Westfälische
Wilhelms-Universität
Münster

Do individual judgment biases diminish in markets? The case of partition-dependence in prediction markets

Inauguraldissertation zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften durch die
Wirtschaftswissenschaftliche Fakultät der
Westfälischen Wilhelms-Universität Münster

vorgelegt von:

Ulrich Sonnemann
aus Hannover

2008

Dekan:	Prof. Dr. Stefan Klein
Erster Berichterstatter:	Prof. Dr. Thomas Langer
Zweiter Berichterstatter:	Prof. Dr. Dr. h.c. Klaus Backhaus
Tag der mündlichen Prüfung:	18. November 2008

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANOVA	Analysis of Variance
ARG	Argentina
AUS	Australia
BRA	Brazil
Caltech	California Institute of Technology
CASSEL	California Social Science Experimental Laboratory
CBoT	Chicago Board of Trade
CDA	Continuous Double Auction
CHI	Chicago Bulls
CIV	Côte d'Ivoire
CLE	Cleveland Cavaliers
CME	Chicago Mercantile Exchange
CPI	Consumer price index
CRC	Costa Rica
CRO	Croatia
CRRA	Constant Relative Risk Aversion
CZE	Czech Republic
DAL	Dallas Mavericks
DAS	Decision Analysis Society
DEN	Denver Nuggets
DET	Detroit Pistons
DJIA	Dow Jones Industrial Average
ECU	Ecuador
ED	Economic derivatives
EMH	Efficient Market Hypothesis
EUT	Expected utility theory
FAQs	Frequently Asked Questions
FIFA	Fédération Internationale de Football Association
GBM	Geometric Brownian Motion
GDP	Gross domestic product

GER	Germany
GHA	Ghana
GM	General Motors
GUI	Graphical User Interface
HICP	Eurozone harmonized index of consumer prices
HSX	Hollywood Stock Exchange
IEM	Iowa Electronic Markets
IJC	Initial jobless claims
IND	Indiana Pacers
IPO	Initial Public Offering
ISE	International Securities Exchange
ISM	Institute for Supply Management
ITA	Italy
ITB	International trade balance
JPN	Japan
LAC	L.A. Clippers
LAL	L.A. Lakers
MAD	Mean Absolute Difference
MAE	Mean Absolute Error
MBA	Master of Business Administration
MEM	Memphis Grizzlies
MIA	Miami Heat
MIL	Milwaukee Bucks
MMS	Money Market Services
MU	Monetary Unit(s)
NASDAQ	National Association of Securities Dealers Automated Quotations
NBA	National Basketball Association
NED	Netherlands
NFL	National Football League
NFP	Non-farm payrolls
NIC	Narrow Interpretation Conjecture
NJN	New Jersey Nets
PDF	Probability Density Function
POL	Poland

RMSE	Root Mean Squared Error
RSX	Retail sales (ex automobiles)
S&P 500	Standard & Poor's 500
SAC	Sacramento Kings
SAS	San Antonio Spurs
SCC	State contingent claim
SCG	Serbia and Montenegro
SEU	Subjective Expected Utility
SPD	Social Democratic Party
SSN	Social Security Number
UCLA	University of California, Los Angeles
Unit PF	Unit portfolio
USA	United States of America
WAS	Washington Wizards
WTA	Willingness to Accept
WTP	Willingness to Pay

List of Symbols

$\beta_{i,P(I_k)}$	Slope coefficient from a linear regression of the average asset prices $P(I_k)$ over trading time for event domain i
$A_{i,t}(I_k)$	Best (lowest) ask quote for interval I_k of team i at time t
$B_{i,t}(I_k)$	Best (highest) bid quote for interval I_k of team i at time t
$\mathrm{E}(\cdot)$	Expected Value
$f_{1/N}$	Ignorance-prior distribution assigning equal probability mass to each interval
$f_{obs}(x)$	Observed probability distribution implied by economic derivatives prices
$f_{true}(x)$	Unobserved unbiased probability distribution
I_k	Interval k
$I_k + I_{k+1}$	Sum of unpacked intervals
$I_k \cup I_{k+1}$	Packed interval (set union of interval I_k and I_{k+1})
λ	Weight on the $1/N$ ignorance prior
$M_{1/N}$	Ignorance prior mean forecast
M_{obs}	Observed mean forecast
M_{true}	Unobserved unbiased mean forecast
N	Sample size or total number of events/intervals
$p_{J,i}(I_k)$	Judged probability for interval I_k of event domain i
$p_{J,i}^{high\_comp}(I_k)$	Judged probability for interval I_k of event domain i (high-competence group)
$p_{J,i}^{low\_comp}(I_k)$	Judged probability for interval I_k of event domain i (low-competence group)
$p_{J,s,i}^{after}(I_x \cup I_{x+1})$	Subject s's judged probability (after trading) for the packed interval $I_x \cup I_{x+1}$ of event domain i
$p_{J,s,i}^{before}(I_x \cup I_{x+1})$	Subject s's judged probability (before trading) for the packed interval $I_x \cup I_{x+1}$ of event domain i
$P_i^*(I_k)$	Equilibrium market price for an asset that corresponds to interval I_k of event domain i

$P_{sports}^{high\_comp}(I_k)$	Equilibrium market price for an asset in the high-competence sports markets
$P_{sports}^{low\_comp}(I_k)$	Equilibrium market price for an asset in the low-competence sports markets
$\overline{PD}_{s,2}(I_k \cup I_{k+1})$	Time-weighted average size of partition-dependence in the time between two messages
$\overline{PD}_{s,3}(I_k \cup I_{k+1})$	Time-weighted average size of partition-dependence in the time between the second message and the end of the trading period
$PD^*_{s,3}(I_k \cup I_{k+1})$	Partition-dependence expressed in mean equilibrium market prices
$PD_{s,t}(I_k \cup I_{k+1})$	Partition-dependence in slot $s = 1, \ldots, 12$ at point $t = 1, 2$

1 Introduction

In its "Advanced information on the Prize in Economic Sciences 2002", the Royal Swedish Academy of Sciences (2002, p. 21) writes:

> *"Experimental evidence indicates that certain psychological phenomena – such as bounded rationality, limited self-interest, and imperfect self-control – are important factors behind a range of market outcomes. […] A challenging task in financial economics is to consider the extent to which the effects of systematic irrationality on asset prices will be weeded out by market arbitrage."*

This passage expressively describes an ongoing debate among researchers whether and under which conditions individual biases transfer to asset market outcomes and therefore contribute to market-wide anomalies (e.g., De Bondt and Thaler (1985, p. 548), Shleifer (2000), Hirshleifer (2001)), or whether those individual errors are eliminated by market forces that are able to produce collectively unbiased market prices (e.g., Rubinstein (2001)). A number of researchers have contributed to this discussion to date by balancing the pros and cons of forces (e.g., arbitrage, learning, and competition) and institutional factors that may or may not cause markets to drive out individual biases (see Kluger and Wyatt (2004)).[1] The growing field of behavioral finance research is characterized by its skepticism about the neoclassical approach of efficient markets in which rational agents derive their decisions by "simply" trading off risk and return of a given investment opportunity. Contrary to the neoclassical perception, the behavioral approach acknowledges the existence and persistence of individual biases in human behavior. "Rational" behavior can be characterized (i) by judgments that are consistent with the laws of statistics and probability like, for instance, Bayes' rule, and (ii) by choices that are in line with expected utility. A bias, by contrast, is defined a systematic inconsistency of judgments and/or choices compared to a normative benchmark, in which "systematic" means that this deviation is predictable in direction. A "market outcome" is defined to be "irrational" if it is incompatible with collective behavior ex-

[1] See Arrow (1982), Blume and Easley (1982), Russell and Thaler (1985), De Long et al. (1990), Camerer and Ho (1999), Shleifer and Vishny (1997), Fama (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), Daniel, Hirshleifer, and Subrahmanyam (2001), Gervais and Odean (2001), and Bossaerts (2002).

pressed by rational individuals (Camerer (1992, p. 239)). By now, there is great demand for empirical work that clarifies which of the individual cognitive biases, if any, influence investor behavior and whether these biases are only marginal phenomena by some sporadic investors or whether they have a significant impact on prices. The present thesis aims to contribute to this question.

Among other biases, recent psychological experiments have documented a rather fundamental cognitive bias: the judged probability distribution of a continuous variable, such as the closing price of a stock index, can depend on the particular intervals into which the variable's possible values are divided, a phenomenon called "partition-dependence." In particular, judged probabilities seem to reflect reliance on a diffuse or "ignorance" prior probability of $1/N$ for each of the N intervals into which the state space is partitioned, plus an adjustment up or down for specific likelihood of each event. This implies that "unpacking" an interval $[I_1, I_2]$ into two separate sub-intervals $[I_1, I_1+x)$ and $[I_1+x, I_2]$ increases the total judged probability from adding the two sub-intervals which is in stark contrast to what normative theory of rational decision making predicts.

However, economists are instinctively skeptical of psychology experiments that use simple abstract or hypothetical events, modest (or no) performance-based financial incentives, and little opportunity for learning, to establish limits on rational behavior. This skepticism has led to several studies testing whether individual biases shown in lab experiments persist under larger incentives, and persist in market trading in both lab and field settings (e.g., Fehr and Tyran (2005), Camerer and Fehr (2006), and DellaVigna (2009)). The present thesis continues in this tradition by testing, for the first time, whether partition-dependence affects prices in experimental and field "prediction markets" for three types of naturally-occurring event domains (financial, sports, and weather outcomes). In prediction markets, people typically trade a set of "all-or-nothing" contingent claims on actual events. A claim pays off if and only if its associated event occurs. The price of the contingent claim is thought to reflect the market's collective probability judgment about the event's likelihood (Manski (2006), and Wolfers and Zitzewitz (2007)).

Common skeptical concerns about the generalizability of psychology experiments to economic settings are addressed in the experiments of this thesis. All three studies include prediction-market bets on actual events, with substantial payoffs linked to choices and outcomes, and trading takes place over time periods lasting from ten minutes to several weeks, which provide a substantial opportunity for learning. Taking advantage of the complementarities of lab and field methods, a lab experiment, a field experiment, and some analysis of naturally-occurring field data will be reported.

A wide range of data was deliberately collected to contribute to the ongoing debate about the importance of psychological phenomena in asset markets. On the one hand, one may argue that the effects reported here are not terribly surprising; one may further argue that it would be highly surprising if the behavioral effects found in the lab did not influence market prices. On the other hand, for example, Levitt and List (2008, p. 910) recently wrote that:

> *"To be empirically relevant, the anomalies that arise so frequently and powerfully in the laboratory must also manifest themselves in naturally occurring settings of interest. Further exploring how markets and market experience influence behavior represents an important line of future inquiry."*

So the type of individual-to-market generalization studied here is "an important line of future inquiry" and it can be assumed that there is still ongoing controversy and more data are therefore welcome.

In addition, these results may interest both psychologists and economists. For psychologists, the magnitude and persistence of these effects in prediction markets says something about their psychological nature: Are they transient slips of the mind that are quickly displaced by effortful thought, and erased by competition? Or do the concrete boundaries of a presented partition persistently influence cognition? For economists, partition-dependence is a particular type of framing effect—the way in which an event is described or "framed" influences its judged likelihood. This phenomenon attacks a basic principle of rationality that Arrow (1982) referred to as *extensionality* and Tversky and Kahneman (1986) called *description invariance*: "The chosen element depends on the opportunity set from which the choice is to be made, independently of how that set is described" (Arrow (1982, p. 6)). In fact, there are already examples of empirical large-scale effects in economic field data that are consistent with a partition-dependent $1/N$ bias, in personal and corporate resource allocation, and racetrack odds.

Some background about both prediction markets and partition-dependence is useful to present before proceeding to the details of what was found.

Prediction markets. The assets (or shares) used in the presented experimental studies are usually referred to as “winner-take-all” contracts (or “all-or-nothing” contracts, or contingent claims) in prediction markets studies (Wolfers and Zitzewitz (2004)). Scientists are interested in prediction markets because the examination of trading activity in these markets is useful to generate important insights into the general effectiveness of markets and trading institutions. Under reasonable assumptions, the prices from prediction market assets can be interpreted as market generated collective probability estimates of the occurrence of these events (Wolfers and Zitzewitz (2007)). The first large-scale prediction markets were created in 1988 at the University of Iowa (Forsythe et al. (1992), and Berg and Rietz (2006)), to trade assets linked to political events. The Iowa markets were inspired by lab evidence that simple abstract experimental markets can aggregate diverse information well (see Plott and Sunder (1982), and Sunder (1995)). Over the years, prices in the Iowa markets have been shown to forecast political outcomes more accurately than many polls about 75% of the time, in hundreds of elections. Around 2001, websites emerged where people can trade contingent claims on a wider range of event domains including political, financial and entertainment events, such as “American Idol” outcomes and when Osama Bin-Laden will be captured (Intrade: http://www.intrade.com, see Wolfers and Zitzewitz (2004)). Firms have also created internal markets to predict outcomes of corporate interest, such as new product sales (Chen and Plott (2002), and Ho and Chen (2007)). In marketing research, prediction markets have been used in the form of preference markets to elicit consumer preferences for new products and product features (Dahan, Soukhoroukova, and Spann (2007), and Dahan et al. (2007)).

Partition-dependence. It is now well established in the psychology literature that limited attention and awareness can lead to reliance on judgmental heuristics, which can deviate systematically from normative logical standards (see Kahneman, Slovic, and Tversky (1982), and Gilovich, Griffin, and Kahneman (2002)).

An early example is “fault-tree bias” (which set the stage for later studies). A fault tree is a hierarchical display with branches showing possible mechanical explanations for an observed system failure (such as an airplane crash or a car failing to start). Increasing levels of detail are shown further down the tree branches. Engineers often create fault trees and assign likelihoods to the branches representing possible causes of a system failure to spot likely flaws and improve designs.

Normatively, when statistically important branches are omitted from a fault tree, the subjective probability assigned to those fault branches should be reassigned to a residual “all other causes” branch. However, experiments showed that the increase in “all other causes” probability is too small when large fault tree branches are omitted, even when the subjects are highly knowledgeable about likely faults. For instance, when experienced auto mechanics were asked to estimate the relative frequency of six categories of reasons why a car might fail to start (battery, starting system, fuel system, ignition system, engine, mischief, all other problems) the mean proportion assigned to “all other problems” was .060. However, in another treatment where three of these categories (starting system, ignition system, mischief) were pruned from the original tree the proportion assigned to “all other problems” was only .215, about half the value of .441 which would have been expected based on the responses given to the unpruned tree (Fischhoff, Slovic, and Lichtenstein (1978, Experiment 6: Experts)).

Four psychological mechanisms have been proposed for bias of this sort: (i) enhanced psychological “availability” of explicitly mentioned categories,[2] resulting in higher judged probability; (ii) “ambiguity” or vagueness about omitted branches;[3] (iii) an ecologically valid belief that the presented fault tree conveys information about likelihood, because omitted branches are likely to be rare (“credibility”);[4] and (iv) a bias toward an ignorance prior of $1/N$ on each of the N identified events.[5] Which mechanism is driving the bias is important because different mechanisms imply different limiting conditions, moderators and “de-biasing” techniques.

Fox and Clemen (2005) distinguish among these explanations by asking participants to judge the likelihood of chance nodes of decision trees that had been partitioned in one of two different ways. In one study expert members of the Decision Analysis Society (DAS; an international association of professional decision analysts and leading scholars of decision analysis) were asked to assess the probabilities that the total number of members of their society would fall into different ranges five years in the future. (The current number was 764.) Fifty-eight of 169 contacted members participated (34%). They were randomly assigned to either a low group or a high group. The low

2 See Tversky and Kahneman (1973), Fischhoff, Slovic, and Lichtenstein (1978), van der Pligt, Eiser, and Speark (1987), Dubé-Rioux and Russo (1988), Russo and Kolzow (1994), and Ofir (2000).

3 See Hirt and Castellan (1988).

4 See Fischhoff, Slovic, and Lichtenstein (1978), and Dubé-Rioux and Russo (1988); see also Sher and McKenzie (2006).

5 See Fox and Clemen (2005).

group was asked to assign likelihoods of membership falling in each of the intervals [0, 400], [401, 600], [601, 800], [801, 1000], [1001+]. The high group was asked the likelihoods for the membership intervals [0, 1000], [1001, 1200], [1201, 1400], [1401, 1600], [1601+]. The judged probability that future membership will reside in the upper interval (>1000) was 10% in the low group, for whom that interval was represented by a single event. The comparable judgment was 35% in the high group, for whom the highest (>1000) interval was partitioned into four separate events.

This example is notable because the subjects are highly expert and responded self-selectively. The first three psychological mechanisms described above ("availability", "ambiguity", and "credibility") cannot explain the difference in judgments between the low and high partition groups. The categories cover all possible ranges of events (i.e., there is no "all other numbers of members" interval), categories are not ambiguous, and participants were told about both possible partitions so that no information was conveyed by a single partition structure. Only the remaining explanation, a natural bias toward an ignorance prior across the categories, can explain the effect. A pure ignorance prior over presented categories would yield $1/N$ judgments of 20% and 80% in the low and high groups, respectively. The actual results of 10% and 35% are partway between those $1/N$ judgments and a common subjective probability for the interval (>1000) that is partition-independent.

Other experiments have shown substantial robustness of partition-dependence to many variables. Partition-dependence was exhibited in controlled learning environment (See, Fox, and Rottenstreich (2006)), using "linguistic priming",[6] in solving probability puzzles, such as a version of the Monty Hall three-door problem (Fox and Levav (2004)), in valuation of insurance policies (Johnson et al. (1993)), and with incentive-compatible payoffs (Fox and Rottenstreich (2003), Fox and Levav (2004), and Fox and Clemen (2005)). Partition-dependence has also been shown when resources, rather than probability, are allocated to categories. Benartzi and Thaler (2001) show bias toward

[6] Linguistic priming means that different wordings of the same event can influence the subjective salience of alternative partitions of the state space. For example, when subjects are asked the likelihood that "tomorrow General Motors' (GM's) stock price will rise by more than any other stock in the DJIA" their judgments are higher than when the event is phrased "Tomorrow in the DJIA, the stock whose price will rise by the greatest amount will be General Motors (GM)." The first phrasing, by mentioning the target event at the outset, frames a partition into the target event and its complement ("GM stock will be the highest" or "GM stock will *not* be the highest") and an ignorance prior of 1/2, whereas the second phrasing, by mentioning the equivalence class at the outset suggests a partition of the state space into 30 stocks and an ignorance prior of 1/30 (Fox and Rottenstreich (2003); see also Fox and Levav (2004)).

$1/N$ in experimental 401(k) investment decisions and in an empirical analysis of retirement savings plan data. Langer and Fox (2005) show partition-dependence more explicitly, in allocations among investments in simple gambles with incentive-compatible payoffs. Other experiments on risky choice are showing that splitting positive-outcome events into sub-events seems to increase preference for those choices (e.g., Humphrey (1996)). Bardolet, Fox, and Lovallo (2007) find in archival data that corporations allocate less capital to divisions when there are more divisions under the same corporate parent, consistent with a $1/N$ bias (see also Scharfstein and Stein (2000)). They also find experimental evidence that experienced managers are statistically biased toward $1/N$ in their hypothetical capital allocations even though they are not aware of their bias. Many studies with many years of data in different countries show a favorite-longshot bias in horse race betting odds: Unlikely winners (longshots) are generally overbet and favorites are underbet, which is consistent with a bias toward $1/N$ probability for every horse (e.g., Wolfers and Zitzewitz (2006)).[7] Fox, Ratner, and Lieb (2005) show $1/N$ bias in experiments allocating money to beneficiaries, consumption to time periods, and choices to menus of options that are grouped by different attributes.[8]

Note that the basic phenomenon underlying partition-dependence is the tendency of concrete, salient categorization to influence attention, thought, and judgment even when the categorization is exogenously imposed and serves no normative purpose. (Many of the earlier experiments, and those reported in the present work, make clear that the partitions are imposed exogeneously by telling subjects that some have one partition and other subjects have a different one; so any effect of alleged information conveyed by the presented partitions should be the same in the two groups.)

This effect of salience based on how possible outcomes are described is ubiquitous in human communication, because complicated ranges of outcomes are rarely categorized naturally. Instead, a discrete categorical structure is invariably *chosen*, or implicitly conveyed by the choice of words. For example, in February 2003, a month before the onset of the Iraq war, U.S. Defense Secretary Donald Rumsfeld said "It is un-

7 However, Plott and Roust (2005) find that in some lab settings with parimutuel betting on abstract events, the favorite-longshot bias is a disequilibrium phenomenon which disappears when some institutional changes are made.

8 Note that in many cases, allocating resources equally could be optimal (e.g., consumption smoothing over time when utility of consumption is time-separable and concave). But the point of these studies is that the way in which categories are unpacked or combined influences allocations, which is not optimal. For instance, allocating consumption equally among *months* will produce (slightly) different results than allocating consumption equally among *weeks*.

knowable how long that conflict [the war in Iraq] will last. It could last six days, six weeks. I doubt six months" (Page (2003)). Rumsfeld's wording invites consideration of a partition of possible war lengths into intervals of [0, 6 days], (6 days, 6 weeks], (6 weeks, 6 months], and (6 months+). Partition-dependence then anchors beliefs around the idea that the war was likely to last 6 weeks. If Rumsfeld had worded his sentence differently (e.g., "six months, six years, or six decades"), it could have established a different partition, with a different public perception of likely outcomes, with different political ramifications.

In most cases, partition-dependence is difficult to entirely expunge because talking about continuous variables often leads to a division of possible outcomes into lumpy natural-language categories. So if partition-dependence is prominent when there are clear historical frequencies lurking behind the cognitive walls of the presented partition, as in the naturally-occurring event domains used in the experiments of the present thesis (financials, sports, and weather), then it might be even more prominent when "unknowable" distributions such as the length of a war are divided into discrete numerical intervals. Tversky and Koehler (1994, p. 565) note that "the need to consider unavailable possibilities […] is perhaps the fundamental problem of probability assessment". They suggest that immunity of judgments to a particular partition is "normatively unassailable but practically unachievable", because people "cannot be expected to think of all relevant conjunctive unpackings or to generate all relevant future scenarios".

Economic theorists have also recognized the importance of cognitive availability that underlies partition-dependence and begun to model it formally. Dekel, Lipman, and Rustichini (1998, p. 524) note that "an unforeseen contingency is not necessarily one the agent *could not* conceive of, just one he *doesn't* think of at the time he makes his choice." Interest in "unforeseen contingencies" is generated by potential applications like simplified employment contracts and the desire for flexibility when it may be difficult to imagine all future events or judge their likelihood (Kreps (1979)). Ahn and Ergin (2007) show that partition-dependence revealed by choices can be modeled by allowing subjective probability to be non-additive in a particular way.[9]

It is important to note that partition-dependence often contributes to a tendency to overweight low-probability events and underweight high-probability events, but is

[9] Their paper contains a particularly thoughtful review of the psychological literature motivating their axiomatization.

not necessarily precisely the same phenomenon. Partition-dependence predicts that in an N-event partition, events with expressed probabilities lower than $1/N$ are likely to be overweighted. Since partitions of the here presented lab and field studies involve 3 or 4 sub-events, overweighting should occur even when probabilities are fairly large (less than .25). Most other evidence of overweighting, however, comes from events with probabilities that are quite low (e.g., below .10). For example, in a comparison of prediction markets forecasts of S&P 500 stock index futures and probabilities inferred from option prices, only probabilities of about .02 seem to be overweighted (Wolfers and Zitzewitz (2004, Table 4)). It is quite possible that details of trading microstructure, some influence of extreme optimism, and other forces are responsible for overweighting of very low probabilities while partition-dependence is more basic and spans a fuller range of probabilities.

To tie together the two strands of prediction markets and partition-dependence, it is important to stress that partition-dependence creates a challenge for the design of prediction markets. In these markets, continuous variables such as political vote shares, movie box office grosses, new product sales, timing of event occurrences, and values of macroeconomic indicators, must be necessarily partitioned into numerical intervals by the market designers. Unlike categorical markets, such as the winner of the Academy Award for Best Film or the winner of the Super Bowl, there is typically no natural partition for events with continuous distributions. If the way in which partitions are constructed matters for actual prediction-market prices, this could affect the quality of the resulting market-wide estimates (as shown in chapter 5). Designers should treat partition-dependence as a cognitive constraint that must be understood and anticipated in the design, much as website screen displays and menu features are chosen to satisfy design goals based on an understanding of visual and motor activity.

A well-designed prediction market will eliminate ambiguity in the definition of events, and can control for the information conveyed by the partition choice if traders know how partitions are created. However, the natural bias toward $1/N$ in the assessment of interval probability cannot necessarily be designed away. Indeed, in naturally-occurring prediction markets, only a single partition for an event is always used. So without experiments in the spirit of those presented here that compare market prices for different partitions of the same state space, there is no way to know for sure whether there is a bias toward $1/N$ in a competitive market setting.

Organization of the thesis and preview of results. Part of the present thesis is included in a recent working paper which is joint work with Colin Camerer, Craig Fox, and Thomas Langer.[10] However, the scope of this thesis goes beyond that of the mentioned working paper as it presents more detailed analyses throughout, intensely explores the potential impact of market participants' self-perceived competence on the obtained results, and studies partition-dependence in a modified lab market design to further examine the robustness of the effect size under conditions which are potentially more "de-biasing" than those in the basis setup.

The next chapter is to describe the fundamental concepts that underlie the subsequent studies and to give a comprehensive literature review. The chapter explains the basic principles of prediction markets and focuses on the theoretical background of these markets. In particular, it is discussed (i) under which circumstances prediction markets are able to produce efficient and accurate outcomes, (ii) whether and under which conditions a market price from "winner-take-all" prediction markets can be interpreted as the aggregate, market-derived probability for the occurrence of a future state of the world, and (iii) which hypotheses have been derived that give reason to assume that individual judgment errors could be eliminated in asset market outcomes. In addition, the chapter identifies partition-dependence as a violation of the description invariance principle in uncertain situations and reviews the explanations that have been proposed for this phenomenon. Support theory is presented as a descriptive model of belief and judgment under uncertainty that is able to capture partition-dependence and related phenomena formally. Furthermore, the chapter describes the basic principles of experimental methods, in particular experimental asset markets and reviews the experimental literature on asset markets with focus on the conditions under which these markets prove able to aggregate and disseminate information well. This is useful to understand the market microstructural design features that were chosen in the experimental studies that follow.

The subsequent three chapters report analyses of three types of data. In chapter 3, short-run experimental markets (two 10-minute trading periods) are described for three naturally-occurring event domains in which one can compare judgments and prices for different partitions of the same numerical interval. These data largely repli-

---

[10] Working paper: Sonnemann, Ulrich, Colin F. Camerer, Craig R. Fox, and Thomas Langer, 2008, Partition-dependent framing effects in lab and field prediction markets, Working paper, University of Muenster, Caltech, UCLA.

cate the magnitude of partition-dependence reported in many psychology experiments (like the canonical Fox-Clemen study mentioned above) and demonstrate the persistence of partition-dependence in short-run prediction market prices. Second order analyses are carried out to further explore the role of competence on the effect size and the trading behavior of differently knowledgeable participants. Beyond this, in an information treatment of this study the experimental design is modified to favor the conditions under which partition-dependence potentially could be reduced. Chapter 4 describes a longer-run experiment conducted on the web, lasting several weeks. Subjects traded assets linked to team victories in the NBA Playoffs and to FIFA World Cup soccer goal scoring. There is noticeable partition-dependence but its magnitude is smaller than in the first lab study. The longer-run study is supplemented by analyzing additional second order competence effects in subjective probability judgments provided by the participants. Chapter 5 describes data from naturally-occurring markets for numerical values of important statistics that macroeconomists follow, called an "economic derivatives market", created by Goldman Sachs and Deutsche Bank. A structural model of these data, which assumes that observed prices mix a $1/N$ ignorance prior belief with other information, allows to back out a de-biased distribution that predicts more accurately actual outcomes than observed prices and suggests some degree of partition-dependence. Chapter 6 summarizes the results and provides a brief outlook for future research.

All three analyses have strengths and weaknesses that are partly compensated for by the other studies (i.e., they are scientific complements). The lab experiments are the easiest to run and replicate, and they provide an initial estimate of whether partition-dependence exists and persists in the short-run and how partition-dependence is influenced by the competence of traders that comprise these markets. However, lab experiments make no statement on whether the effects would persist in the longer run. The field experiments on the NBA Playoffs and soccer World Cup involve a longer span of trading and self-selection of traders who know a lot about the event domains and follow them closely (if not fanatically). The field data on economic derivatives do not compare different partitions for the same variable, as this is possible in the controlled lab environment, but they involve higher implicit stakes and attract more sophisticated (and highly-paid) participants than can ordinarily be used in the lab.

All three studies show evidence of partition-dependence. They provide an example in which a simple observation first discovered in straightforward psychology experiments is robust to market learning opportunities, and varies in the span of trading and the sophistication of traders. These results do not imply that partition-dependence can never be eliminated under any conditions. The results simply establish that some conditions that *might* eliminate partition-dependence do not appear to do so, although in some cases (e.g., the first lab study) evidence of partition-dependence also seems to decrease over time.

2 Fundamental concepts and literature review

2.1 Prediction markets

2.1.1 Basic principles

2.1.1.1 Definition

"Prediction markets"[11] are incentive-based *markets* which are in principal designed to gather and aggregate information on specific uncertain future events, which is then used to obtain *predictions* on these events from resulting market outcomes (Berg and Rietz (2003)).[12] Many of the market microstructure characteristics in prediction markets are similar to those in common financial asset markets. Unlike financial asset markets, though, prediction markets offer a kind of virtual stocks[13] whose terminal payoff depends on the realization of an uncertain future event, such as the outcomes of political elections, corporate sales figures, or sports events (Wolfers and Zitzewitz (2004)). The number of traders who participate in these markets varies from only a few traders to hundreds and thousands of participants in some field prediction markets. Recent studies suggest that it is possible to obtain excellent forecasts of future "states of the world" from trading activity in prediction markets and resulting market prices.[14] Remarkably, these predictions are found to be even more accurate than those derived from opinion polls or expert judgments what is mainly attributed to the fact that prediction market traders back their expectations and opinions by their own money, or to put it differently, traders have to put their money where their mouths are. With respect to Eugene Fama's efficient market hypothesis (EMH; Fama (1970), see also, e.g., Fama (1991)), a prediction market is truly efficient, if the market price is the best predictor of an event's outcome. This, in turn, implies that no combination of available polls, surveys, expert opinions or other information can be used to improve on the market-generated forecasts.

---

[11] Henceforth, the terms "information markets", "decision markets", and "event futures" are used as synonyms. The terms "electronic markets" and "virtual markets" are also sometimes used to refer to prediction markets.

[12] While this definition captures the core aspects of prediction markets, it is worth mentioning that no commonly accepted definition has been established so far (see Tziralis and Tatsiopoulos (2007)).

[13] The terms "assets", "contracts", and "claims" are also commonly used in place of "virtual stocks".

[14] The forecast accuracy of prediction markets will be further discussed in subsection 2.1.2.

It is generally assumed that the market mechanism of competitive asset markets[15] (including prediction markets) is able to efficiently aggregate and disseminate information that is widely dispersed among traders who comprise these markets. According to this suggestion, the full range of individual information is reflected in market outcomes, particularly in the prices of assets.[16] Further, these markets provide (potential) traders with economic incentives for information production and for divulging their true beliefs.

2.1.1.2 Applications

Some of the first significant prediction markets were the Iowa Electronic Markets (IEM) run by the University of Iowa in 1988. These markets aimed to predict the outcome of U.S. presidential elections (Forsythe et al. (1992)). The IEM markets were small-scale, real-money event futures markets, open 24 hours a day, using a continuous, double-auction trading mechanism (CDA).[17] Since then, they have been used to predict presidential electoral results in the U.S., but have also extended the variety of their markets to other elections.[18] This development has also attracted other researchers to set up political stock markets in different countries and to analyze market behavior and forecasts.[19] Besides, commercial providers of prediction markets like Tradesports, Betfair or Intrade have developed. These platforms offer a great diversity of markets and assets that particularly refer to “underlyings“ in the domains of politics, financials, climate and weather, and sports.[20] These markets can be considered as financial and sports betting markets. Some providers of prediction markets use a virtual currency instead of real money, among them Newsfutures, Foresight Exchange or the Hollywood Stock Exchange (HSX).[21] Whereas Newsfutures and Foresight concentrate on political, financial,

---

[15] A market is called “competitive” if it consists of many buyers and many sellers, each of whom individually does not exert much influence on the market price, see Camerer (1992).

[16] Other market outcomes are, e.g., trading volumes or bid/ask quotes.

[17] Traders invest their own funds in the IEM markets; however, the maximum investable amount is typically limited to $500 per trader. Effectively, stakes generally average to less than $50.

[18] See Berg, Forsythe, and Rietz (1996), Berg, Forsythe, and Rietz (1997), Forsythe et al. (1992), Forsythe et al. (1994), and Forsythe, Rietz, and Ross (1999).

[19] Experience gathered from political stock markets in other countries is reported for Germany (Beckmann and Werding (1996), and Kuon (1991)), Canada (Antweiler and Ross (1998), Forsythe et al. (1995), and Forsythe et al. (1998)), Austria (Murauer (1997), and Ortner, Stepan, and Zechner (1995)), and Sweden (Bohm and Sonnegard (1999)).

[20] URLs: http://www.tradesports.com, http://www.betfair.com or http://www.intrade.com.

[21] URLs: http://us.newsfutures.com/home/home.html, http://www.ideosphere.com, http://www.hsx.com.

current events, and sports markets, the HSX offers contracts on the success of movies, movie stars and awards (like opening weekend performances, total box office returns and who will win Oscars). Another appealing application were the "economic derivatives" markets, set up in 2002 by Goldman Sachs and Deutsche Bank which allow institutional traders like banks, pension funds, etc. to take direct positions on the outcome of influential macroeconomic indicator releases such as non-farm payrolls, ISM manufacturing index, U.S. initial jobless claims and retail sales. Furthermore, there have been corporate applications of prediction markets like employees trading on internal trading platforms to predict future printer sales at Hewlett Packard (Chen and Plott (2002)) or adherence to delivery dates at Siemens Austria (Ortner (1997), and Ortner (1998)). Forecasts from these internal prediction markets turned out to be more accurate than target figures derived from the managerial planning process. Berg, Neumann, and Rietz (2008) successfully used prediction markets to forecast the post-IPO market capitalization in Google's auction-based initial public offering. In marketing research, prediction markets have been used in the form of *preference markets* to elicit consumer preferences for new products and product features (Dahan, Soukhoroukova, and Spann (2007), and Dahan et al. (2007)).[22]

2.1.1.3 Types of contracts

An important feature credited to prediction markets is their presumed ability to yield insights into the "market's" expectations about probabilities, means and medians for certain events, and also uncertainty about these parameters. Basically, three different types of assets are used in prediction markets (Wolfers and Zitzewitz (2004)). First, consider so-called "winner-take-all" contracts[23] that offer the prospect of receiving a predefined payoff, say $1, if and only if (iff) the associated (or "underlying") event occurs, and nothing otherwise. For example, one can think of a "winner-take-all" contract referring to the event that "candidate *X* wins the next presidential election" or that "the Dow Jones year-end closing price falls within the interval [6,200; 6,400]". Under reasonable assumptions, the price of a winner-take-all contract can be interpreted as mar-

---

22 Preference markets, however, do not predict actual outcomes, nor are they based on external information. Rather, they aim to elicit expectations of others' preferences, using individual self preferences and insights about others (Dahan, Soukhoroukova, and Spann (2007)).

23 This type of contracts is also named "all-or-nothing" contracts. They are in fact binary options, i.e., options with a discontinuous payoff (see Hull (2009, p. 561)).

ket-derived probability $p(y)$ that an event y will occur.[24] Second, there are so-called "index" contracts, whose payoffs are proportionally linked to the outcome of a specific target value. For instance, there could be a contract paying \$0.01 for each percentage point of votes reached by a party Z in the next elections for the German "Bundestag". Prices of index contracts reflect a market-derived expected value $E[y]$ of outcome y.

Table 2.1: Contract types: Estimating uncertain quantities or probabilities (Wolfers and Zitzewitz (2004, p. 110, Table 1)).

Contract	*Example*	*Details*	*Reveals market expectation of . . .*
Winner-take-all	Event y: Al Gore wins the popular vote.	Contract costs \$$p$. Pays \$1 if and only if event y occurs. Bid according to value of \$$p$.	Probability that event y occurs, $p(y)$.
Index	Contract pays \$1 for every percentage point of the popular vote won by Al Gore.	Contract pays \$$y$.	Mean value of outcome y: $E[y]$.
Spread	Contract pays even money if Gore wins more than y^*% of the popular vote.	Contract costs \$1. Pays \$2 if $y > y^*$. Pays \$0 otherwise. Bid according to the value of y^*.	Median value of y.

Third, there are so-called "spread" betting contracts whose price is fixed at, say \$1, and whose predefined payoff is, for instance, \$2, if a particular event occurs and nothing otherwise.[25] The concrete definition of the associated event, however, is subject to the market participants' beliefs about this event and usually refers to a certain spread, like whether, in the next election, a party will receive a vote share that *exceeds* a certain value *by y'* percent, or like whether a team will win a soccer game *by more than y'* goals. Accordingly, the size of the spread is subject to variations. In an even-money bet like the one described above, the spread reveals the market's expectation about the median outcome, since, assuming risk neutrality, this contract is only a fair gamble if the expected probability of occurrence is one half (i.e., the median). Note that spread betting contracts can also be used to elicit any other percentile of a distribution: if, for ex-

[24] In subsection 2.1.4.3, this issue will be discussed in further detail. Note that the notation in this subsection is adopted from Wolfers and Zitzewitz (2004).

[25] This example represents an even-money bet, in which the owner either doubles (payoff \$2) or loses (payoff \$0) her initial stake (\$1), depending on which state of the world finally occurs.

ample, a contract that costs \$2, pays \$3 if the spread is exceeded and \$0 otherwise, it discloses the market's perception of which spread is representative for the $^2/_3$ percentile of the distribution. Table 2.1 summarizes the most commonly used contract types in today's prediction markets, and what kind of market expectation they reveal.

Additionally, winner-take-all contracts are suited to evaluate the uncertainty about the market-implied expectations. By partitioning the state space of an underlying event into a multiplicity of individual winner-take-all contracts, each of which representing a tiny interval of the whole state space, one can obtain almost the entire probability distribution of the market's expectations. Moreover, offering non-linear index contracts, e.g. two contracts paying "y" and "y^2", respectively, would produce information on $E[y]$ and $E[y^2]$, which, in turn, can be used to calculate the standard deviation or standard error of $E[y]$.

2.1.1.4 Selected market design issues

In their survey article, Wolfers and Zitzewitz (2004) also discuss results on some market design issues. These include the matching mechanism of buyers and sellers, and the use of real-money or virtual-money incentives.

Matching. In principle, one can distinguish between two major matching mechanisms. First, a *continuous double auction* (CDA) is characterized by the fact that bidders place bid quotes stating their maximum willingness to pay for a given asset, and sellers place ask quotes stating their minimum willingness to accept (i.e., their limit prices). A (bilateral) trade between a buyer and a seller occurs whenever a bid quote is at least as high as an existing ask quote, or an ask quote is at least as low as a standing bid quote. Usually, matching of quotes follows a strict price and time priority. In an oral continuous double auction all bidding behavior is publicly observable and can be attributed to individual traders. Contrariwise, in most computerized environments, a trader's identity remains undisclosed and the order book information is limited to the best bid/ask quote and the current market price of each asset. Second, in a *parimutuel* market algorithm, which is commonly used in horse-race betting, "bettors" wager a certain amount to one of the available alternatives. All bets are then collected in a pot and finally (after subtracting potential overhead, commission, or fees) divided among the "winners" (proportionally according to their stakes). Thus, the final payoff (per dollar wagered) is not fixed until the true state of the world is finally known and depends on the amount invested by all other bettors. However, in most cases indicative payoffs are published while the market is open; these indicative payoffs reflect the potential outcome, if no further bets were placed.

Incentives. Since the basic idea of prediction markets is that traders put "*their own money* where their mouths" are, thereby adding weight to their opinions, it seems obvious to use real-money, performance-linked financial incentives to compensate traders. In that case, traders invest their own money with the prospect of earning substantial profits, and the risk of suffering real losses. It seems reasonable to assume that the use of real-money stakes constitutes a strong motivation for the participants to think thoroughly on their decisions and, as the case may be, gather new information, while participants in virtual-money prediction markets may be tempted to pursue risk-enhancing strategies or simply may not be serious enough in their trading actions. However, many attempts have been made so far to use artificial incentives like play-money, non-cash prizes or just a ranking of traders. With play-money one can possibly circumvent problems that may arise in the context of prospect theory (Kahneman and Tversky (1979)), namely participants' individual perception of reference points and a different perception of gains and losses. This circumstance may induce or enforce behavioral biases like overweighing of small probabilities or risk seeking in situations where participants face potential losses, etc. Whereas these biases, in many cases, may consistently describe how people behave in natural-occurring situations, they may negatively affect market outcomes in terms of prediction accuracy.[26] These behavioral biases may potentially be alleviated in a setting in which the participants trade with virtual-money. Up to now, empirical evidence on real- vs. play-money is mixed: In a study of the 2003 NFL football season, Servan-Schreiber et al. (2004) find that play-money markets were just as accurate as real-money markets. Rosenbloom and Notz (2006) compare the forecast accuracy from NewsFuture (play-money) markets to that from TradeSports (real-money) markets and find that for popular sports events both markets predict similarly well. In other domains, though, like financials (DJIA forecasts) or the occurrence of specific political events, real-money markets turn out to perform significantly better. Eventually, it seems that there is still a lack of trustworthy data to understand the extent

[26] For a discussion of how individual biases may affect market outcomes in prediction markets, see section 2.3.3. For a discussion of the predictive accuracy of prediction markets, see the next subsection (2.1.2).

to which (real) money makes predictions more accurate. The question of forecast accuracy in prediction markets is addressed in the following subsection.

2.1.2 Forecast accuracy

Much of the academic interest in prediction markets is attracted by the fact that predictive efficiency and accuracy of these markets seem to be very high according to a significant number of studies so far. Berg et al. (2008) summarize the results from a total of 49 markets of their political IEMs, covering 41 elections in 13 countries. With respect to the absolute market accuracy they find no obvious biases in the market forecasts of index contracts ("vote shares") and, on average, considerable accuracy. With regard to the relative accuracy they find that market predictions are at least as accurate and stable as large-scale poll results in the majority of cases,[27] a few times worse, but often better, albeit only slightly. In terms of the *mean absolute difference* (MAD) between index contract prices and actual vote shares in the week prior to the elections, market-derived forecasts erred by around 1.5 percentage points, while poll data (like the final Gallup poll) were wrong by around 1.9 percentage points. Note that polls and prediction markets use fundamentally different mechanisms to accomplish their forecasts. Polls ask the question: "For which party/candidate would *you* vote, if election were being held today?" to a representative subsample of citizens eligible to vote. Afterwards, these results are sometimes adjusted using statistical methods to correct for potentially biased answers (e.g., strategic response behavior, demand effects, etc.; see Morgan and Stocken (2008)). In contrast, prediction markets implicitly answer the question: "Who will *everyone* vote for on election day?", and traders do self-select to trade on this question, and may do so regardless of their age, eligibility to vote, or other demographic characteristics. Forsythe et al. (1992) analyze the relationship between prediction markets and polls and conclude that polls are obvious information channels for market traders, but market prices do not follow poll results; hence, the availability of poll data influences market prices only marginally, if at all. Wolfers and Leigh (2002) show that prediction markets can even predict well in local elections where no polls are conducted (or where the results remain unpublished).

[27] With regard to the relative accuracy, they restrict the sample to a subset of national elections, in which poll data was available.

Hollywood Stock Exchange contracts have been successful in predicting opening weekend box office returns. Furthermore, its predictions were about as precise as expert panels in foreseeing Oscar winners (Pennock, Nielsen, and Giles (2001)). A comparison of mean economic derivatives predictions and "consensus estimates" based on professional analysts' forecasts shows that both are on a par with respect to their forecast performance, either measured as the correlation with actual outcomes, or in terms of average absolute forecast errors (Gürkaynak and Wolfers (2006)). Besides the large number of promising results in terms of market efficiency and accuracy, Wolfers and Zitzewitz (2008) also report some evidence for prediction markets remaining susceptible to behavioral anomalies like bubbles, the existence of irrational equilibriums and excess volatility which may affect forecast accuracy adversely.

Overall, as Wolfers and Zitzewitz (2004) conclude, prediction markets that tend to produce efficient forecasts of unknown future events, are to a great extent driven by the fact that they provide (i) incentives for a truthful revelation of private information, (ii) incentives for producing and collecting information, and (iii) by the fact that they offer a functioning algorithm to aggregate different opinions in market prices.

2.1.3 Arbitrage strategies

Much research in the field of prediction markets so far has also focused on exploiting possible arbitrage strategies. In this respect, a number of different arbitrage strategies have been analyzed (Wolfers and Zitzewitz (2004)). One of these strategies refers to cross market or cross asset arbitrage, i.e., exploiting price differences across two different markets or related assets at a given point in time. Arbitrage is possible, if the ask price (posted offer to sell an asset) in one market is lower than the bid price (posted offer to buy an asset) for the same asset in another market. Another strategy strives to make use of predictable price patterns in prediction market prices, if they exist. However, it seems as if prediction market prices do *not* systematically follow a specific price path, what makes it a hard task to derive profitable trading strategies based on past price movements. A third strategy aims to take advantage of potential deviations from rationality. In fact, those deviations have been observed in prediction markets. A prominent example is the "favorite-longshot bias" in horse racetrack betting and other sports betting prediction markets. It is reflected in the fact that assumed underdogs (longshots) are likely to be overpriced compared to their actual chance of winning (objective probability), and assumed favorites, by contrast, are likely to be underbet. This phenomenon is well-documented in horse racetrack betting (Thaler and Ziemba (1988)). Empirical evidence finds this bias to become more pronounced, as objective probabilities decrease, so that it is particularly evident in prediction securities that are linked to small-probability events. Thus, the favorite-longshot bias is consistent with the behavioral finding that people tend to overestimate very small probabilities, and to underrate probabilities close to certainty. This can lead to negative abnormal returns from betting on longshots, which is partially offset by returns from betting on favorites, though still negative. In general, it cannot be ruled out that prediction markets on small probability events suffer from cognitive biases. However, it remains unclear to which extent, if at all, these deviations could be exploited by profitable arbitrage strategies.

2.1.4 Theoretical background

2.1.4.1 Hayek hypothesis

This subsection deals with the fundamental question of why markets, and in particular prediction markets, should be able to produce efficient outcomes and accurate predictions of future states of the world.

A particular characteristic of financial markets is the information value of bids, offers, and market prices. Hayek (1945) was one of the first to introduce a view of prices inherently being transmitters of information, besides its role of just being an accounting unit. This characteristic is essential for the ability of markets to aggregate widely dispersed information. Pursuing profit motives is what leads individual agents to reveal their information. According to this idea, prices for scarce resources reflect all relevant information that is necessary to act upon economic principles. Prices do not reflect a good’s attributes *per se*, but signalize its relative valuation among all other goods in an economy. Thus, the pricing system is a means of communication that is used to aggregate and pass on decentralized knowledge. Competitive markets in conjunction with a frictionless price system are assumed to be best suited for (i) efficiently aggregating the knowledge dispersed throughout the economy, (ii) for economizing resources, and (iii) for ideally reacting to changes.

Hayek (1946) emphasizes that a “competitive rational expectations equilibrium” is not a given fact, but rather the outcome of a dynamic competitive process. Based on these considerations, Smith (1982a, p. 167) shaped the expression "Hayek hypothesis" by characterizing it as follows: "Strict privacy together with the trading rules of a market institution are sufficient to produce competitive market outcomes at or near 100% efficiency." Beckmann and Werding (1994) identify three main propositions of the Hayek hypothesis: The first proposition claims *static Pareto efficiency*, i.e., competitive markets result in allocations that leave no room for further gains from trade, and prices that reflect all available information. This proposition underscores the *static* efficiency of competitive market equilibriums and can be found in most textbook versions. The second proposition focuses on the *coordination function* of competitive markets. The coordination function guides the actions of individual agents in a way such that they make their decisions as if they possessed the full set of information available to the economy, while, in fact, each individual is only endowed with a diminutive fraction of this set of information. In rational expectations theory (e.g., Grossman (1981)) the possibility of knowledge transfer between two or more individuals is explicitly considered, which was not yet the case in classical models of the Walrasian auctioneer. Information exchange via the market mechanism starts with an agent's individual belief about the future state of the world. These expectations are then reflected in market outcomes for goods or assets like bid and ask quotes submitted by this individual. These market outcomes, in turn, send signals to all other market participants and stimulate them to reconsider and, if necessary, adjust their own beliefs. Adjusted expectations, again, are incorporated in market variables by (modified) bid/ask quotes and trades that, at the same time, produce new signals to all other traders. This cycle of implied interaction continues until equilibrium is reached, i.e., until no one feels any more prompted to adjust her expectations or to participate in any further trades. With the second proposition it becomes clear that the pricing system contributes to information *revelation* on the one hand, but at the same time facilitates information *aggregation* as well. The third proposition deals with *competition as a discovery process*, i.e., the prospect of receiving financial incentives serves as a strong motivation for individuals to gather and acquire new information and, by this means, to contribute to an efficient market outcome. In this context, price signals offer an informative basis for the direction of where to look for new information. The second, but even more the third proposition, hence, include a *dynamic* view in achieving a state of equilibrium.[28]

---

[28] For a more detailed discussion of the three propositions of the Hayek hypothesis and, in particular,

2.1.4.2 Information paradox and the “no-trade theorem”

In their discussion of the Hayek hypothesis, Beckmann and Werding (1994) highlight the discrepancy of the *static* and the *dynamic* part of Hayek’s hypothesis. Besides, they particularly stress the conflict between prices always reflecting all available information, and prices that motivate people to seek for new information at the same time. This controversy results in a public goods dilemma in which people are free-riding on each other’s information-gathering activities. This dilemma goes back to Grossman (1976), and Grossman and Stiglitz (1980) who introduce (and resolve) their well-known *information paradox*. The paradox refers to Eugene Fama’s efficient market hypothesis whose core message is that a “market in which prices always ‘fully reflect’ available information is called ‘efficient.’” (Fama (1970, p. 383)). By contrast, Grossman (1976), and Grossman and Stiglitz (1980) argue that markets are unable to be efficient, if prices are always expected to fully reflect available information. In a situation of competitive market equilibrium, everyone feels comfortable with his or her current holdings and nobody intends to participate in any further trade. Now, if market prices reacted instantaneously and completely to the arrival of new information, there would be no incentives at all for market participants to seek for new information, because they could not expect to earn any excess returns from doing so. Rather, every investor could expect to yield a return that equals the average return of the market, regardless of whether or how hard they tried to discover new information. This holds particularly true if information seeking is not costless, as the marginal benefit from information search would even be negative. Consequently, no one is any longer willing to accumulate new information. But then, in turn, there is no way of how the pricing mechanism could incorporate news in prices—which finally means that markets cannot be efficient if information is costly. Grossman and Stiglitz (1980) resolve their paradox by introducing increased “noise” which reduces the information content of the price system, but attracts people to become informed, since noisy prices offer potential profits from gathering information. In equilibrium the marginal gains from exploiting noise are exactly offset by the marginal costs of information, so that the informativeness of the pricing system, at the end, is unaffected by noise traders.[29]

their mutual compatibility, see Beckmann and Werding (1994, pp. 7-13).

[29] Another solution was introduced by Kyle (1989) who relaxes the assumption of perfect competition and assumes that each trader takes into account the effect that his or her demand would have on the equilibrium price.

A closely related question is whether information is diverse enough to offer possible trading opportunities, since the motivation of agents to trade is an essential prerequisite of prediction markets to work well. Abstracting away from motivations such like the pure thrill and excitement of trading and/or gambling *per se*, there must be some "disagreement" in traders' beliefs or preferences that motivate two agents to trade "against" one another. This discrepancy may be enforced if traders are overconfident about the quality of their information or their ability to process outcome-related information, or simply if they have priors that substantially differ. It is worth mentioning that, while trading activity in prediction markets may be inspired by traders' heterogeneous prior beliefs, those who set up the markets are typically interested in extracting the information that these traders possess.[30] In this context, it seems important to distinguish between individuals' non-common *prior beliefs* (initial opinions) and different *information* about the event's realization (outcome). Prior beliefs are subjective and thus assumed to be uncorrelated with the outcome. Information, in turn, has an objective nature and is correlated with the outcome. Both components have some influence on traders' heterogeneous beliefs (Ottaviani and Sørensen (2007)). According to Sunder (1995, p. 445), it is an important premise for market prices being able to fulfill an informational role that information is asymmetrically distributed among individuals. As Wolfers and Zitzewitz (2006) point out, the problem of "[H]ow to attract uninformed traders?" is one important question among others about prediction markets which is not finally answered yet, but plays an important role in determining whether current optimism about prediction markets is justified. The need for "uninformed" traders (noise or liquidity traders) seems odd at first glance, but well functioning prediction markets depend to some extent on the existence of uninformed order flow. Otherwise, with only rational traders in the market (i.e., traders with rational expectations), trading activity could collapse according to the "*no-trade theorem*" (Milgrom and Stokey (1982)): In their general model of voluntary trade, they focus on analyzing price adjustments as a reaction to the arrival of new information. Given that the initial allocation of assets was *ex ante* Pareto-optimal among traders,[31] and given that all traders had common priors, then all traders' posterior beliefs would become the same after the market opened, and

[30] As Aumann (1976, p. 1238) notes, "reconciling subjective probabilities makes sense if it is a question of implicitly exchanging information, but not if we are talking about 'innate' differences in priors."

[31] This situation may have been achieved, for example, by a preceding trading round on complete, competitive markets.

no active trading would occur in equilibrium. In particular, obtaining any piece of new private information cannot generate any motivation to trade. The general idea behind their theorem is as follows: if there is common knowledge about the structure of a market (including the way how traders obtain information), then any bid or offer immediately reveals a bidder's private information and will be incorporated into market prices *even before* anyone accepts the initial bid or offer. This, in turn, means that even though a trader may possess private information, she is not able to exploit it profitably, unless there are some noise traders in the market. Wolfers and Zitzewitz (2006) propose a simple modification of the familiar Kyle (1985) model,[32] in which they explain trading activity by the attendance of uninformed outsiders. Demand of outsiders is either driven by hedging or entertainment motives, or by the existence of manipulators who try to influence market prices.

2.1.4.3 Interpretation of prices as market-derived probabilities

The issue whether and under which conditions market prices from a winner-take-all prediction market can be interpreted as market-derived probability for the occurrence of an event deserves somewhat more attention. In Wolfers and Zitzewitz (2004, p. 109), the authors argue that winner-take-all contracts are in fact state-contingent claims, whose prices represent a market-derived estimate of a certain event's probability assuming risk neutral market participants. They further argue that risk neutrality, in most cases, is a reasonable assumption given the small amounts of money usually wagered in these markets. Admittedly though, if agents exhibit some sort of risk-aversion or risk-proneness, probabilities and state prices can differ. Manski (2006) challenges the (lack of) theoretical foundation behind the approach of interpreting prediction market prices as "market probabilities"; in particular, he doubts market prices to coincide with traders' mean belief. Based on a simple model, assuming that traders are *risk-neutral price takers* who have *heterogeneous beliefs* and *limited trading budgets*, he shows that prediction markets are not able to fully aggregate information. According to this malfunction, he casts doubt on the common practice of extracting probabilities

[32] The Kyle (1985) model is characterized by the existence of three different types of traders: a single risk neutral insider, random noise traders, and competitive risk neutral market makers. The insider earns positive profits by exploiting her monopoly power optimally in a dynamic context, where noise trading provides a veil which hides her trading from market makers.

from prediction market prices. In a response to earlier versions of Manski's paper, Wolfers and Zitzewitz (2007) present a formal analysis in which they derive sufficient conditions under which winner-take-all contract prices are consistent with traders' "mean beliefs". The authors identify the *degree of risk aversion* and the *distribution of beliefs* to be the key factors driving trading activity in prediction markets. Their basic model rests upon the following assumptions: heterogeneity in traders' beliefs,[33] beliefs being independent from individual wealth levels, traders who are price-takers and who try to maximize their subjective expected utility (SEU), and the absence of any hedging motives. Under a log utility function, they endogenously model each trader's individual demand for assets. Under this framework, it can directly be derived that equilibrium market prices (i.e., when demand equals supply) do, in fact, equal the mean of beliefs among traders. Let y be a trader's initial wealth, and define q_j a trader j's belief that an event will occur.[34] Then, with log utility, each trader optimizes his or her demand, x, for a winner-take-all contract (paying \$1 if the event occurs and nothing otherwise) that trades at a given price P:

$$\max_x \; SEU_j = q_j \cdot Log\left(y + x_j \cdot (1-P)\right) + (1-q_j) \cdot Log\left(y - x_j \cdot P\right) \tag{2.1}$$

That is, with subjective probability q_j, initial wealth y is increased by x_j times the difference between the payoff of \$1 and the price that was paid for each contract (= profit), and with complementary probability $(1 - q_j)$, initial wealth is reduced by x_j times the price paid (= loss). Differentiating subjective expected utility with respect to the demand for contracts, x, yields:[35]

---

[33] See subsection 2.1.4.2. Possible sources of belief heterogeneity are *uncommon priors* among traders (possibly due to the fact that the event under consideration could not be observed frequently enough in the past), *information asymmetries* (however, no-trade theorems—like those derived by Milgrom and Stokey (1982)—would doubt that any trade occurred under the latter condition), or *behavioral biases* like anchoring and insufficient updating of beliefs (see Wolfers and Zitzewitz (2007)).

[34] Note that the notation corresponds to that used by Wolfers and Zitzewitz (2007).

[35] Equation (2.2) has some quite intuitive implications. As Wolfers and Zitzewitz (2007) point out, demand is positive (negative) when a trader thinks the event is more (less) likely than indicated by the current price ($q > P$, $q < P$, respectively), and demand is zero, when the price equals belief. In addition, demand is decreasing in risk: for prices around 0.5 (indicating greatest possible uncertainty about whether the event occurs or not) the denominator is reaching its maximum of 0.25, thus demand is getting smaller for any given expected return.

$$x_j^* = y \cdot \frac{q_j - P}{P \cdot (1-P)} \tag{2.2}$$

Equilibrium requires supply to equal demand, thus:[36]

$$\int_{-\infty}^{P}\int y \cdot \frac{q-P}{P \cdot (1-P)} dG(y)dF(q) = \int_{P}^{\infty}\int y \cdot \frac{P-q}{P \cdot (1-P)} dG(y)dF(q) \tag{2.3}$$

Recall that—in the simple baseline model—beliefs (q) are assumed to be uncorrelated with individual wealth levels (y), hence (2.3) implies:

$$\frac{y}{P \cdot (1-P)} \int_{-\infty}^{P} (q-P)\, f(q)dq = \frac{y}{P \cdot (1-P)} \int_{P}^{\infty} (P-q)\, f(q)dq \tag{2.4}$$

$$\Rightarrow P^* = \int_{-\infty}^{\infty} q\, f(q)dq = \overline{q} \tag{2.5}$$

Thus, aggregate demand and supply is simultaneously satisfied, if and only if the price of the prediction market contract equals the *average of all traders' beliefs*, i.e. their subjective probability assessments.

Relaxing the assumption of beliefs being independent from individual wealth levels yields the intuitively appealing result that prediction market prices are a *wealth-weighted average of beliefs* among market participants. Based on equation (2.2), and with traders that obey a distribution $F(q, y)$, where $E[q, y] \neq 0$, it follows:

$$\int y \cdot \frac{q-P}{P \cdot (1-P)} dF(q \leq P, y) = \int y \cdot \frac{P-q}{P \cdot (1-P)} dF(q \geq P, y) \tag{2.6}$$

$$\Rightarrow P^* = \int q \cdot \frac{y}{\overline{y}} F(q, y) \tag{2.7}$$

As can be seen from equation (2.7), in this situation the equilibrium market price reflects the mean of traders' beliefs, weighted by each trader's individual wealth in rela-

[36] $G(y)$ and $F(q)$ are to denote the distributions of wealth levels and beliefs, respectively.

tion to the overall wealth that exists in that market. As Wolfers and Zitzewitz (2007) note, the wealth-weighted average of beliefs may even be a better predictor of probability in recurrent or long-run prediction markets, if it is assumed that wealth is accumulated over time by those traders who had accurate beliefs in the past. One of the implications of assuming a log utility function is *demand being linear in beliefs*, i.e. for market participants with a given wealth level, demand rises linearly with their beliefs (cf. equation (2.2)). This implication is central, but not necessary, for the congruence of market prices and the (weighted) mean of beliefs. In a special case, equilibrium prices correspond to mean beliefs, too, if distributions of beliefs and thus demand functions are symmetric. According to Wolfers and Zitzewitz (2007, pp. 6-7), this symmetry condition does not appear to be implausible for beliefs being distributed around $\overline{q} = 0.5$, "as long as traders are not affected by framing issues."

For other cases, in which demand is not linear in beliefs, Wolfers and Zitzewitz (2007) demonstrate that prediction market prices are nonetheless reasonably close to the mean of market participants' beliefs in a great multitude of different models. That is, for most plausible parameter variations of their model, they confirm that all-or-nothing contracts reflect "the central tendency of the distribution of beliefs of traders" very well. All of their model variations result in a monotonic transformation of prediction market prices to the mean of traders' beliefs. Most importantly, they criticize Manski's representation to be just a special case of their model in a sense that it focuses on some misleading assumptions: due to the risk neutrality assumption in Manski's line of argumentation, traders always invest (divest) their *complete* funds (assets), if individual beliefs are greater (smaller) than the current market price. Hence, demand does not evolve endogenously. As a result, the equilibrium market price turns out to reflect a particular *quantile* of the budget-weighted distribution of traders' beliefs rather than the market's "mean belief". To conclude, Wolfers and Zitzewitz (2007) do not deny that prices may be biased in some cases and under some conditions, but provide some convincing evidence for the fact that they typically do produce valuable estimates of average beliefs about an event's probability of occurrence, by this means suggesting that winner-take-all prediction markets, after all, are in fact able to aggregate information accurately.

Gjerstad (2005) also contributes to this discussion and confirms that—for coefficients of relative risk aversion in close proximity to those obtained from empirical evidence and for plausible distributions of beliefs—equilibrium prices turn out to be pretty close to traders' mean beliefs. In his analyses, he considers a two-asset (i.e., two possible states of the world) all-or-nothing prediction market where arbitrage conditions are assumed to hold.[37] In a risk neutral case, where beliefs are symmetrically distributed around its mean, he finds equilibrium prices to be biased toward the price of one half, i.e. the equilibrium price levels off between the true mean of traders' beliefs and an ignorance prior of 0.5. When market participants express *constant relative risk aversion* (CRRA) in their expected utility functions, the aforementioned bias shrinks as the coefficient of relative risk aversion increases (i.e., as people become more risk-averse). Interestingly, there is no bias when the coefficient of relative risk aversion equals one, and the bias is reversed for coefficients of relative risk aversion above one. Within a range of typical estimates for this coefficient (0.5 to 1.5), Gjerstad (2005) demonstrates numerically that market predictions are close to the mean of the distribution of traders' beliefs. However, his finding may contribute in explaining the favorite-longshot bias, where underdogs (having a probability of winning below 0.5) are overbet, and favorites (having a probability of winning above 0.5) are underpriced.

Ottaviani and Sørensen (2007) present a theoretical approach to analyzing the aggregation of information and beliefs in prediction markets, which also contributes to the question of interpreting prediction market prices as probabilities. Assuming risk neutral traders,[38] who are restricted to invest only part of their wealth in the market (which makes this situation different from that in Manski's line of reasoning), they focus on the role of information incorporation into the market price. Their prediction market design is a simple two-state setup with two Arrow-Debreu contingent claims. They demonstrate that the resulting equilibrium price is a generalized average of the participants' posterior beliefs which, in turn, are a combination of their prior heterogeneous beliefs and the information that is disclosed through the trading process. Their main finding is that the market price underreacts to pre-trade information. If there is information that appears to increase the probability of an event, the market price of the corresponding asset will increase, but the price adjustment will be smaller compared to the rational benchmark of a Bayesian probability updating process, so that the effect of information is understated. This finding is driven by a certain wealth effect which arises due to the fact that traders with heterogeneous beliefs take speculative net positions.

---

[37] In the given context, this in particular means that prices for the two assets sum to \$1 (which equals the sure payoff that is received from a bundle consisting of both assets).

[38] Although assuming risk neutral traders in a first step, they show that their results extend to the case of risk-averse participants with decreasing absolute risk aversion (see section 5 of their paper).

Ottaviani and Sørensen's result can be seen as a further account of the favorite-longshot bias, since information that "supports" the favorite is insufficiently incorporated in the market price, so that the contract turns out to be underpriced. Information, in turn, "against" the longshot is also inadequately incorporated in the market price, so that the contract still remains overbet.

2.1.4.4 Individual judgment errors and market outcomes

With respect to the main focus of the present work and the results of prediction markets to date in terms of forecast accuracy and efficiency the questions arise of (i) how individual errors in information processing may affect market outcomes and thus produce false equilibriums, and (ii) whether markets are able to result in efficient equilibriums in spite of traders who are subject to individual biases, thus eliminating individual errors, and if so, under which conditions this may happen.

With reference to the first question, anomalies in markets like bubbles or false equilibriums are often preceded by individual errors in information processing. This, in turn, is closely related with traders' formation of beliefs and expectations in a world of uncertainty. A problem with the formation of expectations may arise if traders, for whatever reason, do not rely on their *own* beliefs and expectations, but rather trust in what they expect others to expect, or to put it in the oft-quoted words of Keynes (1936, p. 156):

> *"Or, to change the metaphor slightly, professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees."*

One possible explanation for situations in which traders may disregard their own beliefs and are geared to the beliefs of others, is that they have reason to assume the market to possess better information on the future state of the world than they selves have. If this is, however, not the case (possibly due to institutional constraints like short selling restrictions that prevent the market prices to reveal the true state), this can result in bubbles or false equilibriums. Market variables then may become self-fulfilling.

With respect to the second question, Duh and Sunder (1986), Camerer (1987), Camerer (1992), Ganguly, Kagel, and Moser (1994), Anderson and Sunder (1995), and Ganguly, Kagel, and Moser (2000), among others, asked whether individual judgment errors transfer to the outcomes of experimental market environments or whether and under which circumstances the effectiveness of market forces is able to correct these biases in equilibrium prices. In general, one may argue that agents in a market setting are more likely to avoid mistakes because they have the prospect of earning an incentive-compatible compensation based on their decisions. In a market, experience can be expected to be high if traders self-select and if there are learning opportunities. Besides, one may argue that market institutions can produce rational market outcomes *despite* the fact that traders who build the market are individually biased. The following discussion aims to explain *how* aggregate rational behavior may arise from mostly irrational people, but also what problems may result. For formal representation, assume the market outcomes to be whatever kind of average of biases in individual judgments and actions (with b_i denoting a measure of bias for the *i*th individual), weighted by each trader's market impact and/or activity w_i, and thus resulting in an average market bias of $\Sigma b_i \cdot w_i$ (cf. here and below Camerer (1987), and Camerer (1992, pp. 240-244); see also Fehr and Tyran (2005)).[39]

Cancellation hypothesis. One of the arguments in favor of why markets could be a rational, undistorted combination of individual opinions is based upon the conjecture that *unsystematic, random errors* of irrational agents may cancel out, thus resulting in rational equilibrium market prices (the "*cancellation hypothesis*"). If individual judgments are randomly distributed around the rational benchmark judgment, then the error terms b_i should be independently distributed around zero and the expected value of individual errors will be zero [$E(b_i) = 0$]. By consequence, the average market error will tend toward zero [$\Sigma b_i \cdot w_i = 0$], too, given that there are enough traders, each of whom is exerting not too much of impact on the market. The cancellation hypothesis, however, should not apply to most psychological biases, since these distortions, by definition,

[39] In the following, the notation is adopted from Camerer (1992).

appear to be *systematic* in the sense that the deviations from a normative benchmark are parallel (i.e., the biases b_i being positively correlated). For that reason, bias in the market average may even cumulate and thus may even be more pronounced than in individual judgments.

Learning hypothesis. Another argument is based on the conjecture that traders who are less rational may implicitly *learn* from rational agents by observing their actions, quotes and trade prices (the "*learning hypothesis*"). In formal notation, this would imply that $b_j \rightarrow b_i$ if $b_i = 0$ and finally would result in $\Sigma b_i \cdot w_i = 0$. In some markets, traders can straightforwardly learn how to avoid mistakes by mimicking the actions of rational traders. In other markets, though, like markets for durable goods in which trades are very rare, learning is unlikely to occur. Alternatively, biased agents may buy advice from rational agents to avoid errors. Consequently, this may lead to reduced market-wide biases. However, learning may be a difficult task. It is often hard to reflect and learn from someone's own experience, but it may be even harder to learn from others. Additionally, biased traders must be aware of their error-proneness, must be able to identify the actions of their rational counterparts, if at all this behavior is observable. Finally, learning requires unbiased traders to be successful, since it seems unlikely that erring traders will imitate actions that turn out to be unsuccessful (even if they are correct with respect to a normative benchmark). With buying advice there is the problem of how much information is sufficient to make wise decisions. Moreover, buying advice suffers to some extent from quality uncertainty, since ex ante the value of information is hard to assess and in many cases, there is not enough history to evaluate the quality of advice.

Evolutionary hypothesis. Furthermore, there is the possibility of biased traders being driven out of the market by *insolvency* that results from permanently trading against unbiased traders at unfavorable prices (the "*evolutionary hypothesis*"). By consequence, only rational traders will "survive", while biased agents go bankrupt over time, resulting in unbiased market outcomes in the long run; this means, if $b_i \neq 0$, then $w_i \rightarrow 0$ over time, and thus $\Sigma b_i \cdot w_i = 0$. However, it may take a very long time before poorly performing traders will be squeezed out of the market. Also, it may not hold under all circumstances that rationality is the best reply to irrational behavior. If there are, for instance, many "noise traders" in the market, then "the market can stay irrational longer than you can stay solvent."[40] Moreover, in an open economy it may happen that exterminated traders are replaced by market participants who are in almost the same manner susceptible to systematic judgment or choice distortions, thereby hindering fully rational traders to prevail.

Smart few hypothesis. Finally, it may be argued that only a few unbiased traders is enough to result in rational market outcomes, as long as these traders have sufficient capital under disposal to exploit arbitrage opportunities[41] (the "*smart few hypothesis*"). After all, unbiased traders must exert more influence on market prices and must be more active than biased traders to generate equilibrium prices that turn out to be less biased than average individual judgments. Or, to put it more formally, an activity-weighted market average will be less biased [$\Sigma b_i \cdot w_i \approx 0$] than the average individual, if more active traders are less error-prone than less active market participants, thus $\sigma^2(b_i)$ and w_i being negatively correlated [$\rho(\sigma^2(b_i), w_i) < 0$]. It is essential for this argument to hold that people who make fewer errors are (i) aware of their relative "unbiasedness", are (ii) endowed with sufficient funds to exploit flawed prices by arbitrage, and are (iii) allowed to trade as much as they want.

Forsythe et al. (1992), Forsythe, Rietz, and Ross (1999), and Oliven and Rietz (2004) deal with the issue of individually biased traders in the Iowa Electronic Markets and their potential impact on market dynamics. They find considerable judgment biases among traders (e.g., over-optimism and wishful thinking with respect to the preferred candidate) that affect trading behavior *on average* and lead to obviously flawed trades.[42] In accordance with the "smart few hypothesis" they argue, though, that this is not necessarily irreconcilable to the very precise estimates generated by these markets, since those traders who usually set bid and ask quotes at the "surface" of the order book generally turn out to be less biased than the average trader. In the same line Wolfers and Zitzewitz (2004) argue that *all* market participants being rational is no necessary condi-

---

[40] This quote is attributed to John Maynard Keynes. Shleifer and Vishny (1997) take the same line in their prominent article titled "The limits of arbitrage".

[41] These arbitrage opportunities refer to a "statistical arbitrage", a strategy which aims to exploit a statistical mispricing of one or more assets based on the expected value of these assets (and which relies on the law of large numbers).

[42] In their political stock markets they identified two behavioral anomalies: the "assimilation-contrast effect" and the "false-consensus effect". The assimilation-contrast effect "states that an individual's preference for an outcome biases his or her interpretation of information about the likelihood of the outcome occurring", the false-consensus effect "states that individuals tend to overestimate the extent to which their views are representative of the population" (Forsythe et al. (1992, p. 1154)). These psychological phenomena can result in a well-pronounced "wishful thinking bias" (Oliven and Rietz (2004, p. 339)).

tion for markets to be efficient, as long as marginal trades are driven by some rational agents. And it is these less biased traders who ensure that market prices, too, are rather unbiased. This idea also came into the literature as the "marginal trader" hypothesis (Forsythe et al. (1992), and Forsythe, Rietz, and Ross (1999)) and is explained in some more detail in the following. The "marginal trader" hypothesis states that even in the presence of some individually biased agents, it is the less error prone traders who mainly account for the best bid and ask quotes and therefore set market prices, which then turn out to be consistent with rational expectations predictions, or to put it differently, it is the marginal traders who drive market prices and thus predictions, not the average traders. Thus, the marginal traders can be thought of as a group of arbitrageurs who trade when market prices do not coincide with rational forecasts of future events. This definition implies that there are at least some agents who do *not* suffer from judgment biases, who realize that they are bias-free while others err, and who are motivated by monetary incentives to exploit the mispricing in a competitive market environment. From a theoretically perspective, it seems unlikely that there is a single trader in such a market who has enough information on objective forecasts and the biasedness of others. However, Forsythe et al. (1992, p. 1157) consider it possible that some traders have "a sufficient intuitive grasp to play the arbitrageur's role successfully." To operationalize their hypothesis, they define a set of trading activity-related criteria to identify marginal traders empirically in their markets. In detail, they classify those traders as marginal traders who frequently submit limit orders at quotes which are close to the current market prices. Non-marginal traders, by contrast, are traders who are quite inactive and place market orders, or limit orders far away from the current price level.[43] Technically speaking, a trader is considered as acting like a marginal trader if, at the end of a trading day, she either has got a pending bid or ask quote in the order book that deviates from the last market price of that day not more than two cents, or if she placed at least one limit order that was accepted by another trader on the same day. Finally, a trader is classified as marginal trader, if she meets at least one of the two criterions for at least three days during the relevant trading period (which, in that case, was 21 days). The authors themselves admit that their definition of a marginal trader is somewhat arbitrary, on

[43] A limit order is an order for which the initiator specifies an upper or lower reservation price at which she is just about ready to buy or sell the asset. A market order, on the contrary, is an order for which the initiator seeks immediate execution at whatever best price is currently available, thus accepting any given price.

which can be agreed. Nonetheless, they find their results to be essentially the same when applying slightly modified definitions of a marginal trader. Ultimately, they identify a subset of 22 out of 192 participants (11.5%) to fulfill the marginal trader condition in their market. The main difference between marginal and non-marginal traders is that the former invested more money in the market than the latter (more than twice as much in terms of total investments); they also traded more shares and accessed the market more often than non-marginal traders. The authors find this result to be supportive for their hypothesis that judgment bias is generally reduced as the investment exposure increases. They also find that marginal traders yielded higher returns, were able to learn in later stages of the market and, as opposed to the remaining traders, did not show any clues of judgment bias in their transactions. This result is mainly attributed to the facts that (i) marginal traders seemed able to differentiate between "news" and "no news", i.e., they could better assess the relevance of new information, and (ii) that they did not suffer from "wish-fulfillment". Summarizing their conclusions, one can say that market forces and arbitrage motives are well-suited to produce efficient (i.e. unbiased) market outcomes, since market prices are not set by the *average* investor, but rather by marginal traders who tend to be less prone to behavioral judgment biases, thus supporting the Hayek hypothesis. Oliven and Rietz (2004) also contribute to the "marginal trader" discussion in their analysis of Iowa Electronic Markets data. They find that marginal traders constitute a more rational subset of traders who (i) self-select into the role of market makers (that is, they set quotes on both sides of the order book) and who are (ii) able to produce superior estimates. Other traders, who are prone to behavioral biases, in contrast, act as price-takers and thus provide the market makers with liquidity and profits. Instructively, the authors point out that behavioral biases in fact seem to transfer from individual choice situations to the market setting which can be seen from the fact that systematic mistakes by biased agents can be observed in their data. Yet, these errors do *not necessarily* impact prices. In this vein they conclude (Oliven and Rietz (2004, p. 350)) "that efficient forecasts from prediction markets do not rely on a representative sample of traders."

The "marginal trader" hypothesis, though, has been challenged by some researchers. In Brüggelambert (1999) and Brüggelambert (2004), for instance, the author tests for the existence of marginal traders in a number of virtual election markets in Germany during the period 1990–1998, applying essentially the same criteria as did Forsythe et al. (1992). In fact, it was possible to identify people who met the marginal trader condition in his dataset. While most of the characteristics of a marginal trader (e.g. higher stakes, higher trading frequency, etc.) also applied to Brüggelambert's marginal traders, they did not yield significantly higher returns than the rest of the population. Rather, it was the "political insider" who earned abnormal returns; the term "political insider" aims to describe participants who were exceptionally interested and more experienced in political topics than others, thereby profiting from lower information costs. Hence, the author criticizes the identification of marginal traders to be somewhat arbitrary.[44] In the end, abnormal returns had not been earned by marginal traders exploiting other traders' behavioral biases, but rather by political insiders taking advantage of existing information asymmetries. From his point of view, it is the institutional framework that is mostly relevant to the degree of market efficiency achievable by "rational" traders. Hence, it largely depends on the institutional framework and the cost of information production, whether a political insider is able to act rationally, thereby accounting for efficient market prices. With respect to the analyzed election markets in Germany, he finds that even political insiders are not in a position to make market-derived forecasts more accurate than opinion polls.

In his widely-known book "The wisdom of crowds", Surowiecki (2005, pp. 288-290) doubts, although intuitively appealing, the existence of a single or a small group of marginal investor(s). He argues that in most markets, particularly in the context of large financial markets, a small group of rational investors is unlikely to have sufficient market impact to stem against the erring "crowds". According to the author, it is more likely that the mass of people, if they disagree, prevail over a small number of marginal traders acting rationally. He emphasizes the applicability of this argument to markets such as the IEM, in which the maximum possible amount of money invested by a single trader is restricted to a $500 stake. Besides, he criticizes the approach of *ex post* classifying some traders as the smart ones to be misleading (unless these traders can be assumed to be more rational in general), since, by chance, there will almost always be some agents who are right, and others who are not. Being correct only by chance, however, is not consistent with the concept of the marginal investor. Further, if marginal traders were responsible for prices to be unbiased, then another puzzle would arise in pari-mutuel markets[45] (which turn out to be quite accurate in practice as well): under a

[44] Even more controversially, he criticizes the marginal trader-identifying characteristics in some ways tautological, since, by definition, it is the marginal traders who are the most active investors in a market.
[45] For a description of a pari-mutuel market mechanism, see subsection 2.1.1.4.

pari-mutuel market regime, there is no explicit marginal trader who could influence prices based on directly observable trades. Therefore, Surowiecki does not attribute the accuracy of predictions from these markets to the "wisdom" of a small group of marginal traders, but rather to the "collective intelligence of the traders as a whole"; hence, he doubts most of the agents to underlie behavioral biases, which otherwise would transfer to the market outcomes. Institutional constraints like short selling restrictions may add to the conjecture that just a few participants are not enough to correct biased prices. Obviously though, external validity of this line of reasoning is subject to the relative strength of marginal traders and non-marginal traders, their respective endowments as well as the characteristics of the market design in a given market.

To summarize, the above mentioned arguments provide an informative basis on possible ways in which market aggregation may mitigate individual errors. However, as was figured out, it depends on the concrete market environment whether rational expectations equilibriums can be reached—in particular, institutional constraints like short selling or budget restrictions, the auction mechanism, etc. may hinder rational traders from incorporating their unbiased views into market prices.

2.2 The description invariance principle and partition-dependence

2.2.1 Preliminaries

Normative theory of rational decision making under risk or uncertainty is built on a logical deduction of *expected utility theory* (EUT; von Neumann and Morgenstern (1944)) and *subjective expected utility theory* (SEU; Savage (1954)).[46] In either case, decisions that are considered as rational from a normative perspective are derived by maximizing (subjective) expected utility.[47] While expected utility theory applies to de-

---

[46] Note that most decision problems can be represented in a decision matrix that is composed of "acts", "states", "probabilities", and "consequences". An act $a_i \in A$ is an action or option that the decision maker may choose to take. All of the available acts are linked to a set of certain consequences $x_{ij} \in X$. Each set indicates—for a given act—which outcome will occur in different future states of the world $s_j \in S$. Subsets of the state space S are called events. Each event arrives with (objective or subjective) probability p_j, $\Sigma\, p_j = 1$. Thus, for act a_i, state s_j, and consequence x_{ij}, we obtain $a_i(s_j) = x_{ij}$ (see Fox and See (2003, pp. 273-274); notation borrowed from this source).

[47] According to the subjective expected utility theory, a decision maker maximizes the following expression: $SEU(a_i) = \sum_{j=1}^{n} u(x_{ij}) \cdot p(s_j)$, with $u(x_{ij})$ = utility of outcome x_{ij}, and $p(s_j)$ = subjective probability that state s_j will obtain (Fox and See (2003, p. 274)).

cisions under *risk*, subjective expected utility theory applies to decisions under *uncertainty*. To point out the difference between risky and uncertain decision problems, it can be said that, in classical decision theory, “risk” refers to a situation in which *objective* probabilities for all future states of nature are available, whereas “uncertainty” refers to a situation in which the decision maker herself has to think of *subjective* probabilities, because objective probabilities are not available.[48] Thus, in uncertain situations decision makers have to weigh the perceived attractiveness of each potential outcome (utility) by its perceived likelihood (subjective probabilities or beliefs). Both expected utility theory and subjective expected utility theory owe their great significance to the fact that they rely on some intuitively appealing axioms. This means, accepting a set of reasonable-seeming propositions, a decision maker inevitably *has* to maximize his or her (subjective) expected utility to behave in accordance with rational decision theory.

In expected utility theory, these axioms are: cancellation, transitivity, dominance, and (description) invariance.[49] *Cancellation*[50] refers to the possibility of eliminating any state of the world from further consideration that yields the same outcome for all available actions. Thus, the choice between different alternatives only depends on states that yield different outcomes for different actions. The cancellation axiom is central to the representation of preference between prospects as the maximization of expected utility. *Transitivity* of preferences refers to the possibility of representing preference by an ordinal utility scale u in a sense that, if an action a_i is preferred to another action a_j ($a_i \succ a_j$) and action a_j is preferred to a_k ($a_j \succ a_k$), then action a_i is also preferred to a_k ($a_i \succ a_k$), thus the utility of each option does not depend on the outcomes of any other option, but can rather be determined on its own. *Dominance* simply asserts that—among two alternatives—an alternative that is preferable in one state of the world and at least as preferable in all other states is under all circumstances preferred (i.e. dominant) to the other alternative. Finally, and most important in the context of the present work, *(description) invariance* assumes that different characterizations of the same

48 “Ambiguity”, in contrast, is used to describe a situation in which the decision maker is not even able to determine subjective probabilities.

49 See, e.g., Tversky and Kahneman (1986), or Tversky (1996). After the axiomatic foundation of expected utility theory had first been introduced by von Neumann and Morgenstern, these axioms were further developed by numerous scientists. One of these axiomatic systems includes: complete ordering (i.e., completeness and transitivity), continuity, and independence (see Eisenführ and Weber (2003, pp. 211-217) whose classification is based on Herstein and Milnor (1953)). Although most of these axiomatic systems slightly differ, almost all of them lead to the same criterion of expected utility.

50 The cancellation axiom is often also referred to as the “independence axiom”.

choice problem should not at all influence subjective probability judgments nor preferences of the decision maker. In other words, the preference between different options should not depend on their respective *description*. Hence, different representations of a certain decision problem, which in effect are logically equivalent, should result in the same choices regardless of how (un-)consciously a decision maker reflects on the given decision problem.

Subjective expected utility theory is built on a similar set of axioms. One of these axioms—called the "sure-thing principle"—is central to the applicability of subjective expected utility. Basically, the "sure-thing principle" is congruent with the cancellation/independence axiom of expected utility theory and states that preference between two actions should not be influenced by a common consequence which—in a certain state—is exactly the same for the two actions.[51] Although normatively appealing, the *descriptive validity* of the cancellation/independence axiom and the "sure-thing principle" have early been challenged by some prominent counterexamples (e.g., Allais (1953), and Ellsberg (1961)) and transitivity was found to be questionable in many decision problems, too (e.g., Bell (1982), Fishburn (1982), Fishburn (1984), and Loomes and Sugden (1982)).[52] Dominance and (description) invariance, in turn, had been generally accepted for a long time and were considered to be essential for (subjective) expected utility theory. However, it turned out that even the latter two principles do not provide an accurate *descriptive* model of choice behavior in most cases.

2.2.2 Violations of the description invariance principle

2.2.2.1 Violations in risky situations

An early example that demonstrates a violation of the invariance principle stems from McNeil et al. (1982). The authors studied the preference between different medical treatments of lung cancer. Respondents received statistical information on two methods of treatments (radiation therapy and surgery), but statistical information was presented either in terms of mortality rates (mortality frame) or in terms of survival rates (survival

---

[51] To be more precise, the "sure-thing principle" is *implied* by expected utility theory (see Fox and See (2003, pp. 277-278)).

[52] By contrast, Birnbaum and Schmidt (2008) who analyze violations of transitivity in individual choice experiments using an error model find evidence in favor of the hypothesis that their participants exhibited a transitive preference order in repeated presentations of the same choices.

frame). From a logical point of view, both descriptions included exactly the same information and normatively, the preference for one or the other treatment should be the same irrespective of how the problem was *framed*. However, there was a considerable difference in the number of respondents who favored the radiation therapy over surgery depending on which of the two descriptions they faced (only 18% of participants favored radiation therapy in the survival frame against 44% in the mortality frame).[53] Obviously, the way in which the statistical information was described or "framed" influenced respondents' preferences. In this case the difference in answers has to be unambiguously attributed to the different framings of statistical information, suggesting that subjects did not transfer the given representation of the problem to whatever standard canonical representation. If there was such a *canonical* or *natural* representation of a decision problem (like a cumulative probability distribution of a random variable), the invariance principle is assumed to be more likely to hold than if such a description is not available. But most real-life decisions neither do entail such a canonical description of the problem, nor does this description intuitively come to a decision maker's mind. For that reason, these decisions remain susceptible to inherent violations of (description) invariance.

2.2.2.2 Violations in uncertain situations

Whereas the just mentioned example included preferences based on objective probabilities (through the presentation of statistical information), the invariance principle may also be hurt in *decision problems under uncertainty*, in particular when it comes to individually judged probabilities. In real-life situations, people are—implicitly or explicitly—required to make subjective probability judgments day in, day out. This includes, for example, probability assessments of the outcome of a trial, the results of a surgery, the occurrence of an insured event, stock market up- or downturns, the success of a business venture, the winner of the next basketball game, or whether to take along an umbrella in the morning, to name just a few. While Savage denied direct judgments of likelihood in favor of measuring subjective probability and utility simultaneously from observed preferences, some psychologists lend credence to direct expressions of

[53] The subject pool initially consisted of a large group of clinic patients. However, results were generally confirmed when asking experienced physicians or statistically sophisticated business students.

subjective probabilities (beliefs) and trust in using them to predict willingness to act under uncertainty (see, e.g., Spetzler and Staël Von Holstein (1975), Kahneman, Slovic, and Tversky (1982), and Clemen and Reilly (2001, in particular Chapter 8)).

Fischhoff, Slovic, and Lichtenstein (1978) document the sensitivity of judged probabilities to the specific representation of the decision problem in the context of *fault trees*. Fault tree analysis is used to identify the concrete cause(s) of a technical error. A fault tree is a graphical representation of conceivable error sources for a given (typically technical) problem with more detailed information in each "branch" of the "tree". Although fault trees appear in many different formats, they all have in common that possible error sources are organized into functional categories. For instance, a fault tree may be used by car mechanics to detect why a car could fail to start. Possible reasons, i.e. branches of the tree, are: "battery charge insufficient", "starting system defective", "fuel system defective", "ignition system defective", "other engine problems", "mischievous acts or vandalism"—each of them providing further sub-categorization—, and a residual category capturing "all other problems". Fault trees do not only serve as a kind of "checklist" for possible causes of defect, but are also used to estimate failure rates for complex systems in situations in which historical data are unavailable. In those cases, experts assign probabilities to each of the available failure categories; combining these probabilities then yields an overall failure rate. However, one might suspect that the resulting overall failure probability depends to a great extent on (i) how the different categories are presented and arranged, on (ii) whether relevant categories are omitted and thus implicitly subsumed under the "all other problems" category, and also on (iii) whether categories are implicitly combined with other categories. The grouping of categories, in some cases, is subject to some arbitrariness. This suspicion motivated Fischhoff, Slovic, and Lichtenstein (1978) to conduct a series of fault tree experiments that addressed these "arbitrary" aspects systematically. Different fault trees of the above mentioned car starting failure problem were presented to both professional automobile mechanics and laypeople, asking them to judge the probability of each of the listed error causes to be the actual source of the technical problem. The basic design of an (unpruned) fault tree with possible reasons why a car might fail to start is shown in Figure 2.1.

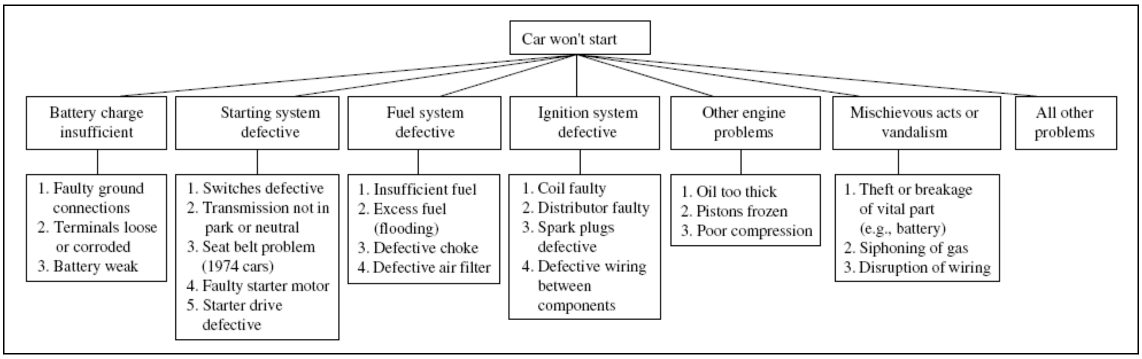


Figure 2.1: (Unpruned) fault tree with possible reasons why a car might fail to start as used by Fischhoff, Slovic, and Lichtenstein (1978); taken from Fox and Clemen (2005, p. 1418, Figure 1).

One of these experiments (Experiment 1) included two treatments, both of which provided a *different number* of failure categories in a fault tree: one treatment included six explicitly mentioned failure causes plus an additional "all other problems" category (i.e., a total of *seven* branches, see Figure 2.1); in the other treatment three of the categories (e.g., "fuel system defective" and two other categories) were collapsed and implicitly moved to the "all other problems" category. "Pruning" the fault tree like this resulted in three explicitly mentioned failure causes plus the residual category (i.e., a total of *four* branches).[54] It turned out that, on average, judged probabilities in the different treatments did not match, or in other words, judged probabilities were inconsistent with respect to the likelihoods assigned to a certain failure cause. Concretely, judged probability of the "all other problems" category in the pruned tree treatment did not fully absorb the probabilities of the pruned branches (as measured by the judgments of those subjects who faced the unpruned tree).[55] Rather, it seemed as if the difference in probability assessment for the "all other problems" branch was distributed among the explicitly mentioned categories, thereby significantly increasing judged probabilities of the remaining components. Higher expertise based on a self-rating of subjects did not alter the results. This pattern has subsequently come into the literature as the "pruning bias" (see, e.g., Russo and Kolzow (1994)).

[54] The pruned tree treatment was further divided into two sub-treatments which differed in that each of them regrouped three different failure causes (out of a total of six causes) to the residual category. Subjects were not informed about the number of categories included in the other treatment.

[55] Judged proportion of the "all other problems" category increased from .078 (unpruned treatment) to .140 (pruned tree 1) and .227 (pruned tree 2), respectively. Assuming insensitive answers though, one would have expected proportions of .468 (pruned tree 1) and .611 (pruned tree 2), respectively, depending on which branches had been removed from the tree. In a variation of this experiment, the authors find that focusing subjects' attention to the question which failure causes are missing in the tree only partially improves their awareness.

Another experiment (Experiment 5) presented by Fischhoff, Slovic, and Lichtenstein (1978) examined possible effects from *splitting* and *fusing* certain branches of the "car won't start" fault tree. In this context, splitting and fusing means that the *same amount of information* is presented either as two independent branches of the fault tree where it was originally presented as a single branch (splitting), or information is presented as a single branch where it was originally presented in two different branches (fusing), always keeping the overall amount of information constant. For example, in one of the treatments the "ignition system defective" branch was split into "ignition system defective" (including "coil faulty" and "spark plugs defective") and "distribution system defective" (including "distributor faulty" and "defective wiring between components", cf. Figure 2.1). This kind of splitting resulted in a fault tree containing seven branches plus the residual category (i.e., a total of *eight* branches). In another treatment the "starting system defective" and "ignition system defective" branches were fused to build a single branch, thus resulting in a fault tree containing only five branches plus the "all other problems" category (i.e., a total of *six* branches). When subjects were asked to judge proportions of failure causes, it turned out that a set of problems was found to be more important when it was presented as two branches than when it was presented as one, compared to the mean data from the unpruned tree judgments in the abovementioned experiment (Experiment 1). On the other hand, a failure cause was perceived to be less important when it was listed as a single branch than when it was listed as two separate branches, compared to the unpruned tree treatment in the experiment described above. The increase in allocated proportions, comparing the sum of split branches to the comparable fused branch, amounted to about one-third.[56] Thus, the more "pieces" into which a group of failure pathways is arranged, the more important this group overall appears.

Evidence from psychological experiments like those conducted by Fischhoff, Slovic, and Lichtenstein (1978) suggests that the particular choice of categories into which a decision problem is divided (i.e., "framed" or "partitioned") may substantially affect subjective probability assessments over categories. It appears that human limitations of memory and information processing capacity can lead to subjective probabilities that are poorly calibrated or internally inconsistent. It is important to stress that

[56] Note, though, that the calculation of this difference is based on fault trees offering a different overall number of branches (six versus eight branches compared to the seven-branch unpruned tree from the aforementioned experiment).

these patterns seem to be sort of a more general phenomenon that does not only apply to *qualitative* descriptions of discrete categories of events, but also applies to *quantitative* accounts of continuous variables which are typically represented by a number of exclusive, but collectively exhaustive intervals of a dimensional state space. Fox and Clemen (2005, Study 5), for instance, asked members of the Decision Analysis Society (DAS, a group of academics and practitioners who study decision analysis) to judge probabilities concerning the total number of members of their society five years in the future. Members, who self-selected to take part in this study, were randomly grouped into one out of two treatments, a low partition group (that included the following intervals: [0, 400 members], [401, 600], [601, 800], [801, 1000], [1001+]), and a high partition group (that included the following intervals: [0, 1000 members], [1001, 1200], [1201, 1400], [1401, 1600], [1601+]). As a result, the median judged probability that the future membership number will fall into the upper interval (>1000) was 10% in the low partition group, in which this interval was represented by one out of five intervals, and the sum of judgments was 35% in the high partition group, in which this interval was represented by four out of five intervals. The fact that different representations of the state space (like the low partition and the high partition) lead to differences in judged probability for a given event ("more than thousand members five years in the future") is inconsistent with normative theories of rational judgment and decision making. This phenomenon usually is referred to as "*partition-dependence*".[57] Partition-dependence means that judged probabilities vary *systematically* with the number and the way a given state space is (deliberately or unconsciously) *partitioned* into single events. Generally speaking, partition-dependence can be conceived a particular manifestation of a framing effect, as this is the case, for instance, with the reference point in prospect theory that divides possible outcomes into gains and losses (Kahneman and Tversky (1979), Tversky and Kahneman (1992), and Tversky and Kahneman (1986)). As with framing effects, people seem to accept the partition at hand and seem to be somehow insensitive to the question of *how* and *why* it came to the presented partition. Accordingly, people seem to "anchor" on the number and the concrete boundaries of events. Partition-dependence thus is in stark contrast to what rational decision theory assumes in terms of "description invariance". Whereas this subsection was mainly to reveal the

[57] Robustness of partition-dependence in psychological studies was further documented by Fox and Rottenstreich (2003), Fox and Levav (2004), Fox, Bardolet, and Lieb (2005), Fox and Clemen (2005), Fox, Ratner, and Lieb (2005), and See, Fox, and Rottenstreich (2006).

basic phenomenon, the next subsection addresses possible psychological accounts that have been presented to explain *why* it comes to violations of description invariance and partition-dependence.

2.2.3 Explanations for violations of description invariance

While Fischhoff, Slovic, and Lichtenstein (1978) identified the "*availability*" heuristic to be the most suitable explanation for the demonstrated phenomena, further assertions have been developed in a series of replications and extensions of the fault tree experiments, among them "*ambiguity*" and "*credibility*" (see, e.g., Fox and Clemen (2005, p. 1419) for a brief survey). Fox and Rottenstreich (2003), and Fox and Clemen (2005) add the "*anchoring and insufficient adjustment*" explanation based on an "*ignorance prior*" of judged frequencies or probabilities. In the following, these four explanations will be discussed in some further detail.

Availability. The "availability" heuristic (Tversky and Kahneman (1973)) may result in a cognitive bias that is reflected in the fact that judged probabilities of a category or an event depend on how easily a decision maker can think of certain instances or scenarios that match the category or event under consideration. In cases in which people are asked to judge probabilities, availability refers to the ease with which a person can retrieve occurrences or construct scenarios that correspond to the category or event at hand. By consequence, explicitly mentioning a particular category in the fault tree experiments increases its salience to the respondent, so that this category becomes mentally more "available" and thus increases a person's judgment of probability. Accordingly, "availability" provides a mechanism by which occurrences with higher salience may appear more likely than they actually are in terms of objective probability. Besides Fischhoff, Slovic, and Lichtenstein (1978), the availability mechanism has been proposed by a number of researchers, including van der Pligt, Eiser, and Speark (1987), Dubé-Rioux and Russo (1988), Russo and Kolzow (1994), and Ofir (2000).

Ambiguity. Another explanation of the pruning bias in fault trees is "ambiguity" with respect to the mapping of specific problems to existing categories as proposed by Hirt and Castellan (1988). Consider, for example, the category "battery charge insufficient" which, in the original Fischhoff, Slovic, and Lichtenstein (1978) fault tree, explicitly includes "faulty ground connections", "terminals loose or corroded", and "battery weak". If this category is pruned from the fault tree, a respondent who faces the pruned version of the tree and who thinks of specific causes, like “faulty ground connections” or “loose connection to alternator,” could just as well assign these causes to the branch labeled “ignition system defective” (including the more specific constituent “defective wiring between components”) or to the residual “all other problems” branch. This ambiguity about the affiliation of specific failure causes to available categories may even be greater, the less knowledgeable a respondent feels within a particular domain. Most notably, though, this could result in judged probabilities of omitted failure causes being distributed amongst some of the available categories instead of being fully subsumed under the “all other problems” category as hypothesized. This can explain why the residual catchall category does not fully capture judged probability of pruned categories.

Credibility. A “credibility” mechanism may also contribute to explaining the pruning bias (Dubé-Rioux and Russo (1988), and Fischhoff, Slovic, and Lichtenstein (1978)). It is assumed that respondents in fault tree experiments rely to some extent on the fault tree itself, believing that the way in which the tree was designed by the experimenter conveys some information about the most relevant failure causes that come into question. If people effectively believe that a particular composition of the presented fault tree contains some information on important failure causes, simply because these causes were included in the fault tree, then they may suppose that each of the explicitly mentioned causes in effect has some significant likelihood of occurrence, while the “all other problems” category is perceived to just cover negligible problems with small probability of occurrence. In this context, the credibility mechanism is closely related to a “demand effect”,[58] in which participants consider the experiment to be an implicit conversation with the experimenter. This includes subjects expecting this kind of implicit dialogue to follow some general conversational norms. In particular, it implies the anticipation that any message transferred from the experimenter to the participants contains some *meaningful* information. Transferred to the fault tree experiment this means that participants expect the experimenter to having designed the fault tree such that each of the explicitly mentioned branches represents a meaningful and non-trivial failure cause. Hence, respondents may feel constrained to assign some significant probability to explicitly listed branches, whereas the “all other problems” category may seem to be of

[58] In general, a “demand effect” in experimental economics describes a phenomenon in which subjects try to behave in a way they believe the experimenter expects or wants them to act (or the opposite), thus the demand effect refers to subjects who attempt to help (or hinder) the experimenter.

little importance. Russo and Kolzow (1994) experimented with the credibility of fault trees by varying and emphasizing the (fictitious) source of the presented tree using different cover stories. However, they did not find a noteworthy effect on the size of the pruning bias.

Against the background of previous research results, Fox and Clemen (2005, p. 1419) argue that, among the three aforementioned accounts ("availability", "ambiguity", and "credibility"), the availability heuristic is the most persuasive explanation for the observed pruning bias and related phenomena. In some instances, though, even the explanatory power of the availability account seems to be overstrained. As an example, they cite one of the Fischhoff, Slovic, and Lichtenstein (1978) variations of the fault tree experiments. As stated above, in their "splitting and fusing branches" version of the experiment the authors split some of the categories into two separate groups, and fused some of the categories to form a single group. However, the overall amount of information was kept constant in all cases. This means that the three to five constituent problems subsumed under each of the main categories (e.g., "faulty ground connections", "terminals loose or corroded", and "battery weak" for the main category "battery charge insufficient") were preserved and rearranged with the main categories when splitting or fusing the branches. Keeping the overall amount of information constant, though, generally implies the same level of availability for each of the listed component problems. If availability was the only reason for the pruning bias, then one would expect judged probabilities of combined categories to match the sum of judged probabilities of corresponding split categories, since the ease with which a person can retrieve occurrences or construct scenarios that correspond to these categories should be the same. In other words, possible failure causes were not brought "out of sight", they were just arranged differently. In fact, though, the authors observed a well pronounced pruning bias even in this version of their experiments, thus availability is unlikely to be solely responsible for the bias.

With respect to description-dependence expressed in probability judgments for dimensional spaces (where different intervals are presented for evaluation), none of the three abovementioned accounts seems to be suited as an exclusive explanation. Remember the example of members of the Decision Analysis Society assessing probabilities that the future number of members of their society will fall into particular ranges. *Availability* of instances or scenarios that match the event under consideration can be excluded, because the categories cover all possible intervals of events (i.e., there is no "all other membership numbers" category). In addition, in such a case there is no *ambiguity* about the categories since they are represented by concrete numerical intervals. And *credibility* with respect to the conveyance of some meaningful information about likely intervals of future membership numbers from the experimenter to the respondents is also invalid, because both the low and the high group partitions were introduced to the participants (and they self-allocated to one of the partitions by whether the last digit of their primary home telephone number was even or odd). For that reason, Fox and Rottenstreich (2003), and Fox and Clemen (2005) bring forward another explanation which they call the "anchoring and insufficient adjustment" account. It focuses on an "ignorance prior" heuristic based on the number of available categories on the main top level.

Anchoring and insufficient adjustment of an ignorance prior. Fox and Clemen (2005) argue that respondents who are asked to assign probabilities to branches of fault trees or events proceed in *two subsequent steps*. In the first step, it is assumed that they notionally distribute probabilities *evenly* across all available (main) categories. This leads to a uniform distribution of probability: If there are N branches (including the "all other problems" category) in the tree, then, without considering its informative value, each branch initially receives a probability of $1/N$, called the "ignorance prior". That means, by default and in the absence of any obvious alternative, people apply the so-called "principle of insufficient reason" which had been ascribed to Leibniz and Laplace (Fox and Clemen (2005, p. 1420)). $1/N$ allocations seem to be a general phenomenon and have been documented, for instance, for employees allocating retirement savings among potential investments in defined contribution plans (Benartzi and Thaler (2001)). In this context the phenomenon is referred to as *naïve diversification*, but it is also observed within a broader context, whenever it comes to the allocation of some scarce resource (probability, attribute weight, or money) over a fixed set of possibilities (events, attributes, or investments) (see Fox, Bardolet, and Lieb (2005), and Langer and Fox (2005)). The $1/N$ allocation then serves as an *anchor* for further considerations. In the second step, people adjust $1/N$ probabilities by deliberately thinking about the different categories which are presented. That is to say, based on the uniform distribution, people reassign probabilities from branches which they think are rather unlikely to branches which they think are more likely to occur. Yet, it usually turns out that this adjustment happens to be insufficient in most cases (see, e.g., Kahneman, Slovic, and Tversky (1982), or Epley and Gilovich (2001)). According to this account, judged probabilities generally tend to be too close to $1/N$ at the end. Hence, (objective) probabilities which are below $1/N$ tend to be overestimated, and (objective) probabilities which are above $1/N$ tend to be underrated. This result is consistent with the favorite-longshot bias observed in horse racetrack betting and other sports betting markets (cf. subsection 2.1.3). To repeat, the favorite-longshot bias refers to a phenomenon in sports betting where odds of extreme underdogs (i.e. unlikely winners = longshots) tend to be overbet, whereas odds of likely winners (i.e. favorites) tend to be underbet (see, e.g., Thaler and Ziemba (1988)). However, there is a subtle difference between the ignorance prior bias and the favorite-longshot bias: partition-dependence based on the ignorance prior predicts that in an N-event partition, events with expressed probabilities lower than $1/N$ are likely to be overweighted. Overweighting should occur even when probabilities are fairly large (e.g., around .25 in a four-fold partition). By contrast, overweighting in the favorite-longshot bias refers to events with probabilities that are quite low (e.g., below .10). So with partition-dependence, the direction of the bias does not depend on whether a certain category or a particular event is an extreme "underdog" or a "favorite", but rather depends on the relationship between the objective probability for a category or an event and $1/N$. It is quite possible that details of trading microstructure, some influence of extreme optimism, and other forces induce overweighting of very low probabilities (favorite-longshot bias) while partition-dependence is more basic and spans a fuller range of probabilities.

The anchoring and (insufficient) adjustment account is suited to explain the pruning bias in fault trees and other partition-dependence related phenomena. To illustrate, consider the following situation: if there are seven categories (counting "all other problems"), as in the original (unpruned) fault tree, judged probability of each failure cause is expected to be biased toward $^1/_7$ (14.3%). If, for instance, three of the branches are pruned from the tree and subsumed under the "all other problems" branch, this results in a four-branch tree (counting "all other problems"). Judged probability for each of the remaining categories is then assumed to be biased toward $^1/_4$ (25%). Leaving aside the effects of adjusting probabilities in the second step, the ignorance prior of the residual failure category then only raises from $^1/_7$ to $^1/_4$ (an absolute increase of $^3/_{28} \approx$ 10.7%). This value should be compared to the value of $^4/_7$ (57.1%) that would have been attributed if probabilities of the three removed failure causes had been completely subsumed under the "all other problems" category. By this mechanism, the ignorance prior account can explain why judged probability of the "all other problems" category in the pruned tree treatment did not fully absorb judged probabilities of the pruned branches in the Fischhoff, Slovic, and Lichtenstein (1978) experiments. The ignorance prior account is also suited to explain the median judgment differences in the Decision Analysis Society example mentioned above: Applying an ignorance prior over presented intervals would yield an $1/N$ judgment of 20% for each of the five intervals. With respect to the target event "number of members > 1000" this implies an ignorance prior probability of 20% in the low partition group, where this event was represented by one out of five categories, and an ignorance prior probability of 80% (4 × 20%) in the high partition group, where the target event was represented by four out of five categories. The actual results of 10% and 35%, respectively, are partway between those $1/N$ judgments and a common subjective probability for the interval (> 1000) that is partition-independent.

Note that the expression "partition-dependence" basically refers to the finding that judged probabilities systematically *depend* on the concrete way in which a state space is (deliberately or unconsciously) *partitioned* into single events, thus referring to a judgmental phenomenon. However, in the psychological literature it seems as if the expression "partition-dependence" is mainly used to refer to description-dependence with respect to the "anchoring on the ignorance prior and insufficient adjustment" account, thus placing more emphasis on this particular explanation (or cause) of the judgmental phenomenon.

2.2.4 Support theory

2.2.4.1 The model

All of the previously discussed findings show that alternative descriptions of the same category or event may lead to different subjective judgments of probability ("description-dependence"/"partition-dependence"). As was demonstrated, it seems as if the observation of description-dependence can be extended from the narrow domain of fault trees (judgments of the relative frequency of various categories of fault in a system) to the more general domain of *judged probabilities of uncertain events*. In formal analysis, the pruning bias and related description-dependence biases that arise in the context of subjective judgments of probability are captured by "*support theory*". Support theory is a *descriptive* model of belief and judgment under uncertainty.[59] It was first developed by Tversky and Koehler (1994), and Rottenstreich and Tversky (1997) to account for the empirical findings that

(i) different descriptions of the same event can lead to subjective probability judgments that differ systematically (as demonstrated in Fischhoff, Slovic, and Lichtenstein (1978)),

(ii) judged probability of a conjunction, $p(A \cup C)$, sometimes exceeds judged probabilities of its marginal totals, $p(A)$ and $p(C)$ (known as the "conjunction fallacy", see Tversky and Kahneman (1983)),

and, mainly relevant in the context of the present work,

(iii) judged probability of an event is sometimes less than the sum of judged probabilities of separately judged constituent events (Teigen (1974, Experiment 2 on heights of students), and Tversky and Fox (1995)).

To repeat, support theory is in contrast to the normative Bayesian school which assumes the *degree of belief* to be represented by an *additive* probability measure and which further assumes validity of the so-called *extensionality principle* or *description invariance* (i.e., "[e]vents with the same extension are assigned the same probability." (Tversky and Koehler (1994, p. 547))).

The general idea of support theory can be described as follows: When people are asked to make subjective probability statements on a given event, they typically bring to mind some—but usually not all—of its constituent elements. They then notionally estimate the probability of each of these elements, and aggregate these into a combined value for the event under consideration. Accordingly, judged probability depends on (i) how many and what kind of elements people can recall when they are asked to judge probabilities, and (ii) how the aggregation into a combined value works in detail.

[59] Cf. here and in the following, as well as for a brief survey of support theory, e.g., Fox and See (2003, pp. 288-290).

Let T be a state space with a minimum of two different future states of the world that may occur (and of which it is known that exactly one state will in fact occur).[60] Subsets of T are then called *events*. The main principle of support theory is that judged probabilities do not directly refer to the *events* under consideration, but rather refer to specific *descriptions* of these events. Descriptions of events are referred to as "hypotheses". Accordingly, denote H a set of hypotheses describing the events in T. Then each hypothesis $A \in H$ is assumed to represent a unique event $A' \subset T$. It is important to stress that different hypotheses (e.g., A and B) may correspond to the same event, thus $A' = B'$ (i.e., this is a many-to-one relation).[61] Eventually, different hypotheses of the same event can lead to different probability judgments, depending on the particular degree of explicitness of its description. This fact is taken into account by the supposition that each hypothesis A is linked to a (non-negative) *support value* $s(A)$ which represents the strength of evidence (or: degree of support) for that hypothesis. In other words, probability judgments are *mediated* by individual judgments of evidence or support. The support function s is a vehicle to represent how a judge summarizes the recruited evidence in favor of an event, given the specific description of this event. However, the support function is assumed not to be directly observable. Support itself is either the result of effortful reasoning and/or calculation by the judge, or is derived by applying heuristics of information processing, thus being susceptible to associated biases. In particular, these heuristics may include representativeness, availability, or anchoring and adjustment (Kahneman, Slovic, and Tversky (1982)). To repeat, availability refers to the ease with which instances or occurrences can be brought to mind (e.g., by recalling concrete occurrences in the past or by imagining plausible instances). With representativeness, the support for a given hypothesis depends on how characteristic this description appears for a certain stereotype. Finally, the mechanism of anchoring and adjustment is employed if people initially focus on a certain value or distribution and then derive their estimates by adjusting the anchor. It is assumed that people generally "accept" the respective form of a given judgment or decision problem (i.e., the respective framing) as it is presented to them. In particular, people do *not* seem to intuitively transform the

[60] In the following, the notation is adopted from the original work of Tversky and Koehler (1994), and Rottenstreich and Tversky (1997).

[61] If H is finite and if it contains at least one hypothesis for each event, then A is *elementary* if $A' \in T$. A is *null* if $A' = \emptyset$. A and B are *exclusive* if $A' \cap B' = \emptyset$. If A and B are in H, and they are exclusive, then their explicit disjunction ($A \vee B$) is also in H. $\vee$ is assumed to be associative and commutative so that $(A \vee B)' = A' \cup B'$ (Tversky and Koehler (1994, p. 548)).

given decision frame into a representation of whatever canonical classes. Accordingly, the initial value or distribution that serves as a starting point to further think about the judgment is strongly influenced by the concrete representation of the problem. And since adjustments are found to be typically insufficient, different starting points can cause different judgments, which are then biased toward the initially set anchor.

In formal notation, define (A, B) an *evaluation frame* consisting of a pair of exclusive hypotheses: the *focal* hypothesis A that needs to be assessed, and the *alternative* hypothesis B ($A, B \in H$). In its general form, support theory then expresses the judged probability[62] $p(A, B)$ that a focal hypothesis A occurs (instead of its complement B) by:

$$p(A,B) = \frac{s(A)}{s(A)+s(B)} \tag{2.8}$$

This expression is interpreted as relative support or balance of evidence for a certain hypothesis against its alternative (Tversky and Koehler (1994)). Support theory incorporates two further assertions:

(i) *Unpacking* an aggregate hypothesis A into an *explicit disjunction* of mutually exclusive constituents, e.g. $A_1 \vee A_2 \vee A_3$, leads to increased support for that hypothesis despite the fact that both descriptions refer to the same event, thus $A' = (A_1 \vee A_2 \vee A_3)'$. Unpacking a hypothesis means explicitly mentioning (some of) its components as part of the description like, for example, unpacking the hypothesis “Tomorrow, there will be *precipitation* in the city of Münster” into “Tomorrow, there will be *rain*, *snow*, or *hail* in the city of Münster”.[63] The rationale behind this assertion is that unpacking a hypothesis may remind people of possibilities which they otherwise might have missed, and/or increases the salience of listed subcomponents.

(ii) *Separately evaluating* constituent hypotheses (e.g. A_1, A_2, and A_3) also yields a higher total support than evaluating the aggregate focal hypothesis A. Thus, as-

[62] In formal terms, a person’s probability judgment is a mapping p from an evaluation frame to the unit interval.

[63] Assuming, without being a meteorologist, that any form of precipitation can either occur as rain, snow, or hail.

sessing the *three* constituent hypotheses "Tomorrow, there will be rain in the city of Münster" *plus* "Tomorrow, there will be snow in the city of Münster" *plus* "Tomorrow, there will be hail in the city of Münster" generally results in even more overall support than the above mentioned unpacked (but jointly evaluated) hypothesis "Tomorrow, there will be *rain*, *snow*, or *hail* in the city of Münster."

In formal terms, (i) and (ii) can be expressed by the following (assuming that A_1 and A_2 are a *partition* of A, i.e., they are exclusive and exhaustive constituents, or to put it more formally, A is an *implicit disjunction* of the coextensional explicit disjunction $A_1 \vee A_2$, thus $A' = (A_1 \vee A_2)'$):

$$s(A) \leq s(A_1 \vee A_2) \leq s(A_1) + s(A_2) \tag{2.9}$$

To clarify the terminology, the inequality on the left hand side of (2.9), $s(A) \leq s(A_1 \vee A_2)$, is called *implicit subadditivity* of support values, whereas the inequality on the right hand side, $s(A_1 \vee A_2) \leq s(A_1) + s(A_2)$, is termed *explicit subadditivity* (Rottenstreich and Tversky (1997)). The inequality $s(A) \leq s(A_1) + s(A_2)$, which follows from expression (2.9), is referred to as *generic subadditivity*.

Since the theory postulates the vehicle of "support" as an intermediary step between hypotheses and probability judgments, the features of the support function can be transferred to judged probability. With respect to the *(sub-)additivity* expressed in *judged probabilities*, equation (2.8) implies *binary complementarity* (i.e. *additivity*) at the highest level; that is, judged probabilities for the focal hypothesis A, and its simple negation B (the complement) sum to one:

$$p(A) + p(B) = 1 \tag{2.10}$$

Equation (2.10) indicates that judged probabilities for the two complement hypotheses "Tomorrow, there will be precipitation in the city of Münster" and "Tomorrow, there will be *no* precipitation in the city of Münster" sum to unity. Contrariwise, equation (2.8) in conjunction with the inequalities expressed in (2.9) imply *subadditivity* *of judged probabilities* at lower levels, that is, judged probability of hypothesis A is less than (or equal to) the sum of judged probabilities of its disjoint components:

$$p(A) \leq p(A_1) + p(A_2) \qquad (2.11)$$

That is to say that not only *support*, but also *judged probabilities* for separately assessed hypotheses "…rain…" *plus* "…snow…" *plus* "…hail…" are expected to be (at least as high, but typically) greater in sum than judged probability for the focal hypothesis (i.e., asking for "…precipitation…" in Münster). Expression (2.11) directly corresponds to generic subadditivity in support values and is the most central subject matter of analysis in the present work. For the sake of completeness, it should be mentioned that implicit subadditivity (inequality (2.12)) and explicit subadditivity (inequality (2.13)) directly transfer to the level of judged probabilities, too:

$$p(A) \leq p(A_1 \vee A_2) \qquad (2.12)$$

$$p(A_1 \vee A_2) \leq p(A_1) + p(A_2) \qquad (2.13)$$

Expression (2.12) refers to implicit subadditivity by unpacking a hypothesis A into an explicit disjunction of constituent hypotheses as described above and states that—at the level of probabilities—judgments tend to be (equal or) greater when a hypothesis A is explicitly broken down into its alternatives like "Tomorrow, there will be *rain*, *snow*, or *hail* in the city of Münster". Expression (2.13), in turn, refers to explicit subadditivity and hypothesizes that—also at the level of probabilities—the sum of judgments may even be greater when subcomponents are judged separately. To summarize, *unpacking* of a given hypothesis may increase (but—according to support theory's original form—may not decrease) its overall judged probability.

With respect to the ignorance-prior (and insufficient adjustment) account introduced in subsection 2.2.3 above, Fox and Rottenstreich (2003) demonstrate how partition-dependence can be modeled within the framework of support theory. Imagine, for instance, a focal hypothesis A that refers to the probability that "Sunday will be hotter than any other day of the next week," and its alternative hypothesis B that implies the probability that "Sunday will *not* be hotter than any other day of the next week." Transforming the probability of the focal hypothesis A, which results from equation (2.8), into odds form[64] yields:

$$R(A,B) \equiv \frac{p(A,B)}{1-p(A,B)} = \frac{s(A)}{s(B)} \tag{2.14}$$

The odds form presented in equation (2.14) can be rewritten as:

$$R(A,B) = \left(\frac{N_A}{N_B}\right)^{1-\lambda} \cdot \left(\frac{s^*(A)}{s^*(B)}\right)^{\lambda} \qquad \text{with } 0 \le \lambda \le 1 \tag{2.15}$$

N_A and N_B denote the number of elements in the (subjective) partition of the focal and the alternative hypotheses, respectively, $s^*(A)$ and $s^*(B)$ represent the support values that result from evaluative assessment (adjustment), $(1 - \lambda)$ is the relative contribution of the ignorance prior, and λ is the relative contribution of adjusted support values.[65] Thus, support is divided into two components: one component reflecting the ignorance prior and another component reflecting assessed support values. If $\lambda = 1$, then the judgment completely relies on the evaluative assessment of the focal hypothesis against its alternative hypothesis. If, in turn, $\lambda = 0$ (complete ignorance), then the judgment is determined only by the ignorance prior term of the equation. Transferred to the binary partition of the "Sunday will/will not be hottest day" example, $R(A, B) = 1$ if $\lambda = 0$, thus judged probability is one half. Whereas support theory assumes that subadditivity is driven by the fact that *support* is subadditive in general (which leads to subadditive probability judgments), the Fox and Rottenstreich (2003) model is interesting, because it demonstrates that subadditivity can also arise from the ignorance prior and the way in which it interacts with support (regardless of whether recruited support is subadditive or not) (see Clemen and Ulu (2008)).

[64] Note that odds, in the present form, indicate relative probabilities, i.e. $p/(1–p)$, not absolute probabilities. If, for instance, the probability of a given event is one half, then odds are 1:1; if probability is 1/7, then odds are 1:6.

[65] When using the logarithm of R, $(1 - \lambda)$ and λ can be interpreted as weights on the two components in a linear model (Fox and Rottenstreich (2003)).

2.2.4.2 Further empirical evidence

As noted above, description-dependence does not only apply to *qualitative* descriptions of events, but also applies to *quantitative* accounts of continuous variables which are typically represented by different intervals of a state space. Thus, the assumptions and conclusions underlying support theory do also transfer to partition-dependence in the context of dimensional spaces.[66] Tversky and Fox (1995), for instance, report subadditivity of probability judgments when groups of Stanford students and football fans were asked to assign probabilities to various intervals into which an uncertain quantity might fall (such as the victory margin in the next Super Bowl, the absolute performance of the Dow Jones Industrial Average over a given period of time, or future San Francisco temperature). Their results demonstrate that, by unpacking a particular event into separately judged components, subjects' judgments happen to be substantially subadditive, which holds across all event domains, thereby providing evidence in favor of support theory (and its validity for quantitative descriptions of events). The notion that subadditivity holds for quantitative hypotheses is informative, because failures of retrieval as a possible account can be largely ruled out in these cases: while memory restrictions may lead people to overlook one or more component hypotheses in qualitative settings, this is most unlikely to happen with respect to events of metric-scaled intervals.

A variety of different studies reviewed and discussed by Tversky and Koehler (1994), and Rottenstreich and Tversky (1997) gave reason to develop and to bring forward the theoretical framework of support theory. Among these studies, Fox, Rogers, and Tversky (1996, Study 2) found subadditivity and binary complementarity in judgments of professional option traders and support staff at the Pacific Stock Exchange and Chicago Board Options Exchange. They asked option trader professionals to judge the probability that the closing price of, say, Microsoft stock would fall within a particular range two weeks in the future. On the one hand, subadditivity showed when four disjoint intervals, spanning the complete set of possible future prices, were presented for evaluation, as the sums of assigned probabilities were typically greater than one. In par-

---

[66] Clemen and Ulu (2008) have a slightly different view. They note that partition-dependence, although not inconsistent, is not implied by support theory (p. 836). They present a model in which judged probability is characterized by a linear combination of ignorance prior and support components. According to the authors, their evidence suggests "additivity of the support function for continuous variables." (p. 837).

ticular, when the four intervals were presented in the form of different target events,[67] it turned out that judged probability for a packed interval, $p(A_1 \cup A_2)$, was generally lower than the sum of separately judged probabilities for the unpacked events, $p(A_1) + p(A_2)$. On the other hand, binary complementarity appeared within the same subject pool when binary partitions were presented, since in that case the sums of assigned probabilities were close to one. These results are not only striking because they are in line with the predictions of support theory, but also because the participants were experts whose every-day business is to make probability assessments on future stock market developments. In another study, subadditivity was also documented in odds of British bookmakers (Ayton (1997)): Consistent with support theory, bookmakers' odds for general hypotheses were subadditive, i.e., they were smaller than the sum of odds offered for an explicitly unpacked, but logically equivalent, disjunction of events that was subsumed by the general hypothesis. Thus, support theory generally seems to be suited to accommodate major violations of the normative assumption of descriptive invariance that is observed in human behavior.

Evidence from Fox, Rogers, and Tversky's professional option traders, Ayton's analysis of bookmakers' odds, combined with the Decision Analysis Society example introduced in subsection 2.2.2.2, experimental evidence suggests that even experts are susceptible to partition-dependence in judging probabilities. However, since the "ignorance prior" account implies adjustments of $1/N$ probabilities, this gives reason to assume that the partition-dependence bias could be less pronounced when knowledgeable people have more information on events than others and when they can use this information to adjust ignorance prior probabilities more accurately (Fox, Bardolet, and Lieb (2005)). Evidence from Fox and Clemen (2005, Study 4) supports the conjecture that partition-dependence interacts with competence and knowledge. The authors asked MBA students in a decision models class to evaluate the probability that the Jakarta Stock Index (which was supposed to be quite unfamiliar to the subject pool) and the NASDAQ index (which, by contrast, was supposed to be very familiar to the subject pool) would close in some particular intervals at the end of the year. Results reflected both pronounced partition-dependence and a pronounced competence effect: Judgments for the unknown Jakarta Stock Index almost perfectly coincided with the ignorance

[67] "Different target events" means that the four price intervals were presented either separately or combined (i.e., by implicitly packing two adjacent intervals to form a single interval).

prior probabilities (reflecting $\lambda = 0$ in terms of the model presented in equation (2.15)) and suggests that respondents hardly adjusted $1/N$ probabilities. By contrast, judgments for the more familiar NASDAQ index reflected substantial adjustments of the ignorance prior (reflecting $\lambda > 0$), resulting in a reduced (but still pronounced) bias of partition-dependence.

On the whole, empirical evidence so far suggests that both generic subadditivity as well as explicit subadditivity appear to be robust phenomena of considerable size, whereas evidence for implicit subadditivity is somehow mixed.[68] Thus, it should not be neglected to mention that some recent research (see, e.g., Sloman et al. (2004)) provides evidence for *reverse subadditivity* (called *superadditivity*), where unpacking does not necessarily increase overall judged probability, depending on whether explicitly listed subcategories represent rather typical (characteristic) or atypical (uncharacteristic) instances of the event under consideration. According to the *narrow interpretation conjecture* (NIC), atypical examples of an event with weak support may draw a judge's attention to some special cases at the expense of more meaningful and/or representative exemplars, to which one would have attached more importance in a situation in which the problem was described as an implicit hypothesis. Thus, overall support for the unpacked description can be reduced such that $p(A) \geq p(A_1 \vee A_2)$, which represents *implicit superadditivity*. Hence, probability judgments may be mediated by an anchoring and adjustment process in such a way that people anchor on (atypical) unpacked instances and then adjust insufficiently for other instances. In this context, it also seems to play a role, whether the unpacked descriptions of events do explicitly list *all* possible elements

---

[68] Besides the studies mentioned in the course of the present work, additional evidence in favor of generic and/or explicit subadditivity is presented or reviewed, for example, in Fox and Tversky (1998) [differently (un-)packed prospects that pay a certain amount if a particular team, division, or conference would win the 1995 NBA championship], Brenner and Rottenstreich (1999) [sums of judgments for complementary hypotheses are close to one when the hypotheses are singletons, and are less than one when one of the hypotheses is a disjunction], Redelmeier et al. (1995) [subadditivity in diagnoses of medical doctors], Fox (1999) [expert sports fans predicting game outcomes], Fox and Birke (2002) [attorneys forecasting trial outcomes], Johnson et al. (1993) [willingness to pay for hypothetical insurance policies]. Mixed evidence for implicit subadditivity results from studies by Rottenstreich and Tversky (1997) [(i) forecasting a trial outcome (implicit subadditivity existent) vs. the winner of the next presidential election (no effect), (ii) "homicide by an acquaintance or homicide by a stranger" (implicit subadditivity existent) vs. "daytime homicide or nighttime homicide" (no effect)], Fox and Tversky (1998) [unpacking the teams from the Eastern Conference (implicit subadditivity existent) vs. unpacking the teams from the Western Conference (no effect)], Fox and Birke (2002) [judgments by experienced attorneys], Fox and See (2003) [strong implicit subadditivity in judgments of the probability that various categories of teams would win a conference basketball championship vs. no effect when scenarios were unpacked into disjunctions of dimensional subhypotheses], Fox and Clemen (2005) [no effect from unpacking events into obvious constituents], or Koehler, Brenner, and Tversky (1997) [implicit subadditivity due to unpacking the alternative hypothesis].

of the packed description, or whether some elements are collapsed into a residual "all other members" category (Sloman et al. (2004, p. 581)). Tackling another issue in the context of description-dependence, Bilgin and Brenner (2008) analyze whether the temporal distance (and thereby the degree of abstractness) from the time when a probability judgment is elicited to the moment when the respective uncertainty is resolved has an impact on the degree of description-dependence. They find unpacking to have a greater impact on the degree of description-dependence in probability judgments for uncertainties that are resolved in the distant rather than the near future (applies similarly to both sub- and superadditive hypotheses). Therefore, temporal distance can additionally be regarded as an influencing factor on the size of unpacking effects. However, the last-mentioned findings should rather apply to qualitative hypotheses than to quantitative settings which are the object of investigation in the course of the present work.

2.3 Experimental asset markets

2.3.1 Basic principles

In chapters 3 and 4 of the present work, experimental asset markets (in this case: experimental prediction markets) are used as a research method to examine whether the partition-dependence bias persists or diminishes in a competitive market setting. While carrying out experiments is the obvious research method in most natural sciences, and also has a long tradition in other social sciences, particularly in (cognitive) psychology, this method of research was first made popular in economics by Vernon L. Smith and other researches[69] since the 1960s. Smith was awarded the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel in 2002 "for having established laboratory experiments as a tool in empirical economic analysis, especially in the study of alternative market mechanisms."[70] The introduction of experimental methods in economics is a remarkable innovation, because until then it was generally assumed that economists are obliged to exclusively make use of naturally-occurring field data (as

[69] Among them were (in alphabetical order): Colin Camerer, Robert Forsythe, Daniel Friedman, Teck Ho, Charles A. Holt, John H. Kagel, Charles Plott, Alvin E. Roth, and Shyam Sunder; this list is not exhaustive, though.

[70] Cf. here and below, Royal Swedish Academy of Sciences (2002).

used in chapter 5 of the present work) to test their theories. The possibility of conducting experiments in a controlled laboratory environment was widely neglected by then.

A major benefit of using *laboratory experiments* is the practicability of the *ceteris paribus* condition which allows controlling for all other variables by holding them constant over different treatments while analyzing variations in the variable under investigation. While establishing a controlled laboratory environment is considered quite easy in most natural sciences, it originally seemed a hard task to control for important factors in economics (Samuelson and Nordhaus (1985, p. 8)).[71] The development of experimental economics as a "new" research area, however, fostered ways in which it is now possible to experimentally study the behavior of economic agents in economically relevant settings (like, e.g., negotiations, product- or asset markets, etc.) in a laboratory. Experimental economics aims to generate an economically relevant but self-contained situation in which agents mostly behave *as if* they acted in a real-world setting, while the experimenter is in a position to control for variables like financial endowment, level and type of information, opportunities to communicate with other participants, etc. This makes it possible to isolate the focal (or: treatment) variable(s) and to analyze its impact on any dependent parameter, e.g., the result of a negotiation, the demand and supply of goods or assets, or market prices, thereby testing predictions from economic theories. Experiments generally may be conducted either in a laboratory setting (lab experiment) or in the field (field experiment). Laboratory and field experiments may appear in the form of individual (choice) experiments, game theory experiments, or (asset) market experiments. These types of experiments found their way into economics in a blend of different research areas: decision making, bargaining, coordination, social preferences, public goods, learning, auctions, as well as market mechanisms and market institutions. Smith dedicated most of his work especially the last two mentioned fields of research, auctions and markets. While experiments in cognitive psychology mainly focus on analyzing individual behavior, auctions and (asset) market experiments aim to study market outcomes, i.e., the results from individuals implicitly interacting with each other via bids, offers, prices, demand, and supply.

Being a pioneer Smith also made substantial contributions to methodological (or: design) issues in carrying out experiments in economics (Smith (1976), see also

[71] Interestingly, whereas this quote comes from the 12th ed. (1985) of their famous "Economics" textbook, such a remark no longer exists in the latest edition (18th ed., 2005).

Smith (1982b)).[72] Procedural aspects which are most different compared to experiments, e.g., in cognitive psychology, are *monetary incentives* and *repeated trials*. Monetary incentives are considered important to reimburse subjects for their decision costs and to motivate them to think thoroughly about their decisions in an economic sense.[73] On the one hand, when financial compensation is applied subjects can be paid a flat (i.e., fixed) reward for participating which may help in recompensing them for their decision costs. On the other hand, it is more preferable to pay subjects an incentive-compatible performance-based compensation to stimulate them to think carefully about their actions and decisions.[74] Offering performance-based rewards is the obvious way of compensation in asset market experiments, since payoffs can be directly linked to subjects' portfolio performance. However, it seems at least as important as the concrete shape of the payoff function to offer subjects the prospect of earning an adequate compensation that is perceived not to be negligible small.[75] Repeated trials, in turn, are considered important to give participants the opportunity to get acquainted with the experimental setting *and* to learn from (implicit) interactions with other subjects. Repetition of trials is of particular importance when testing equilibrium theories which is often the case in economics. The later trials then give reason to assume that initial confusion (if it existed) has calmed down and adjustment processes have been completed.[76]

Other methodological issues include, e.g., the question of providing participants with *instructions* (scripts) and how they should look like (vs. forcing subjects to ad-lib), or the question of *honesty* (vs. deception) (Hertwig and Ortmann (2001)). Instructions[77] aim to describe the experimental set-up, explain the role of different agents and their

---

72 For how-to instructions on conducting experimental methods in economics (in particular on experimental asset markets), see, e.g., Friedman and Sunder (1994).

73 This assumes that people are self-interested and mainly motivated by money (Camerer (1997)).

74 Other means of (non-monetary) payment are points, credit points (for students), virtual money, non-cash prizes and the like.

75 Other issues that arise with incentivizing subjects include the questions of whether (i) one should pay all participants or only a few randomly selected subjects, (ii) whether all decisions (or trials) should finally be compensated or only part of it (keeping the overall amount of compensation (that is paid to all participants as a whole) constant in each case), and (iii) how to deal with the risk of losses.

76 Usually, repeated trials are run under a "stationary replication" condition, i.e., the setting is re-established as it originally was at the beginning of the previous trial, including initial endowments, the set of possible actions and decisions, etc. The only thing a subject can preserve is her experience from the earlier trials. Camerer compares the "stationary replication" condition with the movie "Groundhog Day" in which Phil Connors, a famous TV-weatherman, is captured in a seemingly never-ending 24-hours time loop, re-experiencing the same day again and again: "The endless looping of experience, while boring, turns out to be ideal for learning since different actions can be tried while everything else is held equal." Camerer (1997, p. 319, fn. 2).

77 In most cases, instructions are handed out to all participants and are read out aloud by the experimenter at the beginning of an experimental session.

mutual relationship, the set of possible actions/decisions and their consequences, and the composition of subjects' compensation through the payoff function. Economists usually (not always) employ an abstract context for their experimental set-up. Creating a rich and illustrative storyline around the general hypotheses, by contrast, bears the risk of generating non-monetary utility and possibly distracts subjects from focusing their attention to financial incentives (Camerer (1997)). Hence, carefully elaborated instructions are an important design feature in experimental economics to create a level-playing field for all participants, to enhance replicability of results and comparability across sessions. Honesty is considered to be an incontestable basic principle in experimental economics, whereas it is not, e.g., in psychology. In simple terms, honesty means not to actively claim something as being true, when the experimenter knows it is false (Camerer (1997)). Being honest to the subjects is a basic prerequisite to build and preserve his or her good reputation as a scholar, but is also considered as a public good that helps the whole branch of experimental research to maintain its trustworthiness. In addition, honesty is necessary to ensure that subjects make their decisions based on the financial compensation they can expect to earn, rather than based on a psychological reaction to assumed deception, i.e., subjects may suspect manipulation if it is present (Davis and Holt (1993, pp. 23-24)). However, being honest with respect to the true purpose of a study may result in unintentional strategic behavior by the subjects or may cause demand effects. Besides, for some purposes one may be tempted to deceive subjects in order to simulate (and study) situations which rarely occur in natural settings.

However, experimental methods in the lab have to defend against different objections (see, e.g., Falk and Fehr (2003)). These include, e.g., a subject pool bias, or the concern that subjects may not take their task seriously enough due to stakes which are too low, or a lack of generality due to a (relatively) small subject sample size. First, subject pool bias means that experimental results may be biased by the fact that most experiments use student subjects. Students are easy to recruit, (typically) have a fine grasp with respect to the experimental script, and (usually) have low opportunity costs. However, results may not be representative for the cross-section of the true population and thus may fail to accurately predict real-life behavior. Hence, recent studies also use non-student subjects, recruited by characteristics such as their experience, profession, etc., or to make the sample as representative of a given population as possible. Evidence to date on subject pool differences is mixed (see Falk and Fehr (2003) for a brief discussion), thus the question of subject pool bias has not been finally answered yet. Second, as discussed above, choosing a poorly calibrated compensation scheme and/or insignificant stake levels may not motivate experimental participants to think carefully about what decisions to take. In a meta-level study, Camerer and Hogarth (1999) review a multitude of experiments with no, low, or high performance-based financial incentives to analyze the effects of stake size. They find that variance in results is reduced as payments are increased, whereas virtually no effect is observed on *mean* performance. On the whole, higher incentives contribute to improved performance, but at the same time strongly interact with other treatment variables. Interestingly, the authors note that, in their sample of reviewed studies, violations of rationality could not be eliminated in replication studies merely by offering increased incentives. Third, small sample sizes may limit the statistical significance of lab experiments. Although this, too, is certainly a relevant issue, one can (to some extent) increase sample size by recruiting additional subjects or put particular effort on recruiting a representative sample of participants (which is closely related to the subject pool bias concern) to circumvent these concerns.

A more general critique of experimental methods, which founds on the philosophy of science, challenges *internal* and *external validity* of these methods (Falk and Fehr (2003)). Internal validity refers to the question of whether it is possible to infer causalities from the concrete experimental data. This is mainly a question of properly applying the experimental toolbox, i.e., to use appropriate experimental controls, a well thought-out experimental protocol, and correct methods of data analysis. External validity, in turn, refers to the question of whether it is possible at all to generalize conclusions gained from experimental data to the field. On the one hand, external validity depends on whether one accepts the methodical technique of induction (however, this argument applies to all empirical methods and results).[78] On the other hand, external validity depends on whether one is able to exactly capture those conditions in the experiment which prevail in a natural setting (however, to some extent experiments deliberately abstract from reality, as this is common practice in economic models).

Harrison and List "see the beauty of lab experiments within a broader context—when they are combined with field data, they permit sharper and more convincing inference." (Harrison and List (2004, p. 1009); they also present a general discussion of field experiments; see also List (2006), and List and Reiley (2008) for surveys on field ex-

---

[78] The *principle of induction* states that results can be transferred to new situations as long as the most central underlying conditions remain unchanged (Falk and Fehr (2003)).

periments). While lab experiments offer the greatest possible degree of control over variables on the one hand, and empirical research on field data promises best answers to the question of how people do actually behave in natural-occurring environments on the other hand, *field experiments* can be positioned somewhere in-between. With respect to the methodological relationship between lab and field experiments, Harrison and List (2004, p. 1010) argue that field experiments are not "simply less controlled variants of laboratory experiments," but are effective complements instead. Field experiments can generate exogenous variation in the variables of interest which is an advantage compared to conventional empirical economics. This allows establishing causality rather than pure correlation. Compared to lab experiments, in turn, field experiments give up some of the control that a laboratory experimenter may have over her environment. In return, field experiments gain validity from increased realism (List and Reiley (2008)). Defining field experiments (literally, experiments *in the field*) would concentrate on the *natural environment* in which people are used to take their actions. The "natural environment," in turn, can be interpreted with respect to a number of different dimensions: e.g., self-selected (and maybe more experienced) subjects (instead of students recruited from a single class),[79] the exchange of privately held goods by subjects (instead of an exogenously allocated initial endowment), less-intrusive instructions (instead of scripts describing abstract conditions in great detail), or a familiar home environment (instead of an artificial laboratory), to name just a few. With field experiments, it is possible to observe people's choices and decisions in an experimental setting and with significant control, while these people, acting in a more or less natural and/or familiar environment, at the same time do not feel unnaturally controlled.[80]

With respect to the relationship between theory and data, Camerer (1997) characterizes experiments in economics as a means of operationalizing and testing a general theory in a specific (artificial) context. Against the background of the great importance of (prediction) market institutions and (prediction) market mechanisms in the context of the present work, the next two subsections (2.3.2 and 2.3.3) aim to review experimental evidence from testing theories in experimental asset markets. Theories that have widely been addressed so far are market efficiency (by information dissemination, by informa-

---

[79] However, recruiting students from a single class (or a few parallel courses) may be advantageous in generating a homogeneous subject pool with most similar background knowledge, if this is wished to keep constant.

[80] For a more detailed taxonomy of field experiments, see subsection 2.2 in Harrison and List (2004).

tion aggregation and dissemination, and by information production) and the impact of individual judgment errors on rational market equilibriums. These are particular relevant in the context of the present work.

2.3.2 Experimental tests of market efficiency

2.3.2.1 Preliminaries

Testing market efficiency and other market-related characteristics from field data appears to be a hard, if not impossible task. First, it would be almost impossible to control for the information set of a very large number of traders in naturally-occurring asset markets, and to separate, for instance, insiders (and their level of private information) from noise traders. Second, a test of market efficiency against a normative benchmark suffers from the *joint hypothesis* problem, i.e., such a test unavoidably combines the question of market efficiency with the question of whether the assumed theoretical model is correct. Thus, if in a given situation the market deviates from what the model predicts, is it the market that is inefficient or is it rather the model that imprecisely predicts the "true" value of the asset under consideration? And analyzing price *adjustments* instead that occur as a result of news arrival only provides little remedy, since this does not say anything about the efficiency of price levels, i.e. the market as a whole. Additionally, in many cases it is not only single news that affects an asset price at the same time, but rather a bundle of news including sector-specific and market-specific news that overlap and, in practice, this is hard to disentangle.

Analytical models are often concerned with deriving theoretically appealing solutions to problems that result in equilibriums of a comparative statics analysis. However, the *existence* of equilibrium is no sufficient condition for that it is in fact reached, even under the specific assumptions of the given model. Experimental asset markets, in contrast, can explicitly address this issue by mapping the complete process of market activity from the start to the end. If equilibrium is not reached in experimental asset markets, this does not necessarily mean that equilibrium does not exist; possibly the trading period simply stopped too early. Contrariwise, if one cannot observe any substantial changes in market variables at the end of the trading period, there is no reason to assume that equilibrium would be reached if trading time was prolonged. However, analyzing the price path provides the experimenter with valuable insights on *how* the terminal value of an asset was reached, regardless of whether this state represents a situation of equilibrium or not.

Following Sunder (1995), it can be distinguished between three different experimental approaches of testing informational efficiency in asset markets: the *first* approach focuses on *dissemination* of information. It analyzes whether a particular set of information is effectively conveyed from one group of homogenously informed traders (insiders) to another homogenously group of not-informed traders (noise traders) via the market mechanism. If it works, not-informed traders finally are able to effectively trade as if they possessed the insider information. Dissemination of information mainly refers to the first and to the second proposition of the Hayekian hypothesis (cf. subsection 2.1.4.1). The *second* approach includes the *aggregation* of information that is widely dispersed among traders, *and* the dissemination of this information to all traders. The question is how traders, all endowed with a fractional subset of information, can piece together everyone's individual tesserae into market prices which, at the end, reflect the overall available information. Finally, the *third* approach implements the production of information as an endogenous variable so that it comes to analyzing simultaneous equilibriums in both asset markets and information markets.

The early 1980s can be considered as a starting point for experimental economists to conduct asset market experiments. Results gathered from these first studies provided initial support for the core statements of the Hayekian hypothesis. The findings highlighted the important role of market prices for dissemination and aggregation of information by simply showing that it "works" under some idealized conditions of experimental asset markets.

2.3.2.2 Information dissemination

Forsythe, Palfrey, and Plott (1982) investigate the role of learning and experience in repeated trials of experimental asset markets with surely known asset payoffs, i.e. in the *absence of information asymmetry* among traders. One of their experiments included trading in a single asset paying different amounts of dividends to three types of traders twice a trading round. The market mechanism was organized as an oral double auction. Since the amount of dividends was private knowledge to all traders, two equilibrium market prices existed for the first sub-period of each trading round, namely a naïve equilibrium (i.e., only bearing in mind private information on a trader's own dividends in the two sub-periods) and a perfect foresight equilibrium (i.e., additionally considering potential speculative gains which could be earned in the second sub-period). In the second sub-period of each trading round a single unambiguous equilibrium price existed (namely the maximum amount of dividend paid to any of the three groups of traders). As it turned out from a sequence of eight identical replications under a stationary environment, asset prices in the second sub-period, as expected, were close to the equilibrium price of that sub-period in all trials. Asset prices in the first sub-period oscillated around the naïve equilibrium in the first trading rounds. Remarkably, though, prices in the first sub-period started to converge toward the perfect foresight equilibrium after a few repetitions of trading trials, since traders obviously learned what some of their counterparts were ready to pay for the asset in the second sub-periods. Apparently, experience from the second sub-period of a trading round "swung back" to the first sub-period trading activity in later trading rounds. It can be concluded that the price system was able to reveal information (namely, dividend prospects of other traders) which was privately held at first. And agents in this environment were obviously able to learn from their experience across markets and thus were able to exploit profitable trading strategies.

Plott and Sunder (1982) tie in with the aforementioned experiments by Forsythe, Palfrey, and Plott (1982). They design a series of experimental markets in which assets only pay a single dividend once a trading round, but introduce *uncertain payoffs* instead.[81] Besides, they also use an oral double auction mechanism and define three types of traders. The uncertainty was reflected by the fact that the amount of dividend depended lately on which out of two possible states of the world occurred. Probabilities of occurrence were known in advance. An important design feature was the fact that, at the beginning of each trading period, half of the traders (i.e., six out of twelve, two of each trader type) was given reliable hints on which future state of the world would occur, thus making them to insiders, while the other half of participants was left uninformed. Subjects were informed on the procedure in the instructions; however, the identity of insiders remained undisclosed. Again, there was one prior (or: naïve) information competitive equilibrium and one rational expectations (or: perfect foresight) competitive equilibrium. Under the assumption of risk-neutral individuals, probabilities and dividends were chosen adeptly so that these two equilibriums differed in one of the two

[81] Design details described below refer to Market 3 of their study.

possible states (prior information equilibrium to be at a higher level), but were the same in the other state of the world. The general question addressed by their experiments was, whether transaction prices and asset allocations, in the course of a twelve-fold repetition of trading rounds (with different states of the world occurring), were consistent, if at all, with one of the theoretically derived equilibriums. As a result, they find evidence of trade prices converging closely (but neither perfectly nor instantaneously) toward the rational expectations equilibrium (instead toward the naïve equilibrium), which strongly supports the conjecture that uninformed traders are able, after gaining some experience, to infer reliable information on the real state of the world only from observing market prices. Moreover, profits of insiders were nearly the same as profits of non-insiders suggesting that initially uninformed traders finally acted pretty much as if they possessed insider information.[82]

DeJong et al. (1992) replicate two of Plott and Sunder's (1982) markets in a *computerized* double auction environment. In contrast to an oral double auction, traders can follow only the best bid and ask quote for each asset, the current market price, and transaction prices of their own trades. In particular this means that they can neither observe the prices from other traders' transactions, nor do they know the identity of traders behind the current bid and ask quotes. As a result, the authors find that the markets are nonetheless able to convey information from the informed traders to the uninformed and that prices tend toward rational expectations equilibriums, albeit somewhat slower.

Watts (1993) addresses the question of whether the *common knowledge about the presence of insiders* influences convergence toward the equilibrium. She reexamines the study by Plott and Sunder (1982), introducing a half-half chance for the existence of insiders (in that case, six out of twelve traders became insiders), and no insiders otherwise. She finds that the uncertainty about whether there are informed agents in the markets or whether not reduces the accuracy of rational expectations predictions. In the

[82] Convergence of asset allocations, however, emerged somewhat slower. Banks (1985) generally confirms the conclusions drawn by Plott and Sunder (1982) with respect to the price dynamics of the market. However, he was concerned about the fact that always the same traders were provided with insider information throughout the twelve trading periods. Therefore, one cannot rule out the possibility that rational expectations equilibriums in their study were achieved by mimicking the behavior of traders identified as potential insiders, rather than by inferring information solely from current market prices. (Recall that the market institution in Plott and Sunder (1982, p. 666) was an *oral* double auction which enabled agents to unhamperedly observe trading behavior of other traders.) Banks' findings slightly weaken the claim that information dissemination occurs exclusively by learning from market prices, but do not question the general message.

same vein Camerer and Weigelt (1991) test for “information mirages”[83] in trading behavior when no insiders are around and find that traders are generally able to learn from market data whether insiders are present or not (essentially by the pace of trading), but some information mirages occurred in early trading rounds of a session, suggesting that it took the subjects some time to gain enough experience. A related question is to what extent are the transmission of information and the attainment of rational expectations equilibriums affected by the number of insiders in a market. Nöth and Weber (1996), for instance, conclude that market efficiency increases with the number of insiders. Experimental evidence is mixed, but at least suggests that the diffusion of information becomes more coincidental as the number of informed traders decreases.

Copeland and Friedman (1987) conducted the first experimental study in which information was distributed *sequentially* to traders. They implemented a computerized double auction market and varied both the content and the timing of messages across traders. Their hypothesis of more trading volume as information arrives successively to the market (as opposed to a simultaneous distribution of information) was not confirmed, instead trading volume was lower: in order to avoid unfavorable transactions against insiders, uninformed traders waited until they received the information and hence, traded less often all in all. This was also reflected in wider bid-ask spreads in earlier sections of trading, and immediately after new messages arrived. Besides, they found previous results obtained by oral double auctions to be replicable in their computerized double auction setup.

2.3.2.3 Information aggregation and dissemination

Another type of asset market experiments investigates the second approach of testing information efficiency in asset markets, namely the aggregation of widely dispersed information and conveyance of this information to all traders. Plott and Sunder (1988) conducted a series of experiments that are good representatives of this class of experiments. In their studies three possible states of nature (*X*, *Y*, and *Z*) are equally likely to occur. All variations of their experimental markets included repeated trading rounds in one or more assets with a maturity of one period paying a state-dependent

[83] They define “information mirages” to be self-generating trading, “creating price paths that falsely reveal information that no one has”. This can arise when subjects trade, “mistakenly, as if they had learned inside information from others.” (Camerer and Weigelt (1991, pp. 489-490)).

dividend at the end of each trading period. As in earlier studies, the market mechanism employed was an oral double auction. Markets consisted of eight or twelve traders.

To explore potential aggregation of information, every trader was initially provided with only a fragment of information indicating which state would definitely *not* occur (like "not *Y*" or "not *Z*", each of which was privately sent to one-half of the traders, if *X* was the realization). Thus, the true state of the world could have been inferred through a process of elimination if one possessed all bits of information. However, each single trader could only rule out one of three alternatives, what is different to the insider models mentioned above in which a subgroup of traders received perfect information. All design features were made common knowledge. Final payoffs and probabilities were chosen such that the rational expectations equilibrium differed from the prior information equilibrium. In one series (series A)[84] the market offered a single three-state asset that entailed different payoffs in each of the three states. But payoffs did not only differ in the three states of nature, but also for the three types of traders (to which participants had been assigned at the beginning of the experiment). As a result, it is worth mentioning that market prices did *not* converge toward the rational expectations equilibrium. Presumably, the design was too complex to give traders the opportunity of implicitly recomposing the dispersed information via the price system. Moreover, the design was incomplete, since a single three-state asset is not able to span the whole state space. This motivated the authors to modify the design by introducing three state-spanning contingent claim assets in another series (series B). In that experiment, the markets simultaneously offered three "all-or-nothing" assets, each of which exactly referred to one possible state of the world (*X*, *Y*, or *Z*) and paid off a certain dividend if that state occurred and nothing otherwise.[85] Dividends were still different for the three types of traders. In a further series (series C), Plott and Sunder replicated the series A markets with the distinction that dividends were the same for all traders in the three possible states. As a result, they find that market prices and allocations in the series B and C experiments indeed converge toward the rational expectations equilibrium, with the contingent claims markets performing best. One of the discussed explanations is the conjecture that "the purchase or sale of a security could be directly interpreted as a belief about the occurrence of a particular state. Thus traders in the contingent claims markets had a 'natu-

[84] The labeling of the three series (A–C) does not exactly reflect the chronology of the trading sessions.
[85] State-contingent claims are also referred to as "Arrow-Debreu securities".

ral' knowledge about the preferences of other traders that was not present in the single security markets." (Plott and Sunder (1988, p. 1116)).

With these experiments, the authors were one of the first who showed that, under some specific conditions, (experimental) asset markets are in principle able to aggregate fragmented pieces of information held by different traders, and that they are at the same time able to transfer aggregated information to all agents as a whole. However, it appears to depend particularly on the concrete market institutions and instruments whether this task can be achieved. State-contingent claims seem to be well-suited in doing so. Yet, in accordance with O'Brien and Srivastava (1991) it is also noteworthy that markets, which are efficient in the sense that there are no exploitable arbitrage strategies, need not necessarily to be efficient in a rational expectations sense. In their experiments, O'Brien and Srivastava find that successively increasing the complexity of the market design (like multiple, multi-period assets, correlated dividends across assets and periods) makes it more and more unlikely that the market mechanism aggregates information efficiently, even with a uniform and publicly known dividend structure and a population of experienced subjects.

Forsythe and Lundholm (1990) further investigate the extent to which markets actually aggregate and disseminate information. They conduct a series of experimental asset markets modifying a number of market parameters. They find that trading experience together with common knowledge about other traders' payoff structures (i.e., other traders' preferences) are sufficient conditions to achieve rational expectations equilibriums.

The previously discussed experiments suggest that the ability of markets to aggregate and distribute information via the price system works in general, but depends on different market factors like market rules and institutions, initial allocation of information among agents, experience of traders and the number and nature of offered assets. In some environments, information aggregation worked remarkably well, in other settings markets failed to do so.

2.3.2.4 Information production

The studies on aggregation and conveyance of information in experimental asset markets described above refer to the first and the second proposition of the Hayekian hypothesis. The studies reviewed in the following additionally examine the third proposition, i.e., competition as a discovery process. Recall the free-rider problem that may arise with costly information production: if prices always immediately and completely adjusted to new information, no incentives would exist to *produce* such information at positive costs (cf. subsection 2.1.4.2). This problem is usually encountered by introducing *noisy* rational expectations models. These models allow for noise that covers up the true equilibrium to the extent that gains from trading only reimburse information producers for their information procurement costs (e.g., Grossman and Stiglitz (1980), Hellwig (1980), and Verrecchia (1982)). In these models, equilibriums in both the asset market and in the information market are analyzed simultaneously.

In an experimental market design almost identical to that of Plott and Sunder (1982), Sunder (1992) attaches a positive value to private information and does no longer provide some of the agents with costless insider information.[86] Half of the markets (series A) in his experiment were preceded by a separate market for information in which a fixed number of traders (four out of twelve) could decide to purchase insider information in an auction,[87] thereby gaining knowledge on the true state of nature. The identity of insiders remained undisclosed after the auction took place. As a result, market prices for information were initially high, but decreased substantially as trading experience improved in later trading rounds. Obviously, traders realized that it was not profitable to pay for information that, after a short period of trading, was incorporated in market prices anyway. This intuition rendered costly information acquisition unnecessary.[88] Gross profits of the informed traders turned out to be greater than those of the uninformed; however, when the cost of information procurement was subtracted, net profits were almost the same for both types of traders. The price for information thus was about equal to the marginal profit from this information, which is consistent with noisy rational expectation predictions. According to Grossman and Stiglitz (1976, p. 248), for instance, a market is efficient in its strong form only if the equilibrium price for additional information is zero. Given the rather low prices for information in Sunder's (1992) markets, these appeared to be fairly efficient. In a second set of markets

[86] Theoretically, the value of information is equal to the difference between the expected utility making an optimal decision under the best possible information, and the expected utility of an optimal decision where this information is not available (see Copeland and Friedman (1992, p. 241)).

[87] The auction scheme was a uniform price sealed bid auction (at the fifth highest bid price).

[88] However, if only a few traders initially possess insider information, it could be worthwhile again to purchase insider information and to try exploiting this information head start afterwards (see Huber, Kirchler, and Sutter (2006, p. 190)).

(series B), Sunder (1992) changes the design by charging a fixed price for purchasing insider information, allowing all traders to acquire this information. He finds that the number of information buyers varied widely and did not stabilize during the course of the experiment, because traders could not rely on the fact that enough information was bought at the end. When many traders bought insider information, it was pretty soon revealed through market prices and made it difficult for those who bought the information to yield adequate compensation. When only a few traders purchased the information, though, the market sometimes failed to reveal this information.

Copeland and Friedman (1992) conducted experiments with a *sequential provision of costly information* in different settings, one of which is comparable to Sunder (1992) and yield similar results. In the more complex settings of their experiments, however, markets appeared to be incomplete, allowing substantial net profits to information holders. Huber, Kirchler, and Sutter (2006) demonstrate that even costless information does not necessarily yield a positive marginal utility. In their setting, only above-average informed traders can gain from the additional information they receive.

2.3.3 Experimental tests of individual judgment errors and market outcomes

A central aspect of the present work is the question whether individual judgment biases transfer to competitive market outcomes, particularly in the context of partition-dependence. This subsection is to review existing literature on this issue with respect to other judgmental biases than partition-dependence. An early example of addressing these topics experimentally is Camerer (1987) who analyzes experimental asset markets (employing a double-oral auction mechanism) in which prices provide insights on whether traders *collectively* act as Bayesian updaters in processing sample likelihood information, or whether prices rather reflect traders' susceptibility to the "representativeness bias". In addition, he compares individual judgments elicited before trading to market outcomes. Results are mixed: in one part, both average judgments as well as average market prices are very close to the Bayesian benchmark; in another part, though, market prices appear to be even more biased than individual judgments. However, the size of the bias diminishes as traders' experience increases.

Camerer, Loewenstein, and Weber (1989) compare individual-level judgments to laboratory market outcomes with respect to a bias they call the "curse of knowledge". One common assumption in models of asymmetric information like this one is that better-informed people are able to predict the judgments of less-informed agents. However, it turns out that better-informed people are not always able to disregard their additional private information in estimating less-informed people's judgments, even in an incentivized environment. As a result, predictions of other people's judgments are biased. The authors find the "curse of knowledge" bias to be not fully eliminated in market outcomes, but to be reduced by around 50% compared to individual judgments. They also find that bias strength in individual judgments is reduced with market experience. The main reason for the error-correcting ability of market forces, according to the authors, is the fact that less biased subjects more or less *know* that they are less biased. Therefore, this group of agents trade earlier and more actively, thereby signaling and conveying their information to more biased traders (congruent with the "smart few hypothesis", see subsection 2.1.4.4). Contrary to the market forces, they suggest that incentives and experience did *not* play an important role in reducing the bias in their markets.

In a more recent study, Kluger and Wyatt (2004) use the Monty Hall problem[89] to explore whether individual judgment errors survive in equilibrium market prices. In a two-stage experimental design they first elicit participants' subjective probability estimates on this problem to control for the bias in the judgments. In a second stage, they draw on the same subjects to constitute a market in which trading is related to the same decision problem.[90] Afterwards, they compare individual judgments and therefrom derived errors to market prices and allocations. Each of the twelve sessions comprises a total of six participants. If a market was composed of traders, all of whom were subject to the judgment bias, the authors find that markets are biased in a similar way, even in repeated trials and with substantial incentives. Thus, individual judgment errors aggregated into market prices. Contrariwise, if there were at least a few[91] subjects in the mar-

---

89 The Monty Hall problem is named after an American game show in which a candidate faces three closed doors, two of which hide a goat (blank), and one of which hides a brand new car (grand prize). After the candidate has picked a door which she guesses to hide the car, the host opens a door which he knows is hiding one of the two goats. Leaving the remaining two doors closed, the candidate then gets the chance to decide whether she wishes to switch to the other door or whether she wants to adhere to the initially chosen door. While the chance of winning the new car is one third and remains one third, if the candidate hangs on to her initial decision, the chance of winning doubles to two thirds, if she opts to change the door with her final decision, regardless of which door she had initially selected. While this solution can easily be derived in a formal way (e.g., by applying Bayes' theorem), it is counterintuitive to most people who fail to see why they should switch to another door once a goat is revealed; see, e.g., Friedman (1998).

90 Two trading schemes were employed successively: Second-price sealed bid auctions and oral double auctions.

91 Two traders are enough in their setting.

ket who did not err in individual judgments, market results turned out to be consistent with rational theory. Thus, competition among only a few bias-free traders was enough to correct the bias in market prices. Not surprisingly, the authors conclude that trading alone does not always ensure individual biases to be expunged from market prices; however, they present evidence for the conjecture that only a few unbiased agents is sufficient to collectively correct the bias, at least under the conditions of their market settings (also congruent with the "smart few hypothesis").

Coval and Shumway (2005) also contribute to the question whether behavioral biases affect prices. They analyze field data on Chicago Board of Trade (CBoT) proprietary traders and report strong evidence for behavioral biases (in particular: loss-aversion) among market makers.[92] While biased traders seem to exert a substantial influence on prices in the short run, these prices tend to be reverted more quickly than those generated by unbiased traders, indicating that the market is able to distinguish between trades initiated by biased traders and information-based trades. The authors conclude that price impacts that result from behaviorally biased traders are quickly removed by the healing power of arbitrage in a professional, highly liquid, and short-term environment.

---

[92] Concretely, they analyze trading positions of CBoT market makers on a daily basis and find that traders who suffered losses in the morning increase their afternoon risk to a greater extent than traders who accumulated gains by the end of the morning. Thus, losing market makers are willing to assume more risk in the afternoon to avoid losses in their daily settlement of accounts.

3 Study 1: Laboratory study

3.1 Basis treatment

3.1.1 Study design

The first study is designed to see whether partition-dependence occurs and persists in short-run experimental markets. In addition, effects expressed in probability judgments (both before and after trading) are compared with effects revealed by prediction-market trading prices. Furthermore, individual trading behavior is studied in greater detail to assess the quality and the validity of main results and to see whether (self-rated) competence of traders affects the extent of partition-dependence.

In April 2007, 192 undergraduate finance students (134 male, 58 female) were recruited from the University of Muenster (in Germany) to participate in one of twelve two-hour experimental trading sessions. To carry out the competence analysis (see subsection 3.1.4.1) students were asked to self-assess their general competence (scale 1–7) in the field of soccer (e.g. German Bundesliga) during the registration process (and they were asked for some more competence self-ratings in other domains to avoid problems with self-selection). Based on their self-reported soccer competence participants were grouped (without informing them) into high- and low-competence slots.[93] Each of the twelve sessions comprised 16 traders who were randomly divided into two self-contained groups (markets) with 8 traders in each. The sessions spanned one week and took place in a computerized laboratory environment where participants were separated from each other by dividers during the trading periods. The instructions (see Appendix I, translated from German) were handed out to all participants and were read out aloud by the experimenter to ensure that all information about the experiment was common knowledge.

[93] As noted, students were asked to self-rate their competence/knowledge on four other topics (scale 1–7) than soccer; these included: estimating distances, current political power structure, financial ratio analysis, and macroeconomic indicators. While soccer is directly related to one part of the experiment, the other four topics were just asked to avoid any self-selection bias in the students' willingness to participate. Mean (median) self-reported soccer competence was 4.25 (4.5). Students with a competence score of 1–4 (N_1=95) were assigned to the low-competence groups, students with a score of 5–7 (N_2=95) were assigned to the high-competence groups (two backup participants did not complete the questionnaire during the registration process).

The essential part of the experiment consisted of several trading rounds in a set of three "winner-take-all" contracts[94] on the occurrence of specific future events. Three mutually exclusive and collectively exhaustive events were defined for each market. Events that were used in the instructions, for instance, included intervals of the vote share received by the SPD (Social Democratic Party, a major German political party) in the next "Bundestag" elections. Concretely, the events considered in this example were the vote share v to fall into the following ranges: [$0\% \leq v \leq 29.9\%$], [$30.0\% \leq v \leq 34.9\%$], and [$35.0\% \leq v$]. Each of the three events was represented by a single "winner-take-all" asset. If an event occurred (did not occur), the asset that corresponded to that event would pay the owner 100 cents (0 cents) after the uncertainty about the outcome was resolved, i.e., after the election. Thus, exactly one of the three assets would pay 100 cents while the other two assets would expire worthless. By construction, since all events together cover the full state space but at the same time individual events do not overlap, a complete set of assets is certain to pay 100 cents. Such a complete bundle of all available assets is called a "unit portfolio" (unit pf) and could be traded with the experimenter at any time for 100 cents to allow arbitrage when the sum of state space-spanning prices is above or below 100 cents, and to create liquidity in stocks or assets. Recall that under some assumptions the prices of "winner-take-all" contracts can be directly interpreted to reflect market-aggregated probabilities that an event occurs, as discussed in subsection 2.1.4.3.

The experimental setting included trading in the following three event domains:[95] finance, sports, and weather. The relevant event partitions are displayed in Table 3.1. The finance related event domain refers to the closing price of the most important German stock market index (DAX) on the day two weeks after the experiment. In partition 1 (the low partition) the events are that the DAX index value is in the intervals [0, 7248.99], [7249, 7415.99], or 7416 and above (denoted [7416+]). In partition 2 (the high partition) the events are based on the intervals [0, 7415.99], [7416, 7563.99], and [7564+].

---

94 See subsection 2.1.1.3 for a definition of "winner-take-all" contracts in prediction markets.

95 In the following, the terms "event domain" and "stimulus" are used synonymously.

Table 3.1: Sets of asset partitions for the different treatments in Study 1.

Financial markets ---- Event: DAX-closing in 2 weeks		
Asset definition: An asset will pay 100 cents, if the outcome of the underlying event exactly fits the definition of the asset. The other two assets will expire worthless. Relevant for the payment of the assets in the financial markets domain is the Xetra-DAX closing (incl. final auction) in two weeks from today (i.e., on May 8, 2007). Xetra-DAX closing as of April 23, 2007 was 7335.62.		
	Partition 1	Partition 2
Asset 1	DAX.[0 - 7248.99]	DAX.[0 - 7415.99]
Asset 2	DAX.[7249 - 7415.99]	DAX.[7416 - 7563.99]
Asset 3	DAX.[7416+]	DAX.[7564+]

Weather ---- Event: Maximum Temperature in Münster at the end of May		
Asset definition: An asset will pay 100 cents, if the outcome of the underlying event exactly fits the definition of the asset. The other two assets will expire worthless. Relevant for the payment of the assets in the weather domain is the maximum temperature two meters above ground-level (abbreviation: TX) at the weather center Muenster/Osnabrueck (station no.: 10315) on May 31st, 2007 in degrees Celsius determined by Germany's National Meteorological Service ("Deutscher Wetterdienst").		
	Partition 1	Partition 2
Asset 1	Temp.[up to 15.9]	Temp.[up to 19.9]
Asset 2	Temp.[16.0 - 19.9]	Temp.[20.0 - 23.9]
Asset 3	Temp.[20.0+]	Temp.[24.0+]

Sports ---- Event: Total number of goals scored on the 34th game day of German 1. Bundesliga (Season 2006/2007)		
Asset definition: An asset will pay 100 cents, if the outcome of the underlying event exactly fits the definition of the asset. The other two assets will expire worthless. Relevant for the payment of the assets in the sports domain is the total number of goals scored on the final game day of German 1. Bundesliga in the season 2006/2007 (soccer/men).		
	Partition 1	Partition 2
Asset 1	Goals.[0 - 20]	Goals.[0 - 25]
Asset 2	Goals.[21 - 25]	Goals.[26 - 30]
Asset 3	Goals.[26+]	Goals.[31+]

The weather outcome refers to the maximum temperature in Muenster on May 31, 2007, approximately one month after the experiments. The sports outcome is the total number of goals scored by all teams of German "Bundesliga" on the final game day of the current soccer season, three to four weeks after the experiments. The weather partitions are [-, 15.9], [16.0, 19.9], [20.0+] (low partition) and [-, 19.9], [20.0, 23.9], [24.0+] (high partition). The sports partitions are [0, 20], [21, 25], [26+] (low partition) and [0, 25], [26, 30], [31+] (high partition).

As becomes clear from the fact that two partitions of the state space were designed for each event domain, the main treatment variable is the way in which the state space is divided (i.e. partitioned) into events. In a between-subject design, participants in different markets were randomly assigned to trade one of the two different partitions of the state space for each event domain. In order to eliminate the possibility that partition-dependence is driven by information conveyed by the presented partition, both partitions were explicitly described to all participants in the instructions.[96] This guaranteed that each participant did have the same information about the two different partitions and thus ensured that participants did not infer any likelihood information from the events they traded.

As Figure 3.1 that was not shown to the subjects illustrates, to create these partitions, each state space was initially divided into four disjoint and exhaustive intervals (I_1 to I_4). In each partition two of the adjacent intervals were combined to form a single asset. In partition 1 (the low partition) the upper two intervals were combined (forming an asset $a_{1,3}$ with interval denoted $I_3 \cup I_4$), and the lower two intervals were traded separately (I_1 and I_2). In partition 2 (the high partition) the lower two intervals were combined (forming an asset $a_{2,1}$ with interval denoted $I_1 \cup I_2$), and the upper two intervals were traded separately (I_3 and I_4). Both partitions therefore have three separate events. Note that by construction, the first asset in partition 2 ($a_{2,1}$), which refers to the interval $I_1 \cup I_2$, is a fusion of the first two assets $a_{1,1}$ (I_1) and $a_{1,2}$ (I_2) in partition 1. The third asset in partition 1 ($a_{1,3}$), which refers to the interval $I_3 \cup I_4$, is a fusion of assets $a_{2,2}$ (I_3) and $a_{2,3}$ (I_4) in partition 2.[97] Thus, from a rational point of view these combinations are

---

96 This refers to the "credibility" account discussed in subsection 2.2.3.

97 To make the point crystal clear, to the extent that any information is conveyed by the partitions described in the instructions, it is that the experimenter thinks that the dividing point between intervals I_2 and I_3 is special (perhaps it divides the state space into regions of relatively equal expected likelihood). However, because there is no informational asymmetry between conditions, partition-dependence cannot be rationalized on the basis of information conveyed to participants by the partitions presented.

equally likely to occur and hence, should differ neither in terms of judged probabilities nor in terms of prediction market prices.[98]

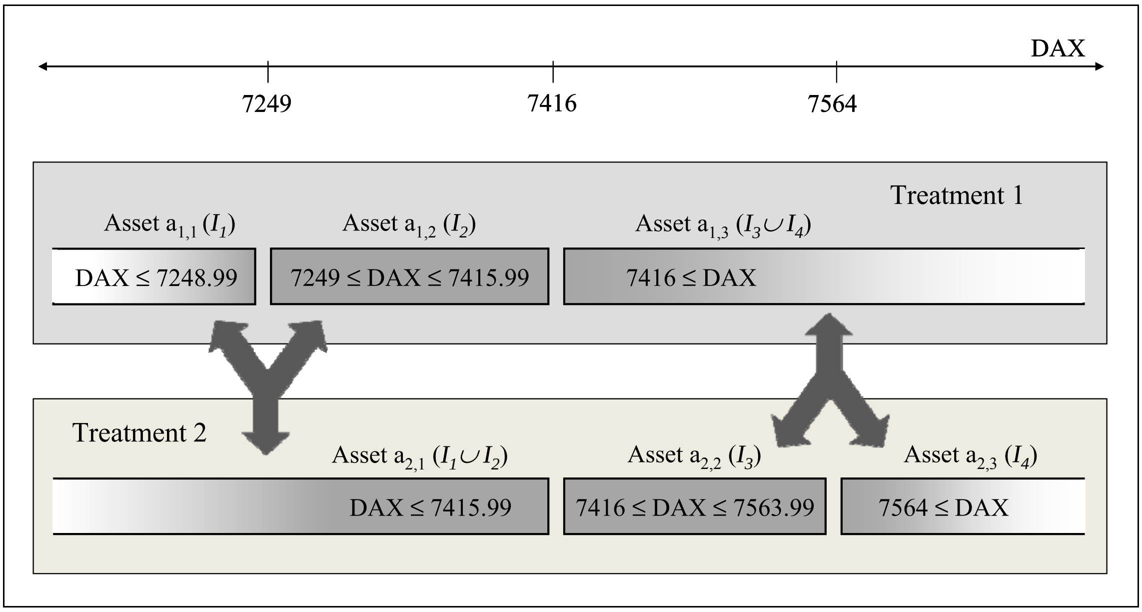


Figure 3.1: Construction of assets for two DAX partitions.

For the sports and weather event domains the interval boundaries were chosen rather arbitrarily based on historical outcomes, so there is no conclusive way to link probabilities expressed by individual judgments or inferred from market prices to objective probabilities. However, for the finance DAX event domain, the four intervals were created from historical data: Given the previous DAX closing price, and the recent short-term historical volatility of the DAX, it is possible to calculate the expected probability density function (PDF) for the DAX close two weeks in the future. The calculation is based on a random-walk theory for stocks and stock index price movements and is consistent with efficient market theory (see, e.g., Hull (2009, pp. 260-272) for a textbook version): The stochastic process usually used to describe stock (index) price movements is called a geometric Brownian motion (GBM). For each infinitesimal small period of time the GBM assumes (i) normally distributed returns and (ii) independent

[98] This experiment corrected a flaw in an earlier pilot study, in which the boundaries of adjacent closed intervals were the same (e.g., [*x*, *y*] and [*y*, *z*], so that *y* is included in both intervals). The use of closed intervals is a design flaw because if subjects are highly confident that the numerical value is *y*, then the sum of the interval prices for [*x*, *y*] and [*y*, *z*] could rationally be higher than the combined interval [*x*, *z*]. For the DAX stock market index (scaled in the hundredth) and for temperatures (rounded to the nearest 0.1°C) the overlap of the closed intervals at their common meeting point *y* is probably a minor problem, but for the sports outcomes (number of goals) it creates an interpretive problem. The results from the earlier pilot study and the one reported here, however, are quite similar.

returns for two non-overlapping intervals. Given this stochastic process the logarithmic future price of a stock (index) ln S_T is normally distributed with mean ln $S_0 + (\mu - \sigma^2/2)\cdot T$ and standard deviation $\sigma\sqrt{T}$:

$$\ln S_T \sim \Phi\left[\ln S_0 + \left(\mu - \frac{\sigma^2}{2}\right)T, \sigma\sqrt{T}\right] \tag{3.1}$$

Against this background, the interval boundaries were defined such that each of the four events represents a certain percentile of the expected PDF. The intervals $\{I_1, \ldots, I_4\}$ represent the following percentiles: {27.5%; 27.5%; 22.5%; 22.5%}. Consistently, intervals $I_1 \cup I_2$ and $I_3 \cup I_4$ represent a percentile of 55.0% and 45.0%, respectively. If prices of the "winner-take-all" contracts were guided by historical frequency, they should correspond to these percentiles. Since different experimental sessions were spread out over a week, the DAX intervals were adjusted for each experimental session (based on the recent DAX index close) to preserve the percentiles.[99] This day-by-day adjustment of interval boundaries enables (i) comparability and aggregation of the data from sessions on different dates and (ii) allows comparing expressed probabilities and market prices with the historical guesses. Traders were not told about the procedure for constructing and adjusting the intervals, since doing so would instruct subjects about expected probabilities and constitute an additional treatment effect. Since there was no such adjustment for the weather and sports events, if similar patterns are evident across all three event domains then there is no harmful influence from the day-to-day adjustment of the DAX intervals. No additional background knowledge was provided concerning the different event domains.[100]

For each of the three event domains two completely identical and independent trading rounds were run successively (i.e., repeated trials were run under a "stationary replication" condition), resulting in six trading rounds per participant and experimental group, as shown in Figure 3.3. Each trading round lasted ten minutes (with short breaks between rounds). The order in which the participants traded assets from the three event domains varied for each experimental session and was perfectly counterbalanced (i.e.,

[99] See Appendix II for a Table with session-individual DAX interval boundaries.

[100] However, participants were told the previous DAX close in the instructions to give them a clue of the current index level.

for each of the six possible event domain orders there were two experimental sessions) to avoid any order effects.[101] In each of the six trading rounds the participants were initially endowed with a combination of assets (i.e., unit portfolios spanning the set of assets) and cash, to the value of €20 in total.[102] Participants were compensated incentive-compatible based on their final cash and asset holdings for an afterwards randomly chosen trading round, at the point when the relevant uncertainty about the future outcome was resolved and asset payoffs (either 100 or 0 cents) became clear.

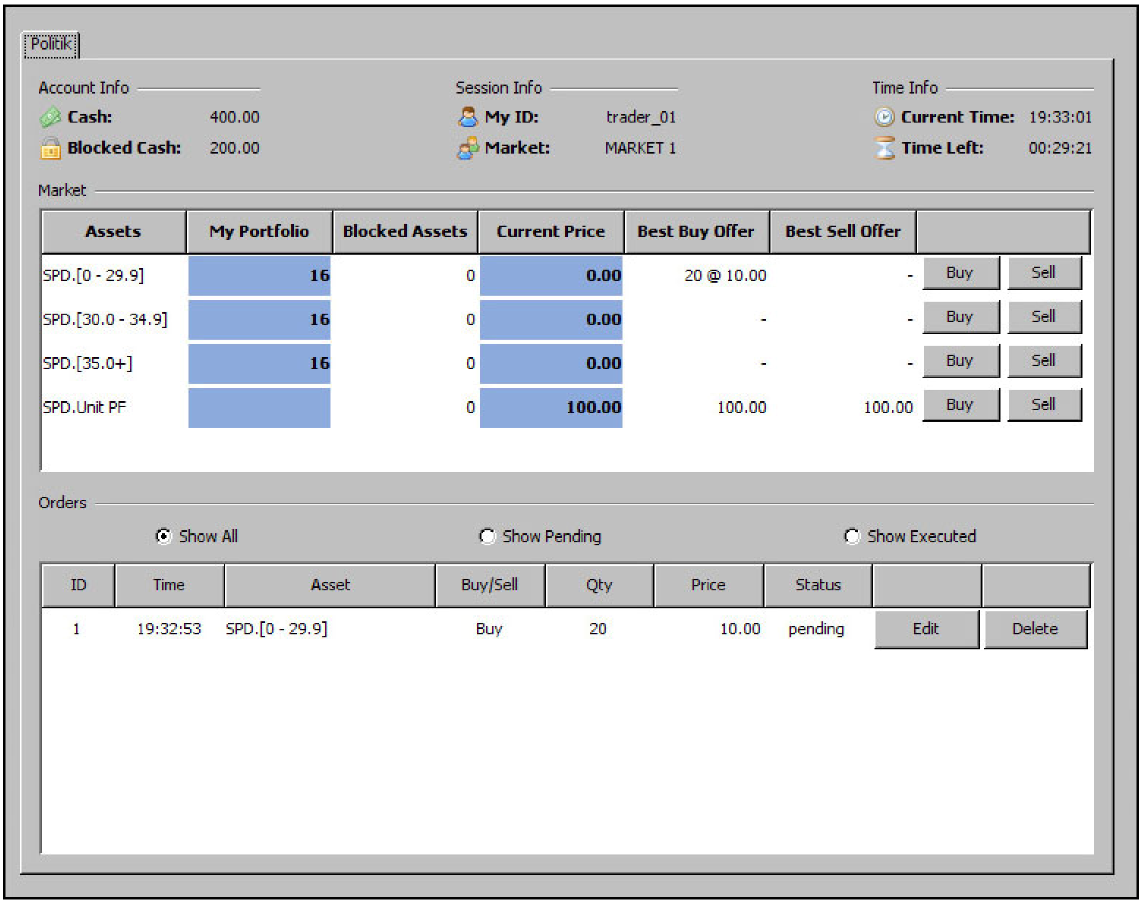


Figure 3.2: Screenshot of the trading application (practice market).[103]

[101] The six possible event domain orders are: (1) 2 × finance, 2 × weather, 2 × sports; (2) 2 × finance, 2 × sports, 2 × weather; (3) 2 × weather, 2 × finance, 2 × sports; (4) 2 × weather, 2 × sports, 2 × finance; (5) 2 × sports, 2 × finance, 2 × weather; and (6) 2 × sports, 2 × weather, 2 × finance.

[102] In each market (of eight participants) groups of two traders were randomly endowed with one of the four different combinations: 16 unit portfolios + 400 cents, 12/800, 8/1200, 4/1600, all of them representing a value of €20. For each trader her initial endowment was the same over the six trading rounds.

[103] The trading software was especially developed for the study. It is based on Java Runtime Environment technology and was set up on a web-based client-server structure. The graphical user interface (GUI) was divided into three areas: the account info area, the market area and the order history. Participants could submit, edit or cancel buy or sell orders via an order form. Orders were processed and executed by the system within split seconds; the trading screen updated in real-time.

Expected value of payment was €20 per person.[104] There was no credit line and no short selling, although traders could use unit portfolio transactions to short sell assets indirectly.[105] No explicit transaction costs were imposed for trading. The trading institution was a multi-unit continuous double auction (CDA) with a hidden order book. Subjects only saw the best bid and ask quotes for each asset (see Figure 3.2 for a screenshot of the trading software). Participants could submit bid and ask quotes for each asset simultaneously, so they could act as effective market makers. Trading took place only among the eight traders that were assigned to the same market;[106] in particular they could not trade across markets with different partitions. During instruction and a practice trading round, participants were provided hints on the handling of the trading software and were told how to exploit arbitrage opportunities (within the market) by buying or selling unit portfolios with the experimenter for cash.

Before the first trading round for each event domain, and after the second (and final) trading round, the participants were asked to provide their individual probability judgments for the occurrence of the three events they traded. These judgments were not incentivized. Participants were also asked in advance for their self-rated competence in making such probability judgments in the domain of the specific stimulus (scale 1 [incompetent] to 7 [very competent]). At the end of the session they were further asked to provide some personal information like age, self-rated knowledge in the field of statistics and econometrics, or trading experience. Some earlier studies have compared individual judgments of probabilities (as often elicited or inferred from psychology experiments) with probabilities expressed by market trades (see subsection 2.3.3). Like those studies, one question the present method can address is whether partition-dependence is expressed by individual judgments, and whether it is moderated by the bundle of institutional and learning properties of markets. Figure 3.3 illustrates one of the six possible courses of an experimental session:[107]

---

[104] The minimum possible payoff per participant was €0, the maximum possible payoff was €160 (which equals the amount that was paid in total to the eight participants of a market). Hence, the total amount of €320 was paid to the 16 participants in each experimental session.

[105] Buying the unit portfolio at 100 cents and selling one of the three assets afterwards is equivalent to short selling that asset. The only difference is that in the latter case a margin of 100 cents (i.e., the worst possible payoff for the shorted asset) is imposed until the final value of the assets is determined. Alternatively, a trader could directly buy the two complement assets (without buying the unit portfolio before) which is equivalent to short selling an asset, too, as long as the prices for all assets sum to 100 cents.

[106] The only exception was trading the unit portfolio which was always executed immediately against the experimenter.

[107] As mentioned above, the order in which the three event domains were traded was perfectly counterbalanced for the twelve session slots.

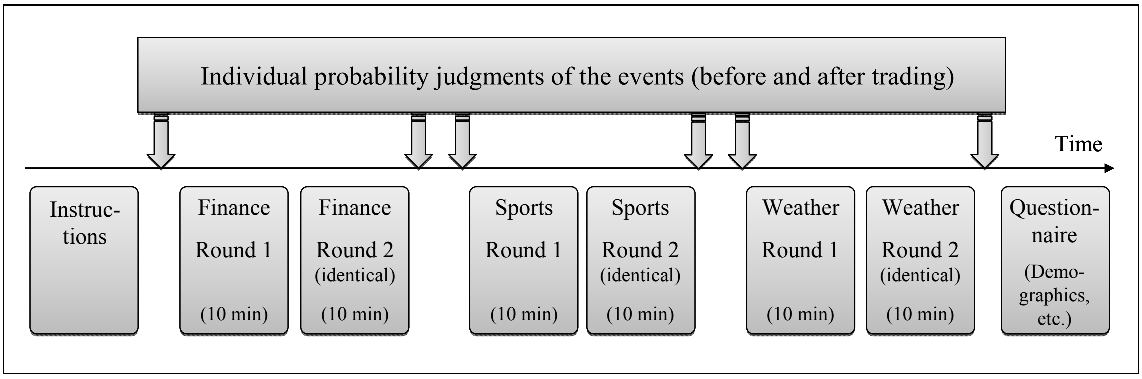


Figure 3.3: Example of the time course of an experimental session.

To summarize, judgments and experimental trading data was gathered from twelve session slots (each of which consisted of two self-contained markets with eight traders per market) with six trading rounds each (three different event domains × two identical and independent trials per event domain), resulting in 144 trading rounds (twelve sessions × six trading rounds per session × two treatments). Half of these markets (72) offered assets from partition 1, the other half offered assets from partition 2. Partitions were randomly assigned to the markets for each event domain, so it was possible, for instance, that a market of eight participants traded assets of partition 1 in the finance and the weather event domain, and traded assets of partition 2 in the sports domain (consequently, the other market in that session traded the complement partitions).

3.1.2 Descriptive analysis

Some general facts on trading activity and market efficiency are presented in the following. The total number of trades (except unit portfolio trades) in the 144 trading rounds was 6,167. Accordingly, the average number of trades per 10-minutes trading round was 42.83, an average of 14.28 for each of the three assets, and total shares traded were about 140 in each market. Table 3.2 presents a breakdown of the total number of trades across event domains and session slots (but pooled across event partitions). It turns out that trades were distributed quite evenly across the three stimuli with some more trades occurring in the second trading round of each event domain (6.7% more trades on average). It also appears that trading happened to be fairly balanced across the twelve session slots, though some slots being more active (like Slot 3 which accounts for 10.3% of all trades) than others (like Slot 9 which accounts for only 6.9% of all trades).

Table 3.2: Breakdown of the total number of trades.

	Number of Trades										
	Finance			Sports			Weather			Total	Percent
	Round 1	Round 2	Total	Round 1	Round 2	Total	Round 1	Round 2	Total		
Slot 1	73	84	157	80	97	177	78	86	164	498	8.1%
Slot 2	94	108	202	90	89	179	79	80	159	540	8.8%
Slot 3	115	95	210	91	100	191	117	115	232	633	10.3%
Slot 4	78	105	183	74	81	155	85	71	156	494	8.0%
Slot 5	71	83	154	81	103	184	71	71	142	480	7.8%
Slot 6	77	76	153	76	77	153	74	79	153	459	7.4%
Slot 7	74	81	155	107	94	201	83	76	159	515	8.4%
Slot 8	83	81	164	82	99	181	63	94	157	502	8.1%
Slot 9	69	70	139	75	70	145	70	71	141	425	6.9%
Slot 10	72	64	136	93	79	172	84	86	170	478	7.8%
Slot 11	103	110	213	83	100	183	77	80	157	553	9.0%
Slot 12	89	113	202	81	110	191	91	106	197	590	9.6%
Total	998	1,070	2,068	1,013	1,099	2,112	972	1,015	1,987	6,167	100.0%

A similarly balanced picture is reflected in the average trading volume per trade, with an average volume of 3.39 assets per trade. Table 3.3 provides some summary statistics on the number of trades per market, the average number of assets per trade, and the total volume per market across event domains.

Table 3.3: Trading volume statistics.

	No. of trades per market			Average No. of assets per trade			Total assets traded per market		
	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.
Finance	21	43.08	62	1.93	3.37	5.42	81	144.17	284
Sports	25	44.00	66	2.07	3.31	4.37	78	143.92	210
Weather	26	41.40	64	2.31	3.49	5.90	85	143.48	242
All	21	42.83	66	1.93	3.39	5.90	78	143.85	284

In addition, trading was relatively continuous across the 10-minute trading period: The fraction of trades that occurred in each one-minute interval averaged between 8.5 and 11.8% (the former in the first, the latter in the second trading minute). As can be seen from Table 3.4, trading pace shows very similar patterns across all three event domains and trading rounds: within a small range of fluctuation, the pattern shows increasing trading activity at the beginning of each trading period that reaches a peak in the second and third minute. Afterwards, trading activity calms slightly down before another step-up occurs during the last two minutes. This suggests that 10 minutes of trading is an appropriate period per trial for two reasons: on the one hand, there is enough time for traders to "calm down" after some minutes of high frequency trading, so they can think thoroughly about prices and their positions; on the other hand, the final increase of trading activity implies that participants are still attentive and eager to trade at the end of the trading period.

Table 3.4: Temporal distribution of trades in a 10-minute trading period.

	Fraction of Trades (in Percent)									
	Finance			Sports			Weather			Total
	Round 1	Round 2	Total	Round 1	Round 2	Total	Round 1	Round 2	Total	
Minute 1	8.3%	9.3%	8.8%	12.8%	5.6%	9.1%	5.9%	9.0%	7.4%	8.5%
Minute 2	11.6%	11.7%	11.7%	12.0%	12.1%	12.1%	12.4%	10.9%	11.7%	11.8%
Minute 3	10.4%	11.7%	11.1%	11.2%	11.6%	11.4%	12.8%	11.5%	12.1%	11.5%
Minute 4	10.4%	10.8%	10.6%	8.3%	11.3%	9.8%	8.5%	10.3%	9.5%	10.0%
Minute 5	9.5%	10.1%	9.8%	10.1%	11.2%	10.7%	10.0%	11.3%	10.7%	10.4%
Minute 6	9.0%	9.7%	9.4%	8.6%	9.0%	8.8%	9.1%	10.3%	9.7%	9.3%
Minute 7	9.8%	10.0%	9.9%	9.7%	9.2%	9.4%	9.3%	8.8%	9.0%	9.5%
Minute 8	9.0%	6.4%	7.6%	9.0%	9.6%	9.3%	9.1%	8.8%	8.9%	8.6%
Minute 9	10.2%	9.3%	9.7%	7.9%	9.5%	8.7%	9.9%	9.9%	9.9%	9.4%
Minute 10	11.6%	11.1%	11.4%	10.5%	10.9%	10.7%	13.2%	9.2%	11.1%	11.1%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

It is also helpful to look at the volatility of trade prices to get a more comprehensive view of trading activity. The mean standard deviation of trade prices per market and asset is 4.99 (based on all trades) and is reduced by nearly half to 2.75 if only the last three trades are considered (N=432).[108] In general, mean volatility is slightly reduced in the second trading round: 5.77 in the first compared to 4.20 in the second trading round (based on all trades) and 3.09 in the first compared to 2.40 in the second trading round (based on the last three trades). The mean volatility of asset prices does not differ systematically across event domains. Mean standard deviation of trade prices seems to be higher for assets of partition 1 (at least for the finance and the weather stimulus) and this difference is indeed statistically significant (1% level by a Kruskal-Wallis test) based on all trade prices, but is not statistically significant for only the last three prices (based on the same test). Table 3.5 and Table 3.6 summarize the mean standard deviation of trade prices.

[108] N=432 results from 144 trading rounds × 3 assets/trading round.

Table 3.5: Mean standard deviation of trade prices (based on all trades).

	Mean Standard Deviation of Trade Prices (Based on all Trades)								
	Trading Round 1				Trading Round 2				Total
	Finance	Sports	Weather	Total	Finance	Sports	Weather	Total	
Treatment 1									
$P(I_1)$	6.57	5.72	5.53	5.94	5.32	3.63	3.16	4.04	4.99
$P(I_2)$	9.17	4.80	5.79	6.59	4.59	4.21	6.42	5.07	5.83
$P(I_3 \cup I_4)$	5.92	5.13	7.22	6.09	6.50	4.91	5.46	5.62	5.86
Total	7.22	5.22	6.18	6.21	5.47	4.25	5.01	4.91	5.56
Treatment 2									
$P(I_1 \cup I_2)$	3.91	5.74	4.93	4.86	2.71	4.12	1.84	2.89	3.87
$P(I_3)$	5.29	6.10	3.45	4.95	3.24	3.72	3.83	3.59	4.27
$P(I_4)$	6.96	6.64	4.99	6.20	4.45	4.23	3.28	3.99	5.09
Total	5.39	6.16	4.46	5.34	3.47	4.02	2.98	3.49	4.41
Total	6.30	5.69	5.32	5.77	4.47	4.13	4.00	4.20	4.99

Table 3.6: Mean standard deviation of trade prices (based on last three trades).

	Mean Standard Deviation of Trade Prices (Based on Last Three Trades)								
	Trading Round 1				Trading Round 2				Total
	Finance	Sports	Weather	Total	Finance	Sports	Weather	Total	
Treatment 1									
$P(I_1)$	2.07	2.65	1.95	2.23	1.35	1.88	1.04	1.42	1.82
$P(I_2)$	3.51	3.95	2.32	3.26	3.04	4.30	3.44	3.59	3.43
$P(I_3 \cup I_4)$	4.11	3.85	2.92	3.63	4.39	2.09	3.54	3.34	3.48
Total	3.23	3.48	2.40	3.04	2.93	2.76	2.67	2.79	2.91
Treatment 2									
$P(I_1 \cup I_2)$	2.25	3.28	3.20	2.91	2.08	1.78	1.63	1.83	2.37
$P(I_3)$	4.20	5.57	0.88	3.55	1.47	1.74	2.72	1.97	2.76
$P(I_4)$	3.07	3.00	2.89	2.99	2.64	1.40	2.66	2.23	2.61
Total	3.17	3.95	2.32	3.15	2.06	1.64	2.34	2.01	2.58
Total	3.20	3.72	2.36	3.09	2.49	2.20	2.51	2.40	2.75

One approach to assess market efficiency is to look at the occurrence and exploitation of arbitrage opportunities. As noted above, buying or selling the unit portfolio could have been used to exploit arbitrage opportunities within a market rapidly. An arbitrage opportunity exists if the sum of one, two or three bid quotes (buy quota) is above 100 cents which is called a *bid arbitrage* opportunity. In that case an arbitrageur could buy a unit portfolio at 100 cents from the experimenter and sell the constituent assets separately to the market for more than 100 cents. Another arbitrage opportunity exists if the three ask quotes (sell quota) sum to less than 100 cents which is called an *ask arbitrage* opportunity. In that case an arbitrageur could buy a full set of assets for less than 100 cents in the market and sell them jointly as a unit portfolio for 100 cents to the experimenter. In the present experimental setting, these arbitrage strategies are not perfectly riskless since the possibility of block trades is not implemented in the trading software, and hence, sequentially executing the necessary trades will take (at least) a few seconds. This bears the risk that arbitrage-permitting quotes are removed from the system or arbitrage is exploited by some other arbitrageur in the meantime. In part, this can explain why arbitrage opportunities were not exploited in any case. Table 3.7 and Table 3.8 present the details of some descriptive statistics on arbitrage opportunities in this study.[109]

Table 3.7: Summary statistics on bid arbitrage opportunities.

	Bid Arbitrage											
	Trading Round 1				Trading Round 2				Total			
Panel A: No. of markets (10-min. trading round) with no arbitrage opportunities (N=144 trading rounds)												
Finance	3				3				6			
Sports	3				2				5			
Weather	2				7				9			
Total	8				12				20			
Panel B: Total arbitrage period [seconds] per market (10-min. trading round), (N=144 trading rounds)												
	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.
Finance	0.00	47.30	84.01	296.11	0.00	37.36	70.46	305.72	0.00	37.36	77.24	305.72
Sports	0.00	72.87	85.25	210.18	0.00	59.68	75.20	230.82	0.00	62.63	80.22	230.82
Weather	0.00	50.90	82.71	436.26	0.00	46.02	75.24	375.25	0.00	48.93	78.98	436.26
Total	0.00	55.40	83.99	436.26	0.00	46.87	73.64	375.25	0.00	50.71	78.81	436.26
Panel C: Time period [seconds] until exploitation												
	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.
Finance	1.16	15.46	26.88	141.87	1.07	8.43	21.14	208.87	1.07	12.13	23.92	208.87
Sports	1.22	16.18	29.65	133.64	1.05	12.42	19.00	126.36	1.05	12.75	23.48	133.64
Weather	1.23	9.95	28.77	369.73	1.00	12.59	24.74	188.38	1.00	11.45	26.70	369.73
Total	1.16	12.84	28.39	369.73	1.00	11.45	21.38	208.87	1.00	12.24	24.62	369.73
Panel D: Time-weighted amount [cents] per arbitrage occurrence (N=461 occurences)												
	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.
Finance	0.49	5.60	11.85	119.00	0.25	3.03	6.77	50.00	0.25	5.00	9.22	119.00
Sports	0.62	8.38	10.54	44.99	0.01	2.55	5.21	51.00	0.01	4.44	7.45	51.00
Weather	0.05	5.00	6.94	33.00	0.62	3.00	3.77	15.40	0.05	3.91	5.31	33.00
Total	0.05	5.60	9.84	119.00	0.01	3.00	5.29	51.00	0.01	4.10	7.39	119.00

[109] Note that a bid quote of 0 cents (any asset could be notionally sold for 0 cents) and an ask quote of 100 cents (any asset could be purchased as part of the unit portfolio for 100 cents) is assumed for missing bid and ask quotes, respectively. By consequence, the bid arbitrage condition may be fulfilled even if a bid quote is missing for one or two assets (e.g., the bid quote for one asset is missing but is 50 and 65 cents for the other two assets, respectively). On the other hand, the ask arbitrage condition cannot be fulfilled if at least one ask quote is missing. Note further that arbitrage opportunities lasting for less than one second were excluded from the calculations for technical reasons.

Table 3.8: Summary statistics on ask arbitrage opportunities.

	Ask Arbitrage											
	Trading Round 1				Trading Round 2				Total			
Panel A: No. of markets (10-min. trading round) with no arbitrage opportunities (N=144 trading rounds)												
Finance	13				10				23			
Sports	15				17				32			
Weather	15				17				32			
Total	43				44				87			
Panel B: Total arbitrage period [seconds] per market (10-min. trading round), (N=144 trading rounds)												
	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.
Finance	0.00	0.00	22.95	106.84	0.00	6.19	28.15	188.90	0.00	2.89	25.55	188.90
Sports	0.00	0.00	20.75	268.31	0.00	0.00	31.47	395.86	0.00	0.00	26.11	395.86
Weather	0.00	0.00	16.46	225.88	0.00	0.00	37.85	484.85	0.00	0.00	27.16	484.85
Total	0.00	0.00	20.05	268.31	0.00	0.00	32.49	484.85	0.00	0.00	26.27	484.85
Panel C: Time period [seconds] until exploitation												
	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.
Finance	1.85	10.05	20.40	101.43	1.24	9.68	21.80	185.09	1.24	9.86	21.15	185.09
Sports	1.85	10.08	33.20	151.67	1.40	15.61	29.05	334.24	1.40	13.08	30.57	334.24
Weather	1.05	6.69	21.95	180.26	1.08	20.77	47.81	455.34	1.05	9.58	35.23	455.34
Total	1.05	9.52	24.06	180.26	1.08	12.42	30.78	455.34	1.05	10.83	27.82	455.34
Panel D: Time-weighted amount [cents] per arbitrage occurrence (N=136 occurences)												
	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.	Min.	Median	Mean	Max.
Finance	0.20	3.50	4.19	17.00	0.38	2.00	3.24	13.77	0.20	3.00	3.68	17.00
Sports	0.50	1.94	2.49	6.36	0.50	2.65	4.42	15.63	0.50	2.00	3.71	15.63
Weather	0.01	1.55	2.57	7.94	0.81	3.56	8.72	37.00	0.01	3.20	5.73	37.00
Total	0.01	2.13	3.28	17.00	0.38	3.00	5.01	37.00	0.01	3.00	4.25	37.00

Panel A of Table 3.7 and Table 3.8 shows a breakdown of the total number of markets (i.e., 10-minute trading rounds) with no arbitrage opportunities at all. In total, out of 144 markets, 20 (13.9%) do not show bid arbitrage opportunities at all and 87 (60.4%) have no ask arbitrage opportunities. Panel B contains a breakdown of summary statistics (minimum—median—mean—maximum) on the total duration of arbitrage opportunities (in seconds) per market. Bid arbitrage occasions sum to a median of 50.71 seconds per 10-minute trading round (i.e., 8.5% of the total trading time). The respective median value for ask arbitrage is 0 since the majority of markets does not show any ask arbitrage opportunities (as reported above). The mean value is somewhat higher for bid arbitrage (78.81 seconds) and is 26.27 seconds for ask arbitrage due to some outlier markets. Panel C addresses the question of how long does it take until an arbitrage opportunity disappears. This question is important because market efficiency can be assumed to be high if existing arbitrage opportunities are exploited rapidly. The median (mean in parentheses) duration of a bid arbitrage event is 12.24 (24.62) seconds after the opportunity first appeared, the corresponding values for ask arbitrage are 10.83 (27.82) seconds. These periods are relatively short for both bid and ask arbitrage occasions given that block trades were not allowed. Panel D finally sheds light on how meaningful arbitrage opportunities were in terms of exploitable profits. The median (mean in parentheses) time-weighted amount of arbitrage profits per arbitrage opportunity was only 4.10 (7.39) cents for bid arbitrage and 3.00 (4.25) cents for ask arbitrage; so even when arbitrage opportunities exist they are small. Given the four metrics of Panels A–D, it appears that neither there is an obvious pattern of arbitrage opportunities across the three event domains nor for the two trading rounds per event domain (figures for bid arbitrage look somewhat "better" for the second trading round, figures for ask arbitrage, though, look slightly "better" for the first trading round).

Table 3.9: Time-weighted average of bid and ask sums.

	Time-weighted Average Sum of Quotes	
	Bid	Ask
Trading Round 1		
Finance	84.57	126.84
Sports	87.05	123.15
Weather	87.89	122.70
Average	86.46	124.05
Trading Round 2		
Finance	90.52	119.74
Sports	90.32	119.83
Weather	91.03	115.64
Average	90.57	118.40
Total average	88.72	121.16

To get an impression of the aggregate bid-ask-spread, Table 3.9 presents the time-weighted average of bid and ask sums, i.e. the sum of all bid and ask quotes on average, respectively.[110] The data is homogeneous across session slots and event domains. However, the average sum of quotes is closer to 100 cents for the second trading round of an event domain (bid sum: 90.57 vs. ask sum: 118.40 cents) compared to the first trading round (bid sum: 86.46 vs. ask sum: 124.05 cents). Pooling data yields a

[110] Periods with missing bid/ask quotes and periods during which the sum of all bid (ask) quotes was above (below) 100 cents (i.e., fulfilling the bid/ask arbitrage condition) were excluded from this analysis to be as conservative as possible. Including arbitrage periods improves the time-weighted average bid (ask) sum to 93.32 (119.47) cents for the pooled data, respectively.

time-weighted average bid sum of 88.72 cents and a time-weighted average ask sum of 121.16 cents. This result suggests that the markup[111] is on average higher for the ask side than for the bid side of market quotes.

To summarize, the descriptive analysis of trading data reflects a balanced view of high trading activity among participants. Trading is quite evenly distributed across session slots and event domains with some more activity taking place in the second of the two consecutive trading rounds. Trading within a trading round reflects a regular pattern suggesting that ten minutes of trading is an appropriate period per trial. The standard deviation of trade prices reflects a "stimulating discussion" among traders about "correct" prices, and is slightly reduced in the second trading round. Volatility of the last three trade prices, in turn, is reasonably small to assume that they are close to equilibrium prices. With respect to the market efficiency of the markets, there was a number of bid arbitrage opportunities (i.e., the market was paying too much for a full set of assets) and a few ask arbitrage opportunities (i.e., the market was asking too less for a full set of assets). However, in most cases arbitrage opportunities were removed after a short period of time (either by exploitation or by cancelation of quotes), and were modest in terms of possible arbitrage profits. Thus, market efficiency can generally be rated high (except for a few outliers): on the one hand, the occurrence of arbitrage opportunities shows that the sum of quotes is generally oscillating around 100 cents (as confirmed by the time-weighted average sum of bid and ask quotes); on the other hand, the fast disappearance of arbitrage opportunities is evidence for market forces being strong enough to eliminate *obvious* irrationalities.

3.1.3 Main results

3.1.3.1 Judged probabilities

The main hypotheses and results refer to the focal treatment variable, i.e., the different partitions into which the state space was divided. The hypotheses are led by the conjecture that subjective probability judgments as well as equilibrium market prices are partition-dependent and biased toward the ignorance prior of $1/N$, thus differ-

[111] The "markup" is defined as the difference between the sum of all bid or ask quotes and the price for the unit portfolio (e.g., 100 cents) and is sometimes also referred to as "overround". It is usually used as a measure of efficiency and competitiveness in (winner-take-all) betting markets.

ing systematically across partitions. Consistent with support theory, judged probabilities as well as market prices are expected to be subadditive, i.e. to be higher for an event if it is unpacked into two separately evaluated intervals than if it is evaluated as a whole.[112] It is further conjectured that market forces may be in part able to diminish partition-dependence; the bias is thus expected to be more pronounced in individually judged probabilities than in probabilities derived from market prices.

At first, the pre-trading and post-trading judged probabilities $p_{J,i}$ will be considered for the different event domains i (judgments are required, by instruction, to sum to 1.0 across the exhaustive set of events).[113] The notation $p_{J,i}(I_1)$ and $p_{J,i}(I_2)$ refer to the judged probability of unpacked intervals I_1 and I_2, respectively (as in partition 1 in Figure 3.1); and $p_{J,i}(I_1 \cup I_2)$ refers to the judged probability of the single packed interval which is the union of intervals I_1 and I_2 (as in partition 2). Accordingly, $p_{J,i}(I_3)$ and $p_{J,i}(I_4)$ refer to the judged probability of unpacked intervals I_3 and I_4, respectively (as in partition 2 in Figure 3.1); and the notation $p_{J,i}(I_3 \cup I_4)$ is used to denote the judged probability of the packed interval which is the union of intervals I_3 and I_4 (as in partition 1). In the present between-subjects design, the main treatment effect for individual judgments is analyzed by comparing the sum of probability judgments for intervals I_1 and I_2 (partition 1) to the judgment for the interval $I_1 \cup I_2$ (partition 2), and analogously by comparing the sum of probability judgments for intervals I_3 and I_4 (partition 2) to the judgment for interval $I_3 \cup I_4$ (partition 1). Hypothesis 3.1 states the partition-dependence prediction for intervals I_1 and I_2 (H_0(a)) and intervals I_3 and I_4 (H_0(b))[114]. The null hypothesis of no partition-dependence is that $p_{J,i}(I_1) + p_{J,i}(I_2) = p_{J,i}(I_1 \cup I_2)$ and $p_{J,i}(I_3) + p_{J,i}(I_4) = p_{J,i}(I_3 \cup I_4)$.

Hypothesis 3.1:

H_0(a): $p_{J,i}(I_1) + p_{J,i}(I_2) > p_{J,i}(I_1 \cup I_2)$ and

H_0(b): $p_{J,i}(I_3) + p_{J,i}(I_4) > p_{J,i}(I_3 \cup I_4)$

---

[112] See subsection 2.2.4. for a presentation of support theory.

[113] Pre-trading (or: before-trading) refers to probability judgments elicited *before* the *first* trading round of each event domain (ex ante). Post-trading (or: after-trading) probability judgments were collected from each participant *after* the *second* trading round of each event domain (ex post).

[114] Hypothesis H_0(b) is reported solely to correspond to later market prices presentation. Note that H_0(b) is, in fact, redundant for judged probabilities, since the results (i.e., effect size) have to be the same for both hypotheses by construction.

with $p_{J,i}(I_k)$ = judged probability for interval I_k of event domain i,
i = {(f)inance, (s)ports, (w)eather},
$k_{partition\_1} = \{1, 2, 3 \cup 4\}$; $k_{partition\_2} = \{1 \cup 2, 3, 4\}$

Table 3.10 shows the average and the median pre-trading individual probability judgments surveyed *before* the *first* trading round of the finance, sports and weather event domains (N=96 participants in each of the two partitions):[115]

Table 3.10: Pre-trading individual probability judgments.

Treatment	Pre-Trading Individual Judgment	Event Domain					
		Finance (N_1=N_2=96)		Sports (N_1=N_2=96)		Weather (N_1=N_2=95)	
		Mean	Median	Mean	Median	Mean	Median
1	$p(I_1)$	0.219	0.200	0.279	0.250	0.144	0.120
1	$p(I_2)$	0.497	0.500	0.398	0.400	0.333	0.320
	$p(I_1)+p(I_2)$	0.717	0.700	0.678	0.700	0.477	0.500
2	$p(I_1 \cup I_2)$	0.405	0.400	0.417	0.400	0.199	0.200
	PD difference	*0.312*	*0.300****	*0.261*	*0.300****	*0.278*	*0.300****
2	$p(I_3)$	0.397	0.400	0.378	0.360	0.349	0.350
2	$p(I_4)$	0.198	0.200	0.205	0.200	0.451	0.400
	$p(I_3)+p(I_4)$	0.595	0.600	0.583	0.600	0.801	0.800
1	$p(I_3 \cup I_4)$	0.283	0.300	0.322	0.300	0.523	0.500
	PD difference	*0.312*	*0.300****	*0.261*	*0.300****	*0.278*	*0.300****

The mean (median in parentheses) difference between summed probabilities of unpacked intervals and the packed interval (e.g., $p_{J,i}(I_1) + p_{J,i}(I_2) - p_{J,i}(I_1 \cup I_2)$) is .312 (.300), .261 (.300) and .278 (.300) for the finance, sports and weather events, respectively. All reported differences are statistically highly significant, based on a Kruskal-Wallis test ($p<.0001$).[116] In line with hypothesis 3.1 the results reveal that participants

[115] In the Table, N_1 refers to participants in partition 1 and N_2 refers to participants in partition 2. For the weather events, one set of judgments in each partition is missing due to technical errors, thus N_1=N_2=95.

[116] The Kruskal-Wallis test is a non-parametric method testing the hypothesis that C samples with N_i observations in the ith sample come from the same population. The test is suited to detect differences among the population means. It is closely related to a one-way analysis of variance, but does not assume a normal distribution. The Kruskal-Wallis test for two samples (like in the present context) is essentially the same as a Wilcoxon rank-sum test (for unmatched data) or the Mann-Whitney two-sample statistic; see Kruskal and Wallis (1952).

on average attribute more probability to an event if it is represented by two out of three intervals than if it is represented by only one out of three intervals. The results provide strong evidence for partition-dependence in ex ante judged probabilities across all event domains.

Hypothesis 3.1 can also be tested for post-trading judged probabilities. Table 3.11 shows the average and the median post-trading individual probability judgments collected *after* the *second* trading round of the finance, sports and weather event domains (again, N=96 participants in each of the two partitions):[117]

Table 3.11: Post-trading individual probability judgments.

Treatment	Post-Trading Individual Judgment	Event Domain					
		Finance (N_1=N_2=96)		Sports (N_1=N_2=96)		Weather (N_1=96; N_2=95)	
		Mean	Median	Mean	Median	Mean	Median
1	$p(I_1)$	0.205	0.200	0.252	0.250	0.116	0.100
1	$p(I_2)$	0.494	0.500	0.432	0.400	0.307	0.300
	$p(I_1)+p(I_2)$	0.699	0.700	0.684	0.700	0.422	0.400
2	$p(I_1 \cup I_2)$	0.442	0.400	0.428	0.400	0.196	0.200
	PD difference	*0.257*	*0.300****	*0.256*	*0.300****	*0.226*	*0.200****
2	$p(I_3)$	0.382	0.400	0.403	0.400	0.352	0.350
2	$p(I_4)$	0.176	0.160	0.169	0.150	0.452	0.400
	$p(I_3)+p(I_4)$	0.558	0.600	0.572	0.600	0.804	0.800
1	$p(I_3 \cup I_4)$	0.301	0.300	0.316	0.300	0.578	0.600
	PD difference	*0.257*	*0.300****	*0.256*	*0.300****	*0.226*	*0.200****

The mean (median in parentheses) difference between summed probabilities of unpacked intervals and the packed interval (e.g., $p_{J,i}(I_1) + p_{J,i}(I_2) - p_{J,i}(I_1 \cup I_2)$) is .257 (.300), .256 (.300), and .226 (.200) for the finance, sports and weather events, respectively. Like the pre-trading judged probabilities, all reported differences are statistically highly significant, based on a Kruskal-Wallis test (p<.0001).

[117] In the Table, N_1 refers to partition 1 and N_2 refers to partition 2. For the weather events, one set of judgments is missing in partition 2 due to a technical error, thus N_2=95.

3.1.3.2 Equilibrium market prices

Of course, probability judgments elicited from individuals might reflect thoughtless errors (particularly since those judgments are not incentivized) which are strongly or weakly diminished in two 10-minute trading periods. As discussed above, markets are, after all, a kind of dollar-activity-weighted opinion polls that also provide substantial time for reflection and opportunities for learning from others. Thus, probabilities inferred from market prices are analyzed in the following.

Figure 3.4 to Figure 3.6 show the aggregated development of asset prices over time for the finance, sports and weather event domains. This means, at each point of time, the relevant market prices (and sums of prices for unpacked interval-assets) are averaged over all twelve (identical) session slots. Figuratively speaking, price charts from the twelve session slots were superimposed and current market prices were averaged at each point of time. Note that price charts for the first and the second trading round are strung together (separated by a vertical line after 600 seconds). These figures shed light on the question of how market prices and partition-dependence, on average, develop *during* the trading rounds and *over* the two identical trials, i.e., how much variation can be observed over the 20 minutes of trading in each event domain. The upper chart of each figure refers to the intervals I_1 and I_2, and the lower chart refers to the intervals I_3 and I_4. In both charts of each figure, the lower path shows the average price for the packed asset and the upper path shows the sum of prices for the corresponding unpacked assets.[118] The gap between the two curves shows the size and persistence of the partition-dependence.

[118] Accordingly, the red curves result from partition 1 markets and the blue curves result from partition 2 markets.

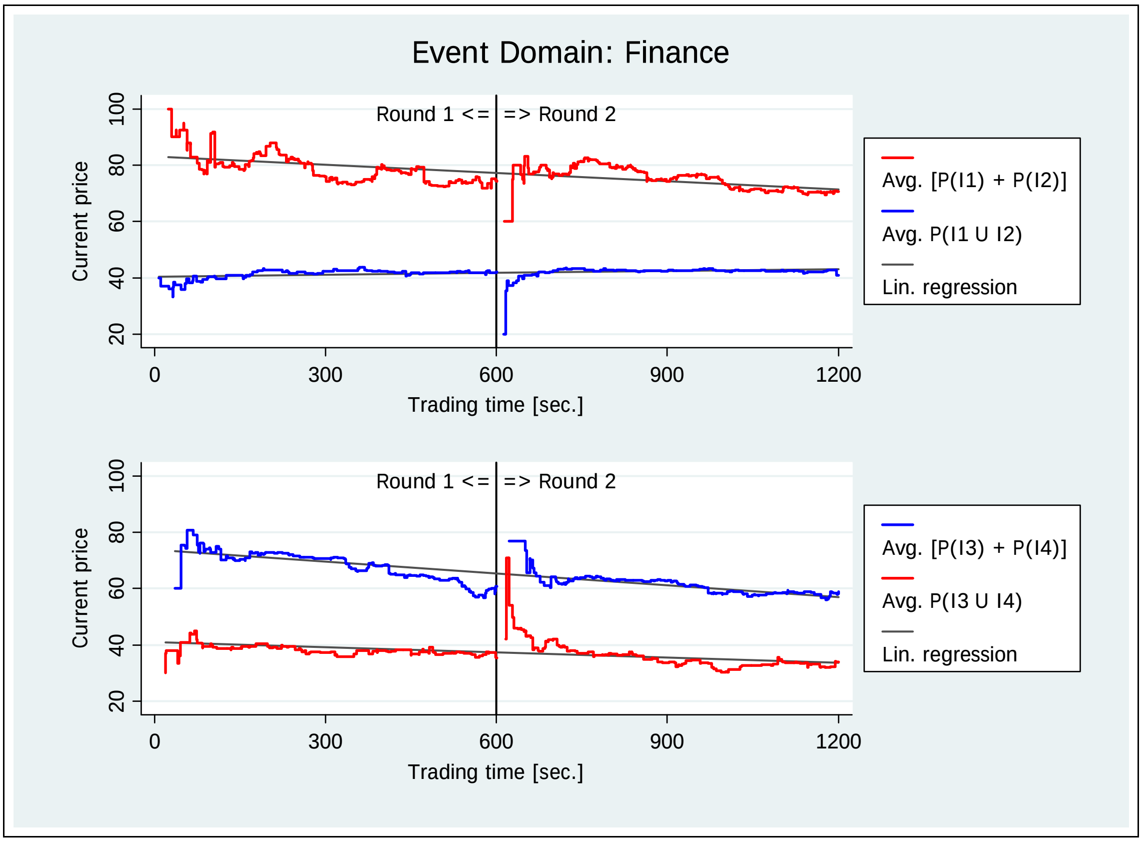


Figure 3.4: Development of price differences over time for the finance assets.

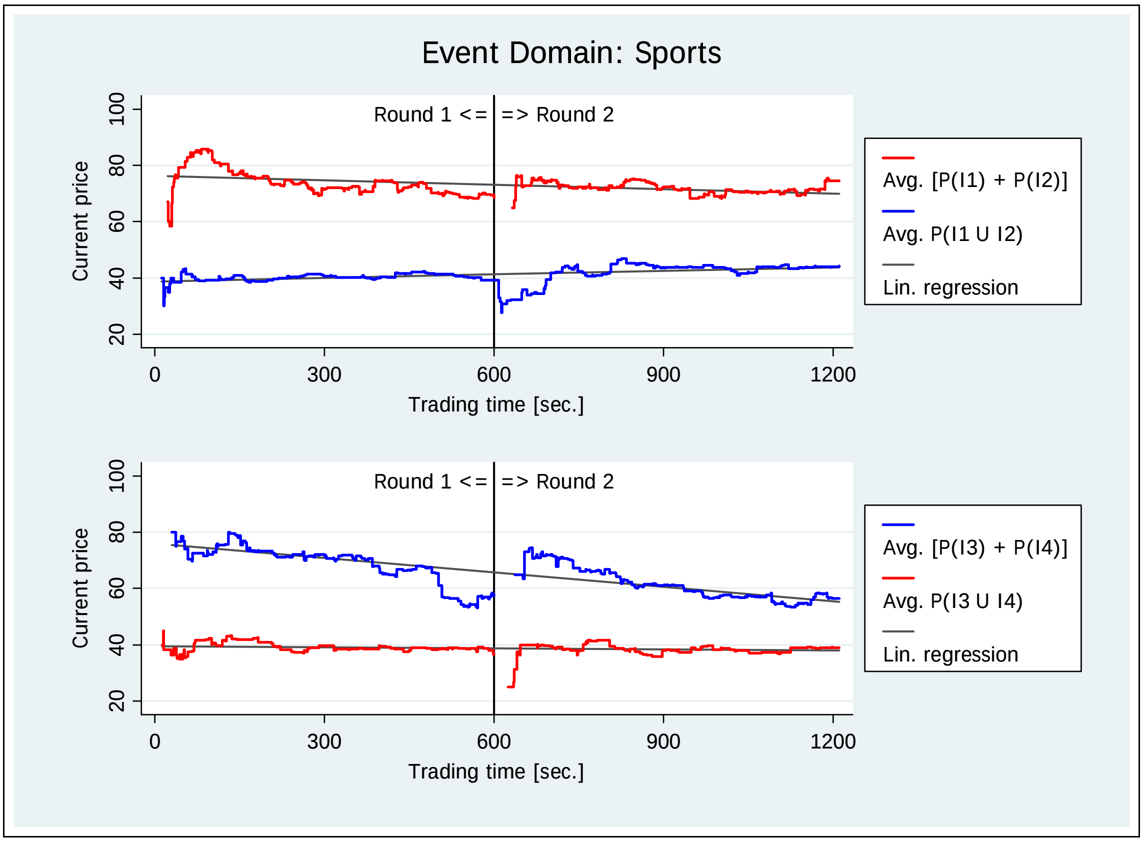


Figure 3.5: Development of price differences over time for the sports assets.

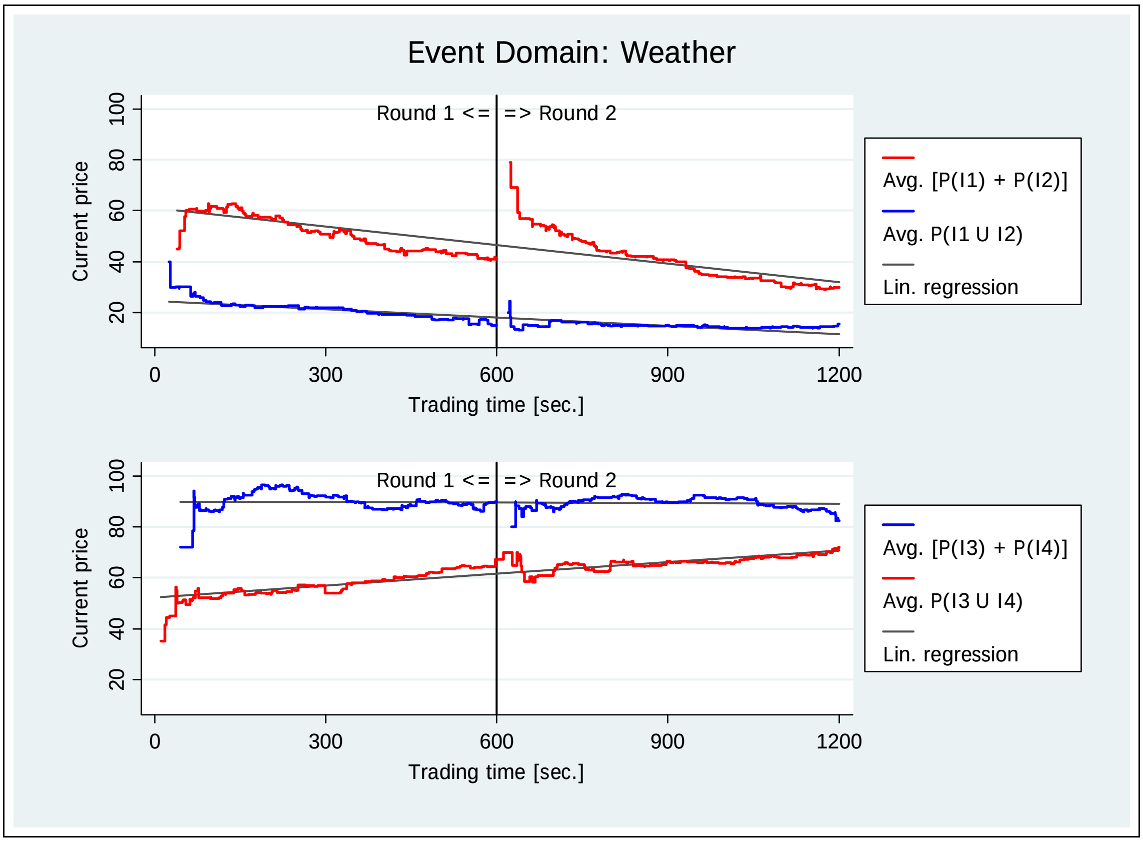


Figure 3.6: Development of price differences over time for the weather assets.

For a formal analysis, linear regressions are estimated for the price paths, and their slopes are compared which gives a rough measure of price convergence over time. Calculate the differences of slopes as a simple measure of convergence over time:

$$\beta_{i,P(I_1)+P(I_2)} - \beta_{i,P(I_1 \cup I_2)} \text{ and } \beta_{i,P(I_3)+P(I_4)} - \beta_{i,P(I_3 \cup I_4)} \tag{3.2}$$

with $\beta_{i,P(I_k)}$ denoting the slope coefficient from a linear regression of the average asset prices $P(I_k)$ over trading time for event domain i.

Results from the convergence measure analysis can be obtained from Table 3.12. A negative (positive) slope coefficient indicates decreasing (increasing) prices over time. If the slope difference is negative (positive), this indicates convergence (divergence) of market prices. The more negative the slope difference, the more do prices from differently partitioned markets converge.

Table 3.12: Measures of price convergence over time.

	Slope Coefficient β from Linear Regression		
	Finance	Sports	Weather
Avg. $[P(I_1)+P(I_2)]$	-0.0096	-0.0053	-0.0243
Avg. $P(I_1 \cup I_2)$	0.0023	0.0042	-0.0109
Slope difference	-0.0119	-0.0095	-0.0134
Avg. $[P(I_3)+P(I_4)]$	-0.0140	-0.0171	-0.0007
Avg. $P(I_3 \cup I_4)$	-0.0061	-0.0012	0.0156
Slope difference	-0.0079	-0.0159	-0.0163
Convergence measure (mean slope difference)	-0.0099	-0.0127	-0.0149

As Table 3.12 reveals, a slight tendency of convergence can be observed in both of the intervals for all three event domains, with the weather domain showing the most and the finance domain showing the least convergence. The finding that there is some slow convergence motivates the later experiments with much longer time periods (Study 2). Despite the fact that there is some convergence of market prices in this lab study, the price difference between the sum of unpacked-interval assets and the packed-interval asset remains large in magnitude and partition-dependence is well pronounced during the whole trading period.

In the following, a more formal analysis of price differences between the two treatments is presented. This analysis concentrates on equilibrium market prices. For each market, define the "equilibrium market price" $P_i^*(I_k)$ to be the quantity-weighted average of the last three trade prices (prices at which trades were executed, not bids and asks) in the second trading round for the interval I_k asset (the focus is on the data from the second trading round here, as prices from this round were less noisy and hence more reliable; however, it turned out that data from the two trading rounds did not differ systematically). For each event domain the experiment generates twelve equilibrium prices per asset for each partition.[119] The hypotheses of partition-dependence in equilibrium market prices are parallel to those above for judgments:

[119] Recall that the whole population of 192 participants was assigned to markets of eight traders each, resulting in 24 trading groups which further divided into twelve markets of each partition.

<u>*Hypothesis 3.2:*</u>

H_0(a): $P_i^*(I_1) + P_i^*(I_2) > P_i^*(I_1 \cup I_2)$ and

H_0(b): $P_i^*(I_3) + P_i^*(I_4) > P_i^*(I_3 \cup I_4)$

with $P_i^*(I_k)$ = equilibrium market price for an asset that refers to interval I_k of event domain i,

i = {(f)inance, (s)ports, (w)eather},

$k_{partition\_1} = \{1, 2, 3 \cup 4\}$; $k_{partition\_2} = \{1 \cup 2, 3, 4\}$

Figure 3.7 and Figure 3.8 show the distribution of equilibrium market prices (or summed equilibrium market prices) for the low intervals (I_1 and I_2 in Figure 3.7) and the high intervals (I_3 and I_4 in Figure 3.8) for the two partition groups (N=12 equilibrium prices per event domain and treatment).[120] Look at Figure 3.7 for the low intervals first: for the finance event domain, equilibrium market prices for the packed-event assets $P_i^*(I_1 \cup I_2)$ range between 15.00 and 60.13 cents (partition 2), while the sums of equilibrium prices for the unpacked-event assets $P_i^*(I_1) + P_i^*(I_2)$ range from 48.01 to 96.96 cents (partition 1). The corresponding range of values for the sports (weather in parentheses) event domain is 17.20 to 69.00 cents (1.37 to 31.20 cents) for prices $P_i^*(I_1 \cup I_2)$ and 58.47 to 90.99 cents (18.03 to 52.17 cents) for the sum of prices $P_i^*(I_1) + P_i^*(I_2)$. Red squares in the figures indicate median values. Figure 3.8 illustrates respective values for the high intervals I_3 and I_4: for the finance event domain, equilibrium market prices for the packed-event assets $P_i^*(I_3 \cup I_4)$ range from 13.75 to 54.50 cents (partition 1), while the sums of equilibrium prices for the unpacked-event assets $P_i^*(I_3) + P_i^*(I_4)$ range from 28.59 to 84.53 cents (partition 2). The range of values for the sports (weather in parentheses) event domain is 16.00 to 66.67 cents (51.89 to 88.00 cents) for prices $P_i^*(I_3 \cup I_4)$ and 33.13 to 81.74 cents (56.18 to 99.83 cents) for prices $P_i^*(I_3)$ + $P_i^*(I_4)$. As hypothesized, it becomes clear by visual inspection that equilibrium market prices for intervals I_1 and I_2 (I_3 and I_4) are systematically higher in partition 1 (partition 2), where these intervals were traded separately.

[120] Figure 3.7 and Figure 3.8 refer to the second trading round of each event domain.

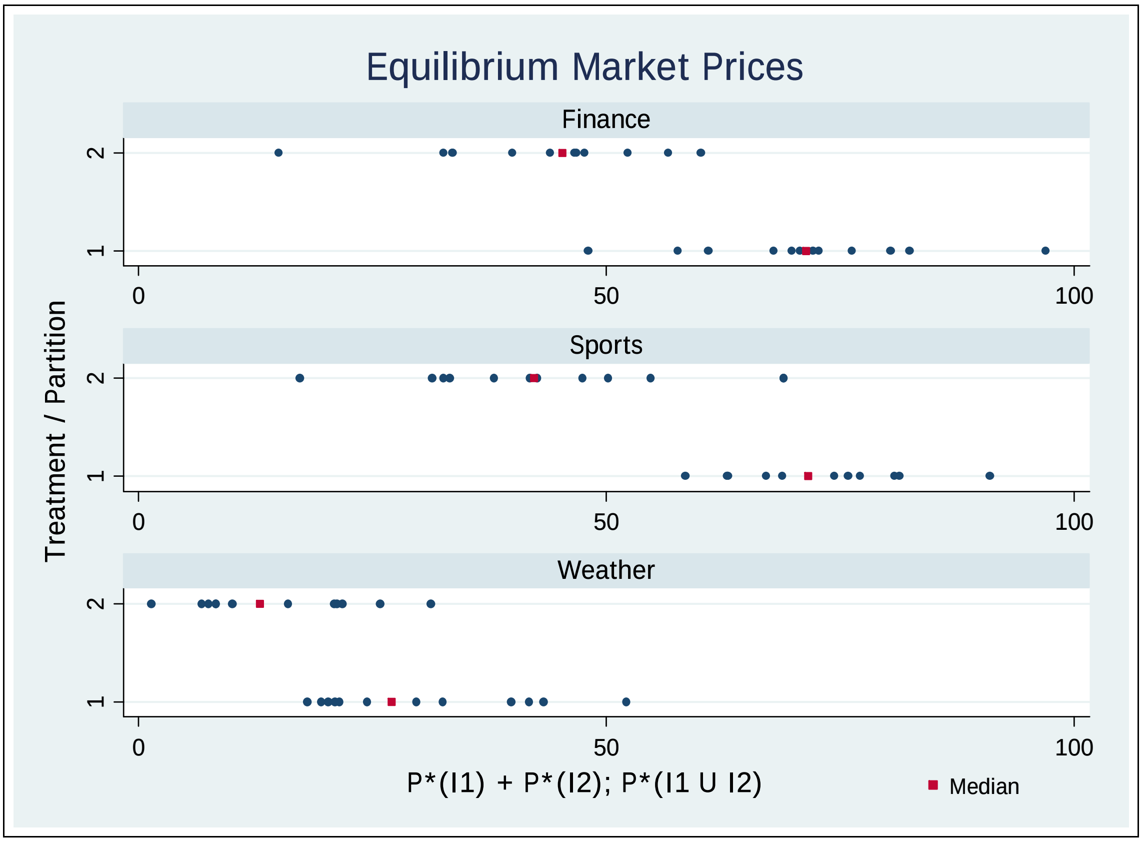


Figure 3.7: (Summed) equilibrium market prices for the two partition markets (intervals I_1 and I_2).

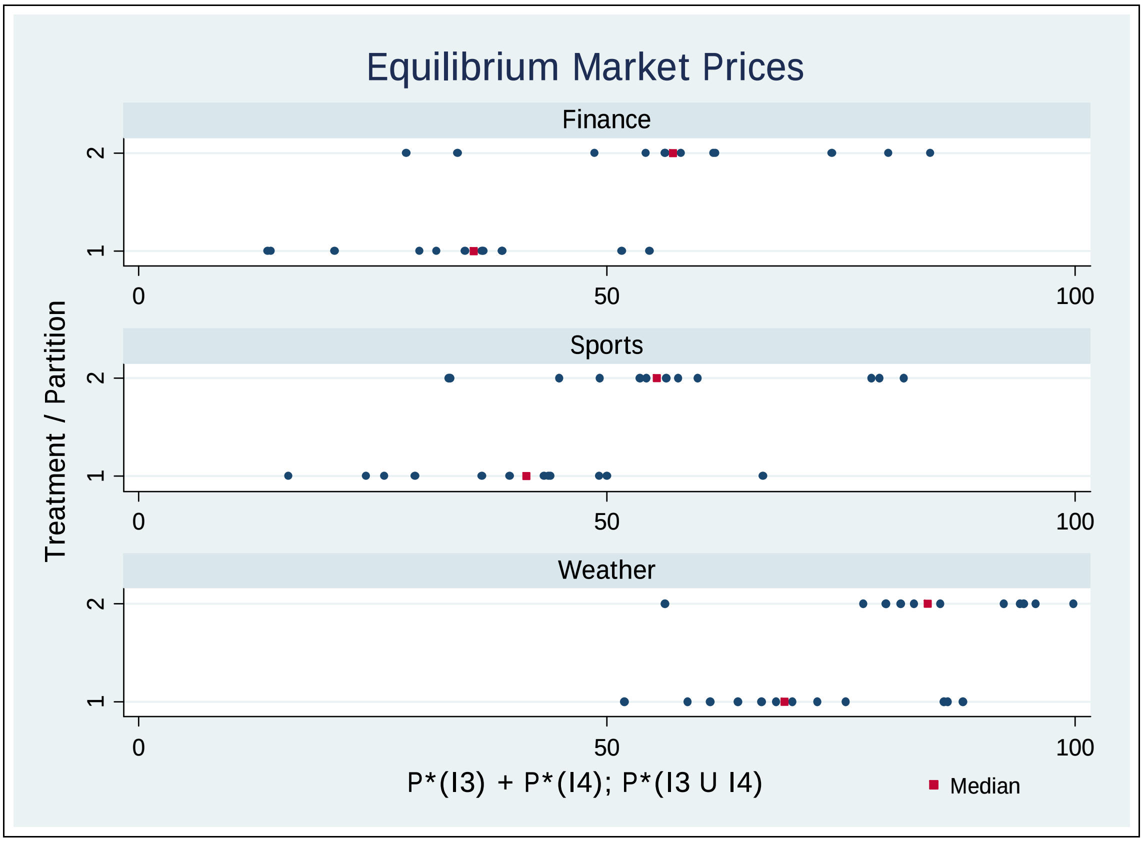


Figure 3.8: (Summed) equilibrium market prices for the two partition markets (intervals I_3 and I_4).

For the low intervals of the sports events, for example, equilibrium market prices from the two treatments hardly overlap, although both, as in all other cases, refer to the same possible outcomes of the underlying variable (in this case, both prices refer to the event that [0, 25] goals would be scored on the final game day).

Aggregating the data, Table 3.13 shows the mean prices (divided by 100 cents to make them comparable to judged probabilities) for the three assets of each partition. For comparison the average *judgments* collected before and after trading are also reported (see Table 3.10 and Table 3.11).

Table 3.13: Mean equilibrium market prices (2nd trading round) and individual judgments (pre-trading and post-trading).

Treatment		Mean Judged Probability/Equilibrium Prices								
		Finance			Sports			Weather		
		Pre-Trading Individual Judgment	Equilibrium Market Prices (Round 2)	Post-Trading Individual Judgment	Pre-Trading Individual Judgment	Equilibrium Market Prices (Round 2)	Post-Trading Individual Judgment	Pre-Trading Individual Judgment	Equilibrium Market Prices (Round 2)	Post-Trading Individual Judgment
1	I_1	0.219	0.152	0.205	0.279	0.230	0.252	0.144	0.048	0.116
1	I_2	0.497	0.561	0.494	0.398	0.490	0.432	0.333	0.256	0.307
	I_1+I_2	0.717	0.713	0.699	0.678	0.720	0.684	0.477	0.303	0.422
2	$I_1 \cup I_2$	0.405	0.424	0.442	0.417	0.439	0.428	0.199	0.149	0.196
	PD difference	*0.312*	*0.289*	*0.257*	*0.261*	*0.281*	*0.256*	*0.278*	*0.154*	*0.226*
2	I_3	0.397	0.404	0.382	0.378	0.416	0.403	0.349	0.354	0.352
2	I_4	0.198	0.177	0.176	0.205	0.152	0.169	0.451	0.496	0.452
	I_3+I_4	0.595	0.581	0.558	0.583	0.568	0.572	0.801	0.850	0.804
1	$I_3 \cup I_4$	0.283	0.336	0.301	0.322	0.391	0.316	0.523	0.707	0.578
	PD difference	*0.312*	*0.245*	*0.257*	*0.261*	*0.177*	*0.256*	*0.278*	*0.143*	*0.226*
Average PD difference			***0.267***			***0.229***			***0.149***	

From Table 3.13 it turns out that the difference between summed equilibrium prices (divided by 100) of unpacked assets, $P_i^*(I_1) + P_i^*(I_2)$, and the packed asset, $P_i^*(I_1 \cup I_2)$, is .289, .281 and .154 for the finance, sports and weather markets, respectively, all of them showing in the expected direction (i.e., indicating subadditivity). The corresponding values for the high intervals, $P_i^*(I_3) + P_i^*(I_4) - P_i^*(I_3 \cup I_4)$, are .245, .177 and .143, respectively. Averaging the magnitude of partition-dependence in equilibrium market prices for the low and the high intervals yields an average effect size of .267, .229 and .149 for the finance, sports and weather domains, respectively, reflecting strong partition-dependence.[121] A similar Table showing differences in medians can be obtained from Appendix III. Results are very similar to those derived from means. Although not marked in the Table, all reported differences are statistically highly significantly different from zero, based on a Kruskal-Wallis test ($p<.01$). Besides, the results are very robust with respect to alternative definitions of the equilibrium market prices: using only the last trade price, for instance, or calculating an unweighted average of the last three trade prices hardly changes the results; Kruskal-Wallis tests are equipollent under all variations and thus there is no reason to suspect that the effect (and its size) just depends on the concrete measure of equilibrium market prices.

To summarize the main results so far, it can be said that partition-dependence is strongly pronounced in subjective probability judgments elicited before and after trading (effect size between .226 and .312). Although there is some slight convergence of market prices over the 20 minutes of trading, equilibrium market prices (at the end of the second trading round) exhibit well pronounced partition-dependence (effect size between .149 and .267). As it seems, the market forces in the present short-run lab study are not able to drive out the bias from market prices. The next subsection aims to further explore the extent to which individually judged probabilities affect market prices and vice versa.

3.1.3.3 Comparison of market prices to individual probability judgments

A relevant question is whether partition-dependence expressed in market equilibrium prices is as much pronounced as in individual probability judgments. One may argue that market forces, if not able to eliminate the bias, are at least able to mitigate partition-dependence by (i) the exchange of information through the price mechanism and/or (ii) by the activity of some (marginal) traders who drive market prices in a direction that is consistent with rational theory. For a theoretical discussion of this question and a review of existing empirical results, see subsections 2.1.4.4 and 2.3.3.

To address this issue with data from the present lab study, the two self-contained markets from each session slot (offering assets from the different partitions) were notionally matched. In the present between-subjects design this allows calculating parti-

[121] Note that the partition difference is not necessarily the same for the low and high intervals in market prices, in contrast to the judgments where this difference has to be the same for the low and high intervals due to the fact that, by instructions, they always summed to 1.0.

tion-dependence per session slot as the difference between the sum of the two unpacked events (from one partition) and the corresponding packed event (from the other partition). Note that this hypothetical matching procedure is somewhat arbitrary, since participants in one treatment traded completely independently from participants in the other treatment. However, this procedure is the only way to test for differences in partition-dependence between before-trading individual judgments and market outcomes. And matching the two markets from the same session slot (compared to any other matching rule) ensures that the same overall conditions prevail for both groups (like most recent DAX index level, weather conditions outside the lab or lab room temperature, and recent news on the soccer teams). Equilibrium market prices again are calculated as the volume-weighted average of the last three trade prices that occurred in the second trading round; and probability judgments are averaged across the eight traders who build a market. This allows calculating the partition bias for both before-trading judgments and market prices in each of the twelve session slots resulting in twelve paired data points per event domain. Then, the null hypothesis of no differences in bias strength between before-trading judgments and equilibrium market prices can be tested. Hypotheses H_0(a) and H_0(b) are specified below:

Hypothesis 3.3:

H_0(a): $P_i^*(I_1) + P_i^*(I_2) - P_i^*(I_1 \cup I_2) = p_{J,i}(I_1) + p_{J,i}(I_2) - p_{J,i}(I_1 \cup I_2)$ and

H_0(b): $P_i^*(I_3) + P_i^*(I_4) - P_i^*(I_3 \cup I_4) = p_{J,i}(I_3) + p_{J,i}(I_4) - p_{J,i}(I_3 \cup I_4)$

with $P_i^*(I_k)$ = equilibrium market price for an asset that refers to interval I_k of event domain i,

$p_{J,i}(I_k)$ = judged probability for interval I_k of event domain i,

i = {(f)inance, (w)eather, (s)ports}.

$k_{partition\_1} = \{1, 2, 3 \cup 4\}$; $k_{partition\_2} = \{1 \cup 2, 3, 4\}$

For the finance event domain, session-based partition-dependence in equilibrium market prices is lower than partition-dependence inferred from before-trading individual judgments in 4 out of 12 cases for the low intervals (and in 8 out of 12 cases for the high intervals), but hypotheses H_0(a) and H_0(b) cannot be rejected by a Wilcoxon matched-pairs signed-ranks test (p=.5829 for $H_0(a)$ and p=.3465 for $H_0(b)$).[122] These results suggest that the partition bias in equilibrium market prices is not systematically different (in neither direction) from the partition bias in before-trading individual probability judgments. In general, this holds true for the sports event domain (bias reduction of market prices in 5 out of 12 cases for the low intervals (p=.5303) and in 8 out of 12 cases for the high intervals (p=.0995). For the weather event domain, in contrast, hypotheses $H_0(a)$ and $H_0(b)$ can be rejected (bias reduction of market prices in 9 out of 12 cases for the low intervals (p=.0186) and in 10 out of 12 cases for the high intervals (p=.0186), indicating that the partition bias is somewhat lower in market prices compared to pre-trading individual probability judgments. Adding the results from Table 3.13 (in section 3.1.3.2), it can be concluded that market-price partition-dependence is slightly reduced compared to before-trading probability judgments for finance and sports event domains, and cut in half for weather events. However, conservative tests using session-level data only show a statistically significant reduction for the weather event domain.

While partition-dependence was analyzed independently in subsection 3.1.3.1 for pre-trading and post-trading individual probability judgments, an obvious question is whether individual judgments elicited after the second trading round of each event domain are influenced by market activity. Thus, post-trading judgments are compared to the judgments collected before the first trading round. As can be seen from Table 3.13 partition-dependence in after-market judged probabilities is .257, .256 and .226 for the finance, sports and weather event domains, respectively, so the bias of partition-dependence is still well pronounced. Compared to the before-market judged probabilities, this is a decrease of –.055 (–17.6%), –.005 (–1.9%) and –.052 (–18.7%) in the finance, sports and weather event domains, respectively. As a result, the bias is slightly reduced for the finance and the weather domain, but is hardly changed for the sports stimulus. Using before-trading and after-trading individual probability judgments of the events in all three event domains allows to perform a powerful within-subjects test that analyzes whether participants, in general, adjusted their ex post judgments systemati-

[122] The Wilcoxon signed-ranks test (Wilcoxon (1945)) is a non-parametric alternative to the paired Student's *t*-test of two related samples or repeated measurements on a single sample. Like the *t*-test, the Wilcoxon test involves comparisons of differences between measurements, so it requires that the data are measured at an interval level of measurement. However, the test does not require assumptions about the concrete shape of the distribution of measurements. It should thus be used whenever the distributional assumptions that underlie the *t*-test cannot be satisfied.

cally and, if they were doing so, in which direction. For each subject the difference between the after- and the before-trading probability judgments is analyzed for the packed intervals.[123] Since the general direction of the bias suggests the judgments for the packed intervals are too low, define a subject's judgments to reflect a bias reduction due to trading experience if this difference is positive, i.e. if post-trading judgments for the packed intervals are higher than pre-trading judgments.[124]

Hypothesis 3.4:

$$H_0: \quad \frac{1}{3}\cdot\sum_{i=1}^{3}\left(p_{J,s,i}^{after}(I_x \cup I_{x+1}) - p_{J,s,i}^{before}(I_x \cup I_{x+1})\right) = 0$$

with $p_{J,s,i}(I_x \cup I_{x+1})$ = subject s's judged probability (after/before trading) for the packed interval $I_x \cup I_{x+1}$ of event domain i,

i = {(f)inance, (s)ports, (w)eather}

$I_x \cup I_{x+1} = I_1 \cup I_2$ or $I_3 \cup I_4$, depending on subject s's assignment to one of the two treatments in event domain i.

Averaging the differences for the three event domains *per subject*, as described in hypothesis 3.4 above, values are positive in 52.1% of the cases (100 out of 192 subjects), and negative in 34.4% of the cases (66 out of 192). Trading does not influence the remaining 26 subjects, on average, in any direction.[125] A conservative sign test shows that the effect is significant (p<.01, one-sided test). Overall, the two 10-minute trading periods apparently have a modest de-biasing effect on individual judgments, though a substantial degree of partition-dependence remains after 20 minutes.

Testing hypothesis 3.4 separately for each event domain (as specified in hypothesis 3.4.(a)) and thus not averaging a subject's judgment differences from the three event domains, loses some statistical power but can address the question whether bias reduction is a general phenomenon in the present data or whether it is possibly driven by a single event domain.

---

[123] Recall that all participants randomly faced one of the two packed intervals, $(I_1 \cup I_2)$ or $(I_3 \cup I_4)$ in each event domain, depending on whether they were assigned to partition 2 or to partition 1.

[124] Calculating the after-before difference of the sums of unpacked intervals would be an equivalent method of testing for a bias reduction. Since the judgments for the unpacked intervals are expected to be relatively too high, a bias reduction would exist if this difference was negative.

[125] 15 subjects did not change their packed-interval judgments at all, i.e., for all event domains they stated the same judgments before and after trading.

Hypothesis 3.4.(a):

H_0: $p_{J,s,i}^{after}(I_x \cup I_{x+1}) - p_{J,s,i}^{before}(I_x \cup I_{x+1}) = 0$ for each event domain i.

It turns out that the null hypothesis of *no* bias reduction can neither be rejected for the finance domain (p=.1577, one-sided test) nor for the sports stimulus (p=.4648, one-sided test). In particular, it becomes obvious that a considerable proportion of subjects (37.5% in finance and 33.3% in sports) did not change their judgments (at least) for the packed event. On the level of event domains, it can also be checked for differences in adjusting judgments for the packed event across low and high intervals (as $I_1 \cup I_2$ in partition 2 is the union of the low intervals and $I_3 \cup I_4$ in partition 1 is the union of the high intervals). A Kruskal-Wallis test for equality of populations, though, shows no significant effects for the finance and the sports event domains. This gives reason to believe that no systematic effect is caused from the fact that the interval $I_1 \cup I_2$ is represented by the "first" asset (i.e., lowest interval) in partition 2, while the interval $I_3 \cup I_4$ is traded as the "third" asset (i.e., highest interval) in partition 1.

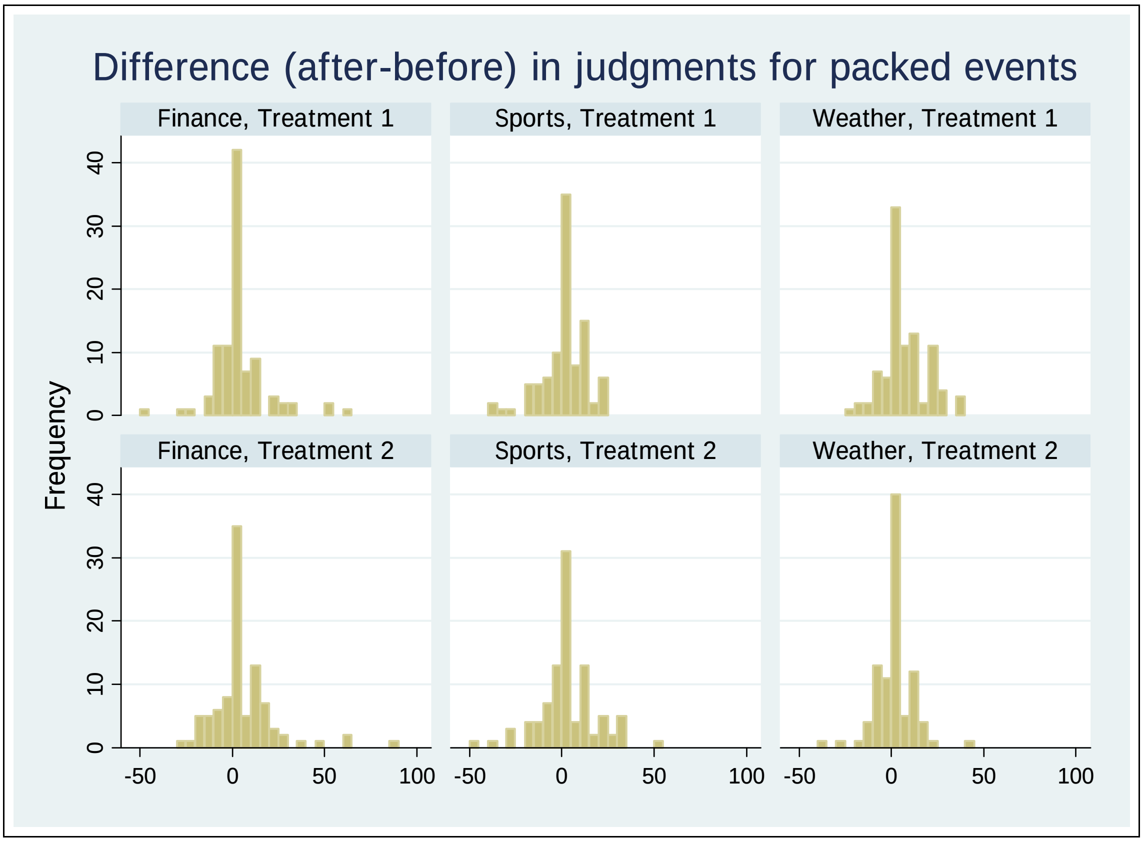


Figure 3.9: Frequency distributions for the difference (after-before) in judgments for packed events.

For the weather event domain, however, bias reduction as measured in terms of hypothesis 3.4.(a) is statistically significant at the 5% level (p=.0274, one-sided test) with 36.5% of subjects confirming their before-trading judgments for the packed event. In this case, the effect is mainly driven by an increase of judgments for the $I_3 \cup I_4$ interval in treatment 1 (+.055 or +10.6%, statistically significant based on a Kruskal-Wallis test), whereas judgments for the interval $I_1 \cup I_2$ hardly differ. Hence, bias reduction for the pooled population seems at least to some extent to arise from a bias reduction in the weather domain, rather than to be a general phenomenon. Figure 3.9 shows the frequency distributions for the difference (after-before) in judgments for the packed interval across event domains and partitions. From this figure it is hard to believe, in fact, that there is a general tendency of raising judgments for the relatively undervalued packed asset after trading.

To summarize, partition-dependence inferred from market prices is slightly reduced compared to the bias expressed in pre-trading judgments for the finance and sports event domains, and cut in half for weather events, though statistical significance is modest. Comparing after-trading judged probabilities to before-market judgments, there is a decline in effect strength of –.055 (–17.6%), –.005 (–1.9%) and –.052 (–18.7%) in the finance, sports and weather event domains, respectively. Hence, the bias is slightly reduced for the finance and the weather domain, but is hardly changed for the sports stimulus. A within-subjects test that analyzes whether participants generally increase their judgments for the undervalued packed event shows that the two 10-minute trading periods apparently have a modest de-biasing effect on individual judgments, though a substantial degree of partition-dependence remains after 20 minutes. In addition, the message of the de-biasing effect is somehow limited due to the fact that it is mainly driven by one of the three event domains (weather). After all, for the finance event domain it seems as if partition-dependence is slightly reduced in market prices and this reduction translates to after-trading judgments. For the weather stimulus, there is a comparable reduction in post-trading judgments, but this reduction does not fully reflect the very strong decline of effect size that is observed in market prices. For the sports domain, market prices reveal a small increase of partition-dependence in market prices for the low intervals and a considerable reduction for the high intervals, while partition-dependence in after-trading judgments is almost at the same level as before-trading.

3.1.3.4 $1/N$ ignorance prior and consistency of prices with a theoretical benchmark

In subsection 2.2.3, anchoring and insufficient adjustment of the $1/N$ ignorance prior was presented as a convincing explanation for partition-dependence, particularly in the context of probability judgments for dimensional state spaces. This account focuses on the “ignorance prior” heuristic that is based on distributing probability evenly across all available “categories” in a first step. Without precise knowledge of objective probabilities and without having (at least) two differently partitioned state spaces, though, it is difficult to decide whether and to what extent judgments or market prices are biased by $1/N$ allocation. (Note, for example, that an allocation of $1/N$ could either reflect full ignorance, but could also be consistent with objective probabilities by chance.) The main design feature of creating two different partitions of the state space allows testing for partition-dependence without the knowledge of objective probabilities of events. That means, regardless of what the true probability of an event is, partition-dependence can be measured by simply comparing judgments and prices from the two treatments. If, in the present study, people were applying an ignorance prior of $1/N$ of the judged probability to each of the three events they faced, the difference between the sum of the two unpacked-interval judgments ($2 \times 1/N = .667$) and the single packed-interval judgments ($1 \times 1/N = .333$) should be one third. So differences in pre-trading judgments (effect size of .312, .261 and .278 for the finance, sports and weather domains, respectively) are close to $(2 - 1)/N = .333$ and thus suggest a very strong impact of the ignorance prior, as demonstrated in earlier psychology experiments. Differences in post-trading judgments (effect size of .257, .256 and .226 for the finance, sports and weather domains, respectively) are still close to .333 but moderately reduced in comparison to pre-trading judgments.

Two alternative explanations for the results might spring to mind: Partitions convey information (the “credibility” account; see subsection 2.2.3); and prices do not reflect mean or median beliefs (see subsection 2.1.4.3 for a discussion). The first alternative explanation is that subjects infer reasonable beliefs from the partition that is presented to them. This explanation is disabled by the fact that *all* subjects were explicitly told about *both* their own event partition *and* the different partition for the same event given to the other subjects. Thus, there might well be an informational effect but it should be the same in both groups (up to sampling error) and therefore cannot explain the between-group differences that are observed.

The second alternative explanation is that prices do not reflect average or median beliefs, which could create an apparent bias in prices. For example, suppose prices reflect the most optimistic beliefs. Then separate prices for events I_1 and I_2 could add to more than the price for $I_1 \cup I_2$. This explanation is closely related to the theoretical debate about what prediction market prices reveal about average beliefs. Recall that Manski (2006) noted that whether the market price is close to the mean or median of a distribution of beliefs in a population of traders depends on the traders' risk attitudes and wealth constraints, and their correlation with belief.[126] Wolfers and Zitzewitz (2007) replied by showing some sufficient conditions under which the market price *is* the mean of the belief distribution, and by showing that under other reasonable conditions prices are likely to be close to the mean belief. The fact that subjects in the present experiment gave individual probability judgments for events, *and* collectively created market prices for the *same* events, provides a rare opportunity to shed some empirical light on this debate: One can compute the quantile of the individual belief distribution that market prices correspond to, i.e., take the second of the two consecutive trading rounds and compare the quantiles of post-market event beliefs elicited from subjects with the quantity/volume-weighted average of the last three trade prices. Remember that in each market there are eight belief judgments and a single (averaged) price for all three events that were traded. The estimated quantile is the percentage of those eight beliefs that are below the market price. Separate quantile estimates can be computed for twelve separate experimental sessions for each of the three event domains. The mean belief quantiles for prices of the finance, sports and weather events are .507, .512, and .483 (all close to the median belief, the .5 quantile).[127] Furthermore, in 54% of the 36 session-stimulus comparisons, the estimated belief quantile is .375, .5, or .625 (i.e., the price is between the highest and lowest three out of eight beliefs about half the time). Thus, the hypothesis that market prices approximately reflect the median belief is reasonably well supported in these data, when individual beliefs and collective prices can be directly compared.

126 For example, in principle one very wealthy risk-neutral trader who is extremely optimistic could effectively set a market price which reflects his extreme optimism if others trade less because of wealth constraints or risk-aversion.

127 The mean belief quantiles on a (not averaged) per asset basis are generally close to the .5 quantile, too. They range between .344 and .594 (finance assets), .396 and .677 (sports assets), and .229 and .729 (weather assets).

With regard to a comparison of prediction market prices to a theoretically derived benchmark, recall that for the sports and weather event domains the interval boundaries were chosen rather arbitrarily, so that there is no conclusive way to link expressed probabilities to objective probabilities. Contrariwise, as is described in subsection 3.1.1, for the finance event domain, the four intervals were created based on a normative theory of stock index price movements, such that the four intervals correspond to certain percentiles of the expected probability density function for the DAX close two weeks in the future. This allows carrying out an additional analysis by comparing market prices from the finance markets to theoretically derived probabilities for the DAX events and sheds light on the question to which extent prices are consistent with theoretically derived probabilities.

Table 3.14 opposes equilibrium market prices of trading rounds 1 and 2, averaged across the twelve session slots, to (i) the 1/*N* ignorance prior of one third for each of the three assets and (ii) (for the finance event markets) to the theoretical benchmark. Note that (rescaled) mean equilibrium prices do not necessarily sum to 1.0 (the range is from .984 (minimum) to 1.127 (maximum)). Since prices are directly compared to the ignorance prior probabilities and to the theoretical benchmark, the obvious procedure is to standardize equilibrium market prices such that they sum to 1.0 in each treatment. Hence, the columns 3 and 4 of Table 3.14 show standardized equilibrium prices. Comparing the (standardized) mean equilibrium trade prices to the corresponding values of the ignorance prior distribution, it appears that prices apparently deviate from one third (at least for two of three assets in each treatment), suggesting substantial adjustments to the ignorance prior. In addition, there seems to be no obvious pattern in *how* prices differ from the ignorance priors. For instance, it is not always the second asset (I_2 in treatment 1 and I_3 in treatment 2) for which prices are the highest, as might be hypothesized. Nor is it the packed asset ($I_3 \cup I_4$ in treatment 1 and $I_1 \cup I_2$ in treatment 2) that trades consistently at the highest price. Regardless of whether participants notionally “start” by a naïve allocation of probabilities to events or not, these results suggest that traders eventually do incorporate their beliefs and private information into market prices. Apparently, though, these adjustments happen to be insufficient, as is reflected in the strong effects of partition-dependence reported above.

Table 3.14: Mean equilibrium trade prices (actual/standardized), 1/N ignorance prior, and theoretical benchmark.

	Mean Equilibrium Trade Price		Mean Equil. Trade Price (Standardized to 1.0)		1/*N* Ignorance Prior	Theoretical Benchmark
	Round 1	Round 2	Round 1	Round 2		
Finance						
Treatment 1						
$P(I_1)$	0.187	0.152	0.166	0.145	0.333	0.275
$P(I_2)$	0.577	0.561	0.512	0.535	0.333	0.275
$P(I_3 \cup I_4)$	0.363	0.336	0.322	0.320	0.333	0.450
Total	1.127	1.049	1.000	1.000		
Treatment 2						
$P(I_1 \cup I_2)$	0.421	0.424	0.405	0.422	0.333	0.550
$P(I_3)$	0.414	0.404	0.398	0.402	0.333	0.225
$P(I_4)$	0.205	0.177	0.197	0.176	0.333	0.225
Total	1.040	1.005	1.000	1.000		
Sports						
Treatment 1						
$P(I_1)$	0.228	0.230	0.212	0.207	0.333	-
$P(I_2)$	0.476	0.490	0.443	0.441	0.333	-
$P(I_3 \cup I_4)$	0.372	0.391	0.346	0.352	0.333	-
Total	1.076	1.111	1.000	1.000		
Treatment 2						
$P(I_1 \cup I_2)$	0.401	0.439	0.408	0.436	0.333	-
$P(I_3)$	0.391	0.416	0.397	0.413	0.333	-
$P(I_4)$	0.192	0.152	0.195	0.151	0.333	-
Total	0.984	1.007	1.000	1.000		
Weather						
Treatment 1						
$P(I_1)$	0.090	0.048	0.084	0.047	0.333	-
$P(I_2)$	0.335	0.256	0.313	0.253	0.333	-
$P(I_3 \cup I_4)$	0.646	0.707	0.603	0.700	0.333	-
Total	1.071	1.010	1.000	1.000		
Treatment 2						
$P(I_1 \cup I_2)$	0.160	0.149	0.153	0.149	0.333	-
$P(I_3)$	0.388	0.354	0.371	0.354	0.333	-
$P(I_4)$	0.499	0.496	0.477	0.496	0.333	-
Total	1.047	0.999	1.000	1.000		

Comparing (standardized) mean equilibrium market prices to theoretically derived probabilities for the finance event domain, it appears as if prices differ substantially but unsystematically from this benchmark in both treatments. In particular, events that represent the same distributional probabilities ($p(I_1) = p(I_2) = .275$ in treatment 1 and $p(I_3) = p(I_4) = .225$ in treatment 2) show differences of .390 (treatment 1) and .226 (treatment 2) in terms of equilibrium market prices. It seems as if market generated probabilities are in no way consistent with the theoretical benchmark. Session-level prices do not show any systematic patterns either.

3.1.4 Second order results

3.1.4.1 Competence effects

A relevant question is whether the degree of partition-dependence depends on the subjects' (self-perceived) level of competence and knowledge in a certain event domain. It can be hypothesized that people who feel more competent in a given event domain are able to provide probability judgments which are better calibrated than those provided by people who feel less knowledgeable. The rationale behind this conjecture is that experts, due to their knowledge and experience, are less sensitive to a concrete partition of the state space and, if the ignorance-prior model applies, are able to adjust the ignorance prior more accurately than others. This should result in reduced partition-dependence since judgments of competent people are likely to be more consistent with whatsoever objective likelihoods no matter which state space partition the experts face. Reduced partition-dependence should also be reflected in prediction-market prices if these markets contain some competent traders. Consistent with the "smart few hypothesis" (or marginal trader hypothesis) this is because expert traders are supposed to be in a position in which they can recognize and exploit potential mispricing of winner-take-all contracts and thereby influence market prices in a direction that is more consistent with objective probabilities. If judgments and market prices were consistent with objective probabilities in both treatments, there would be no partition bias. This section is to analyze whether self-rated competence influences partition-dependence in the present lab environment.

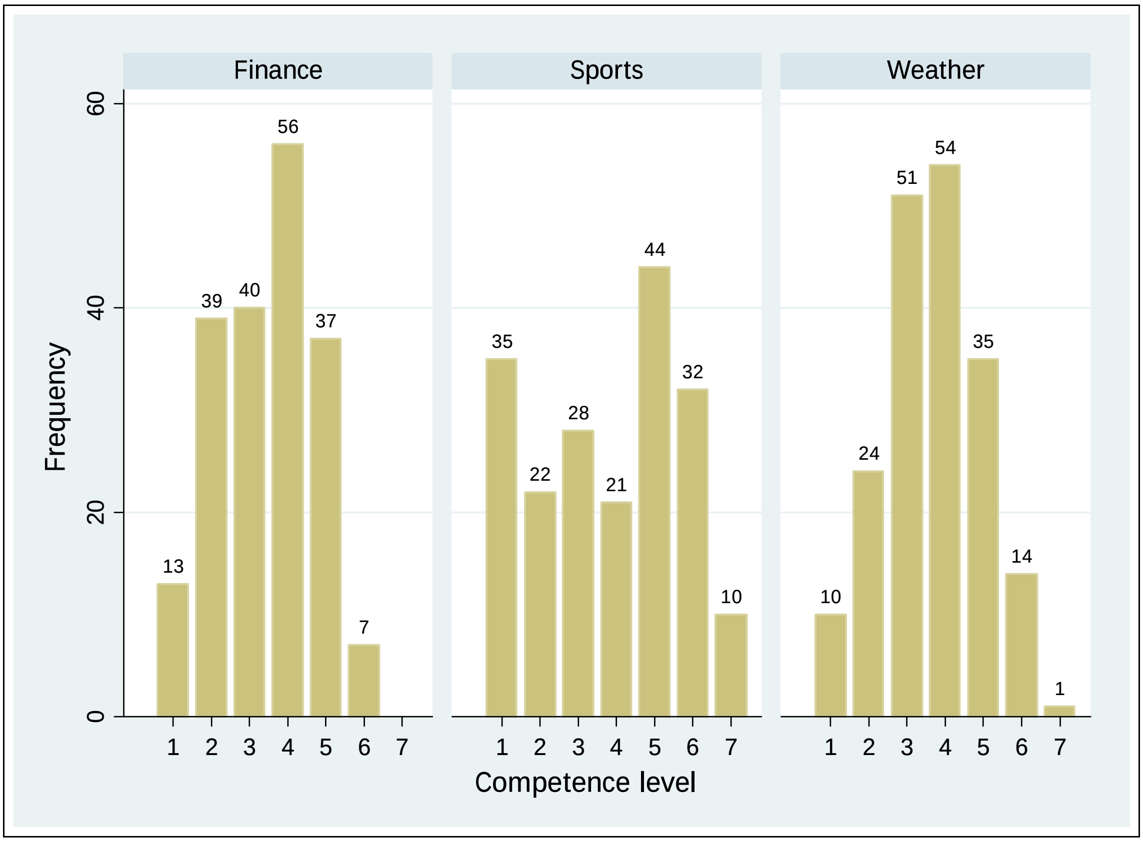


Figure 3.10: Frequency distribution of subjects' self-rated competence.

When participants were asked to provide their subjective probability judgments just before trading began in each event domain they were also asked to self-rate their competence on a [1 (very incompetent) – 7 (very competent)] scale on making such judgments. Figure 3.10 illustrates the frequency distribution of self-assessed competence scores for the three event domains. The mean competence levels for the finance, sports and weather domains are 3.45, 3.80 and 3.67, respectively; the median value is 4 for all event domains. A bimodal distribution of competence levels would be most suitable for second order analyses of competence as such a distribution would generate strong discriminatory power between people who feel very competent and those who feel very incompetent about a given event domain. It appears from Figure 3.10 that there is considerable variation in competence among the subjects and across event domains (with the highest standard deviation in the sports domain), but competence is not at all bimodal distributed. Rather, there is a substantial fraction of people who feel more or less competent (at least for the finance and weather domains) which results in a peak at levels [3] and [4]. Given these distributions, competence effects in individual probability judgments could be distorted by a substantial degree of noise from moderately competent people. To form sub-groups as heterogeneous as possible, moderately competent people are excluded from the following analysis what makes it possible to oppose highly competent people to very incompetent participants. Concretely, in the finance domain levels [1–2] are classified as the low-competence group (N=52), levels [5–6][128] as the high-competence group (N=44), while levels [3–4] are excluded (N=96). In the sports domain levels [1–2] are used as the low-competence group (N=57), levels [6–7] as the high-competence group (N=42), while levels [3–5] are excluded (N=93). Finally, in the weather domain levels [1–2] are used as the low-competence group (N=34), levels [5–7] as the high-competence group (N=50), while levels [3–4] are excluded (N=105).[129] To provide enough statistical power, the groups were created such that a minimum of thirty subjects remains in each of the low- and high-competence clusters.

Second order competence effects can be analyzed by a 2 × 2 factorial analysis of variance (ANOVA). ANOVA is the standard method to analyze the influence of two (or more) independent variables on a single (interval-scaled) variable.[130] In addition, ANOVA is suited to detect *interaction effects* between those variables. The basic idea of ANOVA is to test whether there are significant differences between means by decomposing variances. Since the existence and significance of the main treatment effect (i.e., partition-dependence) is analyzed in sections 3.1.3.1 (for judged probabilities) and 3.1.3.2 (for equilibrium market prices) the focus here is on interaction effects. Interactions exist if the effect of a *focal* independent variable (here: partition type) on a dependent variable (here: judgments for intervals $I_1 \cup I_2$ or $I_3 \cup I_4$) depends on the value of a third variable which is called the *moderator* variable (here: different competence groups). Interaction effects can be thought of as the difference between mean differences. That is to say, the difference between two means in a 2 × 2 factorial design reflects the impact of an independent variable on a dependent variable *at a given level* of the moderator variable, and the fundamental question is whether these differences vary as a function of the levels of the moderator variable.[131] The following hypotheses can be derived:[132]

---

128 Level [7] is unallocated in this event domain.

129 Unfortunately, three competence ratings in the weather domain were lost due to technical errors.

130 For a textbook version of the two-factor analysis of variance (ANOVA) and the role of interaction effects see, e.g., Backhaus et al. (2006, pp. 119-153).

131 See Jaccard (1998, pp. 2-6).

132 Again, the second hypothesis is redundant for judged probabilities because probabilities, by instruction, sum to 1.0. Thus, the effect size is necessarily the same for both intervals.

Hypothesis 3.5:

H$_0$(a): $$\left(p_{J,i}^{low\_comp}(I_1)+p_{J,i}^{low\_comp}(I_2)-p_{J,i}^{low\_comp}(I_1\cup I_2)\right) -\left(p_{J,i}^{high\_comp}(I_1)+p_{J,i}^{high\_comp}(I_2)-p_{J,i}^{high\_comp}(I_1\cup I_2)\right)>0$$ and

H$_0$(b): $$\left(p_{J,i}^{low\_comp}(I_3)+p_{J,i}^{low\_comp}(I_4)-p_{J,i}^{low\_comp}(I_3\cup I_4)\right) -\left(p_{J,i}^{high\_comp}(I_3)+p_{J,i}^{high\_comp}(I_4)-p_{J,i}^{high\_comp}(I_3\cup I_4)\right)>0$$

Table 3.15: Interaction effects between partition type and competence level in pre- and post-trading probability judgments.

	Pre-Trading Judgments		Post-Trading Judgments		N_{low} / N_{high}
Competence Level	Low	High	Low	High	
Mean probability judgments					
Finance					
$p(I_1) + p(I_2)$	76.28	65.08	71.55	60.62	29 / 18
$p(I_1 \cup I_2)$	41.22	41.15	42.30	46.58	23 / 26
Difference	35.06	23.93	29.25	14.04	
Difference of differences	11.13*		15.21**		
(p-value, one-tailed)	(.056)		(.031)		
Sports					
$p(I_1) + p(I_2)$	74.66	55.63	72.46	58.25	25 / 24
$p(I_1 \cup I_2)$	39.51	38.06	43.13	37.50	32 / 18
Difference	35.15	17.57	29.33	20.75	
Difference of differences	17.58**		8.58		
(p-value, one-tailed)	(.023)		(.178)		
Weather					
$p(I_1) + p(I_2)$	51.85	43.91	44.78	37.82	15 / 22
$p(I_1 \cup I_2)$	26.21	16.18	22.79	15.48	19 / 27
Difference	25.64	27.73	21.99	22.34	
Difference of differences	-2.09		-0.35		
(p-value, one-tailed)	(.386)		(.479)		

Table 3.15 shows mean probability judgments (before- and after-trading) for the low intervals ($I_1 \cup I_2$) for the three event domains, broken down into the two competence groups. Calculating the difference of mean judgments between the sum of unpacked events (I_1 and I_2) and the packed event ($I_1 \cup I_2$) yields the effect size of partition-dependence for a given competence level. Calculating the difference of these differences, in turn, gives information about the interaction effect (p-value, one-tailed, in parentheses). If this difference of differences (the interaction term) is positive, this indicates a lower degree of partition-dependence in the high-competence group, as hypothe-

sized. It turns out that there is a substantial interaction between the effect size of partition-dependence and competence level in the finance and sports event domains (statistically significant in three of four cases, as indicated by the asterisks: ***, **, * indicate significance at the 1%, 5%, 10% level, respectively). This means, partition-dependence is on average less pronounced in judgments from the high-competence groups than in judgments from the low-competence groups, and this difference is remarkable in size, as Table 3.15 shows. In pre-trading judgments for the finance domain, this is mainly attributable to a lower sum of judgments for the unpacked events by competent people (65.08 against 76.28). The same is true in pre-trading judgments for the sports events (55.63 against 74.66). Contrariwise, there is no statistically significant interaction effect in the weather stimulus. Interestingly, highly competent subjects attribute consistently less probability to the interval $I_1 \cup I_2$ compared to low competent people (except for post-trading judgments for the packed event in the finance domain).

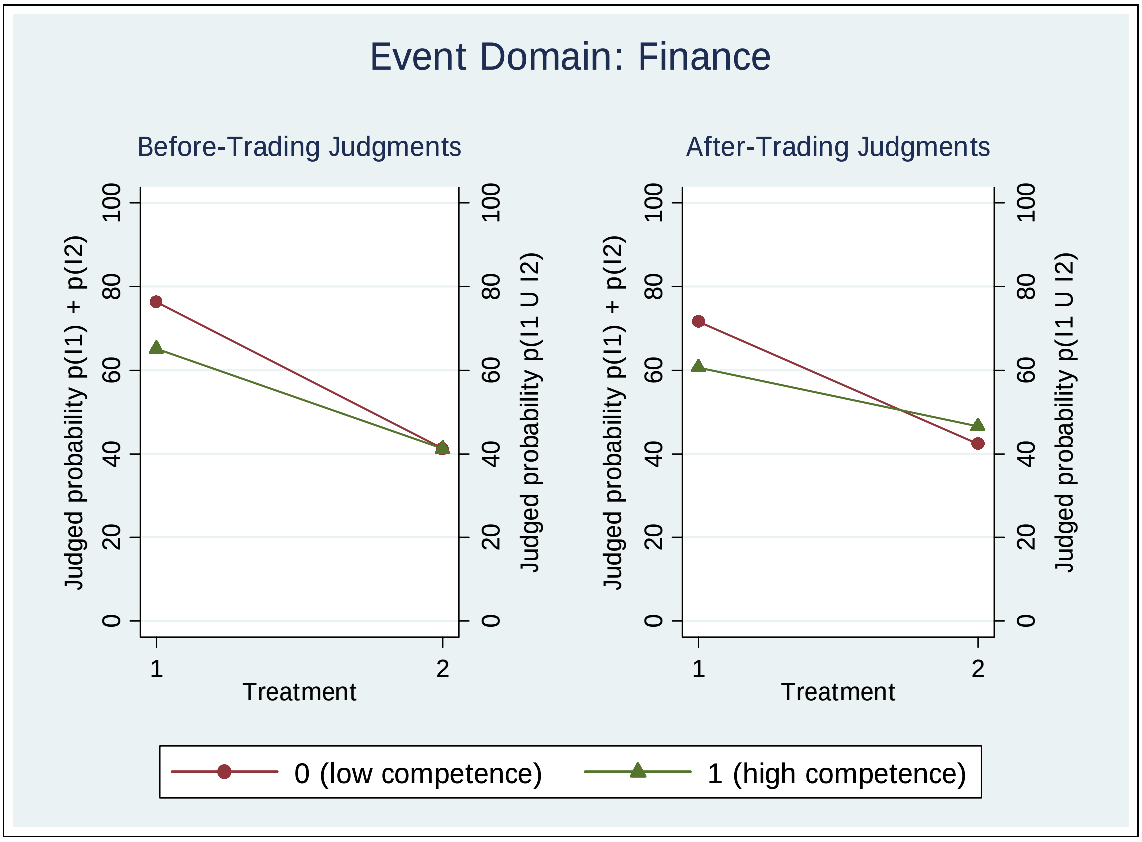


Figure 3.11: Plots of cell means for judgments in the finance event domain (before- and after-trading).

Figure 3.11 presents plots of cell means for the finance event domain, illustrating the interaction effects between the focal treatment variable (i.e. partition type) and competence levels in before- and after-trading probability judgments. The Figure shows the focal independent variable (partition type) on the horizontal axis, the mean values of judged probability for the low intervals on the vertical axes (i.e., the sum of judgments for the unpacked events in partition 1 (left-hand axis) and the judgment for the packed event in partition 2 (right-hand axis)), separately for each category of the moderator variable (i.e., the different levels of the competence proxy: 0 = low competence, 1 = high competence). If the lines connecting cell means are nonparallel, as is the case for the finance event domain, this indicates interaction effects. The slope of the line connecting cell means of highly competent people is much smaller than the slope for the low-competence group. This means that there is less partition-dependence for highly competent subjects than for low competent participants. In after-trading judgments there is even a crossover interaction as the two lines intersect.

To summarize the results on competence effects in judged probability, the data reveals statistically significant competence interactions comparing judgments by highly competent subjects to those provided by very low competent subjects in the finance and the sports event domains. Partition-dependence is thus generally less pronounced for the high competence group. While competence matters in before- and after-trading judgments for the finance and the sports stimulus, there is virtually no competence interaction in the weather judgments. Perhaps this is because the genuine (but unobservable) distribution of weather competence is much more compressed than indicated by the seven-fold scale so that the “competence gap” is much smaller than in the finance and sports domains. In other words, while almost everyone is likely to have at least a rough idea about normal temperatures for the relevant time of the year, the finance and sports events are supposed to be much more information-driven as they do not necessarily come to mind in everyday life. But even if the results suggest that highly competent people are less susceptible to the partition bias in some domains, partition-dependence is yet well pronounced for this subsample (effect size between 14.04 and 27.73).

Consequentially, the next question is whether second-order competence effects are also reflected in prediction-market prices. Recall that participants were allocated to six high- and six low-competence session slots based on their self-reported soccer (“Bundesliga”) competence during the registration process to carry out this analysis. Based on this allocation procedure the mean (median in parentheses) competence level of the six high-competence groups varied between 5.81 and 6.27 (6 for all groups) and ranged from 2.25 to 2.75 (2 to 3) for the six low-competence groups. As a result, the experimental groups happened to be quite homogenous within each competence class, but at the same time were rather heterogeneous for different competence levels (high vs. low). Based on the self-rated soccer competence that was collected just before trading, these competence proxies ranged from 4.88 to 5.50 (5 for all groups) for the high-competence session slots and from 1.88 to 3.00 (1 to 2.5) for the low-competence slots. Spearman's rank correlation coefficient between soccer competence elicited during the registration process and soccer competence collected during the experiment (based on *N*=190 individual statements)[133] is .779. As could have been expected, self-stated competence was consistently lower when concretely asking the subjects for their ability of making probability judgments on the number of goals (right before trading) than asking for their general knowledge in the field of soccer (during registration), but was still very selective during the experiment.[134]

This said, one can compare market prices and partition-dependence from the high- and low-competence markets for the sports-events markets and see whether the partition bias is indeed less pronounced in the high-competence markets. The hypotheses are parallel to those derived for probability judgments:

<u>*Hypothesis 3.6:*</u>

H_0(a):
$$\begin{aligned}&\left(P_{sports}^{low\_comp}(I_1)+P_{sports}^{low\_comp}(I_2)-P_{sports}^{low\_comp}(I_1\cup I_2)\right)\\&-\left(P_{sports}^{high\_comp}(I_1)+P_{sports}^{high\_comp}(I_2)-P_{sports}^{high\_comp}(I_1\cup I_2)\right)>0\end{aligned}$$
and

H_0(b):
$$\begin{aligned}&\left(P_{sports}^{low\_comp}(I_3)+P_{sports}^{low\_comp}(I_4)-P_{sports}^{low\_comp}(I_3\cup I_4)\right)\\&-\left(P_{sports}^{high\_comp}(I_3)+P_{sports}^{high\_comp}(I_4)-P_{sports}^{high\_comp}(I_3\cup I_4)\right)>0\end{aligned}$$

---

133 *N*=2 backup participants did not register via the registration process.

134 A related question is about correlations between soccer competence and competence in the other two event domains (finance and weather). Calculating pairwise correlation coefficients between the competence proxies shows that correlation is only small and statistically insignificant between the weather and sports competence proxies ($\rho_{\text{weather, sports}}$=.0471) and between the weather and finance competence proxies ($\rho_{\text{weather, finance}}$=.0803). The correlation between the sports and finance competence proxies, however, is statistically significant (at the 1% level), but is also modest in magnitude ($\rho_{\text{sports, finance}}$=.2789). Possible effects from this correlation are considered to be negligible since mean (median in parentheses) self-assessed competence in the finance stimulus is 3.60 (4) in the high sports-competence groups and not much different compared to 3.29 (3) in the low sports-competence session slots (a Kruskal-Wallis equality-of-populations rank test is statistically significant only at the 10% level (p=.0852)). Given these correlations, one may conclude that grouping participants based on their soccer competence allows for second order analyses in the sports stimulus, but does not imply any unintended competence treatments in the finance and the weather event domains.

Table 3.16: Interaction effects between partition type and competence level in equilibrium market prices for the sports event domain.

	Trading Round 1		Trading Round 2		N_{low} / N_{high}
Competence Level	Low	High	Low	High	
Panel A: Mean equilibrium market prices					
$P(I_1) + P(I_2)$	68.77	72.00	73.72	70.21	6 / 6
$P(I_1 \cup I_2)$	39.20	41.01	44.37	43.51	6 / 6
Difference	29.57	30.98	29.35	26.69	
Difference of differences	-1.41		2.65		
$P(I_3) + P(I_4)$	54.31	62.22	52.80	60.74	6 / 6
$P(I_3 \cup I_4)$	36.46	37.91	39.56	38.66	6 / 6
Difference	17.85	24.30	13.24	22.09	
Difference of differences	-6.45		-8.85		
Panel B: Mean equilibrium market prices (normalized by 100)					
$P(I_1) + P(I_2)$	65.35	65.50	65.08	64.49	6 / 6
$P(I_1 \cup I_2)$	41.92	39.73	45.66	41.74	6 / 6
Difference	23.43	25.77	19.41	22.75	
Difference of differences	-2.34		-3.34		
$P(I_3) + P(I_4)$	58.08	60.27	54.34	58.26	6 / 6
$P(I_3 \cup I_4)$	34.65	34.50	34.92	35.51	6 / 6
Difference	23.43	25.77	19.41	22.75	
Difference of differences	-2.34		-3.34		

Table 3.16 shows mean equilibrium market prices (first and second trading round) for the two partitions in the sports event domain, grouped by the competence level of markets. Panel A contains the mean equilibrium prices based on the last three trade prices (volume-weighted) of each market. In Panel B mean equilibrium market prices were normalized such that they sum to 100 for each partition type and competence level. This was done to make differences comparable to probability judgments presented above and to control for different "overrounds" (i.e., deviations of the sum of market prices and one hundred) in the markets. The calculation of interaction effects is parallel to that in Table 3.15 above. The main treatment effect corresponds to the results presented in subsection 3.1.3.2. However, it turns out that there is only little interaction between partition type and the level of competence in market prices, i.e., the difference of differences is rather small (ranges between –8.85 and 2.65) compared to differences in probability judgments. Actually, three out of four differences (Panel A) are less than zero and thus reflect greater partition-dependence in the high-competence subsample. Normalizing market prices (Panel B) does not alter the results. Even without performing

an ANOVA (which is problematic due to the small sample size here) it can be concluded that there is no evidence for competence interactions in the sports event domain markets. In particular it seems as if the competence discrepancy diminishes in the market in the sense that the partition bias shrinks for the low-competence group, and increases for the high-competence group, compared to pre-trading judgments. Statistical power of analyzing the market prices is limited, though, as only six markets exist for each combination of partition type and competence level.

To summarize, the data collected in the lab provide evidence for interaction effects between the subjects' self-rated competence and the size of partition-dependence in individually judged probabilities. Highly competent people seem to be less susceptible to the partition bias than low competent people, at least for the finance and sports event domains. By contrast, this effect does not show in judged probabilities of the weather events. In all cases, though, the main treatment effect remains substantial (even for the highly competent subgroup). In addition, the degree of partition-dependence in the markets' equilibrium prices does not depend on the competence level of traders that comprise these markets (as measured for the sports event domain), even though participants were homogeneously allocated to the markets by their level of self-rated competence.

3.1.4.2 Trader-based analysis

To characterize the quality of results inferred from market prices and trading activity in the lab study and to further explore the role of competence in these markets, it is helpful to have a more detailed look at *how* (equilibrium) market prices emerged. A relevant question is, for instance, whether trading was dominated by only a small number of traders who actively traded among themselves, or whether market prices were the result of multilateral trading activity by the majority of participants. Another question is, for example, whether traders tend to have a great exposure at the end of the trading round or whether they rather try to close open positions by accumulating unit portfolios and cash. This subsection is to address these questions.

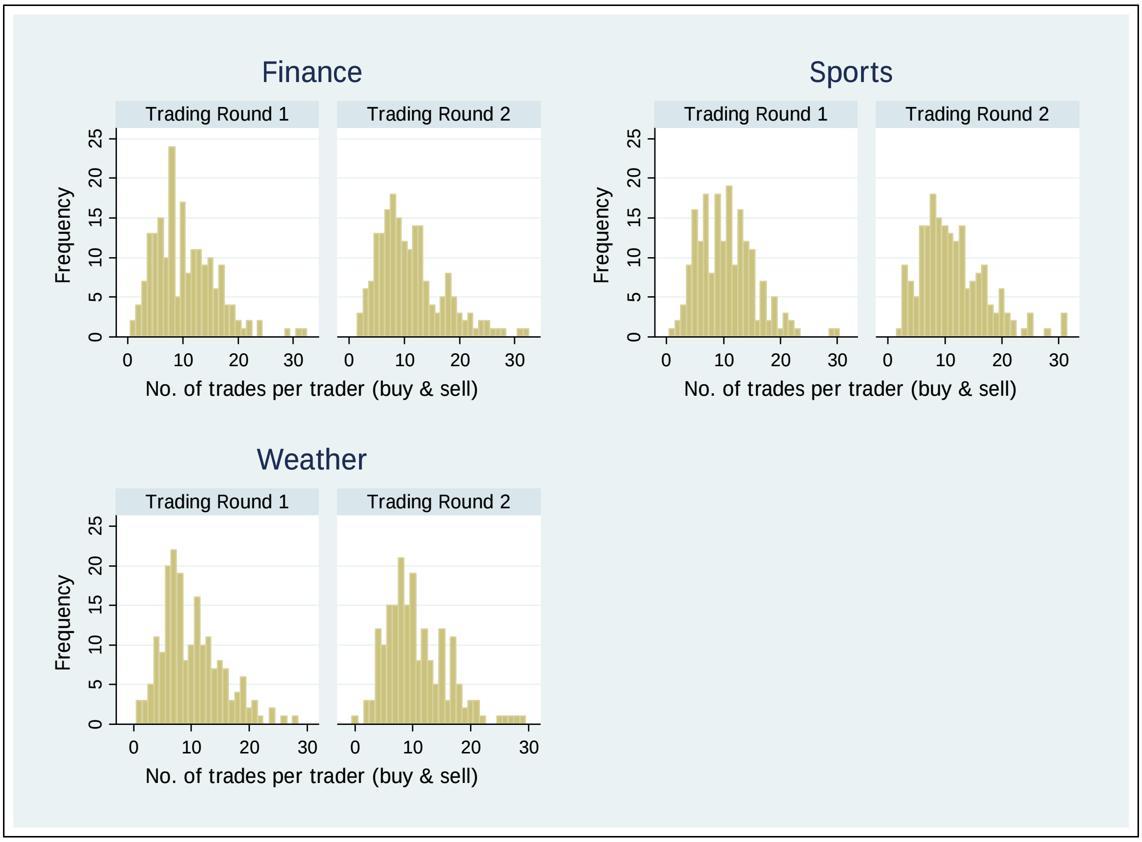


Figure 3.12: Frequency distribution of the number of trades per trader (buy & sell).

Figure 3.12 shows frequency distributions of the total number of trades per trader for each of the six trading rounds, regardless of whether a trader acted as a buyer or a seller (*N*=192 for each of the six trading rounds).[135] Each bin of the chart represents a single number of trades. The mean number of trades (double count) is 10.71 (standard deviation: 5.49; median: 10) which means that the average trader was involved in a trade once in each minute of trading. Remarkably, there is only one trader who did not trade throughout the whole trading round (this occurred in the second trading round of the weather event domain). Generally speaking, trading involvement was extremely high, suggesting that participants were indeed interested in trading. In particular, there were virtually no trade "deniers", i.e. people who traded rarely and confined themselves to just cash in their initial endowment. In most cases, frequency distributions are right-skewed and some subjects bought and/or sold assets up to thirty times (maximum is 32) within a trading round which is equivalent to approximately three trades per minute. To address the question whether market outcomes are largely influenced by a small fraction of very active traders, the following *trade concentration metric* is developed: For each

[135] Trades of the unit portfolio were excluded from the data.

of the eight traders in a market, calculate the proportion of trades in which she appeared as either a buyer or seller. For example, suppose that 73 trades occurred in a given trading round, and a single trader was involved in 14 trades (either purchases or sales), then this metric would result in a 19.2% share of trades in which this trader participated. Since there are always two traders involved in a given trade, this metric sums up to 200% for all traders of a market. Consequently, if trades were uniformly distributed among the eight traders, each of them would have a trade concentration share of 25%. Hypothetically, if a trader was involved in all trades that occurred in a given market, her metric would yield a 100% share.

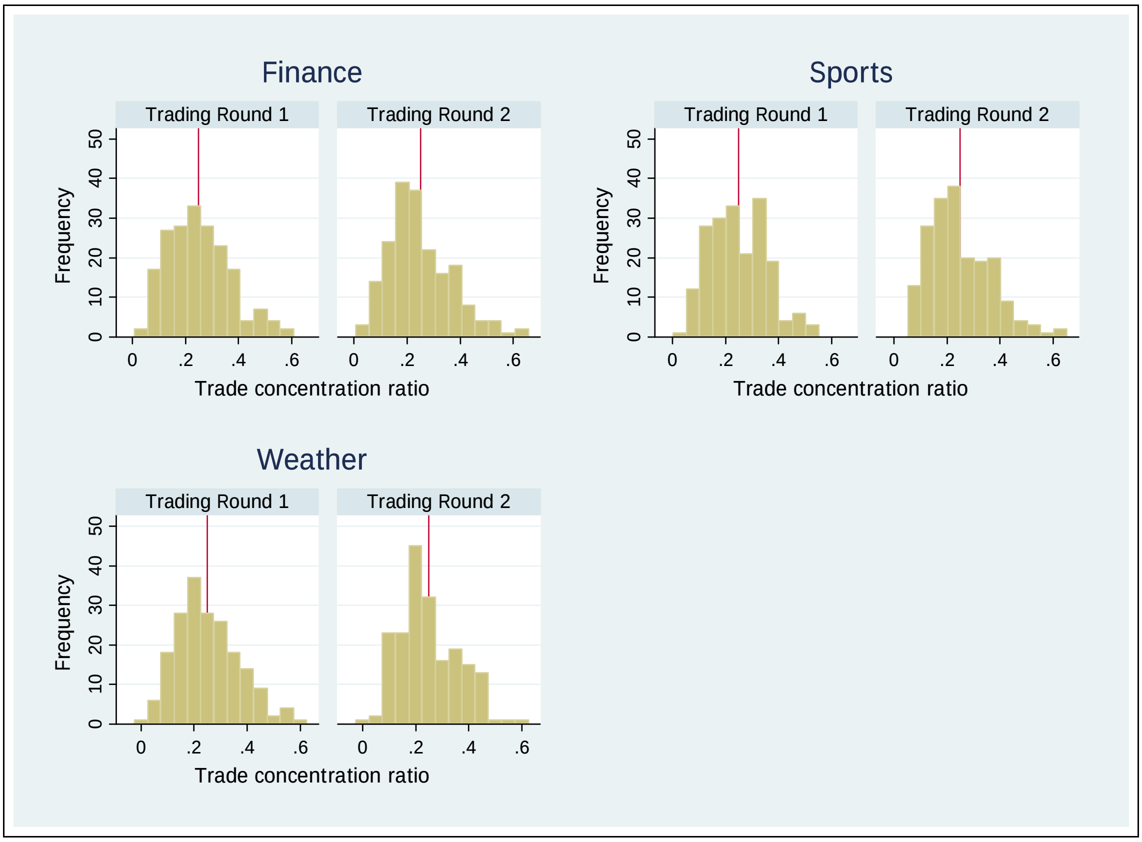


Figure 3.13: Frequency distribution of the proportion of trades per trader (trade concentration metric).

Figure 3.13 shows the frequency distributions of the trade concentration ratio for each of the six trading rounds (N=192 for each of the six trading rounds). Each bin of the chart represents an interval of 5%. The vertical red line at .25 indicates the "benchmark share" of a uniform distribution as described above, which is, by construction, equal to the mean of the distribution. The median concentration ratio is 23.4% and the standard deviation is 11.4%. The highest share was reached by a trader in the second trading round of the finance stimulus in which this trader accounted for 64.3% of the trades. However, the 90%-percentile is much lower (at a 40.0% share) which means that 90% of the traders *each* accounted for less than 40% of the trades. The 10%-percentile, in turn, is 11.7% which means that only 10% of the traders *each* accounted for less than 11.7% of trades. To summarize, market prices are the result of trading activity generated by many different traders, with some traders being particularly active. In addition, there are no systematic patterns that uniquely apply to one of the three event domains or to one of the two trading rounds. The overall view of trading activity gives reason to assume that market outcomes in the present experiment are to a great extent the result of collective information processing, with some traders being even more confident to "put their money where their mouths are" than others.

Another issue with respect to the role of markets in aggregating traders' beliefs and how this may affect partition-dependence deals with the relationship between individual probability judgments and *individual trade prices*. As discussed earlier, in "winner-take-all" prediction markets asset prices can be thought of reflecting market-implied probabilities for the occurrence of an event. Assuming risk-neutral traders, their reservation prices for assets are expected values.[136] Thus, a risk-neutral market participant can be expected to purchase a contract if the current market price (or to be more precise: if the current ask quote) is lower than her individual probability judgment for the corresponding event, and she can be expected to sell a contract if the current price (bid quote) is higher than her individual probability judgment. This guarantees a positive *expected profit* based on a trader's subjective probability judgments. Since individual probability judgments were collected before and after the trading rounds, the judgments can be compared to trade prices at which a given trader did buy and sell the corresponding assets. In doing so one can study whether participants consistently traded according to their likelihood beliefs or whether they adjusted their beliefs as a reaction to the development of market prices. To aggregate trade prices the average buy and sell prices (quantity-weighted) were calculated for each trader and asset based on actual trades of that trader.

Two specific metrics are developed that calculate a trader's expected profit from trading (i) by comparing before-trading judgments to individual trade prices of the first trading round and (ii) by comparing after-trading judgments to individual trade prices of

[136] See Camerer (1987, p. 984); if traders are not risk-neutral, their reservation prices equal certainty equivalents.

the second trading round in each event domain.[137] The basic idea behind this calculation is as follows: If risk-neutral traders always agreed to buy and sell winner-take-all assets at prices that exactly correspond to their individual (i.e., subjective) probability estimates, the expected profit from trading would be zero. Thus, traders have an incentive to create a positive expected return by buying assets at a price lower than their subjective probability judgment, and selling assets for a price higher than their individual probability estimate. Hence, for each trader and asset the quantity-weighted difference between judged probabilities and individual trade prices is calculated, separately for buy and sell transactions. Afterwards, expected profits from purchases and from sales are added up separately and standardized by the overall number of assets traded by a trader. For example, assume that a trader's before-trading judgments for the finance event domain were $p(I_1)$=20%, $p(I_2)$=55%, and $p(I_3 \cup I_4)$=25%. Further assume that, in the first trading round, this trader purchased a quantity of 10 assets at price $P(I_1)$=13.0, 4 assets at price $P(I_2)$=30.5, and 10 assets at price $P(I_3 \cup I_4)$=32.0. At the same time, she sold a quantity of 4 assets at price $P(I_1)$=30.0, 1 asset at price $P(I_2)$=60.0, and 3 assets at price $P(I_3 \cup I_4)$=34.0. Then, her expected profit from the purchases would be: ([20–13] · 10 + [55–30.5] · 4 + [25–32] · 10) = 98, and the expected profit from the sales would be: ([30–20] · 4 + [60–55] · 1 + [34–25] · 3) = 72. To make these profits comparable across subjects, they are standardized by the overall number of traded assets which yields 98/24 = 4.08 [cent] for purchases and 72/8 = 9.00 [cent] for sales. From an economic perspective, these values can be interpreted as the mean expected profit per unit of (purchased or sold, respectively) asset based on subjective probability estimates. Averaging the two profits gives an aggregate metric of the mean expected profit per unit (6.54 [cent] in this example). This metric is calculated for each of the 192 participants in each event domain. Similarly, one can compare after-trading judgments and second trading round transactions which results in the second metric.

The derived metrics are expected to be on average positive, thus indicating that participants do systematically trade according to their beliefs. This means that they buy (sell) an asset when they think the price is relatively too cheap (expensive), as measured by their subjective probability judgments. Positive expected profits would also suggest

[137] It would be also possible to compare before-trading judgments to second round trade prices and/or first round trade prices to after-trading judgments. However, it may be conjectured that (i) ex ante judgments are closely related to transactions in the first trading round, and (ii) transactions in the second trading round are closely related to ex post judgments.

that market prices arise from the interaction of thoughtful acting participants, rather than from noisy transactions by uninterested subjects. A disadvantage of this method is that it cannot control for interim changes in subjective probability judgments as individual transaction prices are only compared to before- or after-trading judgments. This implicitly assumes constant judgments during trading rounds and disregards the possibility of learning effects while trading occurs. However, the method seems to be a second-best alternative based on existing data.

Panel A of Table 3.17 shows the mean expected profit per unit of asset comparing before-trading judgments to individual trade prices of the first trading round, averaged across all traders. Besides, the Table shows the 25%, 50% (median), and 75% percentile of the distribution. For each event domain, one can obtain the expected profits resulting from purchases (buy transactions), sales (sell transactions), and the average expected profit (an aggregate metric based on all trades). All means (except for purchases in the finance and the weather domain) are statistically significantly greater than zero based on a one-sample *t*-test ($p<.001$), as indicated by triple asterisks (***)). In particular, all aggregated metrics are positive and range from 3.83 [cent] for weather-related assets to 5.92 [cent] in the sports domain. For the finance domain, for example, this means that there is an average expected profit of 4.16 [cent] per unit of traded asset. Panel B of Table 3.17, in turn, shows the mean expected profit per unit of asset comparing after-trading judgments to individual trade prices of the second trading round, averaged across all participants. All means are statistically significantly greater than zero ($p<.001$), as indicated by triple asterisks (***). Aggregated metrics are comparable in magnitude to those in Panel A and range from 3.63 [cent] for weather-related assets to 6.45 [cent] in the sports markets. The percentiles indicate that the distributions are positively skewed, i.e., the majority of participants seems to avoid negative expected profits from trades and a substantial number of traders yield significantly higher expected returns from trades. With respect to the main treatment variable one can control for differences in expected profits for the two partitions. However, no systematic differences between the two types of partition can be found. For the sake of clear arrangement, these results are not included in Table 3.17.

Table 3.17: Mean expected profits per unit of traded asset.

	N	Mean	Percentiles		
			25	50	75
Panel A: Before-Trading Judgments vs. Trading Round 1 Trade Prices					
Finance					
Purchases	192	-0.04	-5.86	0.00	5.72
Sales	192	8.36***	0.00	5.74	15.00
All Trades	192	4.16***	-1.01	2.50	8.05
Sports					
Purchases	192	3.40***	-2.46	1.79	8.92
Sales	192	8.44***	0.00	8.10	14.57
All Trades	192	5.92***	-0.35	4.74	11.16
Weather					
Purchases	192	0.93	-4.41	0.39	6.79
Sales	192	6.74***	0.00	4.37	14.15
All Trades	192	3.83***	-1.34	3.21	8.21
Panel B: After-Trading Judgments vs. Trading Round 2 Trade Prices					
Finance					
Purchases	192	3.11***	-2.27	1.08	8.45
Sales	192	7.19***	0.00	3.68	13.51
All Trades	192	5.15***	-0.19	2.31	9.07
Sports					
Purchases	192	4.78***	-1.11	3.10	11.04
Sales	192	8.13***	0.00	6.34	13.92
All Trades	192	6.45***	0.02	4.35	11.22
Weather					
Purchases	192	2.77***	-1.97	0.54	7.77
Sales	192	4.50***	-0.41	2.52	9.12
All Trades	192	3.63***	-0.61	2.20	7.33

Interestingly, the results are consistent with an *endowment effect* what becomes clear from the fact that expected profits were notably higher for sales than for purchases. The endowment effect implies that the willingness to accept (WTA) is greater than the willingness to pay (WTP) for a given good. The effect is caused by an underweighting of opportunity costs (see Thaler (1980, pp. 43–47); see also Knetsch (1989), and Kahneman, Knetsch, and Thaler (1991)). Accordingly, people tend to attribute a greater value to a certain good if they own it (literally speaking, if the good is part of their *endowment*) than if they do not possess it (i.e., the good is *not* part of their endowment). Note that attributing different values to the same good (depending on whether it is part of one's endowment or not) is inconsistent with standard economic theory. Transferred to the present study this means that expected profits generally have to be higher for sellers before they agree to sell than profits have to be for buyers before they are ready to buy, compared to normative standards.

Against the background of the trader-based analysis so far, a related question is whether (i) the trading frequency or (ii) the trading volume is influenced by *individual competence*. While subsection 3.1.4.1 focused on analyzing second order effects in aggregated market outcomes and mean judgments, competence effects are analyzed on a *per trader* basis in the following. It can be hypothesized that traders who are more competent than others are willing to trade more frequently and at higher volumes than less competent people. Due to their high competence and knowledge, they are expected to have a strong opinion of which assets to sell and to buy, and at what prices. As motivated in subsection 3.1.4.1, the analysis is based on a comparison between the most and the least competent people to achieve highest possible discriminatory power. A two-sample *t*-test is performed for each event domain and trading round to test whether (i) the average number of trades (buy or sell) and (ii) the average trading volume (buy or sell) are the same for both the high- and the low-competence group.

Table 3.18: Differences in the average number of trades (Panel A) and the average trade volume (Panel B) per trader by competence level.

Event Domain/	Finance			Sports			Weather		
Trading Round	Round 1	Round 2	*N*	Round 1	Round 2	*N*	Round 1	Round 2	*N*
Panel A: Average Number of Trades (Buy and Sell) per Trader									
HIGH_COMP	12.7	12.1	44	12.6	13.0	42	9.6	11.0	50
LOW_COMP	9.3	11.2	52	10.2	10.6	57	8.8	9.3	34
Difference	3.3***	0.9		2.5**	2.4**		0.8	1.7*	
(*p*-value, one-tailed)	(0.0024)	(0.2488)		(0.0124)	(0.0335)		(0.2195)	(0.0758)	
Panel B: Average Trading Volume (Buy and Sell) per Trader									
HIGH_COMP	41.4	43.9	44	45.6	46.3	42	32.4	38.5	50
LOW_COMP	26.2	33.7	52	30.1	31.3	57	33.0	36.8	34
Difference	15.2***	10.2**		15.5***	15.0***		-0.7	1.7	
(*p*-value, one-tailed)	(0.0002)	(0.0182)		(0.0001)	(0.002)		(0.5644)	(0.3663)	

As can be read from Panel A of Table 3.18 there is some modest support for the hypothesis that highly competent people trade more often than their less competent counterparts. The highest difference in the average number of trades (either purchases or sales) occurs for the first trading round of the finance domain ($\Delta = 3.3$ trades), but is considerably reduced in the second trading round ($\Delta = 0.9$). For the sports stimulus the difference is well pronounced and persists over the two rounds of trading ($\Delta = 2.5$ and $\Delta = 2.4$, respectively). The weakest effect occurs for the weather domain in which the differences amount to $\Delta = 0.8$ and $\Delta = 1.7$, respectively. Even if the differences do not seem to be large in terms of absolute trades, the results indicate that highly competent subjects traded more often by 8 to 36%, which is statistically significant in four out of six cases as indicated by the asterisks in Table 3.18 (***, **, * indicate significance at the 1%, 5%, 10% level, respectively). Furthermore, as can be seen in Panel B of Table 3.18, there is strong support for the hypothesis that highly competent people trade on average a higher volume than their less competent counterparts which proves true for the finance and the sports event domain, but does not hold for the weather stimulus.[138] The difference in the average trading volume (either purchases or sales) is around 15 [assets] and statistically highly significant in three out of six markets, somewhat lower in the second trading round of the finance markets, but close to zero in the weather markets. It appears as if the average number of trades and the average trading volume for less competent people are on a rather constant level across event domains and trading rounds. Corresponding values for highly competent participants are generally higher for the finance and the sports domain, but similar to those of the low-competence group for the weather event domain. A possible explanation for the weak effect in the weather domain (as opposed to the relatively strong effects for finance and sports) could be that a high score of self-rated competence does not result in a “strong opinion” on any of the offered assets. As discussed earlier, trading activity in the finance and the sports event domain is likely to be more information-driven than trading in the weather domain.

Another relevant question deals with a trader’s *portfolio composition* at the end of each trading round. Traders could try to redeploy their portfolio structure by the end of a trading round so that they only hold cash and/or unit portfolios which results in a sure payoff. Another strategy would be to “put all the eggs in one basket”, i.e., to accumulate as many of a particular asset as possible which yields a relatively high payoff if the underlying event occurs and loses everything otherwise. On the one hand, this is a

[138] Since subjects were allocated to session slots by their self-rated sports competence, they basically faced traders of the same competence level in the sports markets, but not in the finance and the weather markets.

question of individual risk aversion (i.e., how much of an expected risky payoff is given up in exchange for a sure profit). In the present setting, though, one cannot control for risk aversion since no such parameters were elicited. On the other hand, the question of a trader's exposure may have something to do with competence and knowledge. Suppose a trader who is quite uninformed in one of the three event domains. If this trader, for instance, is very uncertain about her probability judgments, she could have an interest in eliminating risk from her portfolio and generating a sure payoff. By contrast, if a trader feels very familiar with a particular event domain, and if this trader is very confident about her probability judgments, she could have a keen interest in creating an imbalanced portfolio with the prospect of a high payoff.

In the following, the impact of self-rated competence on portfolio exposure is analyzed. Again, the analysis is based on a comparison between the most and the least competent participants from the population (see subsection 3.1.4.1). For each trader the sure (or minimum) payoff is calculated based on the number of unit portfolios and cash holdings at the end of a trading round.[139] Besides, the maximum possible payoff is calculated that obtains if the most represented asset in a trader's portfolio pays 100 cent. To relate these two values, the sure payoff is divided by the maximum possible payoff. Even if it is not known whether participants have this *payoff ratio* in mind, the metric provides crude information on the risk of a portfolio. For instance, the payoff ratio would be zero if a trader "puts all her eggs in one basket", and it would be equal to 1 for a trader whose portfolio solely consists of unit portfolios and cash. A major drawback of this metric is the fact that no probabilities are taken into account. However, objective probabilities are difficult to derive for the sports and the weather events, and substantially depend on the normative model that is applied to the finance events.

Panel A of Table 3.19 reports the mean sure payoffs [in cent] for highly and low competent traders. Mean sure payoffs (minimum payoffs) range between 999 and 1,481 cent and are generally lower for highly competent subjects which is statistically significant in the sports markets and in the first trading round of the finance domain (based on a two-sample *t*-test (one-tailed); ***, **, * indicate significance at the 1%, 5%, 10% level, respectively). Similar to previous results, statistical significance is marginal for the two weather trading rounds. In general, highly competent traders earn a minimum

[139] The number of unit portfolios equals the quantity of the asset with the smallest holding in the trader's portfolio. For example, if a trader holds a portfolio consisting of 5 × asset 1, 8 × asset 2, and 10 × asset 3, the number of unit portfolios is five, equivalent to a value of 500 cent.

payoff that is considerably lower than sure payoffs earned by less competent people. Panel B, in turn, compares the mean of maximum possible payoffs for traders from the different competence levels. The mean of maximum payoffs ranges between 2,477 and 3,344 cent and is statistically significantly higher for highly competent traders in all finance and sports trading rounds (indicated by the asterisks). The difference for these two event domains ranges between 285 and 749 cent, indicating that the maximum payoff is substantially higher for highly competent traders. Again, no statistically significant effects can be found for the weather domain. The general tendency of these results is confirmed by the results of Panel C that show the payoff ratio metric (i.e., the individual ratio of the minimum to the maximum possible payoff). The sure payoffs as a fraction of the maximum payoffs are lower for highly competent people compared to less competent people. Like in Panel A, these findings are statistically significant for the first trading round in finance, and both sports trading rounds, but are not statistically significant for the weather markets. The mean payoff ratio metric is around .5 for the high-competence group (and even lower for the sports domain), indicating that on average the sure payoff would be doubled if the "correct" event occurs.

Table 3.19: Differences in sure (minimum) payoffs, maximum possible payoffs, and payoff ratio by competence level of traders.

Event Domain/	Finance			Sports			Weather		
Trading Round	Round 1	Round 2	*N*	Round 1	Round 2	*N*	Round 1	Round 2	*N*
Panel A: Mean Sure Payoff [cent]									
HIGH_COMP	1,222	1,226	44	1,048	999	42	1,310	1,255	50
LOW_COMP	1,481	1,287	52	1,399	1,418	57	1,370	1,375	34
Difference	-258**	-61		-351***	-419***		-60	-119	
(*p*-value, one-tailed)	(0.0117)	(0.3238)		(0.0033)	(0.0005)		(0.3057)	(0.1867)	
Panel B: Mean Maximum Possible Payoff [cent]									
HIGH_COMP	2,856	2,999	44	3,122	3,344	42	2,724	2,843	50
LOW_COMP	2,477	2,714	52	2,631	2,595	57	2,756	2,913	34
Difference	380***	285*		491***	749***		-31	-70	
(*p*-value, one-tailed)	(0.0016)	(0.0744)		(0.0009)	(0.0026)		(0.5814)	(0.6194)	
Panel C: Mean Payoff Ratio = Sure Payoff / Maximum Possible Payoff									
HIGH_COMP	0.48	0.47	44	0.39	0.38	42	0.51	0.49	50
LOW_COMP	0.63	0.53	52	0.57	0.60	57	0.54	0.53	34
Difference	-0.15***	-0.06		-0.18***	-0.22***		-0.02	-0.04	
(*p*-value, one-tailed)	(0.004)	(0.1544)		(0.0009)	(0.0001)		(0.3293)	(0.2683)	

To summarize, highly competent subjects were ready to forego some of a sure minimum payoff in exchange for the prospect of very high payoffs if the “predicted” event would occur. Apparently, highly competent traders took a higher exposure as their portfolio structure was more imbalanced at the end of the trading round. This conclusion holds true for the finance and sports markets, but does not hold for the weather events. Controlling for differences in *actual payoffs* between the two competence groups yields statistically significant differences only for the first sports trading round (p=.0863, two-tailed) in which highly competent participants earn on average an additional payoff of about 296 cent (2,239 vs. 1943). The results indicate that for most trading rounds a lower minimum payoff or a higher maximum payoff (i.e., a lower payoff ratio), on average does not yield any differences in actual payoffs. Portfolio structure choice in the present setting seems to be rather a question of risk propensity and safety needs between groups with different knowledge and competence. Furthermore, one can control for differences in *subjectively expected payoff values* based on the traders' individual before- and after-trading judgments and the actual portfolio structure at the end of a trading round.

Table 3.20: Subjectively expected payoff based on individual judgments by competence level of traders.

Event Domain/	Finance			Sports			Weather		
Trading Round	Round 1	Round 2	*N*	Round 1	Round 2	*N*	Round 1	Round 2	*N*
Mean Expected Payoffs based on Individual Judgments									
HIGH_COMP	2,154	2,389	44	2,371	2,395	42	2,175	2,179	50
LOW_COMP	2,109	2,135	52	2,142	2,147	57	2,104	2,150	34
Difference	45	254**		229**	248*		71	28	
(*p*-value, two-tailed)	(0.5459)	(0.0197)		(0.0314)	(0.071)		(0.364)	(0.7438)	

For each first (second) trading round the expected payoff from a trader's portfolio is calculated based on individual probability judgments elicited before (after) trading. Table 3.20 reports mean expected payoffs based on individual judgments for the different competence groups. As a result, expected values for subjects of the low-competence group are relatively constant across event domains and trading rounds and range between 2,104 and 2,150 cent. Thus, they expect to generate an extra 104–150 cent by trading (besides the initial endowment of 2,000 cent) based on individually judged probabilities. Subjects in the high-competence group expect to earn a payoff that turns out to be notably higher (229–254 cent) in the sports markets and the second trading round of the finance markets. In the two weather trading rounds and in the first finance round, they expect to earn a payoff which is similar to low competent participants. This means, for the sports markets and in part for the finance markets, highly competent people are confident to earn a payoff that is substantially higher than that expected by low competent people. For the weather domain and in part for the finance stimuli, however, the effect diminishes and subjectively expected payoffs by highly competent people are similar to those by low competent subjects.

3.2 Information treatment

3.2.1 Motivation and study design

The results of the basis treatment of Study 1 (subsection 3.1) reveal strong partition-dependence in individual probability judgments and prediction market prices. Although some mitigation of effect size could be observed in the experimental markets in the course of a trading round, the partition bias remains well pronounced in equilibrium market prices and in after-trading individual judgments. Although market institutions in the experimental asset markets were designed such that they facilitated best possible trading conditions,[140] market forces seem not to be able to eliminate the systematic distortions expressed in individual judgments. One may argue that this result is in part due to the fact that no further information was provided about the probability of events during the short periods of trading (10 minutes per trading round, 20 minutes per event domain). One may further argue that traders even forgot about the fact that there is another partition of the state space for the other half of participants (as they were told in the instructions) in the heat of the moment. To investigate the role of information conveyance and to address the questions of *whether* and *how* the effect size of partition-dependence in prediction market prices depends on the level of individual information, a follow-up study was designed. Content and type of the information had to be specified such that it *could* influence traders' beliefs and trading decisions. The follow-up study

140 Recall that the trading institution included a computerized market environment based on a continuous double auction and provided the possibility of trading multiple assets, submitting bid and asks simultaneously, accessing the order book, and did not impose any transaction costs. Traders could even short sell assets indirectly using the unit portfolio.

was conducted in May 2007 and the experimental setup was very much the same as in the basis treatment (see subsection 3.1.1) except for two major differences:[141]

1. In the instructions, the Table that contains the description of the different asset partitions was arranged in a way to clarify the relationship between the assets of different partitions. Table 3.21 shows exactly how the description was presented to the participants (translated from German):

Table 3.21: Sets of asset partitions for the different treatments in Study 1 (information treatment).

<table>
<tr><th colspan="4">Financial markets ---- Event: DAX-closing in 2 weeks</th></tr>
<tr><td colspan="4">Relevant for the payment of the assets in the financial markets domain is the Xetra-DAX closing (incl. final auction) in two weeks from today (i.e., on May 21, 2007). Xetra-DAX closing as of May 4, 2007 was 7516.76.</td></tr>
<tr><td>Market A</td><td>Partition 1</td><td>Partition 2</td><td>Market B</td></tr>
<tr><td>Asset A.1</td><td>DAX.[0 - 7437.99]</td><td rowspan="2">DAX.[0 - 7608.99]</td><td rowspan="2">Asset B.1</td></tr>
<tr><td>Asset A.2</td><td>DAX.[7438 - 7608.99]</td></tr>
<tr><td rowspan="2">Asset A.3</td><td rowspan="2">DAX.[7609+]</td><td>DAX.[7609 - 7759.99]</td><td>Asset B.2</td></tr>
<tr><td>DAX.[7760+]</td><td>Asset B.3</td></tr>
</table>

In addition, subjects were given the following hint (translated from German):

"Please note that the intervals of the assets were randomly chosen. Note further that two assets from one market can be described as equal to one asset from the other market; the Table above clarifies this for us. It is further evident that the sum of market prices of two assets from one market (e.g., assets A.1 and A.2) can be compared to the corresponding market price of one asset from the other market (in this case: asset B.1) since they incorporate exactly the same payoff claim. As for the current market prices and bid/ask quotes for the assets from the other market you will be informed of them later at two different points of time during each trading round."

2. During each 10-minute trading round there were two points in time at which subjects were informed on the current market price, the best buy and the best sell offers

[141] A minor difference was the fact that subjects were not allocated to session slots based on their self-stated soccer competence during the registration process. Thus, all participants allocated themselves to one of the twelve session slots. All other parameters of the experiment exactly correspond to the basis treatment design. See Appendix II for a Table with session-individual DAX interval boundaries in the information treatment.

of each of the assets from the corresponding market (that traded assets from the other partition at the same time). The first “market snapshot” appeared after four minutes of trading, the second “market snapshot” appeared three minutes later, after seven minutes of trading. Participants were not informed in advance of the exact time at which news would arrive.[142] The information was presented to the subjects through a message window which popped up on their trading screen automatically (see Figure 3.14 which refers to the practice market).[143]

Figure 3.14: Message window in Study 1 (information treatment).

Against the background of the very strong effect of partition-dependence in the basis treatment of Study 1, the design modifications of the information treatment are suited to test for partition-dependence under “tightened” conditions. First, the fact that the state space was notionally divided into four (more or less) randomly chosen intervals was presented to the participants in a more salient way and it became clear from the instructions that two assets from one partition were packed to form a single asset in the other partition (and that another asset was unpacked into two assets in the other partition). Besides, it was explicitly pointed out to the subjects that the summed market prices of two assets from one partition could be directly compared to the market price of the corresponding packed asset from the other partition, since these positions were equivalent in terms of their final payoff structure. Second, participants were reminded of this design feature twice a trading round by receiving (real-time) market quotes and prices from the other market. Providing the subjects with price information from the

---

[142] Under some circumstances, it can be unfavorable to trade immediately before traders expect to get new information (see Sunder (1995, p. 455)). So informing traders about the exact times at which news arrived could result in discontinuous trading which was not intended.

[143] An alternative way of providing subjects with this information would be to allow them looking at market data from the other market whenever they want. However, the message window was chosen to provide the information in order to control for available sets of information at a given point in time.

other market should also reinforce them to reflect on their beliefs and to reconsider market prices in their own markets. By the type of message participants receive information that reflect a kind of dollar-weighted aggregated opinion about probabilities of events, generated by eight other traders trading a set of assets that is directly related to a trader's own assets. In contrast to the basis treatment of Study 1, in which the two markets of a session slot acted completely independently, the message transfer in the information treatment creates an effective link between the two markets of a session slot from the fourth minute in each trading round.[144]

With reference to the hypotheses derived in subsection 3.1.3, reduced partition-dependence is expected in individual probability judgments due to the more salient presentation of the intervals and the relationship between the two sets of assets in the instructions. Note that this form of presentation could induce a different framing of the events, away from an ignorance prior of one third for each asset and toward an allocation of [.25, .25, .50] for assets of partition 1 and [.50, .25, .25] for assets of partition 2. Furthermore, reduced partition-dependence is expected for equilibrium market prices due to some convergence of market prices for assets (or combination of assets) that refer to the same event. Convergence of market prices could be caused by both rising prices for the packed event and falling prices for the unpacked events. To a certain extent, though, partition-dependence in market prices is expected to remain since traders in the information treatment were still not able to trade across markets to exploit arbitrage opportunities arising from partition-dependence. In addition, subjects do not have any information on *how competent* their *counterparts* are; neither do they have details on the other market's *liquidity* which they could find important to assess the reliability of the received information. As Brüggelambert (1999, p. 109) states, it is the sureness about the reliability of a source of information that turns "news" into "information". Thus, people who receive a market snapshot from another market do not have reason *per se* to assume that *they themselves* are "wrong" with their market prices. Furthermore, the *prior-information model* posits that peoples' expectations are not at all influenced by observed prices. According to this model, market participants only use their prior information instead, which is exogenous to the price formation process (see, e.g., Forsythe and Lundholm (1990)). By consequence, prices would emerge straightly forward from the dynamics of demand and supply as in a Walrasian system (Plott and Sunder (1982)).

[144] Recall that after four minutes of trading the first message was transmitted.

However, many experimental studies show that prices do not concur with the prior-information equilibriums (in particular, see subsection 2.3.2.3) which gives reason to assume that traders do in fact take into account information that is conveyed by prices when updating their beliefs. Hence, one may suppose that traders do not *per se* ignore the information messages in the present study.

In response to the message transmission some traders might adjust their quotes which, in turn, could cause prices to move in a direction of reduced partition-dependence. Possibly, though, this could induce *herd behavior* by other traders.[145] Herding in the context of financial markets refers to the phenomenon that traders sometimes react pro-cyclical and by this means trigger self-energizing price movements; traders then ignore private information and mimic a particular market pattern. This, in turn, could lead to reverse partition-dependence in the context of the present treatment. Furthermore, *overconfidence* of subjects could see to it that traders do not care much about market prices of some other markets and, by consequence, that the degree of partition-dependence is not affected by information exchange. Overconfident people overvalue their private information and tend to ignore new signals (Biais et al. (2005)).

To summarize the hypotheses with regard to the information treatment of Study 1, reduced partition-dependence is expected to be expressed in pre-trading individual probability judgments due to a more obvious presentation (and additional textual explanations) of intervals, assets and their relationship in the instructions. In addition, the crosswise message transmission is expected two have two effects: first, participants should be stimulated to reconsider and adjust their beliefs by comparing them to market prices of the corresponding market. Second, people (in particular less competent subjects) might think that there are more competent traders in the other market and that it is worth trading assets at prices consistent with prices from the other market. This, in turn, would cause convergence of market prices and would result in reduced partition-dependence. However, partition-dependence is not expected to be fully eliminated from market prices due to some uncertainty that remains about the reliability of received messages. In addition, other information related biases like, e.g., herd behavior and overcon-

145 For recent surveys of herding in financial markets, see Hirshleifer and Teoh (2003), and Chamley (2004). Cipriani and Guarino (2005) analyze herd behavior in a laboratory financial market and find that herd behavior seldom occurred in their experimental setting. However, in some cases subjects preferred to ignore their private information and abstain from trading which is not captured by theory.

fidence could affect market prices and could thus overlay the bias-reducing effects on the partition bias.

3.2.2 Main results and impact of information transmission

Table 3.22, which has the same structure as Table 3.13 for the basis treatment of Study 1, summarizes the main results of the information treatment. The Table contains the mean pre-trading individual probability judgments, equilibrium market prices after the second trading round (quantity-weighted average of last three trade prices) and the mean post-trading individual probability judgments. The effect size of partition-dependence in *pre*-trading judgments (i.e., $p_{J,i}(I_1) + p_{J,i}(I_2) - p_{J,i}(I_1 \cup I_2)$) is .263, .211 and .201 for the finance, sports and weather event domains, respectively. The partition bias in *post*-trading judgments is .240, .216 and .212 for the finance, sports and weather events, respectively. A similar Table showing differences in medians can be obtained from Appendix III. Results are very similar to those derived from means. All reported differences are statistically highly significantly different from zero ($p<.0001$). Comparing post-trading judgments to pre-trading judgments in the information treatment, it appears that the effect size is pretty much on the same level for each event domain (i.e., there is not much variability in before- and after-trading judgments). With respect to the effect size expressed in equilibrium market prices, it turns out that the partition bias between summed prices (divided by 100) of unpacked assets, $P_i^*(I_1) + P_i^*(I_2)$, and the packed asset, $P_i^*(I_1 \cup I_2)$, is .153, .189 and .205 for the finance, sports and weather domain, respectively. Corresponding values for the high intervals, $P_i^*(I_3) + P_i^*(I_4) - P_i^*(I_3 \cup I_4)$, are .156, .189 and .191. Averaging the effect size for the low and the high intervals yields an average bias of .154, .189 and .198 for the finance, sports and weather domains, respectively. Like for the judgments, all reported differences are statistically highly significantly different from zero ($p<.01$). Comparing equilibrium market prices to the very similar pre- and post-trading judgments in the information treatment, it is striking that there is a strong reduction of bias size in the finance markets, a small reduction in the sports events, and virtually no effect of bias reduction in the weather markets.

Table 3.22: Mean equilibrium market prices (2nd trading round) and individual judgments (pre-trading and post-trading) in the information treatment of Study 1.

Treatment		Mean Judged Probability/Equilibrium Prices								
		Finance			Sports			Weather		
		Pre-Trading Individual Judgment	Equilibrium Market Prices (Round 2)	Post-Trading Individual Judgment	Pre-Trading Individual Judgment	Equilibrium Market Prices (Round 2)	Post-Trading Individual Judgment	Pre-Trading Individual Judgment	Equilibrium Market Prices (Round 2)	Post-Trading Individual Judgment
1	I_1	0.209	0.161	0.184	0.241	0.176	0.203	0.165	0.116	0.142
1	I_2	0.474	0.451	0.497	0.403	0.405	0.420	0.324	0.316	0.323
	$I_1 + I_2$	0.683	0.613	0.681	0.644	0.580	0.623	0.489	0.432	0.465
2	$I_1 \cup I_2$	0.420	0.459	0.441	0.433	0.391	0.408	0.288	0.227	0.253
	PD difference	*0.263*	*0.153*	*0.240*	*0.211*	*0.189*	*0.216*	*0.201*	*0.205*	*0.212*
2	I_3	0.392	0.380	0.396	0.358	0.376	0.378	0.368	0.358	0.385
2	I_4	0.188	0.153	0.164	0.209	0.228	0.214	0.344	0.399	0.362
	$I_3 + I_4$	0.580	0.533	0.559	0.567	0.605	0.593	0.712	0.756	0.747
1	$I_3 \cup I_4$	0.317	0.378	0.319	0.356	0.416	0.377	0.511	0.565	0.535
	PD difference	*0.263*	*0.156*	*0.240*	*0.211*	*0.189*	*0.216*	*0.201*	*0.191*	*0.212*
Average PD difference			***0.154***			***0.189***			***0.198***	

Table 3.23 compares individual probability judgments, elicited before trading and thereafter, from the basis treatment with the judgments from the information treatment (N=96 in each cell).[146] This allows testing for interaction effects between partition type and information treatment. The analysis is performed by calculating the difference of differences, i.e., the difference in partition bias for the two different "info treatments"; one-sided p-values are reported based on a 2×2 factorial analysis of variance (ANOVA). Looking at pre-trading judgments first, it turns out that partition-dependence in the information treatment is reduced by .048 (a 15.7% decrease), .05 (–19.2%) and .077 (–27.7%) in the finance, sports and weather domain, respectively. The decline is statistically significant at the 10%-level for the finance domain, and at the 5%-level for the weather domain (as indicated by the asterisks), but is not statistically significantly different from zero for the sports stimulus. For the finance and sports event domains, the bias reduction can be mainly attributed to both lower judgments for the sum of unpacked events and higher judgments for the packed event; for the weather stimulus, bias reduction is due to higher judgments for the packed event.

146 Due to technical errors, two judgments are missing for the weather events (basis treatment).

Table 3.23: Interaction effects between partition type and treatment (basis vs. information treatment) in pre- and post-trading probability judgments.

	Pre-Trading Judgments		Post-Trading Judgments		N_{basis} / N_{info}
Treatment	Basis	Info	Basis	Info	
Mean probability judgments					
Finance					
$p(I_1) + p(I_2)$	0.717	0.683	0.699	0.681	96 / 96
$p(I_1 \cup I_2)$	0.405	0.420	0.442	0.441	96 / 96
Difference	0.312	0.263	0.257	0.240	
Difference of differences	-0.048*		-0.017		
(p-value, one-tailed)	(.0839)		(.3255)		
Sports					
$p(I_1) + p(I_2)$	0.678	0.644	0.684	0.623	96 / 96
$p(I_1 \cup I_2)$	0.417	0.433	0.428	0.408	96 / 96
Difference	0.261	0.211	0.256	0.216	
Difference of differences	-0.050		-0.040		
(p-value, one-tailed)	(.126)		(.167)		
Weather					
$p(I_1) + p(I_2)$	0.477	0.489	0.422	0.465	95 / 96
$p(I_1 \cup I_2)$	0.199	0.288	0.196	0.253	95 / 96
Difference	0.278	0.201	0.226	0.212	
Difference of differences	-0.077**		-0.014		
(p-value, one-tailed)	(.0119)		(.3283)		

While partition-dependence remains well pronounced in terms of absolute bias strength across event domains, the effect size seems to be reduced as a result of the modified presentation of intervals and assets in the instructions.[147] For post-trading judgments it turns out that the difference in partition-dependence between the basis treatment and the information treatment almost shrinks to zero for the finance and weather domains, i.e., there are virtually no differences in bias strength between the two "info treatments". For the sports stimulus, partition-dependence in the info treatment is slightly lower than in the basis treatment by .04 (–15.6%). However, none of the interaction terms for post-trading judgments is statistically significant.

The next question is to what extent market prices and therefrom derived partition-dependence is influenced by the real-time price information disseminated to the

[147] Recall that judgments were collected right before trading in each event domain began. Thus, one has to take into account that participants already gathered some trading experience before providing their judgments for the second and the third event domain (the order in which the three event domains appeared was perfectly counterbalanced). Traders in the information treatment additionally gathered experience in receiving information via the message window which might have influenced their before-trading judgments in the subsequent event domains. Therefore, one must be careful in attributing reduced partition-dependence in pre-trading judgments solely to the different presentations in the instructions.

participants while trading. As it was hypothesized that market prices converge in response to the exchange of market data between experimental groups, an obvious starting point is to compare both treatments with respect to the *convergence measure* introduced in subsection 3.1.3.2. Recall that some degree of price convergence due to learning and experience was already found in the basis treatment of the study in which no cross-market information was provided. Price convergence due to message exchange, though, should be more pronounced.

Table 3.24: Measures of price convergence over time (info treatment).

	Slope Coefficient β from Linear Regression		
	Finance	Sports	Weather
Avg. $[P(I_1)+P(I_2)]$	-0.0121	-0.0111	-0.0138
Avg. $P(I_1 \cup I_2)$	-0.0004	-0.0027	-0.0050
Slope difference	-0.0117	-0.0084	-0.0088
Avg. $[P(I_3)+P(I_4)]$	-0.0096	-0.0104	0.0036
Avg. $P(I_3 \cup I_4)$	-0.0028	0.0033	0.0060
Slope difference	-0.0068	-0.0137	-0.0024
Convergence measure (mean slope difference)	-0.0092	-0.0111	-0.0056

Table 3.24 presents the results of the convergence measure analysis for the information treatment that can be directly compared to the results obtained for the basis treatment presented in Table 3.12 (see subsection 3.1.3.2).[148] Comparing the (mean) slope differences from both treatments, though, suggests that convergence of market prices is even more pronounced in the basis treatment.[149] However, linear regression coefficients based on the whole trading period of two trading rounds only provide a rough measure of price convergence.

To further assess whether market price convergence is likely to be message-driven or rather the result of general learning and trading experience (as in the basis treatment), a 2×2 factorial analysis of variance (ANOVA) is suited to shed light on

[148] For graphs that show the aggregated development of asset prices over time for the three event domains in the information treatment, see Appendix IV. These graphs directly correspond to the graphs in Figure 3.4 to Figure 3.6 in subsection 3.1.3.2.

[149] Recall that a negative (positive) slope difference indicates convergence (divergence) of market prices. The more negative the slope difference, the more prices from differently partitioned markets do converge.

possible interaction effects between partition type and information treatment in terms of equilibrium market prices. Table 3.25 compares mean equilibrium market prices of the basis treatment with equilibrium prices of the information treatment (N=12 in each cell) and reports p-values for ANOVA interaction terms (one-tailed; ***, **, * indicate statistical significance at the 1%, 5%, 10% level, respectively). It turns out that results are mixed. There is a considerable bias reduction in equilibrium prices of the info treatment for the first and second trading rounds of the finance and the sports markets (except for the high intervals of the sports markets). This result can be largely attributed to a decline in the sum of market prices for the unpacked assets (note that there is some "overround" in the basis treatment markets, i.e., equilibrium market prices in partition 1 sum to more than 1.0). For the weather event domain, by contrast, results are ambiguous: they indicate a bias reduction in the info treatment based on trading round 1 market prices, but indicate a stronger effect of the partition bias in the info treatment based on trading round 2 market prices. Particularly noticeable for the weather domain, though, is the strong bias reduction between the first and the second trading round in the basis treatment (while the bias between trading rounds is only slightly reduced in the information treatment). In most cases, though, statistical significance of the differences in effect size between the basis treatment and the information treatment is marginal. Obviously, there is a need for a more detailed analysis of price reactions to the arrival of real-time market data from the differently partitioned market.

Since the exchange of current market information constitutes an explicit interrelationship between the two markets of a session slot, it is worth looking at the *development of price differences* over time on a session-based level. Figure 3.15 to Figure 3.20 present overall time-series for the sum of market prices of unpacked assets ($P(I_1) + P(I_2)$ and $P(I_3) + P(I_4)$, respectively) against the corresponding packed asset ($P(I_1 \cup I_2)$ and $P(I_3 \cup I_4)$, respectively). Each Figure refers to either the low or the high-interval assets of one of the three event domains and contains a price chart for each of the twelve session slots. The red curves belong to partition 1 and the blue curves belong to partition 2. The black vertical line (at 600 sec.) separates the first from the second trading round and the dashed gray vertical lines (after 240/420 and 840/1,020 sec., respectively) indicate the points of time at which messages were transmitted.[150] Note again that markets are

[150] Due to the limited performance of the application server, messages were usually delivered with a short delay (mean delivery time was 16 seconds).

effectively unrelated during the first four minutes of each trading round, i.e. before the first set of information was transferred.

Table 3.25: Interaction effects between partition type and treatment (basis vs. information treatment) in mean equilibrium market prices.

Treatment	Trading Round 1 Basis	Trading Round 1 Info	Trading Round 2 Basis	Trading Round 2 Info	N_{basis} / N_{info}
Mean equilibrium market prices					
Finance					
$p(I_1) + p(I_2)$	0.764	0.628	0.713	0.613	12 / 12
$p(I_1 \cup I_2)$	0.421	0.455	0.424	0.459	12 / 12
Difference	0.343	0.172	0.289	0.153	
Difference of differences	-0.170**		-0.136**		
(p-value, one-tailed)	(0.0221)		(0.0334)		
$p(I_3) + p(I_4)$	0.619	0.540	0.581	0.533	12 / 12
$p(I_3 \cup I_4)$	0.363	0.352	0.336	0.378	12 / 12
Difference	0.257	0.188	0.246	0.156	
Difference of differences	-0.069		-0.090		
(p-value, one-tailed)	(0.1790)		(0.1155)		
Sports					
$p(I_1) + p(I_2)$	0.704	0.581	0.720	0.580	12 / 12
$p(I_1 \cup I_2)$	0.401	0.410	0.439	0.391	12 / 12
Difference	0.303	0.171	0.280	0.189	
Difference of differences	-0.132**		-0.091		
(p-value, one-tailed)	(0.0409)		(0.1293)		
$p(I_3) + p(I_4)$	0.583	0.644	0.568	0.605	12 / 12
$p(I_3 \cup I_4)$	0.372	0.434	0.391	0.416	12 / 12
Difference	0.211	0.210	0.177	0.189	
Difference of differences	-0.001		0.012		
(p-value, one-tailed)	(0.4950)		(0.4398)		
Weather					
$p(I_1) + p(I_2)$	0.425	0.475	0.303	0.432	12 / 12
$p(I_1 \cup I_2)$	0.160	0.245	0.149	0.227	12 / 12
Difference	0.265	0.230	0.154	0.205	
Difference of differences	-0.035		0.051		
(p-value, one-tailed)	(0.2659)		(0.1974)		
$p(I_3) + p(I_4)$	0.887	0.737	0.850	0.756	12 / 12
$p(I_3 \cup I_4)$	0.646	0.526	0.707	0.565	12 / 12
Difference	0.241	0.211	0.143	0.191	
Difference of differences	-0.030		0.049		
(p-value, one-tailed)	(0.3266)		(0.2224)		

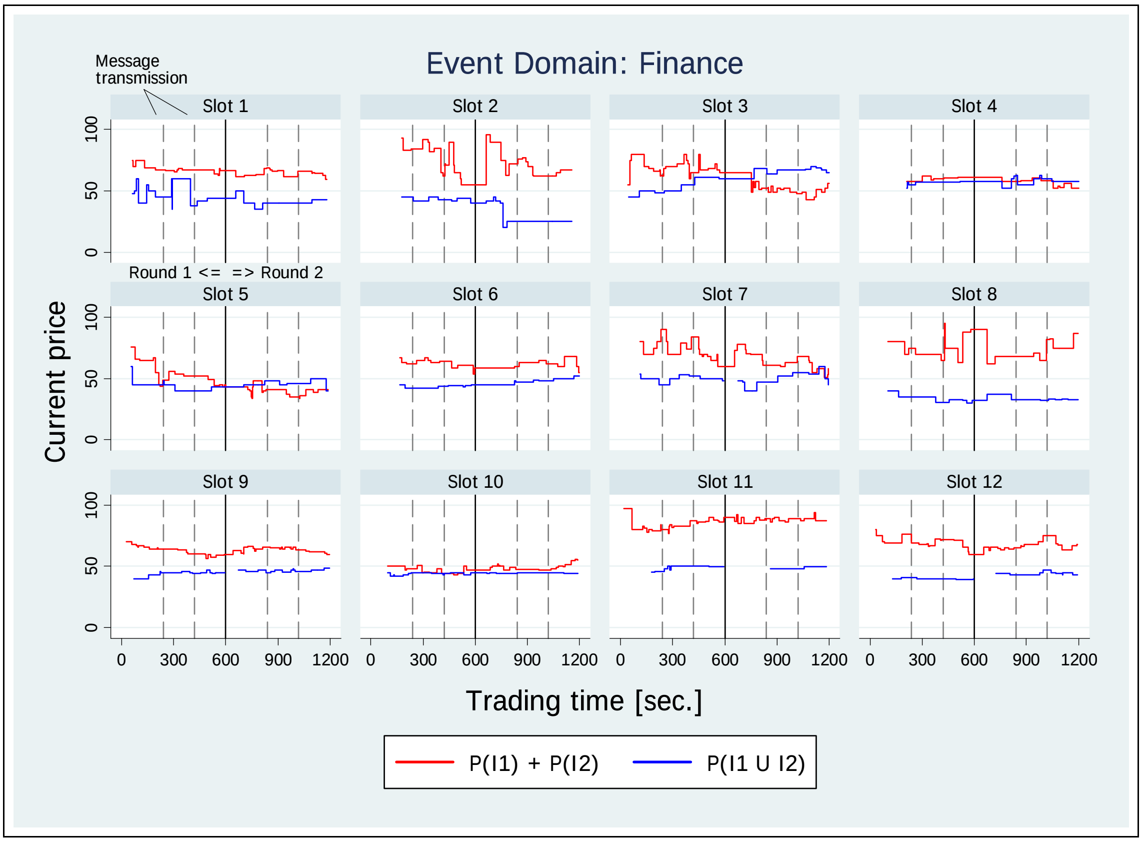


Figure 3.15: Overall time-series price chart (info treatment) by slot (domain: finance, low intervals).

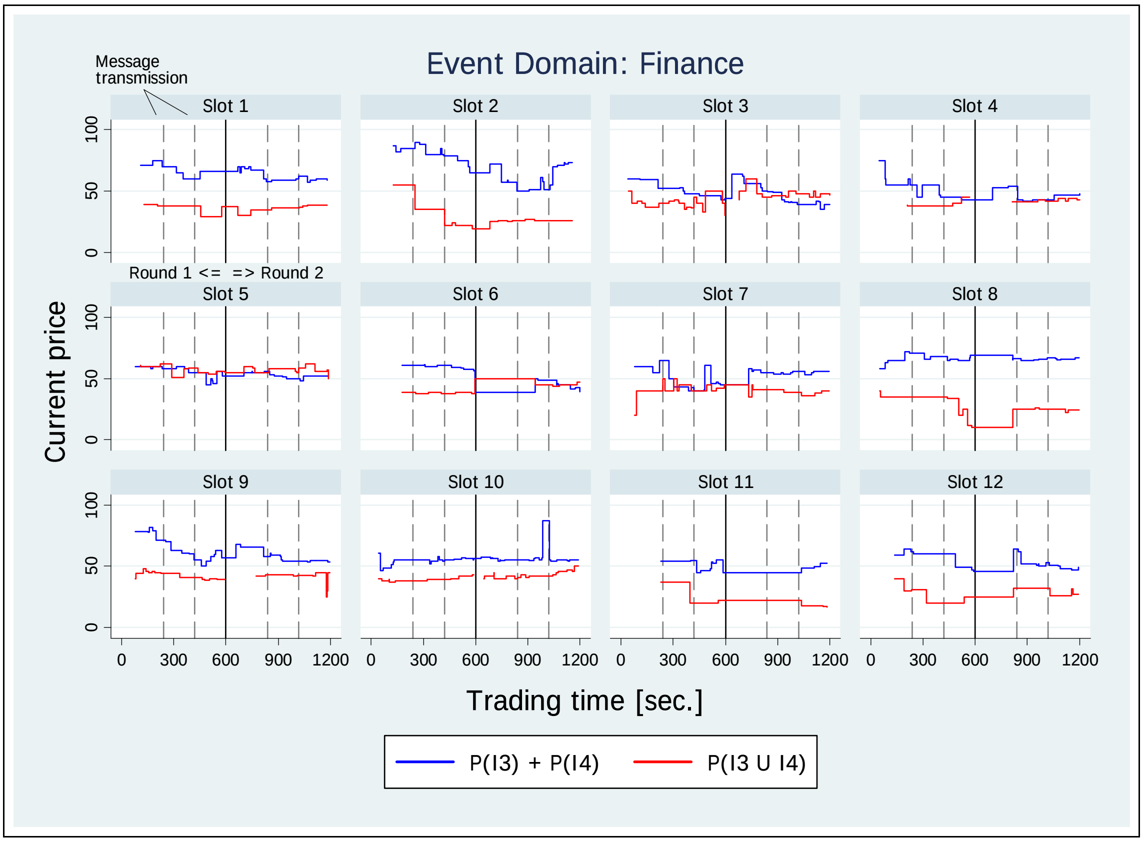


Figure 3.16: Overall time-series price chart (info treatment) by slot (domain: finance, high intervals).

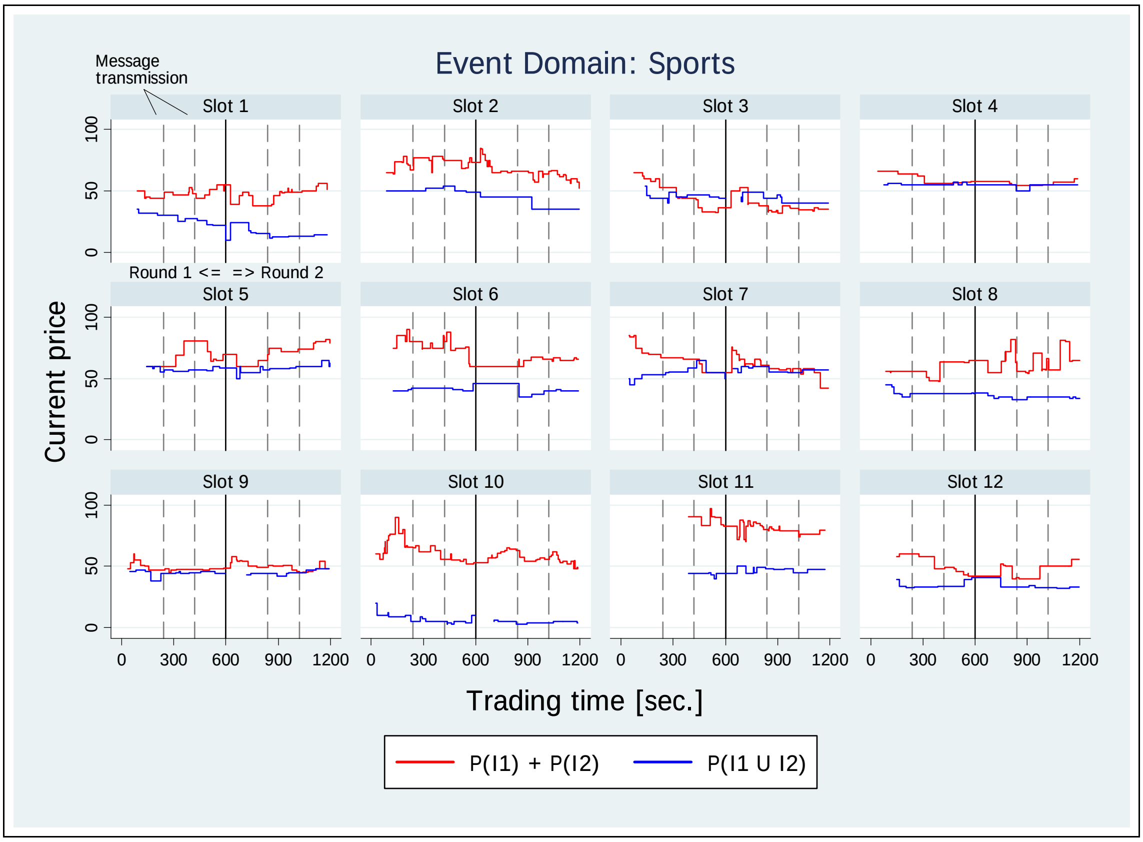


Figure 3.17: Overall time-series price chart (info treatment) by slot (domain: sports, low intervals).

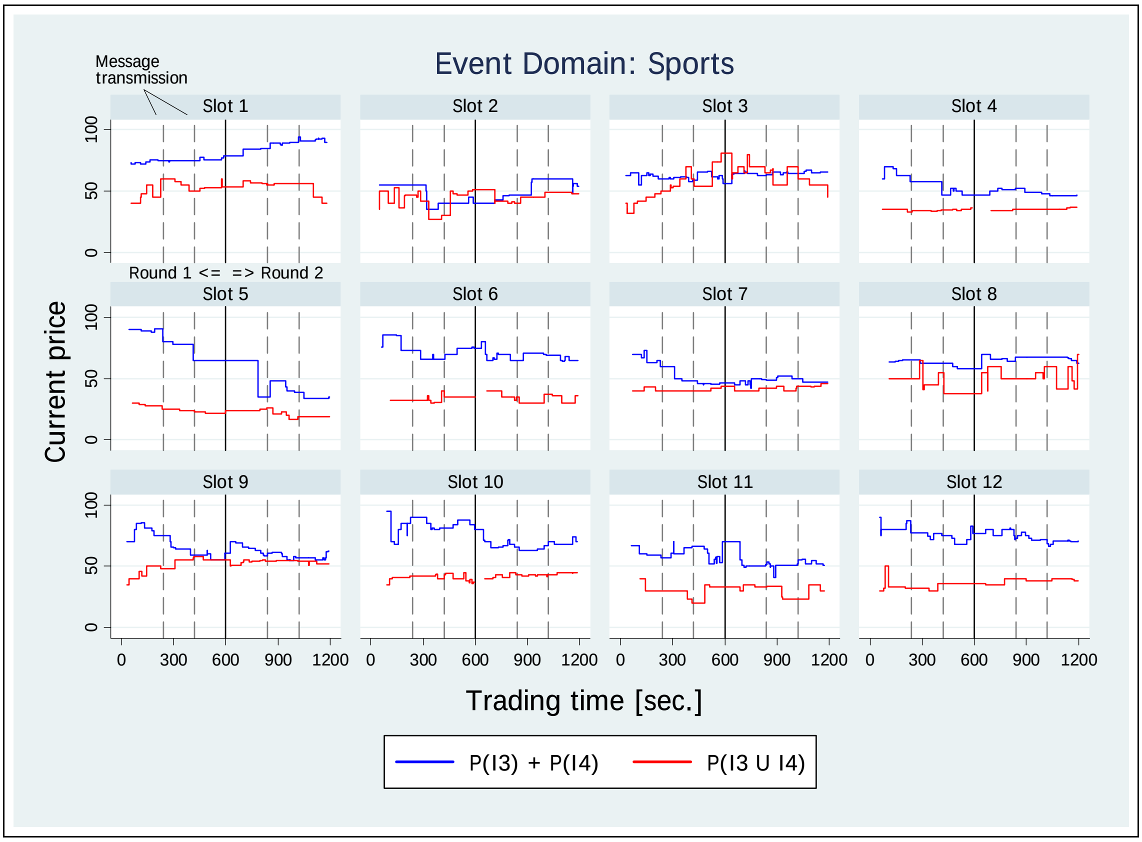


Figure 3.18: Overall time-series price chart (info treatment) by slot (domain: sports, high intervals).

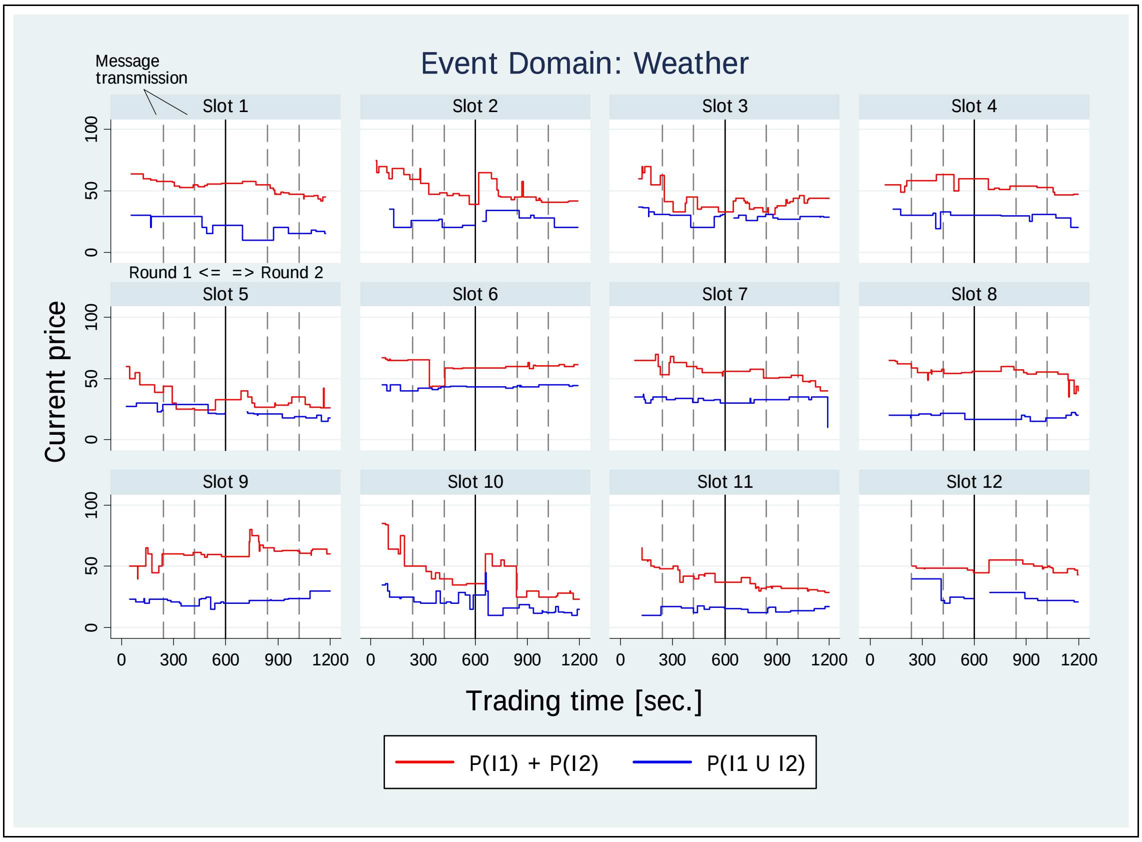


Figure 3.19: Overall time-series price chart (info treatment) by slot (domain: weather, low intervals).

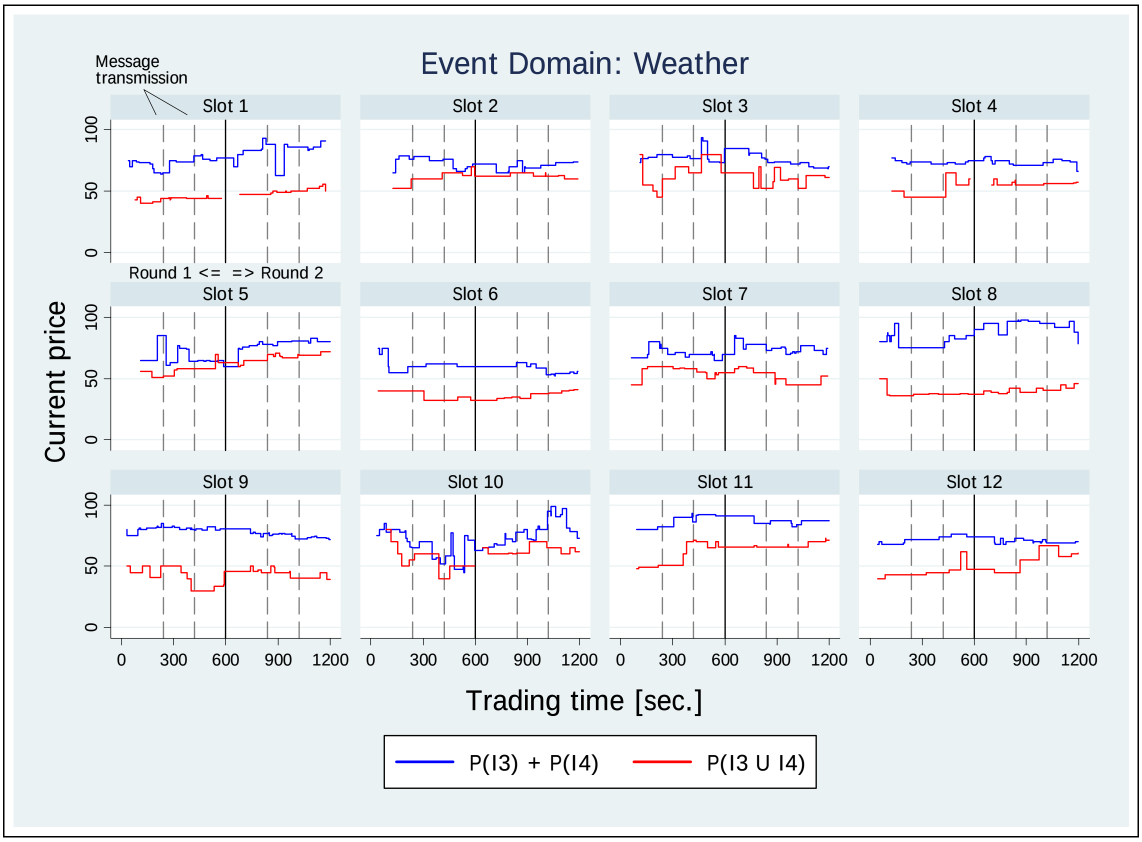


Figure 3.20: Overall time-series price chart (info treatment) by slot (domain: weather, high intervals).

The price charts of Figure 3.15 to Figure 3.20 reveal a great diversity of different patterns of price reactions. In principle, two basis scenarios can be distinguished which are characterized by whether some degree of partition-dependence is implicitly expressed by the first message or not:

(i) If no partition-dependence is reflected by market prices at the time when the first information appears (the first dashed line of each trading round), prices for the considered intervals seem to continue oscillating around their base level (like, e.g., in slots 5 and 10 of the finance domain, or slot 3 of the sports stimulus).

(ii) In most cases, though, partition-dependence is well pronounced in market prices after the first four minutes of trading what became obvious to the participants when they received the first information about market data from the other market.

- ➢ In some of these cases, partition-dependence is temporarily reduced, but reincreases over the course of the trading period (like, e.g., in slot 1 of the finance domain (first trading round)).
- ➢ In other slots, partition-dependence seems to be reduced permanently (like, e.g., the high intervals in slot 4 and 9 of sports).
  - In some of these markets the reduction of partition-dependence appears to be caused by a reduction of prices for the unpacked-event assets (like for the low intervals in slot 2 of finance or the low intervals in slot 2 of weather).
  - In other markets this seems to be due to an increase of the price for the packed-interval asset (e.g., slot 6 of finance (particularly in the second trading round)).
  - In some slots, prices apparently converge both-way (e.g., slot 7 of finance).

In a few instances, there is some weak evidence for reverse partition-dependence (e.g., slots 3 and 5 of finance). In some of the slots, though, information seems to have virtually no impact on prices (e.g., slot 8 of finance, slots 1 and 11 of sports, or slot 9 of weather). On the whole, in many of the experimental markets price developments seem to be a mixture of some of the described patterns. Moreover, price reactions are often different for the low and the high intervals. Accordingly, it is hard to derive a general conclusion from observed price patterns. In addition, if at all adjustments occur, they are quite different in pace and magnitude. Finally, it is not even clear whether price adjustments are a response to new information or rather a result of general learning and experience.

To deal with this question in a more formal way, define

$$PD_{s,t}\,(I_1 \cup I_2) := P_{s,t}\,(I_1) + P_{s,t}\,(I_2) - P_{s,t}\,(I_1 \cup I_2) \quad \text{and} \tag{3.3}$$

$$PD_{s,t}\,(I_3 \cup I_4) := P_{s,t}\,(I_3) + P_{s,t}\,(I_4) - P_{s,t}\,(I_3 \cup I_4) \tag{3.4}$$

to represent the degree of partition-dependence in slot $s = 1, \ldots, 12$ at point $t = 1, 2$ (the two points of time at which the first and the second message was transferred to the other market, respectively). Let $PD^*_{s,3}\,(I_1 \cup I_2)$ and $PD^*_{s,3}\,(I_3 \cup I_4)$ represent the size of partition-dependence expressed in mean equilibrium market prices (based on the quantity-weighted average of the last three trade prices of a trading round). Further define $\overline{PD}_{s,2}(I_1 \cup I_2)$ and $\overline{PD}_{s,2}(I_3 \cup I_4)$ as the time-weighted average size of partition-dependence in the time between the two messages arrived (i.e., between the fourth and the seventh minute), and $\overline{PD}_{s,3}(I_1 \cup I_2)$ and $\overline{PD}_{s,3}(I_3 \cup I_4)$ as the time-weighted average amount of partition-dependence in the time between the second message and the end of the trading period (i.e., from the seventh minute until the end). Figure 3.21 illustrates the five measures of partition-dependence in the information treatment.

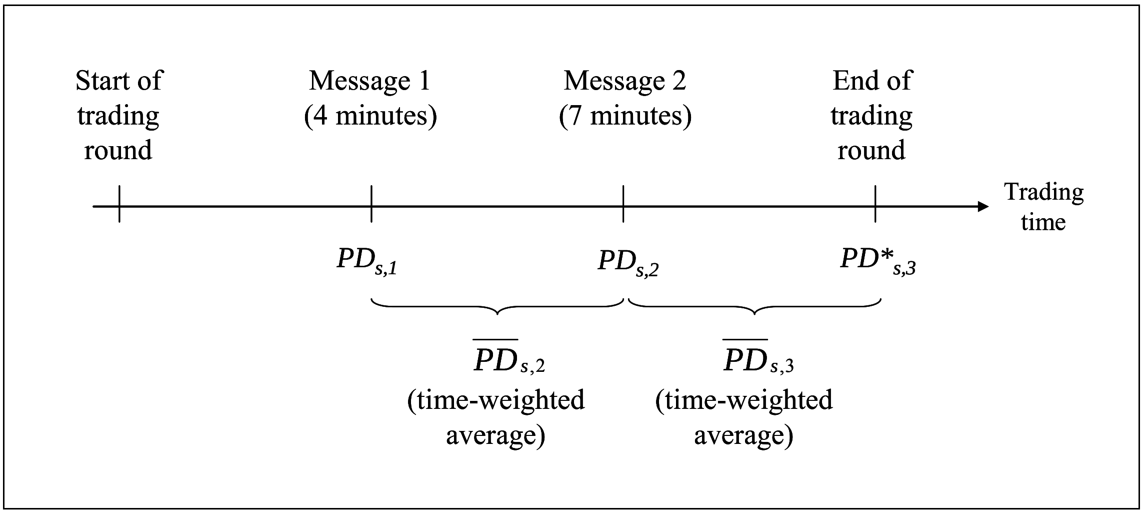


Figure 3.21: Measures of partition-dependence in the information treatment.

It is now possible to carry out a more systematic analysis of how partition-dependence varies throughout the trading period and how it is related to the information received from the differently partitioned market. The following four metrics of partition-dependence can be derived in each market:

(i) The average amount of partition-dependence between the two messages ($\overline{PD}_{s,2}$) compared to the degree of partition-dependence expressed in the first message ($PD_{s,1}$);

(ii) the average size of partition-dependence between the second message and the end of the trading round ($\overline{PD}_{s,3}$) compared to the degree of partition-dependence expressed in the second message ($PD_{s,2}$);

(iii) partition-dependence expressed in mean equilibrium market prices at the end of a trading round ($PD^{*}_{s,3}$) compared to partition-dependence expressed in the first message ($PD_{s,1}$), and finally

(iv) the average degree of partition-dependence in the last sub-period ($\overline{PD}_{s,3}$) compared to the average degree of partition-dependence between the two messages ($\overline{PD}_{s,2}$).

If information transmission has a de-biasing effect on market prices, partition-dependence is expected to be reduced in later stages of a trading period. Accordingly, the following four hypotheses can be derived:[151]

Hypothesis 3.7.(i):

H_0(a): $\overline{PD}_2(I_1 \cup I_2) - PD_1\,(I_1 \cup I_2) < 0$ and

H_0(b): $\overline{PD}_2(I_3 \cup I_4) - PD_1\,(I_3 \cup I_4) < 0$

[151] Note that hypotheses 3.7.(i)-(iv) may induce some interpretational difficulties in cases of *reverse* partition-dependence. While reduced partition-dependence generally leads to negative values for H_0(a) and H_0(b), a negative value may also reflect *increased* reverse partition-dependence. However, instances of reverse partition-dependence are rare in the present data and should not have a major impact on median differences. An alternative approach would be to run a simple linear regression (without constant term) for each of the derived metrics (i)–(iv), and to interpret the resulting slope coefficient as a measure of bias reduction.

Hypothesis 3.7.(ii):

H_0(a): $\overline{PD}_3(I_1 \cup I_2) - PD_2(I_1 \cup I_2) < 0$ and

H_0(b): $\overline{PD}_3(I_3 \cup I_4) - PD_2(I_3 \cup I_4) < 0$

Hypothesis 3.7.(iii):

H_0(a): $PD^*_3(I_1 \cup I_2) - PD_1(I_1 \cup I_2) < 0$ and

H_0(b): $PD^*_3(I_3 \cup I_4) - PD_1(I_3 \cup I_4) < 0$

Hypothesis 3.7.(iv):

H_0(a): $\overline{PD}_3(I_1 \cup I_2) - \overline{PD}_2(I_1 \cup I_2) < 0$ and

H_0(b): $\overline{PD}_3(I_3 \cup I_4) - \overline{PD}_2(I_3 \cup I_4) < 0$

For each of the four hypotheses above, there are twelve paired data points per event domain and trading round. Table 3.26 reports the results of the hypotheses tests (based on a Wilcoxon matched-pairs signed-ranks test), broken down into event domains and trading rounds. Note that all values were rescaled (× 100) for the sake of clarity. For each sub-hypothesis H_0(a) (for the low intervals) and H_0(b) (for the high intervals) the Table contains median values for the respective measures of partition-dependence to gain some feeling for the effect size. The third row shows the median of differences and the fourth row reports *p*-values (one-sided; ** and * indicate statistical significance at the 5% and 10% level, respectively).

In general, the results are supportive of the conjecture that there is some reduction in effect size after new market data have arrived (a negative sign of the median of differences indicates reduced partition-dependence, which is the case for the majority of event domain-trading round combinations). The median difference of effect size across all hypotheses ranges between 2.18 (an increase of partition-dependence) and –8.09 (a reduction of partition-dependence) in terms of absolute differences. For the first trading round of the sports markets all metrics indicate a statistically significant reduction of partition-dependence, which is considerable in magnitude. For the second trading round of the sports markets, though, the data reflects a slight tendency for increased partition-dependence. Some meaningful reductions of effect size can also be observed for the finance and the weather markets.

Table 3.26: Results of hypotheses tests (3.7.(i)–3.7.(iv)).

Median	Finance		Sports		Weather	
	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2
Hypothesis 3.7.(i):						
H_0(a):						
$\overline{PD}_2(I_1 \cup I_2)$	19.58	17.09	18.44	20.20	26.29	18.25
$PD_1(I_1 \cup I_2)$	20.75	17.50	18.50	19.00	28.75	17.50
Median of differences	-0.24	-1.35	-3.38**	1.43*	-3.04*	0.27
(p-value, one-tailed)	(0.4688)	(0.1047)	(0.0250)	(0.0790)	(0.0790)	(0.3475)
H_0(b):						
$\overline{PD}_2(I_3 \cup I_4)$	20.41	14.88	19.44	18.24	27.57	17.98
$PD_1(I_3 \cup I_4)$	19.51	14.25	25.50	17.25	25.00	19.75
Median of differences	-2.25	-0.66	-3.37**	2.18*	-0.81	-1.35
(p-value, one-tailed)	(0.1361)	(0.2669)	(0.0356)	(0.0584)	(0.3769)	(0.1941)
Hypothesis 3.7.(ii):						
H_0(a):						
$\overline{PD}_3(I_1 \cup I_2)$	16.90	14.25	18.28	22.45	23.98	18.08
$PD_2(I_1 \cup I_2)$	24.00	15.50	19.75	19.75	26.00	17.70
Median of differences	-2.84**	-1.68*	-2.75**	0.60	-1.73	-2.54**
(p-value, one-tailed)	(0.0422)	(0.0681)	(0.0498)	(0.4070)	(0.2164)	(0.0356)
H_0(b):						
$\overline{PD}_3(I_3 \cup I_4)$	18.39	14.52	22.28	15.79	19.49	17.92
$PD_2(I_3 \cup I_4)$	19.00	11.25	24.75	16.45	24.25	21.00
Median of differences	-0.40	0.27	-3.10*	-0.06	-2.57**	-0.90
(p-value, one-tailed)	(0.4688)	(0.3190)	(0.0790)	(0.4688)	(0.0356)	(0.1047)
Hypothesis 3.7.(iii):						
H_0(a):						
$PD^*_3(I_1 \cup I_2)$	12.95	11.90	14.48	23.38	22.79	19.78
$PD_1(I_1 \cup I_2)$	20.75	17.50	18.50	19.00	28.75	17.50
Median of differences	-2.33**	-5.80*	-8.09**	-0.32	-3.58	-0.71
(p-value, one-tailed)	(0.0498)	(0.0790)	(0.0299)	(0.4688)	(0.1197)	(0.3769)
H_0(b):						
$PD^*_3(I_3 \cup I_4)$	17.16	12.98	18.49	16.76	18.25	14.45
$PD_1(I_3 \cup I_4)$	19.51	14.25	25.50	17.25	25.00	19.75
Median of differences	-4.68*	0.69	-3.24**	0.29	-5.03	-3.30**
(p-value, one-tailed)	(0.0584)	(0.4377)	(0.0498)	(0.2652)	(0.2164)	(0.0356)
Hypothesis 3.7.(iv):						
H_0(a):						
$\overline{PD}_3(I_1 \cup I_2)$	16.90	14.25	18.28	22.45	23.98	18.08
$\overline{PD}_2(I_1 \cup I_2)$	19.58	17.09	18.44	20.20	26.29	18.25
Median of differences	-2.25	-1.93*	-2.99**	0.61	-1.00	-2.42*
(p-value, one-tailed)	(0.1361)	(0.0912)	(0.0422)	(0.4070)	(0.4377)	(0.0681)
H_0(b):						
$\overline{PD}_3(I_3 \cup I_4)$	18.39	14.52	22.28	15.79	19.49	17.92
$\overline{PD}_2(I_3 \cup I_4)$	20.41	14.88	19.44	18.24	27.57	17.98
Median of differences	-1.90	0.34	-5.46**	-0.61	-1.44	-0.09
(p-value, one-tailed)	(0.3475)	(0.4295)	(0.0299)	(0.3190)	(0.2164)	(0.5000)

In particular it seems as if the reduction is well pronounced when comparing partition-dependence in equilibrium market prices with partition-dependence expressed at the time of the first message (hypothesis 3.7.(iii)), and when comparing the average effect size of partition-dependence in the last sub-period with partition-dependence expressed at the time of the second message (hypothesis 3.7.(ii)). Admittedly, though, the results are by no means consistent. Although most of the medians of differences show a negative sign, statistical significance is marginal in about half the times.

Taken together the results from the information treatment, one may conclude that there is a slight tendency for the partition bias to be reduced over the course of a trading round, particularly after new market data from the differently partitioned market have arrived (most signs of the calculated metrics are negative). The results suggest that traders effectively pay some attention to information received from the other market which, in turn, leads to some convergence of market prices. To a great extent, though, traders seem to adhere to their initially set anchor. Furthermore, it has to be taken into account that the messages did not allow a direct comparison of market prices; participants rather had to "pack" or "unpack" reported prices themselves to make them comparable to prices from their own market. The complexity of messages notwithstanding, the results can be taken as further evidence for the fact that partition-dependence proves quite robust in the short-run experimental markets of the present lab study. Partition-dependence is to some extent diminished, but not eliminated by the "market forces", even under the "tightened" conditions in which participants were (i) reminded of the existence of a different partition and (ii) were informed about obvious price inconsistencies between the two markets.

3.3 Interim conclusions

The lab Study 1 was designed to see whether partition-dependence occurs and persists in short-run experimental asset markets, and to compare effects expressed in probability judgments with effects revealed by market trading prices. Both judgments and prices do show strong effects of partition-dependence across the three event domains that were used. Market prices show a much smaller effect in one of three event domains, and there is a small influence of market experience on post-trading individual judgments. It seems as if there is some convergence of market prices, but market forces are not able to eliminate the bias. The results derived from market prices prove rather reliable since there was vigorous trading among participants and intra-market efficiency can be rated high as temporary arbitrage opportunities *within* markets were quickly removed. The hypothesis that prices are structurally-biased compared to beliefs is refuted because prices are close to the median quantile of measured beliefs.

The *basis treatment* of Study 1 also aimed to analyze the extent to which different competence characteristics of the trading population influences the degree of partition-dependence. Competence of highly knowledgeable people apparently reduces partition-dependence in judged probability for two out of three event domains. A plausible explanation would be that some of the event domains (here: finance and sports) are more information-driven than others (here: weather). However, partition-dependence remains well pronounced even for highly competent subjects and competence effects are hardly reflected in market outcomes of the sports markets to which traders were allocated by their level of self-rated competence. One may argue that differences in competence and knowledge diminish in the market environment, but a trader-based analysis shows that there are significant differences in the trading behavior of differently competent people with respect to the exposure and risk of their portfolios and with respect to their subjectively expected payoffs. Thus, there are significant differences between highly and low competent participants, but these differences hardly transfer to market *prices* in the present lab setting.

In addition, the *information treatment* of Study 1 was designed to examine whether partition-dependence persists under "tightened" conditions in which asset partitions were pointed out more obviously to the participants and in which the subjects received real-time information about current prices from the differently partitioned market. The results of the information treatment suggest that a more salient presentation of the design features may to some extent reduce the bias strength in pre-trading judgments, and the exchange of real-time market data between differently partitioned markets may induce some more convergence of market prices, in particular after new messages have arrived, but statistical significance of the results is modest and partition-dependence proves quite robust, even under the "tightened" conditions of the information treatment. Thus, partition-dependence is to some extent diminished, but is far away from being fully removed by "market forces".

4 Study 2: An NBA/FIFA field experiment

4.1 Experimental design

The very modest effects of trading experience on partition-dependence seen in Study 1, after 20 minutes of trading, suggest the possibility that with much longer trading spans, and perhaps with more knowledgeable traders, partition-dependence could be reduced more strongly or wiped out. Study 2, a field experiment lasting several weeks, was designed to test this hypothesis.

From April to July 2006, an internet-based experimental trading study was conducted. Participants were invited to trade online via the internet in prediction markets for outcomes in the NBA Basketball Playoffs 2005/06 and the FIFA Soccer World Cup 2006.[152] As this study was linked to real sports events, trading markets were open continuously (i.e., 24 hours a day, 7 days a week without interruptions) for approximately nine weeks for the NBA markets (April 20 through June 21, 2006) and approximately 6½ weeks for the FIFA markets (May 24 through July 9, 2006), except for markets that closed when teams were eliminated from the tournament. A number of *N*=317 undergraduate finance students from the University of Muenster (Germany), and *N*=139 students from the CASSEL list at UCLA, Los Angeles (United States) was recruited.[153] Two different channels of recruitment were used in order to analyze second order effects (U.S. students were expected to feel more competent about NBA events whereas German students should feel more competent in the FIFA Soccer World Cup events.

The study was divided into two independent parts: The first part included trading in all-or-nothing contingent claims in experimental prediction markets in which the claims' payoffs depended on the *total number of victories* for a particular NBA team during the Playoffs. For example, a market offered four claims on the San Antonio Spurs, spanning the total number of games the Spurs could win during the Playoffs. One claim would pay a fixed sum of money, 100 cents (€1) if, after the Playoffs, the Spurs

[152] The NBA Basketball Playoffs 2005/06 took place from April 22 through June 21, 2006. The FIFA Soccer World Cup 2006 took place from June 9 through July 9, 2006.

[153] The CASSEL (California Social Science Experimental Laboratory) list contains a pool of students who registered to participate in research studies and experiments. It draws from a large, diverse group of UCLA undergraduates as subjects; http://www.cassel.ucla.edu.

have won 0–3 games (i.e., either zero, one, two, or three games). Another claim would pay 100 cents if their win total is 4–7; a third would pay out if the total is 8–11; and a fourth would pay 100 cents if the total is 12–16. The three claims that do not pay 100 cents expire worthless (note that by construction, there are exactly three claims that expire worthless, whereas a single claim pays 100 cents at the end). The second part of the study included trading in all-or-nothing contingent claims whose payoffs depended on the *total number of goals* scored by a particular national team during the entire World Cup tournament (excluding penalty shoot-out goals). For example, a market offered four claims on Brazil's national team, spanning the total number of goals Brazil could score in the World Cup. One claim would pay a fixed amount of money, 100 cents (€1) if, at the end of the World Cup, Brazil has scored 0–2 goals (i.e., either zero, one, or two goals). Another claim would pay out if, at the end, their goal total is 3–8; a third would pay if the total is 9–11; and a fourth would pay 100 cents if the total is 12 or more.

As in Study 1, the basic idea of these prediction markets is that, after the Playoffs and the World Cup are over, exactly *one* claim (asset) in each market would pay 100 cents, whereas the other *three* claims expire worthless. Again, this is due to the fact that the event intervals represent all possible outcomes of the state space, but do not overlap, i.e. they represent exclusive and collectively exhaustive events.

The main treatment variable is the way in which the state space (number of victories or goals, respectively) is divided into sub-events: For each event domain, there are *two partitions* (treatments) that combine sub-events differently, as shown in Figure 4.1 (Playoffs) and Figure 4.2 (World Cup). For example, in the NBA markets the first partition packs the victory intervals [4, 7] and [8, 11] into a single interval [4, 11], and unpacks the interval [12, 16] into two components of [12, 15] and [16]. More formally, the state space was divided into five disjoint and exhaustive intervals I_0, …, I_4. The first (lowest) interval (I_0) was the same for both partitions ([0, 3] for the Playoffs and [0, 2] for the World Cup). Then, in each of the two partitions two adjoining intervals were packed to form a single asset: In partition 1, intervals I_1 and I_2 were packed to form the interval $I_1 \cup I_2$, whereas the upper two intervals were traded separately (I_3 and I_4). In partition 2, intervals I_3 and I_4 were combined to form $I_3 \cup I_4$, whereas the intervals I_1 and I_2 were traded separately. In the instructions, the participants were explicitly informed what the two different partition sets were (and that they were randomly assigned to only one partition), to control for the concern that offering one partition would convey information to subjects about likelihoods.[154] Every participant was allowed to trade assets based on four different teams—called “team markets”—using the same numerical partitions for each of the four teams. This means, a trader faced the same asset intervals for all of the four teams in each part (NBA, FIFA) of the study. More illustratively, the two partitions used in this study looked as follows:

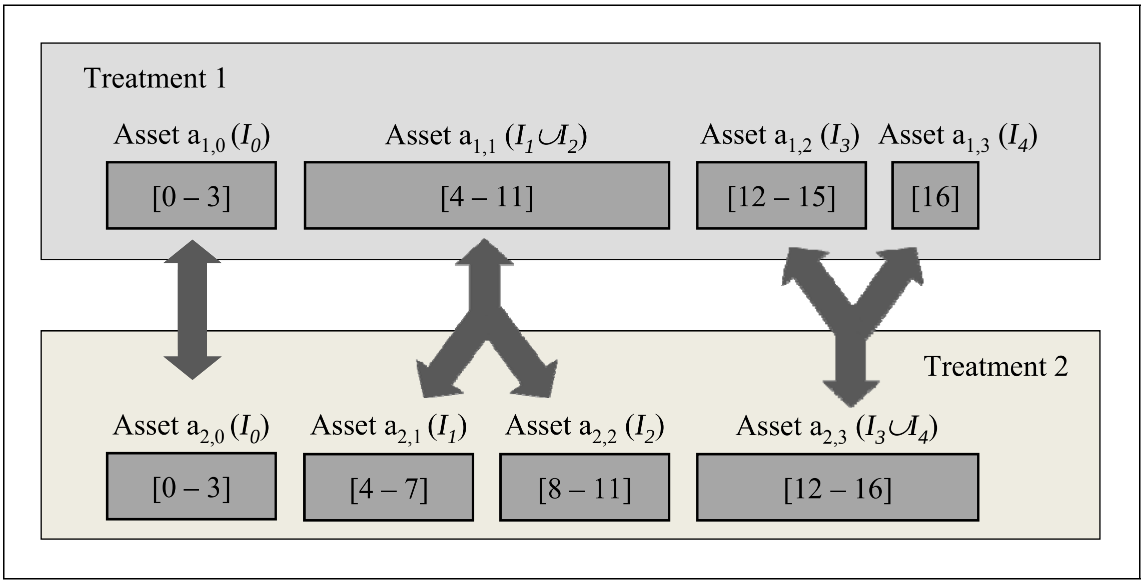


Figure 4.1: Construction of asset partitions (NBA Playoffs victory totals).

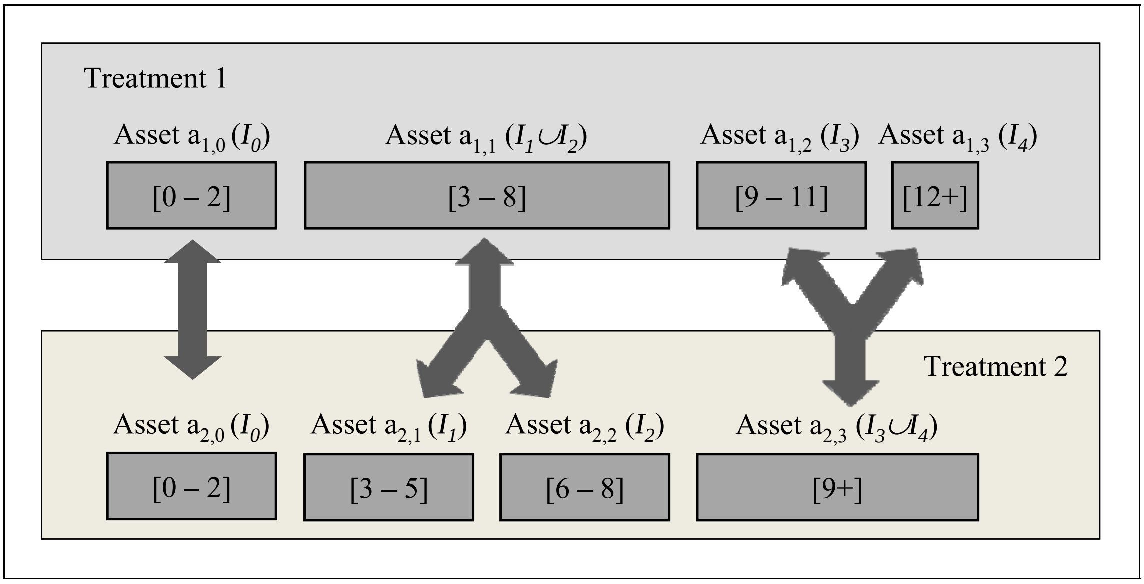


Figure 4.2: Construction of asset partitions (FIFA World Cup goal totals).

[154] Note that the participants did not see Figure 4.1 and Figure 4.2 which show the clear links between packed and unpacked events, but they were informed of the two sets of partition intervals (Appendix V contains the complete instructions of Study 2. Information about the partitions can be found in subsection 3.1.of the instructions.).

Since the NBA Playoffs follow a best-of-seven knock-out bracket mode, the intervals correspond to the number of victories needed by a team to advance across the four playoff rounds, so bets on the various win-total events are equivalent to betting that teams will lose in the first round, the second round, and so forth. Since each fixture is a best-of-seven match, the first team to win four games wins the round and advances. Therefore, betting on the interval [0, 3] is equivalent to betting that the team will leave the Playoffs in the first round, because a team that only wins a total of 0–3 games will be eliminated by an opponent that wins four. The interval [12, 15] is equivalent to winning three rounds (the conference final) but losing in the fourth (and final) round (the Playoffs final). The "interval" [16] is only reached by the NBA champion, who wins four games in all four rounds.[155] The intervals for the number of goals in the FIFA Soccer World Cup were not structured to correspond to advancement across rounds, but were chosen such that they all appeared likely based on the results from the three previous World Cups.

Participants were randomly assigned to one four-team market for NBA Playoffs games, then to another four-team market for FIFA World Cup goals.[156] Groups were reshuffled for the World Cup markets so that students only faced the same traders again by coincidence. Participants from the two different recruitment channels were never assigned to the same market, though. In addition, the students did not know anything about the identity of their counterparts in a market. Each experimental group initially had twenty traders, but some dropped out over time. Each market (each experimental group) consisted of four separate "team markets", so each participant traded in assets based on four different teams in each part of the study. A single "team market" covers the complete state space for a single team. NBA four-team markets included two teams from each of the two Conferences (Eastern and Western).[157] World Cup four-team mar-

[155] Hence, it is known in advance for how many teams a particular claim will pay 100 cents at the end: For eight out of the 16 teams the win total falls into the range of [0, 3] (I_0), four teams end up in the interval I_1, two teams in I_2 (therefore, six teams in $I_1 \cup I_2$), one team ends up in I_3 and one team in I_4 (therefore, two teams in $I_3 \cup I_4$).

[156] There were a number of dropouts during the first part of the study, so that the second part continued with less World Cup prediction markets than Playoffs markets.

[157] Since there are 16 teams constituting the Playoffs bracket at the beginning, this resulted in four different market compositions (market 1: W1, E8, E3, W6; market 2: W2, E7, E4, W5; market 3: W3, E6, E1, W8; market 4: W4, E5, E2, W7 with W=Western Conference and E=Eastern Conference and the digits representing the seed of the team within the Conference).

kets used four official tournament “groups”, out of the eight groups created by FIFA organizers, which were supposed to generate the most interest and trade.[158]

The experimental protocol was similar to Study 1, but was adapted for the Web (see Appendix V). Participants were instructed about the composition of assets and markets (including the partitions of assets they could trade, and the alternative partition), how to use the trading system, payment and incentives, and some details about the NBA Playoffs and the FIFA World Cup, by e-mail. They also had internet access to a homepage (see Figure 4.3) with study details, FAQs,[159] and a practice market in order to allow participants to get acquainted with the trading interface (see Figure 4.4). Participants were kept up to date on the tournaments by e-mail newsletter. The markets were open continuously, so that participants could access the trading platform whenever and wherever they liked.

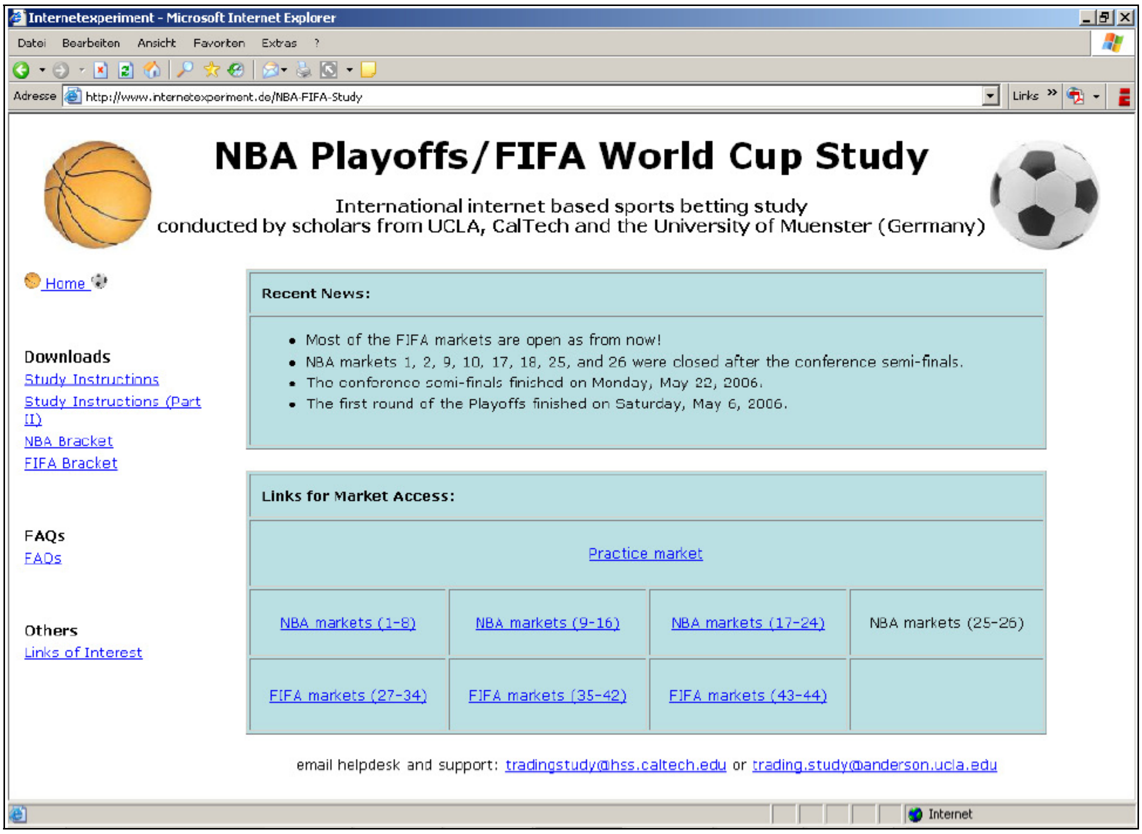


Figure 4.3: Study homepage.

---

[158] This means that teams of a four-team market played each other in the group phase. This guaranteed that exactly two teams per experimental group were active (at least) until the beginning of the knockout phase, since the two top teams from each group advanced to the knockout phase. The groups we used were A (Germany, Costa Rica, Poland, Ecuador), C (Argentina, Côte d’Ivoire, Serbia/Montenegro, Netherlands), E (Italy, Ghana, USA, Czech Republic), and F (Brazil, Croatia, Australia, Japan).

[159] The FAQs can be found in Appendix VI.

As in Study 1, the trading mechanism was a multi-unit continuous double auction (CDA) with a hidden order book, so that subjects could see only the best bid and ask quotes and the most recent trade price for each asset. Traders could submit bid and ask quotes for each asset simultaneously, acting as market makers. Trading took place only among the twenty participants eligible to trade in each market.[160] There was no credit line or short selling opportunity.[161] No explicit transaction costs were imposed for trading.

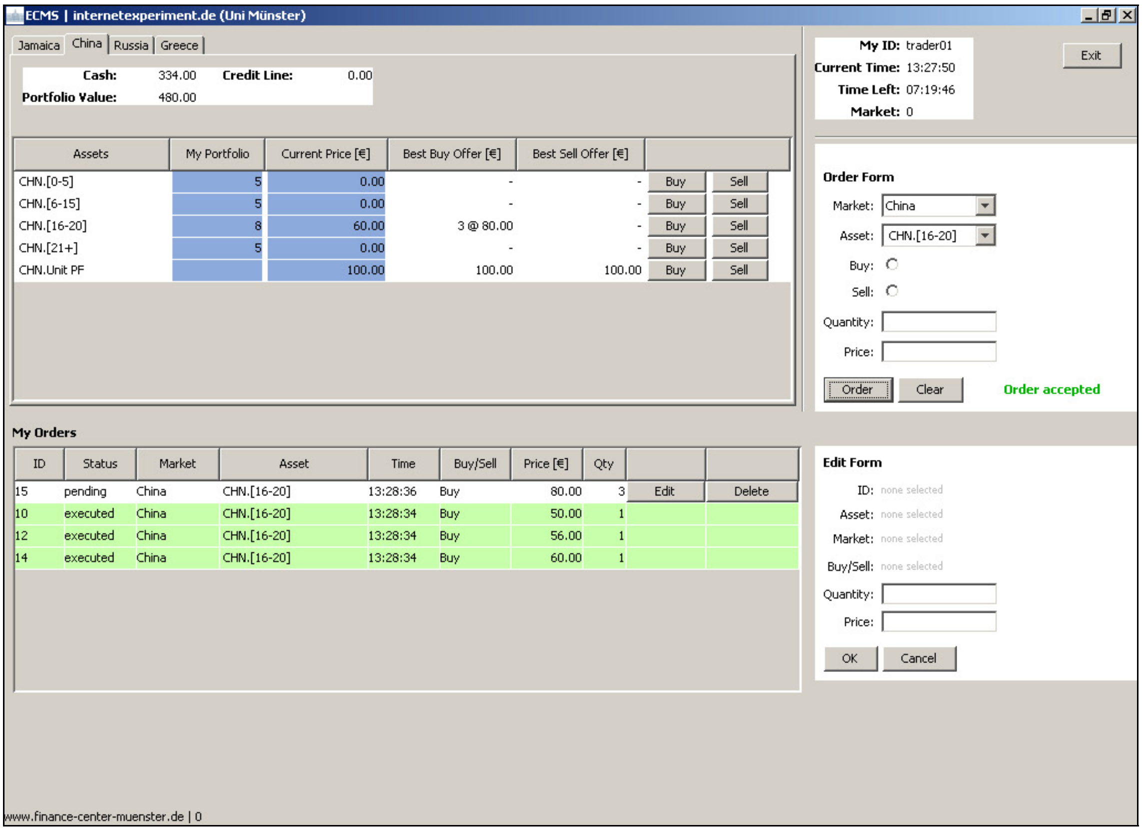


Figure 4.4: Screenshot of the trading application used in Study 2 (practice market).

Participants were initially endowed with different combinations of cash and unit portfolios totaling 1,000 cents (€10) in each "team market" of the NBA Playoffs and were endowed again in the World Cup markets.[162] At the end of the experiment one out

[160] An exception was trading of the unit portfolio which was always executed immediately against the experimenter.

[161] However, the use of the unit portfolio enabled the participants to sell assets short indirectly.

[162] In each market (of twenty participants) always four traders were randomly endowed with one of the five different combinations: 9 unit portfolios + 100 cents, 7/300, 5/500, 3/700, and 1/900, all of them representing an initial value of €10. For each trader the composition of her initial endowment was the same for the four "team markets" in the NBA Playoffs part of the study, but was randomized again for the FIFA World Cup markets.

of the four teams for each experimental group was randomly drawn to compensate the participants incentive-compatibly based on the sum of the actual asset values in their final portfolio (assets being worth either 100 cents or nothing) and their cash balance, for an expected payment of €20 (€10 for Playoffs and €10 for World Cup markets) per person. Questionnaire data was collected from all participants before trading, including individual probability judgments that outcomes would fall into the intervals that corresponded to the assets they traded (only before trading).[163]

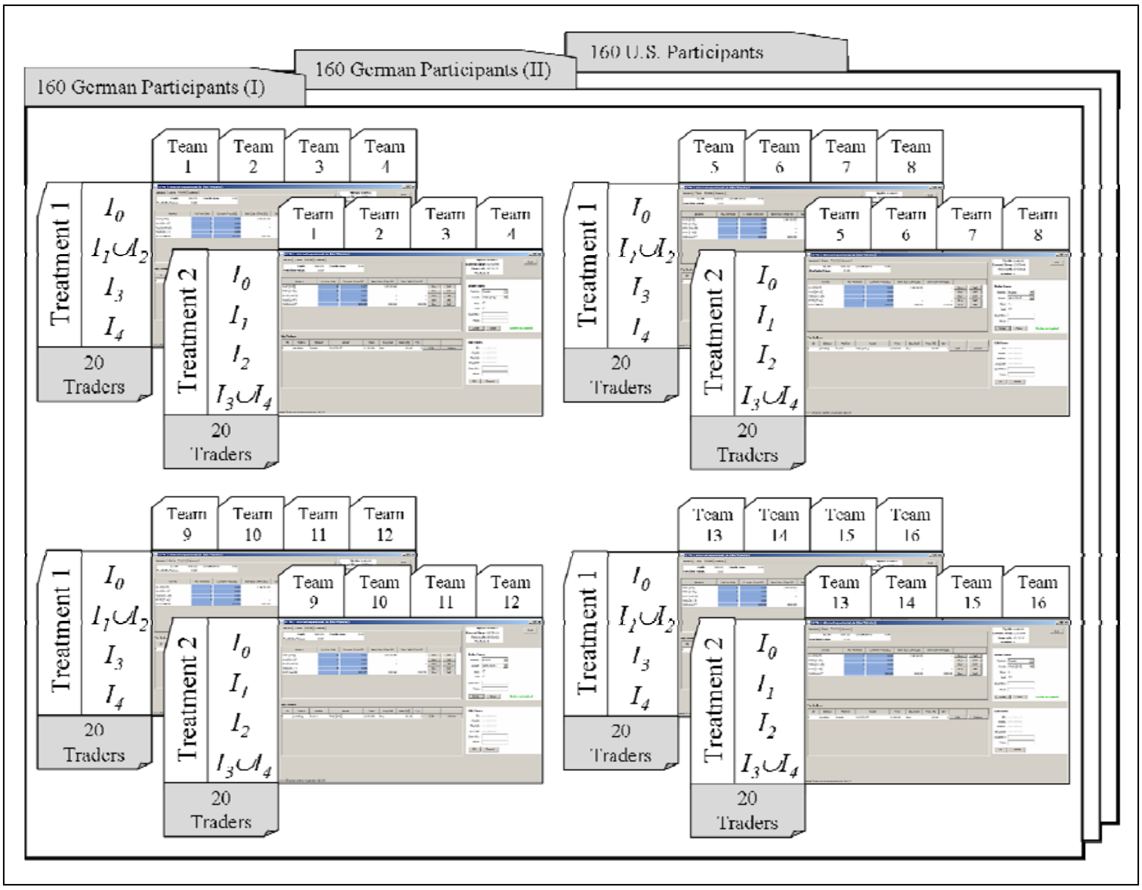


Figure 4.5: Illustration of the experimental setup of Study 2.

It is important to stress that the participants only had direct access to their own four-team NBA and World Cup markets. They could not directly observe market data (like prices or quotes) from other experimental groups trading different partitions.[164] A detailed timetable with all relevant events in the context of this study can be obtained

[163] The questionnaires can be found in Appendix VII.

[164] Of course, it cannot be ruled out the possibility that students were informed about these prices by friends that happen to trade exactly the same teams but the other partition of the state space. However, even in this case, arbitrage opportunities across markets could not jointly be exploited as it was not guaranteed that the same team was chosen for the incentive-compatible payment in both groups.

from the study instructions (see Appendix V, section 2). Figure 4.5 summarizes the experimental setup; in particular it illustrates the composition of the markets. In all cases twenty traders build an experimental group that contains four teams ("team markets") and offers four assets of partition 1 (intervals I_0, $I_1 \cup I_2$, I_3, I_4) or partition 2 (intervals I_0, I_1, I_2, $I_3 \cup I_4$) for each team.

For each part (event domain) of the study—NBA Playoffs and soccer World Cup—assets on 16 teams were traded all in all. Thus trading prices from two different partitions for each of 16 teams in each of the two event domains can be compared. Due to the large number of participants that were recruited, two identical experimental settings ("clones") could be filled with German students and one identical setting could be filled with U.S. students.

4.2 Main results

4.2.1 Judged probabilities

The analysis of results is similar to the analysis from lab Study 1. First, it will be tested for partition-dependence in the individual probability judgments elicited before the beginning of trade. Next, it will be looked for partition-dependence in the bids, asks, and trading prices in the markets. It will also be tested whether the probabilities for the lowest-outcome event ([0, 3] for victories, [0, 2] for goals), which is the same interval in both partitions, happen to differ in the differently-partitioned markets. There is no reason to expect that these probabilities will differ (because the ignorance prior probability is $^1/_4$ for this event in both partitions), but any difference provides a measure of sampling error. This is just a test for whether there are systematically different beliefs in the two markets, and also gives a measure of statistical variability which is useful to assess the size of any partition-dependence effect. Aside from the lowest-outcome event, it is conjectured that subjective probability judgments as well as quotes and trade prices will exhibit partition-dependence even in this large-scale field experiment that is based on real sports events spanning several weeks. As in the lab experiments (Study 1), judged probabilities and market prices are expected to be higher for an event if it is subdivided into two events than if it is traded in aggregation. The notation follows the notation used in Study 1. The main hypotheses are as follows:

Hypothesis 4.1:

H_0(a): $p_{J,i}(I_{0,\, partition\_1}) = p_{J,i}(I_{0,\, partition\_2})$,

H_0(b): $p_{J,i}(I_1) + p_{J,i}(I_2) > p_{J,i}(I_1 \cup I_2)$ and

H_0(c): $p_{J,i}(I_3) + p_{J,i}(I_4) > p_{J,i}(I_3 \cup I_4)$

with $p_{J,i}(I_k)$ = judged probability for interval I_k of team i,
i = {CHI, …, WAS} for NBA Playoffs teams, and
i = {ARG, …, USA} for FIFA World Cup teams, and
$k_{partition\_1} = \{0, 1 \cup 2, 3, 4\}$, $k_{partition\_2} = \{0, 1, 2, 3 \cup 4\}$

Table 4.1: Partition-dependence in before-trading judgments for NBA Playoffs events.

Team	Δ Median, Whole Population (N=302x4), German and U.S. Subjects (Pooled)			
	Event I_0 Equality	$p(I_1) + p(I_2) - p(I_1 \cup I_2)$	$p(I_3) + p(I_4) - p(I_3 \cup I_4)$	N_1/N_2
CHI	15.0	19.0 **	22.5 ***	37/36
CLE	2.0	20.0 ***	15.0 ***	37/36
DAL	0.0	24.5 ***	20.0 ***	40/37
DEN	-2.0	17.5 ***	18.0 ***	41/34
DET	2.5	25.0 ***	36.5 ***	41/34
IND	-10.0	25.0 ***	12.0 ***	41/34
LAC	5.0	5.0	10.0 ***	40/37
LAL	-5.0	22.5 ***	10.0 **	40/37
MEM	15.0	25.0 ***	23.0 ***	37/36
MIA	-10.0	20.0 ***	15.0 ***	40/37
MIL	-5.0	20.0 ***	10.0 **	40/37
NJN	-10.0 *	30.0 ***	20.0 ***	40/37
PHX	0.0	30.0 ***	27.5 ***	37/36
SAC	-12.5	25.0 ***	2.0 *	41/34
SAS	0.0	30.0 ***	40.0 ***	40/37
WAS	-10.0	10.0 ***	10.0	40/37

Notes. The Table presents differences in medians (Δ_{Median}) for interval I_0 and differences in medians for the sum of unpacked events and the packed event per NBA team. N_1 (N_2) indicates the number of participants in partition 1 (partition 2) that provided probability judgments for the team. Each participant (N=302) provided judgments for four different teams (from the four team markets) resulting in a total of 1,208 judgments. *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively (two-tailed) based on a Kruskal-Wallis test for each team.

Because some participants did not submit probability judgments before the first playoff game was played (and their judgments were excluded), there are N=302 (199

German and 103 U.S.) sets of judgments for the NBA teams.[165] For the World Cup, there are N=263 judgment sets submitted by German participants before the opening game was played.[166] Each participant provided ex ante probability judgments for the teams and intervals she would be trading afterwards, providing four probability judgments (four intervals) for four teams (four "team markets" in each experimental group). Judged probabilities summed to 1.0 for each team.

Table 4.2: Partition-dependence in before-trading judgments for FIFA World Cup events.

Team	Δ_{Median}, Whole Population (N=263x4), German Subjects			
	Event I_0 Equality	$p(I_1)+p(I_2)$ $-p(I_1 \cup I_2)$	$p(I_3)+p(I_4)$ $-p(I_3 \cup I_4)$	N_1/N_2
ARG	-1.5	37.5 ***	40.0 ***	30/34
AUS	-5.0	0.0	5.0	33/34
BRA	0.0	22.0 ***	20.0 ***	33/34
CIV	6.5	10.0 *	8.5 ***	30/34
CRC	0.0	15.0 *	9.0 ***	35/32
CRO	10.0 **	10.0 **	20.0 ***	33/34
CZE	8.0 *	35.0 ***	37.5 ***	35/30
ECU	10.0	10.0	9.0 ***	35/32
GER	0.0	35.0 ***	37.5 ***	35/32
GHA	-2.5	12.5 **	5.0 *	35/30
ITA	4.0	25.0 ***	27.5 ***	35/30
JPN	5.0 **	0.0	6.0 ***	33/34
NED	0.0	32.5 ***	30.0 ***	30/34
POL	5.0	30.0 ***	30.0 ***	35/32
SCG	10.0	0.5	14.0 ***	30/34
USA	0.0	15.0 ***	7.0 ***	35/30

Notes. See Table 4.1 notes. Each participant (N=263) provided judgments for four different teams (from the four team markets) resulting in a total of 1,052 judgments.

Table 4.1 and Table 4.2 show the median differences in judgments (×100) for the two different partitions (team by team) and significance by a non-parametric

[165] There were only two days between the NBA bracket was fixed and the first playoff match (CLE-WAS, 2006/04/22, 9.00 P.M.). So N=302 subjects submitted their judgments within the period between 2006/04/21, 12.00 A.M. and 2006/04/22, 9.00 P.M. Judgments submitted afterwards were excluded, since they might be affected by the results of the first games.

[166] In total, N=267 German participants completed the questionnaire. N=4 observations which were submitted after the first game (GER-CRC, 2006/06/09, 6.00 P.M.) were dropped, since they might be affected by the results of the first game, resulting in N=263 remaining datasets. These were completed between 2006/05/24, 12.00 A.M. and 2006/06/09, 6.00 P.M., a period of 17 days. No World Cup results will be reported for U.S. participants since there was an extensive dropout of U.S. students for this part of the study.

Kruskal-Wallis test. Not surprisingly, the differences in the commonly partitioned event between partitions (in the first column) are close to zero and not statistically significant in most of the cases. The differences between the summed probabilities of the unpacked events and the probability for the corresponding packed event (in the second and third column) are positive, almost always highly significant, and are comparable in magnitude to the effects reported earlier (approximately a .20 increase in probability when the interval is unpacked).

4.2.2 Market prices

The next question is whether there is evidence for partition-dependence in observed market prices resulting from trading activity among traders. Remember that several "clone" groups were installed that traded the same events independently. For this purpose, the *most active* team-markets (as measured by the overall number of trades) in each partition for each team were chosen from the market "clones" and matched for further analyses. In the NBA Playoffs, the most liquid markets were the Dallas Mavericks (DAL) and Miami Heat (MIA) (partly because they became the two finalists, so their claims were traded for the longest span of time). For DAL, there were 119 and 129 trades in partitions 1 and 2, respectively. For MIA, there were 102 trades and 101 trades in partitions 1 and 2, respectively.[167]

Figure 4.6 and Figure 4.7 show trade price charts, i.e. the most recent market price (in cents) plotted against the number of days since trading began, for assets corresponding to different partitions. Because there were only about two trades per day across all assets, there are many horizontal flat spots in the time series, which indicate the level of the last trade price when there is no current trading.[168] Gray vertical lines indicate the beginning of a game. The numbers on top of each panel (next to the vertical lines) indicate the cumulated number of victories after each game (if the number is increased by one, the team won the game, if it is unchanged, the game was lost). For example, Figure 4.6 shows that the Dallas Mavericks (DAL) won the first four games, lost the next game, won the sixth game and so on.

[167] Tables showing the overall number of trades (ex unit portfolio trades) for each Playoffs and World Cup team market can be obtained from Appendix VIII.
[168] Note, though, that horizontal flat spots may also hide actual trades at unchanged trade prices.

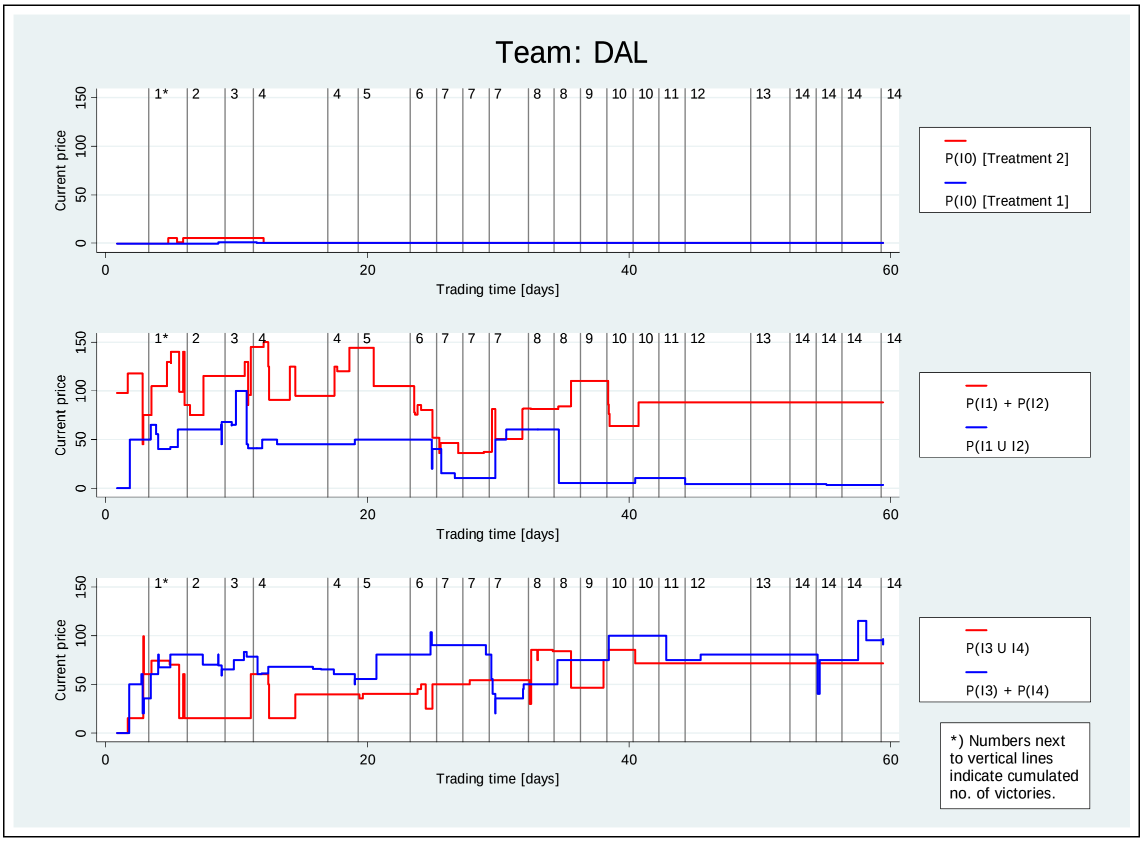


Figure 4.6: Price chart (Dallas Mavericks, DAL).

The upper panel compares prices for the asset intervals I_0 [0, 3] for partition 1 (blue line) and partition 2 (red line). These prices are low, usually zero, since DAL and MIA were expected to win many games, and prices do not differ between the two partitions for each team.[169] In the second panel a blue line indicates the current market price of the packed asset [4, 11] of partition 1 and a red line shows the sum of the market prices for unpacked assets [4, 7] and [8, 11] of partition 2. The fact that the red line lies above the blue line reflects partition-dependence (i.e., the sum of market prices for assets [4, 7] and [8, 11] was usually higher than the market price for asset [4, 11]). The third panel shows a red line for the current market price of packed asset [12, 16] of partition 2 and the blue line represents the sum of the market prices for the unpacked assets [12, 15] and [16] of partition 1. The fact that the blue line is above the red line, for most of the time, indicates partition-dependence, too.

Figure 4.7 shows similar patterns for the two most liquid Miami Heat (MIA) markets. The price charts provide evidence for pronounced partition-dependence in the market prices, as the red curve is above the blue curve for most of the time in the middle

169 Corresponding price charts for all teams can be obtained from Appendix IX.

panel. The effect is also pronounced for the first thirty days in the third panel but diminishes afterwards (for the high-win-total intervals, in fact, there is mild reverse partition-dependence in the last half of the trading span). During the first days of trading prices sometimes summed to more than 100 cents. However, this does not necessarily indicate that there was a lot of potential arbitrage opportunities, since trades for the unpacked assets rarely occurred at the same time.

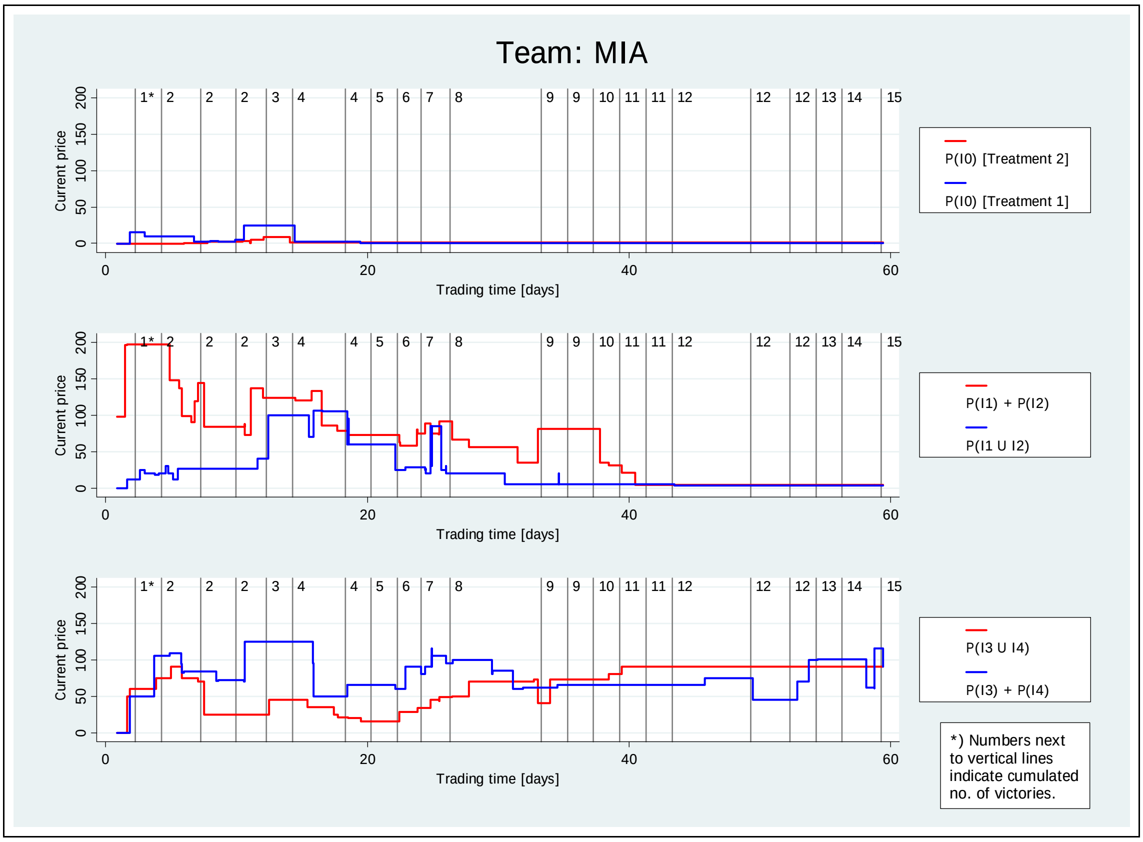


Figure 4.7: Price chart (Miami Heat, MIA).

For the World Cup markets, Figure 4.8 and Figure 4.9 show price charts for Germany (57 and 59 trades in partitions 1 and 2, respectively) and Italy, the eventual World Cup champion (65 trades in both partitions). Note that the numbers on top of each panel (next to the vertical lines) indicate the cumulated number of goals after each game (if the number is increased, the team marked one or more goals in that game, if it is unchanged, no goal was scored). Both markets show persistent partition-dependence of recent prices (although note that trades are rare), i.e., the red line is above blue in the middle panel and vice versa in the lower panel.

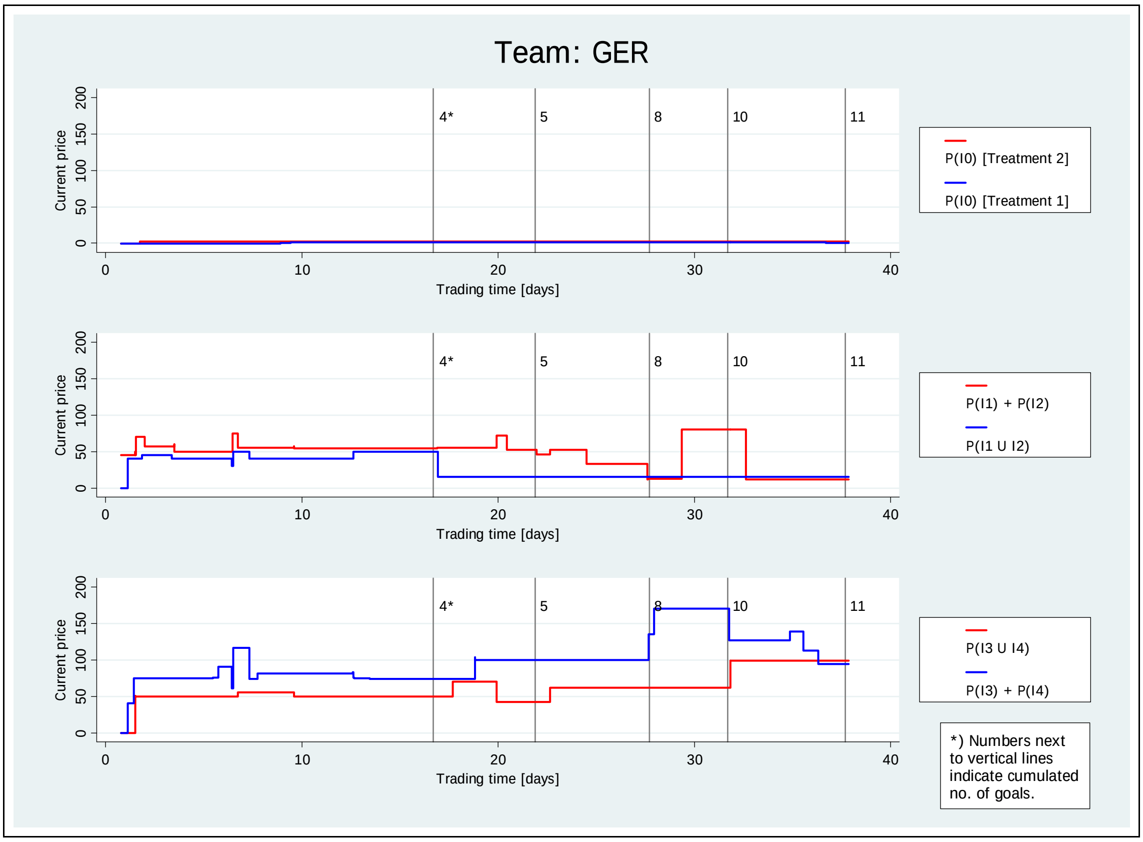


Figure 4.8: Price chart (Germany, GER).

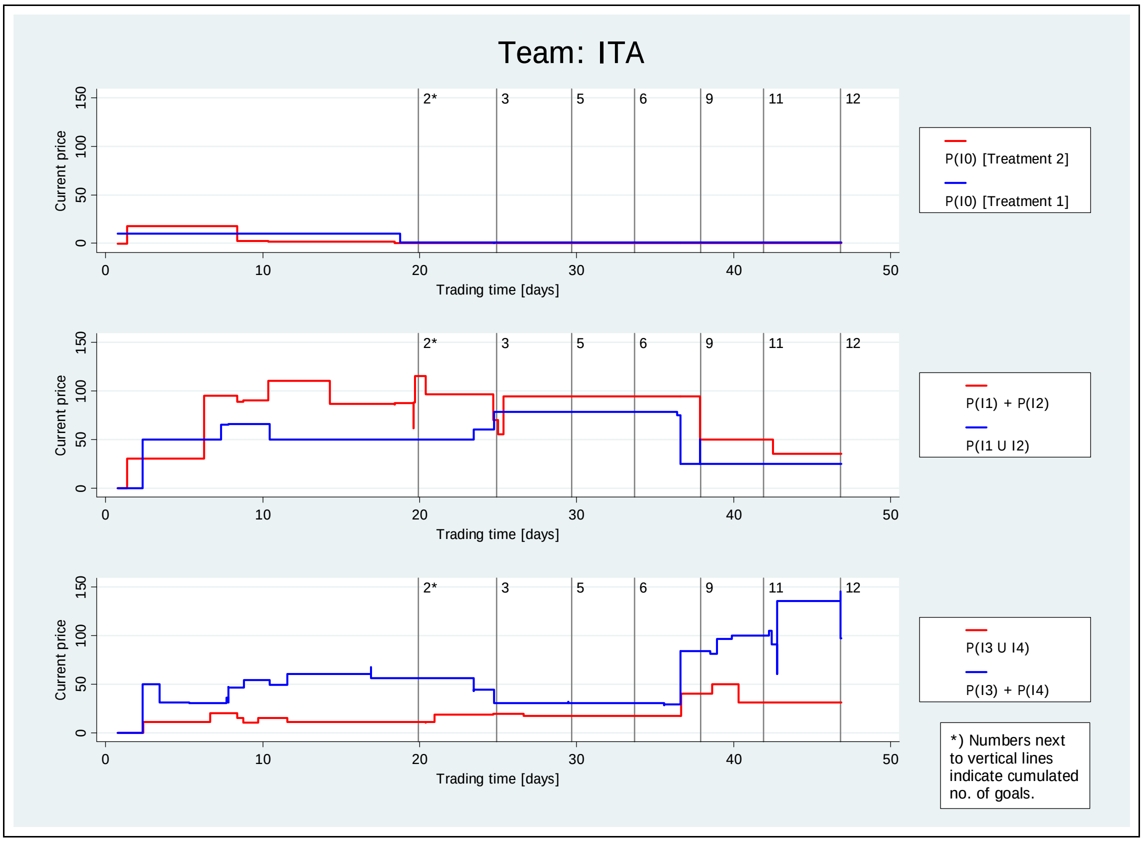


Figure 4.9: Price chart (Italy, ITA).

4.2.3 Hypothetical "pseudo-arbitrage" strategies

4.2.3.1 Preliminaries

Because prices are constantly changing in response to new information over the several weeks of these tournaments, the "equilibrium market prices" for a static event toward the end of trading cannot easily be used to determine the degree of partition-dependence revealed by prices (as in the lab Study 1). Therefore partition-dependence is measured in two more nuanced ways in the next two subsections. Both methods measure the *hypothetical "pseudo-arbitrage"* available by comparing the summed prices for the two unpacked-interval assets (traded in one market) with the price for the equivalent packed-interval asset (traded in a differently partitioned market). These calculations are not true arbitrage opportunities because traders cannot actually trade in markets with different partitions; they simply provide an economically relevant measure of the partition-dependent gap in prices between the two markets.

The first method only looks at available bids and asks and computes whether there is a hypothetical pseudo-arbitrage opportunity across markets for each moment in time ("bid-ask pseudo-arbitrage"). The second method interpolates actual trade prices, assuming one could trade continuously at a price between the last trade price and the next trade price ("interpolated-price hypothetical arbitrage"). Roughly speaking, these two methods provide a lower and upper bound on the possible arbitrage profits from exploiting partition-dependence that would be available to a hypothetical trader with access to multiple markets with different partitions.

4.2.3.2 Bid-ask pseudo-arbitrage

The first test looks at hypothetical cross-market pseudo-arbitrage opportunities using current bid and ask quotes.[170] The method calculates the time-weighted pseudo-arbitrage profit that would result from selling the unpacked-interval assets and buying the equivalent packed-interval asset, at available bid and ask prices. The question is whether there would be arbitrage opportunities if traders could actually trade in both markets, given the available bid and asks which can be used for trade. Remember that participants could only trade in one experimental group at a time and were not informed

[170] For this purpose, the most active markets in each partition and team again were chosen from the experimental "clones" to notionally build a single "slot" (i.e., to match markets with different partitions).

about prices from other markets, so they *could not actually execute these arbitrage trades* (which is the reason why this method is referred to as "pseudo-arbitrage"). Asking how large the arbitrage opportunities are is simply a way to characterize the economic size of partition-dependence in these markets, using all the information on bids and asks. However, because trading is often quite thin, there are long stretches of time when bids and asks are not available on all assets and the measured partition-dependence is zero.

Consider intervals I_1 and I_2 that trade separately (unpacked) in partition 2 and fused (packed) in partition 1. If there is partition-dependence, then the bids for assets underlying events I_1 and I_2 will be high (compared to bids for the packed asset $I_1 \cup I_2$). So one kind of pseudo-arbitrage is to take the sum of the current bids for assets I_1 and I_2 (i.e., the prices at which one could sell those assets) and to subtract the current ask for the equivalent asset $I_1 \cup I_2$ (i.e., the price at which one could buy that asset). If this difference is positive, then a trader with access to both markets could sell the two unpacked assets of intervals I_1 and I_2 for more than she could buy the packed interval asset $I_1 \cup I_2$. If there is reverse partition-dependence, then the opposite strategy would be profitable (i.e., buying the components I_1 and I_2 (at their ask prices) and selling the packed asset $I_1 \cup I_2$ (at its bid quote)). The size of these arbitrage strategies is represented by the following notation:

$$X_{(12)} := B_{i,t}(I_1) + B_{i,t}(I_2) - A_{i,t}(I_1 \cup I_2) \qquad \text{vs.}$$
$$Y_{(12)} := B_{i,t}(I_1 \cup I_2) - [A_{i,t}(I_1) + A_{i,t}(I_2)]$$

with $B_{i,t}(I_k)$ = Best (highest) bid quote for interval I_k of team i at time t
$A_{i,t}(I_k)$ = Best (lowest) ask quote for interval I_k of team i at time t

The value of position $X_{(12)}$ represents a strategy of selling unpacked assets I_1 and I_2 at their bid quotes in one partition and buying the packed asset $I_1 \cup I_2$ at its ask quote in the other partition at each point of time. Position $Y_{(12)}$, in turn, refers to a reverse partition-dependence strategy and is based on selling the packed asset $I_1 \cup I_2$ in one partition and simultaneously buying unpacked assets I_1 and I_2 in the other partition. Arbitrage opportunities exist only if these positions are positive. Analogously, positions $X_{(34)}$ and $Y_{(34)}$ can be calculated for assets I_3 and I_4 against asset $I_3 \cup I_4$:

$$X_{(34)} := B_{i,t}(I_3) + B_{i,t}(I_4) - A_{i,t}(I_3 \cup I_4) \qquad \text{vs.}$$
$$Y_{(34)} := B_{i,t}(I_3 \cup I_4) - [A_{i,t}(I_3) + A_{i,t}(I_4)]$$

with $B_{i,t}(I_k)$ = Best (highest) bid quote for interval I_k of team i at time t
$A_{i,t}(I_k)$ = Best (lowest) ask quote for interval I_k of team i at time t

Figure 4.10 shows these statistics over the life of the experiment for NBA team Dallas Mavericks (DAL). The top panel shows $B_{i,t}(I_1) + B_{i,t}(I_2) - A_{i,t}(I_1 \cup I_2)$ (in blue) and $B_{i,t}(I_1 \cup I_2) - [A_{i,t}(I_1) + A_{i,t}(I_2)]$ (in red). The second panel shows the maxima of each of these spreads and zero (i.e., it only shows their values when they are positive, when pseudo-arbitrage is profitable). The blue spikes in the second panel indicate that there are pseudo-arbitrage opportunities, which are sometimes quite large in magnitude but are sporadic and usually short-lived. The red line at zero (second panel) indicates that there is never a set of available bids and asks consistent with profitable arbitrage against reverse partition-dependence. The horizontal lines at the bottom of the second panel indicate the spans of time during which *any* bid or ask exists in the market for each of the assets in the arbitrage strategy. When those lines are interrupted there is no liquidity and hence no opportunity for arbitrage.[171] Further note that a gray vertical line in the chart shows the point of time at which the displayed event became impossible (zero probability). For instance, when the Dallas Mavericks had won their eighth game, interval I_1 ([4, 7]) became impossible. Hence, the charts are truncated at that point, since assets for intervals $I_1 \cup I_2$ and I_2 exactly represent the same claim as from then, and therefore position $X_{(12)}$ [$= B_{i,t}(I_1) + B_{i,t}(I_2) - A_{i,t}(I_1 \cup I_2) \rightarrow B_{i,t}(I_2) - A_{i,t}(I_2)$] and position $Y_{(12)}$ [$= B_{i,t}(I_1 \cup I_2) - [A_{i,t}(I_1) + A_{i,t}(I_2)] \rightarrow B_{i,t}(I_2) - A_{i,t}(I_2)$] both turn into the negative bid-ask spread for interval I_2 (which is negative by definition). The third and fourth panels show the same time series for the pseudo-arbitrage of intervals I_3 and I_4 against the interval $I_3 \cup I_4$. There are frequent interruptions in the bid-ask existence series (at the bottom of the fourth panel), so pseudo-arbitrage opportunities are rare.

[171] For technical reasons, missing ask quotes were set $+\infty$ and missing bid quotes were set 0. For economical reasons, bid and ask quotes (or the sum of two bid or ask quotes) were capped at 100 cents. Hence, positions $X_{(12)}$ and $Y_{(12)}$ are actually calculated as $X_{(12)} := \min\{B_{i,t}(I_1) + B_{i,t}(I_2), 100\} - \min\{A_{i,t}(I_1 \cup I_2), 100\}$ and $Y_{(12)} := \min\{B_{i,t}(I_1 \cup I_2), 100\} - \min\{A_{i,t}(I_1) + A_{i,t}(I_2), 100\}$. Note that arbitrage opportunities from position $Y_{(12)}$ can only occur if all three quotes are available. By contrast, there may be arbitrage opportunities from position $X_{(12)}$ even if one of the quotes $B_{i,t}(I_1)$ or $B_{i,t}(I_2)$ misses, since it might be possible, for instance, to sell $B_{i,t}(I_1)$ for more than the cost of $A_{i,t}(I_1 \cup I_2)$. The same considerations also apply to positions $X_{(34)}$ and $Y_{(34)}$, respectively.

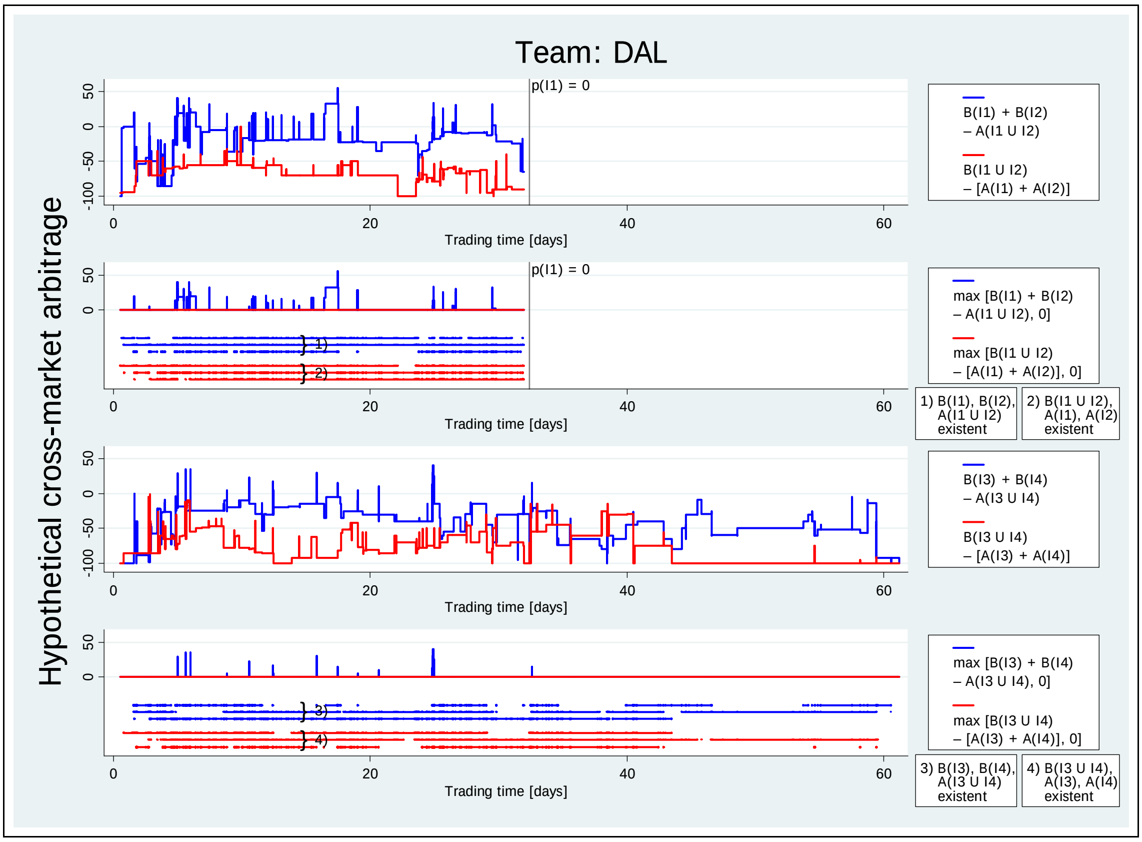


Figure 4.10: Hypothetical cross-market pseudo-arbitrage (Dallas Mavericks, DAL).

Figure 4.11 shows the same time series for the World Cup winning team Italy (ITA). There are few pseudo-arbitrage opportunities for the low-goal intervals I_1 and I_2, but quite a bit of pseudo-arbitrage for intervals I_3 and I_4. From days 9 through 25, there is a persistent gap in the bids of unpacked events I_3 and I_4 and the ask for event $I_3 \cup I_4$. These examples illustrate the advantage of using the continuous bid and ask information. Trades are rather rare for Italy events (only about one trade per day across all assets) but bids and asks are common enough to show persistent gaps in (potential) prices. Again, there is no evidence for profitable arbitrage strategies against reverse partition-dependence, which adds support to the conjecture that partition-dependence in these markets is rather systematic than just a random error.

Figure 4.11: Hypothetical cross-market pseudo-arbitrage (Italy, ITA).

Table 4.3 reports the value of the time-weighted pseudo-arbitrage statistics for all teams. These are the area under the blue and red curves in the second and fourth panels of Figure 4.10 and Figure 4.11, divided by the total trading time.[172] The profitability of strategies exploiting partition-dependence (in columns 2 and 4) is often very low, but is above 1.0 for 9 of 32 teams. Furthermore, pseudo-arbitrage against reverse partition-dependence is much less profitable. For 38 of the 46 team-partition comparisons, arbitraging against partition-dependence is more profitable than arbitraging against reverse partition-dependence (excluding 18 team–partition cases in which both figures are zero), a fraction significantly lopsided by a conservative sign test ($z=5.88$, $p<.001$).

[172] Dividing by the total trading time is to make these statistics comparable across teams. Note that the relevant trading time ends either when the last auction for assets I_1, I_2 or $I_1 \cup I_2$ (or I_3, I_4 or $I_3 \cup I_4$, respectively) occurred or when the corresponding interval asset I_1 (or I_3) expired worthless.

Table 4.3: Per-day profitability of hypothetical bid/ask pseudo-arbitrage strategies.

Team	Low Intervals: Arbitrage PD (Sell I_1, I_2, Buy $I_1 \cup I_2$)	Low Intervals: Arbitrage Reverse PD (Buy I_1, I_2, Sell $I_1 \cup I_2$)	High Intervals: Arbitrage PD (Sell I_3, I_4, Buy $I_3 \cup I_4$)	High Intervals: Arbitrage Reverse (Buy I_3, I_4, Sell $I_3 \cup I_4$)
NBA Playoffs teams				
CHI	0.00	0.04	0.06	0.00
CLE	0.01	0.00	0.11	0.00
DAL	1.95	0.00	0.03	0.00
DEN	0.00	1.08	0.00	0.00
DET	1.36	0.26	0.00	5.06
IND	0.12	0.00	0.20	0.00
LAC	0.16	0.00	0.71	0.00
LAL	0.00	0.00	5.14	0.00
MEM	0.00	0.00	2.61	0.00
MIA	0.96	0.00	0.75	0.00
MIL	0.00	0.04	0.00	0.00
NJN	0.65	0.00	7.90	0.00
PHX	0.25	0.00	1.46	0.00
SAC	0.00	0.00	0.00	0.00
SAS	0.44	0.00	0.24	0.00
WAS	0.05	0.03	0.00	0.00
FIFA World Cup teams				
ARG	0.05	0.00	0.15	0.00
AUS	0.55	0.01	0.48	0.00
BRA	0.00	0.11	0.07	0.00
CIV	0.00	0.00	0.00	0.00
CRC	0.00	0.06	0.00	0.00
CRO	0.03	0.00	0.00	0.00
CZE	0.64	0.00	9.93	0.00
ECU	0.00	0.00	0.00	0.00
GER	0.22	0.00	0.79	0.00
GHA	0.00	0.04	0.00	0.00
ITA	0.35	0.00	8.36	0.00
JPN	3.05	0.00	0.00	0.00
NED	0.29	0.00	0.00	0.00
POL	0.00	0.02	0.23	0.00
SCG	0.01	0.00	0.11	0.00
USA	0.00	0.00	0.00	0.00

4.2.3.3 Interpolated-price hypothetical arbitrage

The bid-ask measures described in subsection 4.2.3.2 above are very conservative because there are substantial periods of time when there are no bids and asks for some of the assets. Since bids and asks are limit orders—any other trader can immediately sell to a posted bid, or buy at a posted ask—traders may be conservative in posting these bids and asks. An active trader posting new bids or asks might be able to initiate trades from subjects who are willing to trade but have not currently posted bids or asks.

One simple way to measure the potential of a more pro-active strategy is to assume that some traders are willing to trade even if they have not posted current bids or asks. Computing the gains from this strategy requires some assumption about the prices at which trades could take place, when there are no bids and asks. A conservative simplifying assumption is that the path of implicit trade prices is smoothly monotonic: That is, if a trade occurs at price P at one point in time t and the next trade occurs at price $P–d$, at a later time $t+n$, then it is assumed that trades could take place in the price interval $[P–d, P]$ continuously between times t and $t+n$ if an agent tried to initiate trade. Implementing this assumption conservatively, one way to bound the possible trades is to assume that somebody could *buy* an asset, at any point in time t, at the *maximum* of the *last* observed trade price and the *next* observed trade price. Similarly, it is assumed that a trader could *sell* an asset at the *minimum* of the last observed price and the next observed price.[173] These assumptions imply that there is unobserved willingness to trade that is not manifested in posted bids and asks, but that the unobserved trade prices are always bounded by the worst prices at which the last and next (unforeseen) trades take place. For example, suppose the trade prices of a thinly-traded asset are 42 at day 20 and 48 at day 25, and there are no trades between those dates. If you are *buying* the asset, it is assumed that you could buy it at the *higher* price of 48 during days 20 to 25 even though there is no trading during those days (and even if there are no posted bids or asks). If you are *selling* the asset, it is assumed that you could sell it for the *lower* price of 42 during days 20 to 25.

The described measure can be termed "interpolated-price hypothetical arbitrage" because the past and future prices are used to *interpolate* a trade price continuously. The arbitrage is *hypothetical* because it just summarizes price differences in separate markets and assumes that trades can take place when there are no standing bids or asks. Because they cannot trade across markets, participants cannot directly act on these pseudo-arbitrage opportunities. This on-paper trading strategy is conservative in the following sense: Suppose there is no trade between time t and the time $t+1$ when a team loses, but at time $t+1$ there is a trade in which the team's asset price plunges because of

[173] Note that the "previous" trade price is assumed to be the first trade price before the first auction occurred, and the "next" trade price to be the latest trade price after the last auction in the market took place.

their loss. Then the assumption is that the trading price available in the time interval [t, t+1] is the post-plunge price. In effect, the strategy assumes that if there is new information which reduces the next executed trade price, the information is capitalized in *all* prices between the last and next prices. Note that basketball games and soccer matches are occurring during the continuous flow of trading, so using the worst of the last and next prices often means that traders are (hypothetically) betting against unfavorable public information, which adds to the conservatism of this measure. This measure is also related to simply taking the area between the red and blue lines (second and third panels) in Figure 4.6 to Figure 4.9. Suppose those curves were altered so that whenever there is a price decrease in an unpacked event, from the last price P at time t to a new price $P–d$ and time $t+n$, the price line between t and $t+n$ is lowered to $P–d$. Similarly, when there is a price increase for the packed-interval asset, the price line is retroactively shifted upward to the new, higher price. Redrawing in this way in the second panel, for example, would reduce the red line (because the unpacked assets are always being sold at weakly worse prices), and increase the blue line. The gap between those redrawn lines (truncated at zero) is the same as the interpolated-price hypothetical arbitrage measure.

In formal notation, the interpolated-price hypothetical arbitrage profit $X_{(12)}$ (and $X_{(34)}$) for the intervals I_1 and I_2 (I_3 and I_4, respectively) at time t is:

$$X_{(12)} := \min\{P_{i,t-r}(I_1), P_{i,t+n}(I_1)\} + \min\{P_{i,t-r}(I_2), P_{i,t+n}(I_2)\} \\ - \max\{P_{i,t-r}(I_1 \cup I_2), P_{i,t+n}(I_1 \cup I_2)\}$$

$$X_{(34)} := \min\{P_{i,t-r}(I_3), P_{i,t+n}(I_3)\} + \min\{P_{i,t-r}(I_4), P_{i,t+n}(I_4)\} \\ - \max\{P_{i,t-r}(I_3 \cup I_4), P_{i,t+n}(I_3 \cup I_4)\}$$

where $P_{i,s}(I_k)$ is the trade price at time s for interval k in event domain i, and $t–r$ and $t+n$ are the times of the most recent and next trades.

Similarly, it can be tested for hypothetical profits $Y_{(12)}$ (and $Y_{(34)}$) from the corresponding reverse arbitrage strategy that is calculated as follows:[174]

[174] However, for economical reasons the following is assumed: for strategy $X_{(12)}$, $\min\{P_{i,t-r}(I_1), P_{i,t+n}(I_1)\} + \min\{P_{i,t-r}(I_2), P_{i,t+n}(I_2)\}$ and $\max\{P_{i,t-r}(I_1 \cup I_2), P_{i,t+n}(I_1 \cup I_2)\}$ each are capped at 100 cents, and for strategy $Y_{(12)}$, $\min\{P_{i,t-r}(I_1 \cup I_2), P_{i,t+n}(I_1 \cup I_2)\}$ and $\max\{P_{i,t-r}(I_1), P_{i,t+n}(I_1)\} + \max\{P_{i,t-r}(I_2), P_{i,t+n}(I_2)\}$ each are also capped at 100 cents. The same considerations also apply to positions $X_{(34)}$ and $Y_{(34)}$, respectively.

$$Y_{(12)} := \min\{P_{i,t\text{-}r}(I_1 \cup I_2), P_{i,t+n}(I_1 \cup I_2)\}$$
$$- \max\{P_{i,t\text{-}r}(I_1), P_{i,t+n}(I_1)\} - \max\{P_{i,t\text{-}r}(I_2), P_{i,t+n}(I_2)\}$$

$$Y_{(34)} := \min\{P_{i,t\text{-}r}(I_3 \cup I_4), P_{i,t+n}(I_3 \cup I_4)\}$$
$$- \max\{P_{i,t\text{-}r}(I_3), P_{i,t+n}(I_3)\} - \max\{P_{i,t\text{-}r}(I_4), P_{i,t+n}(I_4)\}$$

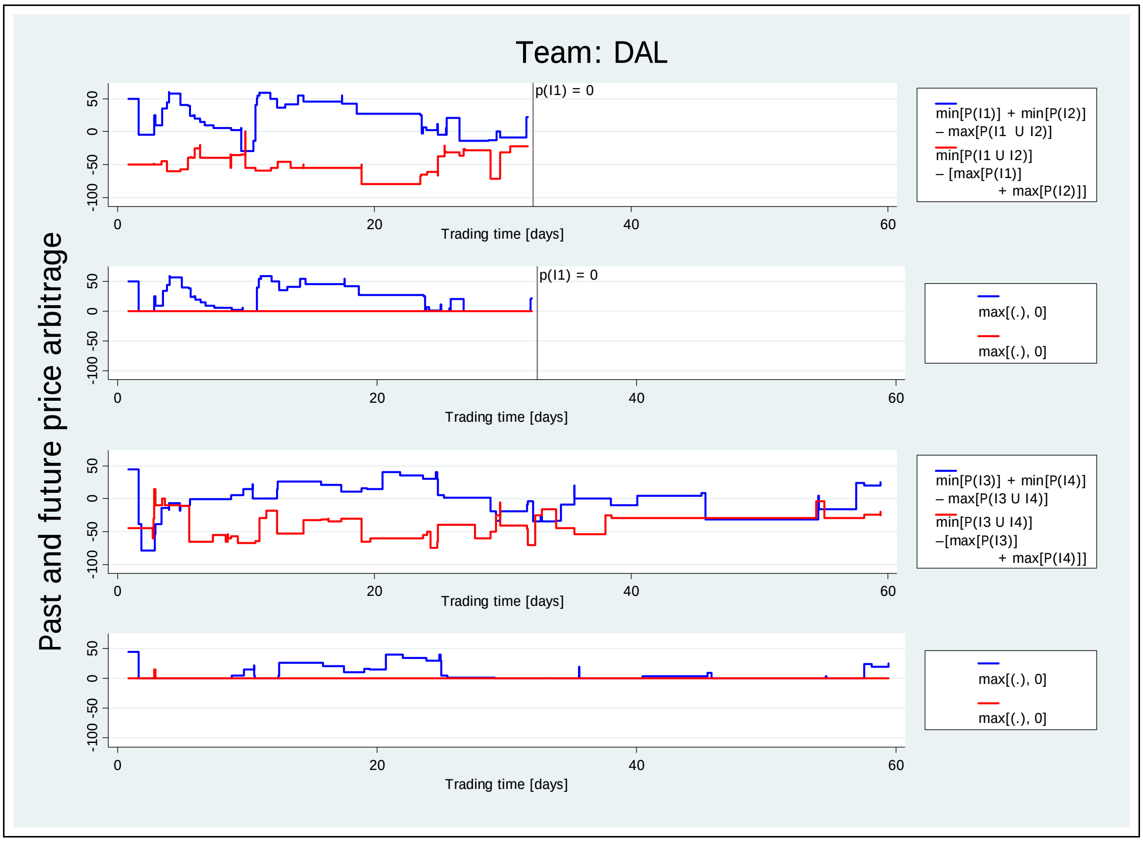


Figure 4.12: Interpolated-price hypothetical arbitrage (Dallas Mavericks, DAL).

Figure 4.12 shows the interpolated-price hypothetical arbitrage profit over time for the Dallas Mavericks (DAL) in the NBA event domain. The blue line in the first panel shows the hypothetical arbitrage profits from selling at the minimum interpolated prices for unpacked intervals I_1 and I_2, and buying at the maximum interpolated price for packed interval $I_1 \cup I_2$ ($X_{(12)}$). The red line, by contrast, shows the hypothetical profits from the reverse arbitrage strategy, i.e., selling at the minimum interpolated price for the packed interval $I_1 \cup I_2$, and buying at the maximum interpolated prices for unpacked intervals I_1 and I_2 ($Y_{(12)}$). Because this "profit" can be positive or negative, the second panel shows the value of this hypothetical profit when it is above zero (i.e., the profit conditional on it being positive). Panels three and four show the same calculations for the assets based on unpacked intervals I_3 and I_4 and the packed interval $I_3 \cup I_4$ ($X_{(34)}$ and

$Y_{(34)}$). The panel two and four hypothetical profits from selling the unpacked-interval assets and buying the packed-interval asset (blue lines) are often positive and large in magnitude.

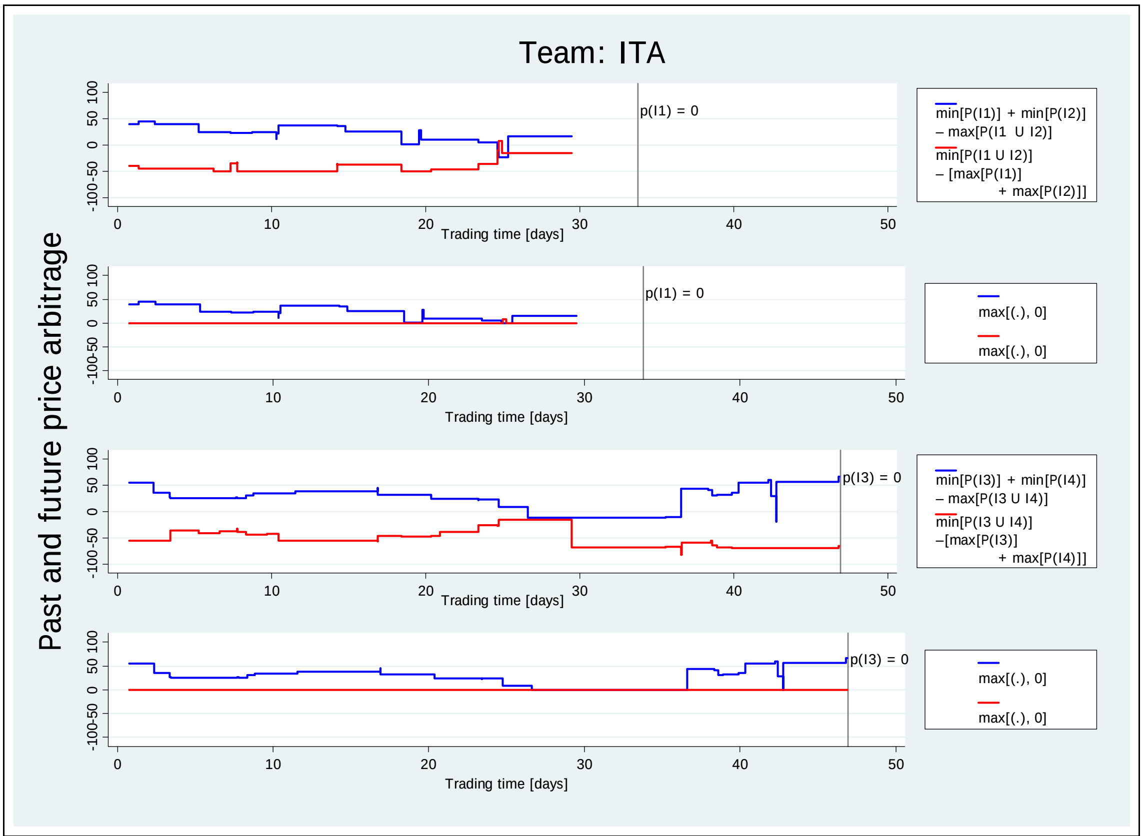


Figure 4.13: Interpolated-price hypothetical arbitrage (Italy, ITA).

Figure 4.13 shows the corresponding data from trades on Italy (ITA) in the World Cup event domain. The results are similar. Note that if there were reverse partition-dependence (the packed-interval asset price is higher) the red lines in Figure 4.12 and Figure 4.13 would be above zero, but this is never the case. The fact that there is virtually no reverse effect proves partition-dependence in the expected direction (as indicated by the blue curves) to be systematically positive and not merely the result of random error.

A way to measure the daily average interpolated-price hypothetical pseudo-arbitrage profit for each team, is to calculate the area under the blue and red curves in the second and fourth panels of Figure 4.12 and Figure 4.13, and divide it by the total trading time (in days).[175] These statistics are provided for each team and interval in Table 4.4. Note the figures are always positive because if the return to pseudo-arbitrage is negative, it is assumed the trade would not be made (i.e., only positive profits are averaged). The average per-day hypothetical profit from exploiting partition-dependence (selling the unpacked-interval assets and buying the packed-interval asset) is higher than for the reverse strategy (buying unpacked and selling packed) for 21 out of 32 teams for intervals I_1 and I_2, and for 26 out of 32 for intervals I_3 and I_4 (significant by sign test at $p<0.1$ and $p<.001$, respectively). The median per-day pseudo-arbitrage profit exploiting partition-dependence, across the 32 teams from both sports, is 5.66 for intervals I_1 and I_2 and 5.89 for intervals I_3 and I_4; the average of this median across intervals is 5.77.

The interpolated-price hypothetical arbitrage profits shown by the time series in this subsection (Figure 4.12 and Figure 4.13, Table 4.4) are larger and more persistent than the sporadic short-lived pseudo-arbitrage profits based on submitted bids and asks shown in subsection 4.2.3.2 (Figure 4.10 and Figure 4.11, Table 4.3). The hypothetical profits from these two measures could be treated as a lower and upper bound on the financial magnitude of partition-dependence. Profitability as measured using simultaneously-available bids and asks (subsection 4.2.3.2) provides a lower bound because there are so many stretches of time with incomplete bids and asks. Profitability as measured by the interpolated-price method (subsection 4.2.3.3) artificially liquefies the market by essentially assuming there is always a latent trade waiting to occur at the right price, so this method provides an upper bound (though it is still conservative because it assumes trades would be executed at the worst of the most recent and next future prices).

[175] Note that the relevant trading time ends either when the last auction for an asset of the relevant interval occurred or when the corresponding interval asset I_1 (or I_3) expired worthless.

Table 4.4: Per-day profitability of hypothetical interpolated-price pseudo-arbitrage strategies.

Team	Low Intervals		High Intervals	
	Arbitrage PD (Sell I_1, I_2, Buy $I_1 \cup I_2$)	Arbitrage Reverse PD (Buy I_1, I_2, Sell $I_1 \cup I_2$)	Arbitrage PD (Sell I_3, I_4, Buy $I_3 \cup I_4$)	Arbitrage Reverse (Buy I_3, I_4, Sell $I_3 \cup I_4$)
NBA Playoffs teams				
CHI	1.24	0.85	2.04	0.00
CLE	3.71	0.72	13.48	0.00
DAL	22.39	0.00	7.38	0.02
DEN	3.49	5.48	2.43	1.54
DET	16.10	2.70	0.48	7.50
IND	7.98	0.19	0.00	0.00
LAC	9.41	0.00	7.76	0.00
LAL	8.56	0.00	8.29	0.00
MEM	0.16	2.14	11.51	0.00
MIA	27.51	0.00	13.52	2.03
MIL	0.00	2.97	2.75	0.00
NJN	5.87	1.69	24.03	0.00
PHX	14.53	0.23	8.17	0.00
SAC	0.38	0.90	0.15	0.58
SAS	28.59	0.00	16.07	0.33
WAS	8.27	0.00	5.23	0.00
FIFA World Cup teams				
ARG	0.53	2.55	13.05	0.00
AUS	2.33	1.07	7.24	0.36
BRA	0.00	5.63	0.57	3.50
CIV	0.04	4.53	0.08	0.72
CRC	1.01	6.30	0.01	0.24
CRO	0.66	2.24	1.22	0.00
CZE	21.85	0.00	29.21	0.00
ECU	12.24	0.05	9.74	0.01
GER	11.87	0.04	9.35	0.20
GHA	0.00	23.55	1.98	0.00
ITA	22.84	0.06	27.66	0.00
JPN	8.11	0.51	1.68	0.00
NED	10.73	0.00	6.54	2.14
POL	0.71	0.19	0.81	0.46
SCG	0.60	0.20	0.25	0.00
USA	5.45	8.08	1.35	0.00

4.3 Second order results

As was mentioned in the description of the experimental protocol (subsection 4.1), two different channels of recruitment were used in order to analyze second order effects: U.S. students (recruited from the CASSEL list at UCLA, Los Angeles (United

States)) were expected to be more competent about NBA events, whereas German students (recruited from an undergraduate finance class at the University of Muenster (Germany)) were expected to be more competent about the FIFA Soccer World Cup events. However, because U.S. participation in the NBA Playoffs markets and the World Cup markets was low, there is little statistical power to detect such effects so they will not be discussed further.[176] The following analyses thus focus on second order effects in subjective probability judgments that were provided by the participants before each part of the study began.

In the run-up to each part of the study, participants were asked to complete a short questionnaire on the Internet.[177] Basically, the questionnaires included some self-rating questions that can be used as additional competence proxies (competence, interest, intention to follow the events on TV, etc.), a brief trivia quiz, and the elicitation of subjective probability judgments (4 events per teams × 4 teams) for the intervals they traded. Additionally, in the first questionnaire (before NBA Playoffs) participants were asked four general questions about attributes that may have an impact on how the subjects behave and derive their decisions during this trading study. These questions may help to decide whether fundamental differences between differently recruited subjects (U.S. vs. German students) exist. The general questions were supposed to be uncorrelated with competence and include the year of birth, self-ratings on general trading experience (e.g., in the stock or bond market) and knowledge in the field of statistics, as well as the question whether a participant has ever been active in sports betting. The differently recruited subsamples (N_{German}=317; $N_{U.S.}$=139) turn out to be fairly balanced with respect to these questions. The average year of birth is 1982 for German students (σ=2.0) and 1984 for U.S. students (σ=3.4). General trading experience (scale 1–7) was moderate in both subsamples (μ=2.5, σ=1.5 for Germans and μ=2.8, σ=1.5 for U.S. students). Knowledge and skills in the field of statistics (scale 1–7) were somewhat

176 At the outset of the study it was intended to compare market data from “U.S. markets” (i.e., markets that comprised U.S. students) to market data from “German markets” (i.e., markets that consisted of German students) with respect to the effect size of partition-dependence expressed in market prices. Unfortunately, the U.S. markets turned out to be fairly illiquid in the NBA Playoffs part of the study so that it is hard to draw any meaningful conclusion in terms of competence effects from these data. In order to be in a position to carry out these analyses for the FIFA World Cup part of the study, German students were allocated to some “high-competence” markets and just as many “low-competence” markets based on their self-reported World Cup competence (scale 1–7) during the registration process. Unfortunately, even in the World Cup part of the study some markets turned out to suffer from thin trading, which makes it difficult again to analyze second order effects from trading data. Therefore, no second order competence results from *trading data* are reported here.

177 See Appendix VII for the questionnaires.

higher (μ=3.9, σ=1.2 for Germans and μ=4.0, σ=1.5 for U.S. students). Finally, 39.7% of German and 32.4% of U.S. students had been active in sports betting before. Thus, potential differences in the degree of partition-dependence are unlikely to be caused by differences in these general attributes.

The remaining questions (including the trivia quiz) were collected separately for both parts of the study. They were mainly chosen to produce some more competence proxies and to check whether U.S. students were indeed more knowledgeable about NBA events (and less knowledgeable about World Cup events). Self-assessment questions included a 1–7 scale and the trivia quiz comprised seven multiple-choice questions (like "Which team has won the most championships in NBA history?"),[178] thus yielding a 0–7 score of correct answers. As it turns out, competence proxies are consistently higher for U.S. students than for German students in the NBA Playoffs event domain (subjects who completed the questionnaire after the first game took place were excluded from the following statistics; remaining sample size: N_{German}=199; $N_{U.S.}$=103). Mean self-rated competence in making judgments regarding the NBA Playoffs was 4.8 (median=5, σ=1.7) for U.S. students, but only 2.7 (median=2, σ=1.6) for German participants. In general, U.S. students are also much more interested in the NBA Playoffs (μ=5.1, median=5, σ=1.8 for U.S. students; μ=3.0, median=3, σ=1.8 for German subjects) and were going to track the Playoffs more intensely on TV, via Internet, etc. (μ=4.9, median=5, σ=1.8 for U.S. students; μ=2.7, median=2, σ=1.6 for German subjects). Similar results obtain for the trivia quiz scores (μ=4.1, median=4, σ=1.8 for U.S. students; μ=2.4, median=2, σ=1.6 for German subjects), on the whole confirming the hypothesis of U.S. students being considerably more competent and knowledgeable about NBA events than their German counterparts (all reported differences are statistically highly significant by a Kruskal-Wallis test (p<.0001)).

[178] Actually, two versions of the trivia quiz existed for each event domain to control for possible differences in the degree of difficulty. However, no such differences were found.

Table 4.5: Interaction effects between partition type and recruitment channel in before-trading probability judgments for each NBA team.

Events (Intervals)	$p(I_1) + p(I_2) / p(I_1 \cup I_2)$		$p(I_3) + p(I_4) / p(I_3 \cup I_4)$		$N_{German} / N_{U.S.}$
Recruitment channel	German	U.S.	German	U.S.	*left --- right col.*
Mean probability judgments (x100) per team					
CHI					
Sum of unpacked events	42.77	27.57	33.25	23.00	22 / 14 --- 24 / 13
Packed event	24.88	21.08	11.77	5.50	24 / 13 --- 22 / 14
Difference	17.89	6.49	21.48	17.50	
Difference of differences	11.40		3.98		
CLE					
Sum of unpacked events	64.64	63.36	32.00	24.15	22 / 14 --- 24 / 13
Packed event	47.58	48.85	10.23	12.00	24 / 13 --- 22 / 14
Difference	17.06	14.51	21.77	12.15	
Difference of differences	2.55		9.62		
DAL					
Sum of unpacked events	65.40	59.35	44.04	37.69	20 / 17 --- 27 / 13
Packed event	46.00	42.69	24.75	27.29	27 / 13 --- 20 / 17
Difference	19.40	16.66	19.29	10.40	
Difference of differences	2.74		8.89		
DEN					
Sum of unpacked events	55.17	60.40	33.15	29.29	24 / 10 --- 27 / 14
Packed event	39.81	33.57	17.21	12.90	27 / 14 --- 24 / 10
Difference	15.36	26.83	15.94	16.39	
Difference of differences	-11.47		-0.45		
DET					
Sum of unpacked events	49.29	49.20	55.04	83.71	24 / 10 --- 27 / 14
Packed event	30.00	12.64	40.42	38.00	27 / 14 --- 24 / 10
Difference	19.29	36.56	14.62	45.71	
Difference of differences	-17.27		-31.09**		
IND					
Sum of unpacked events	56.00	56.40	30.26	23.29	24 / 10 --- 27 / 14
Packed event	37.89	27.07	11.04	12.60	27 / 14 --- 24 / 10
Difference	18.11	29.33	19.22	10.69	
Difference of differences	-11.22		8.53		
LAC					
Sum of unpacked events	50.24	43.75	24.77	28.79	29 / 8 --- 26 / 14
Packed event	46.88	44.57	13.59	6.25	26 / 14 --- 29 / 8
Difference	3.36	-0.82	11.18	22.54	
Difference of differences	4.18		-11.36		
LAL					
Sum of unpacked events	55.45	50.65	37.85	19.77	20 / 17 --- 27 / 13
Packed event	41.19	37.54	21.55	13.94	27 / 13 --- 20 / 17
Difference	14.26	13.11	16.30	5.83	
Difference of differences	1.15		10.47		

to be continued on the next page

Table 4.5 continued

Events (Intervals)	$p(I_1) + p(I_2) / p(I_1 \cup I_2)$		$p(I_3) + p(I_4) / p(I_3 \cup I_4)$		N_{German} / $N_{U.S.}$
Recruitment channel	German	U.S.	German	U.S.	*left --- right col.*
Mean probability judgments (x100) per team					
MEM					
Sum of unpacked events	45.59	37.43	34.00	16.85	22 / 14 --- 24 / 13
Packed event	27.92	25.85	10.41	7.93	24 / 13 --- 22 / 14
Difference	17.67	11.58	23.59	8.92	
Difference of differences	6.09		14.67		
MIA					
Sum of unpacked events	64.85	61.53	38.07	35.38	20 / 17 --- 27 / 13
Packed event	42.81	44.23	17.75	25.41	27 / 13 --- 20 / 17
Difference	22.04	17.30	20.32	9.97	
Difference of differences	4.74		10.35		
MIL					
Sum of unpacked events	48.62	33.75	25.00	21.79	29 / 8 --- 26 / 14
Packed event	30.38	29.86	12.07	3.75	26 / 14 --- 29 / 8
Difference	18.24	3.89	12.93	18.04	
Difference of differences	14.35		-5.11		
NJN					
Sum of unpacked events	64.93	73.13	34.46	27.79	29 / 8 --- 26 / 14
Packed event	39.27	45.57	17.86	8.13	26 / 14 --- 29 / 8
Difference	25.66	27.56	16.60	19.66	
Difference of differences	-1.90		-3.06		
PHX					
Sum of unpacked events	60.91	54.29	44.08	51.08	22 / 14 --- 24 / 13
Packed event	32.92	32.00	21.00	23.57	24 / 13 --- 22 / 14
Difference	27.99	22.29	23.08	27.51	
Difference of differences	5.70		-4.43		
SAC					
Sum of unpacked events	42.88	45.90	24.22	19.36	24 / 10 --- 27 / 14
Packed event	29.52	18.86	11.79	12.60	27 / 14 --- 24 / 10
Difference	13.36	27.04	12.43	6.76	
Difference of differences	-13.68		5.67		
SAS					
Sum of unpacked events	56.66	45.63	55.73	60.00	29 / 8 --- 26 / 14
Packed event	30.77	20.29	29.62	40.63	26 / 14 --- 29 / 8
Difference	25.89	25.34	26.11	19.37	
Difference of differences	0.55		6.74		
WAS					
Sum of unpacked events	57.40	52.88	21.33	20.38	20 / 17 --- 27 / 13
Packed event	42.89	36.15	10.60	13.29	27 / 13 --- 20 / 17
Difference	14.51	16.73	10.73	7.09	
Difference of differences	-2.22		3.64		

Table 4.5 shows interaction effects between partition type (treatment) and recruitment channel in mean before-trading probability judgments for each NBA team. For each team the first row shows the sum of means for unpacked events (low intervals

in the two left-hand columns and high intervals in the two right-hand columns). The second row shows mean values for the packed event. The left-hand sub-columns of each main column show the means for subjects recruited in Germany, the right-hand sub-columns show respective values for participants recruited in the U.S. The rightmost column indicates sample size (i.e., number of participants who provided judgments for a team) broken down for each of the eight sample cells. Contrary to the Lab Study (section 3), low and high intervals are not complementary since some probability was assigned to the I_0 interval in both partitions (which is not displayed in the Table). The row "difference" shows mean effect size of partition-dependence in each sub-group, and the row "difference of differences" calculates the difference in bias strength between the two sub-groups. This value is above zero, if partition-dependence is lower for U.S. students, and it is negative, if partition-dependence is lower for German students.

As Table 4.5 reveals, bias strength varies widely within the different sub-groups and intervals. For the low intervals ($p(I_1) + p(I_2) - p(I_1 \cup I_2)$) it ranges between 3.36 (LAC) and 27.99 (PHX) for German subjects, and varies from –.82 (LAC, indicating slight reverse partition-dependence) to 36.56 (DET) for U.S. students; effect size for the high intervals ($p(I_3) + p(I_4) - p(I_3 \cup I_4)$) ranges between 10.73 (WAS) and 26.11 (SAS) for German participants, and spans from 5.83 (LAL) to 45.71 (DET) for U.S. subjects. Note that complete ignorance would imply a difference of 25.0 in terms of absolute probability since each partition is divided into four intervals ($1/N$=.25, so when a packed event is unpacked, ignorance prior judgments rise by 25.0). With respect to the presumed competence effect, the difference in partition bias between groups ranges between –17.27 (DET) and 14.35 (MIL) for low intervals, and varies from –31.09 (DET, an obvious outlier) to 14.67 (MEM) for high intervals. As it turns out, no systematic competence effects are recognizable, which is also reflected in a 2×2 factorial ANOVA that shows no statistically significant interaction effects between the degree of partition-dependence and channel of recruitment among the 32 team-interval comparisons, except the above mentioned outlier (p-value .0189 indicated by two asterisks).[179]

[179] In part, statistical inference of ANOVA results may be limited in the present case due to possible violations of assumptions (independence of observations within and between samples, normality of sampling distribution, homogeneity of variances) and/or due to small sample sizes. Including those judgments that were submitted after the Playoffs had begun increases the overall sample size by 118 German and 36 U.S. data sets. As a result, a few more interaction terms turn out to be statistically significant, but the general tendency of the findings does not change, and results may be affected by scores from the first games. For the sake of clarity, these results are not reported here.

In particular, it is noticeable that neither U.S. students nor German participants seem to be consistently less biased than their counterparts, and bias strength is generally high in both sub-samples. The channel of recruitment was chosen as a proxy for competence about the events, so maybe this proxy is not selective enough to detect interaction effects. However, carrying out the same analysis using different competence proxies (self-rated competence, interest in the Playoffs, intention to track the Playoffs on TV, etc., and score of the trivia quiz) does not alter the results with regard to the interaction effects.[180]

To reduce variance within sub-groups and to enhance statistical power by increasing the number of observations, one may pool participants' judgments (per event) for the four different teams of a four-team market (e.g., DAL, WAS, MIA, LAL).[181] To further reduce variance, one may even pool the judgments over all sixteen teams, which gives a single 2×2 comparison of the low and the high intervals, respectively. Results are displayed in Table 4.6. Note that N_{German} and $N_{U.S.}$ refer to the number of judgments (four per person) rather than to the number of subjects. However, the null hypothesis of no interaction effects still cannot be rejected for most of the interaction terms. Partition bias is sometimes lower, but sometimes even higher for U.S. students, and the difference of differences is generally not high, suggesting that competence interactions between German and U.S. students in the present data are very weak, if at all existent.

[180] For each of these competence proxies, participants were grouped into two categorical "competence" groups according to whether their self-rating/quiz score was less than or equal to the median self-rating/quiz score ("low-competence" group) or whether it was above the median ("high-competence" group).

[181] This implicitly assumes a subject's judgments for different teams to be uncorrelated. Remember that each participant faced the same partition of events for the four teams of her four-team market.

Table 4.6: Interaction effects between partition type and recruitment channel in before-trading probability judgments, pooled over four-team markets (market clones).

Events (Intervals)	$p(I_1) + p(I_2) / p(I_1 \cup I_2)$		$p(I_3) + p(I_4) / p(I_3 \cup I_4)$		$N_{German} / N_{U.S.}$
Recruitment channel	German	U.S.	German	U.S.	*left --- right col.*
Mean probability judgments (x100) per four-team market (pooled)					
SAS, MIL, NJN, LAC					
Sum of unpacked events	55.11	49.06	34.99	34.59	116 / 32 --- 104 / 56
Packed event	36.83	35.07	18.28	14.69	104 / 56 --- 116 / 32
Difference	18.28	13.99	16.71	19.90	
Difference of differences	4.29		-3.19		
PHX, CHI, CLE, MEM					
Sum of unpacked events	53.48	45.66	35.83	28.77	88 / 56 --- 96 / 52
Packed event	33.32	31.94	13.35	12.25	96 / 52 --- 88 /56
Difference	20.16	13.72	22.48	16.52	
Difference of differences	6.44		5.96		
DEN, IND, DET, SAC					
Sum of unpacked events	50.83	52.97	35.67	38.91	96 / 40 --- 108 / 56
Packed event	34.31	23.04	20.11	19.02	108 / 56 --- 96 / 40
Difference	16.52	29.93	15.56	19.89	
Difference of differences	-13.41**		-4.33		
DAL, WAS, MIA, LAL					
Sum of unpacked events	60.78	56.10	35.32	28.31	80 / 68 --- 108 / 52
Packed event	43.22	40.15	18.66	19.99	108 / 52 --- 80 / 68
Difference	17.56	15.95	16.66	8.32	
Difference of differences	1.61		8.34*		
ALL 16 TEAMS					
Sum of unpacked events	54.84	51.33	35.45	32.80	380 / 196 - 416 / 216
Packed event	37.02	32.42	17.68	16.71	416 / 216 - 380 / 196
Difference	17.82	18.91	17.77	16.09	
Difference of differences	-1.09		1.68		

What follows is a parallel analysis of probability judgments in the World Cup part of the study. As was pointed out in fn. 176, only data from German students were used from the second part of the study. Thus, the following analyses of second order effects will focus on the self-rated World Cup competence proxy of German students ("In general, how competent do you feel in making judgments regarding the FIFA Soccer World Cup 2006?", scale 1–7) instead of the recruitment channel. Without notifying them, participants were grouped into a "low-competence" group (if their competence proxy was less than or equal to the median of self-assessments) or into a "high-competence" group (if their proxy was above the median). Median self-rated competence was fairly high among German students (median=5, σ=1.6, N=263), just like the other two competence proxies (interest in the World Cup: median=7, σ=1.6; intention to track the World Cup on TV, etc.: median=6, σ=1.5) which were collected in a Web-based World Cup questionnaire. Trivia quiz scores were somewhat lower (median=4, σ=1.4).[182]

Analogous to Table 4.5, Table 4.7 shows interaction effects between partition type (treatment) and self-rated competence (either "low-competence" group or "high-competence" group) in mean before-trading probability judgments for each World Cup team. Notes to Table 4.5 similarly apply to Table 4.7. Like for the NBA events, Table 4.7 reveals that the extent of partition-dependence bias varies greatly within the different sub-groups and intervals. For the low intervals ($p(I_1) + p(I_2) - p(I_1 \cup I_2)$) bias size ranges between –9.56 (JPN, reflecting reverse partition-dependence) and 29.52 (NED) for the low-competence group, and varies from 1.27 (SCG, indicating almost partition-independence) to 41.97 (CZE) for the high-competence group; effect size for the high intervals ($p(I_3) + p(I_4) - p(I_3 \cup I_4)$) ranges between –6.31 (AUS, reverse partition-dependence) and 27.78 (ARG) for the low-competence group, and spans from 3.70 (GHA) to 45.50 (CZE) for the high-competence group. With respect to the presumed competence effect, the difference in partition bias between groups ranges between –21.62 (ITA, indicating the low-competence group being less biased) and 13.10 (ARG) for low intervals, and varies from –23.20 (CZE) to 9.49 (ARG) for high intervals.[183] It is striking that for 19 out of the 32 team-interval differences the low-competence group appears to be less biased than those subjects who were allocated to the high-competence group. Some of these "reverse" competence effects are statistically significant (by a 2 × 2 factorial ANOVA) on a 10%-level, whereas none of the positive interaction terms shows statistical significance.[184] However, none of the two competence groups seems to be consistently better calibrated than the other. The power of results also seems to be restricted by the fact that for some teams (e.g., CIV, CRC, SCG) competence interactions point in a different direction for the low and high intervals.

---

[182] Again, no differences in the degree of difficulty were found for the two versions of the trivia quiz.

[183] Note that the difference of differences is difficult to interpret for team AUS and for the low intervals of team JPN, as the low-competence groups reveal *reverse* partition-dependence, whereas the high-competence groups indicate partition-dependence in the expected direction.

[184] Again, ANOVA results may be limited due to possible violations of assumptions and/or due to even smaller sample size compared to the NBA Playoffs part of the study. Only N=4 participants provided their judgments after the World Cup had begun, so including their judgments does not at all alter the results.

Table 4.7: Interaction effects between partition type and self-rated competence in before-trading probability judgments for each FIFA World Cup team.

Events (Intervals)	$p(I_1) + p(I_2) / p(I_1 \cup I_2)$		$p(I_3) + p(I_4) / p(I_3 \cup I_4)$		$N_{low\_comp.} / N_{high\_comp..}$
Competence Level	Low	High	Low	High	*left --- right col.*
Mean probability judgments (x100) per team					
ARG					
Sum of unpacked events	55.84	51.87	67.52	63.22	19 / 15 --- 21 / 9
Packed event	26.43	35.56	39.74	44.93	21 / 9 --- 19 / 15
Difference	29.41	16.31	27.78	18.29	
Difference of differences	13.10		9.49		
AUS					
Sum of unpacked events	58.00	65.21	12.69	9.15	15 / 19 --- 13 / 20
Packed event	63.77	53.05	19.00	4.95	13 / 20 --- 15 / 19
Difference	-5.77	12.16	-6.31	4.20	
Difference of differences	-17.93*		-10.51*		
BRA					
Sum of unpacked events	19.33	38.63	93.77	88.05	15 / 19 --- 13 / 20
Packed event	5.31	10.30	77.27	59.16	13 / 20 --- 15 / 19
Difference	14.02	28.33	16.50	28.89	
Difference of differences	-14.31		-12.39		
CIV					
Sum of unpacked events	60.63	57.40	14.24	18.11	19 / 15 --- 21 / 9
Packed event	45.71	53.33	5.53	6.27	21 / 9 --- 19 / 15
Difference	14.92	4.07	8.71	11.84	
Difference of differences	10.85		-3.13		
CRC					
Sum of unpacked events	46.87	51.53	19.95	12.50	15 / 17 --- 19 / 16
Packed event	37.74	40.94	4.73	3.18	19 / 16 --- 15 /17
Difference	9.13	10.59	15.22	9.32	
Difference of differences	-1.46		5.90		
CRO					
Sum of unpacked events	48.67	75.79	46.31	28.85	15 / 19 --- 13 / 20
Packed event	44.85	57.95	25.33	7.95	13 / 20 --- 15 / 19
Difference	3.82	17.84	20.98	20.90	
Difference of differences	-14.02		0.08		
CZE					
Sum of unpacked events	61.35	71.30	46.60	65.00	20 / 10 --- 20 / 15
Packed event	39.80	29.33	24.30	19.50	20 / 15 --- 20 / 10
Difference	21.55	41.97	22.30	45.50	
Difference of differences	-20.42*		-23.2*		
ECU					
Sum of unpacked events	42.27	54.18	18.79	15.13	15 / 17 --- 19 / 16
Packed event	38.63	47.06	9.07	3.18	19 / 16 --- 15 /17
Difference	3.64	7.12	9.72	11.95	
Difference of differences	-3.48		-2.23		

to be continued on the next page

Table 4.7 continued

Events (Intervals)	$p(I_1) + p(I_2) / p(I_1 \cup I_2)$		$p(I_3) + p(I_4) / p(I_3 \cup I_4)$		$N_{low\_comp.} / N_{high\_comp..}$
Competence Level	Low	High	Low	High	*left --- right col.*
Mean probability judgments (x100) per team					
GER					
Sum of unpacked events	45.67	59.76	64.32	59.81	15 / 17 --- 19 / 16
Packed event	27.32	30.31	41.00	34.71	19 / 16 --- 15 /17
Difference	18.35	29.45	23.32	25.10	
Difference of differences	-11.10		-1.78		
GHA					
Sum of unpacked events	52.30	56.90	14.60	9.80	20 / 10 --- 20 / 15
Packed event	40.60	47.87	9.30	6.10	20 / 15 --- 20 / 10
Difference	11.70	9.03	5.30	3.70	
Difference of differences	2.67		1.60		
ITA					
Sum of unpacked events	52.95	68.40	53.85	62.67	20 / 10 --- 20 / 15
Packed event	40.10	33.93	39.30	28.50	20 / 15 --- 20 / 10
Difference	12.85	34.47	14.55	34.17	
Difference of differences	-21.62*		-19.62		
JPN					
Sum of unpacked events	49.67	63.89	20.00	18.25	15 / 19 --- 13 / 20
Packed event	59.23	56.05	10.67	4.42	13 / 20 --- 15 / 19
Difference	-9.56	7.84	9.33	13.83	
Difference of differences	-17.4*		-4.50		
NED					
Sum of unpacked events	62.00	58.27	59.10	54.11	19 / 15 --- 21 / 9
Packed event	32.48	37.22	33.05	36.87	21 / 9 --- 19 / 15
Difference	29.52	21.05	26.05	17.24	
Difference of differences	8.47		8.81		
POL					
Sum of unpacked events	56.20	71.88	39.21	29.69	15 / 17 --- 19 / 16
Packed event	39.21	53.31	16.47	8.65	19 / 16 --- 15 /17
Difference	16.99	18.57	22.74	21.04	
Difference of differences	-1.58		1.70		
SCG					
Sum of unpacked events	59.00	66.27	20.48	21.44	19 / 15 --- 21 / 9
Packed event	49.52	65.00	8.95	6.80	21 / 9 --- 19 / 15
Difference	9.48	1.27	11.53	14.64	
Difference of differences	8.21		-3.11		
USA					
Sum of unpacked events	66.25	75.00	24.20	16.33	20 / 10 --- 20 / 15
Packed event	50.25	60.33	8.25	3.60	20 / 15 --- 20 / 10
Difference	16.00	14.67	15.95	12.73	
Difference of differences	1.33		3.22		

Carrying out the same analysis using the remaining competence proxies (interest in the World Cup,[185] intention to track the World Cup on TV, etc., and score of trivia quiz) does not alter the general results with respect to the direction and statistical inference of analyzed interaction effects.[186]

Table 4.8: Interaction effects between partition type and self-rated competence in before-trading probability judgments, pooled over four-team markets (market clones).

Events (Intervals)	$p(I_1) + p(I_2) / p(I_1 \cup I_2)$		$p(I_3) + p(I_4) / p(I_3 \cup I_4)$		$N_{low\_comp.} / N_{high\_comp.}$
Competence Level	Low	High	Low	High	left --- right col.
Mean probability judgments (x100) per four-team market (pooled)					
GER, POL, ECU, CRC					
Sum of unpacked events	47.75	59.34	35.57	29.28	60 / 68 --- 76 / 64
Packed event	35.72	42.91	17.82	12.43	76 / 64 --- 60 / 68
Difference	12.03	16.43	17.75	16.85	
Difference of differences	-4.40		0.90		
NED, ARG, CIV, SCG					
Sum of unpacked events	59.37	58.45	40.33	39.22	76 / 60 --- 84 / 36
Packed event	38.54	47.78	21.82	23.72	84 / 36 --- 76 / 60
Difference	20.83	10.67	18.51	15.50	
Difference of differences	10.16*		3.01		
ITA, USA, CZE, GHA					
Sum of unpacked events	58.21	67.90	34.81	38.45	80 / 40 --- 80 / 60
Packed event	42.69	42.87	20.29	14.43	80 / 60 --- 80 / 40
Difference	15.52	25.03	14.52	24.02	
Difference of differences	-9.51*		-9.50		
BRA, JPN, CRO, AUS					
Sum of unpacked events	43.92	60.88	43.19	36.08	60 / 76 --- 52 / 80
Packed event	43.29	44.34	33.07	19.12	52 / 80 --- 60 / 76
Difference	0.63	16.54	10.12	16.96	
Difference of differences	-15.91**		-6.84		
ALL 16 TEAMS					
Sum of unpacked events	53.15	61.00	38.09	35.33	276 / 244 - 292 / 240
Packed event	39.79	44.10	22.95	17.61	292 / 240 - 276 / 244
Difference	13.36	16.90	15.14	17.72	
Difference of differences	-3.54		-2.58		

[185] Since median interest in the World Cup was 7, the low-interest group was built by subjects whose self-rating was below the median, and the high-interest group contained the subjects whose self-rating was equal to the median (cf. fn. 180).

[186] Interestingly, though, using the „interest in the World Cup" proxy yields an interaction effect (namely in the expected direction) that is statistically significant at the 5%-level for the high intervals of team GER. This is interestingly, because it may be assumed that German subjects were particularly interested in the German national team and thus competence effects might be exceptionally pronounced for this team.

Pooling participants' judgments (per event) for the four different teams of a four-team market (e.g., GER, POL, ECU, CRC), and pooling judgments for all sixteen teams (ALL 16 TEAMS) yield the results displayed in Table 4.8. (Note that N_{low_comp} and N_{high_comp} refer to the number of judgments (four per person) rather than the number of participants.) Even if effects turn out to be somehow more consistent across low and high intervals, and despite the fact that three of the interaction effects for the low intervals turn out to be statistically significant (*, ** indicate statistical significance at the 10% and 5% level, respectively), results remain mixed. Results from pooling judgments for all sixteen teams show a small "reverse" competence effect, but partition-dependence is quite balanced across sub-groups, and bias size generally remains high. Thus, the null hypothesis of no interaction effects cannot be rejected for most of the interaction terms, suggesting that competence effects in the present data are rather random errors than a general phenomenon. So either competence effects, if they exist, are not accentuated enough in the analyzed population, or competence proxies are not selective enough to detect such effects, or statistical significance is limited due to small sample sizes. Another explanation would be that subjects possibly did not think carefully enough about their judgments, as these were not incentivized.

4.4 Interim conclusions

Study 2 documents the persistence of partition-dependent pricing effects in a field experiment in which self-selected participants trade assets whose value depend on the outcomes of events in which they take great interest—the NBA Playoffs and the World Cup tournament—and for which trading lasts for several weeks. These experiments address concerns about the generalizability of lab experiments due to the limited involvement of traders and short span of trade. Probability judgments before trading begins exhibit partition-dependence that is similar in magnitude to previous psychological studies—for instance, the sum of unpacked intervals (e.g., [4, 7] playoff wins plus [8, 11] wins) is judged to about 20 percent larger in absolute probability than the packed interval [4, 11]. In addition, the bias appears to be quite insensitive to variations in the level of participants' self-perceived competence in both parts of the study.

The partition-dependence revealed by actual prices of event assets can be roughly bounded by two different methods. Using the possibility of hypothetical cross-market arbitrage at available bids and asks yields an average daily profit of about 1% (largely because there are long stretches of time where there is not a simultaneous set of bids for the unpacked assets and an ask for the packed asset). Using an interpolated-price procedure, which assumes that trades could take place continuously, but only at the worst price from the last trade and next future trade, gives hypothetical arbitrage profits around 6%. The two measures represent likely lower and upper bounds on the practical profitability from exploiting partition-dependence, and therefore bound its likely economic magnitude in markets like these. For both measures and a large majority of team markets, these potential profits are much larger than profits from executing the opposite strategy (buying unpacked intervals and selling the equivalent packed interval asset), indicating that partition-dependence is a systematic bias rather than an artifact of random error.

5 Study 3: Naturally-occurring prediction markets for economic derivatives

5.1 Institutional structure of economic derivatives

The lab and field experiments document the existence of partition-dependence when different partitions are traded (in separate markets) for the same event domain. An open question is whether these effects can be inferred from naturally-occurring prediction markets that rely on a single partition. Study 3 addresses this question.

In October 2002, Deutsche Bank and Goldman Sachs launched large-scale prediction markets for bets on the outcomes of macroeconomic indicators such as the change in U.S. non-farm payrolls (NFP), levels of the Institute for Supply Management's (ISM) purchasing manager index (a measure of business confidence), U.S. initial jobless claims (IJC) (adjusted to reflect seasonal hiring patterns), retail sales (RSX) (ex automobiles, adjusted for normal seasonal variations), and others.[187] These "economic derivatives" (ED) markets are designed to give professionals such as institutional traders (hedge funds, proprietary traders, pension funds, large banks, etc.) the opportunity to take direct positions in unexpected fluctuations of macroeconomic risks, and potentially to provide better widespread distributional forecasts of the underlying variables.[188] With economic derivatives, traders can take *direct* positions in macroeconomic risks,[189] as

[187] Other macroeconomic indicators that serve as underlying for economic derivatives markets are: Eurozone harmonized index of consumer prices (HICP) (a measure of inflation in Europe), U.S. international trade balance (ITB), and U.S. gross domestic product (GDP); see CME (2005).

[188] Initially, economic derivatives were traded over-the-counter based on an agreement with ICap to distribute the securities to the inter-dealer-market. As from September 2005, economic derivatives were organized by the International Securities Exchange (ISE) and clearing was operated and guaranteed by the Chicago Mercantile Exchange (CME). In June 2007, though, ISE and CME have decided to discontinue economic derivatives auctions due to low trading volume. Before closing the markets, some 120 participants were active in these auctions. At any auction there were about 40 participants (some 80% of them were large and small hedge funds); see Parker (2007, p. 11). The highest trading volume was generated in NFP derivatives, with an average nominal value of approximately $9 million per auction (Gadanecz, Moessner, and Upper (2007)).

[189] Macroeconomic or business cycle risks are a fundamental source of risk in corporate activity, as these factors have a major impact on sales, orders received, wage demands, interest rates, earnings, and the like; but they do also affect households, as macroeconomic factors may influence job security, consumption, or price stability. Early attempts have been made in the 1970ies to trade macroeconomic risks using derivatives, for instance by Lovell and Vogel (1973), who proposed the introduction of specific futures contracts to provide everyone with the opportunity to hedge against fluctuations in the general consumer price index (CPI). In the 1980ies, financial engineering was used to introduce investment innovations (e.g., "macro swaps" and "macro options") based on macroeconomic risks (see, e.g., Marshall et al.

they can avoid to assume the "basis risk" that is usually associated with positions in proxy securities (such as bonds, equities, foreign exchange, and other derivatives which are supposed to react in a certain direction as a result of a data release (surprise)). The basis risk is the risk that a chosen proxy security does not exactly react to the data outturn as was expected by the trader (see CME (2005), and Gadanecz, Moessner, and Upper (2007)). The markets may potentially provide *better forecasts*, as prices of economic derivatives reflect information on the entire probability distribution of underlying expectations, and not just point estimates as this is usually the case with survey-based measures (Gadanecz, Moessner, and Upper (2007)). Economic derivatives markets may attract both speculators (giving them the opportunity to express their view directly on economic statistics) and hedgers (who wish to reduce the discrete event risk associated with an economic release).

For each underlying numerical variable (i.e., the release of a specific numerical macroeconomic indicator) a diverse set of contracts is available for trading, including capped options (capped calls and floored puts), forwards (range forwards), and digital (binary) options (digital calls, digital puts, and digital range options) (see CME (2005)).[190] Capped options are "plain-vanilla" style options paying money equal to the amount by which the released indicator exceeds (call) or falls below (put) the exercise price, but are capped at the highest (call) and lowest (put) strike prices available in the auction. Range forwards pay the difference between the actual outcome of an indicator and the trade price, capped at the highest and lowest strike prices. If this difference is negative, the value falls due for payment. Digital calls and puts are options that pay out a fixed amount ($1) if they end up in-the-money and nothing otherwise. The fixed amount does not depend on *how deep* an option is in-the-money. Finally, digital range options comprise two strike prices (a lower and an upper bound) which define a certain interval. These options pay out if the released statistic falls within this range and nothing otherwise. Therefore, digital range options in economic derivatives exactly coincide with the "all-or-nothing" contracts used in Studies 1 and 2. With digital range options a trader is able to take a position in any interval bounded by two adjacent strike prices (and combinations of that). Usually, there are about 10 to 20 equidistant strike prices for

---

(1992), Bansal, Marshall, and Yuyuenyongwatana (1994), and Bansal, Marshall, and Yuyuenyongwatana (1995)), and influential work on macro markets was also done by Shiller (1993), and Shiller (2003).

[190] Trading in digital options represents about three quarters of trades, but less than half in terms of transaction volume. Hedgers usually prefer plain-vanilla options which account for much larger volumes; see Beber and Brandt (2009, p. 12).

each upcoming data release, dividing the state space in 11 to 21 mutually exclusive and collectively exhaustive ranges. It is important to stress that the ranges are deliberately set by the organizers of these markets to reflect the likely range of the outcomes. In more detail, strike prices are set to cover at least two to three standard deviations based on historical volatility of the indicator and the scale midpoint is chosen to reflect mean survey expectations.[191] For example, the strike prices for an U.S. non-farm payrolls auction may range from –250,000 (i.e., a decrease in payrolls from the previous level) to +100,000 in increments of 25,000 jobs. This generates a total of sixteen intervals and a scale midpoint of –75,000 jobs (]–∞; –250,000], [–249,999; –225,000], …, [75,001; 100,000], [100,001; +∞[).

In most financial derivatives markets, trading occurs as a continuous auction, with market makers providing liquidity to the audience by continuously posting bid and ask quotes at which they are ready to buy and sell assets. Usually, market makers hedge their positions by countertrades in the respective underlying or using other derivatives. However, hedging economic derivatives is difficult since there is no such "underlying" for macroeconomic statistics and proxy securities bear the basis risk as described above. Without market makers, though, liquidity is low and trading may suffer from thin trading. For this reason, the economic derivatives markets use a parimutuel system which is also common in horse race betting.[192] In this mechanism the prices of the instruments are based solely on relative demand for their implied outcomes and enables market clearing without discrete matching of buy and sell orders (see Baron and Lange (2003), and Lange and Economides (2005)). In parimutuel markets investors who bet on event *A* and win (i.e., event *A* occurs and associated options are in-the-money) share the winnings from those who bet on all other ("losing") events (i.e., those options which end out-of-the-money). As parimutuel clearing applies to all kinds of derivatives in the economic derivatives markets (capped options, forwards, and digital (binary) options), these instruments are decomposed into a combination of several "state contingent claims" (SCC) for valuation. The state contingent claims are in fact digital range options based on available exercise prices, a fact that highlights the relevance of the different strike-price intervals of the state space for the pricing of these derivatives. As in horse betting, the trading system periodically discloses interim prices showing what the

[191] See Filippov (2005, p. 47), and CME (2005, p. 5).

[192] See subsection 2.1.1.4 for an introductory note on continuous double auctions (as used in Studies 1 and 2) and parimutuel markets.

payouts would be if no further orders were submitted.[193] These auctions typically take place in the morning of the day on which the economic statistic is released and are sometimes preceded by other auctions on the same statistic release one or two days before (e.g., non-farm payrolls auctions are held on the morning the data are released and on the two days before). Thus, these markets usually have a very short-term forecast horizon and thereby offer hedging opportunities against so-called event risks.[194] Each auction lasts for 1 to 2 hours.

Most important in the context of the present work, the economic derivatives market prices can be used to derive a risk neutral probability density function of the market's aggregated beliefs about the outcome of every single data release.[195] Throughout the auction period the trading application continuously discloses indicative prices of available derivatives and lists filled orders, as well as an implied probability distribution derived from current market prices (strictly speaking, these are prices of the digital range options on all intervals).[196] Figure 5.1, for instance, shows the implied probabilities from one set of digital options, for a retail trade statistic announced in April 2005. Gürkaynak and Wolfers (2006), who report data covering the first 2½ years of these markets, conclude that market-generated forecasts based on prices from the economic derivatives markets are slightly more accurate than the "survey forecasts" released by Money Market Services (MMS) on the Friday before a data release, much as the Iowa political market prices are typically more accurate than comparable political polls (Berg and Rietz (2006)). To derive the market-based forecast they use the distribution of outcomes implied by market prices and calculate the distribution's mean by assuming that the probability distribution is uniform within each bin (boundaries of bins are defined by adjacent strike prices). For the MMS forecast, the "consensus" forecast typically averages across around 30 professional analysts. Their finding of market-based forecasts being slightly more accurate than the survey estimates is measured in terms of the

---

193 Since investors in the economic derivatives market are allowed to place limit orders, the parimutuel trading mechanism may result in multiple equilibriums. This problem is encountered by using a special auction-clearing tool that chooses the equilibrium prices such that the number of total trades will be maximized. (This clearing-algorithm was developed by Longitude Inc. and is called "Parimutuel Derivative Call Auction technology"; see Baron (2004).) As in many traditional auctions, all trades (at a given strike) that occur are executed at the same price, regardless of the limit price.

194 By contrast, the „macro securities" proposed by Shiller (1993) were designed to protect economic agents against long-term macroeconomic risks which may affect their livelihood.

195 See subsection 2.1.4.3. for a discussion of whether and under what circumstances prediction market prices can be interpreted as market-derived probabilities.

196 See Appendix X for a screenshot of the trading interface.

mean absolute error (MAE) and the root mean squared error (RMSE) from both mechanisms.

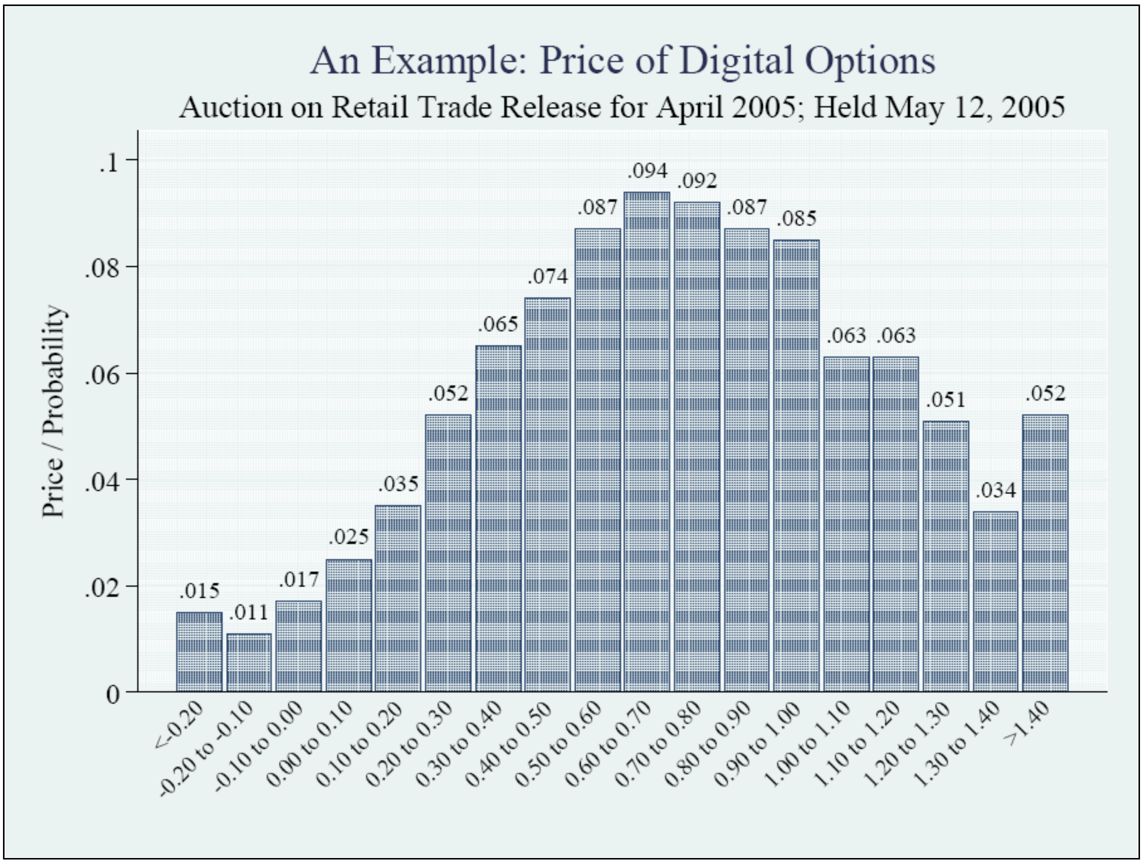


Figure 5.1: Digital option prices on ranges of retail trade statistics (taken from Gürkaynak and Wolfers (2006)).

Essentially, there are two reasons to believe in the superiority of market-derived forecasts (see Gadanecz, Moessner, and Upper (2007)): first, economic derivatives auctions take place shortly before the actual data are released, whereas surveys are generally published with a lead of one week or more, so one may argue that in the meantime more information is potentially available and is incorporated in the market-derived estimates. Second, economic derivatives prediction markets offer particularly strong incentives to trade upon one's true beliefs, as millions of dollars are involved. Analysts, in turn, may be tempted to misrepresent their views in surveys in order to position themselves relative to consensus estimates.

5.2 Ignorance-prior analysis: The "mixture" model

First note that each economic derivatives market presents a single partition of possible event outcomes to participants (the digital option outcome ranges). As a result, one cannot compare prices in two differently-partitioned events on the same interval to estimate the degree of partition-dependence, as was done through experimental manipulation in Studies 1 and 2. However, one can think of a simple econometric "mixture" model to estimate the degree of partition-dependence. For each event category x, assume

$$f_{obs}(x) = (1 - \lambda) \cdot f_{true}(x) + \lambda \cdot f_{1/N}(x), \tag{5.1}$$

where $f_{obs}(x)$ is the observed probability distribution implied by economic derivatives prices, $f_{true}(x)$ is the unobserved unbiased probability distribution, $f_{1/N}$ is an ignorance-prior distribution assigning equal probability mass to each interval, and λ is the weight on the $1/N$ ignorance prior.[197]

If each event was traded repeatedly, the empirical distribution of realized outcomes could be compared to the distribution of implied probabilities and the $1/N$ distribution to produce a sharp estimate of the apparent weight on the $1/N$ component. However, there is only one observation of implied probabilities for each point of time and each economic statistic. Therefore, the data for the different points of time and across the different statistics are pooled.[198] A mean forecast $M_{obs} = \mu(f_{obs})$ is computed for each auction by weighting the interval midpoints by the observed probabilities f_{obs}, and a respective ignorance prior mean $M_{1/N}$ is determined by assigning equal weight to each

[197] Note the structural similarity of the "mixture" model presented here and partition-dependence modeled within the framework of support theory (cf. equation (2.15) in subsection 2.2.4.1). Note also the structural similarity of the presented model and the "information model" introduced by Sobel and Raines (2003) in their analysis of the favorite-longshot bias in racetrack betting. Their model is based on a Bayesian updating process that starts with an ignorance prior probability of $1/N$ for all of the N entrants. Posterior beliefs then result from a weighted combination of $1/N$ probability and probability implied by the information signals received by the bettor.

[198] In the following, the same data on the economic derivatives are used as in Gürkaynak and Wolfers (2006). To make the four statistics they use (non-farm payrolls, ISM business confidence, retail sales (ex auto), and initial unemployment claims) comparable one may follow Gürkaynak and Wolfers (2006) and normalize the data by the historical size of the forecast error.

interval midpoint.[199] For symmetric partitions, $M_{1/N}$ thus coincides with the scale midpoint of available strike prices.

From (5.1) it follows that M_{obs} is a linear function $(1 - \lambda) \cdot M_{true} + \lambda \cdot M_{1/N}$. Call the actual realization of the economic statistic X and apply a little algebra to see that the observed forecast error can be written as:[200]

$$M_{obs} - X = [M_{true} - X] - \lambda / (1 - \lambda) \cdot [M_{obs} - M_{1/N}] \tag{5.2}$$

That is, the observed forecast error, $M_{obs} - X$, has two components. The first component is the error term from a de-biased forecast based on $f_{true}(x)$ (which is expected to have expectation zero). The second component is a negatively-weighted term which reflects the degree of partition-dependence (through the weight λ). Intuitively, suppose the forecast from market data M_{obs} is above the equal-weight forecast $M_{1/N}$. If there is partition-dependence contaminating $f_{obs}(x)$, then $f_{obs}(x)$ is biased downward (toward $M_{1/N}$) relative to the de-biased ideal forecast $f_{true}(x)$ (which is an unbiased predictor of X). This downward bias means the forecast error, $M_{obs} - X$, is likely to be negative. Thus, when $[M_{obs} - M_{1/N}]$ is positive $M_{obs} - X$ is likely to be negative (and vice versa). The strength of the negative correlation can be used to estimate $-\lambda / (1 - \lambda)$ and the implied λ. For cases in which $M_{obs} > M_{1/N}$, Figure 5.2 illustrates how the forecast error $M_{obs} - X$ becomes negative by M_{true} being biased downward (toward $M_{1/N}$) as a result of partition-dependence. Roughly speaking, the model predicts that the unbiased forecast M_{true} is actually further away from $M_{1/N}$ than indicated by the observed forecast M_{obs}.

[199] With regard to the midpoints of the *tail* intervals, one may follow Gürkaynak and Wolfers (2006, p. 6, fn. 9): “For the tails we impute an upper- and lower-bound so that the midpoint would be equal to the mean of that bin if the PDF were normal.”

[200] The derivation of the regression model can be obtained from Appendix XI.

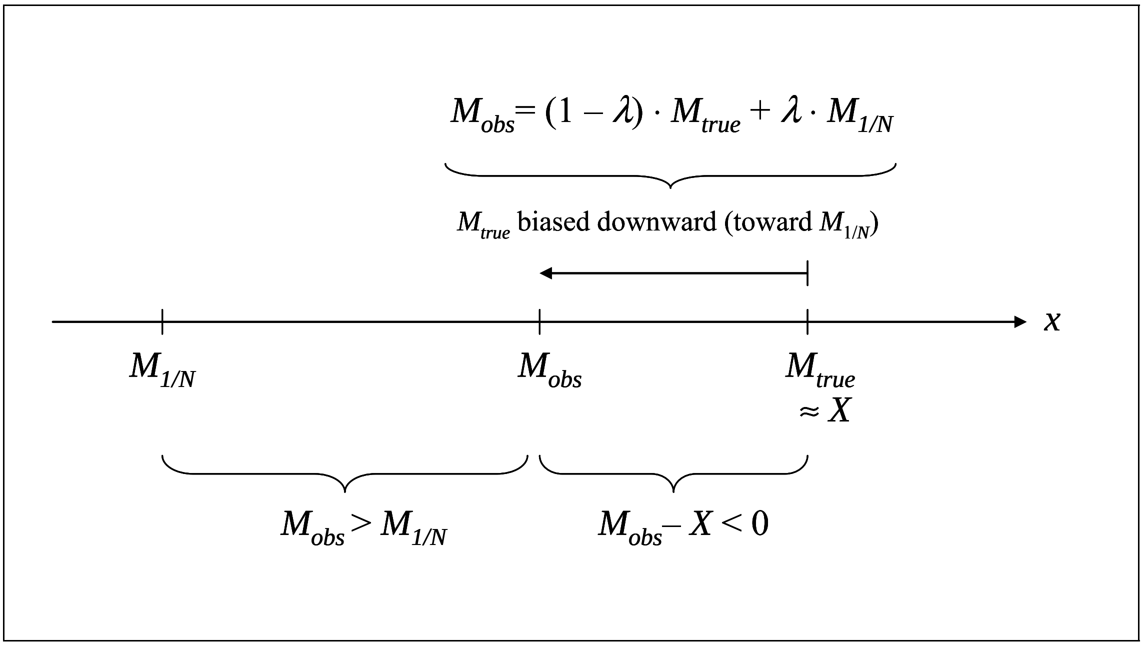


Figure 5.2: Illustration of a negative forecast error ($M_{obs} - X$) < 0 if $M_{obs} > M_{1/N}$.

Table 5.1 summarizes the results of estimating regression (5.2) for markets for four different statistics.

Table 5.1: Results of regressions of forecast errors on the difference between observed forecast and 1/N forecast.

		Regression (5.2) Results			Implied Weight λ on 1/*N*	
	No. of Events	Coefficient –λ / (1 – λ)	*t*-Statistic	*p*-value (one-tailed)	λ Implied by Regression	Error-minimization
Initial jobless claims (IJC)	64	0.13	0.16	0.44	-0.15	0.06
Business confidence (ISM)	30	-0.64	-1.88	0.04	0.39	0.08
Non-farm payrolls (NFP)	33	-1.29	-1.53	0.07	0.56	0.56
Retail sales (excl. autos) (RSX)	26	-1.01	-1.32	0.10	0.50	0.50
All statistics pooled	153	-0.77	-2.60	0.01	0.44	0.39

Note. *P*-value for test of regression coefficient different than zero is one-tailed.

There is modest support for a negative correlation between forecast errors and the "forecast – 1/*N* gap", which is consistent with bias toward a 1/*N* prior. One of the event domains (initial jobless claims) shows no bias; the other three event domains show substantial bias. However, small sample sizes for individual event domains make the effects statistically marginal. The coefficient estimated from pooling all the event domains (–.77) is more significant and implies a value of the weight on 1/*N* of λ=.44 (because –.77 is an estimate of $-\lambda/(1-\lambda)$). Three of the four event domains imply values of λ from .39 to .56.

A second analysis uses the "mixture" model introduced in (5.1) to infer the unobserved unbiased probability distribution $f_{true}(x)$ from the observed distribution $f_{obs}(x)$ for various weights λ. As mentioned above, under partition-dependence the observed forecast M_{obs} is biased toward the 1/*N*-forecast. Thus, the model can be used to *inflate* the difference from 1/*N* to *undo* this compression. Furthermore, one can ask whether bias-corrected forecasts $M_{true}(\lambda)$ are more accurate than observed forecasts M_{obs}. From (5.2) it follows:

$$M_{true} - X = [1/(1-\lambda)] \cdot (M_{obs} - \lambda \cdot M_{1/N}) - X \tag{5.3}$$

That is, the true forecast error $M_{true} - X$ can be computed as a function of λ, the unknown weight on the 1/*N* ignorance prior. Using the Gürkaynak and Wolfers (2006) data, one may compute the mean absolute error (MAE) between the actual realization of the economic statistic, and the λ-weighted combination of the forecast from the observed probability, M_{obs}, and the forecast $M_{1/N}$, for various weights λ. The results are displayed in Figure 5.3: For each of the four available economic statistics the graph shows the mean absolute error (MAE) over different values of λ. Mean absolute errors are normalized by historical standard errors of survey-based forecasts to make them comparable across event domains. For example, $\lambda = 0$ implies that no weight is attached to the ignorance-prior; in that case M_{true} coincides with M_{obs}. For $\lambda \to 1$ it is assumed that M_{true} consists almost solely of the 1/*N* prior. Assuming $M_{true} - X$ to have expectation zero, one can ask for which implied weight of λ the true forecast error reaches a minimum. The values of λ that minimize the error from an λ-weighted combination are provided in the rightmost column of Table 5.1 and are indicated by a circle for each economic indicator in Figure 5.3. For two of the statistics (initial jobless claims (IJC) and business confidence (ISM)) the weights are low, but positive. For the other two statistics (non-farm payrolls (NFP) and retail sales (RSX)) the weights are close to .50. For all statistics pooled (indicated by the vertical gray line in Figure 5.3), the error-minimizing λ weight is about .39.

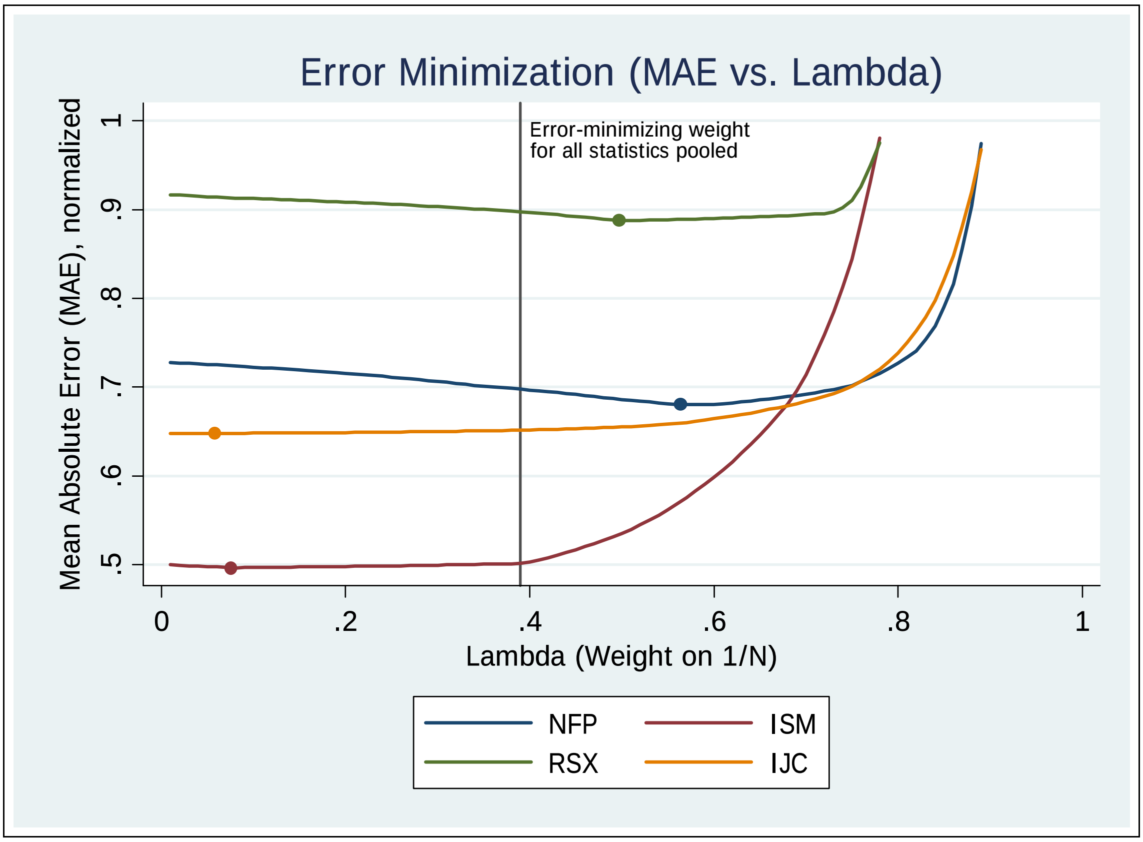


Figure 5.3: Error Minimization (MAE vs. Lambda).

It may be possible that the low weights for the ISM and IJC statistics on the one hand, and the much higher weights for RSX and NFP on the other hand reflect some knowledge or information effect. One may argue that the latter two indicators are more difficult to predict than the former as their forecast errors (MAE) turn out to be generally higher. In addition, retail sales figures (RSX) are usually capable of big surprises and are still quite imprecise when first released (they are the first report of the month on consumer spending, do not include the services sector, are measured only in nominal dollars, etc.), and for non-farm payrolls (NFP) only a few reliable data are available for that month at the time when the indicator is released; initial jobless claims (IJC), by contrast, are released on a weekly basis and are only subject to minor revisions (see Baumohl (2007), p. 74 for RSX, pp. 25-26 for NFP, and p. 40 for IJC). If an economic indicator is more difficult to predict, though, then one may expect people to implicitly put more weight on the ignorance prior (i.e., insufficient adjustment) than when the statistic is reasonably well predictable (i.e., stronger adjustment due to more reliable information). This is consistent with the results of the presented error-minimization analysis.

Admittedly, though, as can be seen from Figure 5.3 the MAE reacts rather insensitive to variations of λ, at least for weights below .70 (except for the ISM indicator of business confidence which is quite sensitive for $\lambda > .40$). That means, even for sizeable weights on the ignorance prior distribution, the forecast error does not change much. Accordingly, great uncertainty remains in determining how much weight is attached to the ignorance prior empirically, that is, to what extent values of M_{obs} are actually biased toward $M_{1/N}$. However, for other event domains than those considered here, identified values of λ may have a much greater impact on forecast accuracy. Another issue that arises with the regression (5.2) model is that it does not control for other variables that may explain why the expectation of an economic number derived from economic derivatives auctions is biased toward the average of the midpoints of the strike ranges. One thing that comes to mind is that observed forecasts M_{obs} may rather be biased toward the prior outcome X_{t-1} of the indicator in question (assuming that the market designer will be guided by the last realization of the indicator when deciding how to design the strike prices for the next auction). However, as was mentioned above, for each series of auctions strike prices are adjusted to reflect current survey expectations (and to span at least two standard deviations based on historical volatility), and in fact there is considerable variation in the absolute level of strike intervals over time. Running a simple regression of the observed market-based forecast M_{obs} on (i) the current scale midpoint $M_{1/N}$, and (ii) on the most recent outcome X_{t-1} suggests that much more of the total variance is explained by the scale midpoint of current intervals (R^2_{IJC} =.96; R^2_{ISM} =.98; R^2_{NFP} =.96; R^2_{RSX} =.87) than is explained by the most recent outcome of the economic statistic (R^2_{IJC} =.52; R^2_{ISM} =.90; R^2_{NFP} =.53; R^2_{RSX} <.01).[201] Thus, anchoring on the latest realization of the economic number can be largely ruled out as explanation of why observed market forecasts are biased toward $M_{1/N}$. Alternatively, as strike intervals are chosen symmetrically around a midpoint to reflect mean survey expectations, market-derived forecasts could be simply biased toward the mean survey forecast. In fact, correlation between the scale midpoint and survey consensus forecasts is very high,[202] but

[201] Since the release of macroeconomic indicators is often subject to later revisions, the latest (revised) estimate of last month's actual number (reported by Yahoo) was used in the regression as the most recent outcome of the economic statistic.

[202] The values of ρ range between .96 and .99 for the four indicators.

the values do not perfectly coincide. Given that Gürkaynak and Wolfers (2006, Section 3) find market-derived forecasts to be slightly more accurate than survey expectations (and taking into account that market-derived forecasts could even be more precise when correcting for the implied weight λ on $M_{1/N}$ as demonstrated by the error-minimization analysis), there is no reason to assume that anchoring on mean survey expectations is better suited to account for the observed bias than pure partition-dependence, though this possibility cannot be entirely neglected. Finally, despite the fact that the way in which strike price intervals are chosen is common knowledge among the market participants, the credibility account (assuming that some meaningful information on likely outcomes is conveyed by the way in which the state space was partitioned by the organizers) cannot be completely ruled out in the context of economic derivatives, since unlike in Studies 1 and 2 only a single partition exists at each point in time.

Taken together, these calculations suggest considerable degree of partition-dependence in two of the four macroeconomic statistics, and in the pooled economic derivatives market prices (a total sample of 153 separate markets). The results are instructive, as they demonstrate the existence of partition-dependence even in large-scale prediction markets with high stakes at play and professional market participants. In part, the results can be interpreted to be consistent with a knowledge or information effect in which the relative weight on the ignorance prior is reduced as more reliable information are available to foresee the outcome of the indicator.

6 Conclusion

Partition-dependence is the finding that judged probabilities—as expressed by individuals directly or implied by asset market prices for event-contingent claims in prediction markets—vary systematically with the set of exclusive and exhaustive events into which a state space happens to be "partitioned". This phenomenon was first discovered in a cumulative series of psychology experiments beginning in the late 1970s. The basic finding in those studies is that "unpacking" a single category or interval into two or more component intervals which are logically equivalent increases the original interval's total expressed probability. This finding is inconsistent with the description invariance principle postulated by rational theory of decision making. The present thesis tackles the question of to what extent market forces are able to eliminate (or at least diminish) partition-dependence observed in individual judgments. The experiments show that the bias transfers to competitive markets and that the phenomenon never disappears even under variations in (i) the event domains that are being traded, (ii) the salience of the relation between the differently partitioned assets and the obviousness of price discrepancies, (iii) the self-selection of participants, (iv) the length of the markets, and (v) whether the markets are experimentally-created or are created by large market firms.

Study 1 demonstrated pronounced partition-dependence under standard lab conditions for short-run (20 minutes) markets. Furthermore, although market experience diminishes partition-dependence it does not eliminate the bias. Unpacking one event interval (of three) into two component intervals (out of three) increases its judged probability by about .25. The degree of partition-dependence varies with the competence level of participants in subjectively judged probabilities for two of three event domains, but competence effects do not show when comparing equilibrium prices from markets that comprised differently competent traders. It is possible that some event domains used in this study are more information-driven (and thus competence-driven) than others. It is also likely that the range of the ordinal scaled competence proxies reflects differently pronounced differences in terms of absolute competence across event domains. In addition, competence differences among traders are also reflected in the risk of their portfolios and in general trading activity. Creating an effective link between differently partitioned markets by disclosing price discrepancies during the trading sessions further reduced the effect size of partition-dependence, though the reduction was moderate and unsystematically across event domains. Study 2 documents similar partition-dependence in longer-run markets (several weeks) on events in which the participants took great interest (NBA Playoffs and FIFA soccer World Cup). Unpacking event intervals led to hypothetical arbitrage profits of 1–6%, which is considerably smaller than in the lab markets. Controlling for the competence of participants in individual probability judgments suggests that the bias appears to be quite insensitive to variations in the level of participants' self-rated competence for both parts of the study. Study 3 examined field data from prediction markets for macroeconomic statistics with a single partition in each market. Econometric techniques suggest that probabilities implied by prediction market prices are a convex combination of partition-independent probabilities and an "ignorance prior" ($1/N$ for each of N intervals) with a weight λ for the prior distribution of around .50 in two of four markets and .05–.10 in the other two markets. In addition, there are slight indications that the degree of uncertainty in predicting a certain indicator is positively correlated with the weight λ attributed to the ignorance prior.

First note that if markets were opened with two different partitions, and traders were allowed to trade in both markets, there is little doubt that arbitrage would erase obvious differences. That is, if the price of events I_1 and I_2 were both higher than the identical packed event $I_1 \cup I_2$, arbitrage would bring the sum of event I_1 and I_2 prices into line with the price of $I_1 \cup I_2$. This (conjectured finding) simply means that if different partitions were actually traded on the same events no partition-dependence would be observed in comparing certain sets of event prices. However, in practice there is usually no reason why two different partitions would be created and traded simultaneously. Therefore, the relevant question is whether revealed prices could conceivably be a combination of highly accurate prices for each interval and a $1/N$ error term for each interval when a single partition is traded. The analysis of the 153 separate economic derivatives markets reported in Study 3 suggests the answer is "Yes", that such partition-dependence could influence prices.

A central principle in including all three studies in this thesis is to search for a single explanation that can apply to all the results. If one does prefer a single explanation for the many findings, the combination of experimental and empirical methods used in the three studies suggests that a basic behavioral propensity to bias judgments over N intervals toward $1/N$ is a component of what is observed. No other plausible explanation can explain the results of *all* three studies:

The apparent bias cannot be due to information conveyed by the choice of partition because the subjects in Studies 1 and 2 were told about *both* partitions. Any information conveyed by the choice of partition should affect both markets equally and therefore any partition-dependence observed in cross-market comparison cannot be traced to shared information. The apparent bias is also not likely to be entirely due to the nature of trading institutions since Studies 1 and 2 used double auctions and the Study 3 data come from a pari-mutuel auction mechanism. The apparent bias cannot be entirely due to naïveté of subjects, since traders in Study 1 participated actively in trading markets (and they deliberately opted for doing so, rather than just cashing in on their initial endowment for sure without trading) and since there is self-selection of active traders in Study 2 (NBA Playoffs and World Cup) and Study 3 (economic derivatives) markets. The apparent bias cannot be entirely due to overweighting of low probabilities (say, below .10) because many of the sub-event probabilities are much higher than that threshold.

More generally, these studies suggest two important themes in thinking about the implication of psychology for economics. One theme is that attention is grabbed by salient presentations of intervals, and people probably do not always compensate for the effect of attention-grabbing. For example, Morwitz, Johnson, and Schmittlein (1993) report that simply asking people whether they will buy a car in the next year increases their tendency to buy a car—by 50%. Similarly, unpacking an interval into the two components increases attention to those components and seems to increase implied probability. Because there is typically no canonical normative way to partition a continuous variable, how the variable's possible outcomes are divided into intervals can inevitably influence perceptions of the likely value of that variable.

The second theme is that the extent to which individual psychological processes influence market prices will depend on the processes and on the markets. As Camerer and Fehr (2006) note, in some market institutions the biases of a small number of traders will be amplified by strategic complementarity, and in other institutions biases are reduced because unbiased traders can profit by extinguishing biases created by other traders. The partition-dependence discussed here seems to exist to various extents in different experimental and field markets, but its robustness and persistence over time should be explored in further studies. It is remarkable that partition-dependence survives in the competitive asset markets of the presented studies even though highly competent traders were present in these markets. This evidence is a challenge to the "smart few hypothesis” according to which only a few unbiased traders are enough to result in rational market outcomes (as long as these traders have sufficient capital under their disposal to exploit mispricing). Thus, a good starting point for further research would be to explore in greater detail which market forces (like “cancellation”, “learning”, “evolutionary”, “marginal traders”, etc.) effectively contribute to eliminating individual biases from asset markets and under what circumstances. Moreover, while the microstructural design features of the presented experimental markets were chosen such that they offer best conditions (based on previous evidence from experimental asset markets) for the partition bias to be expunged, future research should further examine the robustness of partition-dependence with respect to systematic changes in market microstructure (like, e.g., different auction mechanisms or short selling constraints) and with respect to the obviousness of existing price divergence.

Of course, for financial economists the next step is to explore whether an understanding of this bias is manifested by institutional practices. As noted above there is evidence that capital budgeting decisions across corporate divisions tend to be biased toward $1/N$ (when N is the number of divisions). Similarly, in corporate finance one can imagine, say, 10 tranches of debt (i.e., groups of similarly-risky debt obligations) which vary from high-risk to low-risk (like this is common for asset-backed securities and other securitization transactions). If the first 9 tranches are small in size, holders of debt in the 10^{th} tranche might feel safer than they should, thinking there are 9 tranches ahead of them which need to default before those defaults imply a macroeconomic single that their 10^{th}-tranche debt is in danger (rather than adding up the *amount* of debt in those tranches). Firms that bundle debt into tranches might exploit this perceived bias to create many small high-risk tranches to artificially reassure those holding the larger amount of low-risk debt. Whether this type of phenomenon occurs is an empirical question. But the here presented results at least suggest an interesting hypothesis of this sort that could be explored in this setting and in others.

7 Appendices

Appendix I: Instructions for Study 1: Laboratory study

WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

Instructions for the experiment „Stock market in the laboratory“

Prof. Craig R. Fox (UCLA), Prof. Colin Camerer (Caltech),
Prof. Thomas Langer, Ulrich Sonnemann
(Universität Münster)

0. Introduction

Welcome to this experiment and thank you very much for taking the time to support our research.

Your task for approximately the next two hours will be to trade assets on a computer based trading system.

Before starting the experiment, we would like to familiarize you with the experiment and how the trading software that you will use works.

Therefore we would like to fully explain the three following questions:

- What will be traded?
- How does the trading system work?
- How will you be paid?

In studies like this one it is standard to provide the participants with written instructions and to go through them together, in order to guarantee that every group has exactly the same information about how the experiment works. Please keep general questions until the end of the introduction, as they may be answered during the following explanations. Although if there is something which you do not understand at all, please ask the people in charge, so that they can explain the point in further detail.

1. What will be traded?

In this experiment you will trade very simple assets which you can probably better imagine as bets on the occurrence of a certain future event (e.g. „Germany will be European Soccer Champion 2008“). Depending on whether the event occurs or not the asset will pay the owner either 100 or 0 monetary units (MU). Due to the fact that the outcome of the event is uncertain during the trading period, the price usually varies between the two extremes.

In a market you can always trade three assets simultaneously which all depend on the same basic event. At the end of the event, only one of the three assets will pay 100 MU. To clarify this, we will give you an example which will also be used later in a practice trading round:

Example. The uncertain future event is what share of votes the SPD party will receive in the next election of the German “Bundestag“. The three assets relating to this event can be described in the following way:

Asset 1: SPD.[0-29.9]	This asset will pay exactly 100 MU if the SPD share of votes is smaller than 30.0 % or will expire worthless otherwise. The final outcome for this, and the following assets, is based on all valid secondary votes (indirect votes for the party), which is determined by the “Bundeswahlleiter.”
Asset 2: SPD.[30.0-34.9]	This asset will pay exactly 100 MU if the SPD share of votes is between 30.0 % and 34.9 % or will expire worthless otherwise.
Asset 3: SPD.[35.0+]	This asset will pay exactly 100 MU if the SPD share of votes is equal to or more than 35.0 % or will expire worthless otherwise.

Please note that after the election of the German “Bundestag” only one out of the three assets will pay 100 MU, while the other two assets will expire worthless (since the intervals cover every possible outcome of the event, but at the same time do not overlap).

Before you start trading you should think about how much you are willing to pay for these assets and for how much you are willing to sell them for in the market. Trading will only occur between the participants of each experimental session. There are no external and/or computer-controlled participants in the market (one small exemption is the unit-portfolio, which will be explained later). Transaction costs do not exist for these assets.

During the experiment you will not trade assets on political events, but on events concerning financial markets, weather and sports. The order in which you will deal with these three topics will be determined randomly and varies for each group.

Furthermore, for each topic the groups will not trade the exact same three assets, since there are two different partitions (i.e. subset of assets), which will be assigned to your experiment session randomly. In order to guarantee that each participant has the same information, the following table shows all groups of three assets used in this study (even those that you will not be dealing with) including an accurate description of the payment conditions.

Financial markets ---- Event: DAX-closing in 2 weeks		
Asset definition: An asset will pay 100 cents, if the outcome of the underlying event exactly fits the definition of the asset. The other two assets will expire worthless. Relevant for the payment of the assets in the financial markets domain is the Xetra-DAX closing (incl. final auction) in two weeks from today (i.e., on May 8, 2007). Xetra-DAX closing as of April 23, 2007 was 7335.62.		
	Partition 1	Partition 2
Asset 1	DAX.[0 - 7248.99]	DAX.[0 - 7415.99]
Asset 2	DAX.[7249 - 7415.99]	DAX.[7416 - 7563.99]
Asset 3	DAX.[7416+]	DAX.[7564+]

Weather ---- Event: Maximum Temperature in Münster at the end of May		
Asset definition: An asset will pay 100 cents, if the outcome of the underlying event exactly fits the definition of the asset. The other two assets will expire worthless. Relevant for the payment of the assets in the weather domain is the maximum temperature two meters above ground-level (abbreviation: TX) at the weather center Muenster/Osnabrueck (station no.: 10315) on May 31^{st}, 2007 in degrees Celsius determined by Germany's National Meteorological Service ("Deutscher Wetterdienst").		
	Partition 1	Partition 2
Asset 1	Temp.[up to 15.9]	Temp.[up to 19.9]
Asset 2	Temp.[16.0 - 19.9]	Temp.[20.0 - 23.9]
Asset 3	Temp.[20.0+]	Temp.[24.0+]

Sports ---- Event: Total number of goals scored on the 34th game day of German 1. Bundesliga (Season 2006/2007)		
Asset definition: An asset will pay 100 cents, if the outcome of the underlying event exactly fits the definition of the asset. The other two assets will expire worthless. Relevant for the payment of the assets in the sports domain is the total number of goals scored on the final game day of German 1. Bundesliga in the season 2006/2007 (soccer/men).		
	Partition 1	Partition 2
Asset 1	Goals.[0 - 20]	Goals.[0 - 25]
Asset 2	Goals.[21 - 25]	Goals.[26 - 30]
Asset 3	Goals.[26+]	Goals.[31+]

At the beginning of each trading round, you will receive an initial endowment of 2000 MU, which partly consists of assets and partly consists of cash. If you spend all your liquid funds or sell all your assets, you are not able to buy or sell any more; i.e. debts or short selling are not permitted. However, in order to obtain liquidity in cash or assets you can use a so called “unit-portfolio”, which will be explained in the next paragraph.

Unit-portfolio („Unit-PF“)

The unit-portfolio is based on the idea that a full set—consisting of all 3 assets of one event—will always lead to a certain payment of 100 MU. This is independent of the outcome of the uncertain event, since one of the assets will always pay an amount of 100 MU, while the other two assets will expire worthless.

Therefore the person in charge of the experiment changes 100 MU for a complete set of assets at any time during the experiment. This trade takes place via the so called unit-portfolio, which can be bought or sold like the other assets via the trading system. However, the trading partner for this kind of deal is always the person in charge of the experiment and the price is fixed at 100 MU, i.e. it does not depend on bids and asks. In particular it should be clear that this deal does not depend upon the value of a single asset, but the fact that all 3 assets combined represent a certain payment of 100 MU. To clarify this, we return to our example. If you buy one “unit-PF”, one asset of each of the available assets is added to your portfolio. Thus, exactly one asset of type SPD.[0-29.9], one asset of SPD.[30.0-34.9] as well as one asset of type SPD.[35.0+] will be added, while your cash balance will be lowered by 100 MU. After the purchase of the „unit-PF“, you are able to trade and sell these assets in the market independently.

By selling a “unit-PF”, 100 MU will be added to your account, while simultaneously your portfolio will be reduced by one of each type of asset (of course, this sale is only permitted if you own at least one asset of each type). Hence, by using the “unit-PF” you can either acquire liquidity in assets, if you have enough cash or liquidity in cash, as long as you own a positive quantity of all types of assets.

In addition, by using the “unit-PF” you can exploit arbitrage opportunities in the market more directly. Arbitrage means completing a transaction, which results in a sure profit.

Arbitrage opportunity 1. If the market prices for individual assets enable you to purchase a complete set of assets for less than 100 MU, you can obtain a certain profit by buying these 3 assets for less than 100 MU and selling them via the “unit-PF” for 100 MU to the person in charge of the experiment.

Arbitrage opportunity 2. Alternatively, you can obtain a certain profit, if it is possible to sell the 3 assets individually in the market for more than 100 MU. Then you can purchase a complete set of assets via the “unit-PF” for 100 MU and sell them in the market as individual assets for a higher price afterwards.

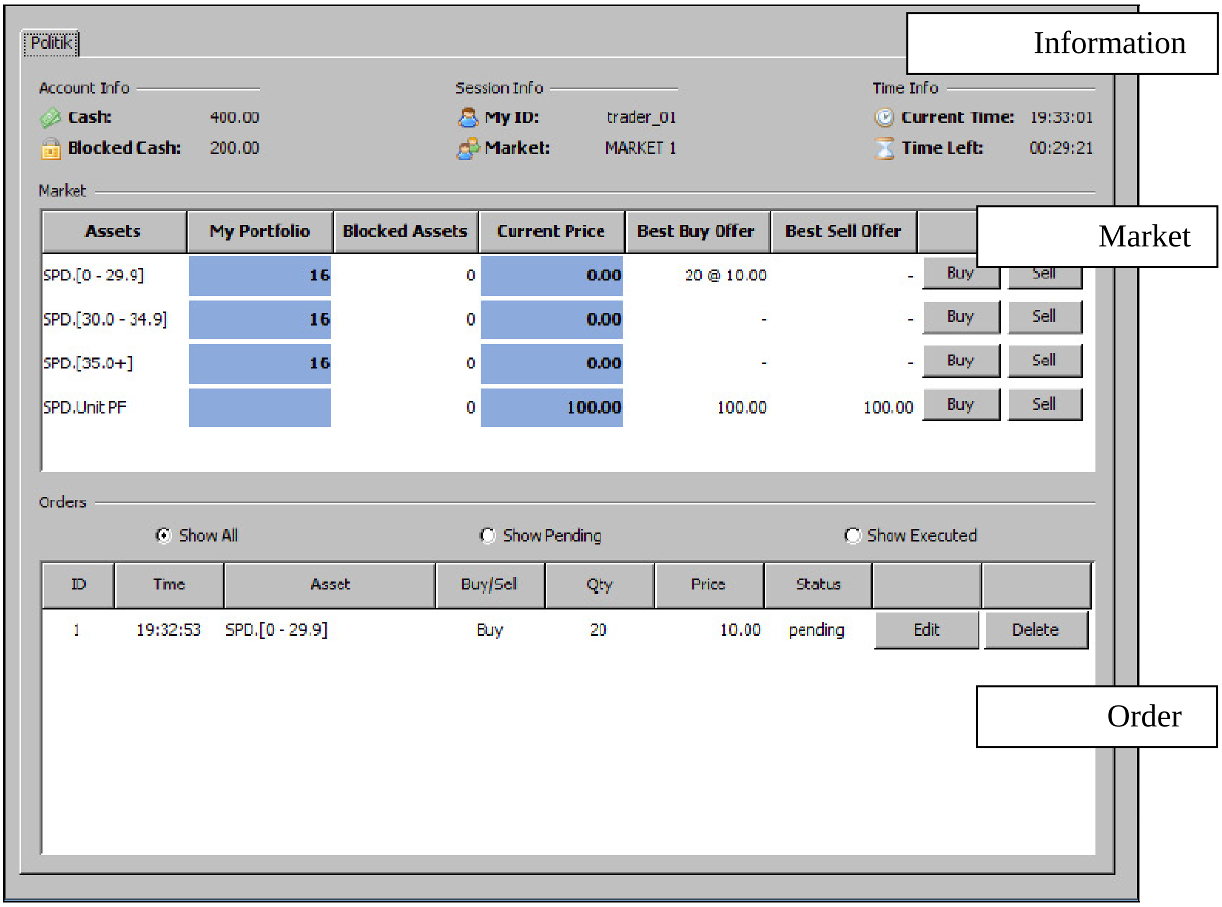
Such arbitrage opportunities are not excluded by the trading system automatically. Therefore, you should be aware that arbitrage may be possible!

2. How does the trading system work?

The trading system is a so called continuous double auction, i.e. at any time during the trading round you can act as buyer or seller. After a test trading phase you will trade a total of 6 rounds, each of 10 min. Each scenario appears twice, so you will deal with 3 different scenarios in total.

We will now explain the trading software in detail.

Trading screen



The trading screen is divided into three areas:

1. “Information Area“:

In this area you can find “account info”, “session info” and “time info”. The “account info” shows information on your cash balance and blocked cash which will be explained in more detail later on.

2. “Market Area“:

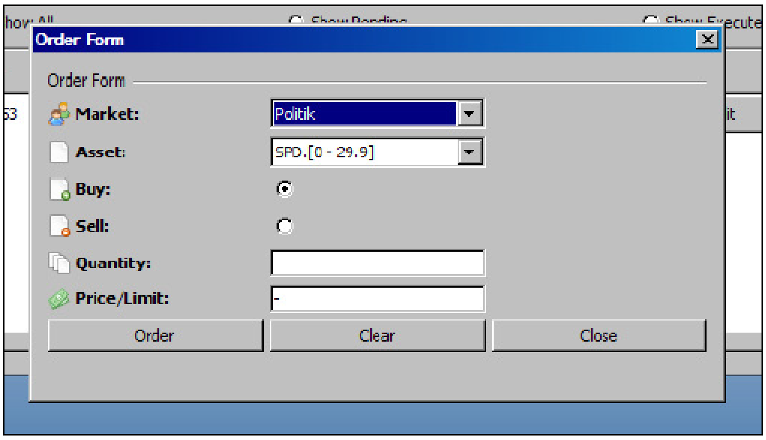
This area gives an overview about the tradable assets, your own portfolio, blocked assets, the recent market price (“current price”) and the best (highest) buy and the best (lowest) sell quote. The best buy order (sell order) indicates the volume and the price limit of the highest buy order (lowest sell order) at present, separated by an “@” symbol. For example, the display “20 @ 10” means, that one or more traders are ready to buy a quantity of 20 assets SPD.[0 - 29.9] at a price of 10 MU each.

3. “Order History”:

This area shows all of your orders which are pending or executed (currently there is a quantity of 1 order in the system). By pushing the “delete” button an outstanding order will be deleted. Using the “edit” button means all data regarding this order will be transferred into the “edit form” which subsequently appears.

Order placement

You can place an order by pushing one of the “buy”/”sell” buttons in the market area. Thereupon the “order form” appears:



You can complete your order by adding/modifying how many assets you would like to trade and for which price limit. By pushing the “clear” button you will delete all values previously entered, by pushing the “order” button you will confirm your order and it will be processed by the system (“order accepted”).

The “order form” remains active until you press the “close” button. While the “order form” is still active you can configure and submit further buy or sell orders.

After confirming your order, it will be processed by the system. When trading, this can last for 1-2 seconds. Your trading screen will be blocked during this time.

Regarding order placement, you should pay particular attention to the following two error messages:

Error “not enough cash” and error “short selling restriction”:

These error messages will appear, if you:

1. do not possess enough cash and/or if you have already placed too many active buy orders (“not enough cash”), or, if you

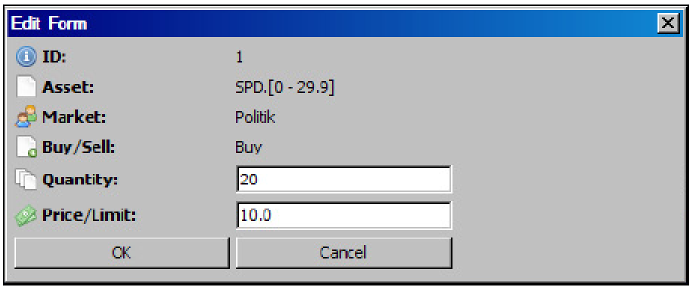
2. do not possess enough assets and/or if you have placed too many active sell orders (“short selling restriction”).

This is because any order placed will block your cash or assets until the orders are executed or cancelled. Blocked cash will be shown in the “account info”, blocked assets will be displayed in the associated column of the market area. This prevents you from placing orders which cannot be executed. Current orders can be cancelled or edited anytime, in order to access the blocked cash/assets again.

In addition, the error message “*do not trade with yourself*” may occur. This happens if you try to place a buy and a sell order for the same asset simultaneously and if the limit price of the buy order is the same or higher than the limit price of your own sell order.

Order editing and cancelling

Orders not yet executed can be cancelled and edited anytime, in order to do this; pick a current order, which is marked as “pending” in the section “orders”, and push one of the related buttons “edit” or “delete” which correspond to this order. “Delete” immediately deletes the order from the system, “edit” removes the order from the system as well, but at the same time inserts the data into the “edit form” that appears:



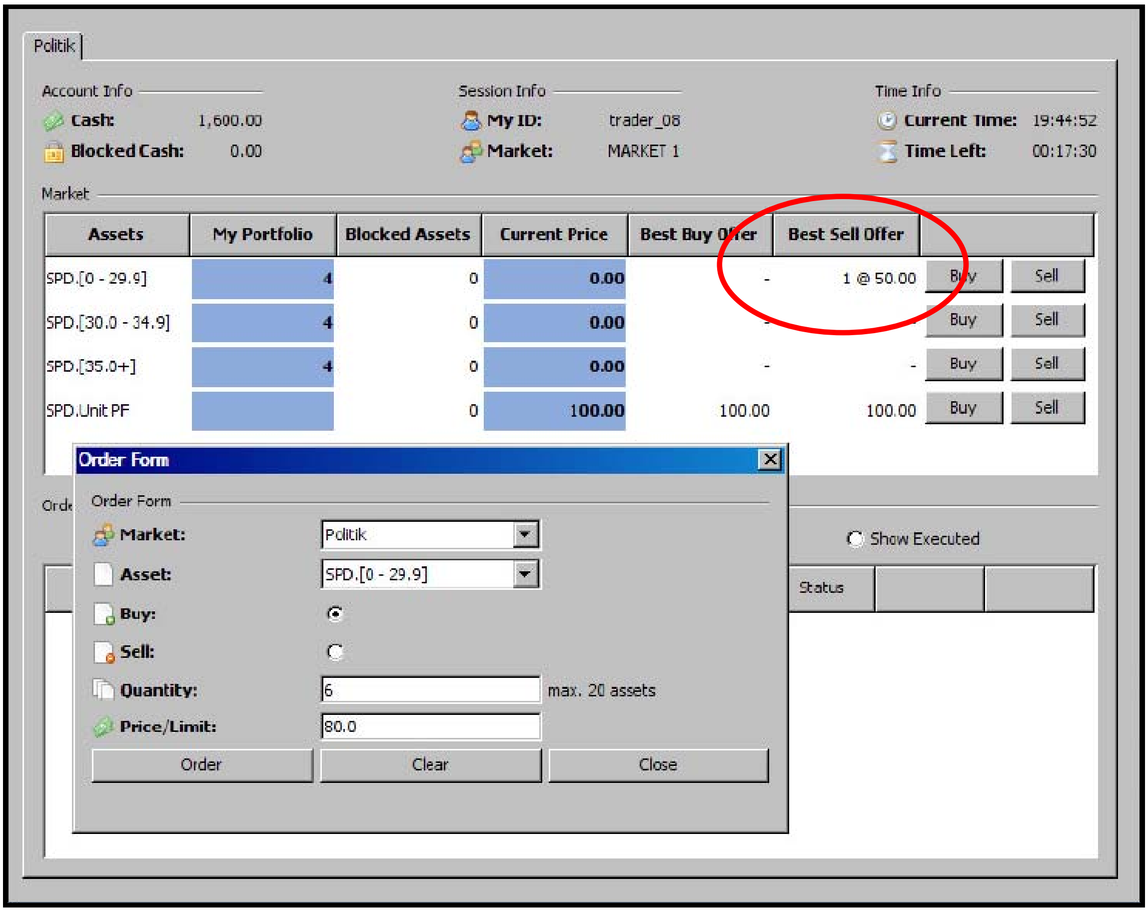
In this “edit form” you can modify the quantity and the limit price of your order. By pressing “OK” the modified order will be placed to the trading system, by pressing “cancel” the original order will be placed to the system again.

Order execution

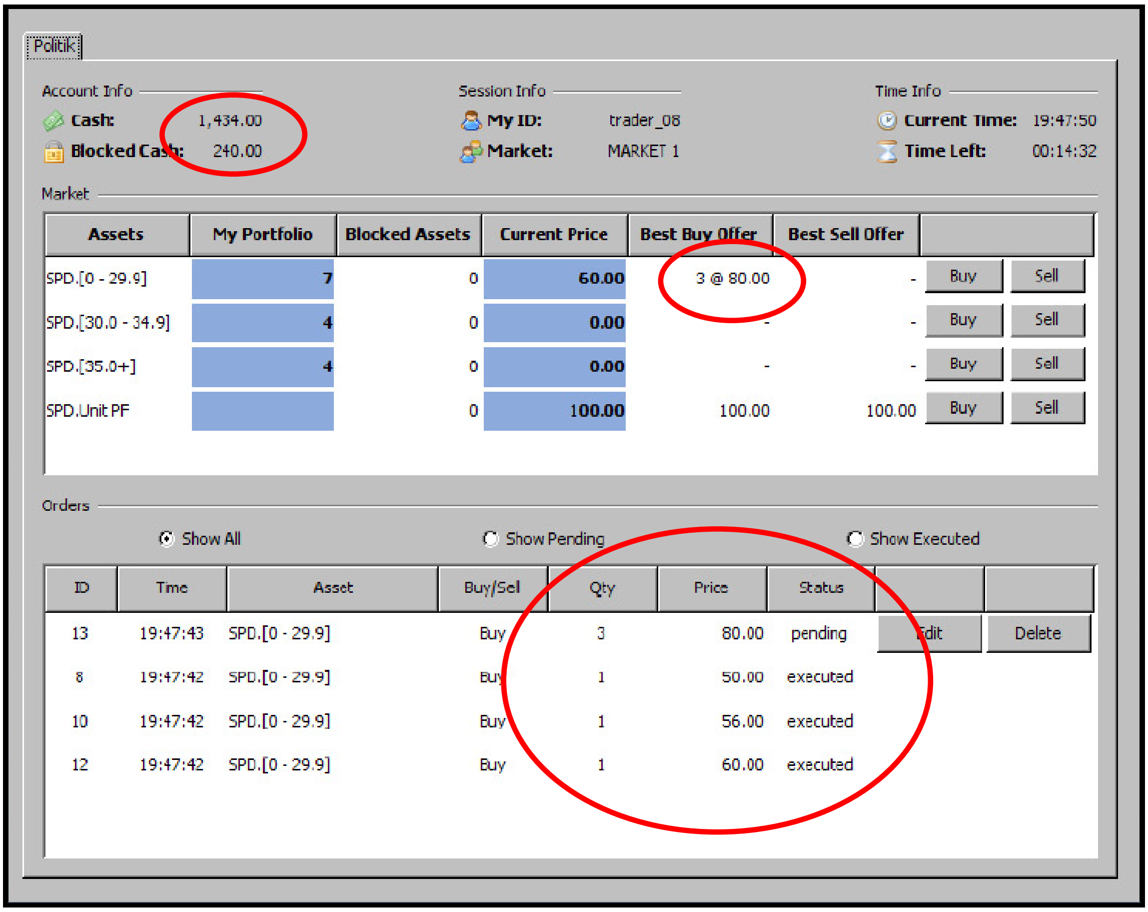
As previously mentioned, the trading system is based on a continuous double auction. Two orders will be executed against each other simultaneously, only if the buy order has the same or a higher price limit as the corresponding sell order. Please look at the following order book, i.e. the data sheet of current buy and sell orders, which are entered into the system but not yet executed:

buy order	*sell order*
	1 @ 60 MU
	1 @ 56 MU
	1 @ 50 MU
-none entered-	

Assume you would like to buy assets and enter a buy order "6 @ 80" MU into the trading system, i.e. you are willing to buy a quantity of 6 assets with a price limit of 80 MU each. You definitely know that you are going to buy 1 asset as you can see on your "trading screen" that a quantity of 1 asset is offered at a price of 50 MU in the market (best sell order).



The remaining part of the order book is not visible on the trading screen. Due to the fact that you are willing to pay up to 80 MU per asset, you will buy a quantity of 3 assets in total (see order book).



Important: Your original order quantity of 6 has been split up, so that 3 orders, each consisting of 1 asset, have been completed, while one order of 3 assets remains in the trading system. The price which is shown in the “price” column is your limit price (for the pending order) or the trading price (for the executed orders).

The part of your order which was not yet executed (3 @ 80) now appears as the best buy offer in the order book. Therefore, an amount of 240 MU was blocked and is displayed as “blocked cash” in your “account info”.

If two orders can be executed against each other, the transaction will always take place at the price limit of the previously requested buy or sell order (time priority).

3. How will you be paid?

To encourage you to carefully think about your decisions, your compensation/payment will be incentive-compatible. This means, your trading performance (as well as a little bit of luck) will determine your payment.

For each experimental group we will randomly select one out of the six trading rounds and use this as the basis for your payment. For this round we will determine the value of your portfolio and add it to your cash balance (to decide which asset will pay 100 MU and which assets will be worthless, we have to wait until the relevant uncertainty (financial markets, weather, or sports) is resolved).

The total value of your portfolio will be divided by 100
and equals your payment in Euro!

We will inform you by e-mail about how much you earned. At a later date, you will be able to pick up your money from room 263 during office hours (daily 10 till 12) and you will need to bring identification (student ID, identity card, etc.).

Good luck!

Appendix II: Session-individual DAX interval boundaries in Study 1

Table A.II.1: Session-individual DAX interval boundaries in Study 1: Basis treatment.

Session Slot	Date	Partition 1			Partition 2		
		I_1	I_2	$I_3 \cup I_4$	$I_1 \cup I_2$	I_3	I_4
1	5/2/2007	[0 - 7327.99]	[7328 - 7496.99]	[7497+]	[0 - 7496.99]	[7497 - 7646.99]	[7647+]
2	4/24/2007	[0 - 7248.99]	[7249 - 7415.99]	[7416+]	[0 - 7415.99]	[7416 - 7563.99]	[7564+]
3	4/25/2007	[0 - 7181.99]	[7182 - 7347.99]	[7348+]	[0 - 7347.99]	[7348 - 7494.99]	[7495+]
4	4/26/2007	[0 - 7256.99]	[7257 - 7423.99]	[7424+]	[0 - 7423.99]	[7424 - 7572.99]	[7573+]
5	4/26/2007	[0 - 7256.99]	[7257 - 7423.99]	[7424+]	[0 - 7423.99]	[7424 - 7572.99]	[7573+]
6	4/27/2007	[0 - 7300.99]	[7301 - 7468.99]	[7469+]	[0 - 7468.99]	[7469 - 7618.99]	[7619+]
7	5/2/2007	[0 - 7327.99]	[7328 - 7496.99]	[7497+]	[0 - 7496.99]	[7497 - 7646.99]	[7647+]
8	4/25/2007	[0 - 7181.99]	[7182 - 7347.99]	[7348+]	[0 - 7347.99]	[7348 - 7494.99]	[7495+]
9	4/25/2007	[0 - 7181.99]	[7182 - 7347.99]	[7348+]	[0 - 7347.99]	[7348 - 7494.99]	[7495+]
10	4/26/2007	[0 - 7256.99]	[7257 - 7423.99]	[7424+]	[0 - 7423.99]	[7424 - 7572.99]	[7573+]
11	4/27/2007	[0 - 7300.99]	[7301 - 7468.99]	[7469+]	[0 - 7468.99]	[7469 - 7618.99]	[7619+]
12	4/27/2007	[0 - 7300.99]	[7301 - 7468.99]	[7469+]	[0 - 7468.99]	[7469 - 7618.99]	[7619+]

Table A.II.2: Session-individual DAX interval boundaries in Study 1: Info treatment.

Session Slot	Date	Partition 1			Partition 2		
		I_1	I_2	$I_3 \cup I_4$	$I_1 \cup I_2$	I_3	I_4
1	5/7/2007	[0 - 7437.99]	[7438 - 7608.99]	[7609+]	[0 - 7608.99]	[7609 - 7759.99]	[7760+]
2	5/7/2007	[0 - 7437.99]	[7438 - 7608.99]	[7609+]	[0 - 7608.99]	[7609 - 7759.99]	[7760+]
3	5/7/2007	[0 - 7437.99]	[7438 - 7608.99]	[7609+]	[0 - 7608.99]	[7609 - 7759.99]	[7760+]
4	5/8/2007	[0 - 7443.99]	[7444 - 7613.99]	[7614+]	[0 - 7613.99]	[7614 - 7764.99]	[7765+]
5	5/8/2007	[0 - 7443.99]	[7444 - 7613.99]	[7614+]	[0 - 7613.99]	[7614 - 7764.99]	[7765+]
6	5/8/2007	[0 - 7443.99]	[7444 - 7613.99]	[7614+]	[0 - 7613.99]	[7614 - 7764.99]	[7765+]
7	5/9/2007	[0 - 7357.99]	[7358 - 7525.99]	[7526+]	[0 - 7525.99]	[7526 - 7675.99]	[7676+]
8	5/9/2007	[0 - 7357.99]	[7358 - 7525.99]	[7526+]	[0 - 7525.99]	[7526 - 7675.99]	[7676+]
9	5/9/2007	[0 - 7357.99]	[7358 - 7525.99]	[7526+]	[0 - 7525.99]	[7526 - 7675.99]	[7676+]
10	5/11/2007	[0 - 7329.99]	[7330 - 7497.99]	[7498+]	[0 - 7497.99]	[7498 - 7647.99]	[7648+]
11	5/11/2007	[0 - 7329.99]	[7330 - 7497.99]	[7498+]	[0 - 7497.99]	[7498 - 7647.99]	[7648+]
12	5/11/2007	[0 - 7329.99]	[7330 - 7497.99]	[7498+]	[0 - 7497.99]	[7498 - 7647.99]	[7648+]

Appendix III: Median equilibrium market prices and judgments

Table A.III.1: Median equilibrium prices (2nd trading round) and individual judgments (pre-trading and post-trading): Basis treatment.

Treatment		Finance: Pre-Trading Individual Judgment	Finance: Equilibrium Market Prices (Round 2)	Finance: Post-Trading Individual Judgment	Sports: Pre-Trading Individual Judgment	Sports: Equilibrium Market Prices (Round 2)	Sports: Post-Trading Individual Judgment	Weather: Pre-Trading Individual Judgment	Weather: Equilibrium Market Prices (Round 2)	Weather: Post-Trading Individual Judgment
1	I_1	0.200	0.129	0.200	0.250	0.244	0.250	0.120	0.051	0.100
1	I_2	0.500	0.563	0.500	0.400	0.469	0.400	0.320	0.244	0.300
	I_1+I_2	0.700	0.714	0.700	0.700	0.716	0.700	0.500	0.271	0.400
2	$I_1 \cup I_2$	0.400	0.453	0.400	0.400	0.422	0.400	0.200	0.130	0.200
	PD difference	*0.300*	*0.261*	*0.300*	*0.300*	*0.294*	*0.300*	*0.300*	*0.141*	*0.200*
2	I_3	0.400	0.417	0.400	0.360	0.385	0.400	0.350	0.361	0.350
2	I_4	0.200	0.177	0.160	0.200	0.149	0.150	0.400	0.474	0.400
	I_3+I_4	0.600	0.570	0.600	0.600	0.553	0.600	0.800	0.842	0.800
1	$I_3 \cup I_4$	0.300	0.358	0.300	0.300	0.414	0.300	0.500	0.690	0.600
	PD difference	*0.300*	*0.212*	*0.300*	*0.300*	*0.139*	*0.300*	*0.300*	*0.152*	*0.200*
Average PD difference			***0.237***			***0.217***			***0.147***	

Table A.III.2: Median equilibrium prices (2nd trading round) and individual judgments (pre-trading and post-trading): Info treatment.

Treatment		Finance: Pre-Trading Individual Judgment	Finance: Equilibrium Market Prices (Round 2)	Finance: Post-Trading Individual Judgment	Sports: Pre-Trading Individual Judgment	Sports: Equilibrium Market Prices (Round 2)	Sports: Post-Trading Individual Judgment	Weather: Pre-Trading Individual Judgment	Weather: Equilibrium Market Prices (Round 2)	Weather: Post-Trading Individual Judgment
1	I_1	0.200	0.121	0.150	0.215	0.161	0.200	0.150	0.091	0.100
1	I_2	0.500	0.425	0.500	0.400	0.409	0.400	0.300	0.343	0.300
	I_1+I_2	0.700	0.594	0.700	0.700	0.570	0.670	0.500	0.444	0.500
2	$I_1 \cup I_2$	0.450	0.453	0.450	0.400	0.422	0.400	0.250	0.219	0.250
	PD difference	*0.250*	*0.141*	*0.250*	*0.300*	*0.148*	*0.270*	*0.250*	*0.225*	*0.250*
2	I_3	0.350	0.388	0.400	0.350	0.395	0.400	0.350	0.350	0.400
2	I_4	0.175	0.140	0.100	0.200	0.255	0.200	0.300	0.387	0.395
	I_3+I_4	0.550	0.532	0.550	0.600	0.615	0.600	0.750	0.730	0.750
1	$I_3 \cup I_4$	0.300	0.412	0.300	0.300	0.447	0.330	0.500	0.584	0.500
	PD difference	*0.250*	*0.119*	*0.250*	*0.300*	*0.169*	*0.270*	*0.250*	*0.146*	*0.250*
Average PD difference			***0.130***			***0.158***			***0.185***	

Appendix IV: Aggregated development of price differences over time (info treatment)

(see next page)

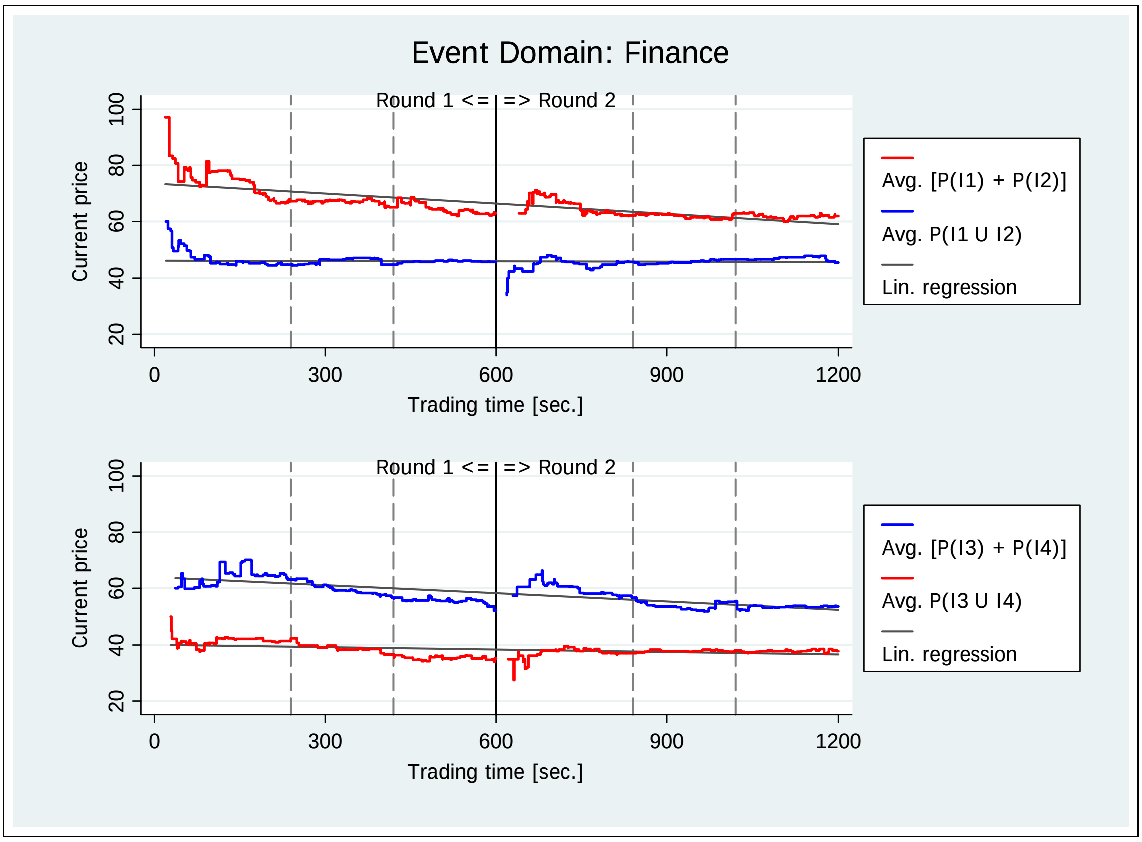


Figure A.IV.1: Development of price differences over time for the finance assets (info treatment).

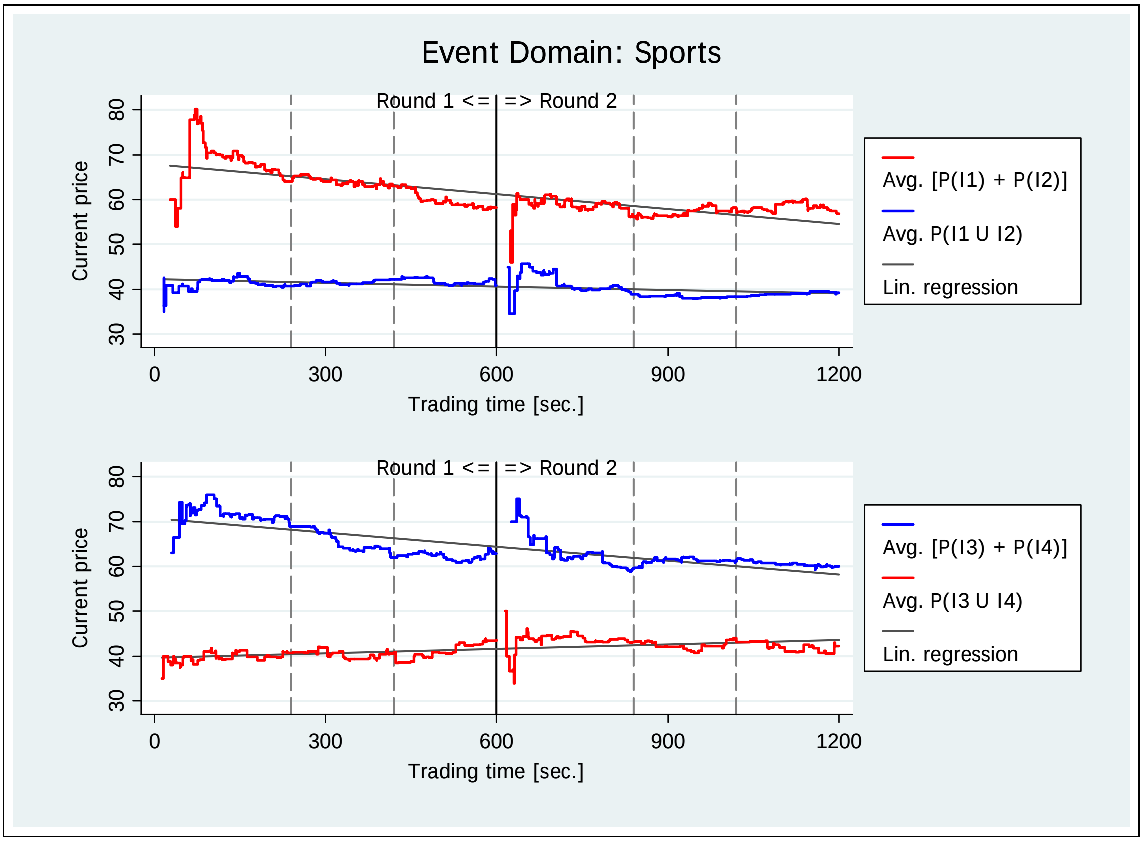


Figure A.IV.2: Development of price differences over time for the sports assets (info treatment).

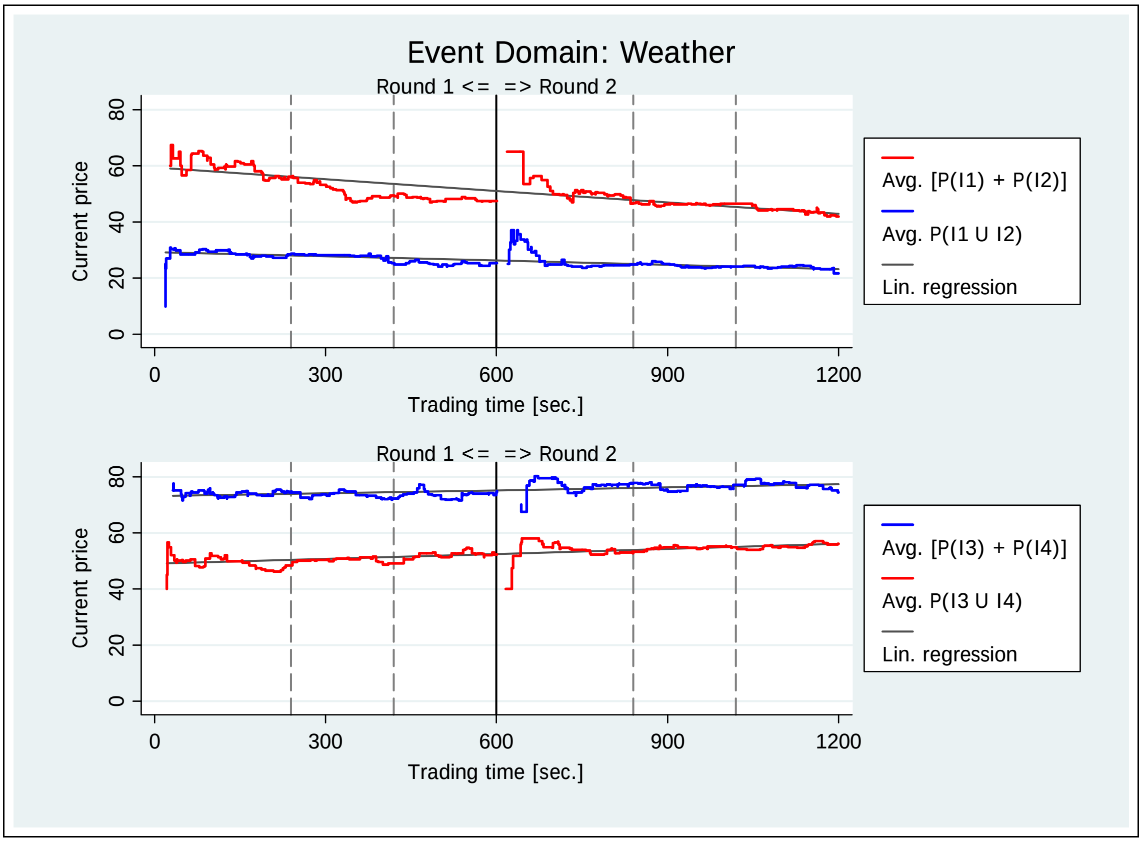


Figure A.IV.3: Development of price differences over time for the weather assets (info treatment).

Appendix V: Instructions for Study 2: An NBA/FIFA field experiment

NBA Playoffs/FIFA World Cup Study – Study Instructions

Prof. Craig R. Fox (UCLA), Prof. Colin Camerer (Caltech),
Prof. Thomas Langer, Ulrich Sonnemann
(University of Muenster, Germany)

Table of Contents

1. Introduction

Thank you very much for registering for our NBA Playoffs/FIFA World Cup Study!

In the next weeks you will have the opportunity to trade online contracts whose payoffs depend on the outcomes of this year's **NBA Basketball Playoffs** as well as the **FIFA Soccer World Cup '06** in Germany. In the course of this experiment you will be assigned to both one NBA market **AND** one FIFA market. Thus, in the first market you will trade on the NBA Playoffs and in the second market you will trade on the FIFA World Cup. But don't worry, you need not be both a basketball and soccer expert in order to successfully participate in this experiment!

Because the study is based on real sports events, the trading period for the NBA Playoffs will be from April 20 through June 22, 2006 (at the latest) and for the FIFA World Cup from the middle of May through July 9, 2006. Thus, the **overall trading period** lasts **from April 20 through July 9, 2006** and we kindly request that all participants actively trade throughout this period.

You can expect to earn cash payments of €20 (about $24) per person on average, depending on your decisions, the outcomes of the games, and your individual trading performance. It is therefore possible to earn more or less than €20/$24 (including earning nothing), but there is NO way for you to lose any money!

In order to make sure that all participants have the same information about the procedure of this study and on the trading software, we **strongly recommend that you go through this instruction file very carefully**. However, if you feel familiar with some of the information provided in this document, e.g. the NBA Playoffs mode or the trading system, you can skip those sections.

We set up an NBA/FIFA Study homepage where you can access the trading software, and get news and additional study-related information. There will also be a **FAQs section**, where you can get more detailed information on the trading system etc. You can reach this webpage at:

http://www.internetexperiment.de/NBA-FIFA-Study/

There will also be a support helpdesk which can be contacted via:

tradingstudy@hss.caltech.edu or
trading.study@anderson.ucla.edu

We hope that you will enjoy participating in this study--thank you very much for supporting our research!

Prof. Thomas Langer (University of Muenster, Germany)
Prof. Craig R. Fox (UCLA, USA)
Prof. Colin Camerer (Caltech, USA)
Ulrich Sonnemann (University of Muenster, Germany)

2. Timetable

In the following Table you can find a summary of all key data related to this study:

Date/Period	Study	NBA Basketball Play-offs	FIFA Soccer World Cup
Wed, April 19		▪ regular season ends, ▪ set of qualified teams will be set (Playoff Matchups)	
Thu, April 20	▪ link to NBA ques-tionnaire sent to participants via email, ▪ practice market opened ▪ all NBA markets opened, start of NBA trading period		
Sat, April 22		▪ Playoffs begin	
Sat, May 20	▪ link to FIFA ques-tionnaire sent to participants via email		
Mon, May 22	▪ all FIFA markets opened, start of FIFA trading period		
Thu, June 8		▪ NBA Finals start date (winner East-ern vs. winner Western Confer-ence)	
Fry, June 9			▪ opening game, ▪ start of group phase
Fry, June 9 through Fry, June 23			▪ group phase
Thu, June 22	▪ last NBA markets closed	▪ latest possible end date of NBA Finals	
Sat, June 24 through Tue, June 27			▪ round of last 16 (1/8-finals)
Fry, June 30 through Sat, July 1			▪ 1/4-finals
Tue, July 4 through Wed, July 5			▪ 1/2-finals
Sat, July 8			▪ game for 3rd place
Sun, July 9	▪ last FIFA markets closed		▪ World Cup Final
Mon, July 10	▪ link for final ques-tionnaire sent to participants via email, ▪ start paying partici-pants		

Tab. 1: Timetable.

3. Assets and Markets

Note: This section explains the most important and crucial aspects of the experimental design and should be read carefully by ALL participants!

3.1 Contingent Claims

The basic activity in this experiment is to trade claims which pay money depending on the outcomes of this year's NBA Playoffs and FIFA Soccer World Cup 2006. Since the first market you will be trading in is based on the NBA Playoffs, we will concentrate on these markets right now and tell you the details for the FIFA World Cup markets later on.

A typical market will consist of four intervals/ranges for the **total number of wins** for a particular team. **For example**, one market might trade four claims on the San Antonio Spurs, spanning the total number of games the Spurs will win in the 2006 Playoffs. One claim will pay a fixed sum of money, 100 Eurocents (= €1 which is about $1.20) if, after the Playoffs are over, the Spurs have won 0-3 games. If you think the Spurs are likely to win very few games, you should buy this claim. If you think they are likely to win more than 3 games, you should sell this claim. Another claim will pay 100 Eurocents if their win total is 4-7; a third will pay 100 Eurocents if the total is 8-11; and a fourth will pay 100 Eurocents if the total is 12-16.

General Question:
"How many games does a particular team win in the 2006 NBA Playoffs?"

For the second part of this experiment we will now concentrate on the details of the FIFA World Cup markets (please refer to the first part of the instructions file for the assets and markets description of the NBA markets).

A typical market will consist of four intervals/ranges for the ***total number of goals scored*** *by a particular national team during the entire tournament (ex penalty shoot outs).* ***For example****, one market might trade four claims on Brazil's national team, spanning the total number of goals Brazil will score in the World Cup. One claim will pay a fixed sum of money, 100 Eurocent (= €1 which is about $1.20) if, after the World Cup is over, Brazil has scored 0-2 goals. If you think Brazil is likely to score very few goals, you should buy this claim. If you think they are likely to score more than 2 goals, you should sell this claim. Another claim will pay 100 Eurocent if, at the end, their goal total is 3-8; a third will pay 100 Eurocent if the total is 9-11; and a fourth will pay 100 Eurocent if the total is 12 or more.*

General Question:
"How many goals (ex penalty shoot outs) does a particular national team score during the entire FIFA Soccer World Cup 2006 tournament?"

The basic idea of these markets is that after the Playoffs (and the World Cup) are over, exactly one asset/claim in each market will pay 100 Eurocents per claim, while the three remaining claims will become worthless. This is due to the fact that the claim intervals represent all possible outcomes, but do not overlap. However, since the outcome of the event is uncertain during most of the trading period, the price for a claim usually varies between 0 and 100 Eurocents. Before you start trading you should carefully think about how much you are willing to pay for these claims and how much you are willing to sell them for in the market!

You will be randomly placed into one market for NBA Playoff games and (later on) another for FIFA World Cup results. **Important**: Each market consists of four separate "team markets", so you can trade in four different teams in each market. In each "team market" you will be endowed with a set of cash and assets representing an initial wealth of 1,000 Eurocents (= €10 or appr. $12). You do not have to trade if you do not wish to.

Please note that there will be different partitions, i.e. asset intervals in the markets. The partitions in the NBA markets are as follows:

Partition 1: [0-3], [4-11], [12-15], [16] games won.
Partition 2: [0-3], [4-7], [8-11], [12-16] games won.

Please note that there will be different partitions, i.e. asset intervals in the markets. The partitions in the FIFA markets are as follows:

Partition 1: [0-2], [3-8], [9-11], [12+] goals.
Partition 2: [0-2], [3-5], [6-8], [9+] goals.

You will be randomly assigned to a market of twenty participants with one of the partitions above. *The market composition was reshuffled for the FIFA markets, thus it is likely that you will face other traders in your FIFA market than in the NBA markets.* There are no external and/or computer-controlled participants in the market. There are no explicit transaction costs for your trading.

3.2 Unit Portfolio

For each team there is an "extra asset" in the market which is called the unit portfolio (Unit PF). The Unit PF is based on the idea that a full set of all four assets will always guarantee a certain payment of 100 Eurocents. This payoff is independent of how many games a particular team has won, since exactly one (and only one) of the claims will pay 100 Eurocents, while the other three assets will expire worthless. *This payoff is independent of how many goals a particular team scores, since exactly one (and only one) of the claims will pay 100 Eurocent, while the other three assets will expire worthless.*

Therefore you will be able to exchange 100 Eurocents for a complete set of assets at any time in the experiment. Practically, this trade with the experimenter takes place via the unit portfolio, which can be purchased or sold like the other assets via the trading system. However, the trade partner for this deal will always be the experimenter and the price for the unit portfolio is fixed at 100 Eurocents, i.e. it does not depend on bids and asks in the market. In particular it should be clear that this trade does not allow any assessment of the value of a single asset, but the fact that all four assets combined represent a certain payment of 100 Eurocents. To clarify this, we return to our example of the San Antonio Spurs. If you buy one unit portfolio, there will be one asset of each claim in the market added to your portfolio. Thus, exactly one asset of type 0-3, one asset of 4-7, one asset of 8-11 as well as one asset of 12-16 will be added, while your cash will be reduced by 100 Eurocents. *To clarify this, we return to our example of Brazil's national team. If you buy one unit portfolio, there will be one asset of each claim in the market added to your portfolio. Thus, exactly one asset of type 0-2, one asset of 3-8, one asset of 9-11 as well as one asset of 12+ will be added, while your cash will be reduced by 100 Eurocent.*

The Unit PF allows you to buy or sell a complete set of all four assets for a particular team at a fixed price of 100 Eurocents. You might want to do this in order to increase your cash on hand or stock in assets, without taking any risks.

After buying a unit portfolio, you can trade and sell the purchased assets in the market independently! By selling a unit portfolio, 100 Eurocents will be added to your cash account, while simultaneously your portfolio stock will be reduced by one of each asset (please note that this sale is only permitted if you own at least one unit of each asset). Hence, by using the unit portfolio you can either enhance liquidity in assets (as long as you have enough cash), or enhance liquidity in cash (as long as you have at least one of all the four assets in the market).

By using the unit portfolio you can exploit arbitrage opportunities in the market more directly. Arbitrage means performing a transaction which results in a sure profit.

Arbitrage opportunity 1. If the market prices enable you to purchase a full set of assets for less than 100 Eurocents, you can realize a sure profit by buying these four assets for less than 100 Eurocents and selling them via the unit portfolio for 100 Eurocents to the experimenter.

Arbitrage opportunity 2. If it is possible to sell the four assets separately in the market for more than 100 Eurocents, you can buy a full set of assets via the unit portfolio for 100 Eurocents and sell them in the market for a higher price afterwards.

Such arbitrage opportunities are not excluded by the trading system automatically. Therefore, you should carefully watch for arbitrage possibilities!

4. Payment/Incentives

In order to motivate you to thoroughly think about your trading activities in this experiment, you will be paid based on your decisions, the outcomes of the games and thus, your individual trading performance. Keep in mind that there is NO WAY FOR YOU TO LOSE MONEY.

At the end of this experiment we will randomly draw **one** out of the four teams from your NBA market and **one** out of the four teams from your FIFA market. These “team markets" then serve as the basis for your payment. Please note, that you will not know which of the teams in your markets will be relevant for your payment until the end of this study, so you should think carefully about your decisions in all "team markets"!

We will then add your cash account to your final portfolio value (with the assets worth either 100 Eurocents or nothing) for both of the drawn NBA and FIFA "team markets", and will pay you the resulting overall value in Eurocents (or US$ at the current exchange rate; today's exchange rate is about 1.20 $/€).

Example: If at the end of the experiment you have 500 Eurocents and three assets worth 100 Eurocents each in the selected NBA "team market" plus 1,000 Eurocents and five assets worth 100 Eurocents each in the selected FIFA country's "team market", you will receive a payment of (500 + 3* 100) + (1,000 + 5* 100) = 2,300 Eurocents (€23 ~ $27.60).

As you will initially be endowed with cash and unit portfolios totaling 1,000 Eurocents (€10 ~ $12) in each market you would earn 2* €10 = €20 ~ $24 if you do not trade at all.

Further details on the means of payment (check, PayPal etc.) will be provided later via email.

5. Trading System

5.1 Installation and Login

The trading system is based on Java technology, so you should have a recent Java version installed to run the trading software on your computer (Java 1.5, recently named Java5). If your Java version is outdated, you can get the latest version at:

http://www.java.com/en/

You can access your NBA resp. FIFA market via the study homepage at:

http://www.internetexperiment.de/NBA-FIFA-Study/

by following the link of your particular NBA or FIFA market number (we informed you of your personal market affiliation by email). Please note that on the first login the trading software (5.5 MB) has to be downloaded to your computer. This will take a while but is not necessary on later logins. You should accept the security certificate issued by the "Chair of Finance, University of Muenster". You will then see the "ECMS | Login" screen and will be asked for your "User:" ID (which we told you by email) and your personal "Password:" (which you chose after completing the first questionnaire). After clicking the "Login" button, the trading screen ("ECMS | internetexperiment.de (Uni Münster)") will appear.

Note: The following sections can be skipped if you have well-founded trading experience and feel comfortable with the continuous double auction trading mechanism!

5.2 Trading Screen

The trading screen is divided into five areas (see figure 1):

Information. In this area you can find information on your current cash balance, portfolio value, and credit line. At the top of this box are tabs which allow you to switch between the four different “team markets”. (The screenshots presented in this document refer to the practice market which can be accessed via the study homepage!)

Market. This area gives you an overview of the assets (and intervals) that are traded in this “team market”, your own portfolio, the current market price as well as the best (i.e. highest) buy and the best (i.e. lowest) sell offers (which is called the order book). The best buy offer (sell offer) indicates the volume and the price limit of the highest buy offer (lowest sell offer) at the current time, separated by an “@” symbol. To give you an example: “2 @ 30.00” means that one or more trader(s) are ready to buy a quantity of two assets “RUS.[0-5]”at a price of 30.00 Eurocents each.

Order Form. In this area you can enter your own orders.

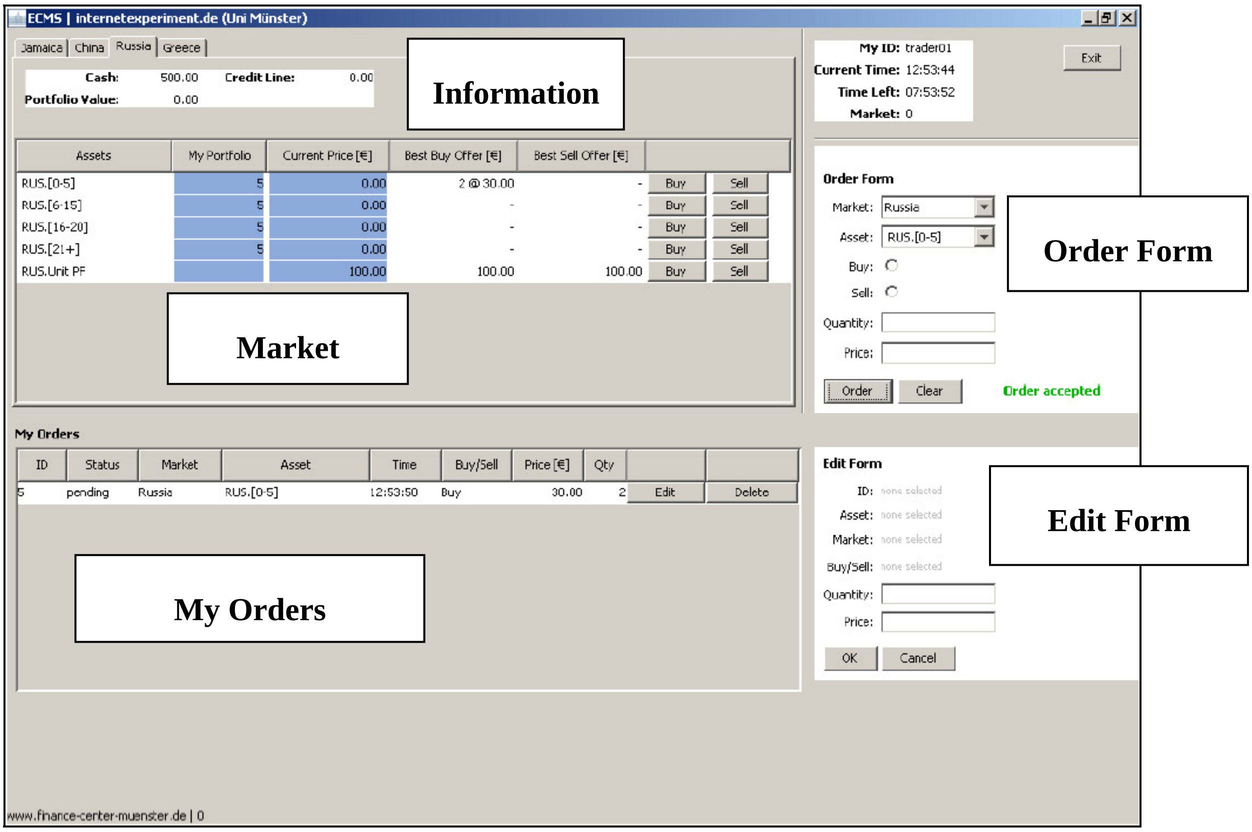


Fig. 1: Trading Screen.

My Orders. This area shows all of your orders which are pending or which were executed (currently there is one pending order in the system) in chronological order. In this area pending orders can be transmitted to the “Edit Form” on the right hand side or they can be deleted and thus, removed from the trading system.

Edit Form. Here, you can modify your pending orders by altering the price limit or the quantity.

5.3 Order Submission

In principle, there are two different ways to submit an order:

1. You can enter all the values (market, asset, buy/sell, quantity and price) in the “Order Form” on your own and click the “Order” button afterwards to submit the order.

2. You can push one of the buttons “Buy”/“Sell” in the “Market” section. The respective asset will be transmitted to the “Order Form” automatically; if available, the best sell offer (if you clicked on the “Buy” button) or the best buy offer (if you clicked on the “Sell” button) will be set as default price. To complete your order, just indicate the quantity you want and click the “Order” button. Of course, all default values can be modified manually.

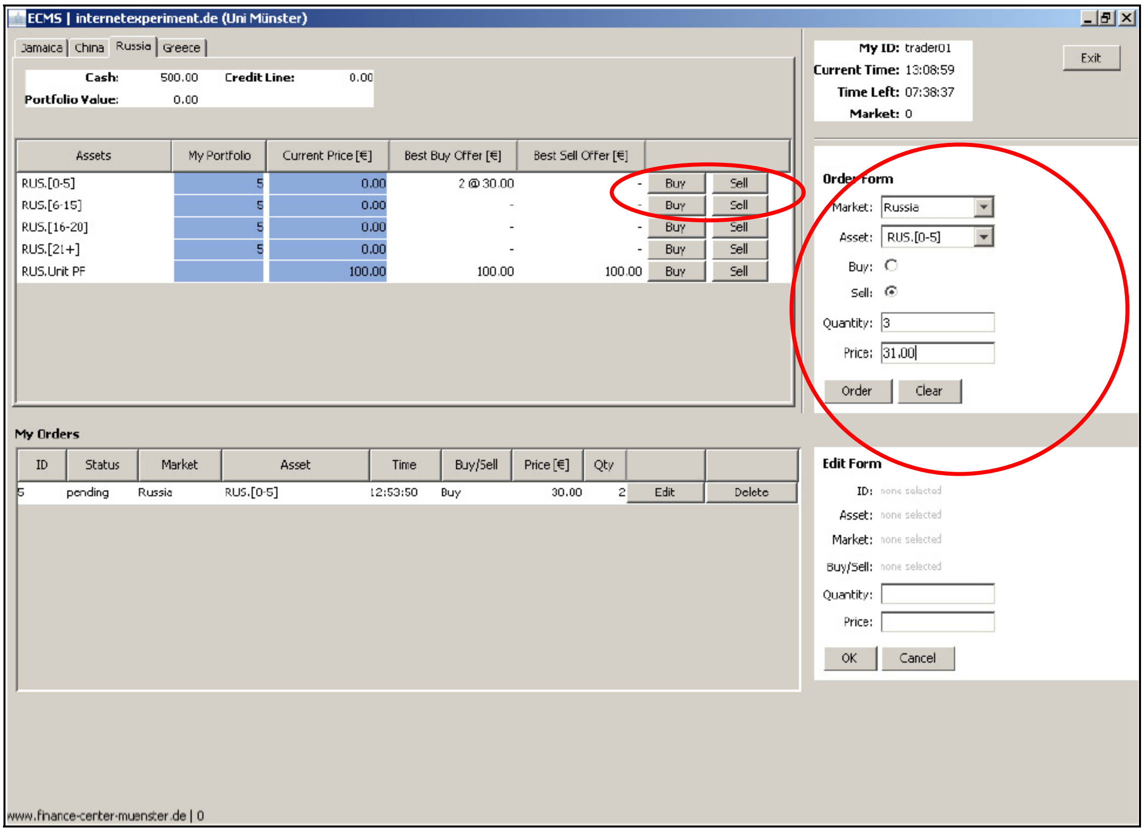


Fig. 2: Order Submission.

After submitting your order, it will be processed by the system. If your order is going to result in a trade, this can last for 1-2 seconds. Your trading screen will be blocked during this time.

In this context you should pay attention to the following two error messages:

Error "not enough cash" and error "short selling restriction"

These error messages will occur, if

1. either you do not have enough cash and/or if you already have too many pending buy orders ("not enough cash"), or if

2. you do not have enough assets and/or if you already have too many pending sell orders ("short selling restriction").

Please note that any pending order will tie-up your cash (buy order) or assets (sell order) until these orders are executed or deleted. This prevents you from placing orders which might not be possible due to a lack of cash/assets. Pending orders can be deleted or edited at any time in order to make the tied-up cash/assets available again.

From figure 2 above you can see that an amount of 2* 30.00 = 60.00 Eurocents is tied-up for the pending buy order at the moment. The disposable cash amount reduced to 500 - 60 = 440 Eurocents.

5.4 Order Execution

The trading system is based on a so-called "continuous double auction". This means that a buy and a sell offer will be executed if the respective buy order has at least the same price limit as the corresponding sell order. Please look at the following order book, which contains a list of all active buy and sell offers in the market:

Buy Offers	Sell Offers
	1 @ 60.00
	1 @ 56.00
	1 @ 50.00
- none -	

Fig. 3: Order book.

Assume you would like to buy assets and enter a buy order for "6 @ 80.00" into the trading system, i.e. you are willing to buy a quantity of six assets at a price limit of 80.00 Eurocents each. In general, only the best offers will appear in the order book of your "Market" section.

In this case, you know for sure that you will buy one asset at a price of 50.00 Eurocents in the market (best sell offer), as you can see from the trading screen.

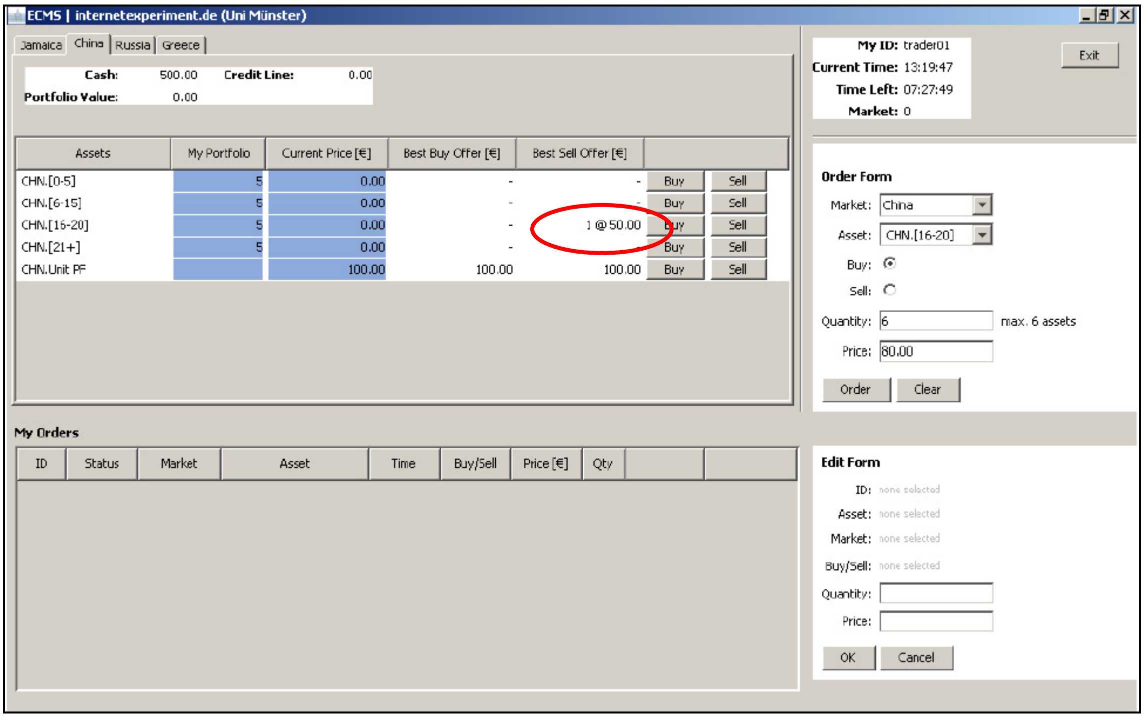


Fig. 4: Order book example.

The remaining part of the order book is hidden to you. However, since you are willing to pay up to 80.00 Eurocents per asset, you will buy a quantity of three assets in total (see Fig. 3).

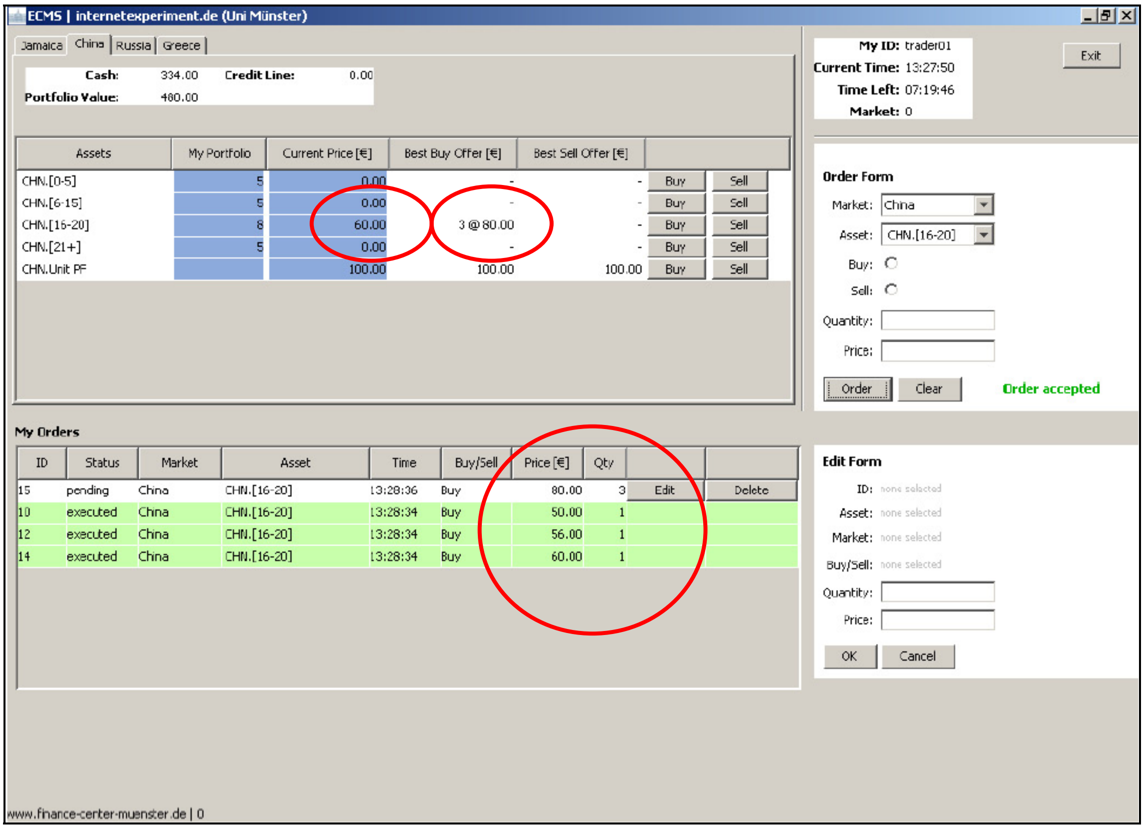


Fig. 5: Splitted Order.

As you can see from the “My Orders” section in Fig. 5, your original order of six assets was split, so that three orders, each consisting of one asset, were executed (highlighted in green colour), while an order of three assets remains in the trading system. The price which is shown in the column “Price” equals the trading price (for the executed orders) and equals your limit price (for the pending order).

The “Current Price” in the “Market” area indicates that the last trade was executed at a price of 60 Eurocents. The remaining part of your order (“3 @ 80.00”) appears in the order book as the best buy offer.

Please note: Whenever two orders can be matched, they will always be executed for the price limit of the buy or sell order which was submitted the earliest (time priority).

Note: For more detailed information on the trading system see the FAQs section on our study homepage!

6. NBA Basketball Playoffs Mode

Note: This section can be skipped if you are an NBA Playoffs expert and already know all the details!

Overall, in the National Basketball Association (NBA) there are 30 teams which are divided into Eastern and Western conference (15 teams each). Each conference consists of three divisions with five teams in every division. Within a so-called regular season, each team plays in total of 82 games (which are divided evenly between home and road games). During the regular season, each team faces opponents in its own division four times a season (4x4 = 16 games), teams from the other two divisions within the same conference three or four times a year (in total 36 games) and all 15 teams of the other conference twice (15x2 = 30 games).

Afterwards, the NBA uses a particular seeding system in order to set the Playoffs match-ups: The top three seeds (1-3) for each conference are determined by taking the winners of the conferences’ three divisions and ranking them by their regular season record (i.e., their relation between wins and losses). The remaining five seeds (4-8) for each conference are determined by taking the five best teams from the remaining pool of 12 teams.

The Playoffs themselves follow a tournament format: Each team plays a rival in a best-of-seven series, with the first team to win four games advancing to the next round. This means that two teams play against each other, until one of the teams wins four times. While the losing team is eliminated from the Playoffs and does not play any more games, the winning team advances to the next round and has the next match-up according to the following preset bracket:

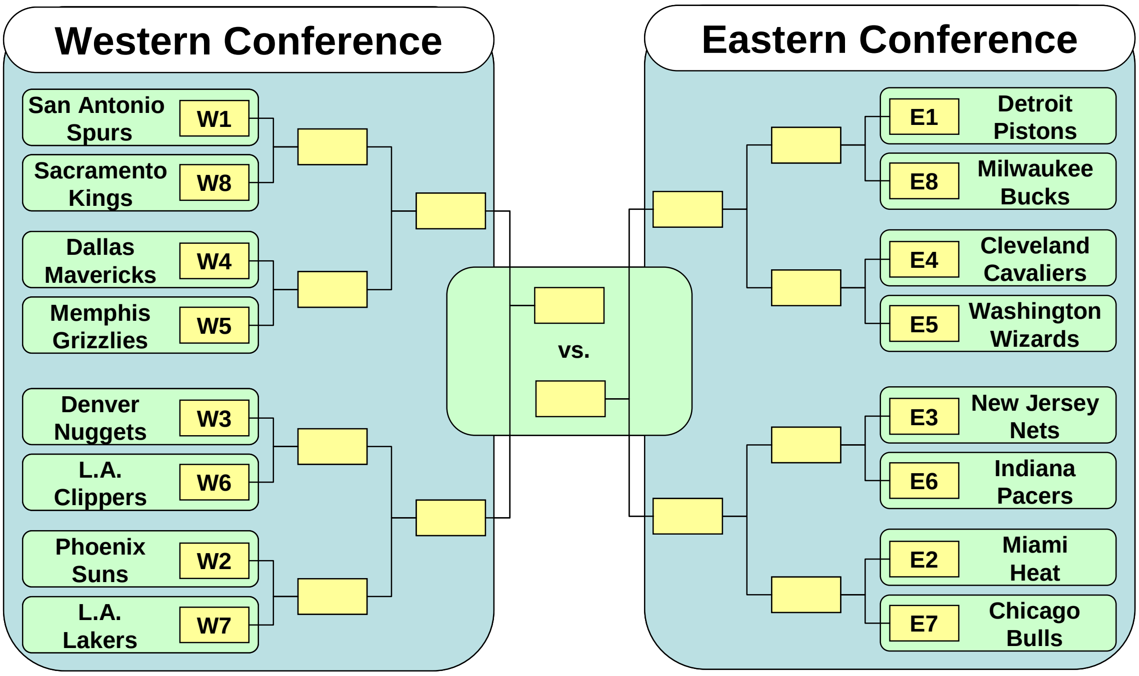


Fig. 6: Traditional NBA Playoffs Bracket with this year's teams.

In the final round (the NBA Finals), the winner of the Western conference faces the winner of the Eastern conference.

Home-court advantage in the NBA Playoffs follows a particular pattern and is based strictly on a team's regular-season record, without regard to whether a team won its division or not. In each round except the NBA Finals, the best-of-seven series follows a 2-2-1-1-1 pattern, meaning that the team with the better won-lost record has home court in games 1, 2, 5, and (the decisive game) 7, while the opponent plays at home in games 3, 4, and 6. In the NBA Finals, home-court advantage follows a 2-3-2 pattern, meaning that one team has home court in games 1, 2, 6, and 7, while the other team plays at home in games 3, 4, and 5.

Please note that the winning team in the NBA Finals always wins exactly 16 games in total; the team that makes it to the NBA Finals but loses, always wins 12 to 15 games etc., while the first-round losers win from 0 to 3 games.

7. FIFA Soccer World Cup Mode

Note: This section can be skipped if you are a FIFA World Cup expert and already know all the details!

After a two-year period of qualification, 32 national teams have qualified for the final tournament (FIFA Soccer World Cup 2006).

In December 2005, the qualified teams were drawn and assigned to eight groups of four nations each. Eight teams (Germany, England, Argentina, Mexico, Italy, Brazil, France and Spain) were seeded as group heads, based on their so-called FIFA ranking and their success in previous FIFA World Cups. The remaining teams were drawn randomly from the different continental zones (and by geographical criteria). They will now compete for the World Cup in two phases, a group phase and a knockout phase.

Group phase. From June 9 on, in all of the eight groups, each team plays every other team in the group once (pure round-robin schedule), guaranteeing that every team will play (at least) three matches. Every match consists of two periods of 45 minutes each and can end in a decision (three points for the winner) or a draw (one point for both teams). At the end of this phase, the two top teams from each group advance to the knockout phase.

Knockout phase. In this phase, teams play each other in single-elimination matches. If there is a draw at the end of the regular playing time (i.e. after the second period of 45 minutes has finished), there are two overtime periods, 15 minutes each. If there is still a draw after the 30 minutes of overtime, a penalty shootout is used to decide the winner. Please note that goals scored in a penalty shootout do not count for the (official FIFA) result of the game (unlike penalty-kicks during the game, which do count for the final score of the game).

In the round of 16 (1/8-finals), the winner of each group plays against the runner-up from another group. This is followed by the quarterfinals, the semifinals and the World Cup final, as can be seen from the following figure:

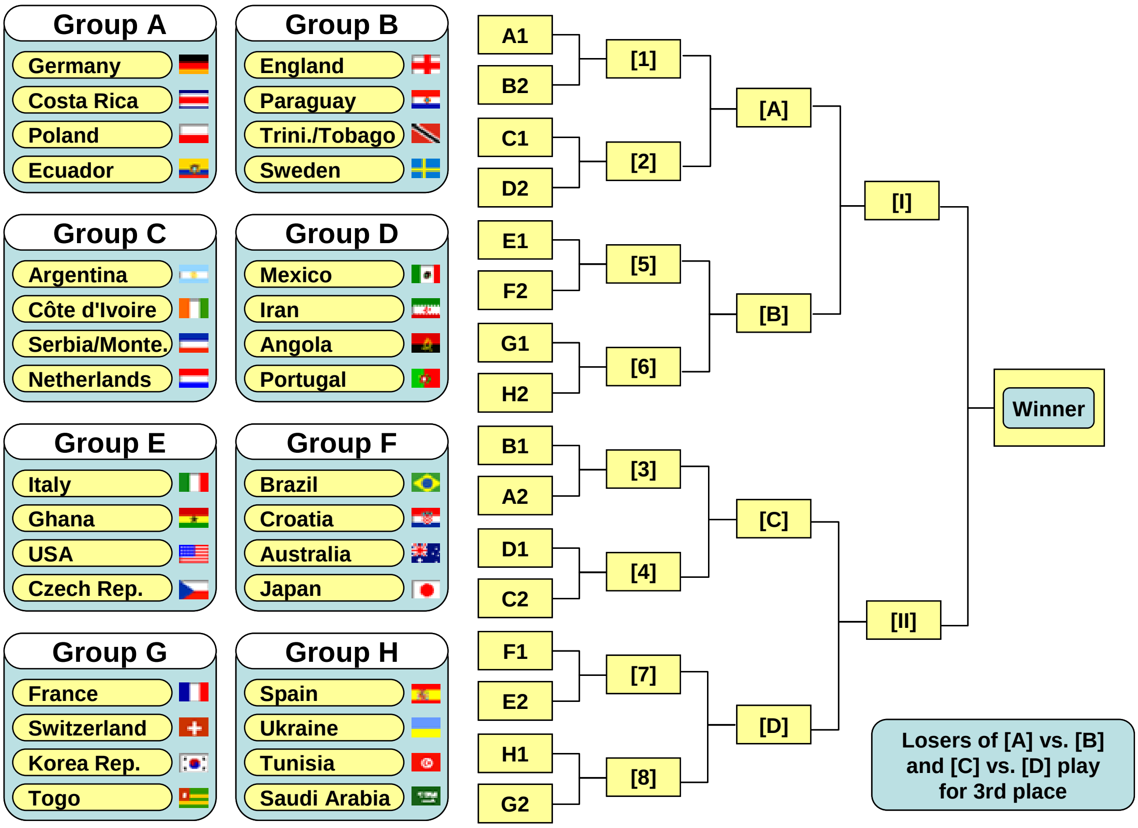


Fig. 7: FIFA Soccer World Cup Bracket.

The losers from the semifinals play each other in a third place match.

Appendix VI: NBA/FIFA study FAQs (Study homepage)

General FAQs

Why did you ask for my Social Security Number (SSN) in the registration form?

Since this is a paid experiment, we can't pay you without reporting tax info.

What is the purpose of this study?

The purpose of this study is to help social scientists better understand decision-making by observing your decisions.

How will I be paid?

At the end of this experiment we will randomly draw **one** out of the four teams from your NBA market and **one** out of the four teams from your FIFA market. These “team markets” then serve as the basis for your payment. Please note that you will not know which of the teams in your markets will be relevant for your payment until the end of this study, so you should carefully think about your decisions in all “team markets”!

We will then add your cash account to your final portfolio value (with the assets being worth either 100 Eurocents or nothing) for both of the drawn NBA and FIFA “team markets”, and will pay you the resulting overall value in Eurocents (or in US$ at the current exchange rate; today’s exchange rate is about 1.20 $/€).

When will I be paid?

You will be paid in the middle of July after all NBA and FIFA markets are closed. Further payment details will be sent to you via email.

Can I lose real money in this experiment?

No! There is no way for you to lose money. The worst possible outcome is for you to walk away with nothing.

Is this experiment about fantasy teams?

No! You are trading in contracts whose payoffs depend on the outcomes of the "real" 2006 NBA Playoffs and the "real" FIFA World Cup being held this year in Germany.

Assets and Markets FAQs

What kind of contracts do I trade in the NBA market?

You trade in contingent claims whose payoffs depend on how many games a particular team wins in the 2006 NBA Playoffs.

What is the basic idea of these markets?

The basic idea of these markets is that after the Playoffs (and the World Cup) are over, exactly one asset/claim in each "team market" will pay 100 Eurocents per claim, while the three remaining claims will become worthless. This is due to the fact that the asset intervals encompass all possible outcomes, but do not overlap. However, since the outcome of the event is uncertain during most of the trading period, the price for an asset usually varies between 0 and 100 Eurocents.

What is a unit portfolio (Unit PF)?

The Unit PF allows you to buy or sell a complete set of all existing assets for a particular team at a fixed price of 100 Eurocents in order to increase either liquidity in cash or stock in assets without taking any risks. This is an easy way to transform cash into assets (buy the unit portfolio) or to transform assets into cash (sell the unit portfolio).

What is meant by arbitrage opportunities?

Arbitrage means doing a transaction, which results in a sure profit. If, at any time, the best sell offers of all assets in a "team market" sum to less than 100 Eurocents, you can buy these assets in the market and sell them as a unit portfolio to the experimenter at a fixed rate of 100 Eurocents afterwards. As you bought the assets for less than 100 Eurocents, you realize a riskless profit.

If, at any time, the best buy offers of all assets in a "team market" sum to more than 100 Eurocents, you can buy a unit portfolio from the experimenter at a fixed rate of

100 Eurocents and sell the assets separately in the market afterwards. As you will receive more than 100 Eurocents, you realize an (almost) riskless profit.

You should hurry to exploit any arbitrage opportunities, since other market participants may try to act in the same direction!

What is an order book?

An order book is a list of all pending buy and sell orders that are active in the market. In this experiment, however, only the best (highest) buy offer and the best (lowest) sell offer will be disclosed to you in the “Market” section of your trading screen. If a buy or a sell offer from the order book is completely executed the next best buy or sell offer will appear on the trading screen, if they exist.

What is a price limit?

The prices you enter in the “Price” field of the “Order Form” or the “Edit Form” are always limit prices, meaning that you won’t buy any assets at a price HIGHER than your fixed price (limit) and you won’t sell any assets at a price LOWER than your fixed price (limit). However, orders may be executed even at a “better” price than your price limit, please see FAQ At what trading price will orders be executed? for further details.

What will happen to my pending orders while a game takes place?

Nothing, i.e. they remain pending in the order book at their original limit prices. Thus, you should think very carefully about which orders you wish to remain pending and which orders should be deleted before a game, because some claims may become worthless immediately following the game.

Please note that trading will not be suspended during the games, thus your pending orders might be executed at their original limit prices even if they will be worthless within minutes!

What is the trading period?

In general, the trading period for the NBA markets lasts from April 20 through June 22, 2006 (at the latest) and the trading period for the FIFA markets lasts from the middle of May through July 9, 2006. However, some markets can expire earlier, as you will be trading in four NBA teams and four FIFA countries’ teams respectively. Thus, as teams drop out of the tournament, everyone will know which assets are worthless and which will pay out the 100 Eurocents for sure, so those respective “team markets” will be closed.

Unless a market expires as described above, they will be open for trading 24/7, i.e. 24 hours a day, 7 days a week.

Can I use my cash from “team market” A to buy assets in “team market” B, C or D?

No! You can’t transfer money from one “team market” to another in order to buy more assets in that market, since your four “team markets” are treated separately and independently.

What is this “practice market” useful for?

The practice market is a market where you can get familiar with the trading system before entering the markets for real money. We strongly recommend you to enter the practice market AT LEAST ONCE before accessing your real NBA market. You can access the practice market with your User ID and your password (after you fill out the questionnaire). In a similar style to the real markets in this experiment, the practice market allows you to trade contracts based on how many gold medals four different nations will win in the next Olympic Summer Games.

Please try to submit, edit and delete at least three buy and sell orders. Also try to buy and sell the unit portfolio at least once. Watch for the changes in your cash account and your portfolio! The actions taken in this practice market will not at all influence your payments at the end of the experiment. Just feel free to try the functions and features of the trading software and learn how the market system works!

What happens if a game is cancelled or has to be stopped before the official time or ….?

In case of superior forces or unforeseeable events (team disqualification, doping etc.) we will always refer to the official NBA or FIFA results when fixing the payoffs for the assets.

Trading System FAQs

Why doesn't my trading software start when I click the market link?

Probably, you do not have a recent Java version installed. Java 1.5, recently named Java5 is necessary to run the trading software on your computer. If your Java version is outdated, you can get the latest version at: http://www.java.com/en/

If I make a mistake trading, can I take my order back?

As long as your order is still pending, you can edit or delete it to alter the price and/or quantity or totally remove it from the trading system. However, if your order was executed immediately after your submission (because it matched another pending order), there is no way of cancelling your order. So, please check your entries carefully before submitting an order to the system!

How can I edit or delete pending orders in the trading system?

Orders which have not been executed yet can be edited or deleted from the order book at any time. You just need to select an active order, which is marked "pending" in the "Status" column within the "My Orders" area and press one of the related buttons "Edit" or "Delete" in the respective row.

By editing a pending order your initial order will (temporarily) be removed from the order book (and from the list "My Orders") and the order details will be transmitted to the "Edit Form". In this form, you can modify your order by altering the quantity and/or the price limit. Afterwards, you can resubmit your order by clicking "OK" (it will then be treated as a new order with respect to its time priority). By clicking the "Cancel" button you will be asked whether you want to delete or to restore the edited order.

By pushing the "Delete" button in the "My Orders" section the respective order will be deleted immediately!

What is meant by the Error: "not enough cash" in the trading system, when I try to submit a buy order?

The error message "not enough cash" will occur if you do **not** have **enough cash** to execute the intended buy order and/or if you already have **too many pending buy orders** in the system.

Please note that any pending buy order will tie-up that amount of cash (limit price x quantity) until the order is executed or deleted. This prevents you from submitting orders for which you might not have the cash on hand to cover. Pending orders can be deleted or edited at any time in order to free-up some cash. (As pending AND executed orders are recorded chronologically in the same list ("My Orders"), please keep in mind that there might be pending buy orders that are down the list.)

What is meant by the Error: "short selling restriction" in the trading system when I try to submit a sell order?

The error message "short selling restriction" will occur, if you do **not** have **enough assets** to sell and/or if you already have **too many pending sell orders** of that asset in the system.

Please note that any pending sell order will tie-up that number of assets (quantity) until the order is executed or deleted. This prevents you from submitting orders for which you might not the assets on hand to cover. Pending orders can be deleted or edited at any time in order to free-up some assets. (As pending AND executed orders are recorded chronologically in the same list ("My Orders"), please keep in mind that there might be pending sell orders that are down the list.)

What is meant by the Error: "do not trade with yourself" in the trading system?

The error message "do not trade with yourself" will occur, if you try to place a buy and a sell order for the same asset at the same time, where the limit price of your buy order is EQUAL or HIGHER than the limit price of your sell order. This means that you are willing to PAY the same price or more for the asset than you are willing to RECEIVE for the same asset, which is inconsistent.

However, you can have buy orders and sell orders for the same asset at the same time in order to act as a "market maker," if the limit price of your buy offer is LOWER than the limit price of your sell offer!

How is the "Portfolio Value" in the trading system calculated?

The portfolio value is calculated by taking the number of assets multiplied by their current market prices, and then summed across all the assets of a particular "team market". Please note that at the end of the experiment there will be one asset worth 100 Eurocents and three assets worth 0 Eurocents, in each "team market".

What is meant by "Credit Line" in the trading system?

Don't worry about the credit line, as it is zero for all participants in this experiment and simply means that you are not allowed to purchase assets on credit.

What is the "Current Price" in the trading system?

The current price gives you the last trading price for that asset.

At what trading price will orders be executed?

In general, the trading mechanism follows a so-called "continuous double auction" mechanism. This means that a buy and a sell offer will only be executed if the respective buy order has AT LEAST THE SAME price limit as the corresponding sell order.

Please note that whenever two orders can be matched, they will always be executed at the price limit of the order that was submitted the earliest (time priority).

Appendix VII: NBA/FIFA study questionnaires

Questionnaire data elicited before the NBA Playoffs part of the study:

ECMS | internetexperiment.de

Questionnaire

Dear Ulrich Sonnemann,

As announced in our first email a few days ago, we would like to ask you to complete a short questionnaire before you can start trading. **Note that all collected personal information will only be used in anonymous form for research purposes and will not be transmitted to third parties.** At the end of this questionnaire you can choose your **password** for your NBA market, which will, in connection with your **User ID**, allow you to access the trading system. We would like to reiterate that you need not to be a basketball or a soccer expert in order to successfully participate in this experiment!

So, let us first ask you some general questions:

1. What is your birth-year?

   (yyyy) ☐

2. Please self-rate your general trading experience (e.g. stock market, bond market etc.)!

   (low) ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 (high)

3. Please self-rate your knowledge and skills in the field of statistics!

   (low) ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 (high)

4. Have you ever been active in sports betting?

   ☐ Yes ☐ No

In the first part of this experiment you will have the opportunity to trade in various contracts whose payoffs depend on the **total number of wins** for a particular team in the 2006 NBA Playoffs. A more detailed description will be provided later on (in the study instructions). In order to give you an idea of which factors may affect the number of wins for a particular team, we will briefly explain the mode of the NBA Playoffs:

Open window for NBA Playoffs description...
(Note: You do not have to read the NBA Playoffs description if you are an expert and already know all the details!)

Against this background, let us now ask you some questions concerning NBA Basketball:

5. In general, how competent do you feel in making judgments regarding the NBA Playoffs?

   (little) 1 2 3 4 5 6 7 (highly)

6. In general, how much are you interested in the NBA Playoffs?

   (little) 1 2 3 4 5 6 7 (highly)

7. How closely/intensely are you going to follow the 2006 NBA Playoffs (on TV, via Internet etc.)?

   (not at all) 1 2 3 4 5 6 7 (very intensely)

8. If any, which is your favorite NBA team?

9. Which of the following four NBA teams do you like the best?

   - Dallas
   - L.A. Lakers
   - Miami
   - Washington
   - none

10. In the following, you will be asked seven multiple choice trivia questions concerning NBA Basketball. For each question, there are five choices, but only one answer is correct! **Note:** We ask you these questions for research purposes only. Your answers will neither affect your payment nor any other aspect of the experiment! We ask you to answer honestly, on your own, and without cheating! Please, do not get discouraged if you have any problems answering these questions!

a. Which team lost the NBA Finals last season (2004/05)?

   - Detroit Pistons
   - Washington Wizards
   - Philadelphia 76ers
   - Los Angeles Clippers

- Denver Nuggets
- I don't know

b. Which team has won the most championships in NBA history?

- New York Knicks
- Boston Celtics
- Utah Jazz
- Phoenix Suns
- Portland Trail Blazers
- I don't know

c. How many seconds does a team have to take a shot?

- 10
- 16
- 24
- 35
- 60
- I don't know

d. Which team did Michael Jordan play for until his second retirement?

- Boston Celtics
- Chicago Bulls
- New York Knicks
- Los Angeles Lakers
- Detroit Pistons
- I don't know

e. What was the longest streak of consecutive NBA championships by one team in NBA history?

- 4
- 5
- 6
- 7
- 8

- I don't know

f. Over the last three seasons, how many games did the Detroit Pistons win on average (2002/03-2004/05)?

- 33
- 43
- 53
- 63
- 73
- I don't know

g. Over the last three seasons, how many of the 45 playoff match-ups ended in a decisive game 7?

- 3
- 12
- 21
- 30
- 39
- I don't know

10. In the following, you will be asked seven multiple choice trivia questions concerning NBA Basketball. For each question, there are five choices, but only one answer is correct! **Note:** We ask you these questions for research purposes only. Your answers will neither affect your payment nor any other aspect of the experiment! We ask you to answer honestly, on your own, and without cheating! Please, do not get discouraged if you have any problems answering these questions!

a. Which team won the NBA title last season (2004/05)?

- Utah Jazz
- Los Angeles Lakers
- San Antonio Spurs
- Chicago Bulls
- Orlando Magic
- I don't know

b. Which of the following teams has not won the NBA Finals in the last 10 years?

- Indiana Pacers
- San Antonio Spurs
- Detroit Pistons
- Los Angeles Lakers
- Chicago Bulls
- I don't know

c. After how many fouls is a player ejected?

- 3
- 4
- 5
- 6
- 7
- I don't know

d. For which team has Dirk Nowitzki played since the 1998/99 season?

- Boston Celtics
- Los Angeles Lakers
- Seattle Supersonics
- Dallas Mavericks
- Cleveland Cavaliers
- I don't know

e. What was the jersey number worn by Michael Jordan in all (except one) his NBA games before his first retirement?

- 3
- 13
- 18
- 23
- 34
- I don't know

f. Over the last five seasons, what is the average number of games won by the team with the best regular-season record?

- 50
- 55
- 60
- 65
- 70
- I don't know

g. Over the last three seasons, how many of the 24 first-round playoff match-ups have been sweeps (4-0)?

- 2
- 5
- 8
- 11
- 14
- I don't know

11. Lastly, we will ask you for your **individual probability judgment** of the **total number of wins** for the following four teams in this year's NBA Playoffs. The given range in brackets corresponds to the range of wins of the particular team. In other words: What do you think is the probability that team XYZ will have 0 to 3 wins in this year's Playoffs and so on. **Note:** Your probability judgments must sum to 100% for each team!

Open window for NBA Playoffs description...

1. Dallas	[0-3]: %	[4-7]: %	[8-11]: %	[12-16]: %
2. L.A. Lakers	[0-3]: %	[4-7]: %	[8-11]: %	[12-16]: %
3. Miami	[0-3]: %	[4-7]: %	[8-11]: %	[12-16]: %
4. Washington	[0-3]: %	[4-7]: %	[8-11]: %	[12-16]: %

Next...

Questionnaire data elicited before the FIFA World Cup part of the study:

ECMS | internetexperiment.de

Questionnaire

Dear Ulrich Sonnemann,

Before trading in the FIFA World Cup markets begins, we would like to ask you to fill out another short questionnaire. **Note that all collected personal information will only be used in anonymous form for research purposes and will not be transmitted to third parties.** We would like to reiterate that you need not be a basketball or a soccer expert in order to successfully participate in this experiment!

In the second part of this experiment you will have the opportunity to trade in various contracts whose payoffs depend on the **total number of goals scored by a particular national team in the FIFA Soccer World Cup 2006**. A more detailed description will be provided later on (in the second part of the study instructions). In order to give you an idea of which factors may affect the number of goals for a particular team, we will briefly explain the mode of the FIFA World Cup:

Open window for FIFA World Cup mode...
(Note: You don't have to read the FIFA World Cup mode description if you are an expert and already know all the details!)

Against this background, let us now ask you some questions concerning the FIFA World Cup:

1. In general, how competent do you feel in making judgments regarding the FIFA Soccer World Cup 2006?

   (little) ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 (highly)

2. In general, how much are you interested in the FIFA World Cup 2006?

   (little) ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 (highly)

3. How closely/intensely are you going to follow the FIFA World Cup 2006 (on TV, via Internet etc.)?

   (not at all) ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 (very intensely)

4. If any, which is your favourite national team?

5. Which of the following four soccer World Cup teams do you like the best?

- Czech Rep.
- Ghana
- Italy
- USA
- none

6. In the following, you will be asked seven multiple choice trivia questions concerning the Soccer World Cup. For each question, there are five choices, but only one answer is correct! **Note:** We ask you these questions for research purposes only. Your answers will neither affect your compensation nor any other aspect of the experiment! We ask you to answer honestly, on your own, and without cheating! Please, don't get discouraged if you have any problems answering these questions!

a. Which country/countries was/were the host of the last FIFA World Cup in 2002?

- South Korea/Japan
- Austria/Switzerland
- Belgium/Netherlands
- South Africa
- France
- I don't know

b. Which country has won the most World Cup championships?

- Argentina
- Brazil
- Germany
- Italy
- England
- I don't know

c. How many times has the host country of the World Cup tournament won the championship in its own country (17 World Cups so far)?

- 0
- 3
- 6
- 9
- 12

- I don't know

d. How many substitutions does FIFA allow a team per game?

- 1
- 3
- 5
- 7
- unlimited
- I don't know

e. Who was the top scorer in the FIFA World Cup in 2002?

- David Beckham
- Ronaldo
- Raúl González
- Henrik Larsson
- Robbie Keane
- I don't know

f. Over the last 3 FIFA World Cups (1994, 1998, and 2002), what is the **average** number of goals (ex penalty shootouts) scored by all the teams combined, during the whole tournament?

- 88
- 98
- 118
- 158
- 238
- I don't know

g. In how many of the 64 games of the FIFA World Cup in 2002 did the winning team score four goals or more ([4+])?

- 1
- 4
- 8
- 13
- 19

- ☐ I don't know.

6. In the following, you will be asked seven multiple choice trivia questions concerning the Soccer World Cup. For each question, there are five choices, but only one answer is correct! **Note:** We ask you these questions for research purposes only. Your answers will neither affect your compensation nor any other aspect of the experiment! We ask you to answer honestly, on your own, and without cheating! Please, don't get discouraged if you have any problems answering these questions!

a. Which team won the FIFA World Cup in 2002?

- ☐ Brazil
- ☐ Netherlands
- ☐ Italy
- ☐ France
- ☐ England
- ☐ I don't know

b. Which of the following countries has **not** won the FIFA World Cup within the past 20 years?

- ☐ Brazil
- ☐ Spain
- ☐ Germany
- ☐ France
- ☐ Argentina
- ☐ I don't know

c. Which national team has never won the World Cup **in its own country**?

- ☐ Germany
- ☐ England
- ☐ Uruguay
- ☐ Brazil
- ☐ France
- ☐ I don't know

d. Within a single game, how many yellow cards does it take before a player is kicked out of the game by a red card?

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ I don't know

e. Which country does Zinedine Zidane play for?

- ☐ Italy
- ☐ Croatia
- ☐ Spain
- ☐ Portugal
- ☐ France
- ☐ I don't know

f. Over the last 3 FIFA World Cups (1994, 1998, and 2002), what is the **average** number of goals (ex penalty shootouts) scored by the world champion in the entire tournament?

- ☐ 7
- ☐ 11
- ☐ 15
- ☐ 19
- ☐ 23
- ☐ I don't know

g. In how many of the 64 games of the last FIFA World Cup in 2002 did at least one of the two opponents not score a single goal?

- ☐ 6
- ☐ 14
- ☐ 22
- ☐ 34
- ☐ 46
- ☐ I don't know

7. Lastly, we will ask you for your **individual probability judgment** of the **total number of goals (ex penalty shootouts)** for the following four teams in the

whole FIFA World Cup tournament. The given range in brackets corresponds to the range of goals scored by the particular team. In other words: What do you think is the probability that team A will score 0 to 2 goals in this year's World Cup and so on. **Note:** Your probability judgments must sum to 100% for each team!

1. Czech Rep.	[0-2]: ___ %	[3-5]: ___ %	[6-8]: ___ %	[9+]: ___ %
2. Ghana	[0-2]: ___ %	[3-5]: ___ %	[6-8]: ___ %	[9+]: ___ %
3. Italy	[0-2]: ___ %	[3-5]: ___ %	[6-8]: ___ %	[9+]: ___ %
4. USA	[0-2]: ___ %	[3-5]: ___ %	[6-8]: ___ %	[9+]: ___ %

Next...

Appendix VIII: Overall number of trades for each Playoffs and World Cup team market

Table A.VIII.1: Overall number of trades (ex unit portfolio) for each NBA Playoffs team market.

Team	Market Nr.	Subjects	Treatment (Partition)	Intervals							Total
				I_0	$I_1 \cup I_2$	I_3	I_4	I_1	I_2	$I_3 \cup I_4$	
Number of trades											
CHI	3	German	1	10	15	13	8				**46**
CHI	4	German	2	16				32	18	22	**88**
CHI	11	German	1	14	9	6	5				34
CHI	12	German	2	12				11	13	17	53
CHI	19	U.S.	1	6		1					7
CHI	20	U.S.	2	3				4	2	2	11
CLE	3	German	1	12	7	13	1				33
CLE	4	German	2	36				35	28	16	**115**
CLE	11	German	1	14	13	11	2				**40**
CLE	12	German	2	21				22	20	14	77
CLE	19	U.S.	1	1	2	2					5
CLE	20	U.S.	2	7				4	1		12
DAL	7	German	1	4	33	24	16				77
DAL	8	German	2	6				39	42	42	**129**
DAL	15	German	1	5	46	39	29				**119**
DAL	16	German	2	14				23	16	15	68
DAL	23	U.S.	1			1					1
DAL	24	U.S.	2					2		1	3
DEN	5	German	1	22	21	12	6				**61**
DEN	6	German	2	10				11	9	6	**36**
DEN	13	German	1	10	12	7	4				33
DEN	14	German	2	6				10	5	2	23
DEN	21	U.S.	1	1	6						7
DEN	22	U.S.	2	4							4
DET	5	German	1	10	19	24	27				**80**
DET	6	German	2	3				8	29	16	**56**
DET	13	German	1	3	23	16	7				49
DET	14	German	2	4				13	14	19	50
DET	21	U.S.	1	3	4	9	3				19
DET	22	U.S.	2	1				1		3	5
IND	5	German	1	26	17	9	6				**58**
IND	6	German	2	8				7	10		**25**
IND	13	German	1	9	15	6	3				33
IND	14	German	2	8				3	7	6	24
IND	21	U.S.	1	1	2	9	1				13
IND	22	U.S.	2								0
LAC	1	German	1	6	8	10	3				27
LAC	2	German	2	15				20	29	16	80
LAC	9	German	1	21	14	20	11				**66**
LAC	10	German	2	29				38	32	29	**128**
LAC	17	U.S.	1				1				1
LAC	18	U.S.	2	3					1		4
LAL	7	German	1	22	19	20	15				**76**
LAL	8	German	2	17				23	23	6	**69**
LAL	15	German	1	21	24	11	11				67
LAL	16	German	2	17				12	11	15	55
LAL	23	U.S.	1								0
LAL	24	U.S.	2	4				2	4		10

to be continued on the next page

Table A.VIII.1 continued

Team	Market Nr.	Subjects	Treatment (Partition)	Intervals							Total
				I_0	$I_1 \cup I_2$	I_3	I_4	I_1	I_2	$I_3 \cup I_4$	
Number of trades											
MEM	3	German	1	12	8	17	2				**39**
MEM	4	German	2	15				19	19	24	**77**
MEM	11	German	1	12	13	5	3				33
MEM	12	German	2	10				13	7	10	40
MEM	19	U.S.	1	2							2
MEM	20	U.S.	2								0
MIA	7	German	1	11	12	13	11				47
MIA	8	German	2	9				23	30	39	**101**
MIA	15	German	1	12	38	34	18				**102**
MIA	16	German	2	10				18	18	14	60
MIA	23	U.S.	1								0
MIA	24	U.S.	2					1	5	1	7
MIL	1	German	1	16	5	3	1				25
MIL	2	German	2	21				14	13	5	53
MIL	9	German	1	17	19	14	16				**66**
MIL	10	German	2	24				15	18	18	**75**
MIL	17	U.S.	1		4						4
MIL	18	U.S.	2	1							1
NJN	1	German	1	14	7	13	3				37
NJN	2	German	2	14				35	16	6	71
NJN	9	German	1	11	11	15	10				**47**
NJN	10	German	2	16				28	19	14	**77**
NJN	17	U.S.	1								0
NJN	18	U.S.	2	1				1	1	2	5
PHX	3	German	1	10	16	8	7				41
PHX	4	German	2	29				41	56	58	**184**
PHX	11	German	1	13	17	12	11				**53**
PHX	12	German	2	19				20	18	18	75
PHX	19	U.S.	1	1	3	1	1				6
PHX	20	U.S.	2	3				6	10	6	25
SAC	5	German	1	20	14	7	7				**48**
SAC	6	German	2	5				13	3	1	22
SAC	13	German	1	12	12	9	5				38
SAC	14	German	2	11				11	7	6	**35**
SAC	21	U.S.	1	1	3	1	1				6
SAC	22	U.S.	2							2	2
SAS	1	German	1	10	14	12	16				52
SAS	2	German	2	14				40	22	20	96
SAS	9	German	1	20	25	16	12				**73**
SAS	10	German	2	17				44	34	35	**130**
SAS	17	U.S.	1	1		1					2
SAS	18	U.S.	2					1	1	3	5
WAS	7	German	1	15	14	10	10				49
WAS	8	German	2	19				17	9	8	**53**
WAS	15	German	1	22	17	8	8				**55**
WAS	16	German	2	11				26	6	6	49
WAS	23	U.S.	1		1						1
WAS	24	U.S.	2	1				4			5
Total				**947**	**562**	**462**	**301**	**710**	**626**	**543**	**4,151**

Table A.VIII.2: Overall number of trades (ex unit portfolio) for each World Cup team market.

Team	Market Nr.	Subjects	Treatment (Partition)	Intervals							Total
				I_0	$I_1 \cup I_2$	I_3	I_4	I_1	I_2	$I_3 \cup I_4$	
Number of trades											
ARG	31	German	1		2	4	7				13
ARG	32	German	1	6	15	8	9				**38**
ARG	33	German	2	14				8	16	11	49
ARG	34	German	2	8				15	12	14	**49**
AUS	39	German	1	7	2	2					11
AUS	40	German	1	10	13	13	11				**47**
AUS	41	German	2	14				17	14	13	**58**
AUS	42	German	2	7				11	12	8	38
BRA	39	German	1		3	9	4				16
BRA	40	German	1	7	14	10	19				**50**
BRA	41	German	2	11				6	16	13	**46**
BRA	42	German	2	14				7	12	7	40
CIV	31	German	1	1	2	3					6
CIV	32	German	1	16	10	5	5				**36**
CIV	33	German	2	19				9	6	5	39
CIV	34	German	2	16				9	16	5	**46**
CRC	27	German	1	7	3	2	1				13
CRC	28	German	1	22	9	5	3				**39**
CRC	29	German	2	5				5	2	1	13
CRC	30	German	2	19				7	8	11	**45**
CRO	39	German	1	5	9	5	2				21
CRO	40	German	1	9	11	16	8				**44**
CRO	41	German	2	10				17	15	15	**57**
CRO	42	German	2	11				13	8	5	37
CZE	35	German	1	4	10	13	5				32
CZE	36	German	1	5	12	11	10				**38**
CZE	37	German	2	1				16	10	12	**39**
CZE	38	German	2					2	9	8	19
ECU	27	German	1	7	3	9	4				23
ECU	28	German	1	20	2	6	5				**33**
ECU	29	German	2	4				7	8	4	23
ECU	30	German	2	7				18	8	12	**45**

to be continued on the next page

Table A.VIII.2 continued

Team	Market Nr.	Subjects	Treatment (Partition)	Intervals							Total
				I_0	$I_1 \cup I_2$	I_3	I_4	I_1	I_2	$I_3 \cup I_4$	
Number of trades											
GER	27	German	1	2	4	7	12				25
GER	28	German	1	9	15	16	17				**57**
GER	29	German	2					2	15	5	22
GER	30	German	2	4				13	32	10	**59**
GHA	35	German	1	10	9	7	3				29
GHA	36	German	1	14	14	9	6				**43**
GHA	37	German	2	20				8	7	1	**36**
GHA	38	German	2	4					3	1	8
ITA	35	German	1	4	16	12	12				44
ITA	36	German	1	5	19	23	18				**65**
ITA	37	German	2	5				19	20	21	**65**
ITA	38	German	2	1				2	9	2	14
JPN	39	German	1	6	6						12
JPN	40	German	1	10	9	6	5				**30**
JPN	41	German	2	14				18	20	2	**54**
JPN	42	German	2	9				12	13	10	44
NED	31	German	1			2	3				5
NED	32	German	1	9	13	15	5				**42**
NED	33	German	2	17				12	13	14	**56**
NED	34	German	2	9				19	8	12	48
POL	27	German	1	3	6	4	1				14
POL	28	German	1	13	10	6	6				**35**
POL	29	German	2	8				9	4	2	23
POL	30	German	2	20				13	15	11	**59**
SCG	31	German	1	1	4	1					6
SCG	32	German	1	11	8	17	4				**40**
SCG	33	German	2	7				9	7	7	30
SCG	34	German	2	7				12	12	5	**36**
USA	35	German	1	9	18	3	1				31
USA	36	German	1	23	19	5	5				**52**
USA	37	German	2	10				9	10	1	**30**
USA	38	German	2	3				9	1		13
Total				**553**	**290**	**254**	**191**	**333**	**361**	**248**	**2,230**

Appendix IX: Price charts in the NBA/FIFA study

(see next page)

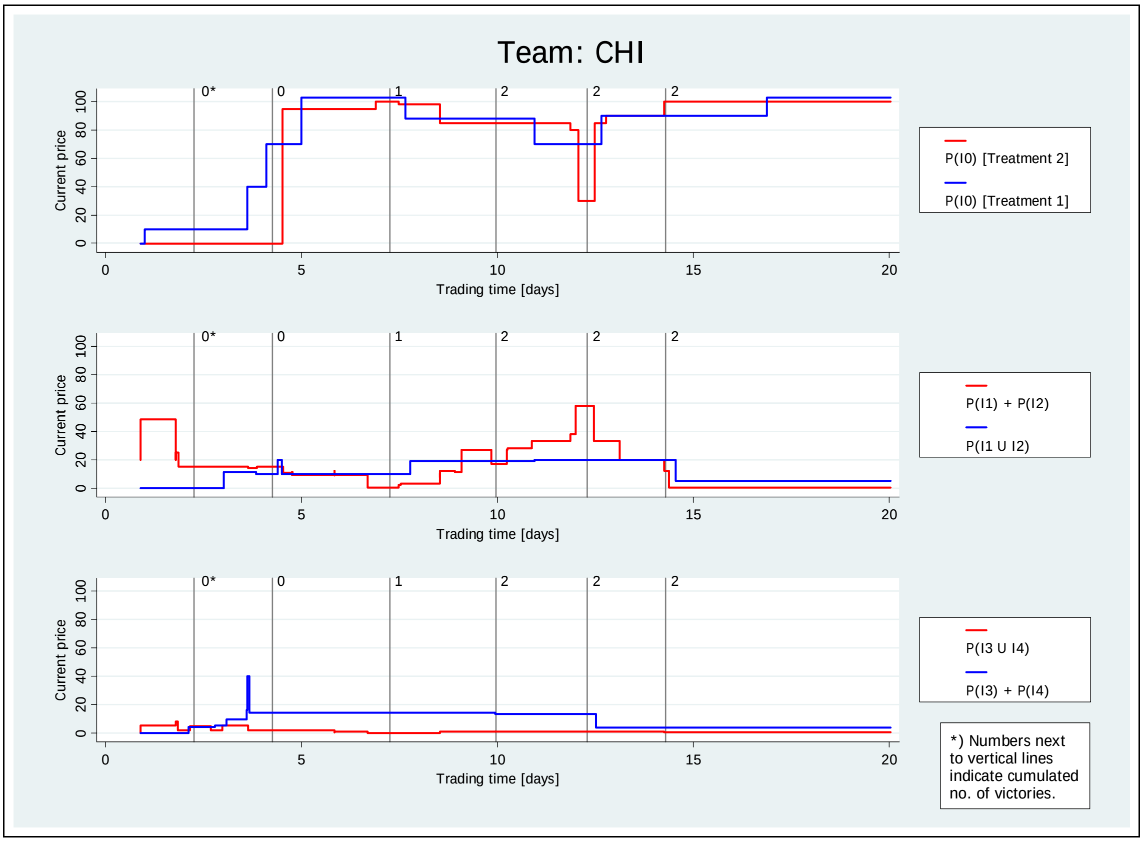


Figure A.IX.1: Price chart (Chicago Bulls, CHI).

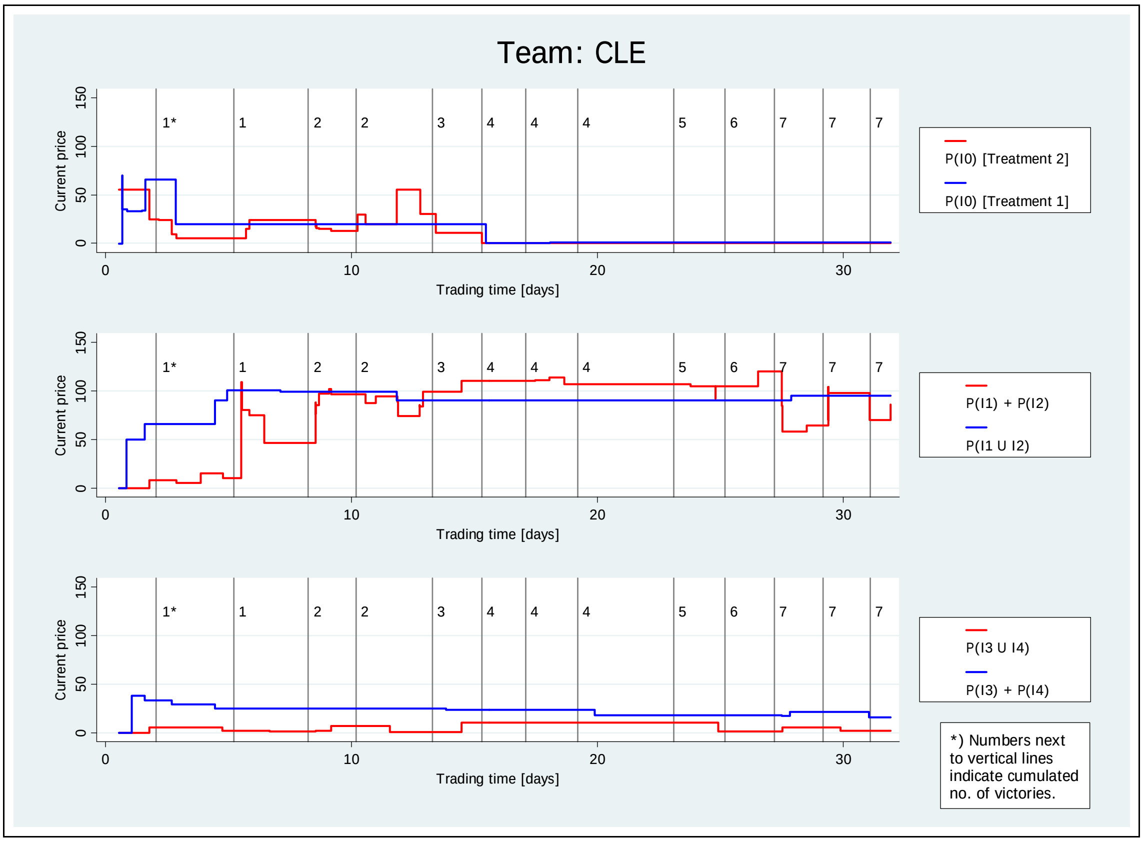


Figure A.IX.2: Price chart (Cleveland Cavaliers, CLE).

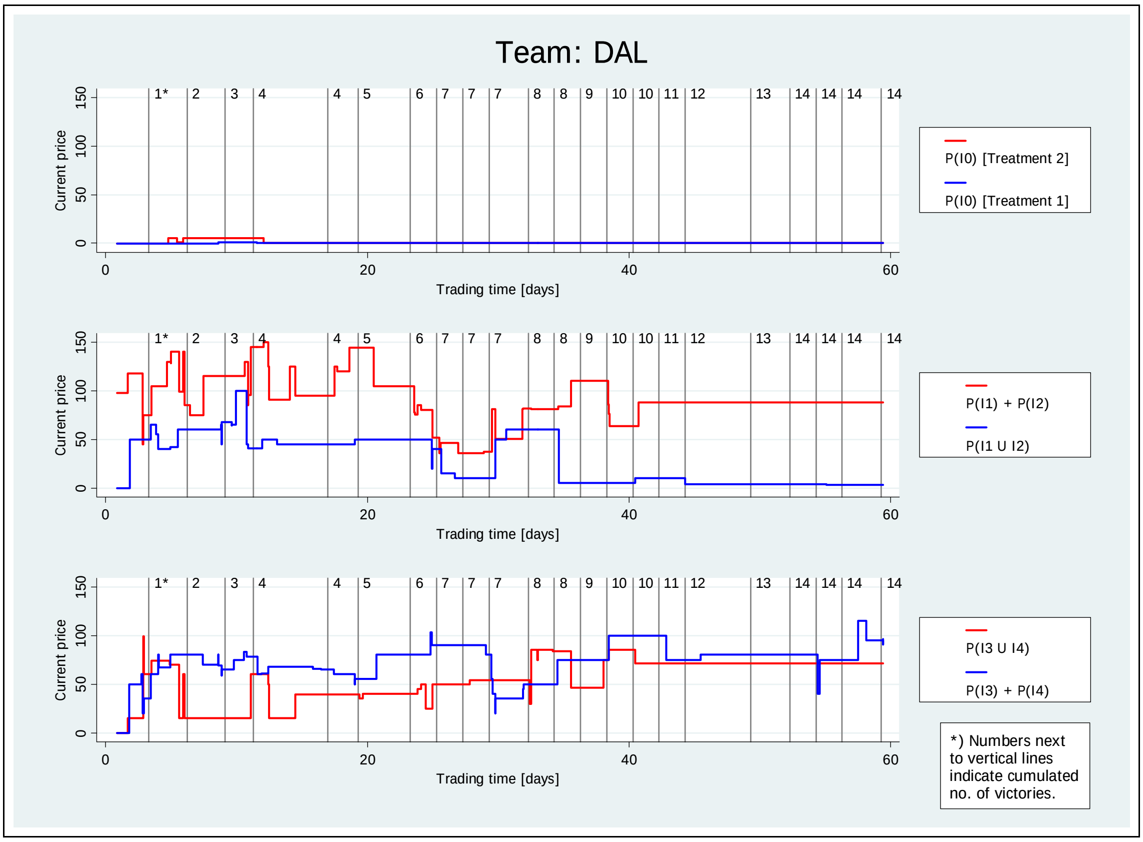


Figure A.IX.3: Price chart (Dallas Mavericks, DAL).

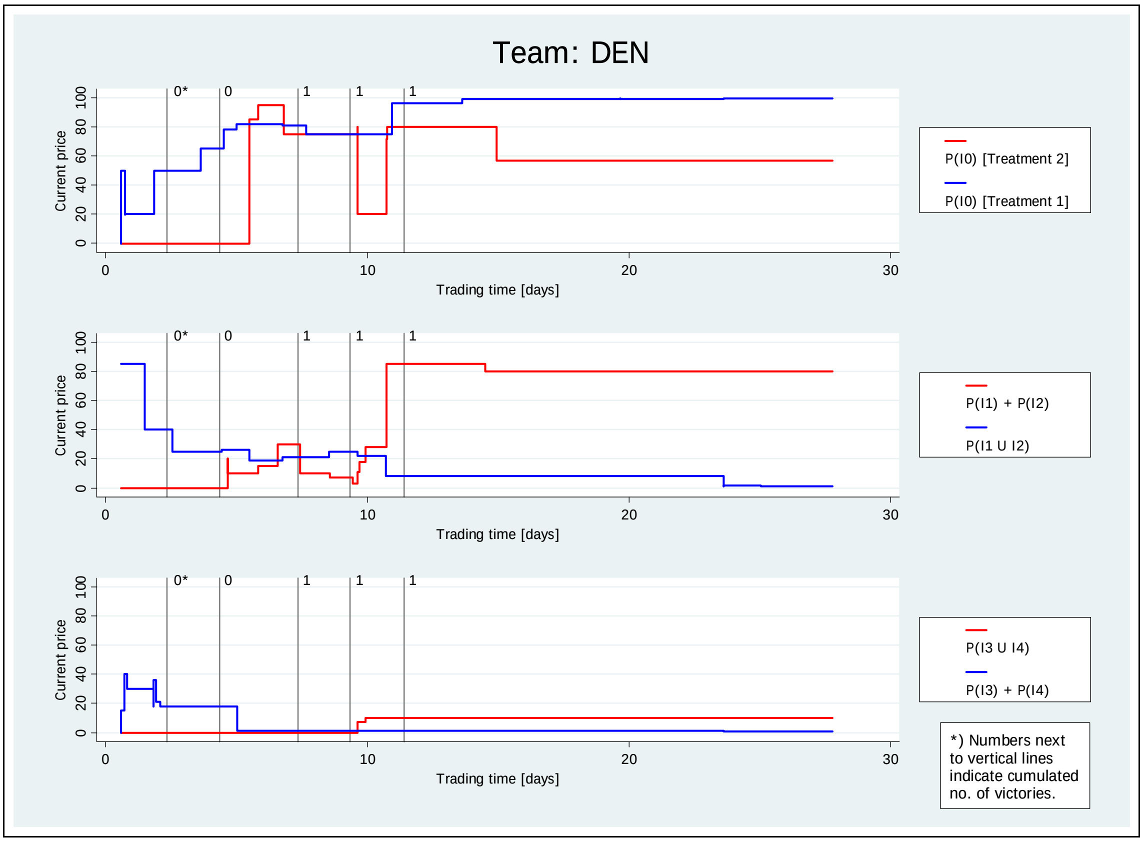


Figure A.IX.4: Price chart (Denver Nuggets, DEN).

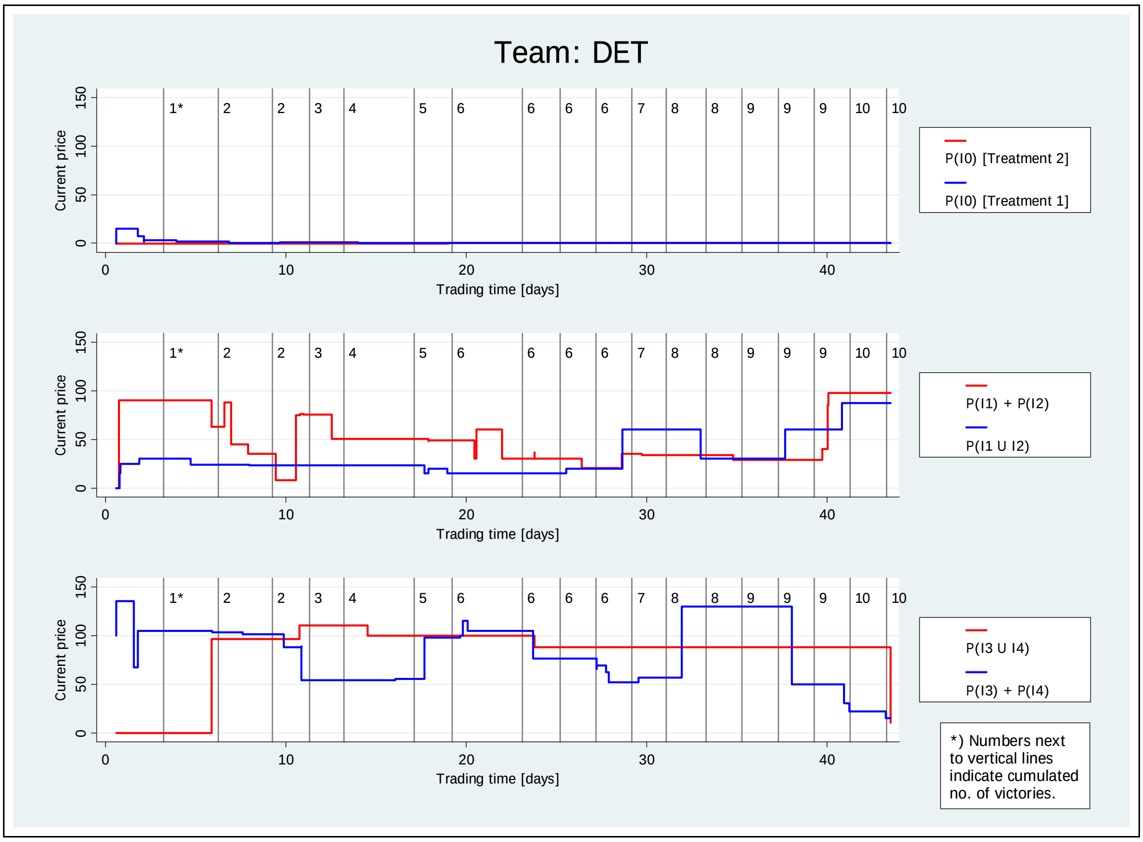


Figure A.IX.5: Price chart (Detroit Pistons, DET).

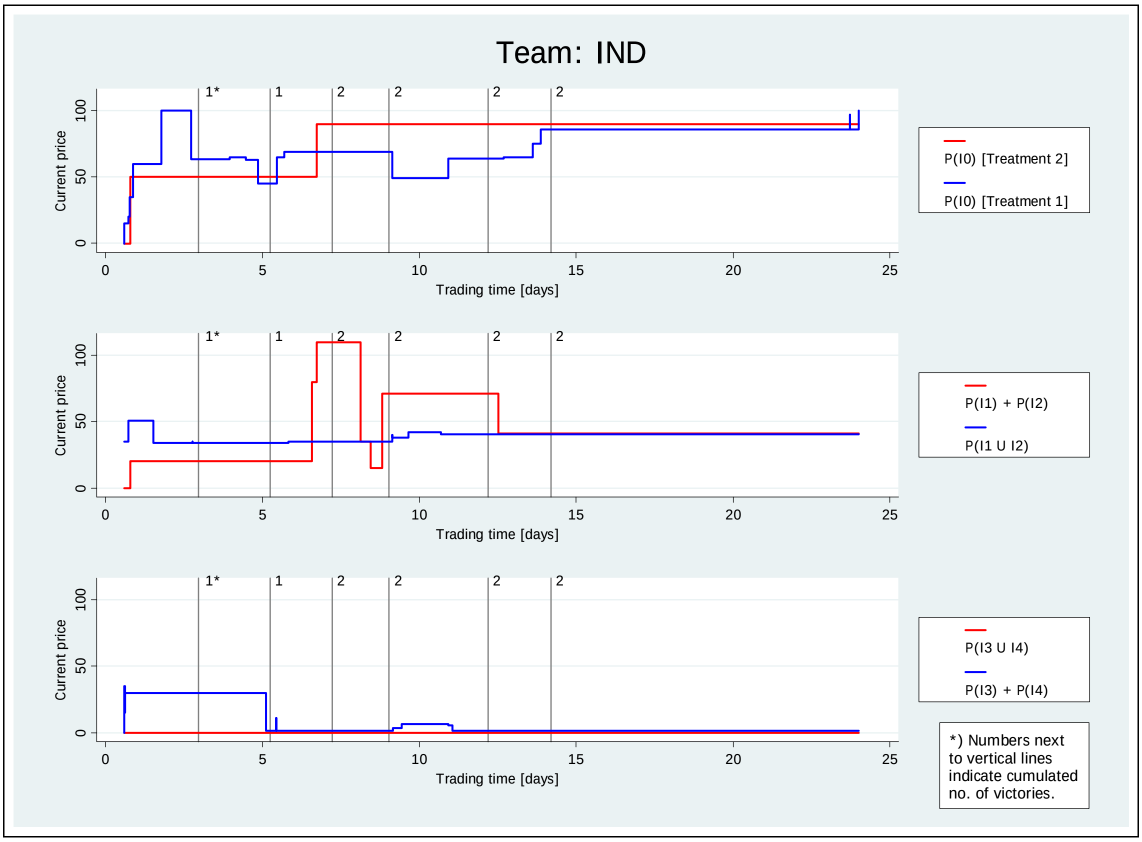


Figure A.IX.6: Price chart (Indiana Pacers, IND).

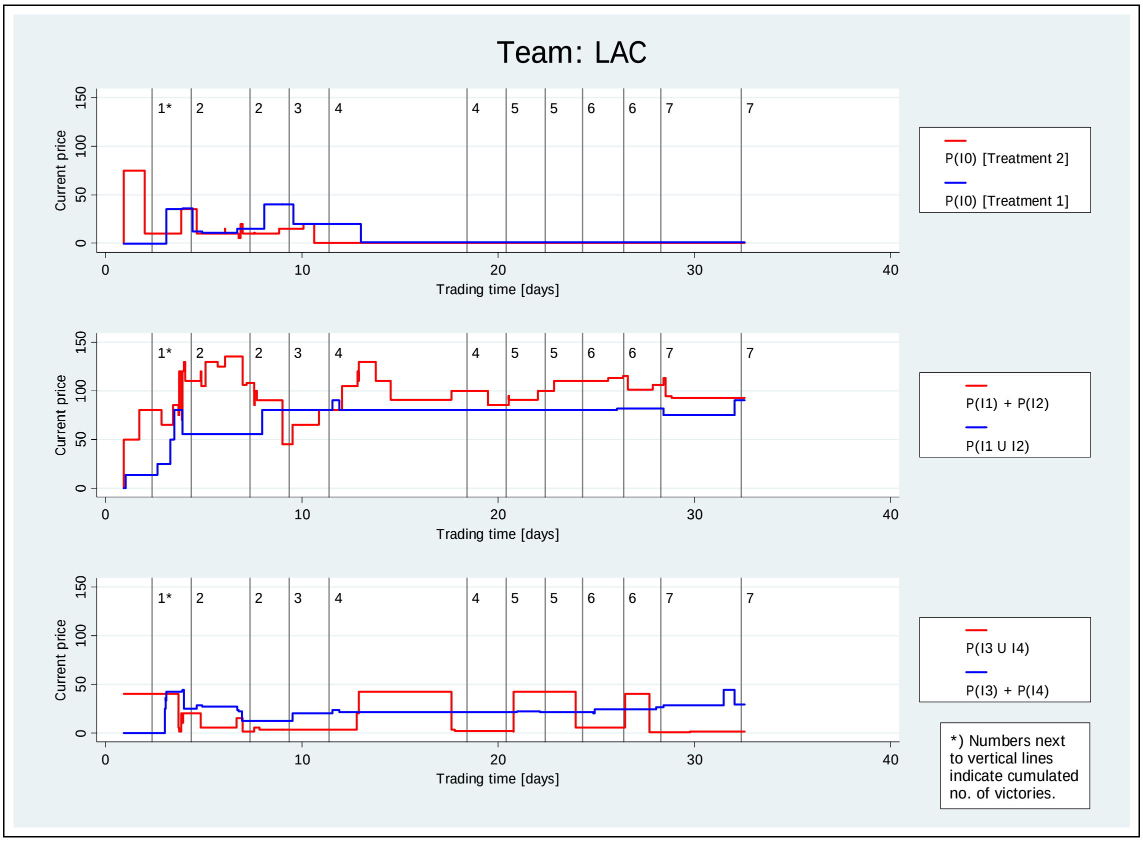


Figure A.IX.7: Price chart (L.A. Clippers, LAC).

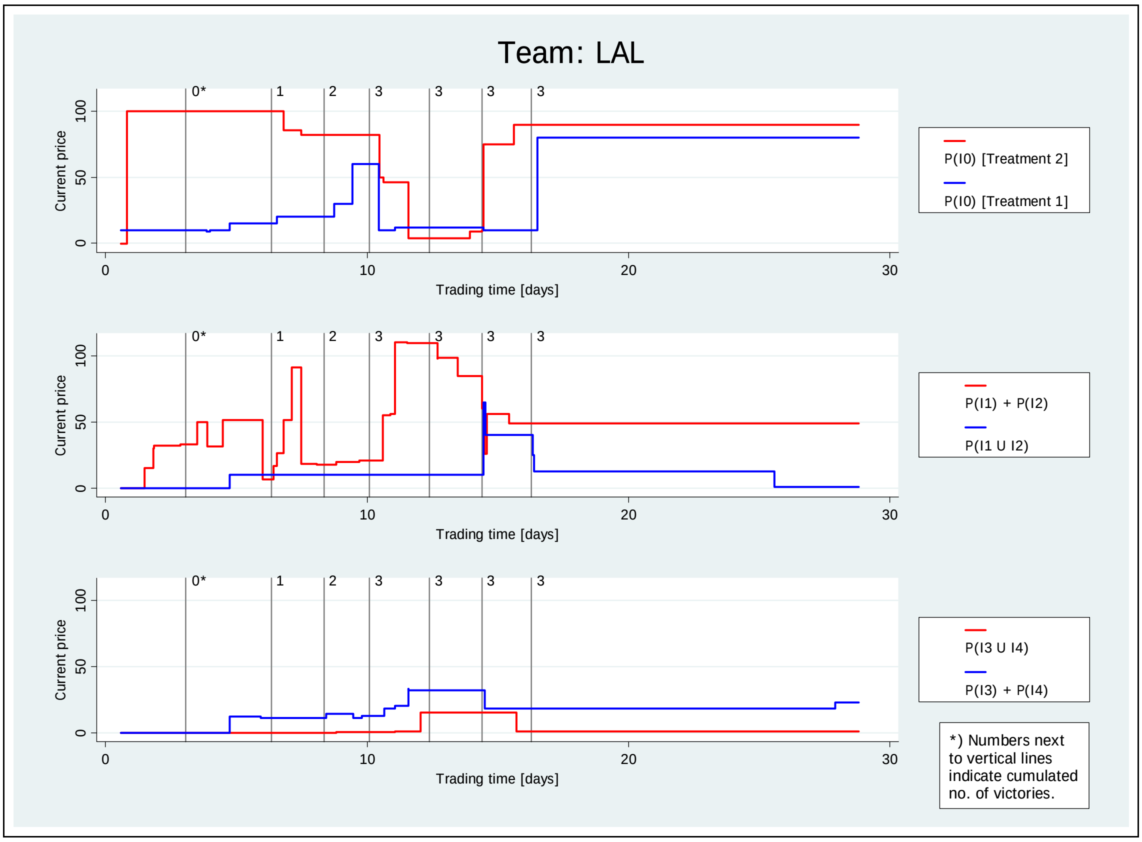


Figure A.IX.8: Price chart (L.A. Lakers, LAL).

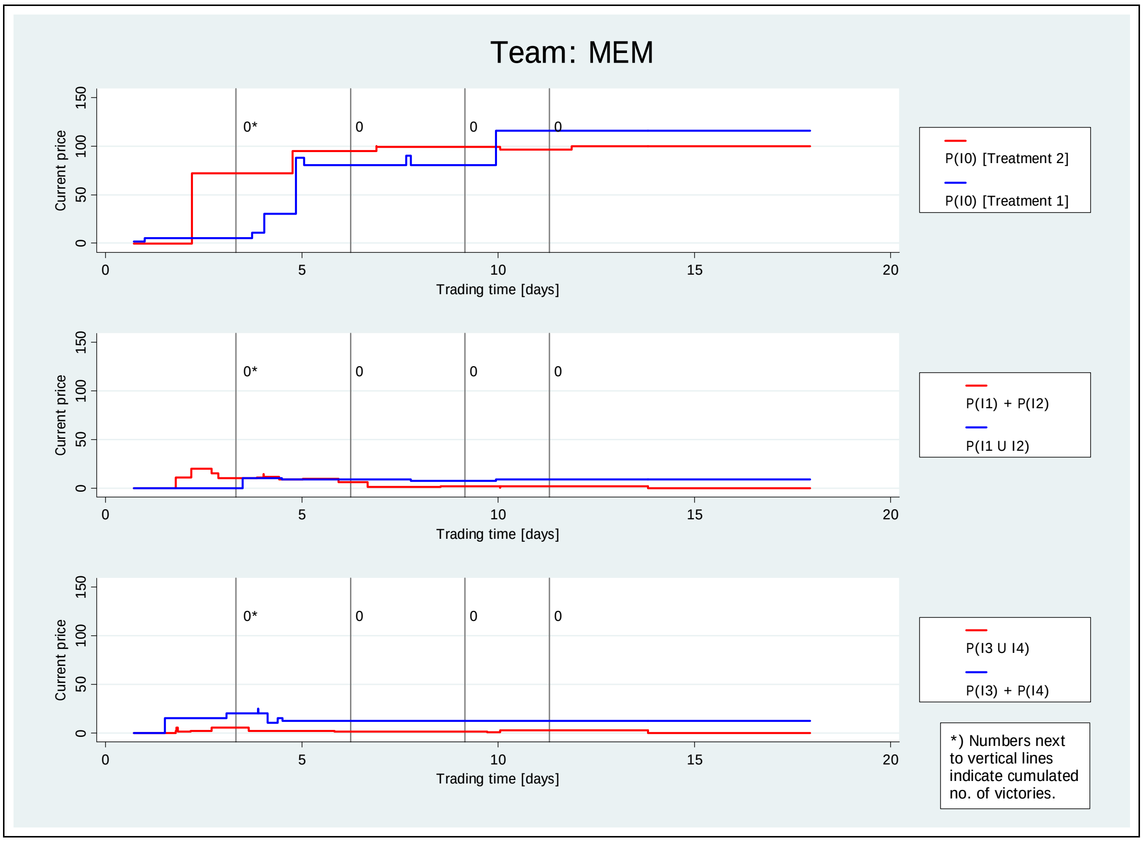


Figure A.IX.9: Price chart (Memphis Grizzlies, MEM).

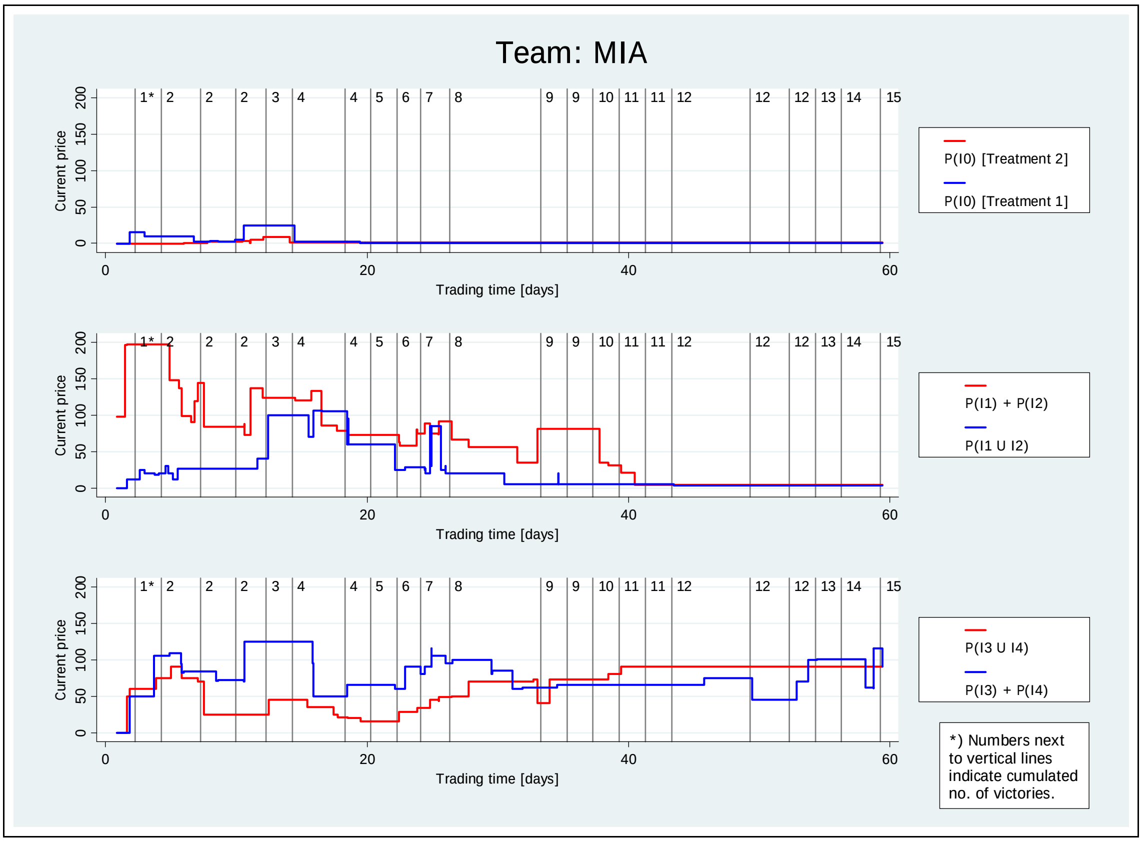


Figure A.IX.10: Price chart (Miami Heat, MIA).

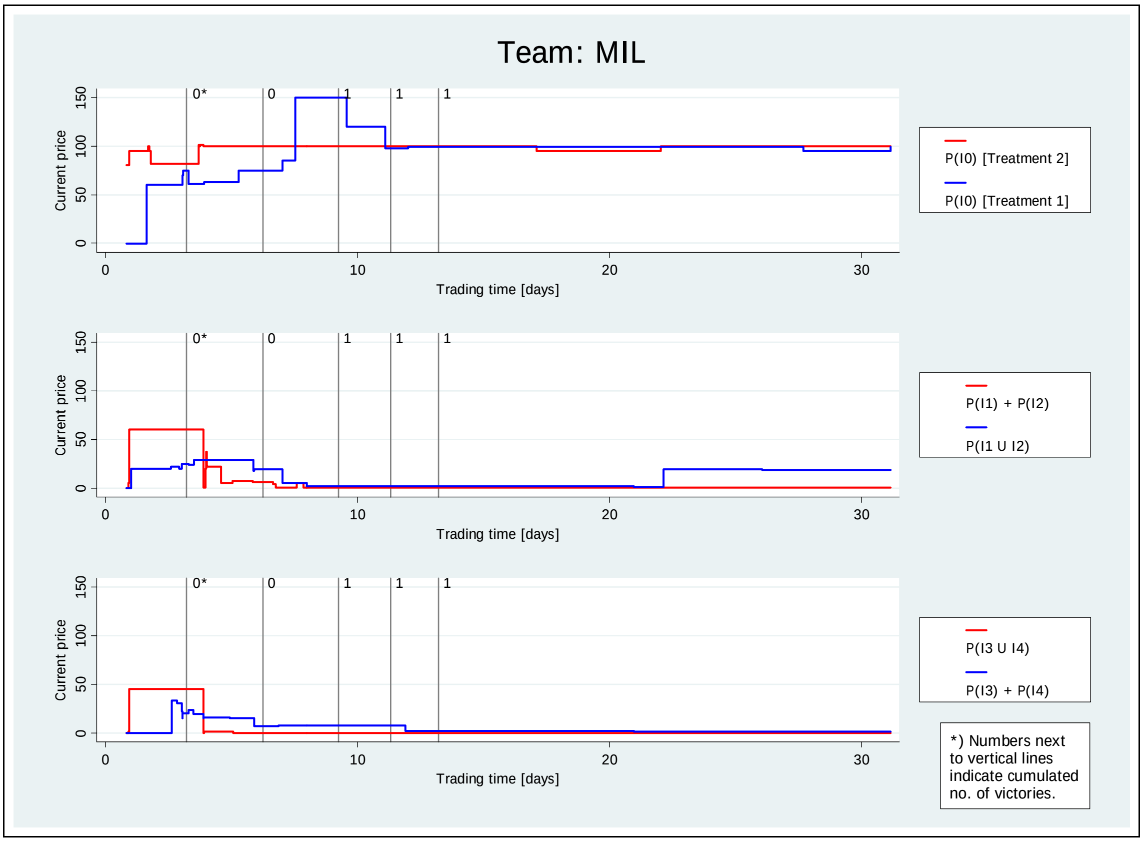


Figure A.IX.11: Price chart (Milwaukee Bucks, MIL).

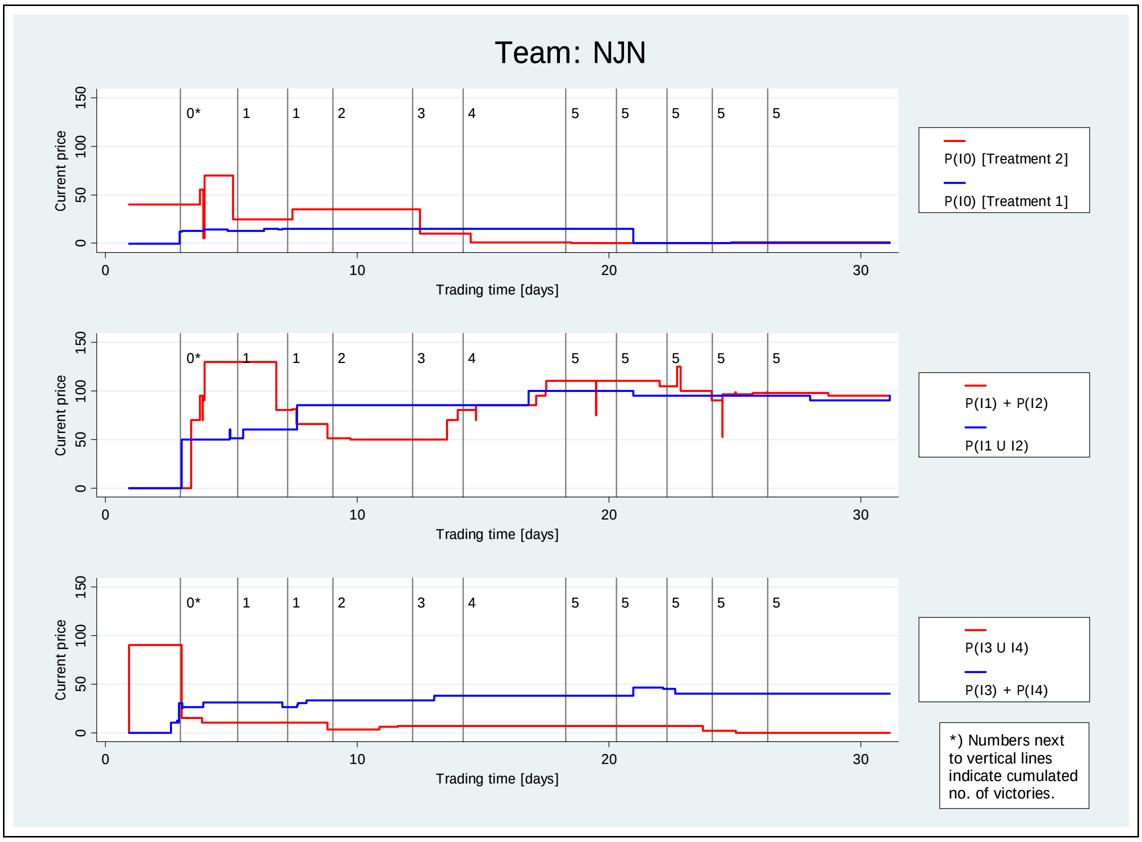


Figure A.IX.12: Price chart (New Jersey Nets, NJN).

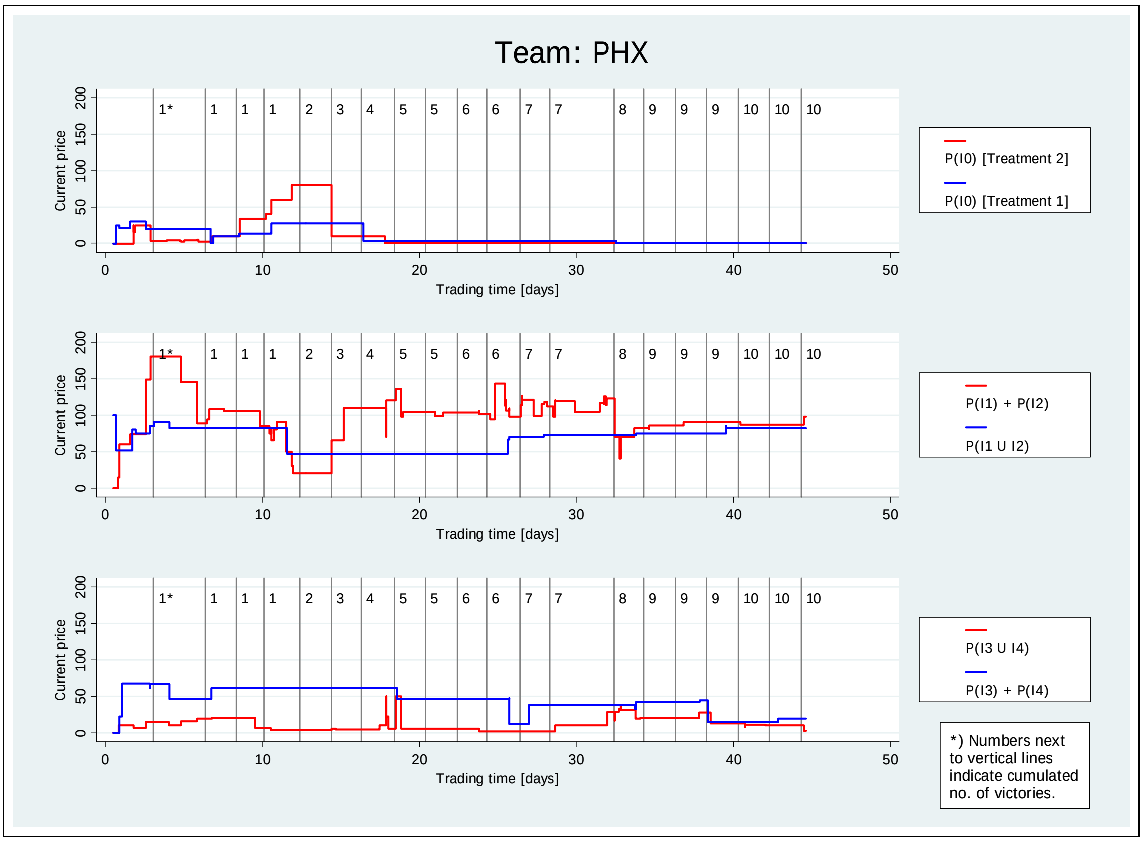


Figure A.IX.13: Price chart (Phoenix Suns, PHX).

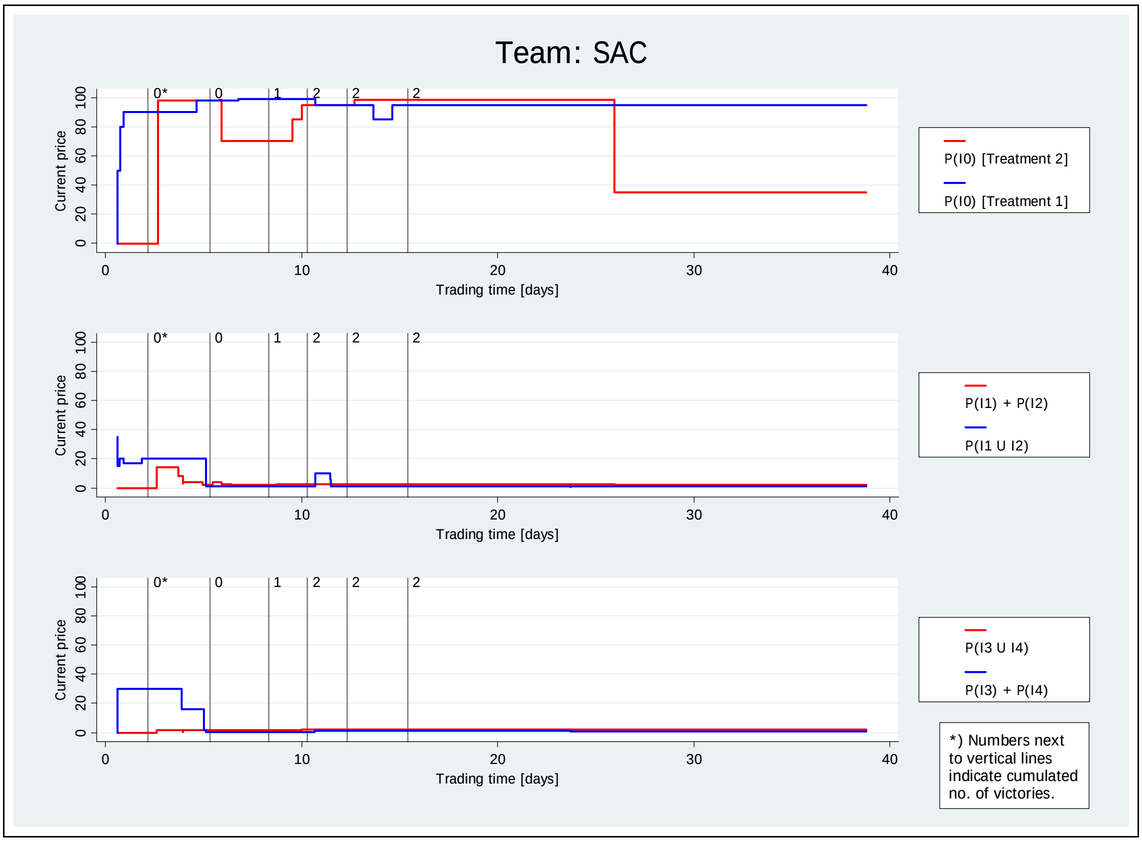


Figure A.IX.14: Price chart (Sacramento Kings, SAC).

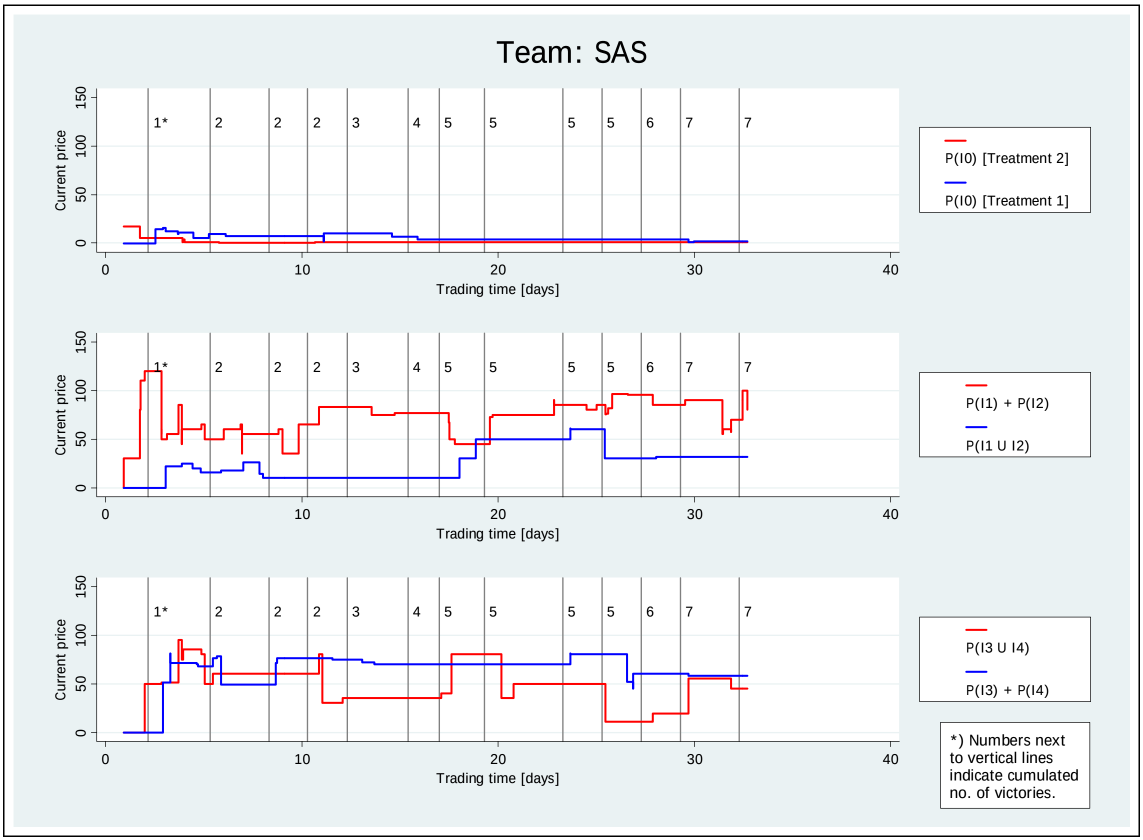


Figure A.IX.15: Price chart (San Antonio Spurs, SAS).

Figure A.IX.16: Price chart (Washington Wizards, WAS).

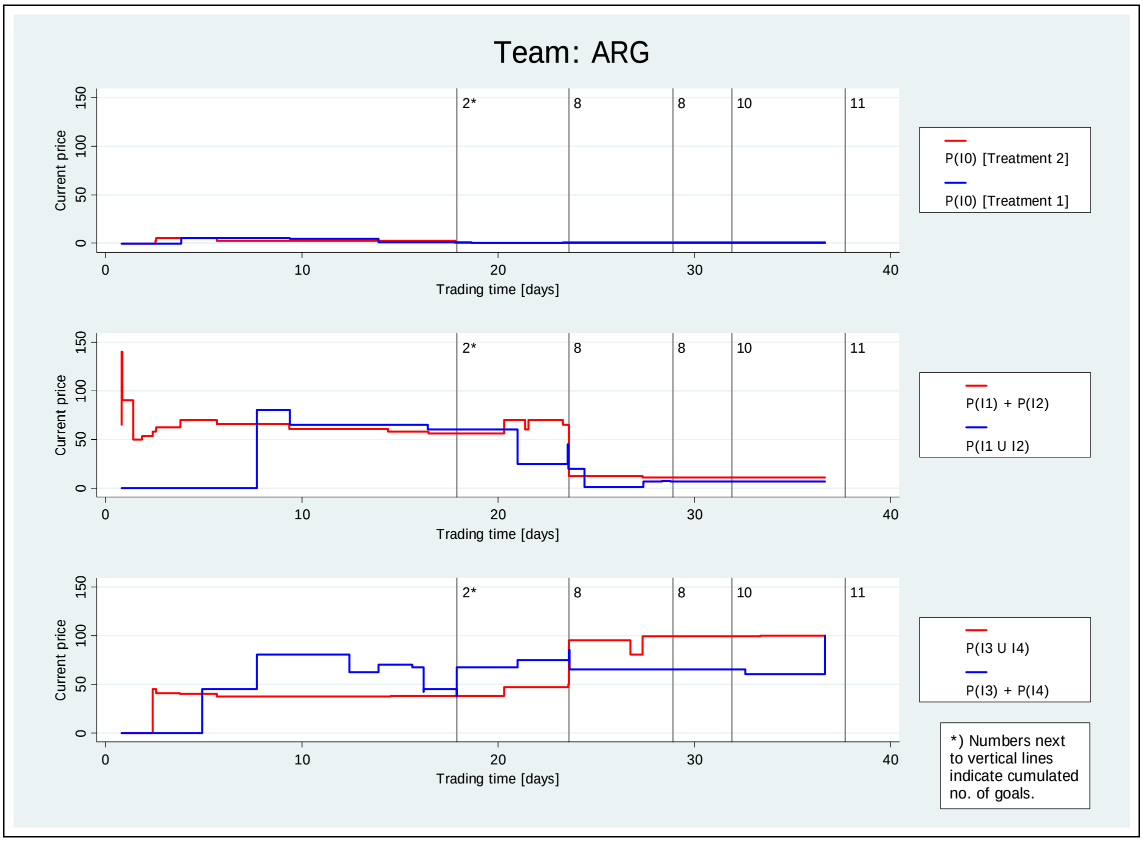


Figure A.IX.17: Price chart (Argentina, ARG).

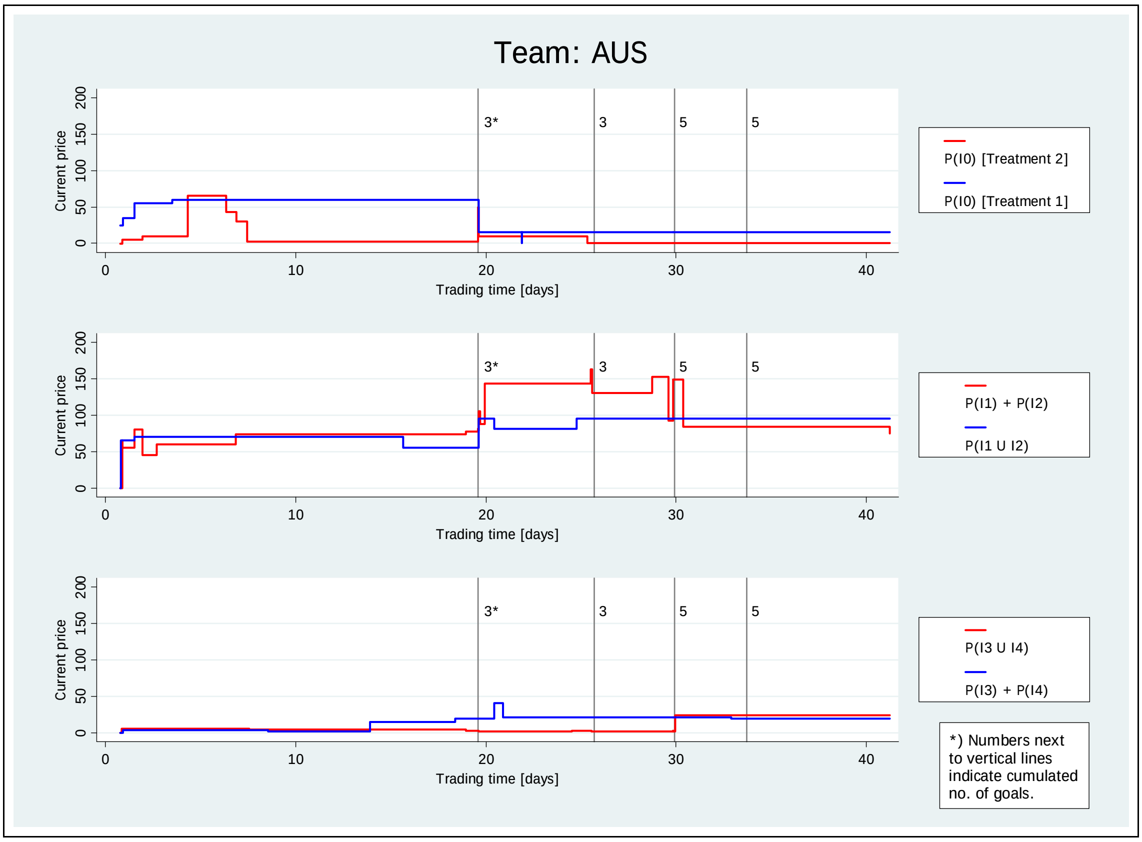


Figure A.IX.18: Price chart (Australia, AUS).

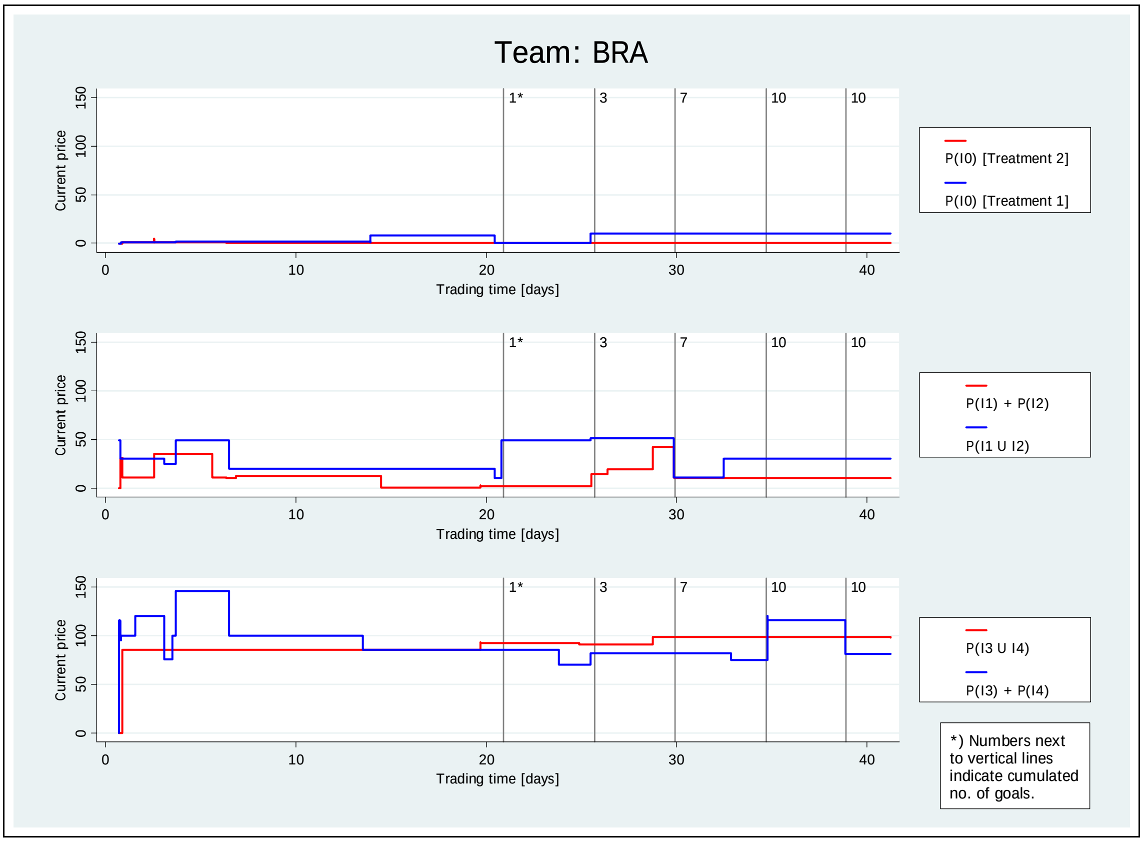


Figure A.IX.19: Price chart (Brazil, BRA).

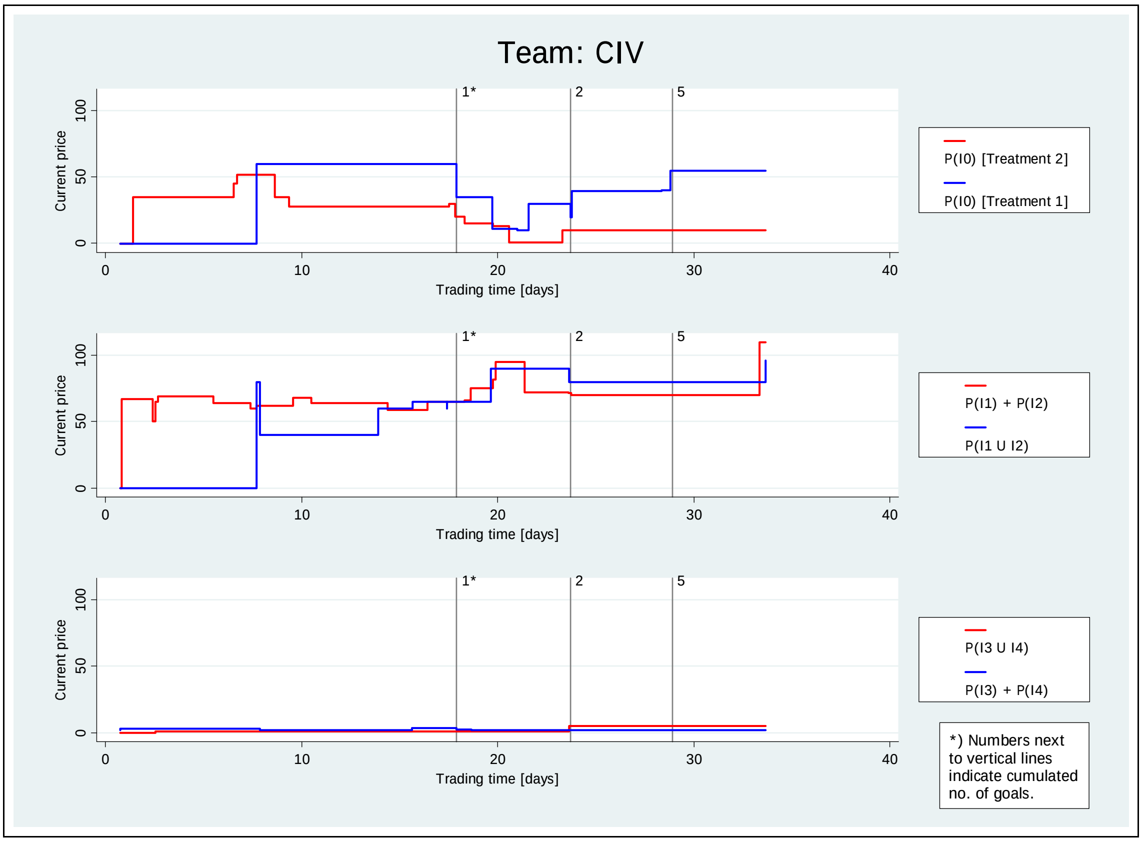


Figure A.IX.20: Price chart (Côte d'Ivoire, CIV).

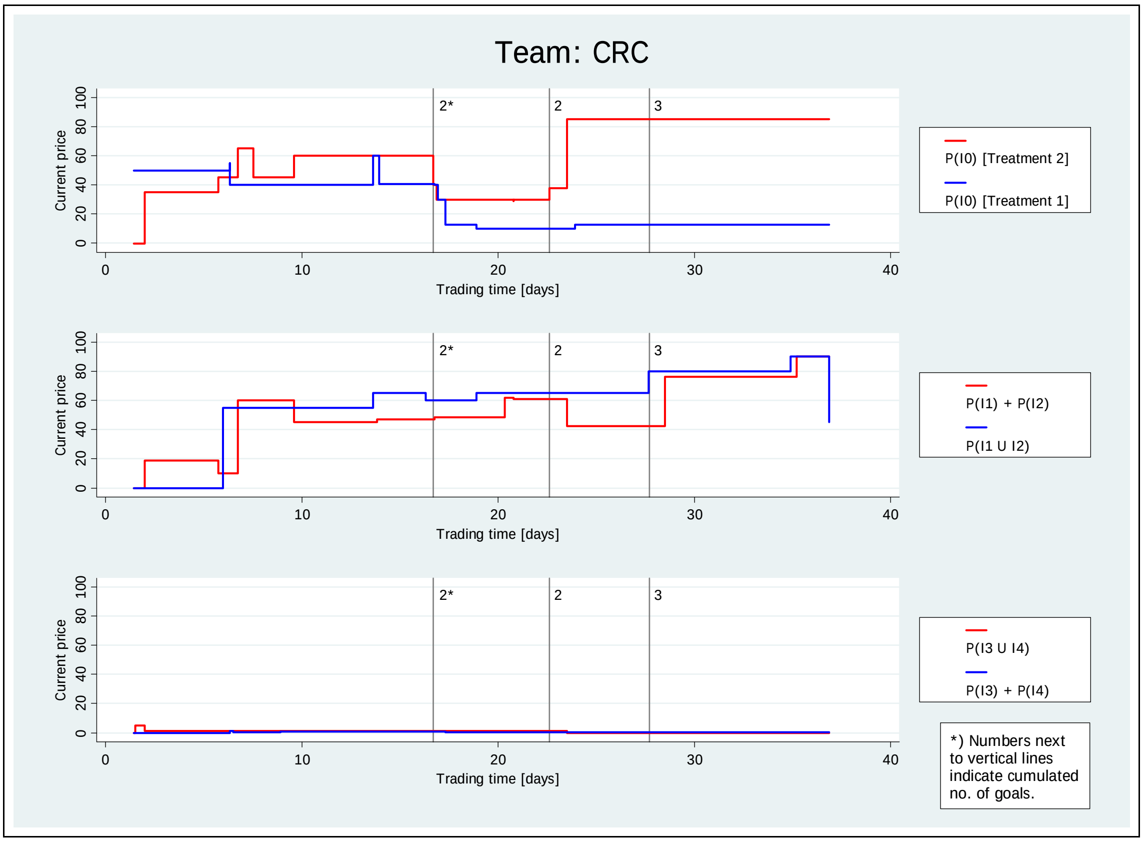


Figure A.IX.21: Price chart (Costa Rica, CRC).

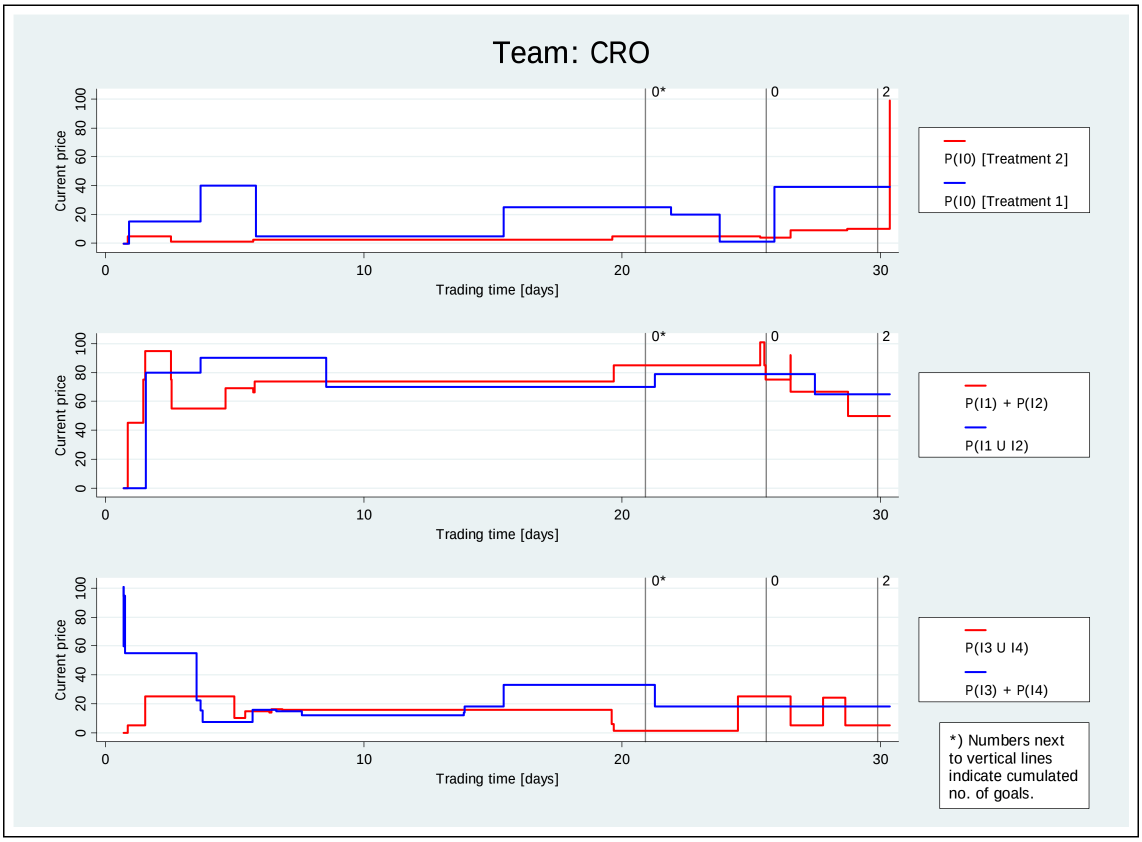


Figure A.IX.22: Price chart (Croatia, CRO).

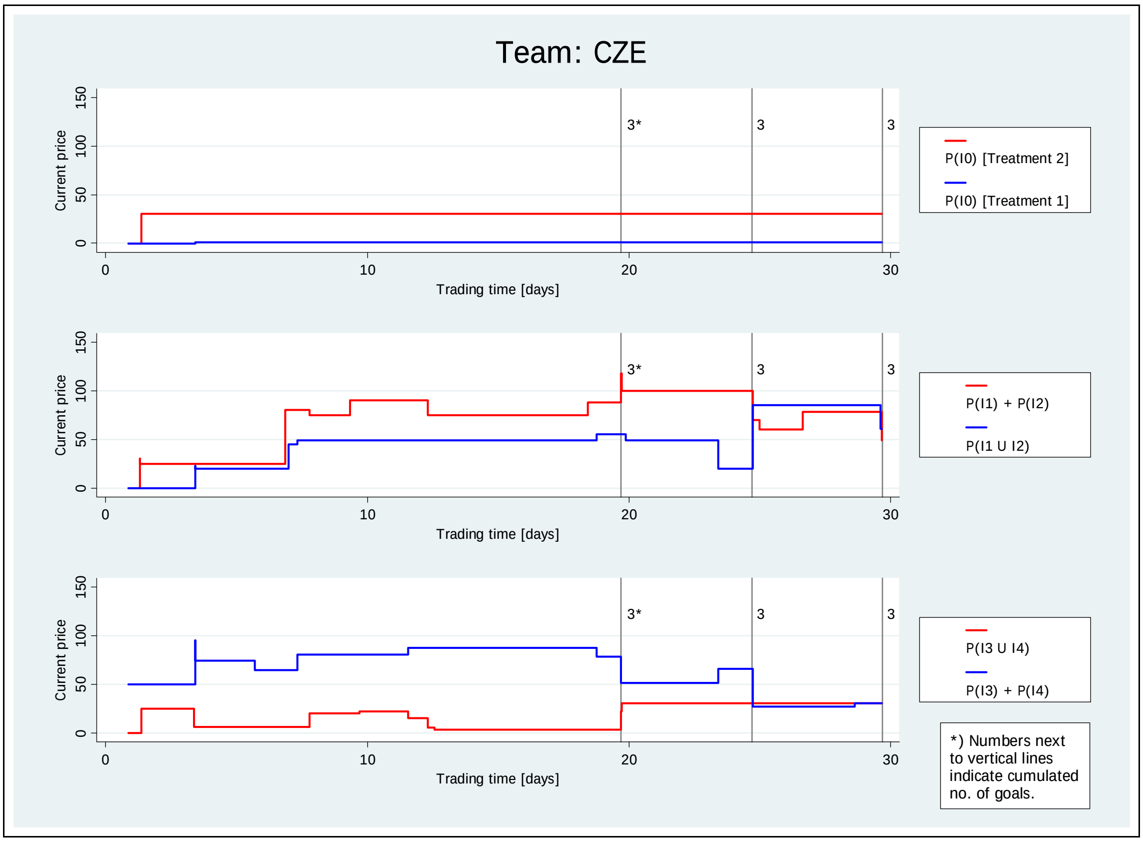


Figure A.IX.23: Price chart (Czech Republic, CZE).

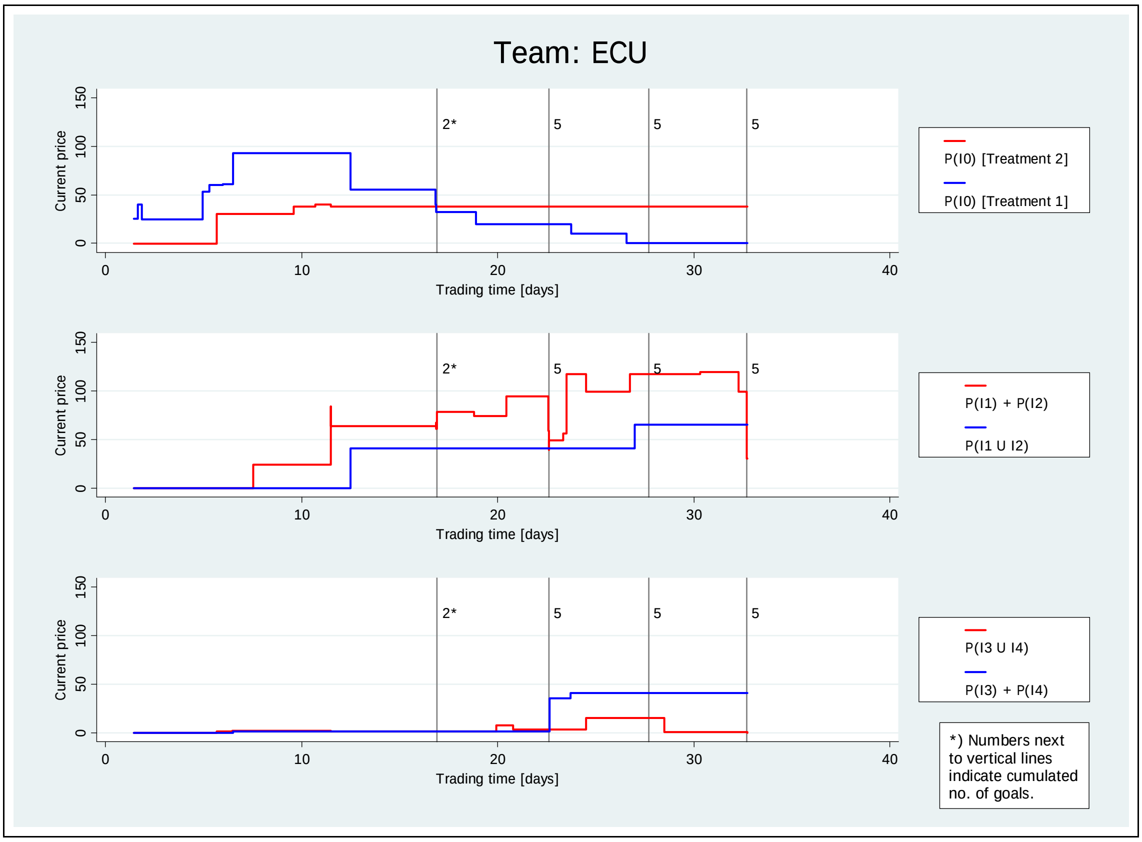


Figure A.IX.24: Price chart (Ecuador, ECU).

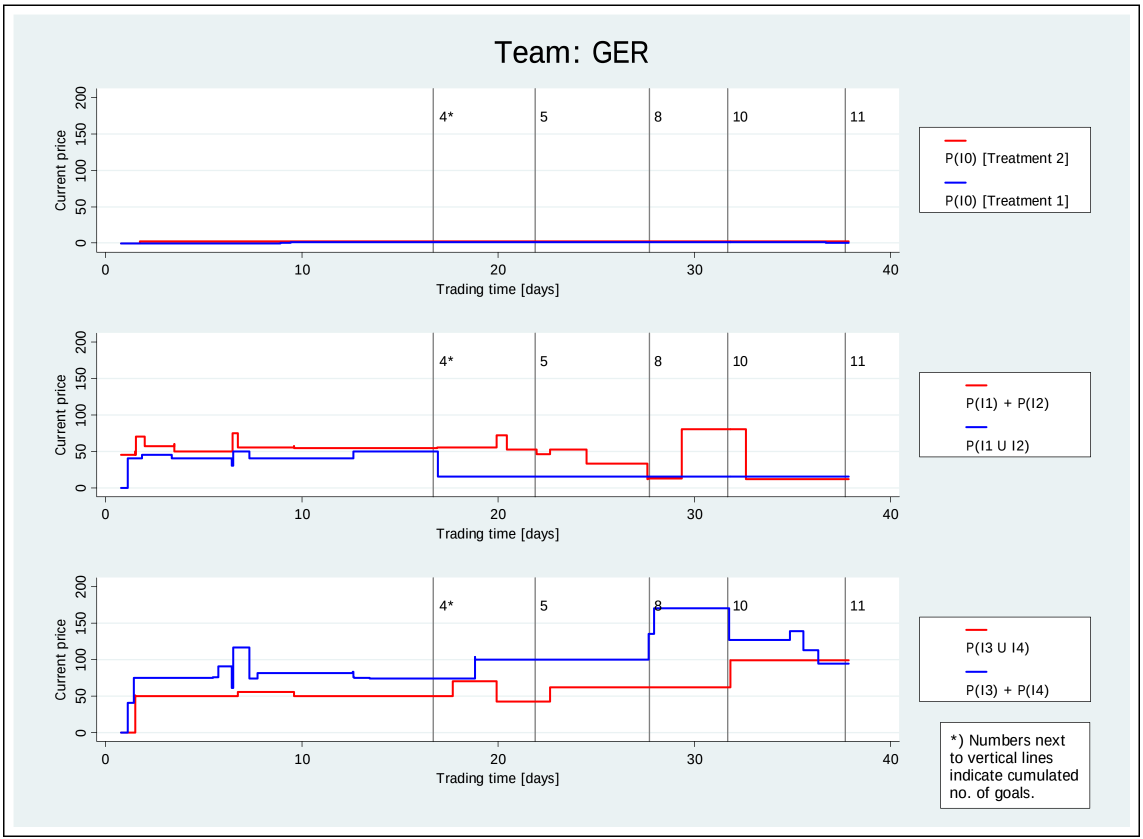


Figure A.IX.25: Price chart (Germany, GER).

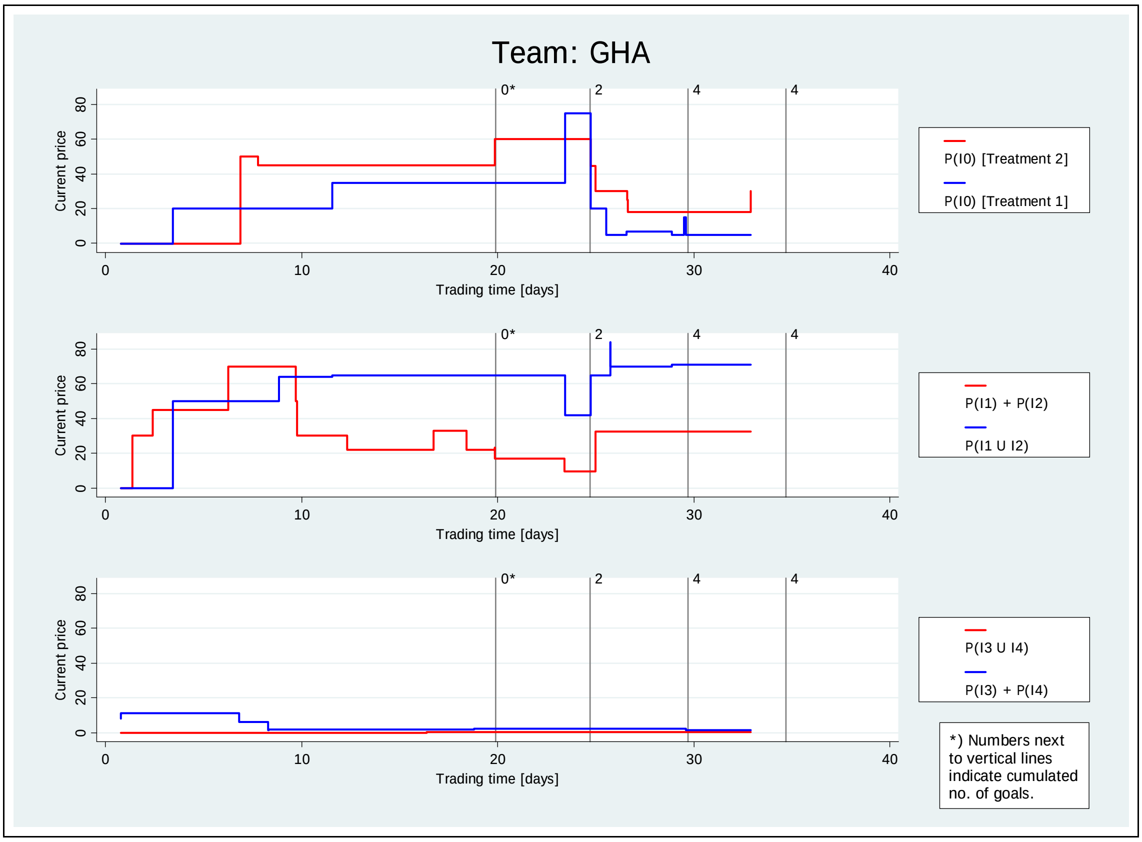


Figure A.IX.26: Price chart (Ghana, GHA).

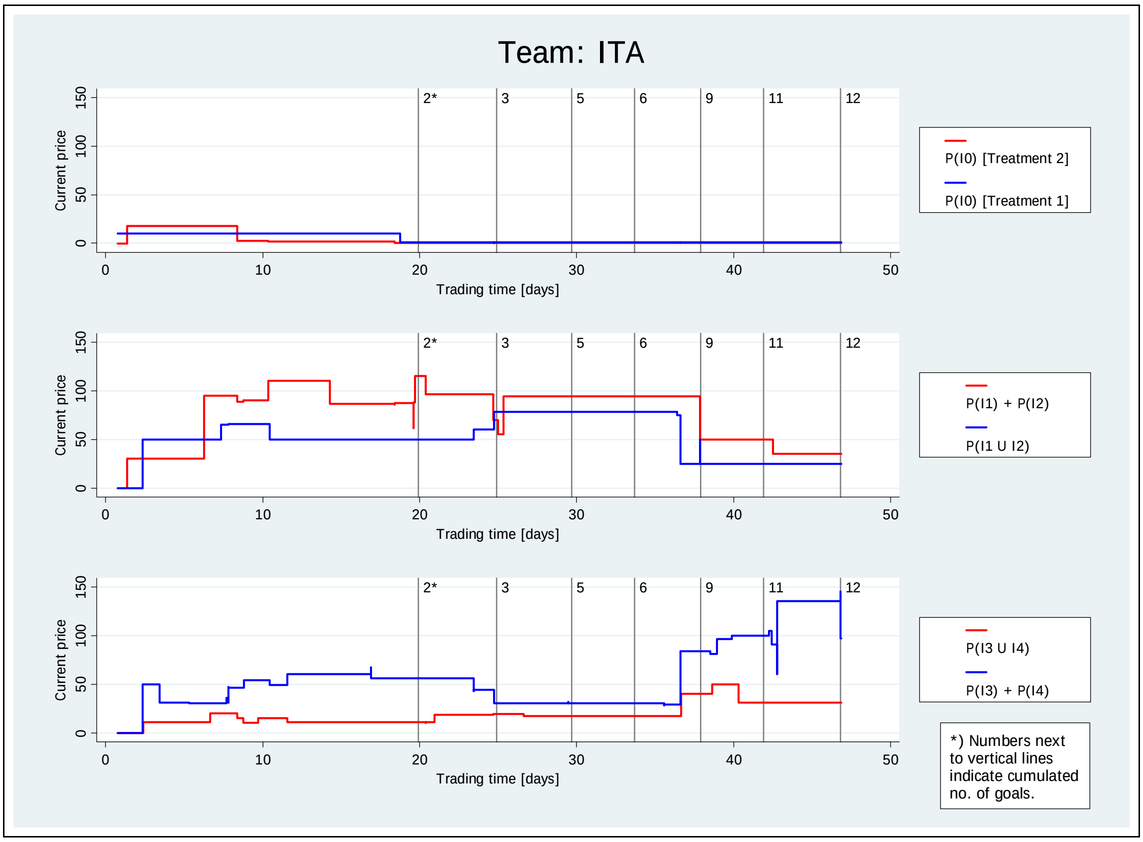


Figure A.IX.27: Price chart (Italy, ITA).

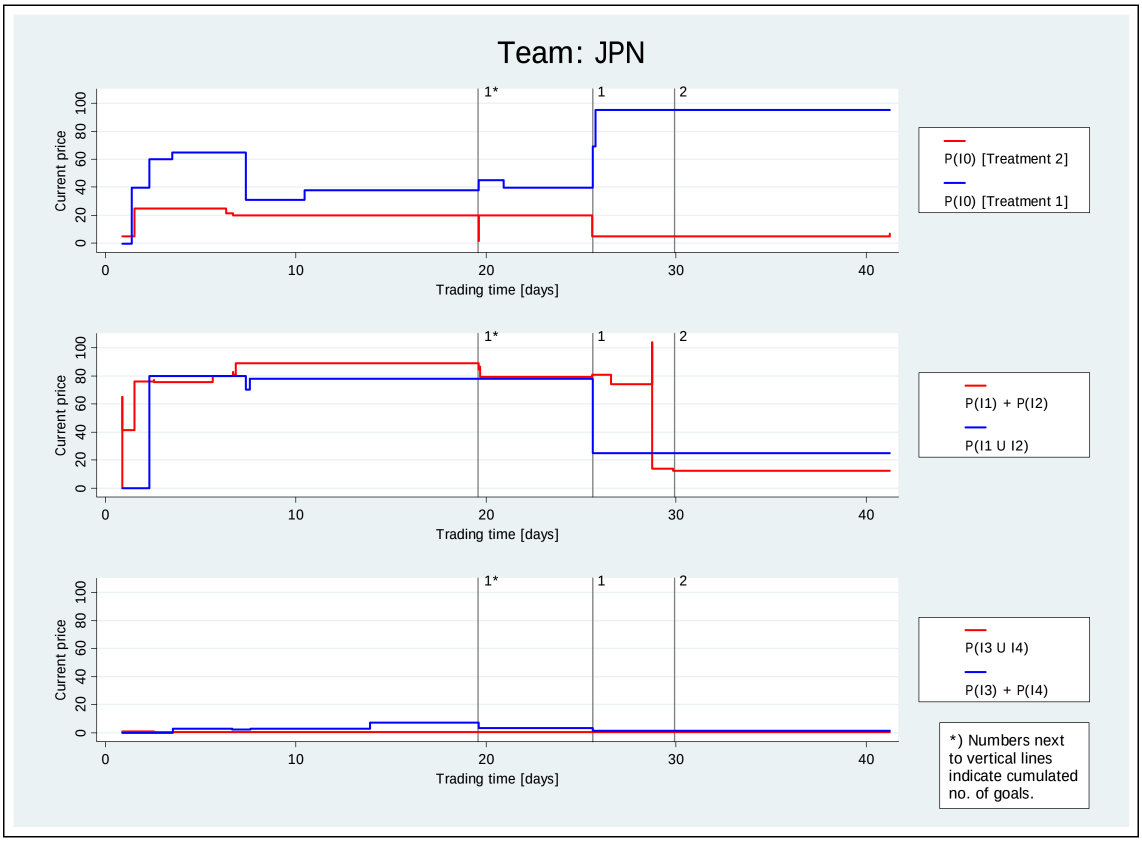


Figure A.IX.28: Price chart (Japan, JPN).

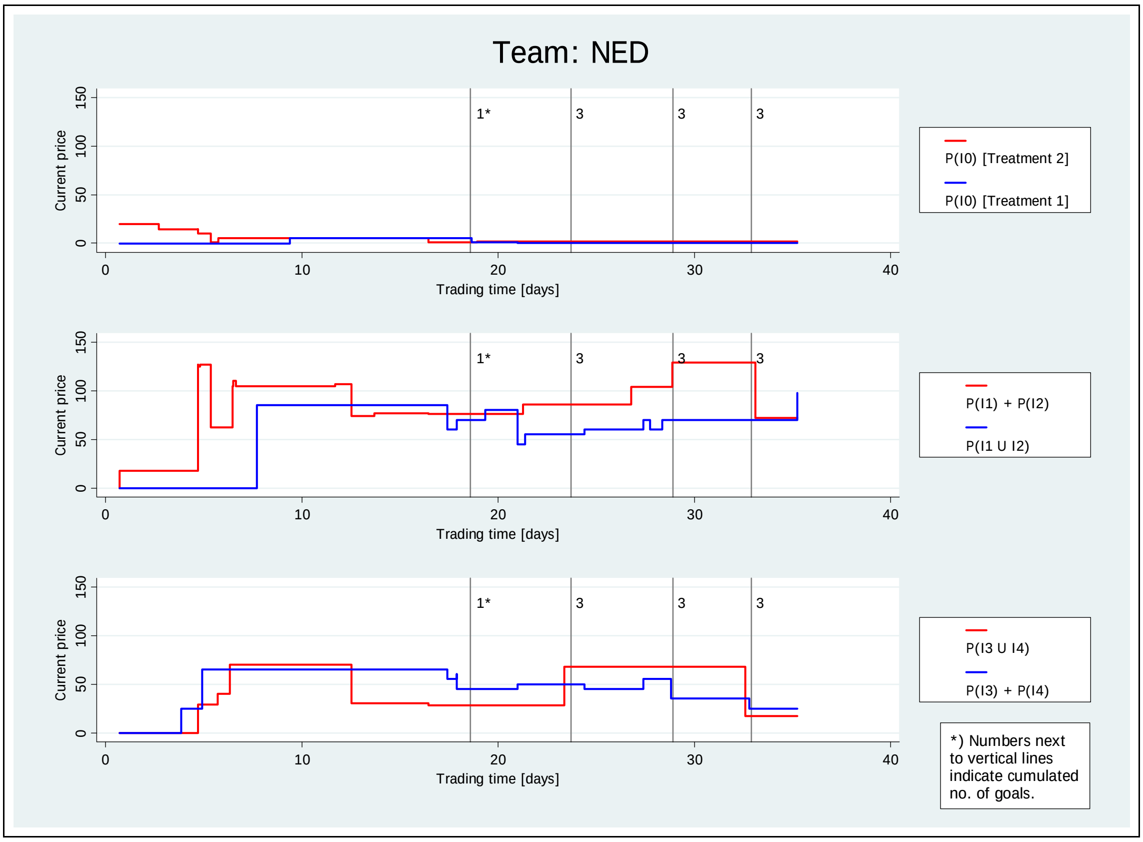


Figure A.IX.29: Price chart (Netherlands, NED).

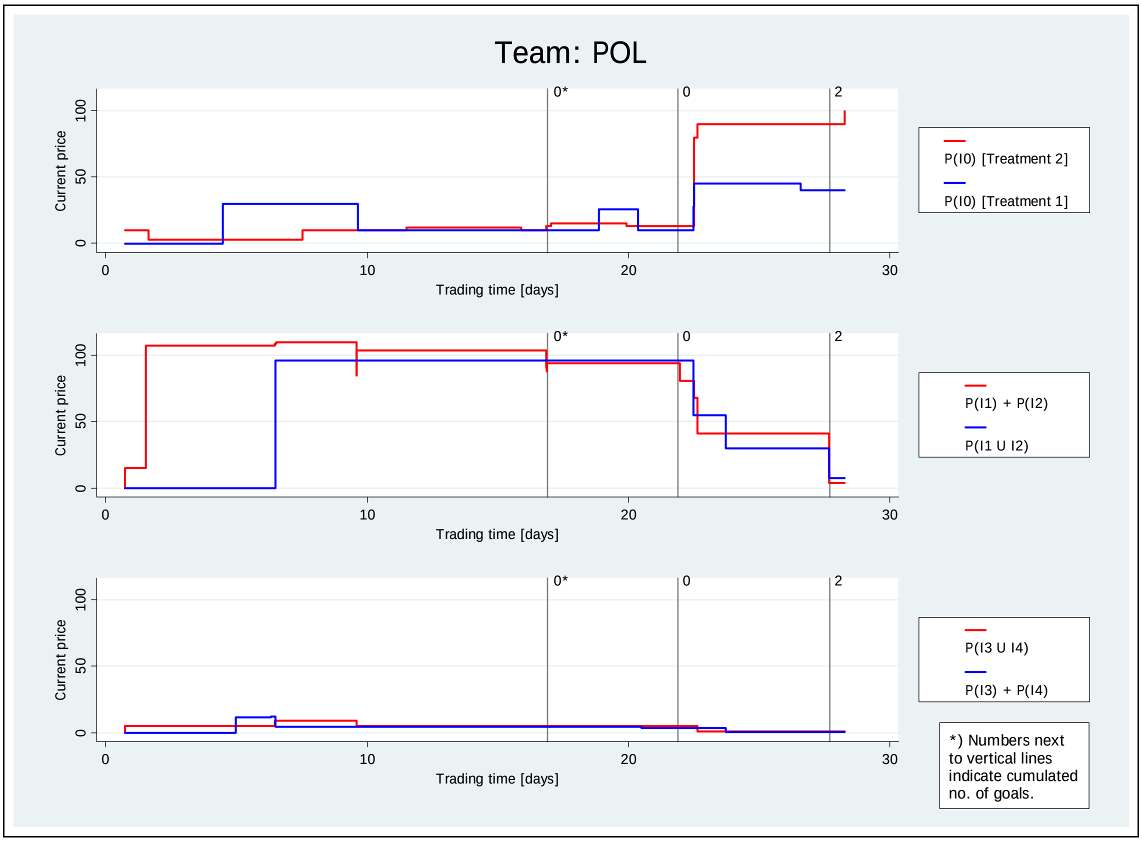


Figure A.IX.30: Price chart (Poland, POL).

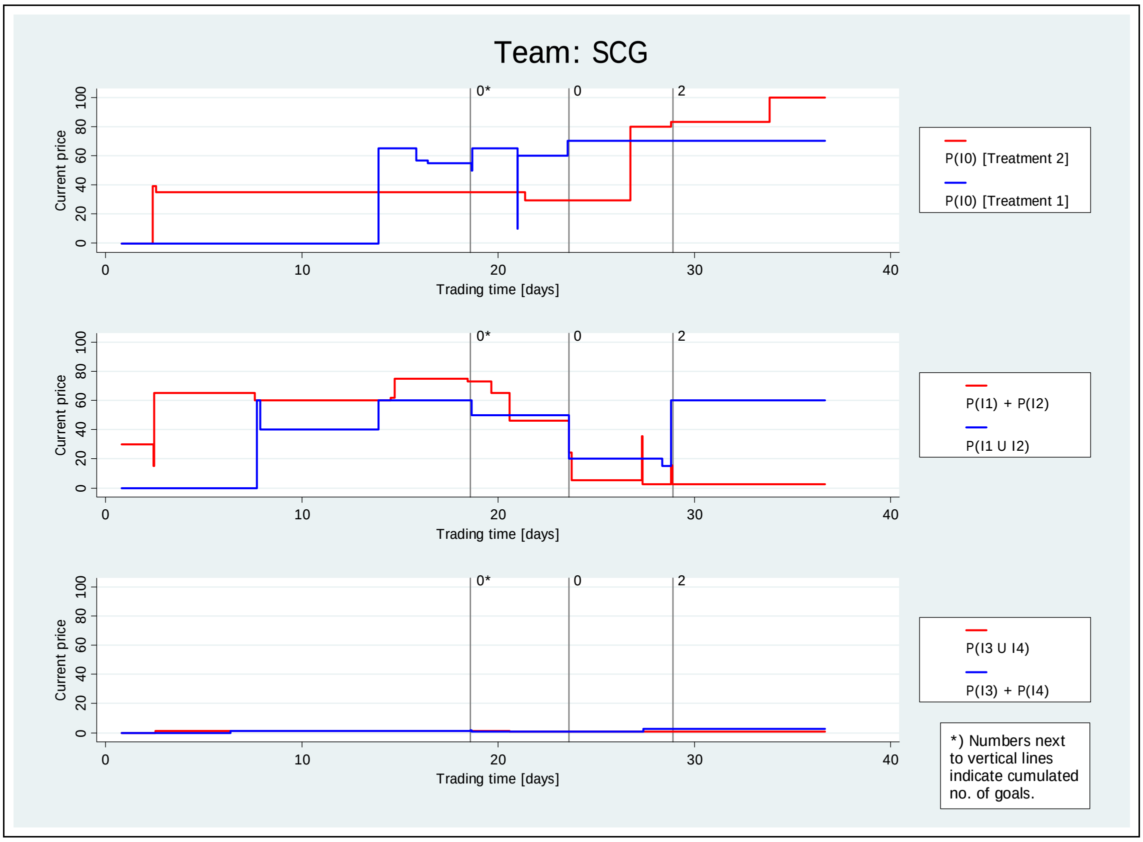


Figure A.IX.31: Price chart (Serbia and Montenegro, SCG).

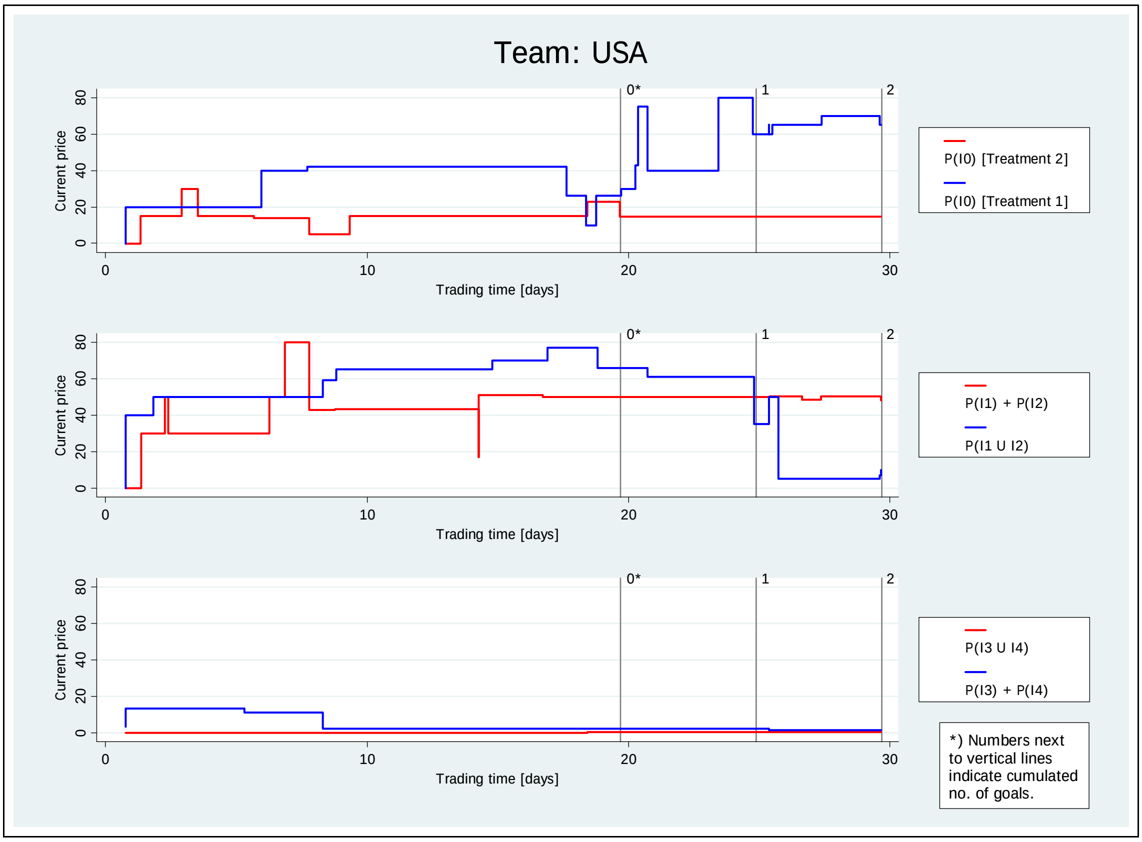


Figure A.IX.32: Price chart (United States of America, USA).

Appendix X: Screenshot of the trading interface

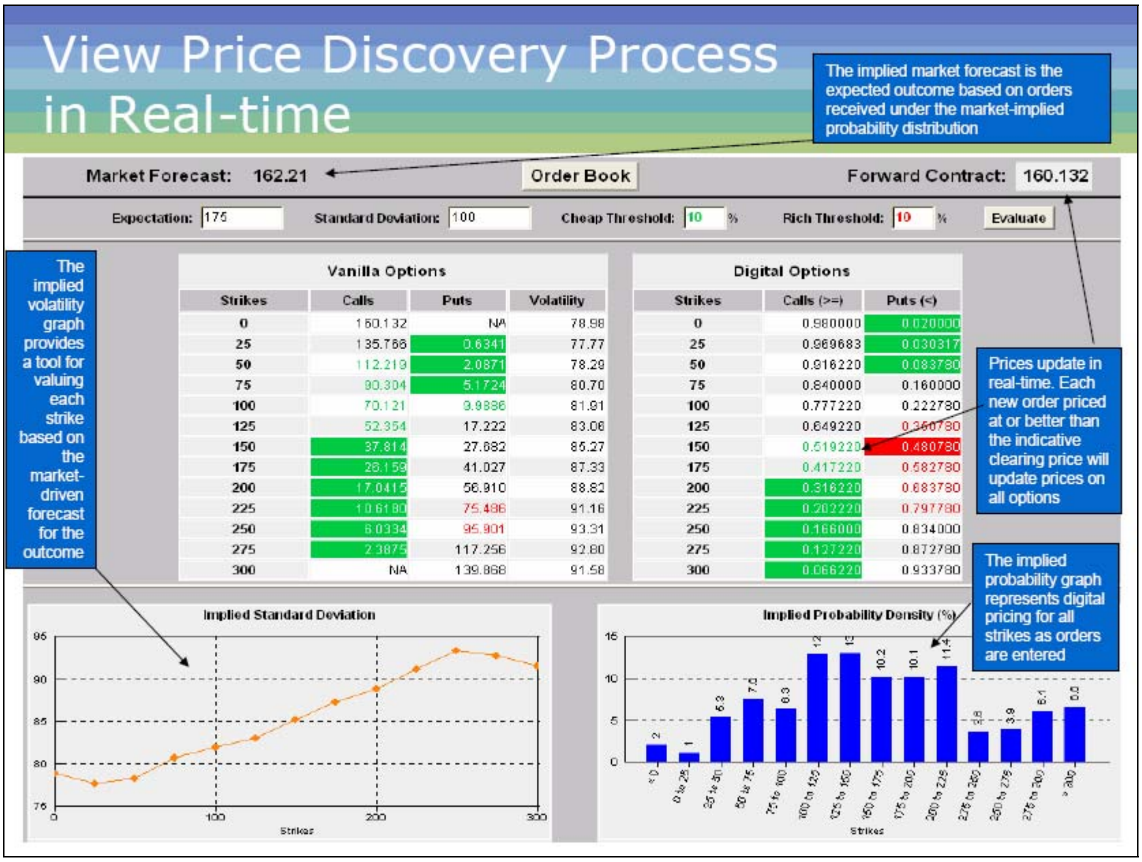


Figure A.X.1: *Screenshot of the trading interface for economic derivatives (taken from “Economic Derivatives. Options on economic statistics”, a presentation prepared by trading and sales groups within Deutsche Bank AG and Goldman Sachs).*

Appendix XI: Derivation of the regression model

From (5.1) it follows that:

$$M_{obs} = (1-\lambda)\cdot M_{true} + \lambda \cdot M_{1/N} \,.$$

Thus,

$$M_{true} = \frac{M_{obs}}{(1-\lambda)} - \frac{\lambda \cdot M_{1/N}}{(1-\lambda)} \,.$$

Further define:

$$c = -\frac{\lambda}{1-\lambda} \,.$$

Then,

$$\begin{aligned}
M_{obs} &= (1-\lambda)\cdot M_{true} + \lambda \cdot M_{1/N} \\
&= M_{true} - \lambda \cdot M_{true} + \lambda \cdot M_{1/N} \\
&= M_{true} - \lambda \cdot \left(\frac{M_{obs}}{(1-\lambda)} - \frac{\lambda \cdot M_{1/N}}{(1-\lambda)} \right) + \lambda \cdot M_{1/N} \\
&= M_{true} + c \cdot \left(M_{obs} - \lambda \cdot M_{1/N} - (1-\lambda)\cdot M_{1/N} \right) \\
&= M_{true} + c \cdot \left(M_{obs} - M_{1/N} \right) \\
&= M_{true} - \frac{\lambda}{(1-\lambda)} \cdot \left(M_{obs} - M_{1/N} \right)
\end{aligned}$$

And finally, the observed forecast error can be written as:

$$M_{obs} - X = \left[M_{true} - X \right] - \frac{\lambda}{(1-\lambda)} \cdot \left(M_{obs} - M_{1/N} \right).$$

8 References

Ahn, David S., and Haluk Ergin, 2007, Framing contingencies, Working paper, University of California, Berkeley and Massachusetts Institute of Technology.

Allais, Maurice, 1953, Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'Ecole Americaine, *Econometrica* 21, 503-546.

Anderson, Matthew J., and Shyam Sunder, 1995, Professional traders as intuitive Bayesians, *Organizational Behavior and Human Decision Processes* 64, 185-202.

Antweiler, Werner, and Thomas W. Ross, 1998, The 1997 UBC election stock market, *Canadian Business Economics* 6, 15-22.

Arrow, Kenneth J., 1982, Risk perception in psychology and economics, *Economic Inquiry* 20, 1-9.

Aumann, Robert J., 1976, Agreeing to disagree, *The Annals of Statistics* 4, 1236-1239.

Ayton, Peter, 1997, How to be incoherent and seductive: Bookmakers' odds and support theory, *Organizational Behavior and Human Decision Processes* 72, 99-115.

Backhaus, Klaus, Bernd Erichson, Wulff Plinke, and Rolf Weiber, 2006, *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (Springer, Berlin et al.).

Banks, Jeffrey S., 1985, Price-conveyed information versus observed insider behavior: A note on rational expectations convergence [Efficiency of experimental security markets with insider information: An application of rational-expectations models], *Journal of Political Economy* 93, 807-815.

Bansal, Vipul K., John F. Marshall, and Robert P. Yuyuenyongwatana, 1994, Hedging business cycle risk with macroeconomic swaps: Some preliminary evidence, *Journal of Derivatives*, Spring, 50-58.

Bansal, Vipul K., John F. Marshall, and Robert P. Yuyuenyongwatana, 1995, Macroeconomic derivatives: More viable than first thought, *Global Finance Journal* 6, 101-110.

Bardolet, David, Craig R. Fox, and Dan Lovallo, 2007, Naïve diversification and partition dependence in capital allocation decisions, Working paper, UCLA Anderson School.

Baron, Ken, 2004, *The Mathematical Principles of PDCA™* (Longitude, Inc., Hoboken, NJ).

Baron, Ken, and Jeffrey Lange, 2003, From horses to hedging, *Risk* 16, 73-77.

Baumohl, Bernard, 2007, *The Secrets of Economic Indicators: Hidden Clues to Future Economic Trends and Investment Opportunities* (Wharton School Publishing, Upper Saddle River, NJ).

Beber, Alessandro, and Michael W. Brandt, 2009, Resolving macroeconomic uncertainty in stock and bond markets, *Review of Finance* 13, 1-45.

Beckmann, Klaus, and Martin Werding, 1994, Markets and the use of knowledge: Testing the «Hayek Hypothesis» in experimental stock markets, Working paper, Universität Passau.

Beckmann, Klaus, and Martin Werding, 1996, "Passauer Wahlbörse": Information processing in a political market experiment, *Kyklos* 49, 171-205.

Bell, David E., 1982, Regret in decision making under uncertainty, *Operations Research* 30, 961-981.

Benartzi, Shlomo, and Richard H. Thaler, 2001, Naive diversification strategies in defined contribution saving plans, *American Economic Review* 91, 79-98.

Berg, Joyce E., Robert Forsythe, Forrest Nelson, and Thomas A. Rietz, 2008, Results from a dozen years of election futures markets research, in Charles R. Plott, and Vernon L. Smith, eds.: *Handbook of Experimental Economics Results, Volume 1* (North-Holland [Elsevier], Amsterdam).

Berg, Joyce E., Robert Forsythe, and Thomas A. Rietz, 1996, What makes markets predict well? Evidence from the Iowa Electronic Markets, in Wulf Albers, Werner Güth, Peter Hammerstein, Benny Moldovanu, and Eric van Damme, eds.: *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten* (Springer, New York).

Berg, Joyce E., Robert Forsythe, and Thomas A. Rietz, 1997, The Iowa Electronic Market, in Dean Paxson, and Douglas Wood, eds.: *Blackwell Encyclopedic Dictionary of Finance* (Blackwell, Oxford, UK).

Berg, Joyce E., George R. Neumann, and Thomas A. Rietz, 2008, Searching for Google's value: Using prediction markets to forecast market capitalization prior to an initial public offering, Working paper, University of Iowa.

Berg, Joyce E., and Thomas A. Rietz, 2003, Prediction markets as decision support systems, *Information Systems Frontiers* 5, 79-93.

Berg, Joyce E., and Thomas A. Rietz, 2006, The Iowa Electronic Market: Lessons learned and answers yearned, in Robert W. Hahn, and Paul C. Tetlock, eds.: *Information Markets: A New Way of Making Decisions* (AEI-Brookings Joint Center, Washington, D.C.).

Biais, Bruno, Denis Hilton, Karine Mazurier, and Sébastien Pouget, 2005, Judgmental overconfidence, self-monitoring, and trading performance in an experimental financial market, *Review of Economic Studies* 72, 287-312.

Bilgin, Baler, and Lyle Brenner, 2008, Temporal distance moderates description dependence of subjective probability, *Journal of Experimental Social Psychology* 44, 890-895.

Birnbaum, Michael, and Ulrich Schmidt, 2008, An experimental investigation of violations of transitivity in choice under uncertainty, *Journal of Risk and Uncertainty* 37, 77-91.

Blume, Lawrence E., and David Easley, 1982, Learning to be rational, *Journal of Economic Theory* 26, 340-351.

Bohm, Peter, and Joakim Sonnegard, 1999, Political stock markets and unreliable polls, *Scandinavian Journal of Economics* 101, 205-222.

Bossaerts, Peter, 2002, *The Paradox of Asset Pricing* (Princeton University Press, Princeton and Oxford).

Brenner, Lyle, and Yuval Rottenstreich, 1999, Focus, repacking, and the judgment of grouped hypotheses, *Journal of Behavioral Decision Making* 12, 141-148.

Brüggelambert, Gregor, 1999, *Institutionen als Informationsträger: Erfahrungen mit Wahlbörsen* (Metropolis-Verlag, Marburg).

Brüggelambert, Gregor, 2004, Information and efficiency in political stock markets: Using computerized markets to predict election results, *Applied Economics* 36, 753-768.

Camerer, Colin F., 1987, Do biases in probability judgment matter in markets? Experimental evidence, *American Economic Review* 77, 981-997.

Camerer, Colin F., 1992, The rationality of prices and volume in experimental markets, *Organizational Behavior and Human Decision Processes* 51, 237-272.

Camerer, Colin F., 1997, Rules for experimenting in psychology and economics, and why they differ, in Wulf Albers, Werner Güth, Peter Hammerstein, Benny Moldovanu, and Eric van Damme, eds.: *Understanding Strategic Interaction. Essays in Honor of Reinhard Selten* (Springer, New York).

Camerer, Colin F., and Ernst Fehr, 2006, When does 'economic man' dominate social behavior?, *Science* 311, 47-52.

Camerer, Colin F., and Teck-Hua Ho, 1999, Experience-weighted attraction learning in normal form games, *Econometrica* 67, 827-874.

Camerer, Colin F., and Robin M. Hogarth, 1999, The effects of financial incentives in experiments: A review and capital-labor-production framework, *Journal of Risk and Uncertainty* 19, 7-42.

Camerer, Colin F., George Loewenstein, and Martin Weber, 1989, The curse of knowledge in economic settings: An experimental analysis, *Journal of Political Economy* 97, 1232-1254.

Camerer, Colin F., and Keith Weigelt, 1991, Information mirages in experimental asset markets, *Journal of Business* 64, 463-493.

Chamley, Christophe P., 2004, *Rational Herds: Economic Models of Social Learning* (Cambridge University Press, Cambridge).

Chen, Kay-Yut, and Charles R. Plott, 2002, Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem, Social Science Working paper 1131, California Institute of Technology.

Cipriani, Marco, and Antonio Guarino, 2005, Herd behavior in a laboratory financial market, *American Economic Review* 95, 1427-1443.

Clemen, Robert T., and Terence Reilly, 2001, *Making Hard Decisions with Decision Tools* (Duxbury Thomson Learning, Pacific Grove, CA).

Clemen, Robert T., and Canan Ulu, 2008, Interior additivity and subjective probability assessment of continuous variables, *Management Science* 54, 835-851.

CME, 2005, *An Introduction to CME Economic Derivatives* (Chicago Mercantile Exchange, Chicago et al.).

Copeland, Thomas E., and Daniel Friedman, 1987, The effect of sequential information arrival on asset prices: An experimental study, *Journal of Finance* 42, 763-797.

Copeland, Thomas E., and Daniel Friedman, 1992, The market value of information: Some experimental results, *Journal of Business* 65, 241-266.

Coval, Joshua D., and Tyler Shumway, 2005, Do behavioral biases affect prices?, *Journal of Finance* 60, 1-34.

Dahan, Ely, Andrew W. Lo, Tomaso Poggio, Nicholas Chan, and Adlar Kim, 2007, Securities trading of concepts (STOC), Working paper, UCLA Anderson School.

Dahan, Ely, Arina Soukhoroukova, and Martin Spann, 2007, Preference markets: Organizing securities markets for opinion surveys with infinite scalability, Working paper, UCLA Anderson School, University of Passau.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839-1885.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 56, 921-965.

Davis, Douglas D., and Charles A. Holt, 1993, *Experimental Economics* (Princeton University Press, Princeton, NJ).

De Bondt, Werner F. M., and Richard Thaler, 1985, Does the stock market overreact?, *Journal of Finance* 40, 793-805.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldman, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703-738.

DeJong, Douglas V., Robert Forsythe, Russell J. Lundholm, and Susan G. Watts, 1992, Do prices convey information? Further empirical evidence, *Research in Experimental Economics* 5, 61-79.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini, 1998, Recent developments in modeling unforeseen contingencies, *European Economic Review* 42, 523-542.

DellaVigna, Stefano, 2009, Psychology and economics: Evidence from the field, *Journal of Economic Literature*, forthcoming.

Dubé-Rioux, Laurette, and J. Edward Russo, 1988, An availability bias in professional judgement, *Journal of Behavioral Decision Making* 1, 223-237.

Duh, Rong Ruey, and Shyam Sunder, 1986, Incentives, learning and processing of information in a market environment: An examination of the base-rate fallacy, in Shane Moriarty, ed.: *Laboratory Market Research* (Norman, OK).

Eisenführ, Franz, and Martin Weber, 2003, *Rationales Entscheiden* (Springer, Berlin et al.).

Ellsberg, Daniel, 1961, Risk, ambiguity, and the Savage axioms, *The Quarterly Journal of Economics* 75, 643-669.

Epley, Nicholas, and Thomas Gilovich, 2001, Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors, *Psychological Science* 12, 391-396.

Falk, Armin, and Ernst Fehr, 2003, Why labour market experiments?, *Labour Economics* 10, 399-406.

Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383-417.

Fama, Eugene F., 1991, Efficient Capital Markets: II, *Journal of Finance* 46, 1575-1617.

Fama, Eugene F., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283-306.

Fehr, Ernst, and Jean-Robert Tyran, 2005, Individual irrationality and aggregate outcomes, *Journal of Economic Perspectives* 19, 43-66.

Filippov, Andrei, 2005, *Macroeconomic Derivatives: Overview and Sovereign Debt Market Applications*, Master Thesis, Norges Handelshoyskole, Bergen.

Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein, 1978, Fault trees: Sensitivity of estimated failure probabilities to problem representation, *Journal of Experimental Psychology: Human Perception and Performance* 4, 330-344.

Fishburn, Peter C., 1982, Nontransitive measurable utility, *Journal of Mathematical Psychology* 26, 31-67.

Fishburn, Peter C., 1984, SSB utility theory and decision-making under uncertainty, *Mathematical Social Sciences* 8, 253-285.

Forsythe, Robert , Thomas A. Rietz, and Thomas W. Ross, 1999, Wishes, expectations and actions: A survey on price formation in election stock markets, *Journal of Economic Behavior & Organization* 39, 83-110.

Forsythe, Robert, Murray Frank, Vasu Krishnamurthy, and Thomas W. Ross, 1995, Using market prices to predict election results: The 1993 UBC election stock market, *Canadian Journal of Economics* 28, 770-794.

Forsythe, Robert, Murray Frank, Vasu Krishnamurthy, and Thomas W. Ross, 1998, Markets as predictors of election outcomes: Campaign events and judgement bias in the 1993 UBC election stock market, *Canadian Public Policy* 24, 329-351.

Forsythe, Robert, and Russell J. Lundholm, 1990, Information aggregation in an experimental market, *Econometrica* 58, 309-347.

Forsythe, Robert, Forrest Nelson, George R. Neumann, and Jack Wright, 1992, Anatomy of an experimental political stock market, *American Economic Review* 82, 1142-1161.

Forsythe, Robert, Forrest Nelson, George R. Neumann, and Jack Wright, 1994, The 1992 Iowa political stock market: September forecasts, *The Political Methodologist* 5, 15-19.

Forsythe, Robert, Thomas R. Palfrey, and Charles R. Plott, 1982, Asset valuation in an experimental market, *Econometrica* 50, 537-567.

Fox, Craig R., 1999, Strength of evidence, judged probability, and choice under uncertainty, *Cognitive Psychology* 38, 167-189.

Fox, Craig R., David Bardolet, and Daniel Lieb, 2005, Partition dependence in decision analysis, resource allocation, and consumer choice, in Rami Zwick, and Amnon Rapoport, eds.: *Experimental Business Research: Marketing, Accounting and Cognitive Perspectives. Volume 3* (Springer, Dordrecht and New York).

Fox, Craig R., and Richard Birke, 2002, Forecasting trial outcomes: Lawyers assign higher probability to possibilities that are described in greater detail, *Law and Human Behavior* 26, 159-173.

Fox, Craig R., and Robert T. Clemen, 2005, Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior, *Management Science* 51, 1417-1432.

Fox, Craig R., and Jonathan Levav, 2004, Partition–edit–count: Naive extensional reasoning in judgment of conditional probability, *Journal of Experimental Psychology* 133, 626-642.

Fox, Craig R., Rebecca K. Ratner, and Daniel Lieb, 2005, How subjective grouping of options influences choice and allocation: Diversification bias and the phenomenon of partition dependence, *Journal of Experimental Psychology* 134, 538-551.

Fox, Craig R., Brett A. Rogers, and Amos Tversky, 1996, Options traders exhibit subadditive decision weights, *Journal of Risk and Uncertainty* 13, 5-17.

Fox, Craig R., and Yuval Rottenstreich, 2003, Partition priming in judgment under uncertainty, *Psychological Science* 14, 195-200.

Fox, Craig R., and Kelly E. See, 2003, Belief and preference in decision under uncertainty, in David Hardman, and Laura Macchi, eds.: *Thinking: Psychological perspectives on reasoning, judgment and decision making* (John Wiley & Sons, Ltd., New York).

Fox, Craig R., and Amos Tversky, 1998, A belief-based account of decision under uncertainty, *Management Science* 44, 879-895.

Friedman, Daniel, 1998, Monty Hall's three doors: Construction and deconstruction of a choice anomaly, *The American Economic Review* 88, 933-946.

Friedman, Daniel, and Shyam Sunder, 1994, *Experimental Methods: A Primer for Economists* (Cambridge University Press, Cambridge, U.K. and New York).

Gadanecz, Blaise, Richhild Moessner, and Christian Upper, 2007, Economic derivatives, *BIS Quarterly Review*, March, 69-81.

Ganguly, Ananda R., John H. Kagel, and Donald V. Moser, 1994, The effects of biases in probability judgments on market prices, *Accounting, Organizations and Society* 19, 675-700.

Ganguly, Ananda R., John H. Kagel, and Donald V. Moser, 2000, Do asset market prices reflect traders' judgment biases?, *Journal of Risk and Uncertainty* 20, 219-245.

Gervais, Simon, and Terrance Odean, 2001, Learning to be overconfident, *The Review of Financial Studies* 14, 1-27.

Gilovich, Thomas, Dale W. Griffin, and Daniel Kahneman, 2002, *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge University Press, Cambridge, UK).

Gjerstad, Steven, 2005, Risk aversion, beliefs, and prediction market equilibrium, Working paper, Economic Science Laboratory, University of Arizona.

Grossman, Sanford J., 1976, On the efficiency of competitive stock markets where trades have diverse information, *Journal of Finance* 31, 573-585.

Grossman, Sanford J., 1981, An introduction to the theory of rational expectations under asymmetric information, *Review of Economic Studies* 48, 541-559.

Grossman, Sanford J., and Joseph E. Stiglitz, 1976, Information and competitive price systems, *American Economic Review* 66, 246-253.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393-408.

Gürkaynak, Refet, and Justin Wolfers, 2006, Macroeconomic derivatives: An initial analysis of market-based macro forecasts, uncertainty, and risk, Working paper, National Bureau of Economic Research.

Harrison, Glenn W., and John A. List, 2004, Field experiments, *Journal of Economic Literature* 42, 1009-1055.

Hayek, F. A., 1945, The use of knowledge in society, *American Economic Review* 35, 519-530.

Hayek, F. A., 1946, The meaning of competition, in F. A. Hayek, ed.: *Individualism and Economic Order* (Routledge and Kegan Paul (1976), London).

Hellwig, Martin F., 1980, On the aggregation of information in competitive markets, *Journal of Economic Theory* 22, 477-498.

Herstein, I. N., and John Milnor, 1953, An axiomatic approach to measurable utility, *Econometrica* 21, 291-297.

Hertwig, Ralph, and Andreas Ortmann, 2001, Experimental practices in economics: A methodological challenge for psychologists?, *Behavioral and Brain Sciences* 24, 383-451.

Hirshleifer, David, 2001, Investor psychology and asset pricing, *Journal of Finance* 56, 1533-1597.

Hirshleifer, David, and Siew H. Teoh, 2003, Herd behaviour and cascading in capital markets: A review and synthesis, *European Financial Management* 9, 25-66.

Hirt, Edward R., and N. John Castellan, 1988, Probability and category redefinition in the fault tree paradigm, *Journal of Experimental Psychology: Human Perception and Performance* 14, 122-131.

Ho, Teck-Hua, and Kay-Yut Chen, 2007, New product blockbusters: The magic and science of prediction markets, *California Management Review* 50, 144-158.

Huber, Jürgen, Michael Kirchler, and Matthias Sutter, 2006, Vom Nutzen zusätzlicher Information auf Märkten mit unterschiedlich informierten Händlern - Eine experimentelle Studie, *Zeitschrift für Betriebswirtschaftliche Forschung* 58, 188-211.

Hull, John C., 2009, *Options, Futures and Other Derivatives* (Pearson Education, Inc., Upper Saddle River, NJ).

Humphrey, Steven J., 1996, Do anchoring effects underlie event-splitting effects? An experimental test, *Economic Letters* 51, 303-308.

Jaccard, James, 1998, *Interaction Effects in Factorial Analysis of Variance* (Sara Miller McCune, Sage Publications Inc., Thousand Oaks et al.).

Johnson, Eric J., John Hershey, Jacqueline Meszaros, and Howard Kunreuther, 1993, Framing, probability distortions, and insurance decisions, *Journal of Risk and Uncertainty* 7, 35-51.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler, 1991, The endowment effect, loss aversion, and status quo bias, *Journal of Economic Perspectives* 5, 193-206.

Kahneman, Daniel, Paul Slovic, and Amos Tversky, 1982, *Judgment under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge, UK).

Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263-291.

Keynes, John M., 1936, *The General Theory of Employment, Interest and Money* (MacMillan and Co., Ltd., London et al.).

Kluger, Brian D., and Steve B. Wyatt, 2004, Are judgment errors reflected in market prices and allocations? Experimental evidence based on the Monty Hall problem, *Journal of Finance* 59, 969-997.

Knetsch, Jack L., 1989, The endowment effect and evidence of nonreversible indifference curves, *American Economic Review* 79, 1277-1284.

Koehler, Derek J., Lyle A. Brenner, and Amos Tversky, 1997, The enhancement effect in probability judgment, *Journal of Behavioral Decision Making* 10, 293-313.

Kreps, David M., 1979, A representation theorem for "preference for flexibility", *Econometrica* 47, 565-577.

Kruskal, William H., and W. Allen Wallis, 1952, Use of ranks in one-criterion variance analysis, *Journal of the American Statistical Association* 47, 583-621.

Kuon, Bettina, 1991, Typical trading behavior in the German election markets 1990, Discussion paper, University of Bonn.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1335.

Kyle, Albert S., 1989, Informed speculation with imperfect competition, *Review of Economic Studies* 56, 317-356.

Lange, Jeffrey, and Nicholas Economides, 2005, A parimutuel market microstructure for contingent claims, *European Financial Management* 11, 25-49.

Langer, Thomas, and Craig R. Fox, 2005, Biases in allocation under risk and uncertainty: Partition dependence, unit dependence, and procedure dependence, Working paper, University of Muenster, University of California at Los Angeles.

Levitt, Steven D., and John A. List, 2008, Homo economicus evolves, *Science* 319, 909-910.

List, John A., 2006, Field experiments: A bridge between lab and naturally occurring data, *Advances in Economic Analysis & Policy* 6, Issue 2, Article 8.

List, John A., and David Reiley, 2008, Field experiments, in Steven Durlauf, and Lawrence Blume, eds.: *The New Palgrave Dictionary of Economics* (Palgrave Macmillan, Basingstoke).

Loomes, Graham, and Robert Sugden, 1982, Regret theory: An alternative theory of rational choice under uncertainty, *The Economic Journal* 92, 805-824.

Lovell, Michael C., and Robert C. Vogel, 1973, A CPI-Futures market, *Journal of Political Economy* 81, 1009-1012.

Manski, Charles F., 2006, Interpreting the predictions of prediction markets, *Economic Letters* 91, 425-429.

Marshall, John F., Vipul K. Bansal, Anthony F. Herbst, and Alan L. Tucker, 1992, Hedging business cycle risk with macro swaps and options, *Journal of Applied Corporate Finance* 4, 103-108.

McNeil, Barbara J., Stephen G. Pauker, Harold C. Sox, and Amos Tversky, 1982, On the elicitation of preferences for alternative therapies, *The New England Journal of Medicine* 306, 1259-62.

Milgrom, Paul, and Nancy Stokey, 1982, Information, trade and common knowledge, *Journal of Economic Theory* 26, 17-27.

Morgan, John, and Phillip C. Stocken, 2008, Information aggregation in polls, *American Economic Review* 98, 864-896.

Morwitz, Vicki G., Eric Johnson, and David Schmittlein, 1993, Does measuring intent change behavior?, *The Journal of Consumer Research* 20, 46-61.

Murauer, Gerald, 1997, Vergleich von Wahlprognosen durch Meinungsumfragen und Bewertung durch Börsen anhand der Oberösterreichischen Landtagswahl 1997, Diploma thesis, Technische Universität Vienna.

Nöth, Markus, and Martin Weber, 1996, Insidererkennung in experimentellen Märkten, *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* 48, 959-982.

O'Brien, John, and Sanjay Srivastava, 1991, Dynamic stock markets with multiple assets: An experimental analysis, *Journal of Finance* 46, 1811-1838.

Ofir, Chezy, 2000, Ease of recall vs recalled evidence in judgment: Experts vs laymen, *Organizational Behavior and Human Decision Processes* 81, 28-42.

Oliven, Kenneth , and Thomas A. Rietz, 2004, Suckers are born but markets are made: Individual rationality, arbitrage, and market efficiency on an electronic futures market, *Management Science* 50, 336-351.

Ortner, Gerhard, 1997, Forecasting markets - An industrial application, part I, Working paper, Technical University of Vienna, Department of Managerial Economics and Industrial Organization.

Ortner, Gerhard, 1998, Forecasting markets - An industrial application, part II, Working paper, Technical University of Vienna, Department of Managerial Economics and Industrial Organization.

Ortner, Gerhard, Adolf Stepan, and Josef Zechner, 1995, Political stock markets - The Austrian experience, *Zeitschrift für Betriebswirtschaft, Ergänzungsband 4/95 - Managment of Structual Change,* 123-136.

Ottaviani, Marco, and Peter Norman Sørensen, 2007, Aggregation of information and beliefs in prediction markets, Working paper, London Business School, University of Copenhagen.

Page, Susan, 2003, Confronting Iraq (USA Today), http://www.usatoday.com/educate/war28-article.htm, September 06, 2007.

Parker, John, 2007, The impact of economic news on financial markets, MPRA Working paper No. 2675.

Pennock, David M., Finn Arup Nielsen, and C. Lee Giles, 2001, Extracting collective probabilistic forecasts from web games, 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2001) (San Francisco, CA).

Plott, Charles R., and Kevin A. Roust, 2005, The design of information aggregation mechanisms: A two-Stage parimutuel market to avoid mirages bubbles, Working paper, California Institute of Technology.

Plott, Charles R., and Shyam Sunder, 1988, Rational expectations and the aggregation of diverse information in laboratory security markets, *Econometrica* 56, 1085-1118.

Plott, Charles R., and Shyman Sunder, 1982, Efficiency of experimental security markets with insider information: An application of rational-expectations models, *Journal of Political Economy* 90, 663-697.

Redelmeier, Donald A., Derek J. Koehler, Varda Liberman, and Amos Tversky, 1995, Probability judgment in medicine: Discounting unspecified possibilities, *Medical Decision Making* 15, 227-230.

Rosenbloom, E. S., and William Notz, 2006, Statistical tests of real-money versus play-money prediction markets, *Electronic Markets* 16, 63-69.

Rottenstreich, Yuval, and Amos Tversky, 1997, Unpacking, repacking, and anchoring: Advances in support theory, *Psychological Review* 104, 406-415.

Royal Swedish Academy of Sciences, 2002, Foundations of Behavioral and Experimental Economics: Daniel Kahneman and Vernon Smith — Advanced Information on the Prize in Economic Sciences 2002.

Rubinstein, Mark, 2001, Rational markets: Yes or no? The affirmative case, *Financial Analysts Journal* 57, 15-29.

Russell, Thomas, and Richard Thaler, 1985, The relevance of quasi rationality in competitive markets, *American Economic Review* 75, 1071-1082.

Russo, J. Edward, and Karen J. Kolzow, 1994, Where is the fault in fault trees?, *Journal of Experimental Psychology: Human Perception and Performance* 20, 17-32.

Samuelson, Paul A., and William D. Nordhaus, 1985, *Economics* (McGraw-Hill, New York et al.).

Savage, Leonard J., 1954, *The Foundations of Statistics* (J. Wiley & Sons, New York).

Scharfstein, David S., and Jeremy C. Stein, 2000, The dark side of internal capital markets: Divisional rent-seeking and inefficient investment, *Journal of Finance* 55, 2537-2564.

See, Kelly E., Craig R. Fox, and Yuval S. Rottenstreich, 2006, Between ignorance and truth: Partition dependence and learning in judgment under uncertainty, *Journal of Experimental Psychology / Learning, Memory & Cognition* 32, 1385-1402.

Servan-Schreiber, Emile, Justin Wolfers, David M. Pennock, and Brian Galebach, 2004, Prediction markets: Does money matter?, *Electronic Markets* 14, 243-251.

Sher, Shlomi, and Craig R. M. McKenzie, 2006, Information leakage from logically equivalent frames, *Cognition* 101, 467-494.

Shiller, Robert J., 1993, *Macro Markets. Creating Institutions for Managing Society's Largest Economic Risks* (Oxford University Press Inc., New York).

Shiller, Robert J., 2003, *The New Financial Order. Risk in the 21st Century* (Princeton University Press, Princeton, NJ).

Shleifer, Andrei, 2000, *Inefficient markets* (Oxford University Press, Oxford).

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35-55.

Sloman, Steven A., Edward J. Wisniewski, Yuval Rottenstreich, and Constantinos Hadjichristidis, 2004, Typical versus atypical unpacking and superadditive probability judgment, *Journal of Experimental Psychology* 30, 573–582.

Smith, Vernon L., 1976, Experimental economics: Induced value theory, *American Economic Review* 66, 274-279.

Smith, Vernon L., 1982a, Markets as economizers of information: Experimental examination of the 'Hayek hypothesis', *Economic Inquiry* 20, 165-179.

Smith, Vernon L., 1982b, Microeconomic systems as an experimental science, *American Economic Review* 72, 923-955.

Sobel, Russel S., and S. Travis Raines, 2003, An examination of the empirical derivatives of the favourite–longshot bias in racetrack betting, *Applied Economics* 35, 371–385.

Sonnemann, Ulrich, Colin F. Camerer, Craig R. Fox, and Thomas Langer, 2008, Partition-dependent framing effects in lab and field prediction markets, Working paper, University of Muenster, Caltech, UCLA.

Spetzler, Carl S., and Carl-Axel S. Staël Von Holstein, 1975, Probability encoding in decision analysis, *Management Science* 22, 340-358.

Sunder, Shyam, 1992, Market for information: Experimental evidence, *Econometrica* 60, 667-695.

Sunder, Shyam, 1995, Experimental asset markets: A survey, in John H. Kagel, and Alvin E. Roth, eds.: *The Handbook of Experimental Economics* (University Press, Princeton).

Surowiecki, James, 2005, *The Wisdom of Crowds* (Anchor Books, New York).

Teigen, Karl Halvor, 1974, Subjective sampling distributions and the additivity of estimates, *Scandinavian journal of psychology* 15, 50-55.

Thaler, Richard H., 1980, Toward a positive theory of consumer choice, *Journal of Economic Behavior & Organization* 1, 39-60.

Thaler, Richard H., and William T. Ziemba, 1988, Parimutuel betting markets: Racetracks and lotteries, *Journal of Economic Perspectives* 2, 161-174.

Tversky, Amos, 1996, Contrasting rational and psychological principles of choice, in R. J. Zeckhauser, R. L. Keeney, and J. K. Sebenius, eds.: *Wise Choices* (Harvard Business School Press, Boston, MA).

Tversky, Amos, and Craig R. Fox, 1995, Weighing risk and uncertainty, *Psychological Review* 102, 269-283.

Tversky, Amos, and Daniel Kahneman, 1973, Availability: A heuristic for judging frequency and probability, *Cognitive Psychology* 5, 207-232.

Tversky, Amos, and Daniel Kahneman, 1983, Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment, *Psychological review* 90, 293-315.

Tversky, Amos, and Daniel Kahneman, 1986, Rational choice and the framing of decisions, *Journal of Business* 59, 251-278.

Tversky, Amos, and Daniel Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty* 5, 297-323.

Tversky, Amos, and Derek J. Koehler, 1994, Support theory: A nonextensional representation of subjective probability, *Psychological Review* 101, 547-567.

Tziralis, Georgios, and Ilias Tatsiopoulos, 2007, Prediction markets: An extended literature review, *The Journal of Prediction Markets* 1, 75–91.

van der Pligt, J., J.R. Eiser, and R. Speark, 1987, Comparative judgments and preferences: The influence of the number of response alternatives, *British Journal of Social Psychology* 26, 269-280.

Verrecchia, Robert E., 1982, Information acquisition in a noisy rational expectations economy, *Econometrica* 50, 1415-1430.

von Neumann, John, and Oskar Morgenstern, 1944, *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, N.J.).

Watts, Susan, 1993, Private information, prices, asset allocations and profits: Further experimental evidence, Vol. 5, in Mark Isaac, ed.: *Research in Experimental Economics* (JAI Press, Greenwich, CT).

Wilcoxon, Frank, 1945, Individual comparisons by ranking methods, *Biometrics Bulletin* 1, 80-83.

Wolfers, Justin, and Andrew Leigh, 2002, Three tools for forecasting federal elections: Lessons from 2001, *Australian Journal of Political Science* 37, 223-240.

Wolfers, Justin, and Eric Zitzewitz, 2004, Prediction markets, *Journal of Economic Perspectives* 18, 107-126.

Wolfers, Justin, and Eric Zitzewitz, 2006, Five open questions about prediction markets, Discussion paper, No. 1975, Institute for the Study of Labor (IZA).

Wolfers, Justin, and Eric Zitzewitz, 2007, Interpreting prediction market prices as probabilities, Working paper, The Wharton School (University of Pennsylvania) and Stanford GSB.

Wolfers, Justin, and Eric Zitzewitz, 2008, Prediction markets in theory and practice, in Steven Durlauf, and Lawrence Blume, eds.: *The New Palgrave Dictionary of Economics* (Palgrave Macmillan, Basingstoke).