Econometric analysis of individual income dynamics

Inauguraldissertation
zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften
durch die

Wirtschaftswissenschaftliche Fakultät
der
Westfälischen Wilhelms-Universität Münster

vorgelegt von Diplom-Volkswirtin
Sarah Meyer
aus Münster

Dekanin: Prof. Dr. Theresia Theurl

Erstgutachter: Prof. Dr. Mark Trede

Zweitgutachter: Prof. Dr. Bernd Wilfling

Tag der mündlichen Prüfung: 11.11.2015

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Labor income always received a lot of attention in the economics literature. As early as 1776, Adam Smith analyzed the determinants of wage differentials among individuals and among employments in his Magnum Opus *The Wealth of Nations*. To date, a large amount of studies has analyzed the structure of wages. While early quantitative work examined differences and changes in wages by occupation (blue vs. white collar workers, Douglas et al. (1930)) and by industry (Slichter (1950)), the human capital revolution in the 1960s and 1970s shifted emphasis to the differences in wages by education and age, and therefore by potential experience. Indeed, models of life-cycle earnings that arise from educational and on-the-job training investments (Ben-Porath (1967), Mincer et al. (1974)) provide an explanation of the timeless qualitative features of wage structures that have been found in most data sets and across countries: more educated workers receive higher earnings, and the age-earnings profiles are upward sloping and concave (Willis (1987)). However, in quantitative terms, the wage structure has been subject to substantial changes in the last decades. In particular since the 1980s, wage inequality and educational wage differentials have increased. This pattern induced a great variety of studies which examine the changes in the wage structure and in earnings inequality, especially for the U.S. (Gottschalk and Moffitt (1994), Katz et al. (1999), Atkinson et al. (2011), among many others). Besides, these studies benefit from large-scale micro data sets which became increasingly available during the last decades. However, all studies mentioned so far aim at explaining earnings inequality and therefore deal with earnings in terms of an aggregate economic level.

A different strand of literature that arose in the late 1970s approaches earnings from the individuals' perspectives. Properly modeling individuals' earnings dynamics is important as it forms the basis for a wide variety of studies. These include studies on the modeling of labor supply (Abowd and Card (1989)), on the prediction of future earnings paths given individual information (Chamberlain and Hirano (1999)), and on earnings inequality, when modeling the time series variation in the earnings distribution (Moffitt and Gottschalk (2002)). Moreover, correctly modeling individual earnings is crucial for the determination of the earnings risk that both individuals and households face.

Various dynamic models on individual incomes have been proposed (Lillard and Weiss (1979), Baker (1997), Guvenen (2009), Hryshko (2012)). To date, two major strands on modeling individual earnings dynamics have prevailed in the literature. They mainly differ in their assumptions on the persistence of income shocks, on the degree of heterogeneity of the individuals and therefore implicitly on the amount of earnings risk that individuals face. Interestingly, the literature has not yet reached a consensus about how to best or even "correctly" model earnings dynamics.

The present thesis extends the existing literature on individuals' earnings dynamics in three aspects. Starting from a general point of view, an alternative approach to the identification of earnings risk is presented. Its main advantage is the fact that it does not only use earnings data to gather information on earnings risk. It moreover includes the economic consumption and portfolio allocation choices of individuals. This approach is even applicable when consumption data are not available. Instead, it can then employ realized capital income data as this results from those very two choices.

Moreover, this thesis ties in with the disaccord in the literature on how to most suitably model individual income dynamics. For this purpose, this thesis first addresses the question if explosiveness needs to be taken into account when modeling income profiles - a question that has been ruled out in previous studies. Exploring whether some incomes evolve in an explosive way accounts for the idea that deviations of earnings from a common trend could be self-reinforcing. To this end, a right-tailed panel unit root test is proposed. Even though many models exist that test for unit roots in panel data, no panel test against explosiveness is known heretofore. Furthermore, this thesis provides new evidence with respect to the two contrasting strands of the literature outlined above. To the best of our knowledge, we are the first to employ a dynamic linear model and to allow for time-varying coefficients. Using Markov chain Monte Carlo (MCMC) methodology, the results provide evidence against one of the two literature strands.

This thesis proceeds as follows. Chapter 2 suggests a method that uses not only earnings data to elicit the amount of earnings risk, but also considers the joint dynamics of capital income and earnings. Since capital income reflects the amount of savings, it contains important information about consumption and savings behavior and therefore implicitly about earnings risk. For this purpose, a life-cycle model of consumption and savings decisions, stated as a dynamic programming problem, is presented. Estimation of the unknown parameters of the earnings process and, if required, of further model parameters is carried out by indirect inference. Simulation results show that the estimates are centered near the true values. The estimation method is variably applicable since it can easily be adapted to estimate different sets of unknown parameters or include different information available on capital income.

Chapter 3 investigates whether explosiveness is a pattern that needs to be taken into account when modeling income profiles. In this context, explosiveness implies that high income trajectories tend to detach from the common trend disproportionately high and vice versa. To this end, a panel test against explosiveness is proposed and applied to German and U.S. earnings data. The null hypothesis of no explosiveness is rejected. However, the proportion of explosive profiles is small. Hence, explosiveness does not need to be considered when modeling labor incomes in the following chapter.

Chapter 4 delves more into the details of how to accurately model labor income profiles. The first part of this chapter provides an overview of the empirical literature on modeling labor incomes in order to depict the two main opposing approaches. To contribute to the disaccord of the literature on these two approaches, a dynamic linear model is proposed in the second part. It allows for both individual-specific and time-varying coefficients. Estimation of the unknown model parameters is carried out using Gibbs sampling, a Markov chain Monte Carlo (MCMC) algorithm. The framework is applied to German earnings data. The key finding of this chapter is the fact that one of the two competing approaches is rejected by the framework. The rejected approach, in particular, assumes that individuals are subject to large and persistent income shocks, while there exist no systematic differences between income profiles.

Chapter 5 summarizes the main results of this thesis. Note that for the implementation of the econometric methods and the empirical applications the software R is used.

Chapter 2

Eliciting earnings risk from labor and capital income

2.1 Introduction

Earnings risk plays a central role in many economic decisions that individuals make. Hence, numerous studies deal with the effects of earnings risk on various economic variables. These include studies which investigate its impact on life-cycle consumption and portfolio allocation (Carroll and Samwick (1997), Campbell et al. (2001)) and studies on how earnings uncertainty affects wealth distribution over the life-cycle (Huggett (1996), Castaneda et al. (2003)). Beyond, its characteristics serve as a basis for modeling earnings dynamics. Two leading views on modeling earnings dynamics have been established in the current literature, which mainly differ in their assumptions on the degree of earnings risk that individuals face. The first view assumes that earnings risk is rather weak and income follows a deterministic path over the life-cycle with shocks of moderate persistence (Lillard and Weiss (1979), Baker (1997), Guvenen (2009)). The other view, however, assumes earnings shocks to have a unit root, therefore having a large and persistent effect on the rest of an individual's working life (MaCurdy (1982), Abowd and Card (1989), Hryshko (2012)). Therefore, quantifying earnings risk is of interest.

Although there is a large range of literature on earnings risk, most of it lacks inclusion of the fact that individuals do indeed have superior information on their very earnings risk. Therefore, Guvenen and Smith (2014) have suggested an estimation method for earnings risk that takes into account economic choices of the individuals. Obviously, the most important economic choice is the decision about the consumption level. If individuals feel insecure about their future earnings, they tend to consume less and, instead, save more to build up a buffer against negative future shocks, see e.g. Carroll (2004) on the theory of buffer saving. Beyond, when feeling insecure, individuals tend to invest more into safe assets rather than into risky ones. Therefore, earnings risk also affects portfolio allocation decisions. However, from an econometric point of view, estimation methods that need all these data are infeasible. There exist hardly any panels providing information jointly on earnings and consumption on an individual level over longer time spans.

Therefore, our main contribution is the suggestion of a method to measure earnings risk from individual earnings and capital income data, whereas further data such as consumption panel data are not required. By doing so, our approach takes into account different sources of capital income, particularly income from risk-free assets and from risky assets. Supposing that an individual knows his earnings risk, our method uses not only information from observing the earnings themselves, but also from the realized capital income resulting from consumption and portfolio allocation decisions. We use this variety of information in order to draw conclusions on the earnings risk, which we expect to increase the estimation accuracy compared with methods that only rely on information from earnings trajectories. Another benefit of our method is that it allows us to estimate the risk aversion of the individual and, if required, further unknown model parameters.

In particular, our model is based on the classical dynamic stochastic optimization described by Samuelson (1969). While the Samuelson model ignores labor income, we assume that earnings are an (exogenous) stochastic process which makes the dynamic optimization problem more complex and analytically intractable. Moreover, we model both the working life of individuals, as well as their retirement. To estimate the parameters of such a life-cycle model, we closely follow Gourinchas and Parker (2002) and Guvenen and Smith (2014). The former authors estimate a life-cycle model of consumption and savings using simulated moments. However, they first estimate income process parameters from earnings data and then the parameter of risk aversion from consumption data. Therefore, they do not use information from individuals' consumption behavior when estimating earnings risk. Guvenen and Smith (2014) estimate all parameters by indirect inference, jointly using both data sources. Their model is much more complex than the one we propose here. Among other things, they allow the individuals to learn about their earnings paths in a Bayesian way, such that the econometrician needs to handle a cumbersome variety of unknown parameters. Following their suggestion, we use indirect inference to estimate the parameters of our model.

The remainder of this chapter is organized as follows. Section 2.2 introduces stochastic dynamic optimization and describes the economic model which is based on such a framework. In Section 2.3 we briefly present the simulation-based estimation method of indirect inference and explain how we use it to estimate the parameters of interest. The results of a simulation study are reported in Section 2.4. Finally, Section 2.5 concludes.

2.2 Stochastic dynamic programming framework

This section first introduces the basic ideas of stochastic dynamic programming, in particular for the case when the number of discrete time periods is finite. This method enables us to solve our life-cycle model of consumption and investment decisions, which is stated as a dynamic programming problem. Its formulation and a short literature overview on life-cycle models are included in a second subsection.

2.2.1 Introduction to dynamic programming

"An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

Principle of optimality, Richard Bellman, 1957

Many economic applications aim at analyzing optimal behaviour under certain constraints. In simple problems, it is often assumed that only the present period is affected by the choice of an agent. This, however, turns out to be a severe restriction since it assumes that all variables of an economic system (called *state variables*) adjust within the present period, therefore excluding any possible dynamic effects of a choice. This restriction can not be maintained for most real world applications, meaning that intertemporal effects have to be taken into account when deciding on the action. Hence, in many cases it is sensible to formulate economic applications as dynamic programming problems. Since our application in the subsequent sections is based on a finite number of periods, we introduce the formulation and solution of dynamic programming problems in discrete time with finite time horizon.

Consider an individual with a concave contemporaneous utility function $r(x_t, u_t)$ that depends on both the state x_t and the control u_t, which is the action taken by the individual. For simplicity, take both to be scalars, however, they could be extended to vectors. The state variable is driven by a stochastic difference equation

$$x_{t+1} = g(x_t, u_t, \varepsilon_{t+1}), \tag{2.1}$$

which is referred to as the *transition equation*.

At any period t, assume the value of x_t to be known. Future values, however, are random variables since they are a function of the current state x_t and the action u_t, but also of a stochastic shock ε_{t+1}. Assume that the process is Markov, thus all (past and current) information relevant for the determination of the distribution of the future states is summarized in the current information set.

In general, the individual's *objective* is to choose the sequence of controls $\{u_0^T\}$ so as to maximize the *expected* sum of its discounted contemporaneous utilities

$$E_0 \sum_{t=0}^{T} \beta^t r(x_t, u_t), \qquad 0 < \beta < 1, \tag{2.2}$$

subject to (2.1). In doing so, assume that the initial state x_0 is known and the error terms ε_t to be independently and identically distributed over time, being independent of preceding states and controls. To solve this problem, we aim at finding a sequence of optimal policies $\{u_t^*\}$, where u_t^* maps state x_t into control u_t,

$$u_t = u_t^*(x_t), \tag{2.3}$$

so that the objective function (2.2) is maximized for any given state and period.

Denote by $V_s(x_s)$ the *value function* which is the objective function, starting from state x_s in period s. Hence

$$V_s(x_s) = \max_{\{u_t\}_{t=s}^{T}} \left[\sum_{t=s}^{T} \beta^t r(x_t, u_t) \right]. \tag{2.4}$$

According to the *principle of optimality* by Richard Bellman, any optimal solution consists of optimal partial solutions. Roughly speaking, this means assuming $\{u_0^*, \ldots, u_T^*\}$ to be the optimal policy of a dynamic problem, then for the subproblem starting at time i, the truncated policy $\{u_i^*, u_{i+1}^*, \ldots, u_T^*\}$ is optimal. This implies that the dynamic optimization problem can be broken down to simpler subproblems. For each period, the decision sequence can be split into two parts, the actual, present period and the entire continuation beyond.

The optimal choice of control u_t in period t is the one which maximizes the sum of the instantaneous utility and the discounted continuation value, assuming optimal decisions in the future. The result, by definition, will be the value function $V_t(x_t)$. This connection yields the *Bellman equation*

$$V_t(x_t) = \max_{u_t} \left\{ r(x_t, u_t) + \beta \cdot \mathbb{E}_t V_{t+1}[g(x_t, u_t, \varepsilon_{t+1})] \right\}, \tag{2.5}$$

which reformulates the optimization problem in a recursive way. Moreover, it is rather easy to handle since only the immediate control u_t is to be chosen optimally, while future optimal choices u_{t+1}, u_{t+2}, etc. are collected in the continuation value. Hence, the dynamic programming approach substantially simplifies the optimization problem, decomposing the problem into a sequence of optimizations, each over the control, rather than optimizing over a whole set of policies at once.

As the time horizon is finite, the dynamic optimization problem can be approached by backward recursion from the very last period. The terminal value function in T is specified as the sum of the contemporaneous utility and some termination value $V_{T+1}(x_{T+1})$, which is often assumed to be zero. Even if it is not zero, the terminal value function is known apart from the realization of the state in $T+1$. Hence, starting in period T, a natural approach is to discretize a continuous state variable into D levels $\{x_T^d\}_{d=1,\ldots,D}$. The optimization problem in T is solved for each possible discrete state to obtain the optimal control strategy u_T^{*d} which maximizes

$$V_T^d(x_T^d) = \max_{u_T} \left\{ r(x_T^d, u_T^d) + \beta \cdot \mathbb{E}_T V_{T+1}^d [g(x_T^d, u_T^d, \varepsilon_{T+1})] \right\}.$$

Hence for each state x_T^d, a corresponding value V_T^d is obtained. Since the realized states are continuous, the values are interpolated. The approximated specification of the value function in T allows us to solve the optimization problem in $T-1$ for the discrete states $\{x_{T-1}^d\}_{d=1,\ldots,D}$. This procedure is repeated until all optimizations in period 0 are done. We thus obtain a sequence of optimal policies, only depending on the possible states for each period. By means of these policies and a given initial state x_0, the decision problem can be solved starting in period 0. Note that when implementing this approach, one has to take into account the tradeoff between precision of the optimization, which can be increased by choosing a small grid step size, and computational costs caused by that.

However, in order to save computational effort, especially when dealing with more than one control, another common approach is to approximate the value function by a functional form $V_t^f(x_t)$. Again we proceed backwards starting in T. As in the naive grid-based method introduced before, the optimization problem in T is solved for each possible discrete state $\{x_T^d\}_{d=1,\ldots,D}$. In contrast to the above solution, we no longer save the value V_T^d for each possible state. Instead, the values $\{V_T^d\}_{d=1,\ldots,D}$ are "compressed" by fitting the functional form $V_T^f(x_T)$ by choosing appropriate coefficients which are saved. In some applications, the functional form can even be derived analytically. When the functional form is known, the precision of the optimization is increased compared to the grid-based method since we no longer need to interpolate between states.

2.2.2 The economic model

In the following, we introduce a life-cycle model of consumption and investment decisions, which is stated as a dynamic programming problem. The model's structure is based on Samuelson (1969) and Viceira (2001). Samuelson was the first to formulate the lifetime planning of consumption and investment decisions as a many-period problem. He considers an individual who maximizes the discounted sum of his contemporaneous utilities subject to his initial wealth. Wealth is either consumed or invested. For the fraction of wealth that is invested, the individual faces a portfolio problem: He has to choose the proportion of wealth he wants to invest into a risky asset; while the remainder is invested into a riskfree asset. It should be noted that Samuelson's optimization problem can be solved in closed-form.

Since then, a variety of studies have emerged which build on Samuelson's basic optimization problem. Here, we focus on the most relevant ones for our model specification. Departing from the lack of labor income in Samuelson's work, Bodie et al. (1992), Heaton and Lucas (1997) and Koo (1998) propose to incorporate labor income into the standard intertemporal model of consumption and portfolio choice. Bodie et al. (1992) examine the effect of labor-leisure choice on consumption and portfolio investment decisions over the life-cycle. They find that labor and investment choices are closely related: the ability to vary labor supply ex post tends to induce an individual to take greater investment risks in his portfolio. This is likely to explain why young workers with greater labor supply flexibility over their entire working lifetime should hold proportionally more risky assets in their portfolios. The authors assume future earnings to be non-stochastic or at least perfectly hedgeable. Since future labor income might be uncertain for most individuals, Heaton and Lucas (1997) and Koo (1998) introduce uninsurable labor income risk - combined with several portfolio constraints in their model. However, they assume individuals to work their entire lifetime - ignoring retirement - and only consider the case in which labor income is uncorrelated with asset returns. Viceira (2001) argues that retirement matters for portfolio choice and suggests to incorporate retirement with zero income into the model. Moreover, he allows labor income to be correlated with asset returns such that consumption may be hedged from negative labor income shocks. He finds both risky labor income and retirement to affect the optimal portfolio choice: increasing labor income risk raises the investor's willingness to save and leads him to increase the proportion invested riskfree. Beyond, the optimal fraction of savings invested into the risky asset turns out to be positively related to the retirement horizon of the investor.

To date, many more applications have come up, in particular in the financial literature. We adopt most of the ideas mentioned above in order to formulate a dynamic optimization problem to estimate labor income risk and, additionally, the degree of risk aversion. Specifically, we include both idiosyncratic labor income and retirement. Otherwise we keep our model rather basic since our focus is to provide a framework to infer the parameters related to earnings risk, rather than deducing an optimal asset allocation.

In particular, we consider an individual whose preferences are described by the standard CRRA utility function over consumption C

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma}$$

where $\gamma > 0$, $\gamma \neq 1$ is the measure of relative risk aversion.

The model includes two state variables: wealth W and labor income Y with initial wealth W_0 and initial labor income Y_0. Let S denote the number of periods in which the individual works. We follow Viceira (2001) and model labor income evolving stochastically according to a geometric random walk

$$Y_{t+1} = Y_t \exp\left(\mu_\varepsilon + \theta_\varepsilon \varepsilon_{t+1}\right), \qquad t = 1, \ldots, S \tag{2.6}$$

with parameters μ_ε and θ_ε and a random innovation $\varepsilon_t \sim N(0,1)$.

In each period, the individual decides about two control variables, namely: the level of consumption C_t and the portfolio composition parameter α_t which denotes the proportion of wealth invested into a risky asset with return $Z_t \sim LN(\mu_Z, \theta_Z)$, where $0 \leq \alpha_t \leq 1$.[1] The remaining proportion $(1-\alpha_t)$ is invested riskfree with interest rate r. Accordingly, the budget constraint and therefore the transition equation of wealth is

$$W_{t+1} = \left((1-\alpha_t)(1+r) + \alpha_t Z_{t+1}\right)(W_t + Y_t - C_t). \tag{2.7}$$

After having worked for S periods, the individual retires and stays pensioner until period T, which is the final period of the model. During retirement, labor income becomes zero and the individual instead receives a pension Y_t for $t = (S+1), \ldots, T$,

[1] The random asset returns Z_t and income innovations ε_t may be dependent; however, for ease of illustration we assume that they are uncorrelated.

which depends only on the last salary and is fully predictable. Strictly speaking, the individual is assumed to receive 80% of his last salary in the first retirement period, which thereafter increases by 2% per period, hence $Y_t = (1.02)^{t-(S+1)} \cdot 0.8 \cdot Y_S$. The rest of the model remains unchanged during retirement, hence the individual still decides on consumption and portfolio composition, while wealth still evolves according to the transition equation (2.7). Moreover, our model contains no bequests, meaning that the termination value V_{T+1} is zero, irrespective of the state in $T+1$. From this follows that the optimal amount of consumption in the last period T equals the sum of the pension and the remaining wealth, hence $C_T = Y_T + W_T$ and $W_{T+1} = 0$.

Finally, we assume that the current states are known when deciding on the controls. Thus, the timing of decisions and realization of the random variables is as follows:

$$\ldots \rightarrow \underbrace{\begin{pmatrix} W_t \\ Y_t \end{pmatrix}}_{\text{observe the state}} \rightarrow \underbrace{\begin{pmatrix} \alpha_t \\ C_t \end{pmatrix}}_{\text{decide the control}} \rightarrow \underbrace{\begin{pmatrix} \varepsilon_{t+1} \\ Z_{t+1} \end{pmatrix}}_{\text{realization of r.v.}} \rightarrow \underbrace{\begin{pmatrix} W_{t+1} \\ Y_{t+1} \end{pmatrix}}_{\text{new state}} \rightarrow \ldots$$

Since earnings Y_t are modeled as an exogenous stochastic process, simply observing $Y_1, \ldots, Y_T$ would, of course, already allow the econometrician to estimate the earnings risk parameter θ_ε. However, taking into account additional information about capital income from risk-free and risky assets may increase the precision of the estimator. This is particularly important if the observation period is relatively short. Besides, this allows us to estimate further parameters such as the risk aversion.

An important practical problem is the fact that savings or, equivalently, consumption or wealth, are not reliably observable for the econometrician in most panels. In contrast, many panels provide accurate and detailed information about different sources of income. If individuals feel insecure, they will increase both their savings to build up a buffer and their portion of wealth invested into the risk-free asset. Hence, in an indirect way, an increase in interest income indicates larger earnings risk.

Regardless whether consumption or wealth data are available or not, the intertemporal optimization problem

$$\max_{\alpha_1, C_1, \ldots, \alpha_T, C_T} \mathbb{E}_t \left(\sum_{t=1}^{T} \left(\frac{1}{1+\rho} \right)^t \frac{C_t^{1-\gamma}}{1-\gamma} \right)$$

subject to (2.7) and (2.6) has to be solved prior to implementing any estimation strategy. The parameter ρ is a subjective discount rate. The value function can be stated according to the Bellman equation as

$$V_t\left(W_t, Y_t\right) = \max_{\alpha_t, C_t} \left\{ \frac{C_t^{1-\gamma}}{1-\gamma} + \frac{1}{1+\rho} \mathbb{E}_t \left(V_{t+1}\left(W_{t+1}, Y_{t+1}\right)\right) \right\}. \tag{2.8}$$

While Samuelson's model (1969) excluding labor income can be solved analytically, adding a stochastic labor income process makes the problem much more complex. Zeldes (1989) derives an approximation to the closed-form solution for optimal consumption with CRRA utility and stochastic labor income. Compared with Samuelson's model, he ignores portfolio investment decisions. Instead, the entire wealth is invested with a risk-free interest rate. However, since our model includes decisions on the portfolio allocation, the problem becomes analytically intractable. Numerical optimization methods are required, and therefore, it is important to keep the number of state variables as small as possible.

In the following, we show that the two state variables W_t and Y_t can be collapsed into a single state variable $w_t = W_t/Y_t$, cf. Carroll (2004). Since the time horizon is finite, we start in the last period T and then proceed backwards. As mentioned before, the optimal amount of consumption then equals the sum of the pension income and the remaining wealth. Hence the value function can be written as

$$V_T(W_T, Y_T) = Y_T^{1-\gamma} \frac{(w_T + 1)^{1-\gamma}}{1-\gamma} \tag{2.9}$$

where $w_T = W_T/Y_T$. Then, the transition equation for the single state w_t is

$$\begin{aligned} \frac{W_{t+1}}{Y_t} &= \frac{((1-\alpha_t)(1+r) + \alpha_t Z_{t+1})(W_t + Y_t - C_t)}{Y_t} \\ \frac{W_{t+1}}{Y_{t+1}} \cdot \frac{Y_{t+1}}{Y_t} &= ((1-\alpha_t)(1+r) + \alpha_t Z_{t+1})(w_t + 1 - c_t) \\ w_{t+1} &= \frac{(1-\alpha_t)(1+r) + \alpha_t Z_t}{Y_{t+1}/Y_t}(w_t + 1 - c_t) \qquad (2.10) \end{aligned}$$

with $c_t = C_t/Y_t$. Proceeding back to period $T-1$, the value function becomes

$$\begin{aligned} V_{T-1}(W_{T-1}, Y_{T-1}) &= \max_{\alpha_{T-1}, C_{T-1}} \left\{ \frac{C_{T-1}^{1-\gamma}}{1-\gamma} + \frac{1}{1+\rho} \mathbb{E}_{T-1}(V_T(W_T, Y_T)) \right\} \\ &= \max_{\alpha_{T-1}, c_{T-1}} \left\{ \frac{(c_{T-1} Y_{T-1})^{1-\gamma}}{1-\gamma} + \frac{1}{1+\rho} \mathbb{E}_{T-1} \left(Y_T^{1-\gamma} \frac{(w_T+1)^{1-\gamma}}{1-\gamma} \right) \right\} \\ &= Y_{T-1}^{1-\gamma} \max_{\alpha_{T-1}, c_{T-1}} \left\{ \frac{c_{T-1}^{1-\gamma}}{1-\gamma} + \frac{1}{1+\rho} \mathbb{E}_{T-1} \left(\left(\frac{Y_T}{Y_{T-1}} \right)^{1-\gamma} \frac{(w_T+1)^{1-\gamma}}{1-\gamma} \right) \right\}. \end{aligned}$$

Proceeding from there and using equation (2.9), we define a new optimization problem with the single state variable w_t

$$v_t(w_t) = \max_{\alpha_t, c_t} \left\{ U(c_t) + \frac{1}{1+\rho} \mathbb{E}_t \left(\left(\frac{Y_{t+1}}{Y_t} \right)^{1-\gamma} v_{t+1}(w_{t+1}) \right) \right\}, \tag{2.11}$$

subject to transition equation (2.10) with terminal value function

$$v_T(w_T) = \frac{(w_T + 1)^{1-\gamma}}{1-\gamma}.$$

Note that for both the single-state transition equation and the value function we have to distinguish retirement and working life since the transition equation of income differs between both. In particular, Y_{t+1}/Y_t equals 1.02 during retirement, 0.8 in the transitional period $t = S$ which is the last period for the individual with labor income and, moreover, it is $\exp(\mu_\varepsilon + \theta_\varepsilon \varepsilon_t)$ during working life.

Then for all $t = 1, \ldots, T$, the relation between the value functions for one and two state variables is

$$V_t\left(W_t, Y_t\right) = Y_t^{1-\gamma} v_t\left(W_t/Y_t\right).$$

After optimization, the policy function $c_t^*(w_t)$ can be re-transformed into

$$C_t^*(W_t, Y_t) = Y_t c_t^*(W_t/Y_t).$$

The policy function $\alpha_t^*(w_t)$, i.e. the share of wealth invested in the risky asset, needs not be re-transformed.

2.2.3 Solution of the dynamic optimization problem

It is sensible to use backward recursion when facing a decision problem in finite time, where the value function in the last period is known. Because of the computational complexity, we approximate the value function by a functional form, in particular by

$$v_t(w_t) = \frac{\left(w_t + a_t\right)^{b_t}}{d_t},$$

which in T exactly equals the value function with $a_T = 1$ and $b_T = d_T = 1 - \gamma$.

We solve the dynamic programming problem for a time horizon of $T = 60$ years, which we refer to as periods. Moreover, we assume the individual to work in the first $S = 45$ years. Table 2.1 shows the parameterization of the model. Following Viceira (2001), we set the value for the standard deviation of innovations in log labor income θ_ε to 15% per year. Expected log income growth μ_ε is set such that the expected income growth rate approximately equals 3% per year.

Parameter	
r, ρ	0.05
μ_Z	0.04
θ_Z	0.2
μ_ε	0.018
θ_ε	0.15
γ	2

Table 2.1: Parameterization

We assume the initial state to be $w_0 = W_0/Y_0 = 1$ and consider an attainable range of 50 equidistant grid points between 0 and 10.

Figure 2.1 shows typical policy functions $C_t^*(W_t, Y_t)$ and $\alpha_t^*(W_t, Y_t)$ solving the dynamic optimization problem. The vertical dashed lines illustrate the period of retirement. In most cases, consumption increases during working life, reaching its maximum near the end. This pattern probably results from income increasing by an average of 3% per year. Just before retirement, consumption drops and stays almost constant during pension, however, on a lower level than during working life since pension payments are rather low. In the last periods of retirement, consump-

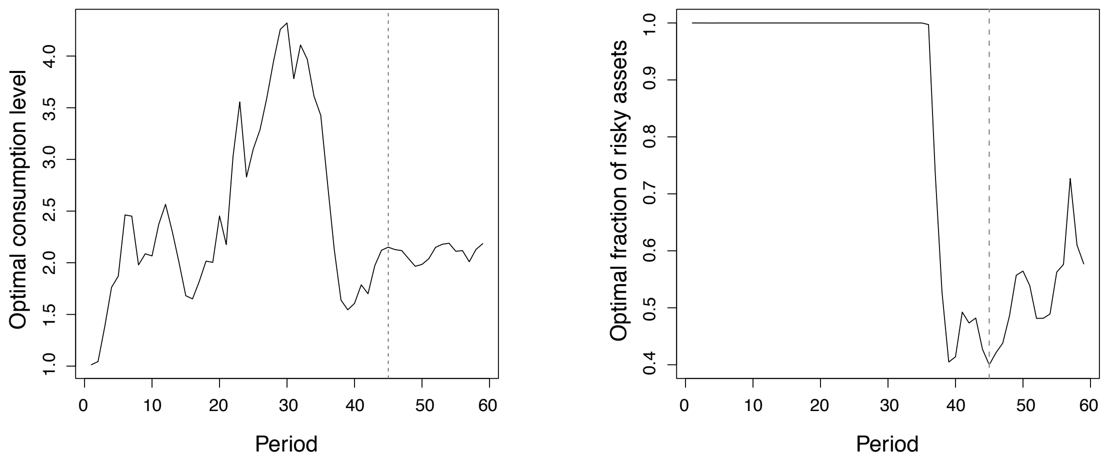

Figure 2.1: Typical optimal policy functions

tion slightly increases again, probably due to the fact that our model contains no bequests.

The second policy function reveals that the optimal fraction of risky assets starts at 100% which holds for almost all individuals in our sample. This may be explained by the fact that the expected value of the risky asset is larger than the expected value of the risk-free asset, while the degree of risk aversion is rather moderate.[2] Regardless of the larger expected value of risky assets, we find a shift towards the risk-free investments during working life. After retirement, facing secure pension payments, the fraction slightly increases again. However, in the last periods, the individuals finally shift their investments towards safe assets. Finding the optimal allocation to risky assets to be larger for employed individuals is consistent with the literature on asset allocation, which typically advises younger people to invest in stocks. An explanation for this suggestion might be that they have more years of labor income ahead in which to recover from the potential losses associated with risky assets, see e.g. Jagannathan and Kocherlakota (1996), among many others. This

[2] Z_t is lognormal distributed with $E(Z_t) = \exp\left(\mu_Z + \theta_Z^2/2\right) = 1.0618$.

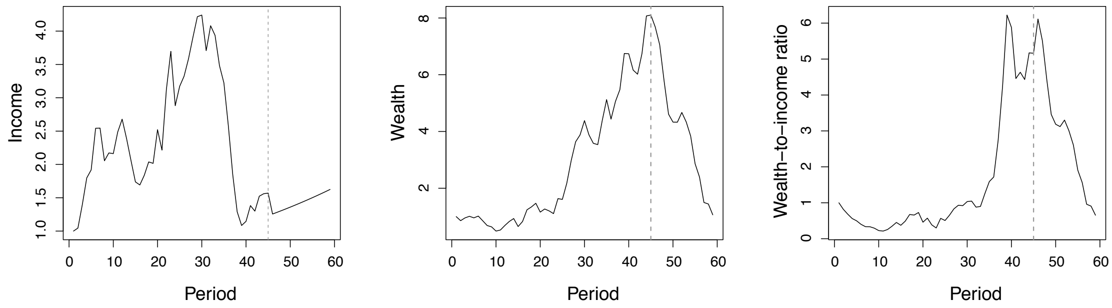

Figure 2.2: Realized paths of states

implies that investments should be shifted towards safer assets as the individuals age.

Figure 2.2 depicts the corresponding income and wealth paths. In this case, "income" means labor income in the first 45 periods and pension payments thereafter. The third panel shows the wealth-to-income ratio $w = W/Y$. Income is exogenous throughout the entire life-cycle. In the first 45 periods, it follows a geometric random walk with drift. During retirement, income evolves deterministically, increasing by 2% per period. The wealth path indicates that the individual starts accumulating wealth in the middle of his working life, attaining its maximum in the last working period $S = 45$. This pattern can be explained by precautionary savings motives since the individual is well informed about the date of retirement and future pension payments. After retirement, the wealth buffer decreases fairly constantly until the final period T.

2.3 Estimation by indirect inference

In the following, the parameters of the earnings process and the risk aversion parameter are assumed to be unknown. The method of indirect inference enables us to estimate all parameters simultaneously. Since we take into account not only the earnings process but also interest income, we suppose the estimates of the earnings processes parameters to be more precise than maximum likelihood estimates, which are based on the earnings process only.

The first part of this section provides a general introduction to indirect inference, presenting the basic idea and a formal definition. Beyond, in the second part, we go into detail on how to use indirect inference to estimate the unknown parameters of the earnings process and the risk aversion parameter γ.

2.3.1 Introduction to indirect inference

Econometric models are often determined by complex links between endogenous and exogenous variables. In such cases, complexity rules out a direct estimation approach since the likelihood function of the model is analytically intractable or too difficult to evaluate. Numerous examples in the literature include nonlinear dynamic models, models with latent or unobserved variables and models with incomplete data.

The goal of indirect inference is to choose the parameters of the economic model such that the generated data and the observed data are as similar as possible, compared by means of the auxiliary model. That is, the parameters of the auxiliary model are estimated using either the generated or the observed data. Indirect inference chooses the parameters of the economic model such that the two sets of estimates of the auxiliary model's parameters are as close as possible. This method is actually similar to the generalized method of moments (GMM). However, its advantage over GMM is that the auxiliary model does not require moment conditions of the economic model for the estimates to be consistent.

Since indirect inference is a simulation-based method, it requires only little analytical tractability. The economic model must, however, allow for simulating data for various values of its parameters. If this condition is met, indirect inference can be used to estimate almost any identifiable model from which it is possible to generate data. In contrast to many other simulation-based methods, the criterion function is an auxiliary model which needs not be an accurate description of the data generating model. However, if it exactly specifies the true model, indirect inference is asymptotically equivalent to maximum likelihood.

Indirect inference was first introduced by Smith (1990, 1993) and later extended by Gourieroux et al. (1993) and Gallant and Tauchen (1996). Since then, there have been many applications to economic models. These include applications to stochastic volatility models (Monfardini (1998)), to semi-parametric models (Dridi and Renault (2000)), to discrete choice models (Keane and Smith (2003)) and studies with Bayesian learning (Guvenen and Smith (2014)).

Mainly following Gourieroux et al. (1993) and Smith (1993), the ideas of indirect inference are now put in a more precise form. Suppose the economic model to be defined as

$$y_t = r(y_{t-1}, x_t, u_t, \theta), \quad t = 1, \ldots, T \tag{2.12}$$

where $x_{1:T}$ is a sequence of observed exogenous variables, $y_{1:T}$ is the corresponding sequence of observed endogenous variables. The error terms $u_{1:T}$ are, however, not observable and further follow a white noise process with known distribution G_0. θ is the unknown parameter vector of the economic model.

Auxiliary model and the first step:

The auxiliary model forms the criterion function. It usually depends on the observations $x_{1:T}$ and $y_{1:T}$ and on the auxiliary parameter vector β which can be determined analytically.[3] In a first step, the parameter vector β of the auxiliary model is estimated using the observed data. The criterion function is maximized with respect to β:

$$\hat{\beta}_T = \arg\max_{\beta} Q_T(x_{1:T}, y_{1:T}, \beta). \tag{2.13}$$

The estimated parameter vector $\hat{\beta}_T$ captures certain features of the observed data. Indirect inference chooses the parameter vector of the economic model θ such that these features are reproduced as closely as possible.

The second step:

To simulate from the economic model (2.12), one first draws a sequence of random errors $\tilde{u}^m_{1:T}$ from their known distribution G_0. Usually, indirect inference uses M such sequences, with $m = 1, \ldots, M$ indicating the number of the simulation step.[4]

Assume the initial value y_0 to be known. Using the simulated error terms and the observed path $x_{1:T}$ it is straightforward to iterate $\tilde{y}^m_{1:T}(\theta)$, given a value of the parameter vector θ. For each of these M paths, the auxiliary parameter vector is estimated by maximizing the criterion function:

$$\tilde{\beta}^m(\theta) = \arg\max_{\beta} Q_T(\tilde{y}^m_{1:T}(\theta), x_{1:T}, \beta). \tag{2.14}$$

Hence, for each value of θ we obtain M estimates of β which are finally averaged over all M repetitions:

$$\tilde{\beta}(\theta) = \frac{1}{M} \sum_{m=1}^{M} \tilde{\beta}^m(\theta) \tag{2.15}$$

[3] Note that the number of parameters in the auxiliary model must be at least as large as the number of parameters in the economic model.

[4] To obtain comparability between data simulated for different parameter values of θ, the M sequences of error terms are drawn only once and then held fixed during the estimation.

The idea of indirect inference is to calibrate the parameter vector θ so as to minimize the distance between $\hat{\beta}_T$ and $\tilde{\beta}(\theta)$. Hence, $\hat{\theta}(\Omega)$ is the solution of the minimum distance problem

$$\hat{\theta}(\Omega) = \arg\min_{\theta} \left[\hat{\beta}_T - \tilde{\beta}(\theta)\right]' \Omega \left[\hat{\beta}_T - \tilde{\beta}(\theta)\right], \tag{2.16}$$

where Ω denotes the positive definite weighting matrix.

Under the assumption that the true parameter θ_0 is the only value of θ which satisfies the previous equation, indirect inference generates consistent estimates as may be argued heuristically: It can be shown that $\tilde{\beta}(\theta)$ converges in probability (with T growing large) to a 'pseudo-true value' that depends on θ. Call it $h(\theta)$, which in the following is referred to as the "binding function". The binding function induces a mapping from the parameters of the economic model θ to the auxiliary parameters β. Similarly, the estimated parameter vector in the actually observed data $\hat{\beta}_T$ converges to a pseudo-true value β_0. Indirect inference chooses θ to satisfy $\beta_0 = h(\theta)$ for $T \to \infty$. Furthermore, (Gourieroux et al., 1993, p. 91f) showed that the indirect inference estimator is asymptotically normal when M is fixed with

$$\sqrt{T}(\hat{\theta}(\Omega) - \theta_0) \xrightarrow[T\to\infty]{d} N(0, W(M, \Omega)), \tag{2.17}$$

where W is the asymptotic variance-covariance matrix.

2.3.2 Estimation of the consumption-savings model

After introducing the most important features of the indirect inference approach, we now turn to the estimation of the economic model described in Section 2.2.2. The three parameters of interest are μ_ε and θ_ε for the stochastic earnings process (2.6) and the risk aversion parameter γ of the CRRA utility function. We collect them in a vector $\theta = (\mu_\varepsilon, \theta_\varepsilon, \gamma)$. Besides, we assume that the other parameters (i.e. interest rate r, subjective discount rate ρ, expected stock return μ_Z, and volatility θ_Z) are known or have been estimated outside our model. For simplicity we also assume that the panel is balanced and that all individuals have the same parameters and known starting values. These restrictions can be relaxed, but simplify the simulation study below.

To estimate the parameters by indirect inference, we have to formulate an auxiliary model which captures the main features of the data. We assume the auxiliary model to consist of several components. In particular, we specify the auxiliary model to be a composition of

- the mean growth rate of earnings (which is the first entry of β),
- the standard deviation of the growth rate of earnings for $t = 1, ..., T$ (second entry of β),
- and additional information about capital income, i.e. the coefficients of a linear regression of *interest payments* on a constant, age and age squared (entry 3-5 of β).

Hence, the auxiliary parameter vector β has length five. The estimation is carried out as outlined in Section 2.3.1. First, the auxiliary parameter vector is estimated for each individual $i = 1, \ldots, N$ using the actually observed earnings and interest payments data. Since we observe N individuals with the same characteristics, we aggregate the results by averaging the estimates over all N individuals. The resulting vector of estimates (from the observed data) refers to $\hat{\beta}_T$ in equation (2.13).

Again, let θ be an arbitrary vector of the parameters of interest and let θ_0 denote the true parameter vector. For parameter vector θ, it is straightforward to compute the corresponding policy functions α_t^*, C_t^*. Assuming that the starting values of the states are given, we simulate M data sets. M can be larger than N to mitigate the influence of sampling errors. The auxiliary parameter vector β is estimated for each of these M data sets, where the average of all M estimates refers to $\tilde{\beta}(\theta)$ in equation (2.15).

As proposed in the previous section, the indirect inference estimate is the solution of the minimum distance problem stated in equation (2.16). In doing so, we assume the weighting matrix Ω to be the identity matrix. Note that no matter how Ω is chosen, the indirect inference estimator is consistent and asymptotically normal. As the asymptotically efficient weighting matrix can be far from optimal in finite samples, we set Ω to the identity matrix.

While the mean and standard deviation of earnings growth are sufficient to estimate the parameters of interest μ_ε and θ_ε, adding information about capital income may increase the precision of the estimates and allows us to estimate the coefficient of risk aversion γ. The amount of information on capital income depends on the question, which variables are observable. In the present setting, the econometrician observes capital income from the risk-free asset in period t, i.e. $(1 - \alpha_t)\, r\, (W_t + Y_t - C_t)$, but has no information on capital income due to changing stock prices. Observing interest payments is, of course, much less informative than observing wealth or consumption, but more common in typical panel data sets. However, no matter which information on capital income is observable, the estimation method can be very easily adapted to the relevant setting since it is very flexible.

		Indirect Inference		ML (Benchmark)	
	True value	Estimate	MSE	Estimate	MSE
μ_ε	0.018	0.01814	$2.599 \cdot 10^{-6}$	0.01807	$1.021 \cdot 10^{-5}$
θ_ε	0.15	0.15028	$7.785 \cdot 10^{-6}$	0.14918	$6.154 \cdot 10^{-6}$
γ	2	1.99705	0.016		

Table 2.2: Mean estimates and MSEs

2.4 Simulation studies

We simulate both the income process and the capital income of $N = 50$ individuals in a dynamically optimal way. Assuming henceforth to be unknown, we aim at estimating the stochastic income parameters μ_ε and θ_ε as well as the risk aversion parameter γ by indirect inference. We set $M = 100$ and therefore obtain 100 data sets for each parameter value of θ. The indirect inference estimate $\hat{\theta}$ minimizes the distance between $\hat{\beta}_T$ and $\tilde{\beta}(\theta)$.

We perform a Monte Carlo procedure and repeat the entire estimation procedure described above 1000 times. Table 2.2 summarizes our average results. Since we know from theory that indirect inference generates consistent estimates, it comes as no surprise that the estimates are centered near the true values. Figure 2.3 shows the corresponding histograms of the estimates, where the true values are illustrated by the vertical black lines. They indicate that the estimates are also symmetrically distributed, even in our finite sample.

As a benchmark, we also estimate the earnings processes' parameters by means of a maximum likelihood estimation that is solely based on earnings data. To make the estimation comparable to our indirect inference approach, we repeat this procedure 1000 times. For each repetition, we generate earnings data for $N = 50$ individuals.

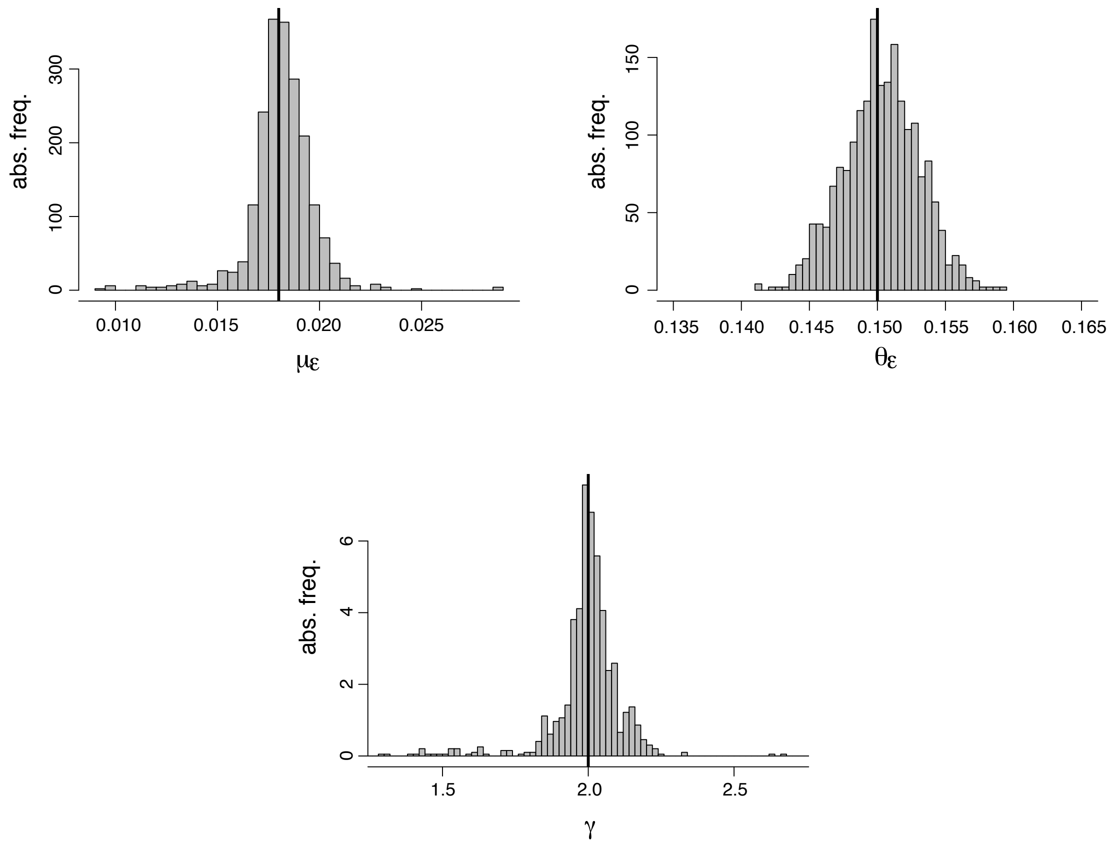

Figure 2.3: Histograms of the estimates

The two right columns of table 2.2 present the average results. Comparing the estimates and the MSEs of the estimation approaches shows that both provide similar results. This applies to the point estimates as well as to the variation of the estimates. However, maximum likelihood is limited to the estimation of the parameters of the earnings process. In contrast, indirect inference is very flexible, providing the possibility of estimating additional parameters of the economic environment such as the risk aversion or the subjective discount factor. Furthermore, it is still well applicable when the length of the sample period is rather small.

2.5 Conclusion

In general, the information set of individuals about their earnings risk differs from the one that is available to the econometrician. Departing from that, in this chapter we suggest a method that uses not only earnings data to elicit the amount of earnings risk, but considers the joint dynamics of capital income and earnings. Since capital income reflects the amount of savings, it contains important information about consumption and savings behavior and therefore implicitly about earnings risk.

To this end, we assume that the true model is a life-cycle model of consumption and investment decisions, stated as a dynamic programming problem. When assuming the parameters of the earnings process and the risk aversion parameter to be known, the optimal controls can be derived by backward recursion. Since these parameters are unknown to the econometrician, we aim at estimating these parameters (from a simulated data set). Using indirect inference enables us to estimate the set of parameters simultaneously. Moreover, it is fairly flexible, meaning that it can easily be adapted to estimate different sets of unknown parameters or include different information available on capital income. The results show that the estimates, obtained from indirect inference, are centered near the true values. Compared with maximum likelihood estimates, no major difference can be observed based on point estimates and MSEs. Hence, our suggestion that a larger information set improves the estimates was not confirmed by the results of our implementation. However, our estimation framework has the distinct advantage that it provides the possibility of estimating further parameters of the economic environment, making it variably applicable. Moreover, it is well employable when the observation period is short.

For future research, it may be interesting to apply our framework to real-world data. Doing so, an interesting modification will be to replace the CRRA utility function by Epstein-Zin preferences which have become quite popular in the recent literature. In contrast to the standard CRRA utility, Epstein-Zin preferences specifiy the utility recursively and allow to separate two of the dimensions along which people care about their allocations; namely, risk aversion and intertemporal substitution. Hence, they might describe the individuals' preferences in a more realistic way.

Chapter 3

Whoever has will be given more: Are earnings subject to a self-reinforcing mechanism?

In the previous chapter we used a consumption-savings model to gather information on earnings risk. Our model's main advantage is the fact that it not only uses earnings data but also considers the joint dynamics of earnings and capital income that we assume to contain information on both consumption and savings behavior and therefore implicitly on earnings risk. Following this application, the next two chapters rather deal with the question of how to model earnings dynamics. To begin with, the present chapter addresses the question if explosiveness should be taken into account when modeling income profiles. Exploring whether incomes evolve in an explosive way accounts for the idea that deviations of earnings from a common trend could be self-reinforcing. Departing from these findings, a more detailed analysis of the nature of individual income profiles will be carried out in the subsequent chapter.

3.1 Introduction

Income and earnings dynamics receive a lot of attention in the economics literature. Understanding the nature of earnings profiles is important as it provides a lot of information on earnings risk. In turn, earnings risk is an important part of the overall economic risk, that individuals and households face. Therefore, several studies examine the structure of earnings. In their seminal work, Nelson and Plosser (1982) found that most macroeconomic variables including wages have a time series structure with a unit root. Beyond, studies like Perron (1988), Carruth and Schnabel (1993) found evidence for the fact that wages follow a random walk with drift.

Incomes following a unit root not only induces a large amount of income risk, but also provides evidence for strong divergence of incomes over time. There is a widespread concern internationally about the rise in earnings inequality since the 1980s. In real terms, workers at the bottom of the income distribution have lost in many countries during the last decades, while we have witnessed a large increase in the share of earnings accruing to the top decile and, even more pronounced, the top percentile of the distribution, especially in English-speaking countries. A whole strand of literature arose from this trend (Gottschalk and Moffitt (1994), Katz et al. (1999), Atkinson et al. (2011)).

In this chapter, we go one step further and suggest to test if explosiveness is a pattern that needs to be taken into account when modeling income profiles. In this context, explosiveness implies that positive deviations tend to boost the income growth rate such that deviations from a common trend will increase even more. In the same way, negative shocks will induce the growth rate to decline or even to become negative. We call this a “self-reinforcing effect”. This could, for instance, occur for managers, who typically start their career earning a relatively high wage, which then increases even more. Explosiveness is ruled out in previous studies on earnings dynamics even though the extreme divergence of some individual income profiles has been noted before (Lillard and Weiss (1979), Guvenen et al. (2015)). This in particular applies to Lillard and Weiss (1979), who found that individuals with greater mean earnings also are subject to greater earnings growth. Moreover, they found the variance of individual mean earnings to increase substantially within the 1970s.

A well specified statistical model of earnings dynamics is of interest not only for economists, but also for policy makers. If, on the one hand, shocks are permanent – and not insurable – policy makers should aim to introduce or facilitate risk sharing mechanisms. On the other hand, if the evolution of earnings is mainly deterministic, an obvious policy response is to improve education in order to shift workers who would end up on low-performing trajectories onto higher profiles.

In our analysis, we employ earnings data which are adjusted for their common trend. We suggest to use a panel unit root test on these data which tests against the alternative of explosiveness. Statistical tests against explosiveness exist for univariate time series (Phillips et al. (2011)) but have not been applied to earnings or other panel data heretofore. Following a suggestion by Hanck (2013) we construct a panel unit root test based on Simes' (1986) intersection test, and apply the test procedure to earnings data from the cross-national equivalent files of the German Socio-Economic Panel (GSOEP) and the U.S. Panel Study of Income Dynamics (PSID) data sets. We find that the null hypothesis of stationarity or unit roots can be rejected in both countries. Explosiveness is evident in the data, but only for a small fraction of the population.

This chapter is organized as follows. Section 3.2 provides an overview of the existing literature on panel unit root tests. In Section 3.3 we formulate an income process which will serve as a basis for the ADF regressions later on when testing for explosiveness. Section 3.4 introduces the right-tailed panel unit root test. The used data are described and the empirical applications are reported in Section 3.5. We apply the panel test to GSOEP and PSID earnings data. Finally, Section 3.6 concludes.

3.2 Literature on panel unit root tests

The main drawback of classical univariate unit root tests is that they often suffer from low power. Adding a cross-sectional dimension could be a way to overcome this problem. Therefore, the application of unit root tests to panel data has attracted much attention in the recent years. The literature thereby distinguishes two "generations" of panel unit root tests. First generation tests are based on the assumption that individual time series are cross-sectionally independent. For instance, Maddala and Wu (1999) applied a method of aggregating individual tests, which was originally suggested by Fisher (1925). Doing so, they tested the joint null hypothesis that all individual processes have a unit root against the alternative that at least one process is stationary. The null is rejected at level α if the test statistic, which combines the p-values of N time series ADF tests, is larger than a given critical value. The critical value was shown to be the $(1-\alpha)$-quantile of the χ^2 distribution with $2N$ degrees of freedom. However, this aggregation method is restricted to test statistics (and hence p-values) which are cross-sectionally independent. Further first generation panel unit root tests were suggested by Levin et al. (2002) and Im et al. (2003). Both approaches test the null hypotheses that all time series are independent random walks. They mainly differ in the set up of the alternative. On the one hand, the autoregressive coefficients are assumed to be identical for all cross section units (Levin et al. (2002)). On the other hand, the autoregressive parameters are supposed to be individual-specific and hence, the alternative states that at least one panel unit is stationary (Im et al. (2003)).

However, for economic applications it is rather inappropriate to assume cross section units to be independent. Especially in macroeconomic applications, time series are often contemporaneously correlated for a variety of reasons such as common factors or spatial spillover effects. Numerous panel unit root tests were developed that allow for different forms of cross section dependence, such as Chang (2002), Phillips and Sul (2003), Bai and Ng (2004), Breitung and Das (2005) and Moon and Perron (2004).

Chang (2002) proposes a test based on a nonlinear instrumental variable estimation of the common augmented Dickey-Fuller regression. With nonlinear transformations of lagged levels used as instruments, she shows that individual ADF statistics are asymptotically independent. The test statistic is defined as a standardized sum of the individual IV t-ratios. However, her test was shown[1] to be valid only if the number of cross section units N is fixed as $T \to \infty$.

Phillips and Sul (2003), Moon and Perron (2004) and Bai and Ng (2004) approach the problem in a similar fashion. They make use of a residual factor model which takes into account cross section dependence. Phillips and Sul (2003) suggest an orthogonalization procedure which may asymptotically remove the common factors[2]. Standard panel tests can then be applied to the transformed series. In line with

[1] See Im and Pesaran (2004).

[2] They address the dependence problem using an iterative method of moments approach to estimate cross section dependence parameters in the factor model.

this, Moon and Perron (2004) suggest to first de-factor the panel data, which is accomplished by means of a principal components estimation of the factor loadings. They show that their proposed test has good asymptotic power properties if there are no deterministic trends. While both previous approaches only allow the idiosyncratic components in the factor model to have unit roots, Bai and Ng (2004) additionally allow for the possibility of unit roots and cointegration in common factors. This method, however, requires large panels with $N/T \rightarrow 0$. Again the common factors are estimated by principal components. Both the idiosyncratic components and common factors are then separately tested for unit roots. The approaches based on factor models are particularly attractive if the number of cross section units (N) is large compared to the number of time periods (T).

Breitung and Das (2005) propose a robust version of the OLS Dickey-Fuller t-statistic which still performs well if the number of time periods is less than the number of cross-sectional units. As an alternative, they suggest a GLS approach obtained from an OLS estimation of a transformed model. The GLS approach, however, is only feasible if $T > N$. Pesaran (2007) introduces the cross section augmented Dickey-Fuller (CADF) test. It augments the standard ADF test with the cross-section averages of both lagged levels and first differences of the individual series. First generation unit root tests can then be applied to the results of the individual CADF tests (e.g. Maddala and Wu (1999)).

All of the panel unit root tests outlined above are tests against stationarity; none of them tests against explosiveness. In the literature, tests against explosiveness are only applied for the detection of financial bubbles, and they are restricted to time series data. Among them are Bhargava (1986), Phillips et al. (2011), Phillips et al. (2015a) and Phillips et al. (2015b).

Our approach follows the idea of Hanck (2013) who proposes a panel unit root test based on Simes (1986) classical intersection test. Hanck sets up the global null hypothesis H_0 that all individual null hypotheses $H_{i,0}$, $i = 1, ..., N$ are true. This method is easy to implement since it only requires p-values of N time series unit root tests. These are ordered ascendingly and compared to increasing critical values. The panel null hypothesis is rejected if at least one p-value is smaller than the corresponding critical value. Hanck's method further enables us to identify those units of the panel, for which $H_{i,0}$ is rejected. Moreover, it accounts for the multiple testing nature since it controls the Familywise Error Rate (FWER), i.e. the probability of falsely rejecting at least one individual null hypothesis, at level α. By combining the p-value approach of Simes (1986) and the p-values of univariate right-tailed unit root tests, we obtain a procedure to test for explosiveness in panel data sets.

3.3 Modeling earnings

The aim of this section is to set up an income process which will serve as a basis for the first-order autoregressive panel model and hence for our test procedure later on.

First, we adjust labor incomes for a common trend. To this end, we assume that labor market experience h_t^i is approximated by age minus school years minus 6. For each period t, we carry out a regression of log earnings in t on a cubic polynomial in potential experience h_t^i and dummy variables for the level of education (with levels less than high school/high school/more than high school), and let $y_{h,t}^i$ denote person i's residual of this regression. This is the usual way of eliminating common effects that affect all individuals in period t (Guvenen (2009)). Since the regressions are run separately for each period, they also capture other time-specific variations in the labor market such as increasing returns to education (Autor (2014)).

In the following, $y^i_{h,t}$ will simply be referred to as "earnings". Earnings are modeled as

$$y^i_{h,t} = \alpha^i + \beta^i h^i_t + z^i_{h,t} + \varepsilon^i_{h,t} \tag{3.1}$$

where α^i and β^i are individual specific random effects with zero expectations, variances σ^2_α and σ^2_β and covariance $\sigma_{\alpha,\beta}$. The random shock $\varepsilon^i_{h,t}$ represents the short-term transitory earnings shocks and is assumed to be homoscedastic and independent of α^i and β^i. Shocks that are longer lasting are modeled by

$$z^i_{h,t} = \rho^i z^i_{h-1,t-1} + \eta^i_{h,t} \tag{3.2}$$

where $\eta^i_{h,t}$ is a (homoscedastic) random shock. Note that the initial value of the persistent shocks is set to $z^i_{0,t} = 0$.

3.4 The right-tailed panel unit root test

As a basis for our test procedure, we first rewrite our earnings model (3.1) and (3.2) as a first-order panel autoregression. Let $\mu^i_{h,t} = E(y^i_{h,t}) = \alpha^i + \beta^i h^i_t$ be individual i's expected earnings in t. Then, our earnings model implies that

$$y^i_{h,t} - \mu^i_{h,t} = \rho^i (y^i_{h-1,t-1} - \mu^i_{h-1,t-1}) + u^i_{h,t} \qquad i = 1, \ldots, N;\ t = 1, \ldots, T \tag{3.3}$$

where T denotes the number of time series observations on each of the N individuals, and $u^i_{h,t}$ is a stochastic process capturing the error terms with

$$u^i_{h,t} = \eta^i_{h,t} + \varepsilon^i_{h,t} - \rho^i \varepsilon^i_{h-1,t-1},$$

and therefore $u^i_{h,t}$ has an MA component.

Before we continue with the panel case, we consider the univariate unit root test against explosiveness, i.e. the null hypothesis $H_0 : \rho^i \leq 1$ against $H_1 : \rho^i > 1$. Tests for explosiveness in time series are commonly used in the literature on the testing for speculative bubbles which are usually referred to as right-tailed unit root tests, for example in Phillips et al. (2011). Pursuant to this approach, we employ the standard univariate Augmented Dickey-Fuller (ADF) test for each time series. To substantiate which version of the ADF test to use, equation (3.3) is transformed into

$$\Delta y^i_{h,t} = \alpha^i(1-\rho^i) + \beta^i(h^i_t + \rho^i - \rho^i h^i_t) + (\rho^i - 1)y^i_{h-1,t-1} + u^i_{h,t}. \qquad (3.4)$$

Since the first two terms are linear in t, we find that the ADF test should be based on a regression including both a constant and a trend. Therefore, the ADF test for individual i is based on the regression

$$\Delta y^i_{h,t} = \gamma^i + \beta^i h^i_t + \phi^i y^i_{h-1,t-1} + \sum_{k=1}^{K_i} \Delta y^i_{h-k,t-k} + v^i_{h,t}. \qquad (3.5)$$

with $\phi^i = \rho^i - 1$ and an appropriate γ^i. The additional lagged differences are included to capture autocorrelations in the error term $u^i_{h,t}$ such that $v^i_{h,t}$ is uncorrelated. Since time series with an MA component can be approximated by AR processes with a large number of lags, the number of additional lagged differences K_i must not be set too low. Gustavsson and Österholm (2014) suggest to use the BIC to determine the number of lags. In the appendix, we show that the BIC may fail to choose the correct number of lags in relatively short time series. Therefore, we set $K_i = 3$ which is large enough to pick up the autocorrelation of the error term, but still small enough to ensure a sufficiently large number of observations.

The regression (3.5) does not only include a constant γ^i but also a trend. The trend is necessary to allow for diverging earnings profiles under the null hypothesis in case of stationarity, i.e. if $\rho^i < 1$. In the absence of a trend, $\rho^i < 1$ would imply a stationary time series around a constant level.

As in ADF tests against stationarity, the test statistic in the test against explosiveness is the t-statistic belonging to the regression coefficient ϕ^i in (3.5). As we consider the right tails of the Dickey-Fuller distribution, the p-values are computed as the probabilities of obtaining larger test statistics than those actually resulting from the ADF tests under the null hypothesis. Since we need precise individual p-values even if they are extremely small, we derived the distribution of the test statistic under the null hypothesis by Monte-Carlo simulations with 1 million replications.

After considering the univariate unit root test against explosiveness, we now come back to the panel case. The global null hypothesis states that *all* N time series are either unit root processes or stationary:

$$H_0 : \rho^1 \leq 1, \rho^2 \leq 1, \ldots, \rho^N \leq 1$$

The global null hypothesis is the intersection over N individual hypotheses, $H_0 = \bigcap_{i=1,\ldots,N} H_{i,0}$ with $H_{i,0} : \rho^i \leq 1$. The alternative hypothesis states that at least one process is explosive. This test setting is based on Simes' intersection test which is a less conservative modification of Bonferroni's procedure for testing multiple hypotheses. The latter lacks power if the test statistics are correlated. The modified version overcomes this problem and controls the FWER even if the test statistics are not independent. If they happen to be independent, the type I error probability is equal to α.

Let $p_{(1)} \leq \ldots \leq p_{(N)}$ denote the ordered p-values of the tests belonging to the individual hypotheses $H_{i,0}$. Simes' test rejects the global H_0 at level α if

$$p_{(j)} \leq j \cdot \frac{\alpha}{N} \text{ for some } j = 1, \ldots, N,$$

i.e., one compares p-values, sorted from most to least significant, to gradually increasing points $j \cdot \alpha/N$. The global null hypothesis is rejected if there exists at least one p-value which is sufficiently small.

The main advantage of this p-value combination approach is that we only require p-values from univariate test statistics, even if these are not independent. At the same time, as found by e.g. Maddala and Wu (1999), p-value combination tests are typically competitive in terms of power and size to computationally much more demanding procedures.

In contrast to previously suggested panel unit root tests, this approach moreover allows to specify the fraction and the identity of the rejected units by means of a procedure suggested by Hommel (1988). He criticizes that Simes' test does not allow statements about individual hypotheses since the FWER is not controlled in this case.[3] To overcome this problem, Hommel introduced an improved multiple test procedure which allows to make statements about individual hypotheses. He applies Simes' test to each intersection hypothesis of a closed test procedure. His procedure controls the FWER at level α, provided Simes' test has level α. The test decisions for each individual hypothesis is performed according to the following algorithm:

1. Compute $j = \max\{(i \in \{1, \ldots, N\} : p_{(N-i+k)} > k\alpha/i \text{ for all } k = 1, \ldots, i\}$.

2. If $p_{(N)} \leq \alpha$, reject all $H_{i,0}$. Else, reject all $H_{i,0}$ with $p_i \leq \alpha/j$.

[3]Simes proposed to reject $H_{1,0}, \ldots, H_{k,0}$ where $k = \max\{j : p_{(j)} \leq j\alpha/N\}$. This procedure was, however, shown not to control the FWER at level α.

Like other p-value combination approaches that are based on transformed sums of p-values, Simes' procedure is consistent as $T \to \infty$ for any $N < \infty$ (e.g. Hanck (2013)). Many other panel unit root tests (e.g. Im et al. (2003), Pesaran (2007)) further require that the fraction of rejected individual null hypotheses must not converge to zero in order to be consistent. However, this is not necessary for tests which are based on the combination of p-values like Simes' test, since the global null is already rejected if one p-value is sufficiently small.

3.5 Empirical applications

We use earnings data from the cross-national equivalent files of the German SOEP (GSOEP) and the U.S. Panel Study of Income Dynamics (PSID) to empirically investigate our research question. More precisely, we employ the annual 1984-2012 waves of the GSOEP and the 1970-1997 waves of the PSID, which comprise a maximum of 29 and 28 years of observations. For both data sets, we restrict our analysis to individuals aged between 20 and 64 who worked at least 520 hours per year. The maximum amount of hours worked is restricted to 5110 and the average number of hours worked is 1852, which corresponds to approximately 35 working hours per week. Moreover, we only consider individuals that have an hourly wage rate of larger than or equal to 3 Euro. Finally, we only take into account individuals with at least 15 observations that, however, need not be consecutive. Following these restrictions, we obtain a data set with size $N = 4270$ for the GSOEP and $N = 4472$ for the PSID. Table 3.1 briefly summarizes these values.

As outlined in Section 3.3, we eliminate possible time effects such as inflation or a potential rise in the skill premium. To this end, we run a cubic regression of log income on experience h_t^i and a dummy variable for the level of education (less than

	GSOEP	PSID
number of observations	4270	4472
∅ years with earnings obs.	20.0	19.8
∅ age in first wave	29.3	24.9
∅ age in last wave	53.8	50.4

Table 3.1: Data description

high school, completed high school, or more than high school). The regressions are carried out separately for each wave in each data set. We approximate experience in the usual way by calculating age minus school years minus 6. The resulting residuals constitute the earnings $y^i_{h,t}$ which are modeled in equation (3.1). To test if explosiveness is evident in earnings, we make use of the panel unit root test suggested in the previous section. We first investigate the GSOEP data set. The first two columns in table 3.2 outline the smallest ordered p-values of the univariate right-tailed unit root tests (left column) and compares them to Simes' cutoff-values (right column).[4]

The results show that the global null hypothesis that all time series are unit root or stationary processes, is rejected at the 5% level. Figure 3.1 (upper panel) illustrates this result: The first 49 p-values are below Simes' cutoff line, represented by the dashed line. Using Hommel's (1988) procedure, we can identify the time series for which non-explosiveness is rejected. At the same time we control the FWER at level $\alpha = 0.05$. With a Hommel's j equal to 4045, all p-values smaller than $\alpha/j = 1.23 \cdot 10^{-5}$ lead to rejections of the corresponding individual hypotheses. This procedure reduces the amount of rejected time series to 13 in case of the GSOEP and only 3 in case of the PSID (see figure 3.2). Apparently, while explosiveness

[4]Some of the time series contain too many consecutive NAs such that we can not carry out unit root tests for them. Excluding these time series reduces the number of individuals to $N = 4061$.

	GSOEP		PSID	
	p-values	Simes' cutoff	p-values	Simes' cutoff
$p_{(1)}$	0.000000	0.000012	0.000000	0.000012
$p_{(2)}$	0.000000	0.000025	0.000000	0.000024
$p_{(3)}$	0.000000	0.000037	0.000003	0.000035
$p_{(4)}$	0.000000	0.000049	0.000013	0.000047
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p_{(12)}$	0.000004	0.000148	0.000071	0.000141
$p_{(13)}$	0.000006	0.000160	0.000077	0.000153
$p_{(14)}$	0.000014	0.000172	0.000087	0.000165
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p_{(28)}$	0.000188	0.000345	0.000293	0.000329
$p_{(29)}$	0.000194	0.000357	0.000321	0.000341
$p_{(30)}$	0.000201	0.000369	0.000389	0.000353
$p_{(31)}$	0.000226	0.000382	0.000449	0.000365
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p_{(48)}$	0.000590	0.000591	0.001081	0.000565
$p_{(49)}$	0.000597	0.000603	0.001081	0.000576
$p_{(50)}$	0.000723	0.000616	0.001084	0.000588
$p_{(51)}$	0.000724	0.000628	0.001087	0.000600
$p_{(52)}$	0.000794	0.000640	0.001099	0.000612
N_{new}	4061		4251	

Table 3.2: First sorted p-values and Simes' cutoff

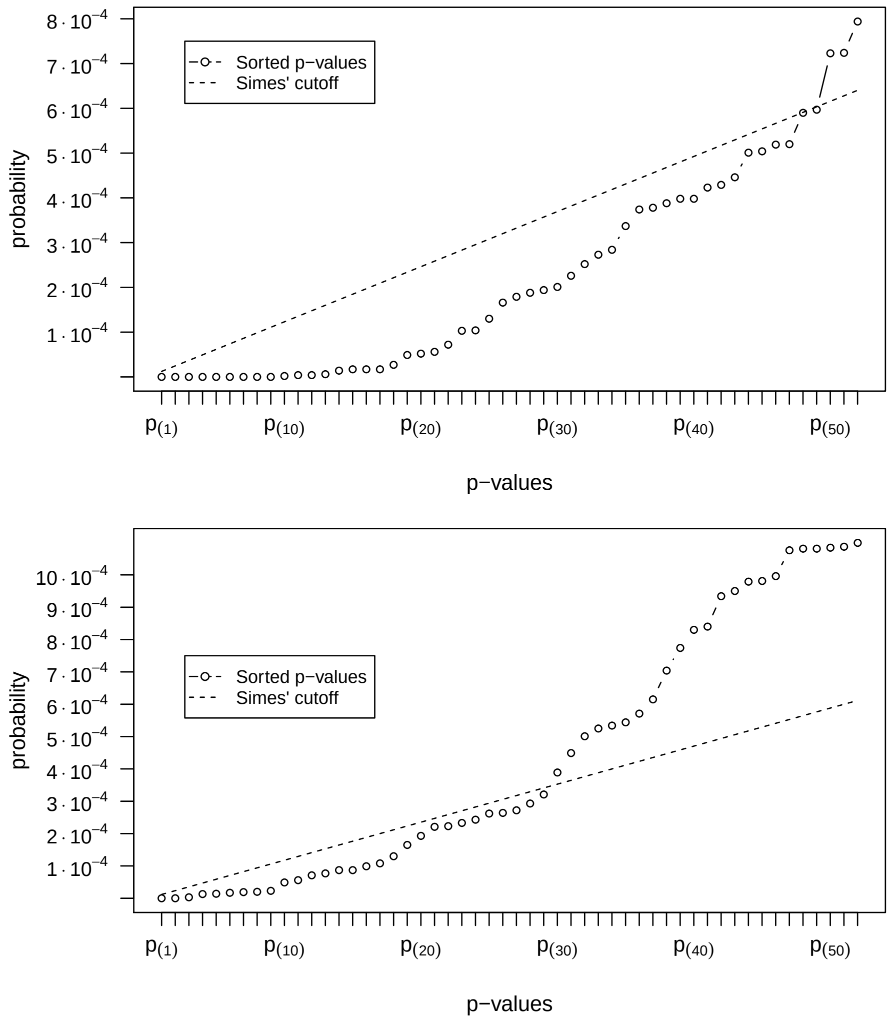

Figure 3.1: First sorted p-values compared to Simes' cutoff line, upper panel: GSOEP, lower panel: PSID

is evident in a statistically significant way in both data sets, it does not play a quantitatively important role. We conclude that explosiveness cannot be ruled out for a very limited number of individuals but needs not to be taken into account when modeling income profiles.[5]

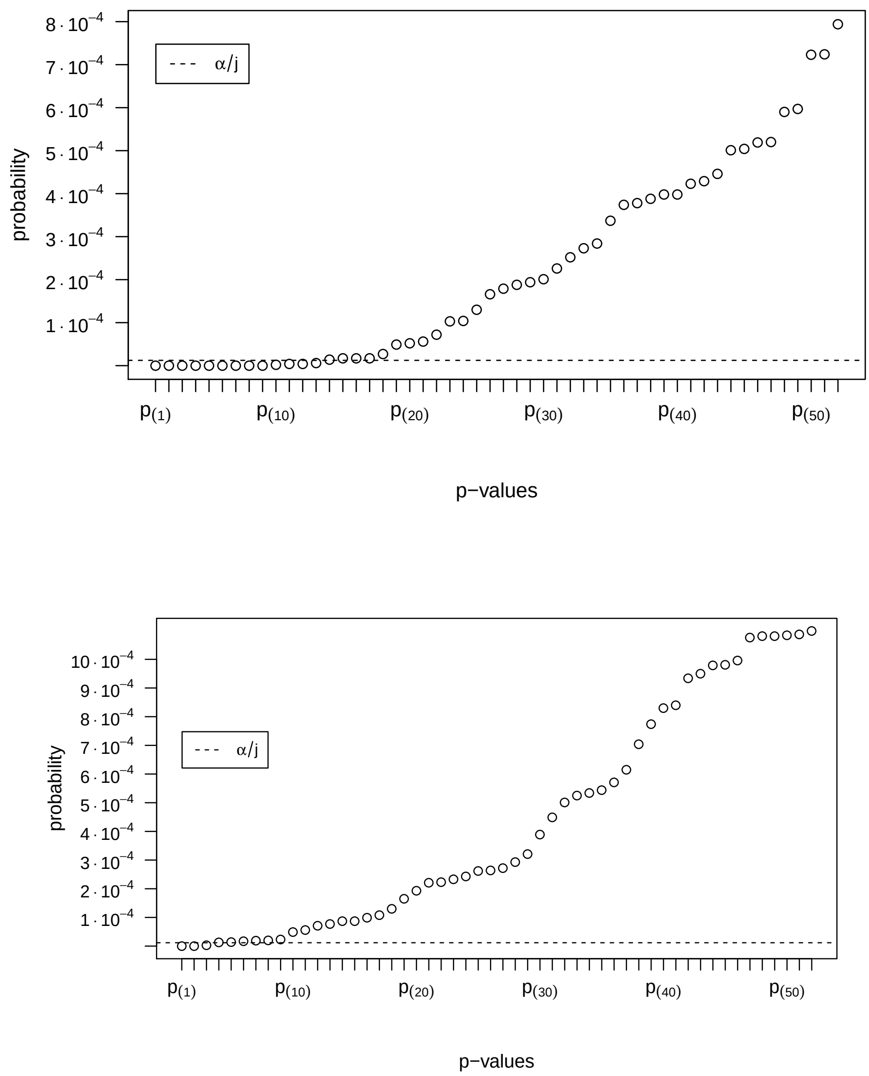

Figure 3.2: First sorted p-values compared to Hommel's cutoff line, upper panel: GSOEP, lower panel: PSID

[5]The null hypothesis that all individuals have non-stationary (i.e. unit root or explosive) processes is also rejected at the 5% significance level. Simes' cutoff yields 85 rejections (i.e. significantly stationary processes) for the PSID and 123 rejections for the GSOEP. Hommel's procedure boils these numbers down to 31 (PSID) and 50 (GSOEP) rejections.

3.6 Conclusion

In many countries earnings inequality has risen sharply in the last decades. Workers at the bottom of the distribution lost, while top earnings increased ever more. Departing from this observation, we suggest to test if individuals are subject to earnings which evolve explosively. In this context, explosiveness implies that earnings above a certain threshold experience higher growth rates than individuals below that very line. If found, such a pattern may be interpreted as a self-reinforcing mechanism of incomes. Moreover, earnings evolving explosively indicates that individuals are subject to a large amount of earnings risk.

A possible method to distinguish between explosive and non-explosive (i.e. stationary or unit root) dynamics is our panel unit root test against explosiveness. We suggest a procedure based on Simes' (1986) intersection test. The global null hypothesis states that all individuals have either stationary earnings processes or unit root processes. This test is robust to cross-sectional dependence and, moreover, rather easy to implement since it only requires p-values of univariate test statistics that are combined into a panel test statistic. Applying this procedure, we may detect the units for which the individual null hypotheses do not hold.

Our test is illustrated for U.S. and German earnings panel data ranging over almost 30 years. We find that the null hypothesis can be rejected at common significance levels, but the number of significantly explosive earnings profiles is very small. Hence, we conclude that explosiveness can be neglected when modeling income dynamics. However, our approach does not provide information as to whether individuals' income profiles are unit root processes or rather stationary around a trend. Therefore, the next chapter aims at investigating individual incomes in more detail. In particular, it explores if incomes are driven by large and persistent shocks or by deterministic components. It therefore provides further information on the nature of earnings risk.

Chapter 4

Is there a systematic fanning out of income profiles?

A correctly specified model of earnings dynamics is important as it provides essential information on earnings risk that individuals face. In the previous chapter we showed that explosiveness is not to be considered when modeling labor incomes. Departing from there, this chapter delves more into the details of how to accurately model labor income profiles. In particular, we now turn to the question if labor incomes mainly evolve deterministically or if they are rather driven by stochastic components. This distinction is closely related to the question if earnings have a unit root or if they are trend-stationary. Two main approaches on modeling earnings dynamics have been established in the literature. Finding evidence for either one of both approaches will be in the focus of the present chapter.

4.1 Introduction

The nature of labor income risk is important for various economic decisions. This equally holds for individuals who participate in the labor market and wish to hedge their income risk as well as for policymakers pursuing the goal of reducing consumption inequality. How individuals respond to variations in their earnings depends to a large degree on the persistence of shocks and whether there are deterministic components that can be anticipated. Both components as well determine optimal policy. When facing persistent shocks which are not insurable, policy makers should aim to introduce risk sharing mechanisms. If the evolution of incomes is mainly deterministic, policy makers rather should improve education and support human capital investments. Hence, a well specified statistical model is of interest.

Two different approaches to modeling labor income profiles have been established in the literature: The heterogeneous income profile model (HIP) and the restricted income profile model (RIP). Both models mainly differ in their assumptions about whether differences in labor income profiles are driven deterministically or stochastically. The restricted income profile model assumes that individuals are exposed to large and persistent income shocks which account for most differences in income courses. The life-cycle labor income profiles, however, are rather similar. According to heterogeneous income profile model, individuals are subject to moderate income shocks, while facing individual-specific income profiles that differ systematically. A literature overview on both models can be found in Section 4.2.

In this chapter we empirically investigate which of the two models is more suitable to describe earnings data. For this issue, we provide a model to estimate labor income courses and thereby distinguish between common coefficients and individual-specific coefficients. Common coefficients encompass effects that affect individuals with similar demographic characteristics in the same way, such as economic or education indicators. By means of individual-specific components the divergence of income profiles can be explained.

More so, extending the models that are used in the literature, our model allows all coefficients to be time-varying, hence enabling us to test if income courses follow a random walk, indicated by varying individual intercepts. If this holds true, it indicates that the RIP model is more suitable to accurately describe labor income profiles than the HIP model.

Estimation of the latent state vector and of the unknown variance parameters is carried out using Gibbs sampling, a Markov chain Monte Carlo (MCMC) algorithm. In our application, the Forward Filtering Backward Sampling (FFBS) algorithm is used as a building block, while further draws result from the conjugate prior distributions. Applying our framework to the German SOEP (GSOEP) data, we find evidence that the earnings data disprove the RIP approach. For validation we employ the Bayes factor, which confirms our finding.

The rest of this chapter is structured as follows. Section 4.2 introduces both HIP and RIP models in more detail. Section 4.3 describes the economic model and the Bayesian estimation framework. In Section 4.4, the estimation of an artificial data set is presented. Section 4.5 describes the GSOEP panel data that will be used for empirical analysis, whose results are presented in Section 4.6. Finally, Section 4.7 concludes.

4.2 Overview - HIP vs. RIP model

Two different approaches for modeling idiosyncratic labor incomes have been established in the literature. The first one assumes that income grows at an individual-specific, deterministic and non-observable rate, facing stationary shocks around this very rate. Since each labor income profile is unique - even in the absence of shocks

- Guvenen (2009) suggested to label this model the "heterogeneous income profile" model (HIP). The human capital model possibly serves as a theoretical basis for this approach, which suggests that the systematic differences in income profiles stem, for example, from different talents in accumulating skills. In contrast, the "restricted income profile" model (RIP) assumes that income contains a random walk and a transitory stationary component. Hence, each individual is subject to large and persistent income shocks, while there exist no systematic differences between income profiles.

Many studies empirically examine both approaches. The HIP model has a long-standing tradition in the literature beginning in the late 1970's. Lillard and Weiss (1979) proposed a parameterization of an individual earnings function whose covariance structure accounts for individual differences in level and growth of earnings as well as for serially correlated transitory differences. Their specification is successful in predicting the pattern of the data. Estimating the parameters of the residual covariance structure, they found that individuals with larger mean earnings also had larger earnings growth. Moreover, they found a substantial increase of the variance of individual mean earnings with increased experience, even though the variance of the growth component is constant. Both results led them to infer that there exists a sizable amount of inequality in their (quite homogeneous) sample. Hause (1980) was primarily motivated by the on-the-job training hypothesis and allowed for deterministic growth rate heterogeneity. Moreover, he discussed a random walk alternative which he found to be inappropriate. More recently, Baker (1997) estimated a nested version of both models using an equally weighted minimum distance estimator. In this nested specification, the hypothesis of a unit root is rejected. Baker also tested for serial correlation in the income growth rates. The presence of serial correlation in the data could only be explained by the HIP, but not by the RIP model. Baker found rather weak evidence in favor of the HIP model. However, since the test has poor small sample properties, it does not seem entirely reliable. Guvenen (2009) formulated an income process which he estimated using a minimum distance approach. To this end, he minimized the distance between the elements of the empirical covariance matrix of income residuals and the covariance matrix implied by the theoretical model for HIP and RIP model, respectively. Moreover, he qualitatively compared the evolution of the empirical variances and autocovariances of the income residuals and the theoretical counterparts implied by HIP and RIP when using the estimated coefficients from the previous step. Overall, he found empirical evidence in favor of the HIP model and showed that the results of the existing literature arguing against profile heterogeneity may have been misinterpreted. Browning et al. (2010) showed that conventional income processes are unable to capture many features of the observed data due to limited allowance for heterogeneity. They suggest a non-linear factor model which, among others, allows for heterogeneity in the starting value, the variance of shocks, in MA and AR parameters and in the measurement error variances. The model is estimated by indirect inference. The authors revealed a substantial amount of heterogeneity among individuals. Moreover, they found strong evidence against unit root models.

In contrast, numerous studies support the RIP model. MaCurdy (1982) fitted an error structure to earnings data which combines factor schemes and time series processes. Since he found the autocovariances of the earnings growth rates to be zero for any lag larger than 2, he rejected the presence of growth rate heterogeneity in earnings. If earnings profiles were heterogeneous, individual income growth should be positively autocorrelated, still for higher-order lags. Furthermore, MaCurdy found that error processes associated with income in levels are non-stationary with increasing variability over time. These results were supported by Abowd and Card (1989) who found that all higher-order autocovariances of earnings growth are jointly equal to zero.

Carroll and Samwick (1997) and Meghir and Pistaferri (2004) focused on the structure of incomes, considering its single components. The former showed that idiosyncratic income can be decomposed into a random walk and a transitory component. As they moreover exclude income growing at a deterministic rate, their approach clearly supports the RIP model. Beyond, Meghir and Pistaferri (2004) tested the hypothesis that there are no permanent shocks in earnings which was strongly rejected. Based on this result, they concluded that earnings should be modeled as the sum of a permanent component and a transitory shock and therefore found empirical evidence in favor of the RIP model. Recently, Hryshko (2012) proposed an income process which encompasses both transitory and persistent components of earnings and, moreover, allows for income profile heterogeneity. His estimation results indicate that the variance of the deterministic income growth is zero. Hence, the HIP model can be rejected. At the same time, the variance of the permanent component is significantly larger than zero such that the RIP model appears to be correct one.

Overall, a great number of studies support one of the two approaches and the literature has not yet reached a consensus as to which model is more suitable to mirror reality. Therefore, the present chapter aims at providing further evidence for one of both models.

The nature of income profiles plays a central role in many economic applications. If earnings have a unit root, predictability decreases dramatically over time.

If income courses evolve differently in a systematic fashion, further covariates must be considered when modeling future outcomes, cf. Baker (1997). As motivated by Hryshko (2012), different income profiles can also lead to very different policy and welfare implications, for instance when estimating models dealing with household consumption or portfolio choice behavior. Based on the HIP model, individuals must be able to update their beliefs about their individual income profiles in order to sequentially decide on their optimal consumption level and portfolio allocation, cf. Guvenen and Smith (2014). However, if the variation in income is due to either transitory and permanent shocks rather than being influenced by heterogeneous income profiles, it is more reasonable to model households' consumption and portfolio choice in an incomplete market model, such as Castaneda et al. (2003). Another reason for distinguishing both models may arise in the context of politics. In the general economic theory, a policymaker's goal is to reduce consumption inequality. If the true income process is expressed by the HIP model, which deals with moderate income shocks and income differences that may be explained by the human capital model, the policymaker may want to implement a subsidy system to support human capital investments in indigent families. As the income risk is rather moderate, self-insurance may be an adequate instrument to protect against shocks. If, however, the true income process is RIP with substantial and long-lasting income shocks, self-insurance might not be sufficient. In that case, it would rather be useful to educate the public about risk-sharing instruments like human capital contracts.

4.3 The model

In order to find empirical evidence for either one of the two approaches outlined above, this section will now present the basic estimation framework. This framework provides a flexible estimation of income profiles, allowing both for individual-specific and time-varying coefficients. As a consequence, it is able to capture both processes. As we permit parameters to evolve through time, we now consider a *dynamic linear model* (DLM). Inference is accomplished using the Gibbs sampler (Geman and Geman (1984)), a Bayesian method, iteratively sampling states, the observational variance and the state variance, each contingent upon the others. As we consider a linear Gaussian model, states can be sampled jointly via forward filtering backward sampling (FFBS) (Carter and Kohn (1994), Frühwirth-Schnatter (2008)), an algorithm based on the Kalman filter (Kalman (1960)).

We now present our income profile model in state-space formulation. It allows for time varying coefficients and controls for the part of variation that is common to all individuals. As typically used in the literature, the state space model consists of an observation equation and a state equation,

$$\begin{aligned} \mathbf{y}_t &= \mathbf{X}_t \cdot \boldsymbol{\beta}_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim N(0, \sigma_v^2\,\mathbf{I}_N) \\ \boldsymbol{\beta}_t &= \mathbf{G} \cdot \boldsymbol{\beta}_{t-1} + \mathbf{w}_t, \quad\ \mathbf{w}_t \sim N(0, \mathbf{W}). \end{aligned} \tag{4.1}$$

$\mathbf{y}_t$ corresponds to the $N \times 1$ endogenous income vector, containing the logarithmic labor incomes of all individuals ($i = 1, \ldots, N$) at time t. The $(2N+3) \times 1$ vector $\boldsymbol{\beta}_t$ consists of the unobservable and time-varying regression coefficients. It is given by

$$\boldsymbol{\beta}_t = \begin{pmatrix} \beta_{\text{GDP}_t} \\ \beta_{\text{MALE}_t} \\ \beta_{\text{EDU}_t} \\ \boxed{\begin{matrix} \alpha_{1,t} \\ \beta_{\text{AGE}_{1,t}} \end{matrix}} \\ \boxed{\begin{matrix} \alpha_{2,t} \\ \beta_{\text{AGE}_{2,t}} \end{matrix}} \\ \vdots \\ \boxed{\begin{matrix} \alpha_{N,t} \\ \beta_{\text{AGE}_{N,t}} \end{matrix}} \end{pmatrix}. \tag{4.2}$$

The first three coefficients correspond to effects that are common to all individuals, capturing the global effects on income driven by GDP, sex and education. These are the so called "fixed effects" as they influence the income of individuals with similar demographic characteristics identically. Additionally, $\boldsymbol{\beta}_t$ contains both individual intercepts and individual slope coefficients, that are referred to as "individual effects" here. The intercepts are related to permanent shocks in income, since an unrestricted $\alpha_{i,t}$ may capture the (permanent) random walk component of income. Thus, if $\alpha_{i,t}$ has a unit root, the RIP model will be concluded to more appropriately describe the data.

Furthermore, we have to allow for heterogeneity in income profiles for mapping the HIP model. Therefore, we use individual-specific slope coefficients that may evolve through time. More precisely, this effect is encompassed by the age of all individuals which serve as our proxy for time. A large variation between individuals' slope coefficients would strongly indicate the HIP model to be more suitable to describe the data.

The $N \times (2N+3)$ matrix $\mathbf{X}_t$ represents all explanatory variables at time t.

$$\mathbf{X}_t = \begin{pmatrix} log(\mathrm{GDP}_t) & \mathrm{MALE}_1 & \mathrm{EDU}_{1,t} & \boxed{1 \quad \mathrm{AGE}_{1,t}} & 0 & 0 & 0 & 0 & \dots & 0 \\ log(\mathrm{GDP}_t) & \mathrm{MALE}_2 & \mathrm{EDU}_{2,t} & 0 \quad 0 & \boxed{1 \quad \mathrm{AGE}_{2,t}} & & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \quad \vdots & \vdots & \vdots & \ddots & \ddots & & \dots \\ log(\mathrm{GDP}_t) & \mathrm{MALE}_N & \mathrm{EDU}_{N,t} & 0 \quad 0 & 0 & 0 & 0 & \dots & \boxed{1 \quad \mathrm{AGE}_{N,t}} & \end{pmatrix} \tag{4.3}$$

The first column of $\mathbf{X}_t$ contains the logarithmic GDP at time t, which is identical for all individuals in the sample. The second and third column contain gender dummies and the number of years of education for each individual. Furthermore, $\mathbf{X}_t$ contains a constant for each individual and the corresponding age.

The observational error $\mathbf{v}_t$ is assumed to be normally distributed with mean zero and unknown covariance matrix $\mathbf{V} = \sigma_v^2 \mathbf{I}_N$. For simplicity, the variances are assumed to be identical for all individuals. The errors are further assumed to be independent, which seems reasonable since common influences are already taken into account by the fixed effects. In terms of the two income models introduced above, the observational error vector $\mathbf{v}_t$ is related to short-term shocks which die out rather quickly. More so, in this present model short-term shocks are assumed to die out within one single period. This assumption accounts for the fact that the number of observed periods in our data is rather small.

The state equation describes the evolution of the latent state vector $\boldsymbol{\beta}_t$. In our framework, $\mathbf{G}$, which is usually referred to as the state evolution matrix, is assumed to be the known, time invariant identity matrix $\mathbf{I}_{(2N+3)}$. Furthermore, $\mathbf{w}_t$ is the innovation at time t and assumed to be normally distributed with zero mean and covariance matrix $\mathbf{W}$ (that in the following is referred to as the "system variance matrix"). Similar to the observational matrix, the system variance matrix is a diagonal matrix, which here is

$$\mathbf{W} = \begin{pmatrix} \sigma^2_{w_{\text{GDP}}} & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & \sigma^2_{w_{\text{MALE}}} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & \sigma^2_{w_{\text{EDU}}} & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \sigma^2_{w_\alpha} & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \sigma^2_{w_{\text{AGE}}} & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{w_\alpha} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \end{pmatrix}. \tag{4.4}$$

Unlike $\mathbf{V}$, which only depends on the single parameter σ^2_v, $\mathbf{W}$ allows the coefficients to vary to different degrees and is therefore assumed to consist of five unknown variation parameters. Besides the one related to each of the common coefficients, there are variation parameters for the intercepts and slope coefficients, which are identical for all individuals.

Note that both the RIP and HIP model are included in the state-space model. Reconsider the HIP model, which assumes individuals' labor incomes profiles to differ systematically while facing moderate shocks. The HIP model is present if there is a large variation in individuals' coefficients $\beta_{i,t}$, along with a rather constant, or at least stationary, $\alpha_{i,t}$ and a transitory component $\mathbf{v}_t$. However, if the individual slope coefficients $\beta_{i,t}$ are almost identical among individuals and the individual intercepts $\alpha_{i,t}$ have a unit root, the RIP model seems to apply. The transitory shock $\mathbf{v}_t$ may as well occur in this model.

However, the state vector $\boldsymbol{\beta}_t$, the observational variance parameter σ^2_v and the system variance parameters $\Sigma^2_w := \{\sigma^2_{w_{\text{GDP}}}, \dots, \sigma^2_{w_{\text{AGE}}}\}$ are unknown in the state-space model. Using a normally distributed prior for $\boldsymbol{\beta}_t$ and Inverse Gamma distributed priors for σ_v^2 and for each element of Σ_w^2 allows to perform a fully conjugate Bayesian analysis, meaning that both the prior and posterior distributions come from the same distribution families. More concretely, this leads to the following prior specification:

$$(\boldsymbol{\beta}_0|\Sigma_w^2, \sigma_v^2, \mathcal{D}_0) \sim N(\mathbf{m}_0, \mathbf{C}_0), \tag{4.5}$$

$$(\sigma_v^2|\mathcal{D}_0) \sim IG(a_0, b_0), \tag{4.6}$$

$$(\sigma_{w_{k[i]}}^2|\mathcal{D}_0) \sim IG(\nu_{0_{k[i]}}, s_{0_{k[i]}}), \tag{4.7}$$

for $k = (\text{GDP MALE EDU } \alpha \text{ AGE})'$ and $i = 1, ..., 5$. The prior for the coefficient vector is centered around $\mathbf{m}_0$ with covariance matrix $\mathbf{C}_0$. a_0 and $\nu_{0_{k[i]}}$ denote the shape parameters of the Inverse Gamma draws and b_0, $s_{0_{k[i]}}$ are the scale parameters. Furthermore, let $\mathcal{D}_t = \{\mathbf{y}_1, ..., \mathbf{y}_t\}$ be the information set available at time t.

The joint posterior density is proportional to the product of the likelihood and the joint prior distribution[1] of $\boldsymbol{\beta}_t$, σ_v^2 and Σ_w^2:

$$\begin{aligned} p(\boldsymbol{\beta}_{1:T}, \sigma_v^2, \Sigma_w^2|\mathbf{y}_{1:T}) \propto \prod_{t=1}^{T} N(\mathbf{y}_t|\boldsymbol{\beta}_t, \sigma_v^2, \Sigma_w^2) \cdot N(\boldsymbol{\beta}_t|\sigma_v^2, \Sigma_w^2) \cdot IG(\sigma_v^2) \\ \cdot IG(\sigma_{w_{\text{GDP}}}^2) \cdots IG(\sigma_{w_{\text{AGE}}}^2) \end{aligned} \tag{4.8}$$

Samples from this joint posterior density can be obtained by Gibbs Sampling, that performs alternating draws from the full conditional densities $p(\boldsymbol{\beta}_{1:T}|\sigma_v^2, \Sigma_w^2, \mathbf{y}_{1:T})$, $p(\sigma_v^2|\boldsymbol{\beta}_{1:T}, \mathbf{y}_{1:T})$ and $p(\sigma_{w_{k[i]}}^2|\boldsymbol{\beta}_{k[i],1:T}, \mathbf{y}_{1:T})$. In order to initialize the Gibbs sampling, one has to set a start value for either σ_v^2 and $\sigma_{w_{k[i]}}^2$ (for $i = 1, ..., 5$) or for $\boldsymbol{\beta}_{1:T}$. Here this is accomplished by drawing σ_v^2 and $\sigma_{w_{k[i]}}^2$ from

[1] Note that all priors that are related to σ_v^2 and Σ_w^2 are assumed to be independent.

their prior distributions, stated in (4.6) and (4.7). These draws can then be used to determine the initial matrices of $\mathbf{V}$ and $\mathbf{W}$.

While drawing from the latter conditional posteriors is specified by the set of conjugate priors, we need an algorithm for sampling from $p(\boldsymbol{\beta}_{1:T}|\sigma_v^2, \Sigma_w^2, \mathbf{y}_{1:T})$. For this purpose we use the FFBS algorithm, an algorithm sequentially updating the information set by new observations *(Forward Filtering)* and then drawing from the joint distribution of the states given the data *(Backward Sampling)*.
This procedure is now presented in more detail, starting with the (Kalman) *filtering*:

- Prediction step: Predict the new state based on the information up to time $t-1$.[2] The prior for the state vector at time t $(\boldsymbol{\beta}_t|\mathcal{D}_{t-1})$ is $N(\mathbf{a}_t, \mathbf{R}_t)$ with $\mathbf{a}_t = \mathbf{G}\mathbf{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}' + \mathbf{W}$.

- The one-step-ahead predictive distribution based on time $t-1$ is then given by:
  $(\mathbf{y}_t|\mathcal{D}_{t-1}) \sim N(\mathbf{f}_t, \mathbf{Q}_t)$ with $\mathbf{f}_t = \mathbf{X}_t\mathbf{a}_t$ and $\mathbf{Q}_t = \mathbf{X}_t\mathbf{R}_t\mathbf{X}_t' + \mathbf{V}$.

- Update step: Given the updated information set $\mathcal{D}_t$, the posterior distribution for the state vector at time t $(\boldsymbol{\beta}_t|\mathcal{D}_t)$ is $N(\mathbf{m}_t, \mathbf{C}_t)$ with $\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t(\mathbf{y}_t - \mathbf{f}_t)$, $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t\mathbf{Q}_t\mathbf{A}_t'$ and $\mathbf{A}_t = \mathbf{R}_t\mathbf{X}_t'\mathbf{Q}_t^{-1}$.

However, the goal is to sample from the joint posterior distribution of the states. It is therefore not sufficient to consider only the past information $\mathcal{D}_t$ at each time t. One rather needs to take into account all information that is available, including the future state vectors.

[2] Keep in mind that the initial distribution of the state $(\boldsymbol{\beta}_0|\Sigma_w^2, \sigma_v^2, \mathcal{D}_0)$ is $N(\mathbf{m}_0, \mathbf{C}_0)$.

In general, the joint posterior distribution can be written as the product of the conditional distributions

$$p(\boldsymbol{\beta}_{1:T}|\mathbf{V},\mathbf{W},\mathcal{D}_T) = \prod_{t=1}^{T} p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1:T},\mathbf{V},\mathbf{W},\mathcal{D}_T), \tag{4.9}$$

where the last factor is $p(\boldsymbol{\beta}_T|\mathbf{V},\mathbf{W},\mathcal{D}_T)$: the known filtering distribution of $\boldsymbol{\beta}_T$ that is $N(\mathbf{m}_T,\mathbf{C}_T)$. Then, for $t = T-1, T-2, ..., 1$, one can recursively draw $\boldsymbol{\beta}_t$ from $p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1:T},\mathbf{V},\mathbf{W},\mathcal{D}_T)$.

Note that the Markovian structure of the model implies that

$$p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1:T},\mathbf{V},\mathbf{W},\mathcal{D}_T) = p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1},\mathbf{V},\mathbf{W},\mathcal{D}_t),$$

meaning that solely the state vector in period $t+1$ is relevant to determine the distribution of $\boldsymbol{\beta}_t$. Based on this, it can be shown that the distribution of $\boldsymbol{\beta}_t$ is $N(\mathbf{m}_t^*,\mathbf{C}_t^*)$ with

$$\mathbf{m}_t^* \equiv E(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1},\mathbf{V},\mathbf{W},\mathcal{D}_t) = \mathbf{m}_t + \mathbf{B}_t(\boldsymbol{\beta}_{t+1} - \mathbf{a}_{t+1}), \tag{4.10}$$

$$\mathbf{C}_t^* \equiv V(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1},\mathbf{V},\mathbf{W},\mathcal{D}_t) = \mathbf{C}_t - \mathbf{B}_t\mathbf{R}_{t+1}\mathbf{B}_t', \tag{4.11}$$

where $\mathbf{B}_t = \mathbf{C}_t\mathbf{G}\mathbf{R}_{t+1}^{-1}$.

Hence, the FFBS algorithm provides a way to generate a sample from the joint posterior distribution of the states.

Sampling from the conditional posteriors $p(\sigma_v^2|\boldsymbol{\beta}_{1:T},\mathbf{y}_{1:T})$ and $p(\sigma_{w_{k[i]}}^2|\boldsymbol{\beta}_{k[i],1:T},\mathbf{y}_{1:T})$ is more straightforward since we know from their conjugate priors that the conditional posterior distributions are Inverse Gamma.

We first derive the conditional posterior distributions of the elements in Σ_w^2. Since they are independent from each other, each can be drawn separately. As the first three variation parameters have the same structure, they are updated in the same way. Accordingly, we first consider $i \in \{1, 2, 3\}$. Assume $\mathbf{K}_i$ to be a matrix whose i'th diagonal element is 1, while the rest of the matrix is 0.

In general, $p(\sigma^2_{w_{k[i]}} | \boldsymbol{\beta}_{k[i],1:T}, \mathbf{y}_{1:T})$ is proportional to the joint likelihood of $\boldsymbol{\beta}_{k[i],1:T}$ multiplied by the prior distribution of $\sigma^2_{w_{k[i]}}$:

$$p(\sigma^2_{w_{k[i]}} | \boldsymbol{\beta}_{k[i],1:T}, \mathbf{y}_{1:T}) \propto f(\boldsymbol{\beta}_{k[i],1:T} | \mathbf{y}_{1:T}, \sigma^2_{w_{k[i]}}) \cdot p(\sigma^2_{w_{k[i]}}). \tag{4.12}$$

For $i \in \{1, 2, 3\}$, this is

$$\propto \prod_{t=1}^{T} \det(\mathbf{W})^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)' \cdot \mathbf{K}_i \cdot \mathbf{W}^{-1} \cdot \left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)\right)$$
$$\times \left(\sigma^2_{w_{k[i]}}\right)^{-\nu_{0_{k[i]}}-1} \cdot \exp\left(-\frac{s_{0_{k[i]}}}{\sigma^2_{w_{k[i]}}}\right),$$

where all parameters in $\mathbf{W}$ and $\boldsymbol{\beta}_t$, that are not related to $k[i]$, are treated as constants. Due to the proportionality in the above expression, the constants need not be considered any further in this derivation. Therefore, $\mathbf{W}$ is replaced by $\sigma^2_{w_{k[i]}}$ and the product equals

$$\left(\sigma^2_{w_{k[i]}}\right)^{-\frac{T}{2}} \cdot \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)' \cdot \mathbf{K}_i \cdot \frac{1}{\sigma^2_{w_{k[i]}}} \cdot \left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)\right\}$$
$$\times \left(\sigma^2_{w_{k[i]}}\right)^{-\nu_{0_{k[i]}}-1} \cdot \exp\left(-\frac{s_{0_{k[i]}}}{\sigma^2_{w_{k[i]}}}\right).$$

Excluding $\frac{1}{\sigma^2_{w_{k[i]}}}$ and rearranging yields

$$\exp\left\{\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)' \cdot \mathbf{K}_i \cdot \left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right) - s_{0_{k[i]}}\right) \cdot \frac{1}{\sigma^2_{w_{k[i]}}}\right\}$$
$$\times \left(\sigma^2_{w_{k[i]}}\right)^{-\frac{T}{2}-\nu_{0_{k[i]}}-1}.$$

Hence, the conditional posterior distribution is

$$IG\left(\nu_{0_{k[i]}} + \frac{T}{2}, s_{0_{k[i]}} + \frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)' \cdot \mathbf{K}_i \cdot \left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)\right),$$

for $i \in \{1, 2, 3\}$.

However, the posteriors of $\sigma^2_{w_\alpha}$ and $\sigma^2_{w_{\mathrm{AGE}}}$ (that is $\sigma^2_{w_{k[i]}}$ for $i \in \{4, 5\}$) are slightly different since each of the two parameters occurs N times in $\mathbf{W}$. To derive their conditional posteriors we again consider the parameters that are not related to $k[i]$, as constants and exclude them from the derivation. The conditional posterior of $\sigma^2_{w_{k[i]}}$ is proportional to

$$\propto \underbrace{\det(\mathbf{W})}_{=\left(\sigma^2_{w_{k[i]}}\right)^N}{}^{-\frac{T}{2}} \cdot \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)' \cdot \mathbf{K}_j \cdot \frac{1}{\sigma^2_{w_{k[i]}}} \cdot \left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)\right\}$$
$$\times \left(\sigma^2_{w_{k[i]}}\right)^{-\nu_{0_{k[i]}}-1} \cdot \exp\left(-\frac{s_{0_{k[i]}}}{\sigma^2_{w_{k[i]}}}\right),$$

for $i \in \{4, 5\}$ and $j = (i \quad i+2 \quad i+4 \quad \ldots \quad 2N+3)'$.

Rearranging yields the conditional posterior distribution of $\sigma^2_{w_{k[i]}}$, which is

$$IG\left(\nu_{0_{k[i]}} + \frac{NT}{2}, s_{0_{k[i]}} + \frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)' \cdot \mathbf{K}_j \cdot \left(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1}\right)\right).$$

Now that we derived the conditional posterior distributions of all the parameters contained in $\mathbf{W}$, it is straightforward to derive the conditional posterior distribution of σ^2_v. This is

$$IG\left(a_0 + \frac{NT}{2}, b_0 + \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_t\right)'\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_t\right)\right).$$

Given the conditional posteriors, the observational variance parameter σ^2_v and the system variance parameters collected in Σ^2_w are sampled independently of one another. Both matrices **V** and **W** can then easily be determined from the draws as their structures are known.

The Gibbs steps are iterated a large number of times. After a burn-in period, the Gibbs Sampler generates draws from the joint posterior distribution of σ_v^2, Σ_w^2 and $\boldsymbol{\beta}_{1:T}$. Finally, the variety of draws is averaged.

4.4 Estimation of simulated data sets

We intend to find evidence in favor of either the RIP or HIP model. To verify our estimation framework, we apply it to a simulated panel data set with known coefficients. To this end, we use the state-space model (4.1) as the data generating process and simulate data for 100 individuals over 50 years of time. The chosen parameters and start values can be found in table 4.1.

Parameter	Value
σ_v^2	$5 \cdot 10^{-5}$
$\sigma_{w_{GDP}}^2, \sigma_{w_{MALE}}^2, \sigma_{w_{EDU}}^2$	$1 \cdot 10^{-5}$
$\sigma_{w_\alpha}^2, \sigma_{w_{AGE}}^2$	$2 \cdot 10^{-5}$
$\boldsymbol{\beta}_0$	$\begin{pmatrix} 0.5 \\ 0.15 \\ 0.02 \\ \boxed{\begin{matrix} 0.5 \\ 0.01 \end{matrix}} \\ \vdots \\ \boxed{\begin{matrix} 0.5 \\ 0.01 \end{matrix}} \end{pmatrix}$

Table 4.1: Parameterization

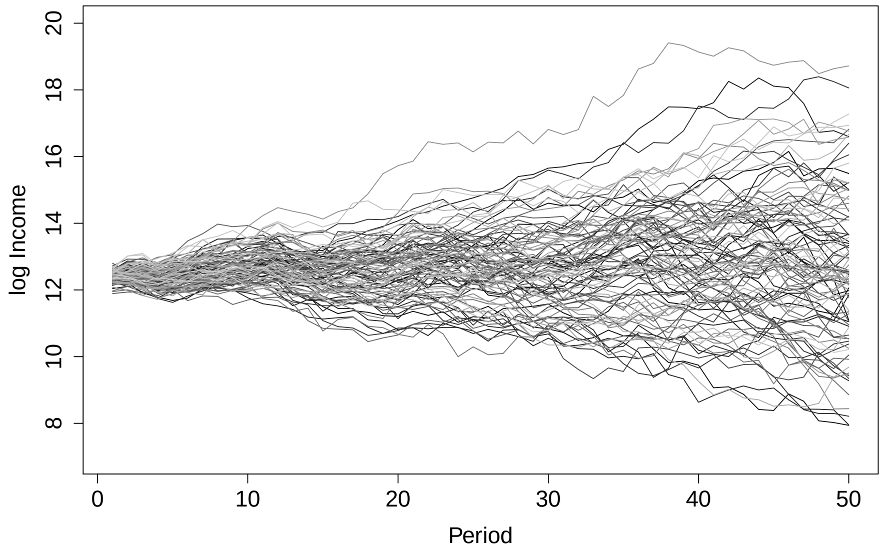

Figure 4.1: Generated log income paths

Figure 4.1 illustrates the artificial data used. It contains the income courses of all individuals which mirror both common and individual-specific effects. Either individual-specific income parameters or idiosyncratic shocks cause a fanning out of income courses over time. More so, individuals are subject to common influences, which is indicated by local peaks in periods 21 and 28.

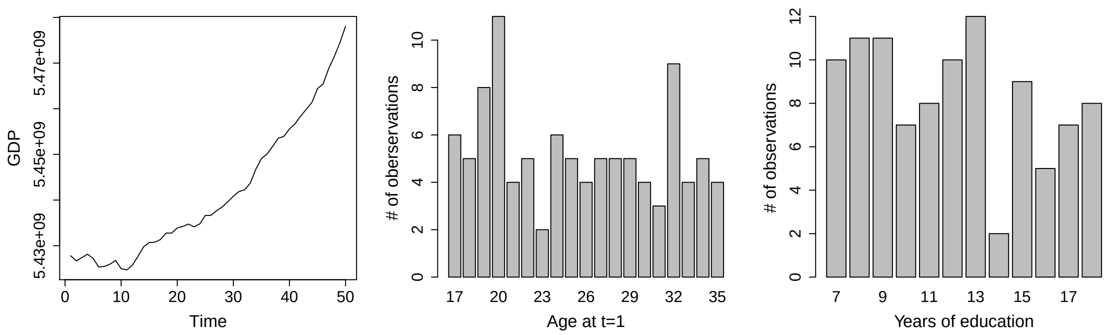

Figure 4.2: Sampled exogenous variables

Figure 4.2 depicts the set of the exogenous variables, namely the randomly generated GDP (following a Brownian motion with drift) and the initial distributions of age and education for all N individuals. Both the initial distributions of age and education are sampled from a uniform distribution within the ranges shown in the corresponding barplots. The gender dummy - not illustrated in this figure - is sampled from a binomial distribution, assuming the probability of drawing a male individual to be 70%.

To initialize our estimation, we choose a (normally/inverse-gamma) prior distribution for the latent states $\boldsymbol{\beta}_t$ and the unknown variance parameters σ_v^2 and Σ_w^2. To test if data patterns are adapted by the method, we center the initial values for the state vector away from the true values. However, they are centered close to them since we have expectations regarding the elasticities between log income and the demographic variables used. Moreover, each coefficient is surrounded by a high degree of uncertainty with $\mathbf{C}_0 = \text{diag}(5, \ldots, 5)$. For the detailed specification of the priors, see table 4.2.

Parameter	Value
$\nu_{0_{GDP}}, \ldots, \nu_{0_{AGE}}$	$2N$
$s_{0_{GDP}}, \ldots, s_{0_{AGE}}$	0.002
a_0	$2N$
b_0	0.01
$\mathbf{m}_0$	$\begin{pmatrix} 0.4 \\ 0.1 \\ 0.01 \\ \boxed{\begin{matrix} 0 \\ 0 \end{matrix}} \\ \vdots \\ \boxed{\begin{matrix} 0 \\ 0 \end{matrix}} \end{pmatrix}$
$\mathbf{C}_0$	$5 \cdot I_{2N+3}$

Table 4.2: Prior specification

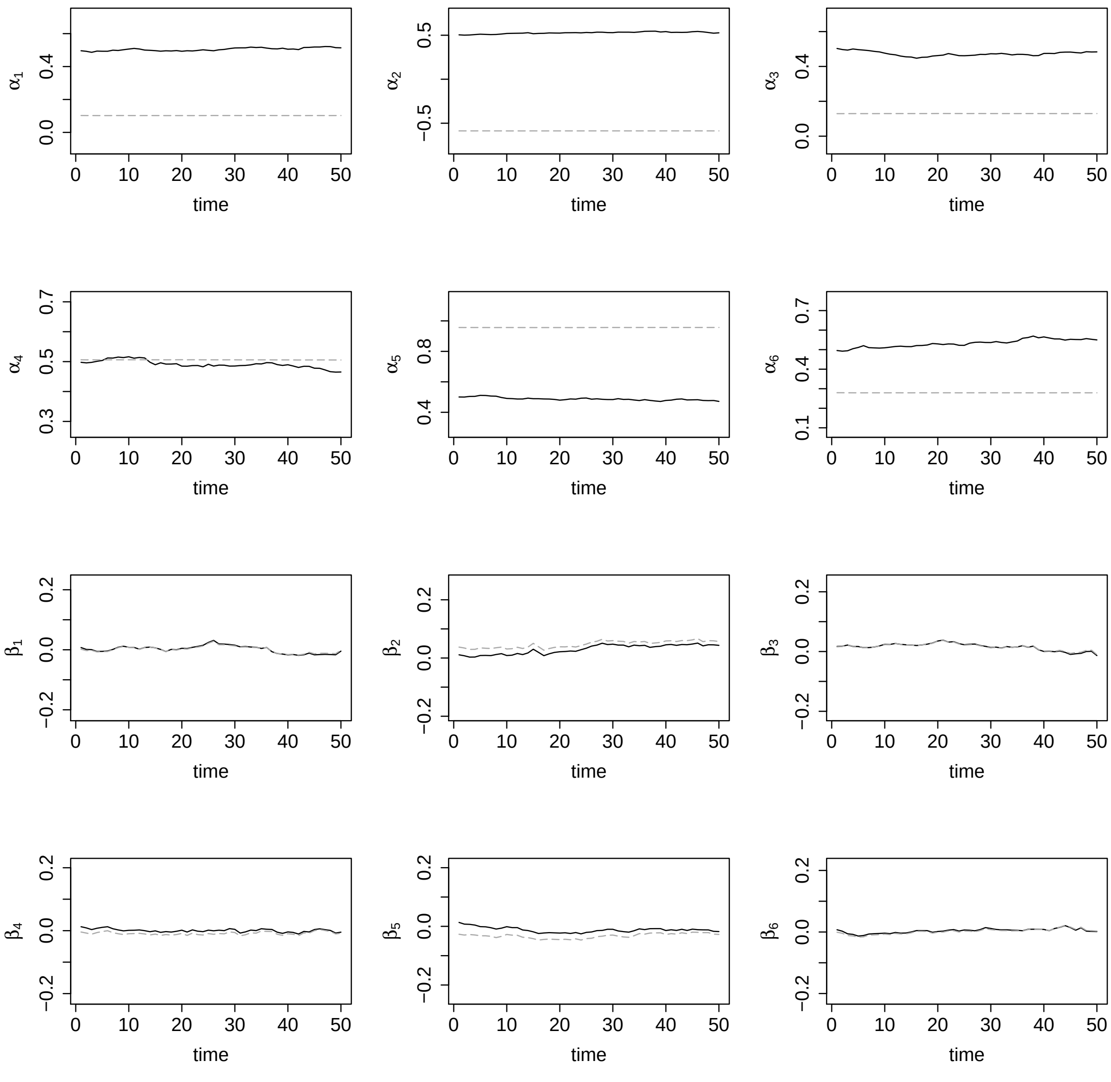

Figure 4.3: True coefficients (black line) vs. estimates (gray dashed line)

We run our estimation with 5000 Gibbs sampling repetitions and assume the burn-in phase to have length 1000. Figure 4.3 compares some of the true (known) values of $\boldsymbol{\beta}_t$ with the corresponding estimates for all periods $t = 1, ..., T$. The results show that the Bayesian method provides quite accurate estimates for the individual slope coefficients since the trajectories are almost identical. In contrast, the estimates for the individual intercepts are almost constant and may obviously not uncover the true coefficients, possibly due to an identification problem. For validation we strongly increase their variance in the data generating process to $\sigma_\alpha^2 = 0.05$. We find that the estimates for the individual intercepts still are almost constant and do not have a unit root.

To focus on the estimates in more detail, we now consider the distribution functions of the estimates of individual intercepts and coefficients.

The left column in figure 4.4 plots the empirical distribution function for α_i (for $i = 1, \ldots, N$) in comparison to the true distribution function (in cross section and for three different periods). It turns out that the median of the estimates is near to their prior of 0 and that the distribution of the estimates is much less centered than the "true" distribution. The right column in figure 4.4 plots the equivalent comparison of the distribution functions for all β_i in the cross section. Here, we find that both distribution functions are quite similar. Moreover, we find that the estimates adapt to the true values away from the priors, and that the medians of both distributions are almost identical.

To derive the expectation of the estimates, we repeat the above estimation 1000 times. However, for ease of calculation, we reduce the number of individuals to $N = 10$. Doing so, it turns out that the average of the differences between all true and estimated cross-sectional individual intercepts converges to 0.5. This exactly is

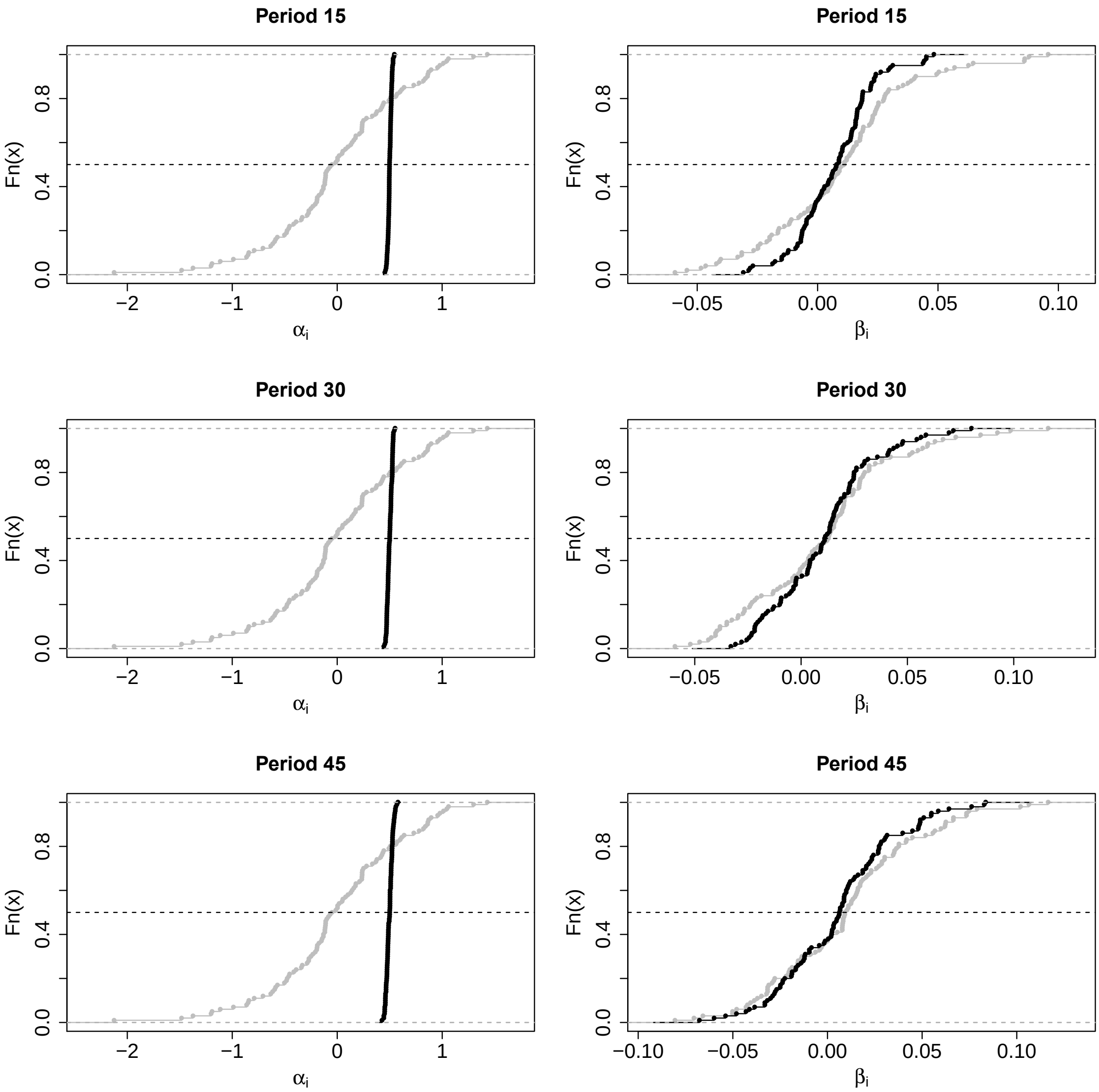

Figure 4.4: Comparison of distribution functions within various periods;
True distribution function (black line) vs. distribution function of estimates (gray line)

the difference between the prior and the true value. A different picture emerges for the individual slope coefficients. Here, the mean difference between the true values and the estimates converges to 0. Figure 4.5 illustrates both.

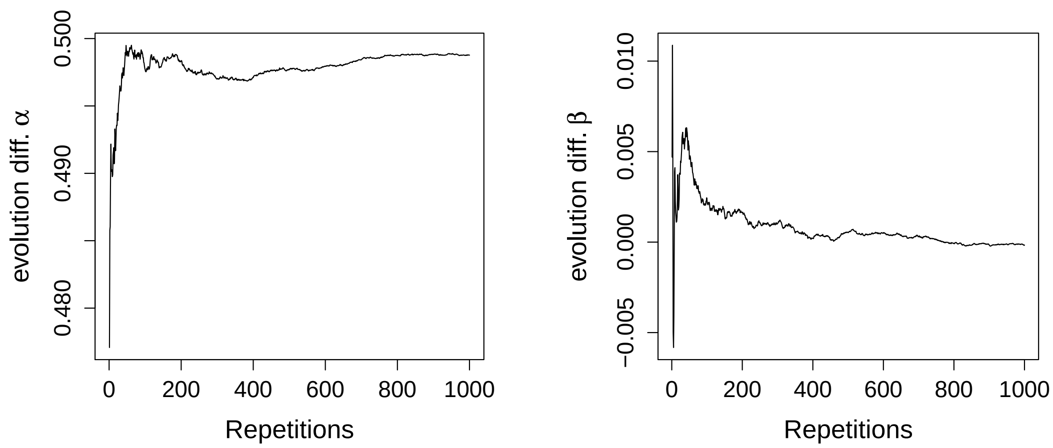

Figure 4.5: Evolution of mean differences

However, regarding the discussion about which model best describes income dynamics, our findings suggest that any interpretation should be based on the individual slope coefficients and not on the intercepts. Indeed, this allows us to examine the extent of heterogeneity in income profiles.

To test whether our results remain unchanged if our data are driven by only one of both effects, we repeat the exercise and simulate data for both the RIP and HIP model separately. We thereby reduce the complexity of our model, because we impose restrictions on the parameters when generating data from the state space model. Firstly, we generate data based on the RIP approach. In this case, the individual slope coefficients $\beta_{AGE_{i,t}}$ are simulated such that they are identical for all individuals within one period. Nonetheless, they are allowed to vary in time and we therefore set $\sigma^2_{w_{\mathrm{AGE}}} = 3 \cdot 10^{-5}$. We also allow for time variation of the individual intercepts by setting $\sigma^2_{w_\alpha} = 0.05$. By doing so, we generate a unit root structure in the intercepts. We simulate data for 50 individuals over 50 years of time.

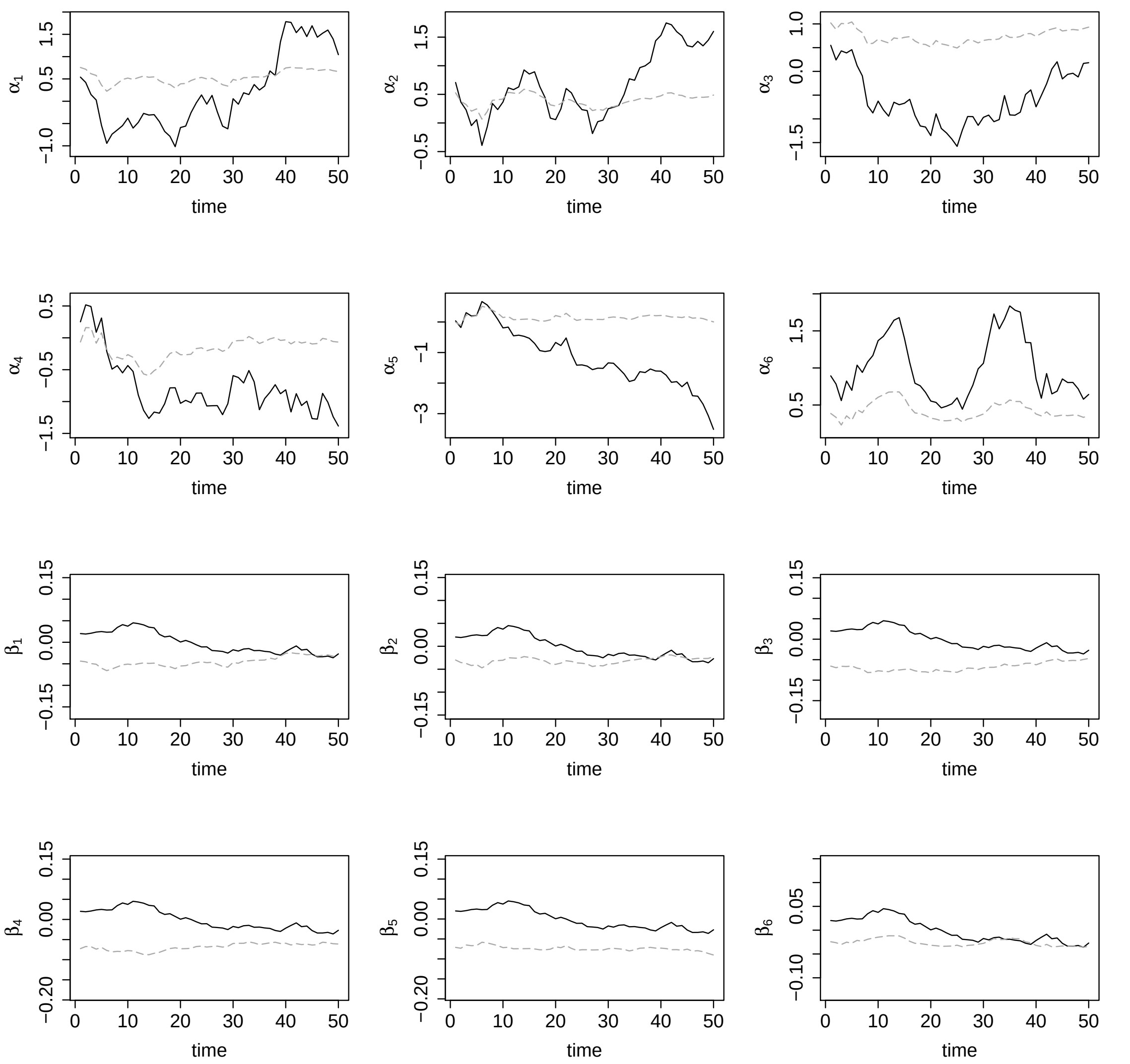

Figure 4.6: RIP: True coefficients (black line) vs. estimates (gray dashed line)

The estimation is initialized with the same priors as before; they are summarized in table 4.2. Figure 4.6 depicts our results of the first six individuals. It turns out that the estimates of the individual intercepts do not entirely reflect the pattern of the true intercepts, however, they still seem to follow a random walk. We fail to reject the null hypothesis of a unit root on every single individual intercept. Moreover, a KPSS test rejects the null hypothesis of stationarity for each trajectory at the 10% significance level. We therefore conclude that the RIP model best describes the earnings data because the unit roots are not detected when both effects are present. The estimates of the slope coefficients, however, seem to differ among individuals.

In a next step, we examine which of the following two models is best supported by our data: either the model which does not impose restrictions on the individual slope coefficients in the estimation, or the model restricting the estimation by assuming $\beta_{AGE_{1,t}} = \beta_{AGE_{2,t}} =, \ldots, = \beta_{AGE_{N,t}}$. To stay in line with our Bayesian framework, we suggest to employ Bayes factors which are the Bayesian analogues of likelihood-ratio tests. They contrast the evidence provided by the data for competing models. In order to compare two models M_j and M_k, the posterior odds in favor of model M_j are computed using Bayes theorem as

$$\frac{p(M_j|\mathbf{y})}{p(M_k|\mathbf{y})} = \frac{p(\mathbf{y}|M_j)}{p(\mathbf{y}|M_k)} \frac{p(M_j)}{p(M_k)},$$

where $p(M_j|\mathbf{y})$ is the probability that model M_j is the correct model, given the data. Moreover, $p(M_j)$ is the prior probability that M_j is the true model. Assigning both models equal prior probabilities as it is commonly done, the prior odds ratio becomes 1. Bayes factor is then given by

$$B_{jk} = \frac{p(\mathbf{y}|M_j)}{p(\mathbf{y}|M_k)},$$

which equals the quotient of the marginal likelihoods of the two models. A large value of B_{jk} is evidence that model M_j is better supported by the data. A small value can be interpreted as evidence in favor of M_k being the better model. Values around 1 indicate that both models are equally well supported. In most cases, the marginal likelihoods can not be derived analytically and must be determined numerically. The

Bayesian information criterion (BIC) was developed as a computationally tractable approximation of the log marginal likelihood of a model. Therefore, the difference between two BIC estimates may be a good approximation of the natural log of the Bayes factor. Formally, the BIC is defined as

$$BIC = -2 \cdot \ln(\hat{L}) + k \cdot ln(N),$$

where k is the number of estimable parameters and $\hat{L}$ is the maximized likelihood.

We compare the BICs of the model which is restricted by $\beta_{AGE_{1,t}} = \beta_{AGE_{2,t}} =, \ldots, = \beta_{AGE_{N,t}}$ and of the unrestricted alternative where the latent state vector equals the estimated one. Note that both likelihoods can easily be extracted from the Kalman filtering of the corresponding models. We exploit that $(\mathbf{y}_t|\mathcal{D}_{t-1}) \sim N((\mathbf{X}_t\mathbf{a}_t), (\mathbf{X}_t\mathbf{R}_t\mathbf{X}_t' + \mathbf{V}))$ and that the sample log likelihood is $\sum_{t=1}^{T} \log f_{y_t|\mathcal{D}_{t-1}}$.

We find vigorous support of the "RIP" data for the restricted model which even remains unchanged when choosing a (non-extreme) prior odds ratio different from 1. Hence, based on the Bayes factor, our data support the assumption that the individuals are subject to a common income profile. For validation of the Bayes factor, we additionally generate data from the HIP model. Again we compare the BICs of the unrestricted model and of the model restricted by $\beta_{AGE_{1,t}} =, \ldots, = \beta_{AGE_{N,t}}$, which we do not expect to be supported by the data. In fact, we find strong evidence in favor of the unrestricted model and therefore against the RIP model.

In sum, our results show that the estimated intercepts have unit roots when the data are generated according to the RIP approach. Moreover, the Bayes factor supports the restriction that individuals are subject to a common income profile. Overall, we conclude that we are able to identify the RIP model if it is indeed the underlying one.

In the heterogeneous case (HIP), the individual slope coefficients $\beta_{AGE_{i,t}}$ are assumed to differ systematically and allowed to vary over time with $\sigma^2_{w_{\mathrm{AGE}}} = 5 \cdot 10^{-5}$. Moreover, the individual intercepts which potentially capture the random walk component of incomes are assumed to be constant over time. However, they may be different among individuals. Again employing the priors in table 4.2, our resulting estimates are presented in figure 4.7. The slope coefficients are well captured and the estimates of the intercepts are constant and again centered around their prior 0. From the figure it clearly appears that there is heterogeneity in income profiles. Moreover, as stressed above, the restriction that the slope coefficients are identical among individuals can be refuted based on the Bayes factor.

As the individual intercepts follow a random walk, we conclude that the RIP approach is the underlying model of the data. In the two other simulation scenarios the unit roots are not detected. If, on the contrary, the estimates of the intercepts are constant, we can reject the RIP model being suitable to describe the data. Beyond, the Bayes factor provides an alternative procedure to examine whether the RIP approach is the underlying model of the data or not. However, if the RIP model is rejected, we can not clearly infer whether elements of both approaches are present in the data or if the fanning out of earnings is solely driven by individual income profiles. Nonetheless, since the slope coefficients are well captured by the estimates in both scenarios, they allow us to determine the extent of heterogeneity in the earnings data.

4.5 The data

For our empirical analysis we use the 1984-2009 waves of the German SOEP (GSOEP). More precisely, we will employ the $PEQUIV-file containing extended

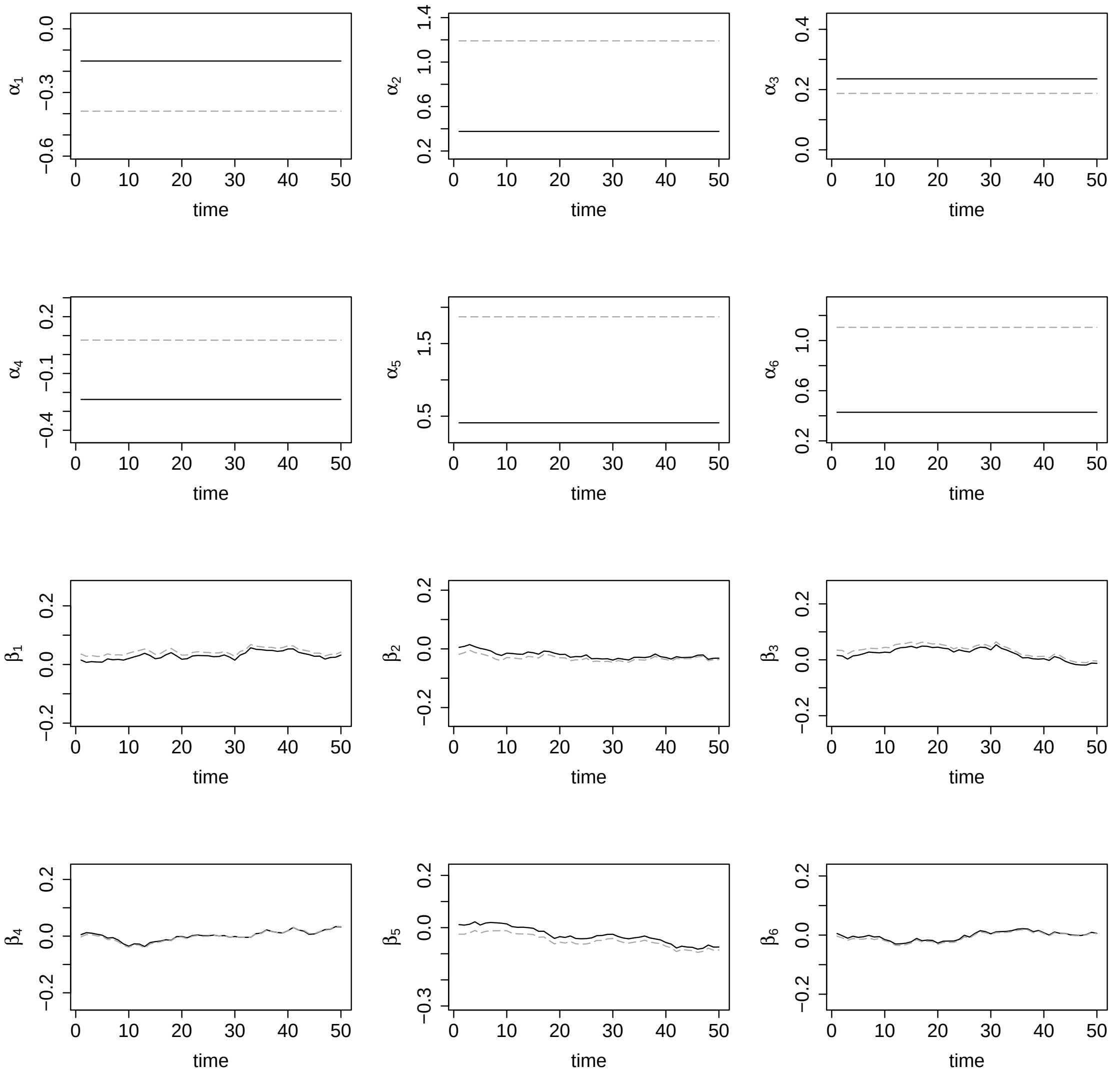

Figure 4.7: HIP: True coefficients (black line) vs. estimates (gray dashed line)

income information on the GSOEP. The \$PEQUIV-file is the German portion of the Cross-National Equivalent File (CNEF) created by the Cornell University in cooperation with the DIW-Berlin, the ISER-Essex and StatsCan-Ottawa. It aims at providing comparable variables from the American PSID, the German SOEP, the British BHPS and the Canadian SLID based on common definitions. Besides the imputation of missing income information, all income variables are annualized and the variable names are matched across all countries. The CNEF further provides income information on an individual basis, while the classical GSOEP contains income information on household basis only.

Since our estimation framework is computationally intensive and the computer's calculation time grows disproportionately with the number of individuals in the data set, the estimation will be carried out for a very limited number of individuals only. This allows the consideration of a balanced panel data set. For this reason (and for obtaining as many periods as possible), we take into account individuals that are present in all waves between 1984 and 2009. More so, the data set consists solely of those individuals that answered all the questions that are crucial to our estimation. Our analysis will be further restricted to individuals that work at least 1000 hours per year in order to exclude those with part time jobs. Following these restrictions, we obtain a data set of size $N = 301$.

However, in order to draw conclusions on the income profiles of the entire population, one additionally has to consider sampling weights to correct for sample disproportionality. The disproportionality may be due to unequal probabilities of selection or to a biased sample by random chance, even if dealing with equal probabilities. Thus, by using weights, a sample can be made representative for the population. As we implicitly deal with time series data for each individual, we need to construct longitudinal weights. In general, they are created by multiplying the cross-sectional weights of the first period by the staying factors of each further period. By definition, staying factors are the inverse of the probability that an individual participates in the named wave, given participation in the previous wave.

Both the cross-sectional weights and the staying factors are contained in the $PEQUIV-file. For simplicity, we deal with one time-invariant weight per individual. Therefore, only the longitudinal weights of the last period are taken into account. From these weights one can easily determine the corresponding individual sampling probabilities. This is achieved by dividing each weight by the sum of all weights.

Finally, the sampling probabilities are applied to the data performing a bootstrapping step. Accordingly a sample of size $N = 301$ is redrawn from the original sample, taking into account the sampling probabilities that were calculated in the previous step. Thus, the new data set implicitly includes the weights and can be used for empirical analysis. Figure 4.8 shows the logarithmic incomes of all individuals in the resampled data set.

4.6 Empirical findings

Our estimation results are presented in the following. Figure 4.9 plots the estimates of three randomly chosen intercepts and six slope coefficients, respectively.

Apparently, the estimates of the individual intercepts do not have a unit root. In particular, they are almost constant and again seem to be subject to an identification problem. Note that this holds not only for the individual intercepts depicted here; this pattern can actually be observed in the entire sample. Figure 4.10 plots the empirical distribution functions for α_i and β_i (for $i = 1, ..., N$), each in cross section

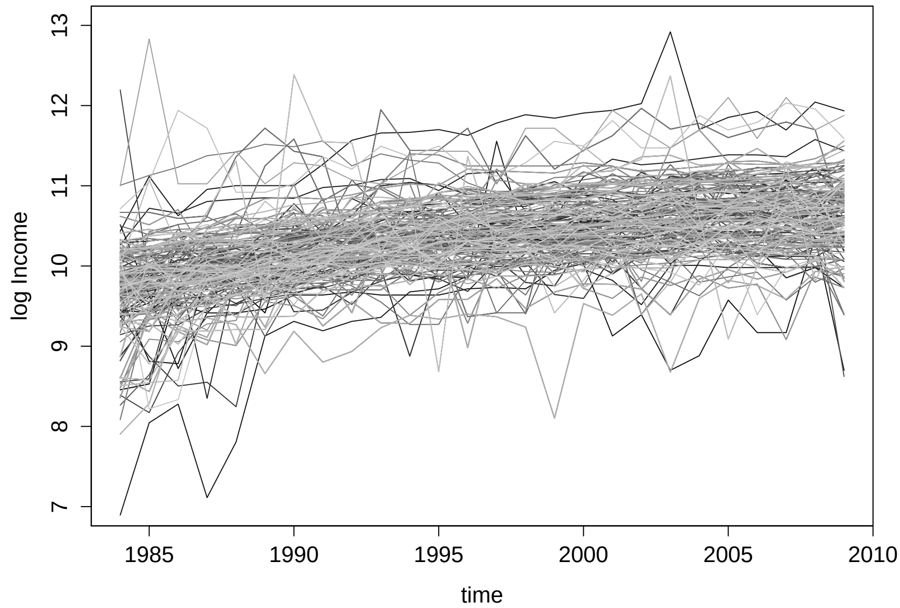

Figure 4.8: Log incomes in the GSOEP sample

and for three different time periods. The left column suggests that the estimates of the intercepts are centered around their prior 0.

Of course, our results need to be interpreted carefully. Overall, they resemble those obtained when the data are simulated according to the HIP model or when elements of both approaches are present in the simulated data: in both simulations, the estimated intercepts do not have a unit root which also applies here. If, on the contrary, the data are simulated according to the RIP approach, we indeed detect a unit root. Therefore, we conclude that the RIP approach is unable to describe our earnings data. This finding is also supported by the fact that there is a large extent of heterogeneity in the individual slope coefficients. These coefficients were reliably uncovered in both simulations - when the HIP model or a mixture of both models is underlying.[3]

[3]Note that foregoing the resampling step in the data preparation does not affect our conclusions.

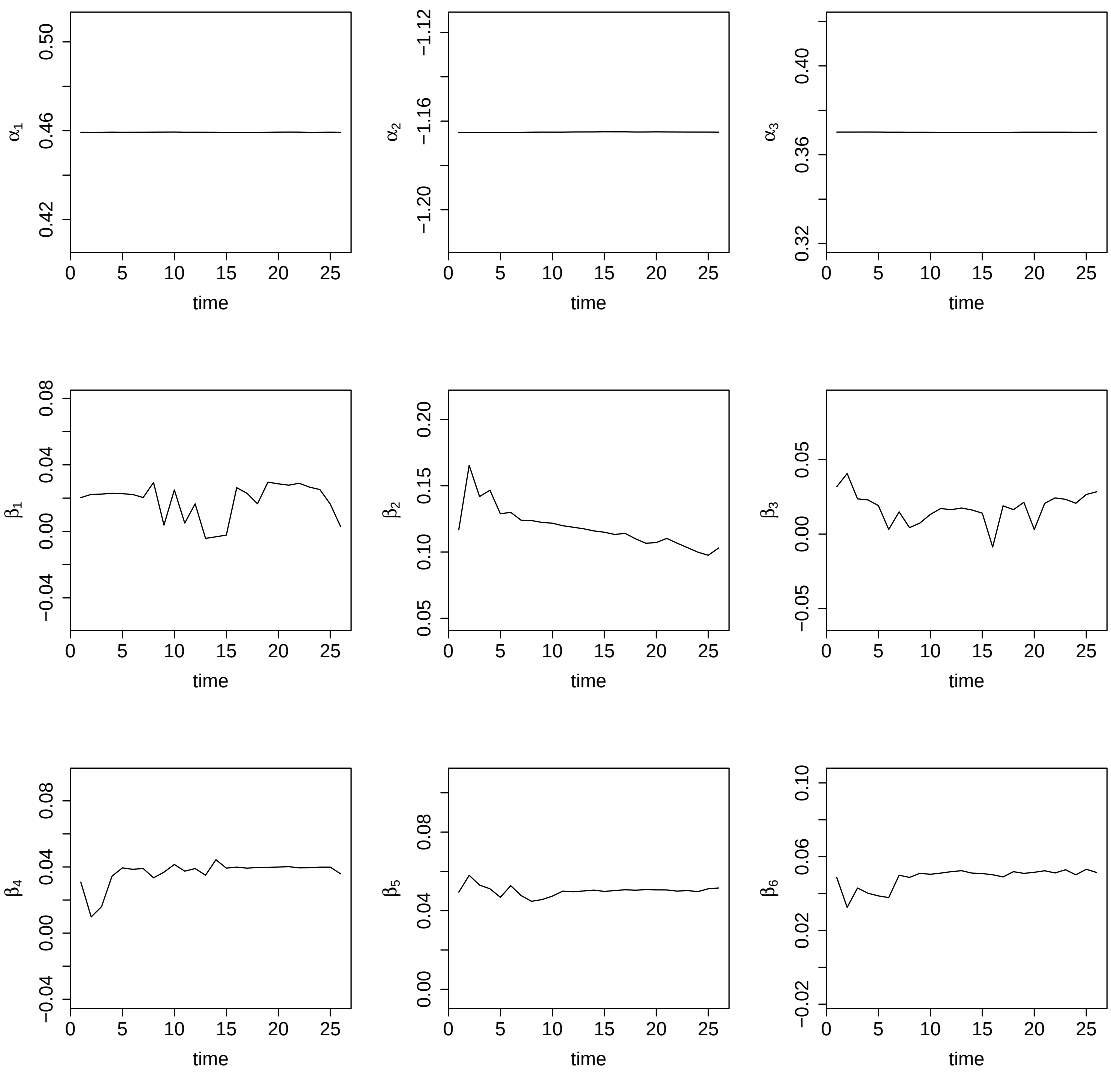

Figure 4.9: Estimates of various intercepts and coefficients (GSOEP)

Calculating the Bayes factor to compare the unrestricted model and the model which is restricted by imposing the individual slope coefficients to be identical in the cross section, it clearly favors the unrestricted model. Hence, we find further confirmation that the data contradict the RIP approach. To sum up, there is no clear evidence in favor of one of the models. Our results, however, suggest that the RIP approach alone is not suitable to describe the present earnings data. We

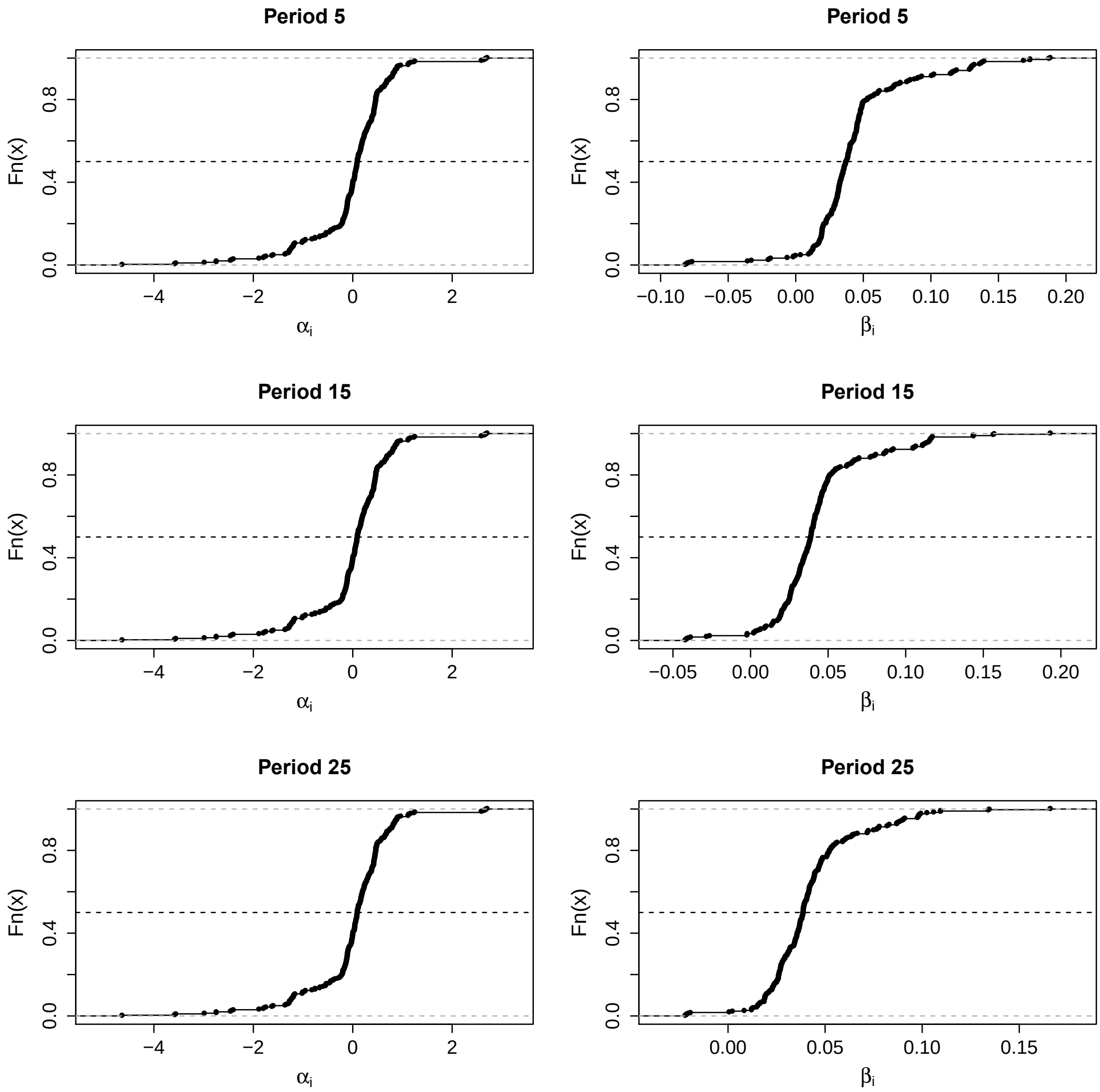

Figure 4.10: Distribution functions of individual parameters (GSOEP)

moreover find a large amount of heterogeneity in the earnings data and deduce that they tend to fan out systematically, rather than due to large and persistent shocks.

4.7 Conclusion

This chapter departs from a present disaccord in the literature about how to model idiosyncratic labor income profiles. Two different approaches have prevailed in the literature that mainly differ in their assumptions on the nature of income growth rates.

The heterogeneous income profile model proclaims that income grows at an individual-specific rate. Income further contains a stochastic, but stationary component, causing the growth rate to fluctuate around the deterministic trend. As a result, income courses differ systematically. The restricted income profile model assumes that individuals are exposed to large and persistent income shocks which account for most differences in income courses. The labor income profiles, however, are similar and hence there is no systematic fanning out of income courses.

To evaluate both approaches, we use a state-space model. This model provides a flexible framework as it enables us to allow for both common and individual coefficients, as well as time-varying coefficients. To estimate the dynamic income model, we use Gibbs sampling, a MCMC algorithm which approximates the joint distribution of the latent states and unknown variance parameters by iteratively sampling from their conditional posteriors. First, a full path of the latent state vector is drawn from its conditional posterior distribution by means of the FFBS algorithm. Second, we employ the conjugacy of the prior distributions. This characteristic enables us to derive the posterior distributions of the variance parameters analytically. After iterating both Gibbs steps, one finally obtains draws from the joint posterior distribution of the latent states and unknown variance parameters.

As an intermediate step, we verify our framework by applying it to simulated data with known coefficients. We first assume that elements of both approaches are contained in the data. In this scenario, the individual-specific slope parameters are estimated precisely, while the intercepts can not be identified. However, in this case conclusions on the nature of income profiles can be drawn from the former as one can determine the variation among individuals' coefficients. Beyond, the same results are found when data are generated according to the HIP approach. When using data obtained from the RIP approach - and only in this case - the estimated intercepts have unit roots and reveal the presence of long-lasting income shocks. Hence not finding a unit root indicates that the data contradict the RIP approach. Moreover, using the Bayes factor provides another way to discriminate between both approaches. We find that it also enables us to reject the RIP model.

For an empirical analysis, our framework is applied to the German SOEP data. Our results indicate that the RIP approach is unsuitable to describe our earnings data - this equally holds when building our interpretation on the structure of the individual intercepts and on the Bayes factor. Instead, we find a large amount of heterogeneity in earnings. Even though we can not completely rule out the possibility that elements of both approaches are present and blurred, it seems that income courses evolve differently in a systematic fashion and do not comprise a high amount of income risk.

Chapter 5

Summary

This thesis investigates labor income dynamics at the individual level and extends the existing literature in several aspects. To begin with, it presents a rather general econometric application on the identification of earnings risk. The two subsequent chapters aim at gaining new insights into the modeling of individual income dynamics. For this purpose, the work first examines if explosiveness is evident in the data and therefore needs to be further taken into account. To this end, a panel unit root test against explosiveness is introduced. Based on the finding that explosiveness can be neglected, the thesis ties in with the disaccord in the literature regarding two major strands of the literature on modeling income profiles. That is, it examines whether the RIP or the HIP approach is more suitable to mirror earnings data. The framework is established in a state space form. For parameter estimation and estimation of the latent state variable MCMC methodology is employed.

Chapter 2 presents an alternative approach to the identification of earnings risk which accounts for economic choices of individuals, in particular the level of consumption and the portfolio allocation. Since there exist hardly any panel data on consumption over longer time spans, the approach allows for employing capital income data instead as these result from consumption and portfolio allocation decisions. Enlarging the information set is expected to increase the estimation accuracy compared to methods that solely rely on earnings data. The economic framework is assumed to be a life-cycle model of consumption and investment decisions, stated as a dynamic programming problem. Earnings risk parameters and, optionally, further parameters of interest are estimated by the simulation-based method of indirect inference. The procedure is applied to simulated data resulting from the (true) life-cycle model. The estimates obtained are centered near the true values and only show little variation. Compared with maximum likelihood estimates solely based on earnings data, no major difference can be observed in terms of the point estimates and MSEs. However, the estimation framework is variably applicable, providing the possibility of estimating further parameters of the economic model. Moreover, it can include different information sets available on capital income and is also employable with short observation periods.

Chapter 3 investigates if explosiveness is a pattern that needs to be taken into account when modeling income profiles. If evident, explosiveness implies that positive deviations tend to boost the income growth rate such that deviations from a common trend will increase even more and vice versa. Here this is called a “self-reinforcing effect”. To this end, this thesis proposes a panel unit root test which tests against explosiveness, using a p-value combination approach. The test procedure is applied to earnings data from the cross-national equivalent files of the German SOEP and the U.S. PSID data sets. The null hypothesis of stationarity or unit roots can be rejected in both countries. Explosiveness is evident, but only for a small fraction of the population. Hence it needs not to be considered when modeling labor incomes.

Chapter 4 proposes a new approach to empirically investigate whether income courses differ systematically due to a large amount of heterogeneity or if the fanning out of incomes is rather driven by persistent income shocks. Pursuant to the literature, the two opposing views are labeled HIP and RIP model, respectively. In order to explore which model is more suitable to correctly mirror labor income data, a dynamic linear model is proposed. It allows for both individual-specific and time-varying coefficients. The latent state and the unknown model parameters are estimated by Gibbs sampling, using artificial as well as real-world data from the German SOEP. A simulation study reveals that the model correctly rejects the RIP approach underlying the data if, in fact, it is not. This conclusion derives from the finding that the individual coefficients do not comprise a unit root structure which would indicate the presence of long-lasting income shocks. Moreover, it is supported by the Bayes factor which evaluates the probability of the RIP approach to be true, given the data. Finally, the framework is applied to the German SOEP data. The results indicate that the real-world income data contradict the RIP approach. However, it is not clear whether the HIP model or a mixture of both approaches is better suitable to describe the data. Since individuals, moreover, are found to be highly heterogeneous, the results are in line with the literature suggesting that incomes rather fan out systematically over time and are not so much driven by large and persistent shocks.

References

Abowd, J. and D. Card (1989). On the covariance structure of earnings and hours changes. *Econometrica 57*(2), 411–445.

Atkinson, A. B., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of Economic Literature 49*, 3–71.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science 344*(6186), 843–851.

Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica 72*(4), 1127–1177.

Baker, M. (1997). Growth-rate heterogeneity and the covariance structure of life-cycle earnings. *Journal of Labor Economics 15*(2), 338–375.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *The Journal of Political Economy*, 352–365.

Bertsekas, D. P. (2000). *Dynamic Programming and Optimal Control* (2nd ed.), Volume 1. Belmont, Massachusetts: Athena Scientific.

Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *Review of Economic Studies 53*(3), 369–384.

Bodie, Z., R. C. Merton, and W. F. Samuelson (1992). Labor supply flexibility and portfolio choice in a life cycle model. *Journal of economic dynamics and control 16*(3), 427–449.

Breitung, J. and S. Das (2005). Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica 59*(4), 414–433.

Browning, M., M. Ejrnaes, and J. Alvarez (2010). Modelling income processes with lots of heterogeneity. *The Review of Economic Studies 77*(4), 1353–1381.

Campagnoli, P., S. Petrone, and G. Petris (2009). *Dynamic Linear Models with R.* Springer-Verlag New York.

Campbell, J. Y. et al. (2001). Investing retirement wealth: A life-cycle model. In J. Campbell and M. Feldstein (Eds.), *Risk Aspects of Investment-Based Social-Security Reform*, Chicago, pp. 439–482. Chicago University Press.

Carroll, C. (2004). Theoretical foundations of buffer stock saving. NBER Working Paper 10867.

Carroll, C. D. and A. A. Samwick (1997). The nature of precautionary wealth. *Journal of monetary Economics 40*(1), 41–71.

Carruth, A. and C. Schnabel (1993). The determination of contract wages in west germany. *The Scandinavian Journal of Economics*, 297–310.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Castaneda, A. et al. (2003). Accounting for the us earnings and wealth inequality. *Journal of Political Economy 111*(4), 818–857.

Chamberlain, G. and K. Hirano (1999). Predictive distributions based on longitudinal earnings data. *Annales d'Economie et de Statistique*, 211–242.

Chang, Y. (2002). Nonlinear IV unit root tests in panels with cross-sectional dependency. *Journal of Econometrics 110*(2), 261–292.

Dixit, A. K. and R. S. Pindyck (1994). *Investment under uncertainty.* Princeton university press.

Douglas, P. H. et al. (1930). Real wages in the united states, 1890-1926.

Dridi, R. and E. Renault (2000). Semi-parametric indirect inference. *LSE STICERD Research Paper No. EM392*.

Fisher, R. A. (1925). *Statistical methods for research workers.* Genesis Publishing Pvt Ltd.

Frick, J. and J. Haisken-DeNew (2005). *Desktop Companion to the German Socio-Economic Panel (SOEP), Version 8.0.* DIW Berlin.

Frühwirth-Schnatter, S. (2008). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*(2), 183–202.

Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory 12*(04), 657–681.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (6), 721–741.

Gottschalk, P. and R. Moffitt (1994). The growth of earnings instability in the U.S. labor market. *Brookings Papers on Economic Activity 2*, 217–272.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, S85–S118.

Gourinchas, P.-O. and J. A. Parker (2002). Consumption over the life-cycle. *Econometrica 70*(1), 47–89.

Greenberg, E. (2007). *Introduction to Bayesian econometrics.* Cambridge University Press.

Gustavsson, M. and P. Österholm (2014). Does the labor-income process contain a unit root? Evidence from individual-specific time series. *Journal of Economic Dynamics and Control 47*, 152–167.

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *American Economic Review 97*(3), 687–712.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics 12*(1), 58–79.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2015). What do data on millions of us workers reveal about life-cycle earnings risk? Technical report, National Bureau of Economic Research.

Guvenen, F. and A. A. Smith (2014). Inferring labor income risk and partial insurance from economic choices. *Econometrica 82*(6), 2085–2129.

Haider, S. J. (2001). Earnings instability and earnings inequality in the United States, 1967–1991. *Journal of Labor Economics 19*(4), 799–836.

Hanck, C. (2013). An intersection test for panel unit roots. *Econometric Reviews 32*(2), 183–203.

Hause, J. (1980). The fine structure of earnings and the on-the-job training hypothesis. *Econometrica 48*, 1013–1029.

Heaton, J. and D. Lucas (1997). Market frictions, savings behavior, and portfolio choice. *Macroeconomic Dynamics 1*(01), 76–101.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika 75*(2), 383–386.

Hryshko, D. (2012). Labor income profiles are not heterogeneous: Evidence from income growth rates. *Quantitative Economics 3*(2), 177–209.

Huggett, M. (1996). Wealth distribution in life-cycle economies. *Journal of Monetary Economics 38*(3), 469–494.

Im, K. S. and M. H. Pesaran (2004). On the panel unit root tests using nonlinear instrumental variables. Working Paper.

Im, K. S., M. H. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of econometrics 115*(1), 53–74.

Jagannathan, R. and N. R. Kocherlakota (1996). Why should older people invest less in stocks than younger people? *Quarterly Review 20*(3), 11–23.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering 82*(1), 35–45.

Katz, L. F. et al. (1999). Changes in the wage structure and earnings inequality. *Handbook of labor economics 3*, 1463–1555.

Keane, M. and A. A. Smith (2003). Generalized indirect inference for discrete choice models. *Department of Economics, Yale University (manuscript)*.

Koo, H. K. (1998). Consumption and portfolio selection with labor income: A continuous time approach. *Mathematical Finance 8*(1), 49–65.

Levin, A., C.-F. Lin, and C.-S. James Chu (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of econometrics 108*(1), 1–24.

Lillard, L. A. and Y. Weiss (1979). Components of variation in panel earnings data: American scientists, 1960-70. *Econometrica 47*(2), 437–454.

Ljungqvist, L. and T. J. Sargent (2000). *Recursive Macroeconomic Theory*. Cambridge: MIT Press.

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics 18*(1), 83–114.

Maddala, G. S. and S. Wu (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and statistics 61*(S1), 631–652.

Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica 72*(1), 1–32.

Mincer, J. A. et al. (1974). Schooling, experience, and earnings. *NBER Books*.

Miranda, M. J. and P. L. Fackler (2004). Applied computational economics and finance. *MIT Press Books*.

Moffitt, R. A. and P. Gottschalk (2002). Trends in the transitory variance of earnings in the United States. *Economic Journal 112*(478), C68–C73.

Monfardini, C. (1998). Estimating stochastic volatility models through indirect inference. *The Econometrics Journal 1*(1), 113–128.

Moon, H. R. and B. Perron (2004). Testing for a unit root in panels with dynamic factors. *Journal of econometrics 122*(1), 81–126.

Nelson, C. R. and C. R. Plosser (1982). Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of monetary economics 10*(2), 139–162.

Perron, P. (1988). Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of economic dynamics and control 12*(2), 297–332.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics 22*(2), 265–312.

Phillips, P. C., S. Shi, and J. Yu (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review 56*(4), 1043–1078.

Phillips, P. C., S. Shi, and J. Yu (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review 56*(4), 1079–1134.

Phillips, P. C. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal 6*(1), 217–259.

Phillips, P. C., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International economic review 52*(1), 201–226.

Prado, R. and M. West (2010). *Time Series: Modeling, Computation, and Inference.* CRC Press.

Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics 51*(3), 239–246.

Seierstad, A. (2009). *Stochastic control in discrete and continuous time*, Volume 1. Springer.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika 73*(3), 751–754.

Slichter, S. (1950). Notes on the structure of wages. *Review of Economics and Statistics 32*, 80–91.

Smith, A. (1776). An inquiry into the nature and causes of the wealth of nations. *The Glasgow edition of the works and correspondence of Adam Smith 2*.

Smith, A. A. (1990). *Three essays on the solution and estimation of dynamic macroeconomic models*. Ph. D. thesis, Duke University.

Smith, A. A. J. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*, S63–S84.

Viceira, L. (2001). Optimal portfolio choice for long-horizon investors with nontradable labor income. *The Journal of Finance 56*(2), 433–470.

Willis, R. J. (1987). Wage determinants: A survey and reinterpretation of human capital earnings functions. *Handbook of Labor Economics 1*, 525–602.

Zeldes, S. (1989). Optimal consumption with stochastic income: Deviations from certainty equivalence. *Quarterly Journal of Economics 104*(2), 275–298.

Appendix A

Lag order determination for the ADF test

The panel test against explosiveness is based on N univariate ADF tests. The number of lags K_i to be included in regression (3.5) can be determined by the Bayesian (or Schwarz) information criterion (Gustavsson and Österholm (2014)). However, if there are measurement errors, the BIC does not, in general, choose a sufficiently large lag order as can be demonstrated by means of a simple Monte-Carlo simulation.

Figure A.1 (solid line) shows the distribution function of the ADF test statistic (with constant and trend) under a unit root when there is no measurement error. The time series is generated as $y_t = z_t + t$ with

$$z_t = z_{t-1} + \eta_t, \qquad t = 1, \ldots, 50$$

where $\eta_t \sim N(0,1)$ is white noise and $z_0 = 0$. The number of lags to be included is chosen by the BIC (of course, the correct number is 0). The number of Monte-Carlo replications is $R = 10000$.

The dashed line in figure A.1 depicts the distribution function of the ADF test statistics if white noise measurement error is added to the time series,

$$y_t = z_t + t + \varepsilon_t$$

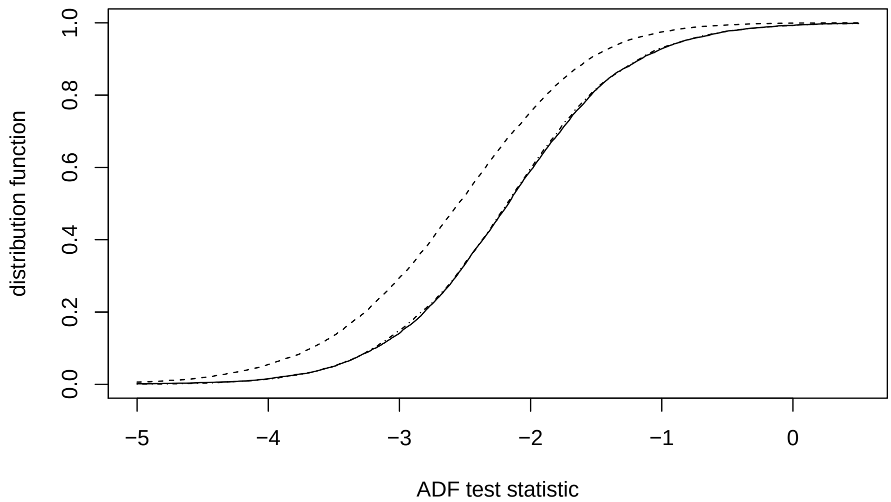

Figure A.1: Distribution function of the ADF test statistic; solid line: distribution under a unit root without measurement error and lag order determined by BIC; dashed line: distribution with measurement error and lag order determined by BIC; dotted-dashed line: distribution with measurement error and lag order set to 3.

where $\varepsilon_t \sim N(0,1)$. Measurement error leads to an MA component in the first differences of y_t. The number of lags has again been selected by the BIC. Apparently, the distribution is shifted to the left and does not equal the null distribution (solid line). We conclude that the BIC does not succeed in selecting the correct lag order. The same result holds for the Akaike information criterion AIC (not shown).

If we set the number of lags to $K_i = 3$ the resulting distribution of the test statistic is shown by the dotted-dashed line in figure A.1. Apparently, it virtually equals the true null distribution (solid line). In our empirical application we therefore set the lag order to $K_i = 3$ for all individuals.