Bayesian Nonparametrics for Financial Volatility Modeling

Inauguraldissertation

zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften

durch die

Wirtschaftswissenschaftliche Fakultät

der

Westfälischen Wilhelms-Universität Münster

vorgelegt von

Martina Danielova Zaharieva M.Sc.

Dekanin: Prof. Dr. Theresia Theurl

Erstgutachter: Prof. Dr. Bernd Wilfling

Zweitgutachter: Prof. Dr. Mark Trede

Tag der mündlichen Prüfung: 09.11.2017

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Financial time series exhibit certain characteristic empirical properties. These *stylized facts* include volatility clustering, serial dependance and heavy tails. An extensive literature on possible approaches to modeling these properties has emerged over the past three decades. Two model classes have predominated, the Generalized Conditional Autoregressive Heteroscedasticity (GARCH) and the Stochastic Volatility (SV) - type models, both being subject to multivariate extensions. Such extensions enable exploring the dependence structure of international financial markets by analyzing the volatility co-movements and by identifying the channels through which volatility transmission occur.

However, more recently, in addition to the advanced econometric techniques, an important role has been played by the increasing availability of computational power. Those developments have made the Bayesian statistical methods generally applicable to complex financial problems, such as volatility modeling and density forecasting under (parameter-) estimation uncertainty.

When it comes to financial market forecasting, the distributional assumptions are of a considerable significance to prediction accuracy. In order to deal with the

non-Gaussian behavior found in financial returns, many alternative distributional assumptions, such as the Student's t-distribution, the generalized hyperbolic skew t-distribution, and the finite mixture of normals, have been proposed. Yet, all of these distribution choices remain parametric and therefore restricted by the respective parameters and properties. In this regard, a new class of semiparametric models have been introduced by combining the existing literature on time-varying volatility models with the advantages provided by Bayesian nonparametric modeling. The resulting modeling framework features a GARCH- or an SV-type specification for the dynamic volatility process and a flexible nonparametric distribution for the error term. The latter is an infinite mixture of Gaussian distributions with weights and component parameters constructed by means of a Dirichlet process mixture (DPM). The Dirichlet process mixture (DPM) prior is a nonparametric Bayesian prior put on an infinite dimensional parameter space, which allows the model to adapt to the complexity of the data, therefore constituting an "infinite capacity model" (Kalli et al. (2013)).

This thesis explores both the methodological and the empirical aspect of certain Bayesian nonparametric methods applied to time-varying financial volatility modeling in the univariate and the multivariate cases. The related issues involve not only designing an appropriate model structure, which provides the desired modeling flexibility, but also accordingly constructing an efficient sampling algorithm. The DPM-based approach allows flexible in-sample estimation and also provides density forecasts for univariate and high-dimensional problems. The flexible framework can be applied to many high-dimensional financial market modeling aspects, such as analyzing volatility co-movements and patterns of volatility transmission.

The thesis comprises three main contributions, in which the complexity of the econometric problem increases gradually, by extending the univariate model to a multivariate one with respect to the volatility framework, as well as to the dimension of the Dirichlet process mixture adopted for modeling the innovation distribution.

The thesis is organized as follows. In Chapter 2, we propose a new and highly flexible Bayesian sampling algorithm for non-linear state-space models under non-parametric distributions. The estimation framework combines a particle filtering and smoothing algorithm for the latent volatility process with a Dirichlet process mixture model for the error term of the observable variables. In particular, we overcome the problem of constraining the models by transformations or the need for conjugate distributions. We use the Chinese restaurant representation of the Dirichlet process mixture, which allows for a parsimonious and generally applicable sampling algorithm. Thus, our estimation algorithm combines a pseudo-marginal Metropolis-Hastings scheme with a marginalized hierarchical semiparametric model. We test our approach for several nested model specifications using simulated data and provide density forecasts. Furthermore, we carry out a real data example using S&P 500 returns.

In Chapter 3, we establish a Cholesky-type multivariate stochastic volatility estimation framework, in which we let the innovation vector follow a Dirichlet process mixture, thus enabling us to model highly flexible return distributions. The Cholesky decomposition allows parallel univariate process modeling and creates potential for estimating highly dimensional specifications. We use Markov Chain Monte Carlo methods for posterior simulation and predictive density computation. We apply our framework to a five-dimensional stock-return data set and analyze international volatility co-movements among the largest stock markets.

In the Chapter 4, we investigate the volatility transmission in bond, foreign exchange and equity markets, by constructing a global trading day divided into three trading zones and revising the concepts of *heat wave* and *meteor shower* effects. In particular, we investigate whether volatility is determined by previous-day volatility in the same region, or rather by volatility in the geographically preceding region. The empirical analysis involves examining the volatility spillovers among trading zones and includes different measures of realized volatility as an explanatory variable. The resulting multivariate GARCH framework is estimated by means of a sophisticated, nonparametric Bayesian approach, designed to deal with the forms of asymmetry and heavy tails found in financial time series. For this purpose, we adopt a Dirichlet process mixture model for the return innovations and parametrically estimate the dynamics of the multivariate volatility.

Chapter 5 summarizes the overall contributions and results of the thesis and outlines ideas for future research. Finally, the Appendix provides details on several methodological aspects discussed in Chapters 2-3.

Chapter 2

Semiparametric Bayesian forecasting with an application to stochastic volatility

2.1 Introduction

Time-varying volatility is a well known stylized fact of financial returns and thus not only its modeling, but especially its estimation and prediction, are of main interest for practitioners and researchers. In particular, Stochastic Volatility (SV) models are widely popular, even though direct estimation by classical Maximum-Likelihood is often infeasible. Nevertheless, Markov Chain Monte Carlo (MCMC) methods, in combination with a sampling algorithm for the latent volatility, as proposed by Jacquier et al. (2004) or Kim et al. (1998), provide a straightforward solution.[1] More recently, Jensen and Maheu (2010) added a further degree of freedom to SV models by augmenting the models with nonparametric distributions based on infinite mixtures. Thus, in addition to the stochastic latent volatility, the error term distribution is highly flexible, which allows, in combination with a Bayesian estimation approach,

[1] See Broto and Ruiz (2004) for a survey.

to learn about the type of distribution from the data. Quite naturally, this enormous flexibility comes at the cost of high complexity as the resulting distributions are possibly non-standard.

The main literature in the field, such as Jensen and Maheu (2010), Delatola and Griffin (2011), Jensen and Maheu (2014), Delatola and Griffin (2013) or Virbickaitė et al. (2014), circumvents this challenge by restricting the model to conjugate distributions and/or transformations of the model equations, but does not offer a generalized solution. Thus, the intended flexibility of a nonparametric model with non-linear effects of stochastic volatility is constrained by analytical feasibility. We argue that this strongly contradicts the motivation of nonparametric/non-linear models. We suggest a new, more general estimation algorithm without artificially pruning the model's dynamics and flexibility.

The point of departure for the present paper is the state-space representation of the (nonparametric) SV model. As such, an SV model is comparable, for example, to a non-linear dynamic stochastic general equilibrium (DSGE) model.[2] For the latter, non-conjugacy and non-standard distributions are widely accepted, and estimation is usually conducted by means of the Metropolis-Hastings (MH) algorithm and particle filter approximations of the likelihood (Fernández-Villaverde and Rubio-Ramírez, 2005). We adopt the same approach and develop our sampling algorithm on an abstract level, using generic distributions without requiring specific distributional assumptions. This allows us to present a modular sampling algorithm which nests nonparametric SV, DSGE, classical SV or even simpler models. Moreover, our presentation is straightforward and strips off the aura of mystery which sometimes surrounds Bayesian nonparametric models. In particular, we use

[2]See Flury and Shephard (2011) for an estimation approach to both model types.

the Chinese restaurant process (CRP) representation of the DPM, which enables an attractive visual representation of the sampling steps.

In our simulation exercises, we show that the new algorithm is highly flexible, reliable and straightforward to apply for several nested model specifications. We also provide a real data example using the semiparametric stochastic volatility of Jensen and Maheu (2010) for S&P 500 data. Furthermore, we demonstrate that our algorithm provides an intuitive way of constructing density forecasts, based on the posterior distributions.

The remainder of the chapter is as follows. Section 2.2 introduces the general setting and preliminary concepts, Section 2.3 presents the sampling algorithm and Section 2.4, an application to the semiparametric stochastic volatility model using simulated and real data. Section 2.5 concludes.

2.2 General setting

2.2.1 Non-linear state-space model

In what follows, we consider an observable variable

$$y_t = g(s_t, \boldsymbol{\theta}, \epsilon_t), \qquad \epsilon_t \overset{iid}{\sim} \mathcal{G}, \tag{2.1}$$

where the latent state variable s_t follows the transition equation

$$s_t = f(s_{t-1}, \boldsymbol{\theta}, \eta_t), \qquad \eta_t \overset{iid}{\sim} \mathcal{F}. \tag{2.2}$$

Furthermore, $g(\cdot)$ and $f(\cdot)$ are potentially non-linear functions, $\boldsymbol{\theta}$ is a parameter vector and $\mathcal{G}$ and $\mathcal{F}$ are continuous random distributions. For parsimony, we work on one-dimensional y_t and s_t, but the above representation applies to multivariate

variables as well. Note that a parametric assumption on $\mathcal{G}$ and $\mathcal{F}$ yields the DSGE model case, and a nonparametric assumption on $\mathcal{G}$ yields the semiparametric SV model case, on which we focus.

2.2.2 Dirichlet process mixture

The Dirichlet process mixture (DPM) represents the distribution of a random variable x_t as an infinite mixture of continuous distributions, for which the mixture component parameters come from a discrete distribution G. In turn, G is constructed from the Dirichlet process prior $\mathrm{DP}(\alpha, G_0)$ (Ferguson (1973)), where α is the *concentration parameter* and G_0 the *base distribution* of the mixture component parameters $\tilde{\mu}_t$ and $\tilde{\sigma}_t$, which we parameterize as a Normal $\mathcal{N}(\cdot)$ and Gamma distribution $\Gamma(\cdot)$, respectively. Throughout Chapter 2, we use mixtures of normals, such that the component parameters are the expected value $\tilde{\mu}_t$ and the standard deviation $\tilde{\sigma}_t$. Following the literature, the hierarchical representation is

$$x_t|(\tilde{\mu}_t, \tilde{\sigma}_t^2) \sim \mathcal{N}(\tilde{\mu}_t, \tilde{\sigma}_t^2), \tag{2.3}$$

$$(\tilde{\mu}_t, \tilde{\sigma}_t^2)|G \overset{iid}{\sim} G, \tag{2.4}$$

$$G|G_0, \alpha \sim \mathrm{DP}(G_0, \alpha), \tag{2.5}$$

$$G_0(\tilde{\mu}_t, \tilde{\sigma}_t^2) = \begin{cases} \mathcal{N}(m_0, v_0^2) \\ \\ \Gamma(a_0, b_0) \end{cases}, \tag{2.6}$$

where a_0, b_0, m_0 and v_0 are hyperparameters.

2.2.3 Chinese Restaurant Process (CRP)

Our estimation algorithm is based on the CRP representation of the DPM, which represents the mixture components as tables in a restaurant, the component parameters as the location inside the restaurant and observations y_t $(t = 1, \ldots, T)$ as customers entering the restaurant.[3] Before introducing the CRP in more detail, we clarify the notation in order to avoid ambiguity.

Let z_t be a label denoting which of the $k \in \{1, 2, \ldots, \infty\}$ tables (components) a customer (observation) y_t is assigned to and $z_t = k$ if customer y_t sits at table k. Furthermore, let c_k be the number of customers sitting at table k in the restaurant and define the *nonparametric* set $\boldsymbol{\phi}_k = \{\mu_k, \sigma_k\}$, which contains the parameters of component k. Thus, we have $\tilde{\mu}_t = \mu_{z_t}$ and $\tilde{\sigma}_t = \sigma_{z_t}$. Given this notation, the CRP can be summarized in two simple steps:

1. For $t = 1$:

   The first customer y_1 sits at the first table with probability 1. Thus we have $z_1 = 1$. The parameters of the first component, indexed by $k = 1$, are sampled from the base distribution, i.e. $\boldsymbol{\phi}_1 \sim G_0$.

2. For $t = 2, \ldots, T$:

   The t-th customer sits at any of the occupied tables $k \in \{1, \ldots, n\}$ with probability $\propto c_k$ or at a non-occupied table with probability $\propto \alpha$. Whenever a new table is chosen, indexed by $n + 1$, sample $\boldsymbol{\phi}_{n+1} \sim G_0$. In particular, it holds that

$$P(z_{t+1} = k | \mathbf{z}_{1:t}, \alpha) = \frac{c_k}{t + \alpha},$$

[3] Alternative representations are the stick-breaking representation (Sethuraman, 1994) or the closely related Pólya urn scheme (Blackwell and MacQueen, 1973). An overview is available in Teh (2011).

$$P(z_{t+1} = n+1 | \mathbf{z}_{1:t}, \alpha) = \frac{\alpha}{t+\alpha}, \tag{2.7}$$

where $\mathbf{z}_{1:t} = \{z_1, \ldots, z_t\}$.

Note that the number of possible tables is unrestricted and the corresponding density is

$$p(z_{t+1} | \mathbf{z}_{1:t}, \alpha) = \frac{c_k}{t+a}\delta(z_{t+1} = k) + \frac{\alpha}{t+a}\delta(z_{t+1} = n+1),$$

where $\delta(\cdot)$ is the Dirac delta function [4]. Therefore, the model capacity in terms of the parameter space is infinite. Nevertheless, the number of occupied tables is constrained by $n \leq T$. We refer to n as the number of *active* tables or *non-neglectable* components. Note that the process outlined above exhibits the typical *rich-gets-richer* property, i.e. clustering of the customers. Furthermore, as the probability of creating a new table is proportional to α, a small (large) value of α leads to fewer (more) non-empty components. Thus, the value of the concentration parameter α is of major importance. For that reason, our estimation approach additionally imposes a hyperprior on α, in order to achieve higher flexibility. The likelihood of the indicators $P(\mathbf{z}_{1:T}|\alpha)$ can be decomposed as

$$\begin{aligned} P(\mathbf{z}_{1:T}|\alpha) &= P(z_T|\mathbf{z}_{1:T-1}, \alpha)P(z_{T-1}|\mathbf{z}_{1:T-2}, \alpha)\ldots P(z_2|z_1, \alpha) \\ &= \prod_{i=1}^{T-1} P(z_{T+1-i}|\mathbf{z}_{1:T-i}, \alpha). \end{aligned} \tag{2.8}$$

[4] Here, the Dirac delta function is used as an indicator function, i.e. given the statement A, we define $\delta(A) \equiv \begin{cases} 1 & \text{if } A \text{ is true,} \\ 0 & \text{else.} \end{cases}$

Essentially, using the CRP, we study the marginalized hierarchical semiparametric model

$$x_t|(\tilde{\mu}_t, \tilde{\sigma}_t^2) \sim \mathcal{N}(\tilde{\mu}_t, \tilde{\sigma}_t^2),$$

$$(\{\tilde{\mu}_1, \tilde{\sigma}_1^2\}, \ldots, \{\tilde{\mu}_T, \tilde{\sigma}_T^2\})|(G_0, \alpha) \sim p(\{\tilde{\mu}_1, \tilde{\sigma}_1^2\}, \ldots, \{\tilde{\mu}_T, \tilde{\sigma}_T^2\}|G_0, \alpha),$$

$$G_0(\tilde{\mu}_t, \tilde{\sigma}_t^2) = \begin{cases} \mathcal{N}(m_0, v_0^2) \\ \\ \Gamma(a_0, b_0) \end{cases},$$

where $p(\{\tilde{\mu}_1, \tilde{\sigma}_1^2\}, \ldots, \{\tilde{\mu}_T, \tilde{\sigma}_T^2\}|G_0, \alpha)$ is the joint density of the component parameters constructed by the CRP. Therefore, recall that the CRP with indicators $\mathbf{z}_{1:T}$ and component parameters $\boldsymbol{\phi}_{1:n}$ straightforwardly implies the conditional density $p(\{\tilde{\mu}_T, \tilde{\sigma}_T^2\}|\{\tilde{\mu}_1, \tilde{\sigma}_1^2\}, \ldots, \{\tilde{\mu}_{T-1}, \tilde{\sigma}_{T-1}\}, G_0, \alpha)$ in closed form. Due to the exchangeability of $\{\tilde{\mu}_t, \tilde{\sigma}_t^2\}$ [5], this enables us to calculate conditional distributions for all other t as well, and is thus the impetus for the Gibbs sampling approach, which we apply in Section 2.3.1.1.

2.3 Bayesian inference

Let $\mathbf{z}_{-t}$ denote the set of table assignments $\mathbf{z}_{1:T} = \{z_1, z_2, ..., z_T\}$ without assignment z_t, i.e. $\mathbf{z}_{-t} = \{z_1, \ldots, z_{t-1}, z_{t+1}, \ldots, z_T\}$, analogously $\boldsymbol{\phi}_{1:n,-k}$ the set of component parameters $\boldsymbol{\phi}_{1:n}$ except for ϕ_k, i.e. $\boldsymbol{\phi}_{1:n,-k} = \{\phi_1, \ldots, \phi_{k-1}, \phi_{k+1}, \ldots, \phi_n\}$. Then, $\{\phi_k, \boldsymbol{\phi}_{1:n,-k}, \boldsymbol{\phi}_{n+1:\infty}\}$ equals the full (infinite) parameter set $\boldsymbol{\phi}_{1:\infty}$. The objective is to sample from the joint posterior density

$$p(\mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha|\mathbf{y}_{1:T}).$$

[5] For the main properties of the DPM, see e.g. Teh (2011).

Our sampling approach extends Algorithm 5 of Neal (2000) to latent variables. In contrast to Jensen and Maheu (2010) or Delatola and Griffin (2011), for example, this imposes no restrictions on the distributions with regard to conjugacy. In particular, we break the sampling algorithm down into four major steps:

(A) DPM,

(B) latent variables,

(C) parameters,

(D) hyperparameter,

where each step deals with several conditional posteriors in the tradition of Gibbs blocking. We discuss each step in detail in the following sections.

2.3.1 Sampling algorithm

We initialize the algorithm by drawing from the priors of $\boldsymbol{\theta}$ and α, simulating the CRP, conditional on α, and subsequently running the particle smoothing algorithm to obtain initial values for the latent variables.

2.3.1.1 Step (A): DPM

In order to obtain a posterior sample from the DPM, we require draws from the posteriors of the table indicators $\mathbf{z}_{1:T}$ and the infinite parameter set $\boldsymbol{\phi}_{1:\infty}$. In particular, we use two Gibbs blocks, i.e. sampling from

A.1. $p(\mathbf{z}_{1:T}|\mathbf{y}_{1:T}, \mathbf{s}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha)$,

A.2. $p(\boldsymbol{\phi}_{1:\infty}|\mathbf{y}_{1:T}, \mathbf{s}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\theta}, \alpha)$.

To sample the table indicators (Block A.1.) we use a version of Algorithm 5 from Neal (2000). Given the states, we iteratively draw from

$$p(z_t|\mathbf{z}_{-t}, \mathbf{y}_{1:T}, \mathbf{s}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha)$$
$$\propto p(\mathbf{y}_{1:T}|z_t, \mathbf{z}_{-t}, \mathbf{s}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha) p(z_t|\mathbf{z}_{-t}, \mathbf{s}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha)$$

for all z_t with $t = 1, \ldots, T$ using an MH algorithm with a proposal equal to the prior. Thus, the acceptance probability reduces to

$$\min \left\{ 1, \frac{p(\mathbf{y}_{1:T}|\tilde{z}_t, \mathbf{z}_{-t}, \mathbf{s}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha)}{p(\mathbf{y}_{1:T}|z_t, \mathbf{z}_{-t}, \mathbf{s}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha)} \right\}, \tag{2.9}$$

where $\tilde{z}_t$ denotes a candidate table indicator. Conditioned on $\boldsymbol{\phi}_{1:\infty}$, states $\mathbf{s}_{1:T}$ and table assignments $\mathbf{z}_{1:T}$, the required likelihood $p(\mathbf{y}_{1:T}|z_t, \mathbf{z}_{-t}, \mathbf{s}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha)$ is straightforward to calculate from Eq. (2.3).

The acceptance probability in (2.9) is valid, if the proposal $\tilde{z}_t$ is drawn from the conditional $p(z_t|\mathbf{z}_{-t}, \mathbf{s}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha)$. Noting that the latter distribution is by construction independent of $\boldsymbol{\phi}_{1:\infty}$, $\mathbf{s}_{1:T}$ and $\boldsymbol{\theta}$, this is equivalent to sampling from $P(z_t|\mathbf{z}_{-t}, \alpha)$. Via the CRP definition, we know that the distribution of the cluster pattern is exchangeable, i.e. the current z_t can be regarded as the last customer entering the restaurant. Thus, a candidate table can be drawn from a multinomial distribution, constructed from the probabilities given in Eq. (2.7). Hence, the current customer re-enters the restaurant filled with the remaining $T-1$ customers and gets assigned either to a new or to an existing table. Denoting table counts, excluding the current customer by $c_{k,-t}$, the probabilities for sitting at one of the occupied tables $k = 1, \ldots, n$ and opening a new table are given by

$$P(\tilde{z}_t = k|\mathbf{z}_{-t}, \alpha) = \frac{c_{k,-t}}{T - 1 + \alpha},$$

$$P(\tilde{z}_t = n+1|\mathbf{z}_{-t}, \alpha) = \frac{\alpha}{T-1+\alpha},$$

respectively.

Block A.2. is designed to sample the infinite parameter set $\boldsymbol{\phi}_{1:\infty}$ from

$$p(\boldsymbol{\phi}_{1:\infty}|\mathbf{y}_{1:T}, \mathbf{s}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\theta}, \alpha).$$

Any empty (non-active) tables can be neglected, since

$$p(\mathbf{y}_{1:T}|\boldsymbol{\phi}_{-k,1:n}, \tilde{\phi}_k, \mathbf{s}_{1:T}, \mathbf{z}_{1:T}, \alpha) = p(\mathbf{y}_{1:T}|\boldsymbol{\phi}_{-k,1:n}, \boldsymbol{\phi}_{n+1:\infty}, \tilde{\phi}_k, \mathbf{s}_{1:T}, \mathbf{z}_{1:T}, \alpha),$$

i.e. a sample from the posterior of an empty table is obtained by simply drawing from the base distribution G_0. Thus, Block A.2. iterates through all active tables $k = 1, \ldots, n$ and samples from

$$p(\phi_k|\mathbf{y}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\phi}_{-k}, \boldsymbol{\theta}, \alpha) \propto p(\mathbf{y}_{1:T}|\boldsymbol{\phi}_{1:n,-k}, \phi_k, \mathbf{s}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\theta}, \alpha) G_0(\phi_k)$$

using a random walk MH step with acceptance probability

$$\min\left\{1, \frac{p(\mathbf{y}_{1:T}|\boldsymbol{\phi}_{1:n,-k}, \tilde{\phi}_k, \mathbf{s}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\theta}, \alpha) G_0(\tilde{\phi}_k)}{p(\mathbf{y}_{1:T}|\boldsymbol{\phi}_{1:n,-k}, \phi_k, \mathbf{s}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\theta}, \alpha) G_0(\phi_k)}\right\}.$$

In the Chinese Restaurant interpretation, this step can be regarded as moving around the occupied tables within the restaurant.

2.3.1.2 Step (B): Latent variables

Step (B) samples from

$$p(\mathbf{s}_{1:T}|\mathbf{y}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha).$$

In particular, we use a particle smoother approximation and draw from a multinomial distribution using the weights and particles from the particle smoother.

The particle filter proceeds in the spirit of Flury and Shephard (2011), see Herbst and Schorfheide (2015) for details. The idea is to approximate all required densities by a particle swarm defined as the set $\{\mathbf{s}_t, \mathbf{w}_t\}$, in which $\mathbf{s}_t \in \mathbb{R}^{N_p}$, $\mathbf{w}_t \in \mathbb{R}^{N_p}$ and N_p is the number of particles. Iterating on forecasting and updating steps, the weights $\mathbf{w}_t$ enable us to track the evolution of the swarm over time. That is, we start with a randomly drawn swarm with weights equal to unity. Subsequently, according to Bayes' Theorem, the weights are updated conditional on the observation y_t. Thus, the particle filter approximates the integral

$$p(y_t|\mathbf{y}_{1:t-1}, \mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha) = \int p(y_t|s_t, \mathbf{y}_{1:t-1}, \mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha) p(s_t|\mathbf{y}_{1:t-1}, \mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha) ds_t,$$

by taking the mean over the appropriate set of particle weights. Given these *incremental likelihoods*, we are able to calculate an unbiased particle filter approximation of the log-likelihood

$$\log p(\mathbf{y}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha) = \sum_{t=1}^{T} \log p(y_t|\mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \mathbf{y}_{1:t-1}, \alpha).$$

Appendix A.3 provides details on the particle filter.

We use the reweighting particle smoother (Doucet et al. (2000)) to obtain draws $s_t|\mathbf{y}_{1:T}$ for all t. The idea behind the smoothing algorithm is to reweight the particles by Bayes' rule, in order to obtain an approximation of the smoothed distribution of s_t, which is given by

$$p(s_t|\mathbf{y}_{1:T}) = p(s_{t+1}|\mathbf{y}_{1:t}) \int \frac{p(s_{t+1}|s_t) p(s_{t+1}|\mathbf{y}_{1:T})}{p(s_{t+1}|\mathbf{y}_{1:t})} \, ds_{t+1}.$$

We refer to Särkkä (2013) for a textbook treatment.

2.3.1.3 Step (C): Parameters

The third block is a canonical random walk MH algorithm, which samples from

$$p(\boldsymbol{\theta}|\mathbf{y}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \alpha) \propto p(\mathbf{y}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \alpha)p(\boldsymbol{\theta}).$$

For brevity, we refer to Greenberg (2008) for details. Note that we integrate out the latent states and use the incremental likelihoods generated by the particle filter in order to obtain an unbiased approximation to the likelihood. That is, Step C is equivalent to the pseudo-marginal method discussed for example, by Pitt et al. (2012) or Doucet et al. (2015).

2.3.1.4 Step (D): Hyperparameter

The last step samples the concentration parameter α. Conditional on the indicators $\mathbf{z}_{1:T}$, the posterior of α is independent of $\mathbf{y}_{1:T}$, $\boldsymbol{\phi}_{1:\infty}$ and $\boldsymbol{\theta}$, i.e.

$$p(\alpha|\mathbf{y}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}) = p(\alpha|\mathbf{z}_{1:T}) \propto p(\mathbf{z}_{1:T}|\alpha)p(\alpha),$$

implying that the concentration parameter depends *exclusively* on the clusters. Furthermore, the CRP yields a straightforward rule for calculating the conditional density of the cluster pattern (likelihood of $\mathbf{z}_{1:T}$ given α) from eq. (2.7), which we use to calculate the acceptance probability

$$\min\left\{1, \frac{p(\mathbf{z}_{1:T}|\tilde{\alpha})p(\tilde{\alpha})}{p(\mathbf{z}_{1:T}|\alpha)p(\alpha)}\right\}$$

for a random walk MH algorithm. Note that the indicators $\mathbf{z}_{1:T}$ are labels, and are exchangeable (label switching), as only the cluster pattern matters for the pro-

bability $p(\mathbf{z}_{1:T}|\alpha)$. The issue of label switching in the context of Dirichlet process mixtures is addressed in more detail by Jensen and Maheu (2010).

2.3.2 Savage-Dickey Density Ratio

Even though our approach deviates from the conjugate priors used in the literature, we are able to calculate the Bayes factors in favor of nested models using the Savage-Dickey density ratio (Dickey (1971)), as in Jensen and Maheu (2010). Nevertheless, a slightly more general definition is required to preserve interpretability.

Consider the nested model specification, $M_2 : \alpha = \alpha_0$, for which the limiting cases $\alpha_0 = \{0, \infty\}$ correspond to a normally distributed error term and a t-distributed error term, respectively. Denoting the unrestricted model by M_1, the Bayes factor is

$$\begin{aligned} \text{BF}(\alpha = \alpha_0) &= \frac{p(\mathbf{y}_{1:T}|M_2)}{p(\mathbf{y}_{1:T}|M_1)} \\ &= \frac{p(\alpha = \alpha_0|\mathbf{y}_{1:T}, M_1)}{p(\alpha = \alpha_0)}, \end{aligned}$$

i.e. the ratio of the posterior density of α to its prior, both evaluated at α_0. As the hypothesis of $\alpha_0 \to \infty$ is not operational, we follow Jensen and Maheu (2010) and define the transformed variable

$$u = \frac{\alpha}{\alpha + 1},$$

and thus $u \to 1$ ($u \to 0$) as $\alpha \to \infty$ ($\alpha \to 0$). Note that u is the probability of a second component. Using the transformation, it holds that

$$\text{BF}(u = u_0) = \frac{p(u = u_0|\mathbf{y}_{1:T}, M_1)}{p(u = u_0)}.$$

In contrast to Jensen and Maheu (2010), we do not impose the restriction $p(u) = \mathcal{U}(0,1)$, such that the approximation of the Savage-Dickey ratio by the posterior draws of u has to be corrected using the prior density of u, which is calculated from the prior of α using the transformation rule

$$p(u) = \frac{p(\alpha = \frac{u}{1-u})}{(1-u)^2}.$$

Thereupon, plots of the rescaled posterior of u carry the same information as in Jensen and Maheu (2010) and can be interpreted equivalently. In particular, the value of the Savage-Dickey ratio can be interpreted as the Bayes factor in favor of the nested models defined by the value on the abscissa.

2.3.3 Density forecast

In line with Jensen and Maheu (2010), we construct the posterior density forecast

$$p(y_{T+1}|\mathbf{y}_{1:T}) = \int p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \mathbf{z}_{1:T}, s_{T+1}, \alpha) p(\boldsymbol{\phi}_{1:\infty}, \boldsymbol{\theta}, \mathbf{z}_{1:T}, \alpha|\mathbf{y}_{1:T})\, d\boldsymbol{\phi}_{1:\infty}\, d\boldsymbol{\theta}\, ds_{T+1}\, d\mathbf{z}_{1:T},$$

by means of the MCMC output

$$\hat{p}(y_{T+1}|\mathbf{y}_{1:T}) = \frac{1}{N}\sum_{i=1}^{N} p_{\mathcal{N}}(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\phi}_{1:\infty}^{(i)}, \boldsymbol{\theta}^{(i)}, s_{T+1}^{(i)}, \mathbf{z}_{1:T+1}^{(i)}, \alpha^{(i)}).$$

Given draw $i = 1, \ldots, N$ from the posterior of $\boldsymbol{\phi}_{1:\infty}$, $\boldsymbol{\theta}$ and $\mathbf{z}_{1:T}$, we run the particle smoother and draw a latent state s_T. Given s_T, we can generate a draw s_{T+1} with the transition Eq. (2.2). Subsequently, we iterate the CRP forward conditional on $\mathbf{z}_{1:T}$, which either yields $z_{T+1} \in \mathbf{z}_{1:T}$ or a new component with probability $\propto \alpha$. In the latter case, we sample $\tilde{\mu}_{T+1}$ and $\tilde{\sigma}^2_{T+1}$ from the base distribution G_0. In either case, given the drawn component parameters $\tilde{\mu}_{T+1}$ and $\tilde{\sigma}^2_{T+1}$, it is straightforward to draw y_{T+1} using the observation Eq. (2.1).

2.3.4 Nested models

Besides the full semiparametric model, which we study in detail in Section 2.4, our sampling algorithm nests several model specifications and is easily adapted.

If $\mathcal{G}$ is nonparametric, and s_t a constant, we can use Steps A, C and D without the filtering step. This case corresponds to a standard DPM model, where, for example, a density estimate is required (Walker (2007)). Appendix A.1 provides an example.

If we assume a parametric distribution $\mathcal{G}$ and latent s_t, we only require the particle filter in combination with Step C. This is, for example, the case in DSGE models, see e.g. Fernández-Villaverde et al. (2016) and Herbst and Schorfheide (2015), or standard SV models as shown in Appendix A.2.

2.4 Semiparametric stochastic volatility model

The semiparametric stochastic volatility model of Jensen and Maheu (2010) is a suitable application of the sampling algorithm, as it incorporates the flexibility of the nonparametric error term into a non-linear state-space representation of the time-varying volatility model. In particular, the model is defined by

$$y_t = \exp(h_t/2)\epsilon_t, \qquad \epsilon_t \sim \mathcal{G}, \tag{2.10}$$

$$h_t = \rho h_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \tag{2.11}$$

where the log volatility h_t is the latent state variable, $\mathcal{G}$ an unknown distribution and $\boldsymbol{\theta} = \{\rho, \sigma_\eta\}$. Note that we set the unconditional expectation of the latent volatility equal to zero, such that the level of the volatility is captured by the nonparametric $\mathcal{G}$, ensuring identification of the SV model.

2.4.1 Simulated data

Prior to adopting our sampling algorithm to real-world data, we implement the approach using simulated data. We simulate 1500 data points according to Eqs. (2.10) and (2.11) with parameter values $\rho = 0.95$ and $\sigma_\eta^2 = 0.04$ and a mixture distribution for the simulated error term given by

$$\epsilon_t \overset{iid}{\sim} \begin{cases} \mathcal{N}(0.2825, 0.3) & \text{with prob. } 0.8, \\ \mathcal{N}(-1.3000, 1.3) & \text{with prob. } 0.2, \end{cases}$$

which scales the distribution of the observation to have zero mean, unit variance, negative skewness (≈ -1.3) and high kurtosis (≈ 8). Figure 2.1 plots the simulated data set.

Figure 2.1 about here

We run the algorithm for 15000 iterations after a burn-in phase of 5000 iterations, using flat priors on the $\boldsymbol{\theta}$-parameters and parameterizing G_0 as $\mathcal{N}(0, 3) \times$ Gamma$(1, 1)$ and $p(\alpha) =$ Gamma$(1, 1)$.

Table 2.1 about here

Table 2.1 gives the posterior means and 90% Bayesian intervals. The posterior mean of the persistence parameter ρ is quite close to the true value, while the volatility σ_η of the log-volatility is slightly underestimated.

Figure 2.2 about here

Figure 2.2 presents the graphical posterior summary. The trace plots in panels (a) and (b) and the marginal posteriors in (c) and (d) indicate that the sampling algorithm has converged to the posterior distribution. The Bayes factor (panel (e)) has the highest support at $u = 0.8$, which is in line with the underlying mixture model. The log-predictive density (blue line) in panel (i) exhibits the desired properties, i.e. asymmetry and fat tails. Furthermore, the log-predictive density resembles the true predictive density (dashed black line) with a slightly more pronounced right tail, which we attribute to the smaller information set. Note that the true number of mixture components is two, while the average number of components n is around ten. However, most of these components are negligibly small, as can be seen in panel (f).

2.4.2 Real-world data application

Given the encouraging results from the simulation exercise, we turn to a real data application. We use daily S&P 500 percentage returns from 03.08.2009 to 01.05.2015 (depicted in Figure 2.3). The objective is to obtain a posterior sample of the parametric part of the model and to construct a one-step-ahead density forecast.

Figure 2.3 about here

Note that the returns exhibit the typical patterns, such as heteroscedasticity and volatility clustering. Additionally, the descriptive statistics displayed in Table 2.2 provide further evidence of non-Gaussian behavior, in particular the negative skewness and high kurtosis. Therefore, applying the highly flexible semiparametric SV model is a natural choice.

Table 2.2 about here

We run the algorithm for 15000 iterations with a burn-in phase of 5000 and adopt the same priors as in Section 2.4.1.

The posterior means and the 90% CI are reported in Table 2.3 and the complete posterior summary is shown in Figure 2.4. Panel (a) shows the trace plots of the $\boldsymbol{\theta}$-parameters, ρ and σ_η, indicating the convergence of the chain. Subplots (c) and (d) give the corresponding marginal posteriors, where the horizontal blue line indicates the prior. The trajectory of the concentration parameter α can be seen in panel (b), and the Bayes factor is depicted in panel (e). Note that the Bayes factors for $u_0 > 0.8$ are zero, which supports the hypothesis of a mixture. Panel (f) plots the mixture weights, and (g) and (h), the trajectories of the mixture parameters μ_k and σ_k. Those plots do not have a direct interpretation related to the model and enable us only to observe the characteristics of the DPM, such as the mixing pattern. Lastly, panel (i) shows the posterior log-predictive density, which captures the high kurtosis and the slight asymmetry observed from the raw data. In a sensitivity analysis (not reported here) we ran the algorithm in eight parallel chains with random starting values drawn from the priors, and confirmed that each chain produced comparable results.

Table 2.3 about here

Figure 2.4 about here

2.5 Conclusion

We presented a new, flexible and general sampling algorithm for non-linear, non-parametric state-space models. In particular, our framework integrates complex

methods, such as the DPM, into a simple and intuitive estimation algorithm. As we do not rely on specific distributional assumptions or conjugacy of the priors, our approach is the first to allow for a comparison of the influence of prior distributions on nonparametric SV models. Furthermore, possible extensions include mixtures of more complex distributions, such as Skew-Slash distributions and/or leverage effects. We leave both extensions for future research.

Tables and Figures

	True	Post. Mean	CI (90%)
ρ	0.95	0.9546	(*0.9000, 0.9821*)
σ_η^2	0.04	0.0333	(*0.0128, 0.0837*)
α	-	1.2751	(*0.4495, 2.4261*)
n	-	9.9943	(*5, 17*)

Table 2.1: Simulated data: Posterior medians and 90% CI in the parentheses.

Mean	Median	St. Dev.	Skewness	Kurtosis
0.0524	0.0736	0.9987	-0.4630	7.2468

Table 2.2: Descriptive statistics of the S&P 500 percentage returns

	Post. Mean	CI (90%)
ρ	0.9505	(*0.9168, 0.9770*)
σ_η^2	0.0849	(*0.0404, 0.1564*)
α	1.1781	(*0.3575, 2.3076*)
n	9.1611	(*4, 16*)

Table 2.3: S&P Data: Posterior means and 90% CI

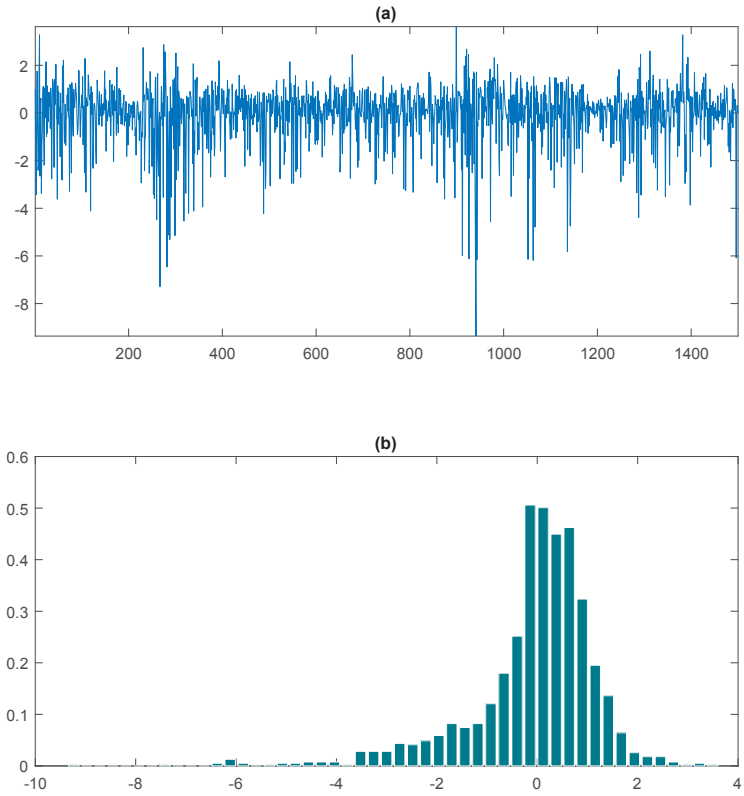


Figure 2.1: Simulated return data plot (a) and histogram (b)

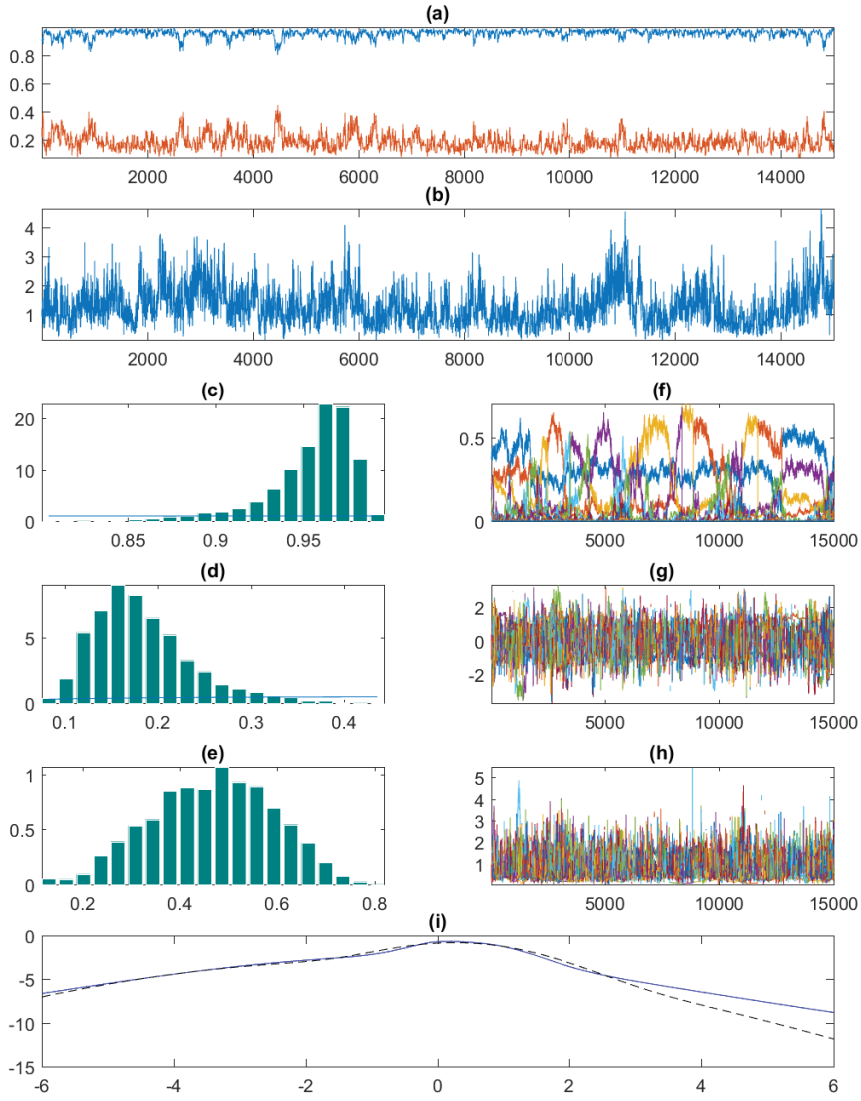


Figure 2.2: Simulated semiparametric stochastic volatility model (Section 2.4.1): (a) trace plots of ρ (in blue) and σ_η (in red), (b) trace plots of α, (c), (d) and (e) priors (in blue) and marginal posteriors of ρ, σ_η and Bayes factors, (f) mixture weights, (g) and (h) trajectories of the mixture parameters μ_k and σ_k, (i) posterior log-predictive density (blue line) and true log-predictive density (dashed black line)

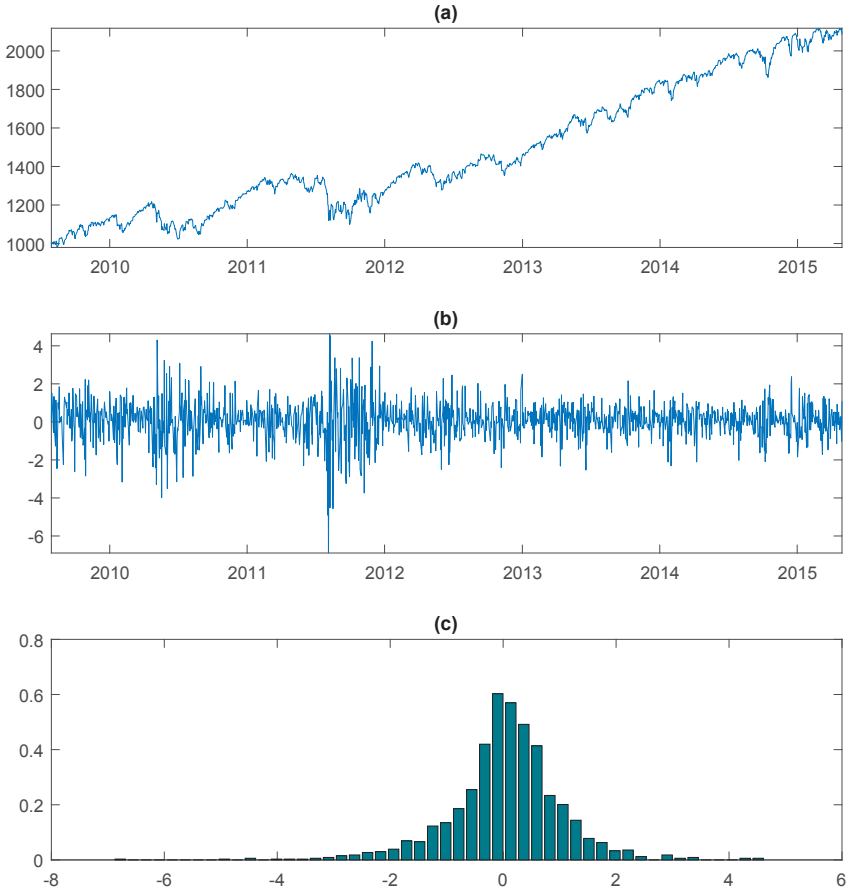


Figure 2.3: S&P 500 price (a) and return data (b) and (c)

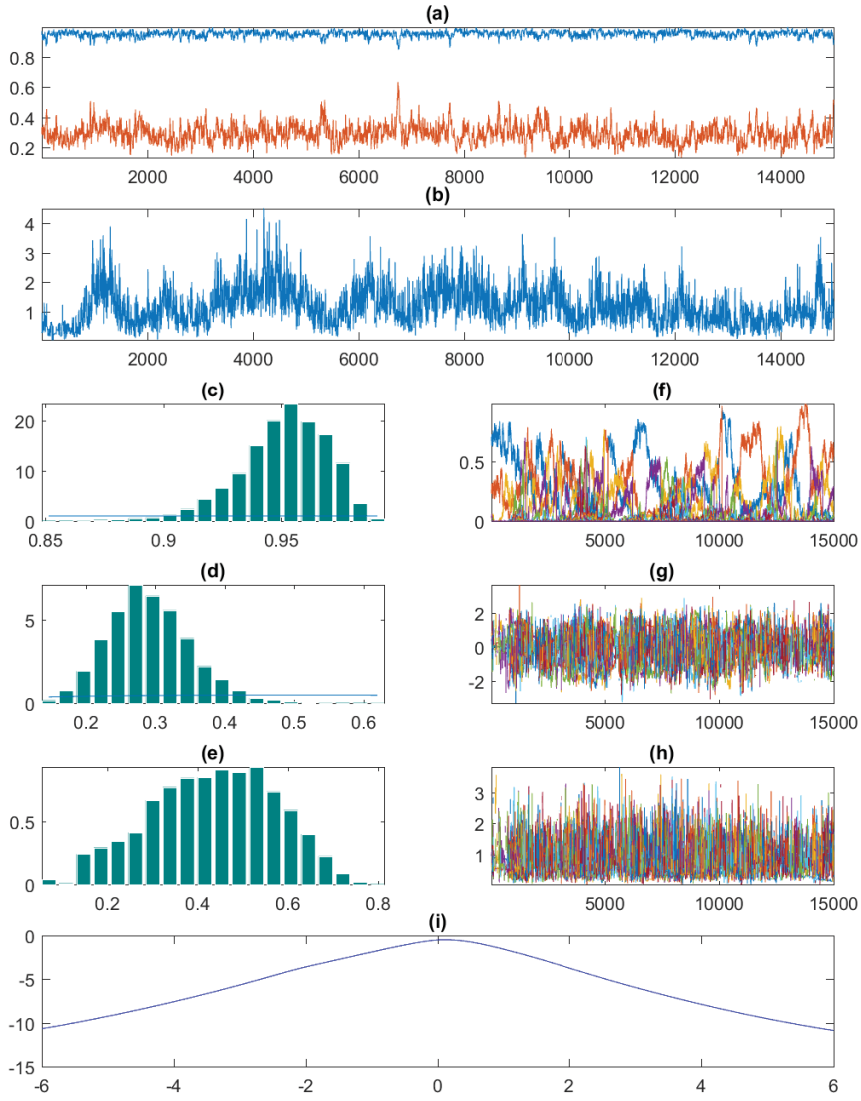


Figure 2.4: Real S&P 500 data application of the semiparametric stochastic volatility model of Section 2.4: (a) trace plots of ρ (in blue) and σ_η (in red), (b) trace plots of α, (c), (d) and (e) priors (in blue) and marginal posteriors of ρ, σ_η and Bayes factors, (f) mixture weights, (g) and (h) trajectories of the mixture parameters μ_k and σ_k, (i) posterior log-predictive density.

Chapter 3

Bayesian semiparametric multivariate stochastic volatility with an application to international volatility co-movements

3.1 Introduction

Owing to increasingly integrated financial markets, both domestically and internationally, volatility modeling and the analysis of volatility co-movements and spillovers among multiple asset returns have become central topics for the last few decades (*inter alia* Ehrmann et al., 2011; Clements et al., 2015). The two by far the most popular volatility model classes discussed in the literature are the generalized autoregressive conditional heteroscedasticity (GARCH-type) models (Engle 1982; Bollerslev 1986) and the stochastic volatility (SV) models (Taylor, 1982; 1986), both in univariate and multivariate variants. Several in-depth overview articles on multivariate GARCH (Bauwens et al., 2006) and SV models

(Chib et al., 2009) document the enormous professional interest in the field. While both model classes have distinct advantages on their own, a major characteristic of SV specifications is that they model the unobserved volatility directly as a separate stochastic process. This converts many SV specifications into discrete-time versions of continuous-time models that are well-established in finance theory, which constitutes the general attraction of SV models (Harvey et al., 1994; Kim et al., 1998, Asai et al., 2006).

Irrespective of model selection issues, various stylized empirical properties of asset returns have been discovered in real-world data, the most prominent being the fat-tail (kurtotic) nature of the return distribution. Cont (2001) reports that "... the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular, this excludes stable laws with infinite variance and the normal distribution.". Interestingly, the fat-tail property even persists after correcting the financial returns for volatility clustering (e.g. via GARCH-type models), although to a less pronounced degree. Numerous attempts have been made to account for the fat-tail property by replacing the Gaussianity assumption with alternative parametric distributions for the return innovation in distinct volatility models. Recently, however, several authors have proposed the nonparametric modeling of return innovation as a Dirichlet process mixture (DPM) and emphasize the flexibility increase associated with this approach, compared to using parametric distributions. In particular, to date, the nonparametric DPM approach has been applied successfully (i) to univariate SV modeling by Jensen and Maheu (2010, 2014) and Delatola and Griffin (2011, 2013), (ii) to univariate GARCH modeling

by Ausin et al. (2014), and (iii) to multivariate GARCH modeling by Jensen and Maheu (2013) and Virbickaitė et al. (2016).

In this chapter, we complete the above-described list by integrating the nonparametric DPM approach into a specific class of multivariate SV models with time-varying covariance components, based on the Cholesky decomposition of volatility matrices (see e.g. Nakajima, 2017). We establish a Bayesian estimation procedure for this semiparametric overall framework and study its predictive abilities by means of predictive density evaluation. In the empirical part, we apply our econometric setup to a five-country data set, in order to analyze volatility co-movements among the most important stock markets worldwide.

The chapter is organized as follows. Section 3.2 reviews (i) the multivariate SV model based on Cholesky decomposition, and (ii) the Dirichlet process mixture. In Section 3.3, we present the Markov Chain Monte Carlo (MCMC) algorithm for Bayesian inference. Section 3.4 presents essential probabilistic features of our econometric framework. Section 3.5 contains the empirical application to daily returns from the five largest international stock markets. Section 3.6 concludes.

3.2 Model development

3.2.1 Cholesky Multivariate Stochastic Volatility (MSV)

In order to introduce Cholesky MSV modeling, we follow the approach of Primiceri (2005) and Nakajima (2017) and consider the $m \times 1$ vector $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})'$ of time series observations at date t, which we assume to follow an m-dimensional multivariate normal distribution with zero-expectation vector, $\mathbb{E}(\mathbf{y}_t) = \mathbf{0}$, and time-varying

covariance matrix $\text{Cov}(\mathbf{y}_t) = \mathbf{H}_t$, i.e. $\mathbf{y}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{H}_t)$. The Cholesky decomposition of $\mathbf{H}_t$ is given by the factorization

$$\mathbf{A}_t \mathbf{H}_t \mathbf{A}_t' = \mathbf{\Sigma}_t \mathbf{\Sigma}_t, \tag{3.1}$$

where $\mathbf{A}_t$ is the lower triangular matrix of covariance components with 1s along the principal diagonal and $\boldsymbol{\Sigma_t}$ is the diagonal matrix of the time-varying standard deviations:

$$\mathbf{A}_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{21,t} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \alpha_{m1,t} & \cdots & \alpha_{mm-1,t} & 1 \end{pmatrix}, \qquad \mathbf{\Sigma}_t = \mathbf{\Sigma}_t' = \begin{pmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{m,t} \end{pmatrix}. \tag{3.2}$$

Via the Eqs. (3.1) and (3.2), the standard Cholesky MSV model is then defined as

$$\mathbf{y}_t = \mathbf{A}_t^{-1} \mathbf{\Sigma}_t \boldsymbol{\epsilon}_t, \tag{3.3}$$

$$\mathbf{H}_t = \mathbf{A}_t^{-1} \mathbf{\Sigma}_t \mathbf{\Sigma}_t (\mathbf{A}_t^{-1})', \tag{3.4}$$

where the innovation vector $\boldsymbol{\epsilon}_t$ is assumed to follow the m-dimensional multivariate standard normal distribution $\mathbf{N}(\mathbf{0}, \mathbf{I})$. Based on Eqs. (3.2) and (3.3), several alternative Cholesky MSV models have been proposed in the literature, by letting the innovation vector $\boldsymbol{\epsilon}_t$ follow distributions other than the multivariate standard normal, for example, the multivariate t (originally Harvey et al., 1994, in a non-Cholesky MSV framework), and the multivariate generalized hyperbolic skew t distribution (Nakajima, 2017). These specifications retain the essential Cholesky structure, but make more realistic distributional assumptions, with the aim of more effectively capturing some stylized facts of financial return data (like leverage effects and skew-

ness). In the next section, we define a new class of Cholesky MSV models by letting $\boldsymbol{\epsilon}_t$ follow a Dirichlet process mixture, in order to account for excess kurtosis in the data.

When it comes to Bayesian estimation of Cholesky MSV models with the time-varying parameters from Eq. (3.2), we adopt the common methodology of reducing the multivariate dynamics to univariate volatility processes that form a state-space representation (Lopes et al., 2012; Nakajima, 2017). Specifically, we collect the parameters from the matrix $\mathbf{A}_t$ in the $[m(m-1)/2] \times 1$ vector $\boldsymbol{\alpha}_t$ and define the stochastic volatility from $\boldsymbol{\Sigma}_t$ in the $m \times 1$ vector $\mathbf{h}_t$ as follows:

$$\boldsymbol{\alpha}_t = (\alpha_{21,t}, \alpha_{31,t}, \alpha_{32,t}, \ldots, \alpha_{m1,t}, \ldots, \alpha_{mm-1,t})', \tag{3.5}$$

$$\mathbf{h}_t = (\log(\sigma_{1,t}^2), \ldots, \log(\sigma_{m,t}^2))'. \tag{3.6}$$

(The parameters in $\boldsymbol{\alpha}_t$ are collected row by row from matrix $\mathbf{A}_t$.) We then specify the dynamics of the Cholesky parameters as the (stationary) AR(1) processes

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \boldsymbol{\Phi}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_{t-1} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}) + \mathbf{e}_t, \tag{3.7}$$

$$\mathbf{h}_t = \boldsymbol{\Phi}_{\mathbf{h}}\mathbf{h}_{t-1} + \boldsymbol{\eta}_t, \tag{3.8}$$

$$\begin{pmatrix} \mathbf{e}_t \\ \boldsymbol{\eta}_t \end{pmatrix} \sim \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\boldsymbol{\eta}} \end{bmatrix}\right), \tag{3.9}$$

where we assume (i) that the matrices $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}, \boldsymbol{\Phi}_{\mathbf{h}}, \boldsymbol{\Sigma}_{\mathbf{e}}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ are all diagonal, and (ii) that the $p = m(m-1)/2$ diagonal entries $\phi_{\boldsymbol{\alpha}1}, \ldots, \phi_{\boldsymbol{\alpha}p}$ of $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}$ and the m diagonal entries $\phi_{\mathbf{h}1}, \ldots, \phi_{\mathbf{h}m}$ of $\boldsymbol{\Phi}_{\mathbf{h}}$ are all less than 1 in absolute value (stationarity conditions).[1]

[1]Note that we specify the AR(1) process for $\mathbf{h}_t$ in Eq. (3.8) without an intercept term. This is due to an identification problem that would arise in the case of a non-zero intercept; see Jensen and Maheu (2010).

3.2.2 Bayesian semiparametric Cholesky MSV

Finally, it remains to specify the distribution of the innovation vector $\boldsymbol{\epsilon}_t$ from Eq. (3.3), which we model as a nonparametric Dirichlet process mixture (DPM). The DPM represents an infinite mixture model and constitutes an extremely flexible extension of finite mixture models (Jensen and Maheu, 2010, 2013; Kalli et al., 2013; Maheu and Yang, 2016; Virbickaitė et al., 2016). In introducing the DPM, we need to consider the Dirichlet process $\mathrm{DP}(c, G_0)$, defined in terms of the base distribution G_0 and the concentration parameter c (Ferguson, 1973). In a Bayesian context, the base distribution G_0 represents the prior distribution of the component parameters in the infinite mixture, while the parameter c, roughly speaking, controls for the number of clusters in the mixture. A small value of c can be thought of as *a priori* indicating a small number of components with relatively large weights in the infinite mixture, whereas large values of c *a priori* assume many mixture components, all with relatively small weights.

Overall, our semiparametric Cholesky MSV specification, in which we model the $m \times m$ matrix $\mathbf{H}_t$ from Eq. (3.4) parametrically, while we let the distribution of the innovation vector $\boldsymbol{\epsilon}_t$ follow the nonparametric DPM as given in Eq. (3.17) below, has the following hierarchical representation:

$$\mathbf{y}_t | \, \boldsymbol{\Lambda}_t, \mathbf{A}_t, \boldsymbol{\Sigma}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Sigma}_t(\mathbf{A}_t^{-1})'), \tag{3.10}$$

$$\mathbf{H}_t = \mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t(\mathbf{A}_t')^{-1}, \tag{3.11}$$

$$\boldsymbol{\Lambda}_t = \mathrm{diag}(\lambda_{1,t}, \ldots, \lambda_{m,t}), \tag{3.12}$$

$$\lambda_{i,t} \overset{\text{i.i.d.}}{\sim} G_i, \qquad (i = 1, \ldots, m) \tag{3.13}$$

$$G_i | \, G_0, c_i \sim \mathrm{DP}(c_i, G_0), \tag{3.14}$$

$$G_0 \overset{\text{d}}{=} \mathrm{Gamma}(\nu_0/2, s_0/2), \tag{3.15}$$

$$c_i \quad \sim \quad \text{Gamma}(a_0, b_0), \tag{3.16}$$

and where the elements of $\mathbf{A}_t$ and $\mathbf{\Sigma}_t$ collected in the vectors $\boldsymbol{\alpha}_t$ and $\mathbf{h}_t$ follow the AR(1) processes from Eqs. (3.7) and (3.8), respectively.[2] In Eqs. (3.10) and (3.12), the $m \times m$ matrix $\mathbf{\Lambda}_t$ is the precision matrix, which we assume to be diagonal, in order to ensure identification of the model.[3] We model the diagonal entries $\lambda_{1,t}, \ldots, \lambda_{m,t}$ as i.i.d. (with respect to t) and place a nonparametric Dirichlet process prior on the distribution of $\lambda_{i,t}$; see Eqs. (3.13) and (3.14). As in Ausin et al. (2014), we specify the base distribution G_0 for the diagonal elements of $\mathbf{\Lambda}_t$ as the gamma distribution in Eq. (3.15).

Following the line of argument in Jensen and Maheu (2013), we emphasize that our hierarchical model (3.10) to (3.16) can be expressed in the Sethuraman (1994) stick-breaking representation of the DPM mixture model. This allows us to write the density function of each component of the innovation vector $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{mt})'$ as an infinite scale-mixture of Gaussian distributions. That is, for $i = 1, \ldots, m$ we have

$$f(\epsilon_{it} | \, \omega_{i1}, \omega_{i2}, \ldots, l_{i1}, l_{i2}, \ldots) = \sum_{j=1}^{\infty} \omega_{ij} f_N \left(\epsilon_{it} | \, 0, l_{ij}^{-1} \right), \tag{3.17}$$

where $f_N \left(\cdot | \, 0, l_{ij}^{-1} \right)$ denotes the density function of the univariate normal distribution with zero mean and variance l_{ij}^{-1}. The mixture parameters l_{ij} prior is given in Eq. (3.15). It follows from the stick-breaking representation that the mixture weights are distributed as $\omega_{i1} = \upsilon_{i1}$, $\omega_{ij} = (1 - \upsilon_{i1}) \cdots (1 - \upsilon_{ij-1}) \cdot \upsilon_{ij}$ for $j > 1$, where $\upsilon_{i1}, \upsilon_{i2}, \ldots$ are i.i.d. Beta$(1, c_i)$ (beta distribution with parameters 1 and c_i). As described above, the choice of the concentration parameter c_i is crucial. In line

[2] In the hierarchical representation, $\overset{\text{d}}{=}$ means 'has the distribution'. The operator $\text{diag}(\lambda_1, \ldots, \lambda_m)$ creates the diagonal $m \times m$ matrix, say $\mathbf{M}$, with $\mathbf{M}_{ii} = \lambda_i$ and $\mathbf{M}_{ij} = 0$ for $i \neq j$ $(i, j = 1, \ldots, m)$.

[3] *Prima facie*, the diagonal structure of $\mathbf{\Lambda}_t$ might appear restrictive. However, as will become evident below, it does not impose any severe restriction on model flexibility.

with Escobar and West (1995), we assume a gamma hyper-prior distribution $c_i \sim \text{Gamma}(a_0, b_0)$; see Eq. (3.16). We remark that each of the m mixtures in Eq. (3.17) is related to its own specific concentration parameter c_i.

For notational convenience, we collect the parameters from the parametric part of our Cholesky MSV model in the vector $\mathbf{\Phi}$ (i.e. $\mathbf{\Phi}$ contains all parameters from $\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \mathbf{\Phi}_{\boldsymbol{\alpha}}, \mathbf{\Phi}_{\mathbf{h}}, \mathbf{\Sigma}_{\boldsymbol{\eta}}, \mathbf{\Sigma}_{\mathbf{e}}$), and all parameters from the nonparametric specification in the infinite dimensional entity $\mathbf{\Omega} = \{\omega_{ij}, l_{ij}\}_{i=1,\dots,m;j=1,2,\dots,\infty}$. In cases where we need to address all model parameters, we merge the partial parameter entities $\mathbf{\Phi}$ and $\mathbf{\Omega}$ into the full-parameter vector $\mathbf{\Theta}$.

3.3 Bayesian inference

In this section, we present the samplers for the single parameter-components of the Cholesky-Dirichlet-Process-Mixture-Multivariate-Stochastic-Volatility (Cholesky-DPM-MSV) established in Section 3.2. In Section 3.3.1, we apply Forward-Filtering-Backward-Sampling (FFBS) to the elements of the matrix $\mathbf{A}_t$ (Carter and Kohn, 1994). In Section 3.3.2, we use the volatility sampler suggested by Jacquier et al. (2002) for the volatility processes in the matrix $\mathbf{\Sigma}_t$. For the nonparametric DPM part of the Cholesky-MSV model, we apply the efficient slice-sampler according to Walker (2007) and Kalli et al. (2011) in Section 3.3.3.

3.3.1 Sampling the $\mathbf{A}_t$-elements

In order to apply FFBS sampling to the elements of the $\mathbf{A}_t$-matrix, we need to embed the $\mathbf{A}_t$-parameters in an appropriate state-space model. To this end, we first rewrite Eq. (3.3) of our Cholesky-MSV-DPM model as

$$\mathbf{A}_t \mathbf{y}_t = \mathbf{\Sigma}_t \boldsymbol{\epsilon}_t, \tag{3.18}$$

where $\mathbf{y}_t$ is observable, and $\mathbf{A}_t$ has the lower triangular form given in Eq. (3.2). As in Primiceri (2005), we next define the $m \times m(m-1)/2$ matrix

$$\mathbf{Z}_t = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ -y_{1t} & 0 & \cdots & 0 \\ 0 & -\mathbf{y}_{[1:2]t} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\mathbf{y}_{[1:m-1]t} \end{pmatrix}, \tag{3.19}$$

in which $\mathbf{y}_{[1:i]t}$ denotes the row vector $(y_{1t}, y_{2t}, \ldots, y_{it})$, so that Eq. (3.18) can be written as

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{\Sigma}_t \boldsymbol{\epsilon}_t, \tag{3.20}$$

where $\boldsymbol{\alpha}_t$, defined in Eq. (3.5), follows the AR(1) process specified in Eq. (3.7). Finally, we replace $\boldsymbol{\epsilon}_t$ in Eq. (3.20) with $\mathbf{\Lambda}_t^{-1/2}\mathbf{u}_t$, where $\mathbf{u}_t$ is assumed to follow the m dimensional multivariate standard normal distribution $\mathbf{N}(\mathbf{0}, \mathbf{I})$, and obtain the desired state-space model via Eqs. (3.20) and (3.7) with observation and transition equations given by

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{\Sigma}_t \mathbf{\Lambda}_t^{-1/2} \mathbf{u}_t \equiv \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\xi}_t, \tag{3.21}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \mathbf{\Phi}_{\boldsymbol{\alpha}} (\boldsymbol{\alpha}_{t-1} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}) + \mathbf{e}_t, \tag{3.22}$$

with $\boldsymbol{\xi}_t \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t})$, $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_t} = \boldsymbol{\Sigma}_t \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\Sigma}_t$ and

$$\begin{pmatrix} \boldsymbol{\xi}_t \\ \mathbf{e}_t \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{e}} \end{bmatrix} \right). \tag{3.23}$$

In order to apply FFBS based on Kalman filter recursion, we denote the entire history of the vector $\mathbf{y}_t$ and the matrices $\mathbf{Z}_t, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t}$ to date s by $\mathbf{y}^{(s)} \equiv \{\mathbf{y}_0, \ldots, \mathbf{y}_{s-1}, \mathbf{y}_s\}$, $\mathbf{Z}^{(s)} \equiv \{\mathbf{Z}_0, \ldots, \mathbf{Z}_{s-1}, \mathbf{Z}_s\}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(s)} \equiv \{\boldsymbol{\Sigma}_{\boldsymbol{\xi}_0}, \ldots, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_{s-1}}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_s}\}$, respectively, and let

$$\boldsymbol{\alpha}_{t|s} = \mathbb{E}(\boldsymbol{\alpha}_t | \mathbf{y}^{(s)}, \mathbf{Z}^{(s)}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(s)}, \boldsymbol{\Sigma}_{\mathbf{e}}) \tag{3.24}$$

$$\mathbf{V}_{t|s} = \mathrm{Cov}(\boldsymbol{\alpha}_t | \mathbf{y}^{(s)}, \mathbf{Z}^{(s)}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(s)}, \boldsymbol{\Sigma}_{\mathbf{e}}). \tag{3.25}$$

Furthermore, we define the $p \times 1$ vector

$$\mathbf{c}_{\boldsymbol{\alpha}} \equiv (\mu_{\boldsymbol{\alpha}1}(1 - \phi_{\boldsymbol{\alpha}1}), \ldots, \mu_{\boldsymbol{\alpha}p}(1 - \phi_{\boldsymbol{\alpha}p}))', \tag{3.26}$$

where $\mu_{\boldsymbol{\alpha}1}, \ldots, \mu_{\boldsymbol{\alpha}p}$ are the elements of the vector $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ and $\phi_{\boldsymbol{\alpha}1}, \ldots, \phi_{\boldsymbol{\alpha}p}$ the diagonal entries of the matrix $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}$ as defined above. Then, given the starting values $\boldsymbol{\alpha}_{0|0}$ and $\mathbf{V}_{0|0}$, the standard Kalman filter can be summarized as follows:

$$\boldsymbol{\alpha}_{t|t-1} = \mathbf{c}_{\boldsymbol{\alpha}} + \boldsymbol{\Phi}_{\boldsymbol{\alpha}} \boldsymbol{\alpha}_{t-1|t-1}, \tag{3.27}$$

$$\mathbf{V}_{t|t-1} = \boldsymbol{\Phi}_{\boldsymbol{\alpha}} \mathbf{V}_{t-1|t-1} \boldsymbol{\Phi}_{\boldsymbol{\alpha}}' + \boldsymbol{\Sigma}_{\mathbf{e}}, \tag{3.28}$$

$$\mathbf{K}_t = \mathbf{V}_{t|t-1} \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{V}_{t|t-1} \mathbf{Z}_t' + \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t})^{-1}, \tag{3.29}$$

$$\boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \boldsymbol{\alpha}_{t|t-1}), \tag{3.30}$$

$$\mathbf{V}_{t|t} = \mathbf{V}_{t|t-1} - \mathbf{K}_t \mathbf{Z}_t \mathbf{V}_{t|t-1}. \tag{3.31}$$

The final entities $\boldsymbol{\alpha}_{T|T}$ and $\mathbf{V}_{T|T}$ contain the mean and variances of the normal distribution, from which we draw $\boldsymbol{\alpha}_T$. We use this value in the first step of the backward recursion that yields $\boldsymbol{\alpha}_{T-1|T}$ and $\mathbf{V}_{T-1|T}$, which we then use to draw $\boldsymbol{\alpha}_{T-1}$. The backward recursion iterates from $T-1$ to 0, and at date t, the update step is given by

$$\boldsymbol{\alpha}_{t|t+1} = \boldsymbol{\alpha}_{t|t} + \mathbf{V}_{t|t}\boldsymbol{\Phi}_{\boldsymbol{\alpha}}'\mathbf{V}_{t+1|t}^{-1}(\boldsymbol{\alpha}_{t+1} - \mathbf{c}_{\boldsymbol{\alpha}} - \boldsymbol{\Phi}_{\boldsymbol{\alpha}}\boldsymbol{\alpha}_{t|t}), \tag{3.32}$$

$$\mathbf{V}_{t|t+1} = \mathbf{V}_{t|t} - \mathbf{V}_{t|t}\boldsymbol{\Phi}_{\boldsymbol{\alpha}}'\mathbf{V}_{t+1|t}^{-1}\boldsymbol{\Phi}_{\boldsymbol{\alpha}}\mathbf{V}_{t|t}. \tag{3.33}$$

As the prior distribution of the initial state $\boldsymbol{\alpha}_{0|0}$ we use a multivariate normal distribution (see Section 3.5) and, as mentioned above, assume the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{e}}$ to be diagonal with entries $\sigma_{\mathbf{e}1}^2, \ldots, \sigma_{\mathbf{e}p}^2$. Note that for each $i = 1, \ldots, p$ the unconditional expectation of the α_{it}-process is $E(\alpha_{it}) = \mu_{\boldsymbol{\alpha}i} = \frac{c_{\boldsymbol{\alpha}i}}{1-\phi_{\boldsymbol{\alpha}i}}$, so that the $3p = 3m(m-1)/2$ parameters to be sampled are

$$c_{\boldsymbol{\alpha}1}, \ldots, c_{\boldsymbol{\alpha}p}, \phi_{\boldsymbol{\alpha}1}, \ldots, \phi_{\boldsymbol{\alpha}p}, \sigma_{\mathbf{e}1}^2, \ldots, \sigma_{\mathbf{e}p}^2.$$

The sampling strategy for these parameters is readily obtained from standard Bayesian estimation of the linear regression model. The prior distributions for the $c_{\boldsymbol{\alpha}i}$- (or $\mu_{\boldsymbol{\alpha}i}$-) and $\phi_{\boldsymbol{\alpha}i}$-parameters are normal distributions (where the prior for the $\phi_{\boldsymbol{\alpha}i}$-parameters have to be restricted to ensure the p stationarity conditions $|\phi_{\boldsymbol{\alpha}i}| < 1$), while the prior distribution for $\sigma_{\mathbf{e}i}^2$ is chosen as the inverse Gamma distribution. We sample the $c_{\boldsymbol{\alpha}i}$- and $\phi_{\boldsymbol{\alpha}i}$-parameters by the Metropolis-Hastings (MH) algorithm, while the $\sigma_{\mathbf{e}i}^2$-parameters are sampled directly (given the conjugate prior). More details are provided in Appendix B.

3.3.2 Sampling the Σ_t-elements

Defining $\widetilde{\mathbf{y}}_t = \mathbf{A}_t \mathbf{y}_t$, we note that $\widetilde{\mathbf{y}}_t$ has a diagonal covariance matrix, what enables us to independently estimate the m univariate stochastic volatility models. The ith univariate stochastic volatility model is given by

$$\tilde{y}_{it} = \sigma_{i,t} \lambda_{i,t}^{-1/2} u_{it}, \qquad (i = 1, \ldots, m) \tag{3.34}$$

with $u_{it} \sim N(0,1)$. At this stage, we consider the matrix $\mathbf{A}_t$ as given and since $\mathbf{y}_t$ is observable, the values of $\tilde{y}_{it}$ can be computed. We note that the associated dynamic model is nonlinear:

$$\tilde{y}_{it} = \exp\{h_{it}/2\} \lambda_{i,t}^{-1/2} u_{it}, \qquad (i = 1, \ldots, m) \tag{3.35}$$

$$h_{it} = \phi_{\mathbf{h}i} h_{it-1} + \eta_{it}, \tag{3.36}$$

with $\eta_{it} \sim N(0, \sigma^2_{\boldsymbol{\eta}i})$ and $\sigma^2_{\boldsymbol{\eta}i}$ being the ith diagonal entry of the matrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$.

The m univariate SV models from Eqs. (3.35) and (3.36) can be estimated separately by consecutively sampling from the following conditionals, in the representation of which we use the m row vectors $\boldsymbol{\vartheta}_i = (\sigma^2_{\boldsymbol{\eta}i}, \phi_{\mathbf{h}i})$:

1. $\pi(\boldsymbol{\vartheta}_i | h_{it})$, yielding the AR parameters.

2. $\pi(h_{it} | \tilde{y}_{it}, \boldsymbol{\vartheta}_i, l_{ij}, \omega_{ij})$, yielding the parametric volatility component.

3. $\pi(l_{ij}, \omega_{ij} | \tilde{y}_{it}, h_{it})$, yielding the nonparametric volatility component.

Sampling from the first conditional is straightforward and analogous to sampling the $\boldsymbol{\alpha}_t$-parameters in the previous section. Assuming a normal prior for the $\phi_{\mathbf{h}i}$- and an inverse Gamma prior for the $\sigma^2_{\boldsymbol{\eta}i}$-parameters, we sample the $\sigma^2_{\boldsymbol{\eta}i}$-parameters

directly (given the conjugate prior), while we apply an MH step in order to sample the $\phi_{\mathbf{h}i}$-parameters from their posterior distribution, where we restrict these latter parameters to meet the stationarity conditions $|\phi_{\mathbf{h}i}| < 1$ (see Appendix B). The third conditional from above involves sampling the infinite mixture parameters, for which we introduce a complete sampling algorithm in Section 3.3.3.

In order to sample from $\pi(h_{it}|\tilde{y}_{it}, \boldsymbol{\vartheta}_i, l_{ij}, \omega_{ij})$, we follow Jensen and Maheu (2010) and apply our log volatility sampler to the transformation $y^*_{it} \equiv \tilde{y}_{it}\sqrt{\lambda_{i,t}}$ yielding the m simplified univariate models

$$y^*_{it} = \exp\{h_{it}/2\}\, u_{it}, \qquad (i = 1, \ldots, m) \tag{3.37}$$

$$h_{it} = \phi_{\mathbf{h}i} h_{it-1} + \eta_{it}, \tag{3.38}$$

so that our task reduces to sampling from $\pi(h_{it}|y^*_{it}, \boldsymbol{\vartheta}_i)$. We accomplish this by using a procedure from Jacquier et al. (2002), who propose a Bayesian approach, in which they construct a Markov chain for drawing directly from the joint posterior distribution of the latent volatility components. Specifically, let $\mathbf{h}^{(i)}_{-t} \equiv (h_{i0}, \ldots, h_{it-1}, h_{it+1}, \ldots, h_{iT})'$ and $\mathbf{y}^*_i \equiv (y^*_{i1}, \ldots, y^*_{iT})'$, which are used to decompose the distribution $\pi(h_{it}|y^*_{it}, \boldsymbol{\vartheta}_i)$ into a set of conditionals of the form $\pi(h_{it}|\mathbf{h}^{(i)}_{-t}, \mathbf{y}^*_i, \boldsymbol{\vartheta}_i)$. The authors suggest a (hybrid) cyclic random walk Metropolis chain which uses a series of independent Metropolis acceptance/rejection chains, which do not directly sample from the univariate conditionals, but still ensure that the posterior is a stationary distribution.

Thus, in order to sample from the target distribution $\pi(h_{it}|y^*_{it}, \boldsymbol{\vartheta}_i)$, we follow the lines of argument in Jacquier et al. (2002) and sample from the auxiliary density $\pi(h_{it}|h_{it-1}, h_{it+1}, y^*_{it}, \boldsymbol{\vartheta}_i)$, which can be factorized for $t = 2, \ldots, T-1$ as follows:

$$\begin{aligned}\pi(h_{it}|h_{it-1}, h_{it+1}, y^*_{it}, \boldsymbol{\vartheta}_i) &\propto \pi(y^*_{it}|h_{it})\pi(h_{it}|h_{it-1})\pi(h_{it+1}|h_{it}) \\ &\propto \frac{1}{\exp\{h_{it}/2\}} \exp\left\{-\frac{1}{2}\frac{(y^*_{it})^2}{\exp\{h_{it}/2\}}\right\} \\ &\quad \times \exp\left\{\frac{-(h_{it} - \phi_{\mathbf{h}i}h_{it-1})^2 - (h_{it+1} - \phi_{\mathbf{h}i}h_{it})^2}{2\sigma^2_{\boldsymbol{\eta}i}}\right\}. \end{aligned} \tag{3.39}$$

The density (3.39) does not have a standard form and we apply a Metropolis-Hastings algorithm for each of the latent volatility components $h_{i2}, \ldots, h_{iT-1}$. We sample the first and last latent volatility components from

$$\begin{aligned}\pi(h_{i1}|h_{i2}, y^*_{it}, \boldsymbol{\vartheta}_i) &\propto \frac{1}{\exp\{h_{i1}/2\}} \exp\left\{-\frac{1}{2}\frac{(y^*_{i1})^2}{\exp\{h_{i1}/2\}}\right\} \\ &\quad \times \exp\left\{\frac{-(h_{i2} - \phi_{\mathbf{h}i}h_{i1})^2}{2\sigma^2_{\boldsymbol{\eta}i}}\right\}, \end{aligned} \tag{3.40}$$

$$\begin{aligned}\pi(h_{iT}|h_{iT-1}, y^*_{it}, \boldsymbol{\vartheta}_i) &\propto \frac{1}{\exp\{h_{iT}/2\}} \exp\left\{-\frac{1}{2}\frac{(y^*_{iT})^2}{\exp\{h_{iT}/2\}}\right\} \\ &\quad \times \exp\left\{\frac{-(h_{iT} - \phi_{\mathbf{h}i}h_{iT-1})^2}{2\sigma^2_{\boldsymbol{\eta}i}}\right\}. \end{aligned} \tag{3.41}$$

As the proposal, used at each step of the random walk Metropolis Hastings algorithm, we use $N(0, \sigma^2_{\boldsymbol{\eta}i})$.

3.3.3 Slice sampling the $\boldsymbol{\epsilon}_t$-DPM-elements

The slice-sampler proposed by Walker (2007) and its more efficient version presented in Kalli et al. (2011) tackle the general issue of sampling the infinite number of DPM parameters. The idea behind the slice- sampler is to introduce appropriate latent

variables, with the objective of finding a finite set of DPM parameters, the sampling of which produces a valid Markov chain with a correct stationary distribution.

The first step of the slice-sampling procedure consists of introducing a latent variable ρ_{it} (with positive support), such that for $i = 1, \ldots, m$ the joint density of the innovation ϵ_{it} and the latent variable ρ_{it} is given by

$$\begin{aligned} f(\epsilon_{it}, \rho_{it}|\boldsymbol{\Theta}) &= \sum_{j=1}^{\infty} \mathbf{1}(\rho_{it} < \omega_{ij}) \cdot f_N(\epsilon_{it}|\, 0, l_{ij}^{-1}) \\ &= \sum_{j \in \mathcal{A}(\rho_{it})} f_N(\epsilon_{it}|\, 0, l_{ij}^{-1}), \end{aligned} \tag{3.42}$$

where (i) $\mathbf{1}(\cdot)$ is the indicator function, which is equal to 1 when its argument is true and 0 otherwise, and (ii) $\mathcal{A}(\rho_{it}) \equiv \{j : \omega_{ij} > \rho_{it}\}$, which becomes a finite set for any given $\rho_{it} > 0$. We note that the conditional distribution of ϵ_{it} given ρ_{it} is a finite normal mixture with equal weights. Based on this result, the slice-sampling procedure then introduces a second latent variable ζ_{it} indicating the mixture component from which ϵ_{it} is observed to yield the joint density

$$f(\epsilon_{it}, \zeta_{it} = j, \rho_{it}|\, \boldsymbol{\Theta}) = f_N(\epsilon_{it}|\, 0, l_{ij}^{-1})\, \mathbf{1}(j \in \mathcal{A}(\rho_{it})). \tag{3.43}$$

Specifically, after initializing the starting values $c_i^{(0)}, \zeta_{i1}^{(0)}, \ldots, \zeta_{iT}^{(0)}$, the slice-sampler proposed by Kalli et al. (2011) and Walker (2007) proceeds as follows in iteration r of the MCMC algoritm ($r = 1, \ldots, R$):

1. Sampling c_i:

   We use the sampling strategy proposed in Escobar and West (1995) and start

by sampling the auxiliary variable $\psi_i \sim \text{Beta}(c_i^{(r-1)} + 1, T)$ and then sample c_i from the Gamma mixture

$$\pi_{\psi_i} \cdot f_\Gamma(c_i | a_0 + \zeta_i^*, b_0 - \log(\psi_i)) + (1 - \pi_{\psi_i}) \cdot f_\Gamma(c_i | a_0 + \zeta_i^* - 1, b_0 - \log(\psi_i)),$$

where $f_\Gamma(\cdot|\alpha, \beta)$ denotes the density function of the Gamma(α, β) distribution, $\zeta_i^* = \max\left\{\zeta_{i1}^{(r-1)}, \ldots, \zeta_{iT}^{(r-1)}\right\}$ and $\pi_{\psi_i} = (a_0 + \zeta_i^* - 1)/(a_0 + \zeta_i^* - 1 + T(b_0 - \log(\psi_i)))$.

2. Sampling υ_{ij}:

   For $j = 1, 2, \ldots, \zeta_i^*$ we sample the υ_{ij} values from the conditional distribution

   $$\upsilon_{ij} | \, \zeta_{i1}^{(r-1)}, \ldots, \zeta_{iT}^{(r-1)} \sim \text{Beta}\left(n_{ij} + 1, T - n_{i\cdot} + c_i\right),$$

   where $n_{ij} = \sum_{t=1}^{T} \mathbf{1}(\zeta_{it}^{(r-1)} = j)$ is the number of observations belonging to the jth component of the ith variable, and $n_{i\cdot} = \sum_{k=1}^{j} n_{ik}$ is the cumulative sum of components in the groups. We compute the associated mixture weights according to the stick-breaking procedure, $\omega_{i1} = \upsilon_{i1}$, and $\omega_{ij} = (1-\upsilon_{ij}) \ldots (1-\upsilon_{ij-1})\upsilon_j$ for $j = 2, \ldots, \zeta_i^*$.

3. Sampling ρ_{it}:

   We sample the latent variables ρ_{it} from the uniform distribution $U(0, \omega_{i\zeta_{it}^{(r-1)}})$ and set $\rho_i^* = \min\{\rho_{i1}, \ldots, \rho_{iT}\}$, which we use to truncate the sequence of mixture weights in the next step.

4. Updating the weights ω_{ij}:

   We determine the smallest integer j_i^* such that $\sum_{j=1}^{j_i^*} \omega_{ij} > (1 - \rho_i^*)$. For those ω_{ij} with $j > \zeta_i^*$, we draw υ_{ij} from the prior Beta$(c_i, 1)$ distribution and compute the associated weights ω_{ij} according to the stick-breaking procedure

for $j = \zeta_i^* + 1, \ldots, j_i^*$. Thus, the latent variable ρ_{it} indicates how many weights need to be sampled.

5. Sampling the mixture parameters l_{ij}:

   The mixture parameters are sampled from the conditional posterior, which given the conjugate priors, has the following Gamma distribution:

$$l_{ij} \sim \text{Gamma}(\bar{\nu}_{ij}/2, \bar{s}_{ij}/2), \tag{3.44}$$

$$\bar{\nu}_{ij} = \nu_0 + n_{ij}, \tag{3.45}$$

$$\bar{s}_{ij} = s_0 + \sum_{t=1}^{T} \epsilon_{it}^2 \cdot \mathbf{1}(\zeta_{it}^{(r-1)} = j). \tag{3.46}$$

   We note that, according to Eq. (3.35), $\epsilon_{it} = \tilde{y}_{it} \exp\{-h_{it}/2\}$ is treated as observable at this stage of the algorithm. As in Step 4, if a new component has been formed, the mixture parameters are sampled from the prior.

6. Updating the indicator variables ζ_{it}:

   According to the weight truncation induced by the variable ρ_{it}, we update the indicator variables ζ_{it} by sampling from

$$\Pr(\zeta_{it} = j | \{\epsilon_{it}\}_{t=1}^{T}, \{l_{ij}\}_{j=1}^{j_i^*}, \{\omega_{ij}\}_{j=1}^{j_i^*}, \{\rho_{it}\}_{t=1}^{T}) \propto f_N(\epsilon_{it}|0, l_{ij}^{-1}) \cdot \mathbf{1}(j \in \mathcal{A}(\rho_{it})).$$

   The updated variables ζ_{it} indicate the component to which each observation belongs. Given ζ_{it}, we set $\lambda_{i,t} = l_{i\zeta_{it}}$.

3.4 Features of the Cholesky DPM-MSV model

3.4.1 Predictive density

A key issue in Bayesian nonparametric inference is the predictive density (Escobar and West, 1995). Denoting the sequence of all observations obtained through date T by $\mathbf{y}_{1:T} = \{\mathbf{y}_1, \ldots, \mathbf{y}_T\}$, we write the one-step ahead predictive density as

$$f(\mathbf{y}_{T+1} | \, \mathbf{y}_{1:T}) = \int f(\mathbf{y}_{T+1} | \, \boldsymbol{\Theta}, \mathbf{y}_{1:T}) \pi(\boldsymbol{\Theta} | \, \mathbf{y}_{1:T}) \mathrm{d}\boldsymbol{\Theta}, \tag{3.47}$$

where (i) the density $f(\mathbf{y}_{T+1} | \, \boldsymbol{\Theta}, \mathbf{y}_{1:T})$ constitutes an infinite scale mixture, given the representation of the innovation term in Eq. (3.17), and (ii) the posterior $\pi(\boldsymbol{\Theta} | \, \mathbf{y}_{1:T})$ is defined on the infinitely dimensional parameter space $\boldsymbol{\Theta}$. Since the integral in Eq. (3.47) is analytically untractable, we approximate the predictive density via the MCMC output,

$$f(\mathbf{y}_{T+1} | \, \mathbf{y}_{1:T}) \approx \frac{1}{R} \sum_{r=1}^{R} f(\mathbf{y}_{T+1} | \, \boldsymbol{\Theta}^{(r)}, \mathbf{y}_{1:T}), \tag{3.48}$$

where R is the length of the Markov chain and $\boldsymbol{\Theta}^{(r)}$ denotes the parameter set in iteration r. We cope with the infinitely dimensional parameter space by introducing the latent variables according to Eq. (3.42) in each iteration r (which we denote by $\rho_{it}^{(r)}$) and thus for $i = 1, \ldots, m$ obtain the following (finite number of) DPM parameters in iteration r:

$$\left\{ \omega_{i1}^{(r)}, \omega_{i2}^{(r)}, \ldots, \omega_{ij_i^{*(r)}}^{(r)} \right\} \qquad \text{and} \qquad \left\{ l_{i1}^{(r)}, l_{i2}^{(r)}, \ldots, l_{ij_i^{*(r)}}^{(r)} \right\}.$$

Next, we implement the 3-step algorithm proposed by Jensen and Maheu (2013), in order to sample a single precision (mixture) parameter $l_i^{(r)}$ in iteration r for $i = 1, \ldots, m$:

1. We sample the random variable a_i from the uniform distribution $U(0,1)$.

2. We compute the sum $\sum_{j=1}^{j_i^{*(r)}} \omega_{ij}^{(r)}$.

3. If $\sum_{j=1}^{j_i^{*(r)}} \omega_{ij}^{(r)} > a_i$, we find the index d_i such that

$$\sum_{j=1}^{d_i-1} \omega_{ij}^{(r)} < a_i < \sum_{j=1}^{d_i} \omega_{ij}^{(r)}$$

and set the precision parameter $l_i^{(r)} = l_{id_i}^{(r)}$; else we draw $l_i^{(r)}$ from the prior distribution G_0 given in Eq. (3.15).

After having run the three steps for each $i = 1, \ldots, m$, we compose the predictive error term covariance matrix at iteration r as $(\mathbf{\Lambda}^{(r)})^{-1} \equiv \text{diag}(1/l_i^{(r)})$.

We now repeat the complete algorithm (i.e the 3 steps for each $i = 1, \ldots, m$) a number of times (say $B^{\max}$ times) and record at each iteration r, the $B^{\max}$ covariance matrices $(\mathbf{\Lambda}_1^{(r)})^{-1}, \ldots, (\mathbf{\Lambda}_{B^{\max}}^{(r)})^{-1}$. Denoting the density function of the m dimensional multivariate normal distribution by $f_{\mathbf{N}}(\cdot|\cdot,\cdot)$ and given sampled parameters, we approximate the one-step-ahead predictive density according to Eq. (3.48) as

$$f(\mathbf{y}_{T+1}|\mathbf{y}_{1:T}) \approx \frac{1}{R} \sum_{r=1}^{R} f^{(r)}(\mathbf{y}_{T+1}|\mathbf{y}_{1:T}) \tag{3.49}$$

with

$$f^{(r)}(\mathbf{y}_{T+1}|\mathbf{y}_{1:T}) = \frac{1}{B^{\max}} \sum_{k=1}^{B^{\max}} f_{\mathbf{N}}\left(\mathbf{y}_{T+1}|\, \mathbf{0}, (\mathbf{A}_{T+1}^{(r)})^{-1}\mathbf{\Sigma}_{T+1}^{(r)}(\mathbf{\Lambda}_k^{(r)})^{-1}\mathbf{\Sigma}_{T+1}^{(r)}[(\mathbf{A}_{T+1}^{(r)})^{-1}]'\right), \tag{3.50}$$

where, for the computation of $\mathbf{A}_{T+1}^{(r)}$ and $\mathbf{\Sigma}_{T+1}^{(r)}$, we draw each $\alpha_{iT+1}^{(r)}$ from $N(\mu_{\boldsymbol{\alpha}i}^{(r)} + \phi_{\boldsymbol{\alpha}i}^{(r)}\alpha_{iT}^{(r)}, \sigma_{\mathbf{e}}^{2(r)})$ for $i = 1, \ldots, p$, and each $h_{iT+1}^{(r)}$ from $N(\phi_{\mathbf{h}i}^{(r)}h_{iT}^{(r)}, \sigma_{\boldsymbol{\eta}}^{2(r)})$ for $i = 1, \ldots, m$. In our empirical application below, we choose $B^{max} = 3$.

3.4.2 Conditional moments

According to the hierarchical representation of our Cholesky DPM-MSV model from the Eqs. (3.10) to (3.17), the conditional mean of $\mathbf{y}_t$ is assumed to equal the zero vector, while the conditional covariance matrix is given by

$$\mathbf{H}_t^* = \text{Cov}(\mathbf{y}_t|\boldsymbol{\Theta}, \mathbf{y}_{1:t-1}) = \mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t \text{Cov}(\boldsymbol{\epsilon}_t|\boldsymbol{\Omega})\boldsymbol{\Sigma}_t(\mathbf{A}_t^{-1})', \tag{3.51}$$

where

$$\text{Cov}(\boldsymbol{\epsilon}_t|\boldsymbol{\Omega}) = \text{diag}\left(\sum_{j=1}^{\infty} \omega_{ij} l_{ij}^{-1}\right).$$

Using our predictive density from the Eqs. (3.49) and (3.50), we may approximate conditional second-moment forecasts of the Cholesky DPM-MSV model by

$$\mathbb{E}(\mathbf{H}_{T+1}^*) \approx \frac{1}{R}\sum_{r=1}^{R} \mathbf{H}_{T+1}^{*(r)}, \tag{3.52}$$

where

$$\mathbf{H}_{T+1}^{*(r)} = (\mathbf{A}_{T+1}^{(r)})^{-1}\boldsymbol{\Sigma}_{T+1}^{(r)} \frac{1}{B^{\max}} \sum_{k=1}^{B^{\max}} (\boldsymbol{\Lambda}_k^{(r)})^{-1}\boldsymbol{\Sigma}_{T+1}^{(r)}[(\mathbf{A}_{T+1}^{(r)})^{-1}]'. \tag{3.53}$$

3.4.3 Ordering of variables

Owing to the lower triangular structure of the $\mathbf{A}_t$ matrix, the ordering of the variables in the vector $\mathbf{y}_t$ of the Cholesky DPM-MSV model is crucial (Primiceri, 2005). In the context of time-varying VAR models, Nakajima and Watanabe (2011) address the problem by analyzing the structure of the Japanese economy and monetary policy. When analyzing multiple financial time series data, it might sometimes appear problematic or arbitrary to use a specific ordering of variables *prima facie*. However, in our empirical application below, an obvious criterion for variable ordering is the chronological sequence, in which the various stock markets start their trading day.

3.5 Empirical application

3.5.1 Data

In this section, we apply the Cholesky DPM-MSV model to stock index data for the five most important international stock markets, with the objective of analyzing volatility co-movements. In particular, our data set includes daily stock index values between 17 February 2012 and 19 February 2016 (1046 observations for each time series) for (i) the US Dow Jones Industrial, (ii) the German DAX 30 Performance, (iii) the European EuroStoxx50 index, (iv) the Japanese Nikkei 225, and (v) the Chinese Shanghai Shenzen CSI 300. All data were collected from *Datastream* (daily closing prices).

Figure 3.1 about here

Figure (3.1) displays the five indices along with their daily returns (computed as the daily first differences in logs $\times 100$). The sampling period does not cover the global financial crisis, but includes two country-specific stock market turbulences, namely the European sovereign debt crisis in early 2012 and the Chinese stock market turmoil between June 2015 and February 2016. Both events are accompanied by phases of high return volatility, as is evident from the right panels in Figure (3.1).

Table 3.1 about here

Table (3.1) contains summary statistics and the sample correlation coefficients among the five return series. All return series exhibit negative skewness and excess kurtosis, indicating non-Gaussian behavior. Although all five sample means are close to zero, we use demeaned data in our estimation procedure. The sample

correlation coefficients are all positive and lead us to expect particularly pronounced co-movements among the European and US markets.

As described in Section 3.4.3, the ordering of the 5 return series within the Cholesky DPM-MSV model matters. As the natural ordering, we choose the chronological sequence, in which the respective stock markets start their trading day, i.e. $y_{1t}, \ldots, y_{5t}$ are the return series for (1) the Nikkei, (2) the Shanghai Shenzen, (3) the EuroStoxx50, (4) the Dax, and (5) the Dow Jones.

Table 3.2 about here

3.5.2 Estimation results

According to Eqs. (3.5) to (3.9) and the exhibition in Section 3.3.3, the estimation of our five-dimensional Cholesky-DPM-MSV model involves the sampling of (i) 5 stochastic volatility processes ($\mathbf{h}_t$-processes), (ii) 10 $\boldsymbol{\alpha}_t$-processes, (iii) 40 AR-parameters (stemming from the $\mathbf{h}_t$- and $\boldsymbol{\alpha}_t$-processes), and (iv) 5 DPM sets $\{\omega_{ij}, l_{ij}\}_{j=1}^{\infty}$. We ran a total of 50000 + 50000 iterations, and deleted the first 50000 results as burn-in phase. As prior distributions, we chose

$$
\begin{aligned}
c_{\boldsymbol{\alpha}i} &\sim N(0,1), \\
\phi_{\boldsymbol{\alpha}i} &\sim N(0.95,25)\mathbf{1}(|\phi_{\boldsymbol{\alpha}i}|<1), \\
\sigma^2_{\mathbf{e}i} &\sim \text{InverseGamma}(10/2, 0.5/2), \\
\phi_{\mathbf{h}i} &\sim N(0.95,25)\mathbf{1}(|\phi_{\mathbf{h}i}|<1), \\
\sigma^2_{\boldsymbol{\eta}i} &\sim \text{InverseGamma}(10/2, 0.5/2), \\
c_i &\sim \text{Gamma}(4,4),
\end{aligned}
$$

and the base distribution G_0 as Gamma$(10/2, 10/2)$. Table (3.2) displays the posterior means and standard deviations of the 40 AR parameters.

Figure 3.2 about here

We assess the volatility co-movements between the five markets via the pairwise in-sample time-varying correlation coefficients (denoted by $\text{Corr}_{\text{INDEX1, INDEX2};t}$), which we obtain from the overall time-varying covariance matrix $\mathbf{H}_t^*$ from Eq. (3.51) computed in each MCMC iteration and at every date t. Figure (3.2) displays the time-varying correlation coefficients for the 10 market pairs. In each panel, the solid line represents the correlation coefficients computed as an average of 333 posterior thinned draws (out of 50000), while the darkly and brightly shaded areas represent 50% and 90% Bayesian intervals, respectively.

Figure (3.2) provides the following major findings: (i) The time-varying in-sample correlation coefficients appear surprisingly volatile. (ii) Except for $\text{Corr}_{\text{DJ, EU};t}$ (US/European markets), $\text{Corr}_{\text{DJ, DAX};t}$ (US/German markets) and $\text{Corr}_{\text{DAX, EU};t}$ (German/European markets), the time-varying correlation coefficients take on negative values strikingly often. (iii) The coefficients $\text{Corr}_{\text{EU, SHA};t}, \text{Corr}_{\text{DAX, SHA};t}$, $\text{Corr}_{\text{DJ, NIK};t}, \text{Corr}_{\text{DJ, SHA};t}$ appear to fluctuate around mean levels close to zero, indicating rather weak correlation among the corresponding markets. (iv) During the Chinese stock-market downturn between 2015 and 2016, the coefficients $\text{Corr}_{\text{SHA, NIK};t}$ take on substantially smaller values (close to zero) than during all other phases of the sampling period. (v) The most stable, positive correlation coefficients are found between the German and the European stock markets ($\text{Corr}_{\text{DAX, EU};t}$), the US and the European markets ($\text{Corr}_{\text{DJ, EU};t}$), and the US and the German markets ($\text{Corr}_{\text{DJ, DAX};t}$).

Figure 3.3 about here

As a robustness check, Figure (3.3) depicts the sample correlations, obtained from a rolling window with a size of 50 and centered around the current observation.

Figure 3.4 about here

Figure 3.5 about here

Table 3.3 about here

Finally, we investigate the predictive ability of our Cholesky DPM-MSV model in terms of predictive density estimation. Figure (3.4) displays the nonparametric predictive densities of the elements of the covariance matrix $\mathbf{H}_t^*$, approximated according to Eqs. (3.52) and (3.53), while Figure (3.5) shows the pairwise density contour plots. The covariances from the one-step-ahead prediction closely follow the patterns obtained from the in-sample estimation. For example, the contour plots for the European and the Chinsese markets (Panel EU, SHA), the German and the Chinese Markets (Panel DAX, SHA), the US and the Japanese markets (Panel DJ, NIK), and the US and the Chinese markets (Panel DJ, SHA) all reflect the lack of linear dependence, as mentioned in the above discussion on Figure (3.2). Table (3.3) summarizes the posterior information of the one-step-ahead predictive density. Our model predicts the highest variance for the Japanese market (with the broadest 90% credibility interval), and the lowest variance for the US market.

3.6 Conclusion

In this chapter, we establish a Cholesky multivariate stochastic volatility model with a highly flexible nonparametric distribution for the innovation vector—based

on the Dirichlet process mixture (DPM)—and implement a Bayesian semiparametric estimation procedure. A striking advantage of our modeling framework is that it allows us to estimate DPM-based volatility models of higher dimensions ($m > 3$), without imposing unnecessarily restrictive assumptions. More concretely, this is due to the Cholesky structure, under which the common assumption of uncorrelated DPM error terms does not entail a flexibility loss, insofar as our overall covariance matrix $\mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Sigma}_t(\mathbf{A}_t^{-1})'$ contains DPM elements in its non-diagonal entries.

In the empirical section, we apply our estimation framework to five daily stock-index return series, with the aim of analyzing volatility co-movements among international stock markets. As two major empirical results, we find (i) a reduction in the co-movement between the Chinese and the Japanese markets during the recent Chinese stock-market downturn, and (ii) distinctively stable, positive co-movements among the European (including the German) and the US stock markets.

Two conceivable extensions of our modeling framework to be tackled in future research are worth mentioning. First, frequently observed volatility asymmetries could be modeled by integrating leverage effects into our Cholesky DPM-MSV framework. Second, our estimation framework could be applied to high-frequency data sets containing realized (co)variances along the lines of Shirota et al. (2016), who suggest estimating Cholesky realized stochastic volatility models.

Tables and Figures

Table 3.1: Descriptive statistics

	NIK	SHA	EU	DAX	DJ
Mean	0.0509	0.0177	0.0125	0.0302	0.0226
Median	0.0086	0.0000	0.0111	0.0614	0.0066
Variance	1.9457	2.7957	1.5524	1.4216	0.6229
Skewness	−0.2386	−0.8491	−0.1151	−0.2329	−0.1961
Kurtosis	6.3634	8.1768	4.4545	4.2553	4.7188
Sample correlation:					
NIK	1.0000				
SHA	0.2160	1.0000			
EU	0.2158	0.1349	1.0000		
DAX	0.2194	0.1390	0.9526	1.0000	
DJ	0.1224	0.1418	0.5929	0.5743	1.0000

Note: The indices are abbreviated as NIK (Nikkei 225), SHA (Shanghai Shenzen CSI 300), EU (EuroStoxx), DAX (DAX 30 Performance), DJ (Dow Jones Industrial).

Table 3.2: Posterior means and standard deviations (in parantheses)

i	$c_{\boldsymbol{\alpha}i}$	$\phi_{\boldsymbol{\alpha}i}$	$\sigma^2_{\mathbf{e}i}$	$\phi_{\mathbf{h}i}$	$\sigma^2_{\boldsymbol{\eta}i}$
1	-0.1455	0.0649	0.0474	0.9610	0.0510
	(0.0582)	(0.3532)	(0.0144)	(0.0158)	(0.0176)
2	-0.1614	0.0601	0.0657	0.9799	0.0364
	(0.0548)	(0.2965)	(0.0180)	(0.0087)	(0.0103)
3	-0.0662	-0.1444	0.0945	0.9323	0.0724
	(0.0338)	(0.1953)	(0.0234)	(0.0372)	(0.0499)
4	-0.0159	-0.0370	0.0261	0.9980	0.0330
	(0.0120)	(0.1403)	(0.0054)	(0.0013)	(0.0135)
5	0.0025	-0.0509	0.0310	0.9938	0.0674
	(0.0124)	(0.1679)	(0.0091)	(0.0039)	(0.0245)
6	-0.4111	0.5478	0.0254		
	(0.1070)	(0.1182)	(0.0069)		
7	0.0008	0.4161	0.0466		
	(0.0127)	(0.1831)	(0.0121)		
8	-0.0018	-0.4898	0.0277		
	(0.0218)	(0.1796)	(0.0107)		
9	-0.1832	0.2271	0.0498		
	(0.0986)	(0.3241)	(0.0159)		
10	-0.1745	-0.1649	0.0488		
	(0.0922)	(0.2732)	(0.0136)		

Table 3.3: Posterior summary of the elements of the one-step-ahead covariance matrix

$\mathbf{H}^{*}_{T+1}$	mean	median	90% CI
$\mathbf{H}^{*\text{NIK}}_{T+1}$	7.2337	6.1811	$(2.9575, 12.7184)$
$\mathbf{H}^{*\text{SHA}}_{T+1}$	3.2385	2.6356	$(1.0457, 6.1222)$
$\mathbf{H}^{*\text{EU}}_{T+1}$	3.7878	3.1497	$(1.4192, 6.8541)$
$\mathbf{H}^{*\text{DAX}}_{T+1}$	3.6246	2.9022	$(1.2207, 6.7965)$
$\mathbf{H}^{*\text{DJ}}_{T+1}$	1.6232	1.1733	$(0.3730, 3.3584)$
$\mathbf{H}^{*\text{SHA, NIK}}_{T+1}$	0.9987	0.7307	$(-0.8603, 3.2080)$
$\mathbf{H}^{*\text{EU, NIK}}_{T+1}$	1.5025	1.1124	$(-0.8145, 4.2972)$
$\mathbf{H}^{*\text{EU, SHA}}_{T+1}$	0.4096	0.2605	$(-0.9820, 1.9723)$
$\mathbf{H}^{*\text{DAX, NIK}}_{T+1}$	1.5457	1.1560	$(-0.8077, 4.3994)$
$\mathbf{H}^{*\text{DAX, SHA}}_{T+1}$	0.4052	0.2579	$(-1.0745, 2.0624)$
$\mathbf{H}^{*\text{DAX, EU}}_{T+1}$	3.5458	2.9134	$(1.2484, 6.5298)$
$\mathbf{H}^{*\text{DJ, NIK}}_{T+1}$	0.1496	0.0923	$(-1.8593, 2.2114)$
$\mathbf{H}^{*\text{DJ, SHA}}_{T+1}$	0.0052	-0.0103	$(-1.0902, 1.1261)$
$\mathbf{H}^{*\text{DJ, EU}}_{T+1}$	1.2784	0.9443	$(-0.1758, 3.1538)$
$\mathbf{H}^{*\text{DJ, DAX}}_{T+1}$	1.2232	0.8752	$(-0.1981, 3.1004)$

Note: The indices are abbreviated as NIK (Nikkei 225), SHA (Shanghai Shenzen CSI 300), EU (EuroStoxx), DAX (DAX 30 Performance), DJ (Dow Jones Industrial).

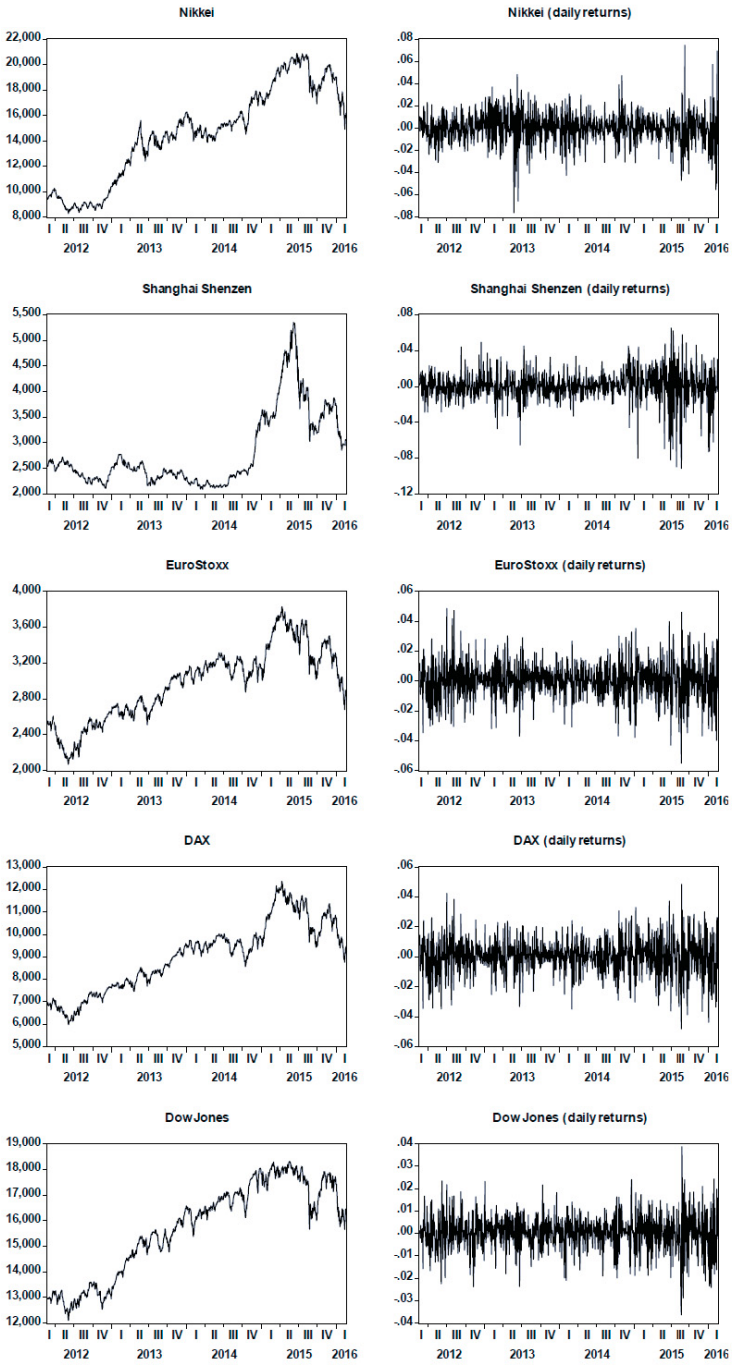


Figure 3.1: Index values and daily returns

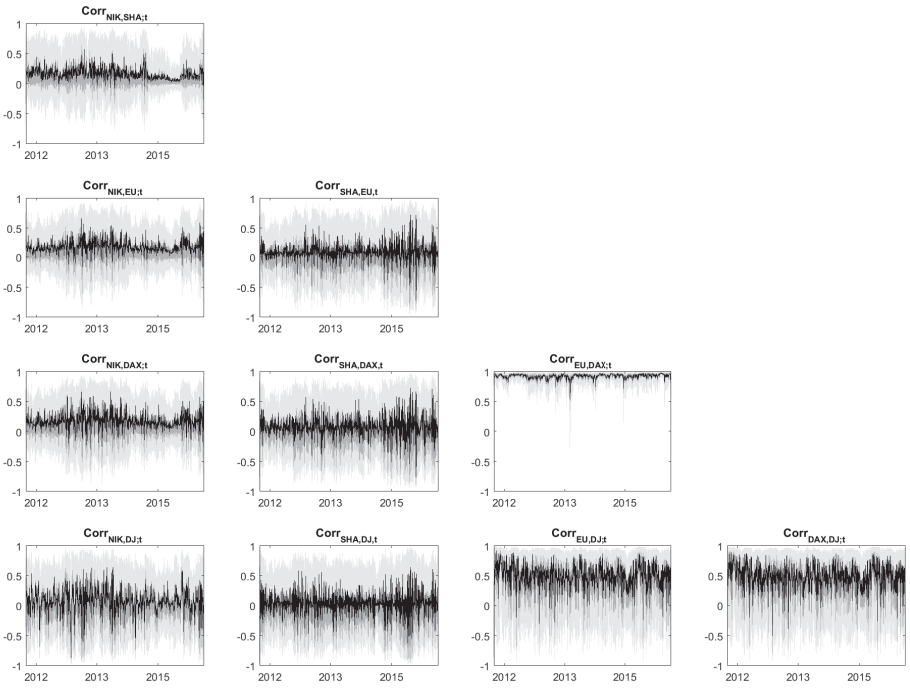


Figure 3.2: In-sample correlations: posterior mean plus 50% and 90% Bayesian intervals

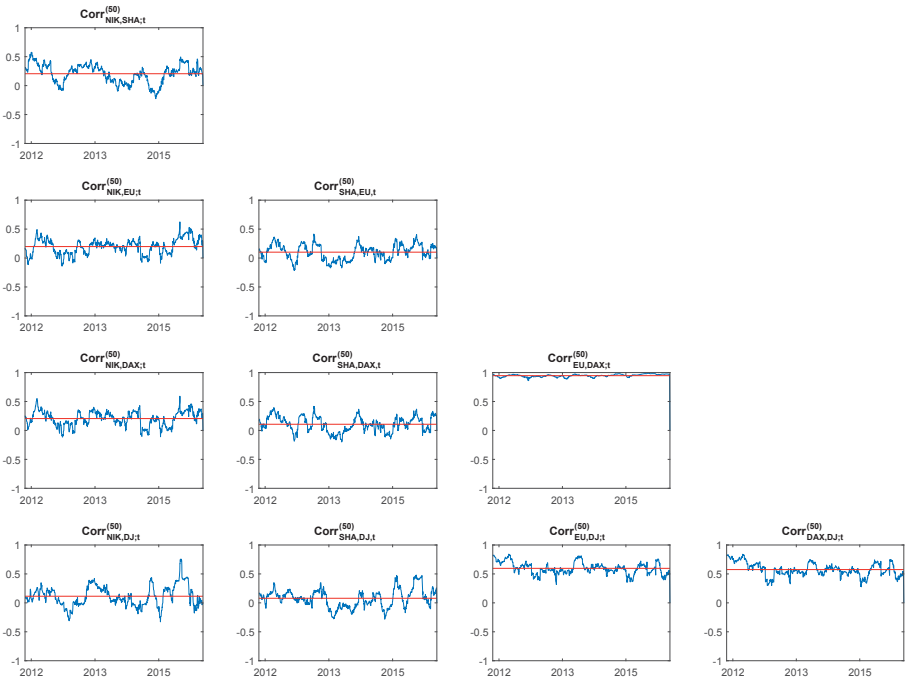


Figure 3.3: Sample correlations obtained from a rolling window of size 50 centered around t (blue). The horizontal line (red) corresponds to the sample-correlation.

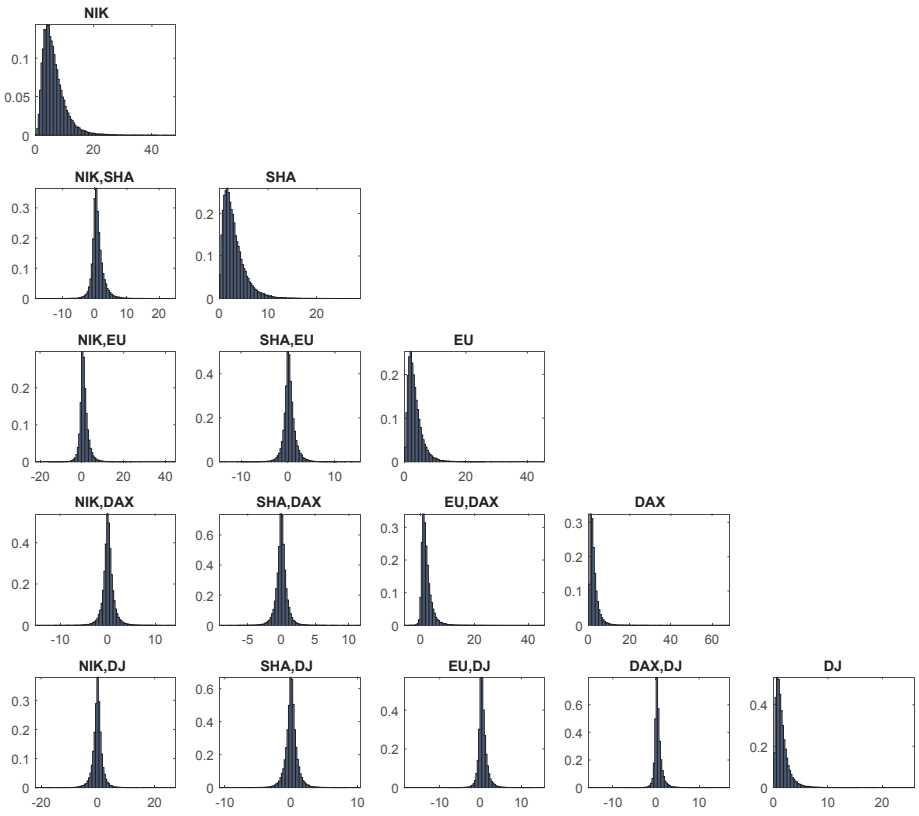


Figure 3.4: One-step-ahead density forecasts (of the elements of $\mathbf{H}^*_{T+1}$)

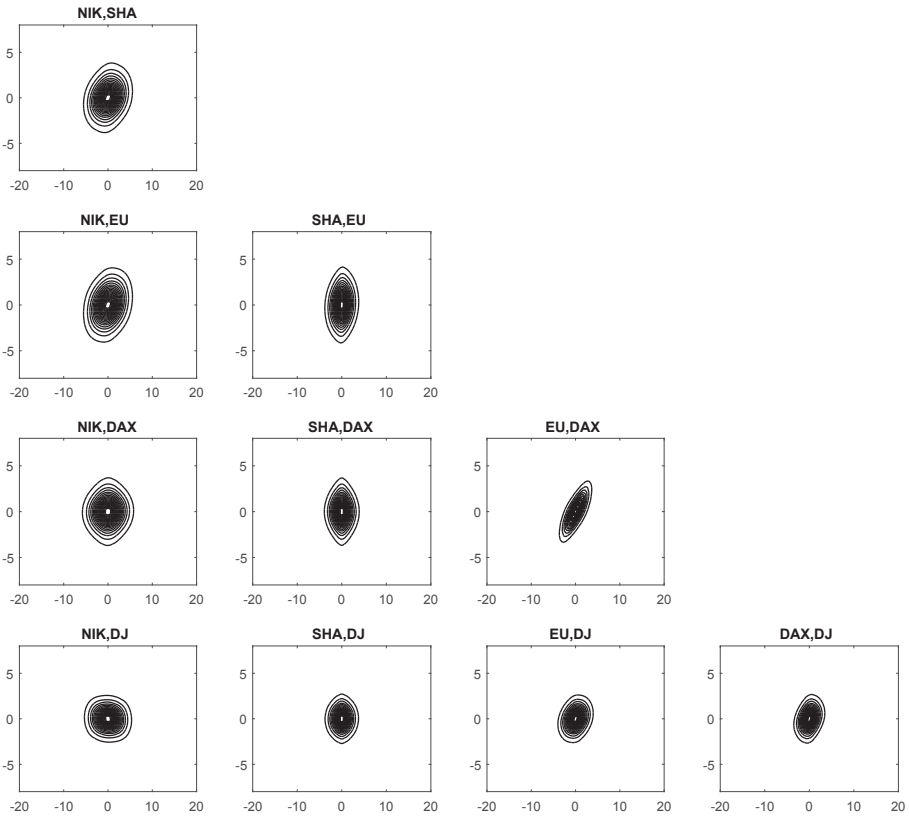


Figure 3.5: Pairwise one-step-ahead density forecasts

Chapter 4

Volatility transmission in global financial markets: A Bayesian nonparametric approach

4.1 Introduction

It is well known that stock markets in different countries can be correlated. However, the investigation of the channels through which this dependence takes place is a rather recent area of interest in the financial market research literature.

Two volatility transmission patterns proposed by Engle et al. (1990), are the *heat wave* and the *meteor shower*. The meteorological analogies are motivated by the statement that "... a hot day in New York is likely to be followed by another hot day in New York but not typically by a hot day in Tokyo" (Engle et al. (1990)). Therefore, volatility in one region is only determined by previous-day volatility in the same region and evolves like a *heat wave*. The *meteor shower*, on the other hand, rains down as the Earth turns and therefore, the alternative suggestion is that the volatility in one region is explained by that in another, which precedes in terms of calendar time. The hypothesis testing procedures conducted by Engle et al. (1990)

based on a Multivariate Generalized Conditional Autoregressive Heteroscedasticity (MGARCH) framework and intra-day exchange rate data support the meteor shower hypothesis.

In the context of volatility spillover effects, several lines of empirical research have also emerged in the past few years. Chang et al. (2015), among others, examine volatility spillovers in agricultural commodities and energy markets, and Hammoudeh et al. (2013) on the other hand investigate credit and market risk transmission using four oil-related Credit Default Swaps (CDS) indexes. Furthermore, with respect to international financial markets, Ehrmann et al. (2011) investigate the directions of volatility transmission between the the money, bond and equity market of the USA and Europe. Yet, not many studies have focused on advanced estimation methodology as most of them rest on Maximum Likelihood estimation techniques (see e.g. Corradi et al., 2012).

A very comprehensive study on volatility transmission in global financial markets is conducted by Clements et al. (2015), in which the findings of Engle et al. (1990) are re-investigated and additional volatility modeling strategies involving realized volatility measures are proposed. However, questions such as model fit diagnostics, model specification and a general discussion on the methodological issues related to the estimation procedure, are not the subject of this chapter. The multivariate GARCH framework relies on the assumption of a normal distribution, which has been widely rejected in favor, for example, of the Student-t, among other more flexible distributions. Misspecifying the unconditional distribution often leads to biased estimation of the volatility dynamics, as, for example, the lack of flexibility in the unconditional distribution can be overcompensated by additional volatility through the conditional.

In this chapter, the volatility transmission patterns examined by Clements et al. (2015) are revisited in a Bayesian framework. The estimation procedure applied here is the semiparametric MGARCH model proposed by Jensen and Maheu (2013). Accordingly, the assumption of a multivariate normal distribution for the return innovation vector is replaced by simply imposing a nonparametric prior over all possible distributions. The Markov Chain Monte Carlo (MCMC) estimation is conducted jointly for all parameters of the conditional variances and the unknown innovation distribution, where a parametric model is assumed for the dynamics of the conditional covariance matrix and a nonparametric Dirichlet process mixture (DPM) prior for the error term. The proposed modeling framework has the advantage of capturing the heavy tails and the asymmetries in the return distribution.

The nonparametric approach makes use of the stick-breaking representation of the Dirichlet process (DP) prior, and the estimation procedure follows the efficient slice-sampler of Walker (2007).

A similar procedure applied to a univariate GARCH model is found in Ausin et al. (2014) and a portfolio decision based on Bayesian nonparametric Asymmetric Dynamic Conditional Correlation MGARCH (ADCC-MGARCH) model is conducted by Virbickaitė et al. (2014). For a complete overview of Bayesian MGARCH models, including the semiparametric MGARCH, see Virbickaitė et al. (2015).

The remainder of the following chapter is organized as follows. Section 4.2 provides five model set-ups for volatility transmission, based on multivariate GARCH frameworks. In addition to the general framework, realized volatility estimates as explanatory variables are introduced. Section 4.3 provides details on the MCMC algorithm used for the estimation of the semiparametric model. Section 4.4 includes

a description of the data used for the estimation and a discussion of the estimation results. Finally, some concluding remarks are presented in Section 4.5.

4.2 Volatility transmission modeling

4.2.1 Heat waves and meteor showers

Engle et al. (1990) propose testing the null hypothesis of heat waves against the alternative of meteor showers. The heat wave hypothesis suggests that volatility has only country-specific autocorrelation and alternatively, the meteor shower is a phenomenon of intra-day volatility spillovers from one market to another. The null hypothesis is tested using the following multivariate GARCH model specification:

$$R_{it} = \sqrt{h_{it}}\epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{F} \tag{4.1}$$

$$h_{it} = k_i + \alpha_{ii} h_{it-1} + \sum_{j=1}^{i-1} \beta_{ij} R_{jt}^2 + \sum_{j=i}^{n} \gamma_{ij} R_{jt-1}^2. \tag{4.2}$$

Here, R_{it} is the return in zone i at date t, h_{it} is the conditional variance in zone i $(i = 1, \ldots, n)$ at time t $(t = 1, \ldots, T)$ and $\mathcal{F}$ is some unknown distribution, which is subsequently modeled nonparametrically. This model specification is different from the classical GARCH, by virtue of allowing for news from preceding zones to influence volatility on the same day in the current zone. Moreover, when trading in zone $i = 1$ ends, trading zone $i = 2$ starts and so on. Hence, the modeling is restricted to non-overlapping trading zones. The intra-day return of each zone, R_{it} is calculated as the difference between the log closing and opening prices (regarding the observed trading period), normalized by the length of the trading period, i.e.

$$R_{it} = (\ln \mathrm{PC}_{it} - \ln \mathrm{PO}_{it})/\sqrt{(T_{it})}, \tag{4.3}$$

where PC_{it} and PO_{it} are the closing and opening price in zone i on day t and T_{it} is the duration of trading in the given zone.

Introducing matrix notation to the model reveals the calendar structure, which allows us to examine the dynamic market reaction to news:

$$\mathbf{h}_t = \mathbf{K} + \mathbf{A}\mathbf{h}_{t-1} + \mathbf{B}\mathbf{R}_t^2 + \mathbf{G}\mathbf{R}_{t-1}^2, \tag{4.4}$$

where $\mathbf{h}_t$ is the $n \times 1$ vector $\mathbf{h}_t = (h_{1t}, \ldots, h_{nt})'$, $\mathbf{R}_t$ - the $n \times 1$ vector $\mathbf{R}_t = (R_{1t}, \ldots, R_{nt})'$ of observed returns at time t, $\mathbf{R}_t^2$ - the $n \times 1$ vector $\mathbf{R}_t^2 = (R_{1t}^2, \ldots, R_{nt}^2)'$ of the squared returns and $\mathbf{K}$ is a vector of constants $(k_1, \ldots, k_n)'$. The parameter matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{G}$ are defined as

$$\mathbf{A} = \begin{pmatrix} \alpha_{11} & 0 & \ldots & 0 \\ 0 & \alpha_{22} & \ldots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \ldots & \alpha_{nn} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ \beta_{21} & 0 & \ldots & 0 \\ \vdots & & & \vdots \\ \beta_{n1} & \beta_{n2} & \ldots & 0 \end{pmatrix}, \mathbf{G} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \ldots & \gamma_{1n} \\ 0 & \gamma_{22} & \ldots & \gamma_{2n} \\ \vdots & & & \vdots \\ 0 & \ldots & \ldots & \gamma_{nn} \end{pmatrix}. \tag{4.5}$$

Owing to the lower triangular structure of the matrix $\mathbf{B}$, it is apparent that news from zone i can be transmitted to zone $i+1$, and that the effect is captured by the coefficient $\beta_{i+1,i}$. However, news from zone $i+1$ can be transmitted to zone i only on the next day. These effects are captured by the off-diagonal elements of the matrix $\mathbf{G}$. The heat wave null hypothesis implies $\beta_{ij} = 0$ and $\gamma_{ij} = 0$ for $i \neq j$. The meteor shower hypothesis is the alternative, implying that at least one of the $\beta_{ij} = 0$ and $\gamma_{ij} = 0$ for $i \neq j$ coefficients is significantly different from zero.

4.2.2 Volatility spillovers

The intra-day volatility effect h_{it} from the preceding zone on the current zone volatility, referred to as a *volatility spillover* effect, can be captured by modifying the model in the following way:

$$\tilde{\mathbf{A}}\mathbf{h}_t = \mathbf{K} + \mathbf{A}\mathbf{h}_{t-1} + \mathbf{B}\mathbf{R}_t^2 + \mathbf{G}\mathbf{R}_{t-1}^2, \text{ with}$$

$$\tilde{\mathbf{A}} = \begin{pmatrix} 1 & 0 & 0 \\ -\tilde{\alpha}_{21} & 1 & 0 \\ 0 & -\tilde{\alpha}_{32} & 1 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \alpha_{11} & 0 & \alpha_{13} \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{pmatrix}. \tag{4.6}$$

The modeling framework above is an example of three trading zones, as in Clements et al. (2015). The two matrices jointly ensure the circularity of the volatility transmission through the system. The matrix $\tilde{\mathbf{A}}$ captures the effect of the conditional variance in preceding zones on the current zone on the same trading day, and the upper right element of the $\mathbf{A}$ matrix closes the circle by introducing the effect of the previous day's conditional variance of zone $n = 3$ on the variance of zone 1 today.

For the estimation, two restrictions are imposed: $\mathbf{B} = \mathbf{0}$ and $\mathbf{G}$ is modeled as a diagonal matrix, given by

$$\mathbf{G} = \begin{pmatrix} \gamma_{11} & 0 & 0 \\ 0 & \gamma_{22} & 0 \\ 0 & 0 & \gamma_{33} \end{pmatrix}. \tag{4.7}$$

Therefore, there are no intra-day effects from the news of the preceding zones, which isolates the spillover effect and at the same time reduces the number of parameters.

4.2.3 Realized volatility, jumps and asymmetry

4.2.3.1 Volatility measures

By using high-frequency data, a realized variance estimator (RV) can easily be computed. According to Liu et al. (2015), the simple 5-minute realized volatility estimator (RV) appears to be superior in comparison to other volatility measures and will therefore be the estimator used here. First, the returns at frequency Δ for each day t, in which trading occurs are computed as

$$r^i_{j,t}(\Delta) = \ln(p^i_{j,t}) - \ln(p^i_{j-1,t}),$$

where $p^i_{j,t}$ is the price at day t in zone i and in the time interval j. The realized volatility is computed by summing the squared returns defined above:

$$\mathrm{RV}^i_t(\Delta) \equiv \sum_{j=1}^{\Delta} (r^i_{j,t})^2.$$

Next, the jump component can be extracted from the realized volatility. For this purpose, following Andersen et al. (2012), the minimum realized volatility estimator (MinRV) and the minimum realized quarticity (MinRQ) are computed as

$$\mathrm{MinRV}^i_t(\Delta) \equiv \frac{\pi}{\pi - 2}\left(\frac{1}{1-\Delta}\right)\sum_{j=2}^{\Delta} \min(|r^i_{j,t}|, |r^i_{j-1,t}|)^2,$$

$$\mathrm{MinRQ}^i_t(\Delta) \equiv \frac{\pi}{\Delta(3\pi - 8)}\left(\frac{1}{1-\Delta}\right)\sum_{j=2}^{\Delta} \min(|r^i_{j,t}|, |r^i_{j-1,t}|)^4.$$

Statistically significant jumps J^i_t (at significance level $\bar{\alpha} = 0.999$) are identified according to

$$J^i_t(\Delta) \equiv \mathbf{1}(Z^i_t(\Delta) > \Phi_{\bar{\alpha}}) \times (\mathrm{RV}^i_t(\Delta) - \mathrm{MinRV}^i_t(\Delta)),$$

which is based on asymptotic results from Andersen et al. (2012), providing

$$Z_t^i(\Delta) = \frac{(\mathrm{RV}_t^i(\Delta) - \mathrm{MinRV}_t^i(\Delta))/\mathrm{RV}_t^i(\Delta)}{\sqrt{1.81\Delta \max(1, \mathrm{MinRQ}_t^i(\Delta)/\mathrm{MinRV}_t^i(\Delta)^2)}} \sim N(0,1).$$

Here $\mathbf{1}(\cdot)$ is the indicator function and $\Phi_{\bar{\alpha}}$ is the $\bar{\alpha}$-quantile of the standard normal distribution.

The continuous component is defined as the difference between the realized volatility and the jump component:

$$\mathrm{CC}_t^i(\Delta) = \mathrm{RV}_t^i(\Delta) - J_t^i(\Delta). \tag{4.8}$$

In order to capture the asymmetry of volatility transmission, the realized volatility can be decomposed into realized volatility related to positive news (RV_t^+), and related to negative news (RV_t^-) such as

$$\mathrm{RV}_t^{i+} = \sum (r_{i,j}^i)^2 \mathbf{1}(r_{i,j}^i \geq 0), \tag{4.9}$$

$$\mathrm{RV}_t^{i-} = \sum (r_{i,j}^i)^2 \mathbf{1}(r_{i,j}^i < 0), \tag{4.10}$$

where the zero returns are treated as positive news.

4.2.3.2 Three models

Finally, using realized volatility estimates from the different regions, the volatility model can be rewritten in the following three alternative specifications:

$$\mathbf{h}_t = \mathbf{K} + \mathbf{A}\mathbf{h}_{t-1} + \mathbf{B}\ \mathbf{RV}_t + \mathbf{G}\ \mathbf{R}_{t-1}^2, \tag{4.11}$$

$$\mathbf{h}_t = \mathbf{K} + \mathbf{A}\mathbf{h}_{t-1} + \mathbf{B}\ \mathbf{CC}_t + \tilde{\mathbf{B}}\ \mathbf{J}_t + \mathbf{G}\ \mathbf{R}_{t-1}^2, \tag{4.12}$$

$$\mathbf{h}_t = \mathbf{K} + \mathbf{A}\mathbf{h}_{t-1} + \mathbf{B}\ \mathbf{RS}_t^+ + \tilde{\mathbf{B}}\ \mathbf{RS}_t^+ + \mathbf{G}\ \mathbf{R}_{t-1}^2, \tag{4.13}$$

where $\mathbf{K}$, $\mathbf{A}$ are defined as in Eqs. (4.4) and (4.5). $\mathbf{G}$ is restricted to be diagonal (see Eq. (4.7)) and for the three-dimensional case the matrices $\mathbf{B}$ and $\tilde{\mathbf{B}}$ are defined as

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & \beta_{13} \\ \beta_{21} & 0 & 0 \\ 0 & \beta_{32} & 0 \end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix} 0 & 0 & \tilde{\beta}_{13} \\ \tilde{\beta}_{21} & 0 & 0 \\ 0 & \tilde{\beta}_{32} & 0 \end{pmatrix}. \tag{4.14}$$

In this case the realized volatility measures are the following 1×3 vectors:

$$\begin{aligned} \mathbf{RV}_t &= (\mathrm{RV}_{1,t}, \mathrm{RV}_{2,t}, \mathrm{RV}_{3,t-1})', \\ \mathbf{CC}_t &= (\mathrm{CC}_{1,t}, \mathrm{CC}_{2,t}, \mathrm{CC}_{3,t-1})', \\ \mathbf{J}_t &= (J_{1,t}, J_{2,t}, J_{3,t-1})', \\ \mathbf{RS}_t^+ &= (\mathrm{RS}_{1,t}^+, \mathrm{RS}_{2,t}^+, \mathrm{RS}_{3,t-1}^+)', \\ \mathbf{RS}_t^- &= (\mathrm{RS}_{1,t}^-, \mathrm{RS}_{2,t}^-, \mathrm{RS}_{3,t-1}^-)' \end{aligned}$$

Equation (4.11) introduces a realized volatility estimate ($\mathbf{RV}_t$), while the models presented in Eqs. (4.12) and (4.13) consider two alternative decompositions of the realized volatility. In Eq. (4.12) the realized volatility is decomposed into a jump component and a continuous component, and in Eq. (4.13), the effects due to upside and downside realized semivariance are incorporated into the model.

4.3 Bayesian semiparametric MGARCH model

The Bayesian model framework for each of the described specifications is a multivariate GARCH model, which follows the general semiparametric MGARCH of Jensen and Maheu (2013). Specifically, a parametric model is introduced for the dy-

namics of the conditional covariance matrix, while a nonparametric prior is placed on the distribution of the error term, in particular, a Dirichlet process prior. The hierarchical structure of the model labeled as DPM-MGARCH is given by

$$\begin{aligned}
\mathbf{R}_t \,|\boldsymbol{\tau}_t, \mathbf{D}_t, \mathbf{H}_t &\sim \mathbf{N}(\mathbf{H}_t^{1/2}\boldsymbol{\tau}_t,\ \mathbf{H}_t^{1/2}\mathbf{D}_t^{-1}(\mathbf{H}_t^{1/2})'), && (4.15)\\
\mathbf{H}_t &= \text{diag}(\mathbf{h}_t), && (4.16)\\
\boldsymbol{\tau}_t, \mathbf{D}_t|\, \mathbf{G} &\overset{iid}{\sim} \mathbf{G} && (4.17)\\
\mathbf{G}|\, \mathbf{G_0}, c &\sim DP(c, \mathbf{G_0}) && (4.18)\\
\mathbf{G_0}(\boldsymbol{\tau}_t, \mathbf{D}_t) &\equiv \mathcal{NW}(\mathbf{m_0}, s_0, \mathbf{W_0}, d_0) && (4.19)\\
c &\sim \text{Gamma}(a_0, b_0). && (4.20)
\end{aligned}$$

Eq. (4.15) defines a multivariate normal kernel for each mixture component, where $\mathbf{H}_t$ is the parametric time-varying conditional covariance with a Cholesky decomposition given by $\mathbf{H}_t^{1/2}$, which is diagonal (4.16) and has a structure defined according the models in Eqs. (4.5), (4.6), (4.11), (4.12) and (4.13).

Equations (4.17)-(4.19) place a nonparametric prior on the distribution of the innovations. The Bayesian nonparametric prior used here produces an infinite mixture of multivariate normals with mixing over both expectation parameter vector and covariance matrix. Both unknown parameters vectors $\boldsymbol{\tau}_t$ and $\mathbf{D}_t$ are distributed according to some unknown nonparametric distribution $\mathbf{G}$, on which a Dirichlet process (DP) prior (Ferguson, 1973) with base measure $\mathbf{G_0}$ and precision parameter c (positive scaling parameter) is placed. Therefore, the unconditional distribution is an infinite mixture of normals with unique mixture expectation vector and covari-

ance matrix of component j being $\boldsymbol{\mu}_j$ and $\boldsymbol{\Lambda}_j$ respectively. The base distribution in Eq. (4.19) is the Normal-Wishart denoted by $\mathcal{NW}(\mathbf{m_0}, s_0, \mathbf{W_0}, d_0)$ with

$$\begin{aligned} \boldsymbol{\mu}_j|\boldsymbol{\Lambda}_j &\sim \mathbf{N}(\mathbf{m_0}, (s_0\boldsymbol{\Lambda}_j)^{-1}) && (4.21) \\ \boldsymbol{\Lambda}_j &\sim \mathcal{W}(\mathbf{W_0}, d_0) && (4.22) \end{aligned}$$

where $\mathbf{N}$ is the multivariate normal distribution, $\mathcal{W}$ denotes the Wishart distribution and $j = 1, 2, \ldots, \infty$.

The resulting model for the return vector $\mathbf{R}_t$ is a DPM mixture with an infinite mixture representation (Sethuraman, 1994) given by

$$f(\mathbf{R}_t|\,\mathbf{H}_t, \{\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\omega}\}) = \sum_{j=1}^{\infty} \omega_j f_{\mathbf{N}}(\mathbf{R}_t|\,\mathbf{H}_t^{1/2}\boldsymbol{\mu}_j,\ \mathbf{H}_t^{1/2}\boldsymbol{\Lambda}_j^{-1}(\mathbf{H}_t^{1/2})'), \tag{4.23}$$

where $\omega_1 = \upsilon_1$ and $\omega_j = \upsilon_j \prod_{s=1}^{j-1}(1-\upsilon_s)$ with $\upsilon_j \sim \text{Beta}(1, c)$ for $j > 1$ are the mixture weights computed by the stick-breaking rule, $f_{\mathbf{N}}$ denotes the multivariate Gaussian density with expectation vector $\mathbf{H}_t^{1/2}\boldsymbol{\mu}_j$ and covariance matrix $\mathbf{H}_t^{1/2}\boldsymbol{\Lambda}_j^{-1}(\mathbf{H}_t^{1/2})'$, and $\boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j$ are the unique component parameters.

Note that the general semiparametric MGARCH framework can be applied to any model specification provided the transformation

$$\begin{aligned} \boldsymbol{\epsilon}_t &= \mathbf{H}_t^{-1/2}\mathbf{R}_t, && (4.24) \\ \boldsymbol{\epsilon}_t|\,\boldsymbol{\tau}_t, \mathbf{B}_t &\sim \mathbf{N}(\boldsymbol{\tau}_t, \mathbf{B}_t^{-1}). && (4.25) \end{aligned}$$

is possible. Therefore, the distribution of the standardized returns in (4.24) is directly the distribution of the error term, which is modeled as a flexible DPM. As noted by Jensen and Maheu (2013), and by Virbickaitė et al. (2014), the mixture model for $\boldsymbol{\epsilon}_t$ does not impose a unit covariance matrix for the DPM. However, in

the application a prior is used, which will essentially center the error term around a unit covariance matrix.

4.4 Bayesian inference

4.4.1 Slice-sampler for multivariate DPM models

Kalli et al. (2011) and Walker (2007) propose a DPM sampler, which deals with the problem of sampling infinite number of parameters by introducing latent variables and truncating the weights of the random Dirichlet distribution. Note that an alternative to the slice-sampler is the retrospective sampler of Papaspiliopoulos and Roberts (2008), and a combination of both is presented by Papaspiliopoulos (2008).

The following section presents a multivariate extension of the slice-sampler described in Chapter 3 (Section 3.3.3). Analogously to the univariate case, a latent variable ρ_t has been introduced, such as

$$\begin{aligned} f(\boldsymbol{\epsilon}_t, u_t|\boldsymbol{\Theta}) &= \sum_{j=1}^{\infty} \mathbf{1}(\rho_t < \omega_j) \cdot f_{\mathbf{N}}(\boldsymbol{\epsilon}_t|\boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j^{-1}) \\ &= \sum\nolimits_{j \in \mathcal{A}_\omega(\rho_t)} f_{\mathbf{N}}(\boldsymbol{\epsilon}_t|\boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j^{-1}), \end{aligned} \tag{4.26}$$

where the set $\mathcal{A}_\omega(\rho_t) \equiv \{j : \omega_j > \rho_t\}$ is finite for all $\rho_t > 0$, $\boldsymbol{\Theta}$ is the full set of parameters and $\boldsymbol{\epsilon}_t$ is standardized according (4.24). Obviously, integrating over the latent variable yields the correct infinite mixture. Therefore, the infinite sum becomes finite and sampling is now replicable.

Again, analogously to the univariate case, a second latent variable is introduced, the indicator variable ζ_t, which indicates from which mixture component each $\boldsymbol{\epsilon}_t$ is taken. The joint conditional density is defined as

$$f(\boldsymbol{\epsilon}_t, \zeta_t = j, \rho_t | \boldsymbol{\Theta}) = f_{\mathbf{N}}(\boldsymbol{\epsilon}_t | \boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j^{-1}) \mathbf{1}(j \in \mathcal{A}_\omega(\rho_t)). \tag{4.27}$$

Finally, the likelihood can be computed, conditional on all latent variables ρ_t and ζ_t.

4.4.2 The sampling algorithm for the DPM-MGARCH model

This section presents the Markov Chain Monte Carlo (MCMC) algorithm for the estimation of the semiparametric multivariate GARCH model described in Section 4.3. The Bayesian inference is based on the estimation procedure developed by Jensen and Maheu (2013) and Virbickaitė et al. (2016). The DPM part of the algorithm employs the stick-breaking representation, the sampling strategy of which is discussed in Section 4.4.1. For the parametric part of the model, a multivariate Random Walk Metropolis Hastings (RWMH) algorithm is applied.

By initializing the starting values $c^{(0)}, \zeta_1^{(0)}, \ldots, \zeta_T^{(0)}$, for each iteration step $r = 1, \ldots, \bar{R}$, the complete sampling algorithm can be summarized in the following steps:

1. Sampling c:

   As in Escobar and West (1995), given the auxiliary variable $\psi \sim \text{Beta}(c^{(r-1)} + 1, T)$, the precision parameter c is sampled from the Gamma mixture

$$\pi_\psi \cdot f_\Gamma(c | a_0 + \zeta^*, b_0 - \log(\psi)) + (1 - \pi_\psi) \cdot f_\Gamma(c | a_0 + \zeta^* - 1, b_0 - \log(\psi)),$$

where $f_\Gamma(\cdot|\alpha,\beta)$ denotes the density function of the Gamma(α,β) distribution, $\zeta^* = \max\left\{\zeta_1^{(r-1)},\ldots,\zeta_T^{(r-1)}\right\}$ and $\pi_\psi = (a_0+\zeta^*-1)/(a_0+\zeta^*-1+T(b_0-\log(\psi)))$.

2. Sampling υ_j:

   For $j = 1,2,\ldots,\zeta^*$ the parameter υ_j is sampled from

$$\upsilon_j|\,\zeta_1^{(r-1)},\ldots,\zeta_T^{(r-1)} \sim \text{Beta}\left(n_j+1, T-\bar{n}+c\right),$$

   where $n_j = \sum_{t=1}^T \mathbf{1}(\zeta_t^{(r-1)} = j)$ and $\bar{n} = \sum_{k=1}^j n_k$. The associated mixture weights are computed according to the stick-breaking procedure, $\omega_1 = \upsilon_1$, and $\omega_j = (1-\upsilon_j)\ldots(1-\upsilon_{j-1})\upsilon_j$ for $j = 2,\ldots,\zeta^*$.

3. Sampling ρ_t:

   We sample from $\rho_t \sim U(0,\omega_{\zeta_t^{(r-1)}})$ and set $\rho^* = \min\{\rho_1,\ldots,\rho_T\}$.

4. Updating the weights ω_j:

   We find the smallest integer j^* such that $\sum_{j=1}^{j^*}\omega_j > (1-\rho^*)$. For all $j > \zeta^*$, we draw $\upsilon_j \sim \text{Beta}(c,1)$ and compute the associated weights ω_j according to the stick-breaking procedure for $j = \zeta^*+1,\ldots,j^*$.

5. Sampling the mixture parameters $\boldsymbol{\mu}_j$ and $\boldsymbol{\Lambda}_j$:

   The mixture parameters are sampled from the conditional posterior, which is a Normal-Wishart distribution (see Virbickaitė et al., 2016). Alternatively, the mean and the covariance matrix can be drawn by a Gibbs sampler, if independent priors are assumed (as in Jensen and Maheu, 2013). Here, for $j = 1,\ldots,j^*$, the conditional posterior is $(\boldsymbol{\mu}_j,\boldsymbol{\Lambda}_j) \sim \mathcal{NW}(\mathbf{m}_j, s_j, \mathbf{W}_j, d_j)$ with

$$\mathbf{m}_j = \frac{s_0\mathbf{m_0}+n_j\bar{\boldsymbol{\epsilon}}_j}{s_0+n_j}, \qquad \bar{\boldsymbol{\epsilon}}_j = \frac{1}{n_j}\sum_{t:\zeta_t=j}^{T}\boldsymbol{\epsilon}_t$$

$$
\begin{aligned}
s_j &= s_0 + n_j, \\
\mathbf{W}_j &= \mathbf{W}_\mathbf{0}^{-1} + \mathbf{S}_j + \frac{s_0 n_j}{s_0 + n_j}(\mathbf{m_0} - \bar{\boldsymbol{\epsilon}}_j)(\mathbf{m_0} - \bar{\boldsymbol{\epsilon}}_j)', \\
\mathbf{S}_j &= \frac{1}{n_j} \sum_{t:\zeta_t = j}^{T} (\boldsymbol{\epsilon}_t - \bar{\boldsymbol{\epsilon}}_j)(\boldsymbol{\epsilon}_t - \bar{\boldsymbol{\epsilon}}_j)', \\
d_j &= d_0 + n_j.
\end{aligned}
$$

Again, if a new component has been formed, the mixture parameters are sampled from the prior.

6. Updating the indicator variables ζ_t:

   According to the weight truncation induced by the variable ρ_t, the indicator variables ζ_t are updated by sampling from

   $$\Pr(\zeta_t = j | \{\boldsymbol{\epsilon}_t\}_{t=1}^T, \{\boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j\}_{j=1}^{j^*}, \{\omega_j\}_{j=1}^{j^*}, \{\rho_t\}_{t=1}^T) \propto f_\mathbf{N}(\boldsymbol{\epsilon}_t | \boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j^{-1}) \cdot \mathbf{1}(j \in \mathcal{A}_\omega(\rho_t)).$$

   The updated variables ζ_t indicate the component to which each observation belongs. Given ζ_t, we set $\boldsymbol{\tau}_t = \boldsymbol{\mu}_{\zeta_t}$ and $\mathbf{B}_t = \boldsymbol{\Lambda}_{\zeta_t}$.

7. Computing the (parametric) time-varying conditional covariance matrix $\mathbf{H}_t = \operatorname{diag}\{\mathbf{h}_t\}$.

   The non-zero elements of the matrices $\mathbf{A}$, $\mathbf{B}$ $\mathbf{G}$, $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ (depending on the model specification) are collected in the parameter vector $\boldsymbol{\Phi}$. The parameters in $\boldsymbol{\Phi}$ are sampled jointly by applying the RWMH algorithm, which can be presented in two sub-steps:

(a.) We generate a candidate vector $\tilde{\mathbf{\Phi}}$ from an multivariate fat-tailed mixture of normals centered around $\mathbf{\Phi}$ (the current value):

$$\tilde{\Phi} \sim \begin{cases} \mathbf{N}(\mathbf{\Phi}, \mathbf{V}) & \text{with prob. } p \\ \mathbf{N}(\mathbf{\Phi}, 100\mathbf{V}) & \text{with prob. } 1-p \end{cases}$$

(b.) We accept the candidate with probability

$$\min\left\{1,\, p(\tilde{\mathbf{\Phi}}, \mathcal{R})/p(\mathbf{\Phi}, \mathcal{R})\right\}.$$

The covariance $\mathbf{V}$ is obtained by running several initial iterations ($\approx$ burn-in) and then adjusting the sample covariance by some factor in order to achieve an acceptance rate between 0.2 and 0.5. The second normal in the mixture proposal allows for large moves in the parameter space with probability $1-p$. In the empirical application, p is set to 0.9.

In sub-step (b), the acceptance probability depends on the posterior ratio, which is a function of the complete information set $\mathcal{R} = \{R_{it}\}_{\substack{i=1,\ldots,n \\ t=1,\ldots,T}}$. The prior is uniform and restricted to the stationary region of the parameter space. Therefore, the acceptance probability reduces to $\min\left\{1,\, \prod_{t=1}^{T} f_{\mathbf{N}}(\tilde{\mathbf{\Phi}}|\mathbf{R}_t) / \prod_{t=1}^{T} f_{\mathbf{N}}(\mathbf{\Phi}|\mathbf{R}_t)\right\}$, where the log-likelihood $l(\mathbf{\Phi}|\mathcal{R})$ obtains as

$$l(\mathbf{\Phi}|\mathcal{R}) = -\frac{1}{2}\sum_{t=1}^{T}(n \cdot \log(2\pi) + \log|\mathbf{H}_t^*| + (\mathbf{R}_t - \boldsymbol{\mu}_t^*)\mathbf{H}_t^{*-1}(\mathbf{R}_t - \boldsymbol{\mu}_t^*)'), \quad (4.28)$$

where $\boldsymbol{\mu}_t^* = \mathbf{H}_t^{1/2}\boldsymbol{\tau}_t$ and $\mathbf{H}_t^* = \mathbf{H}_t^{1/2}\mathbf{D}_t^{-1}(\mathbf{H}_t^{1/2})'$ are the overall expectation and covariance matrix from (4.15).

8. Recalculating the error term $\boldsymbol{\epsilon}_t = \mathbf{H}_t^{-1/2}\mathbf{R}_t$ and returning to the first step.

4.5 Empirical application

4.5.1 Data

The data set used in Clements et al. (2015) contains high-frequency data gathered from Thompson Reuters Tick History for the period 3 January 2005 to 28 February 2013 at a 10-minute frequency rate. The data is available for three markets:

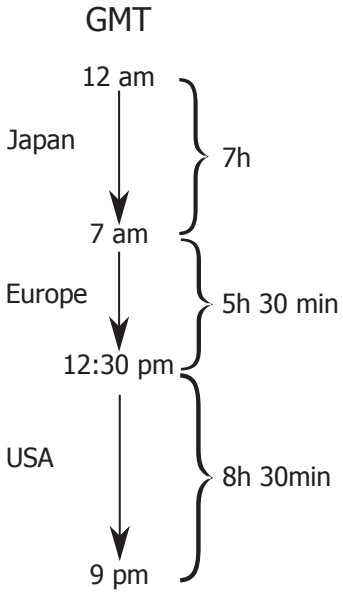
1. Foreign exchange market (FX): €/$ US future contracts, traded on the Chicago Mercantile Exchange
2. Bond market (Bond): US 10 year Treasury bond future contracts
3. Equity market (EQ): S&P 500 future contracts

The described data relate to U.S.-based assets and will be used for the empirical application.[1]

The global trading day is constructed by merging three trading zones. First is the Japanese trading zone, which starts at 12am and ends at 7am Greenwich Mean Time (GMT). Next, 7am to 12:30 pm GMT is the European trading zone and finally, 12:30pm to 9pm GMT is the United States zone. The 21-hour global trading day can be depicted as follows:

There is an obvious overlap of approximately 4 hours between the European and the US market, as the London Stock Exchange is open until 16:25(GMT). Clements et al. (2015) propose subsuming the overlap into the US market. Even though there is not much evidence in the literature to support that assumption, Clements

[1] All data is available at http://www.ncer.edu.au/resources/data-and-code.php, where also additional data sets containing Japanese and German data can be downloaded.



et al. (2015) motivate their choice by analyzing the average trading volume of S&P 500, US 10 year Treasury bond and €/$ futures contracts in that period. After a period of low trading activity in the Japanese trading hours, a large increase in volume is observed between 12:00-14:30(GMT), the pre-trading period in Chicago. Therefore, one could argue that the increased trading volume is due to US news (see e.g. Dungey et al., 2009).

Detailed discussion on the institutional background of the above-listed futures contracts can be found in Volkov (2015).

Table 4.1 about here

Figure 4.1 about here

Figure 4.2 about here

Figure 4.3 about here

Descriptive statistics of the intra-day returns calculated according to Eq.(4.3) are provided in Table 4.1. The corresponding time series plots are provided in Figures

4.1, 4.2 and 4.3 for the bond, foreign exchange and equity markets respectively. The realized volatility and the jump estimates for the three markets are presented in Figures 4.4, 4.5 and 4.6 as in Clements et al. (2015).

Figure 4.4 about here

Figure 4.5 about here

Figure 4.6 about here

The realized volatility exhibits similar patterns in all three zones with a peak around the middle of the sample, which corresponds to the Global financial crisis period. In the equity and foreign exchange markets the the Japanese bond market features lower volatility and jump activity. However, Fig. 4.4 shows that during recorded trading hours the Japanese bond market displays a volatility and jump structure deviating from the one in the other zones. Note that the realized volatility subplots of Fig. 4.6 are scaled higher then the bond and foreign exchange markets, indicating considerably higher realized volatility and jump estimates.

4.5.2 Estimation results

The Bayesian estimation procedure presented in Section 4.4.2, combined with each of the model specifications (4.5), (4.6), (4.11), (4.12) and (4.13), is applied to the bond, foreign exchange and equity market data multiplied by 1000. The initial values of the RWMH algorithm are chosen close to the ML-estimates[2], and the

[2]Note that the ML-estimates provided by Clements et al. (2015) are solely used as a "prior"-kind of information and do not restrict the estimation procedure in any way. In comparison to non-informative starting values the latter speeds up the convergence of the chain.

MCMC algorithm is run for 50K burn-in and a further 50K iterations. The DPM priors are set to $\mathbf{m}_0 = \mathbf{0}$, $s_0 = 0.1$, $d_0 = 5$, $\mathbf{W}_0 = \text{diag}(1/d_0)$, $a_0 = 4$, $b_0 = 4$.

Table 4.2 about here

Table 4.2 reports the coefficient posterior means and their corresponding 95% Bayesian credibility intervals for Model (4.5). The impact of the previous-day volatility (α_{ii}), is in a market-wise comparison, highest in the foreign exchange market and in total, highest in the US bond market. Observing the meteor shower and heat wave effects, the domination of the heat wave becomes apparent. The results for the bond and foreign exchange market are similar to those of Clements et al. (2015). However, in the US equity market, the effect of intra-day news originating from both Japan and Europe appear rather lower, compared to those of the previous day and also much lower then the coefficient estimates found in Clements et al. (2015). Observing the news from the previous day, European news evidently has a larger effect than US news on volatility in Japan.

Table 4.3 about here

The estimation results of the model in Eq. (4.6) subject to the restrictions $\mathbf{B} = \mathbf{0}$ and $\mathbf{G}$ diagonal (see Eq. (4.7)) are reported in Table 4.3. The intra-day effects are excluded from the model in order to isolate the spillover effect. In the foreign exchange and equity markets, a reduction of the heat wave effect in the sense of lower α_{ii} coefficients can be observed in comparison with the previous estimates, which may be a consequence of the inclusion of the conditional variance from the preceding zone (see coefficients $\tilde{\alpha}_{21}$ and $\tilde{\alpha}_{32}$). Once again, the heat wave effect appears to be the predominant one on the bond market. Contrary, on the foreign

exchange and equity market, a shift of the coefficients towards a meteor shower can be observed. In comparison to the estimates of Clements et al. (2015), a stronger effect is found of the US conditional variance of the previous day on the Japanese variance on the current day.

Table 4.4 about here

Furthermore, Table 4.4 summarizes the posterior information of model (4.11), in which the realized variance estimates are included in the volatility equation and their effects collected in the reparametrized matrix $\mathbf{B}$. The new matrix $\mathbf{B}$ (as in (4.14)) incorporates the effects coming from preceding regions: the news effect and the conditional variance effect. The new semiparametric model specification provides results that do not deviate substantially from those in Clements et al. (2015). Similar to the original study, the heat wave hypothesis prevails strongly over the alternative, as all β_{ij} coefficients are close to zero. However, a striking difference occurs in the realized variance effect from the US on the European equity market with a posterior mean of 0.0372, with the Bayesian interval (0.0164, 0.0587) compared to the Maximum-Likelihood estimate 0.2228 provided by Clements et al. (2015).[3] The second interesting finding is the strong evidence of a heat wave taking place in the bond and foreign exchange markets, as all the individual lagged conditional variance coefficients are higher in absolute terms in comparison to those in Table 4.3.

Table 4.5 about here

The model specification in Eq. (4.12) suggests using a decomposition of the realized volatility in a continuous and a jump component as explanatory variables

[3]Note that the usual non-negativity restrictions on the GARCH parameters are imposed and therefore, the confidence intervals cannot contain the 0. As usual, the numbers are rounded to the nearest 4 decimal digits.

in the volatility equation. Table 4.5 reports the posterior summary. The results on the bond and foreign exchange markets are fairly in line with those of Clements et al. (2015). That is, the jumps are not a major factor for intra-day volatility transitions. However, while Clements et al. (2015) do find conclusive results supporting the relevance of both the jump and the continuous component for the volatility transition from Europe to the US in the equity market, the posterior means of the β_{32}- and $\tilde{\beta}_{32}$ coefficients in Table 4.5 have very small values with tight intervals.

Table 4.6 about here

Finally, in Table 4.6 are presented the coefficient posterior means and credible intervals of the volatility model set-up defined in Eq. (4.13). An interesting asymmetry in the response to news from preceding zones can be observed. While the bond and the foreign exchange markets react more strongly to positive news, the equity market exhibits distinctly higher sensitivity towards negative semivariance. In particular, bad news from Japan has a greater effect on the volatility in Europe, and bad news from Europe substantially affects the variance in the US. By comparing these results with the previous study presented in Table 4.3, one could argue that the estimated intra-day effects of the conditional variance in the equity market are mostly due to bad news.

4.5.3 DP precision

The precision parameter of the DP c controls the number of mixture components. It can be seen that its value directly affects the mixture weights through the Beta prior used in the stick-breaking scheme ($\nu_j \sim \text{Beta}(c, 1)$). The precision parameter specifies two important limiting cases. First, if $c \to 0$, the mixture model collapses

to a one-component model, where, in the case of the stick-breaking representation presented in Section 4.3, the mixture weights become $\omega_1 = 1$, $\omega_{j,j\neq 1} = 0$ and the unconditional distribution of $\mathbf{R}_t$ is multivariate Gaussian. Second, as $c \to \infty$, the number of components also becomes infinite, which results in a multivariate Student-t distribution for $\mathbf{R}_t$. Note that c represents the concentration of the DP on the base G_0, but not the concentration on just one component.

In order to examine the behavior of the precision parameter, the transformed precision variable is introduced:

$$u \equiv \frac{c}{c+1}. \tag{4.29}$$

By imposing a uniform prior for the transformation u, one can restrict the two special cases for c in a $[0, 1]$ interval. Here, the value of $u = 0$ corresponds to the case where $c \to 0$ (the multivariate Gaussian case) and the case $u = 1$ corresponds to $c \to \infty$ (multivariate Student-t).[4]

Figure 4.7 about here

Figure 4.8 about here

Figure 4.9 about here

The histograms of the transformed precision u for all models in the bond, foreign exchange and equity markets are provided in Figures 4.7-4.9. The histogram plots the after-burn-in sample of the transformed variable, which is normalized as shown in Chapter 2.3.2. All histograms show consistently average values above 0, below

[4]For a derivation of the transformation u, using the Savage-Dickey Density Ratio, see Chapter 2, Section 2.3.2

1 and around 0.4, which generally supports the DPM specification with several distinct mixture components. Observing both the average transformed precisions and the average nonempty components, it can be concluded that, similar to the daily return data, the equity market intra-day returns deviate most significantly from Gaussianity. By contrast, the transformed precision u in foreign exchange market exhibits the lowest levels, usually below 0.4 on average.

4.6 Conclusion

The chapter provides an empirical investigation of the volatility transmission in global financial markets. The adopted semiparametric Bayesian estimation framework is a relatively new topic in the volatility literature and a flexible tool which outperforms the established parametric models in terms of model fit (see Jensen and Maheu, 2013). Moreover, the Bayesian approach integrates out parameter-estimation uncertainty, and the nonparametric portion of the model allows for a very flexible estimation of the density with respect to asymmetry and thick tails. Furthermore, a model specification test confirms the hypothesis of the infinite mixture model and allows for unbiased volatility dynamics.

The posterior results partially support the findings of Clements et al. (2015), but overall, call into question the evidence that volatility in all three markets should be modeled as a function of both the previous day local volatility and the volatility of neighboring regions on the same day. While stronger evidence for both effects is found for the equity market, the bond and foreign exchange markets are more severely affected by the heat wave.

Possible extensions of the study include computing the predictive distribution and performing density forecast evaluation. In addition, the study could be applied to more recent data, in order to involve more current financial market turbulence after the Global Financial Crisis, such as the European Sovereign Debt Crisis.

Tables and Figures

Table 4.1: Descriptive statistics multiplied by 100 for the bond, foreign exchange and equity markets in Japan, Europe and USA.

	Mean	Median	Variance	Skewness	Kurtosis	Sample correlation		
Bond market								
Japan	-0.0005	0.0000	0.0044	-0.5978	14.0123	1.0000		
Europe	0.0018	0.0000	0.0063	-0.1087	5.2929	0.0244	1.0000	
USA	0.0035	0.0049	0.0121	-0.1329	4.8356	0.0244	0.0123	1.0000
Foreign exchange market								
Japan	0.0032	0.0032	0.0102	-0.0142	5.6396	1.0000		
Europe	-0.0072	-0.0077	0.0244	-0.2930	5.8742	-0.0502	1.0000	
USA	0.0047	0.0072	0.0244	0.0490	5.3783	-0.0118	0.0371	1.0000
Equity market								
Japan	0.0065	0.0070	0.0251	-0.0953	37.6800	1.0000		
Europe	0.0061	0.0084	0.0434	0.0556	10.2371	-0.1121	1.0000	
USA	-0.0066	0.0149	0.1554	-0.6466	12.7935	0.1149	0.0575	1.0000

Table 4.2: Posterior means and 95% credibility intervals (CI) of the parameters, estimated according the model defined in Eq. (4.5).

	BD		FX		EQ	
	Mean	CI	Mean	CI	Mean	CI
κ_1	0.0016	(0.0001 , 0.0045)	0.0031	(0.0001 , 0.0103)	0.0048	(0.0002 , 0.0166)
κ_2	0.0072	(0.0004 , 0.0202)	0.0261	(0.0055 , 0.0638)	0.0046	(0.0002 , 0.0133)
κ_3	0.0065	(0.0007 , 0.0178)	0.0193	(0.0022 , 0.0834)	0.1984	(0.0875 , 0.3235)
α_{11}	0.6240	(0.5529 , 0.6859)	0.9111	(0.8667 , 0.9401)	0.8115	(0.7675 , 0.8559)
α_{22}	0.8825	(0.8360 , 0.9199)	0.9277	(0.8946 , 0.9564)	0.8677	(0.8292 , 0.9128)
α_{33}	0.9428	(0.9123 , 0.9609)	0.8904	(0.8291 , 0.9274)	0.7775	(0.7166 , 0.8264)
β_{21}	0.0081	(0.0004 , 0.0208)	0.0365	(0.0085 , 0.0858)	0.0701	(0.0143 , 0.1379)
β_{31}	0.0109	(0.0008 , 0.0320)	0.0589	(0.0191 , 0.1338)	0.0916	(0.0392 , 0.1785)
β_{32}	0.0149	(0.0009 , 0.0522)	0.0130	(0.0008 , 0.0472)	0.0722	(0.0206 , 0.1705)
γ_{11}	0.1089	(0.0709 , 0.1537)	0.0544	(0.0228 , 0.0843)	0.0873	(0.0495 , 0.1271)
γ_{12}	0.0438	(0.0245 , 0.0642)	0.0075	(0.0009 , 0.0188)	0.0580	(0.0279 , 0.0924)
γ_{13}	0.0202	(0.0116 , 0.0306)	0.0172	(0.0061 , 0.0423)	0.0079	(0.0033 , 0.0133)
γ_{22}	0.0995	(0.0698 , 0.1438)	0.0241	(0.0061 , 0.0573)	0.1038	(0.0582 , 0.1548)
γ_{23}	0.0174	(0.0049 , 0.0304)	0.0230	(0.0047 , 0.0538)	0.0106	(0.0041 , 0.0175)
γ_{33}	0.0466	(0.0295 , 0.0759)	0.0200	(0.0055 , 0.0529)	0.0535	(0.0265 , 0.0992)
c	0.6419	(0.2151 , 1.2840)	0.5601	(0.1930 , 1.1175)	0.7278	(0.2478 , 1.5109)
k^*	4.2637		3.3168		4.7235	

Note: c is the DP precision parameter and k^* is the average number of nonempty mixture components.

Table 4.3: Posterior means and 95% credibility intervals (CI) of the parameters estimated according the model defined in Eq. (4.6).

	BD		FX		EQ	
	Mean	CI	Mean	CI	Mean	CI
κ_1	0.0021	(0.0001 , 0.0075)	0.0062	(0.0001 , 0.0236)	0.0024	(0.0001 , 0.0085)
κ_2	0.0040	(0.0004 , 0.0098)	0.0152	(0.0005 , 0.0436)	0.0059	(0.0003 , 0.0169)
κ_3	0.0041	(0.0002 , 0.0139)	0.0050	(0.0001 , 0.0186)	0.2040	(0.1059 , 0.3162)
α_{11}	0.5395	(0.4055 , 0.6437)	0.4872	(0.3460 , 0.6486)	0.5142	(0.3654 , 0.6784)
α_{13}	0.0458	(0.0221 , 0.0931)	0.3257	(0.1131 , 0.5919)	0.0912	(0.0368 , 0.1806)
α_{22}	0.9054	(0.8650 , 0.9331)	0.7418	(0.5466 , 0.8653)	0.7730	(0.6292 , 0.8600)
α_{33}	0.9345	(0.8990 , 0.9589)	0.6817	(0.5064 , 0.8417)	0.6683	(0.5874 , 0.7644)
$\tilde{\alpha}_{21}$	0.0343	(0.0025 , 0.0729)	0.5571	(0.1546 , 0.9217)	0.2582	(0.0984 , 0.4979)
$\tilde{\alpha}_{32}$	0.0308	(0.0078 , 0.0611)	0.1571	(0.0495 , 0.3172)	0.2681	(0.1362 , 0.4296)
γ_{11}	0.1365	(0.0736 , 0.2290)	0.0967	(0.0424 , 0.1925)	0.1951	(0.0996 , 0.3131)
γ_{22}	0.0832	(0.0569 , 0.1214)	0.0267	(0.0076 , 0.0590)	0.1436	(0.0766 , 0.2155)
γ_{33}	0.0455	(0.0244 , 0.0704)	0.0298	(0.0125 , 0.0517)	0.0693	(0.0381 , 0.1099)
c	0.6264	(0.2202 , 1.2538)	0.5951	(0.2064 , 1.2012)	0.7682	(0.2667 , 1.5604)
k^*	4.2929		3.6426		5.8154	

Note: c is the DP precision parameter and k^* is the average number of nonempty mixture components.

Table 4.4: Posterior means and 95% credibility intervals (CI) of the parameters estimated according the model defined in Eq. (4.11).

	BD		FX		EQ	
	Mean	CI	Mean	CI	Mean	CI
κ_1	0.0034	(0.0002 , 0.0085)	0.0052	(0.0001 , 0.0143)	0.0074	(0.0003 , 0.0170)
κ_2	0.0041	(0.0005 , 0.0095)	0.0238	(0.0050 , 0.0648)	0.0185	(0.0007 , 0.0553)
κ_3	0.0068	(0.0014 , 0.0148)	0.0315	(0.0027 , 0.0722)	0.1981	(0.1151 , 0.2720)
α_{11}	0.5687	(0.5481 , 0.5906)	0.8453	(0.8155 , 0.8794)	0.5515	(0.4533 , 0.7314)
α_{22}	0.9076	(0.8983 , 0.9175)	0.9451	(0.9197 , 0.9621)	0.5902	(0.4818 , 0.7136)
α_{33}	0.9384	(0.9238 , 0.9604)	0.7991	(0.7034 , 0.8719)	0.7161	(0.6589 , 0.7861)
β_{13}	0.0039	(0.0025 , 0.0056)	0.0058	(0.0026 , 0.0095)	0.0048	(0.0025 , 0.0080)
β_{21}	0.0014	(0.0001 , 0.0034)	0.0103	(0.0037 , 0.0231)	0.0899	(0.0493 , 0.1459)
β_{32}	0.0037	(0.0008 , 0.0079)	0.0221	(0.0102 , 0.0393)	0.0372	(0.0164 , 0.0587)
γ_{11}	0.0801	(0.0693 , 0.0939)	0.0677	(0.0370 , 0.0961)	0.1266	(0.0757 , 0.1857)
γ_{22}	0.0826	(0.0711 , 0.0914)	0.0156	(0.0026 , 0.0403)	0.0664	(0.0266 , 0.1203)
γ_{33}	0.0523	(0.0340 , 0.0664)	0.0125	(0.0005 , 0.0380)	0.0194	(0.0051 , 0.0367)
c	0.6248	(0.2225 , 1.2342)	0.5276	(0.1743 , 1.0711)	0.7566	(0.2712 , 1.4963)
k^*	4.1751		2.9843		5.1271	

Note: c is the DP precision parameter and k^* is the average number of nonempty mixture components.

Table 4.5: Posterior means and 95% credibility intervals (CI) of the parameters estimated according the model defined in Eq. (4.12).

	BD		FX		EQ	
	Mean	CI	Mean	CI	Mean	CI
κ_1	0.0034	(0.0001 , 0.0091)	0.0031	(0.0001 , 0.0102)	0.0046	(0.0001 , 0.0194)
κ_2	0.0053	(0.0009 , 0.0120)	0.0172	(0.0058 , 0.0342)	0.0270	(0.0018 , 0.0639)
κ_3	0.0069	(0.0006 , 0.0185)	0.0194	(0.0025 , 0.0400)	0.2847	(0.1133 , 0.4708)
α_{11}	0.6236	(0.6078 , 0.6405)	0.8740	(0.8347 , 0.9053)	0.5879	(0.5170 , 0.6564)
α_{22}	0.9012	(0.8759 , 0.9233)	0.9408	(0.9131 , 0.9621)	0.5548	(0.4011 , 0.7492)
α_{33}	0.9348	(0.9134 , 0.9511)	0.7903	(0.7326 , 0.8898)	0.6571	(0.5382 , 0.7568)
β_{13}	0.0067	(0.0042 , 0.0098)	0.0042	(0.0021 , 0.0069)	0.0054	(0.0032 , 0.0084)
β_{21}	0.0025	(0.0003 , 0.0061)	0.0069	(0.0035 , 0.0125)	0.1172	(0.0761 , 0.1738)
β_{32}	0.0071	(0.0022 , 0.0137)	0.0172	(0.0083 , 0.0281)	0.0610	(0.0230 , 0.1045)
$\tilde{\beta_{13}}$	0.0025	(0.0001 , 0.0080)	0.0051	(0.0003 , 0.0135)	0.0027	(0.0001 , 0.0078)
$\tilde{\beta_{21}}$	0.0012	(0.0001 , 0.0040)	0.0188	(0.0052 , 0.0369)	0.0993	(0.0419 , 0.1841)
$\tilde{\beta_{32}}$	0.0024	(0.0001 , 0.0113)	0.0143	(0.0010 , 0.0356)	0.0223	(0.0008 , 0.0681)
γ_{11}	0.1366	(0.0997 , 0.2047)	0.0427	(0.0237 , 0.0661)	0.1574	(0.0888 , 0.2586)
γ_{22}	0.0907	(0.0664 , 0.1172)	0.0123	(0.0036 , 0.0249)	0.0763	(0.0247 , 0.1485)
γ_{33}	0.0463	(0.0253 , 0.0784)	0.0067	(0.0003 , 0.0170)	0.0214	(0.0024 , 0.0572)
c	0.6115	(0.2193 , 1.2178)	0.5224	(0.1517 , 1.0947)	0.7777	(0.2774 , 1.5342)
k^*	3.8925		3.4448		5.3499	

Note: c is the DP precision parameter and k^* is the average number of nonempty mixture components.

Table 4.6: Posterior means and 95% credibility intervals (CI) of the parameters estimated according the model defined in Eq. (4.13).

	BD		FX		EQ	
	Mean	CI	Mean	CI	Mean	CI
κ_1	0.0069	(0.0004 , 0.0140)	0.0058	(0.0002 , 0.0200)	0.0072	(0.0001 , 0.0315)
κ_2	0.0030	(0.0004 , 0.0068)	0.0193	(0.0050 , 0.0397)	0.0338	(0.0019 , 0.0864)
κ_3	0.0072	(0.0008 , 0.0175)	0.0299	(0.0012 , 0.0956)	0.6846	(0.2541 , 0.9480)
α_{11}	0.5952	(0.5535 , 0.6413)	0.8656	(0.8206 , 0.9071)	0.5570	(0.4440 , 0.6623)
α_{22}	0.9153	(0.8845 , 0.9345)	0.9375	(0.8924 , 0.9608)	0.5890	(0.4393 , 0.7194)
α_{33}	0.9384	(0.9101 , 0.9586)	0.8171	(0.7260 , 0.8906)	0.4908	(0.2047 , 0.6748)
β_{13}	0.0129	(0.0066 , 0.0198)	0.0085	(0.0022 , 0.0172)	0.0007	(0.0000 , 0.0021)
β_{21}	0.0017	(0.0001 , 0.0046)	0.0126	(0.0019 , 0.0287)	0.0392	(0.0017 , 0.1217)
β_{32}	0.0055	(0.0004 , 0.0158)	0.0214	(0.0047 , 0.0455)	0.0599	(0.0052 , 0.1344)
$\tilde{\beta_{13}}$	0.0069	(0.0019 , 0.0122)	0.0047	(0.0004 , 0.0116)	0.0119	(0.0066 , 0.0205)
$\tilde{\beta_{21}}$	0.0016	(0.0001 , 0.0046)	0.0083	(0.0006 , 0.0255)	0.2224	(0.1066 , 0.3473)
$\tilde{\beta_{32}}$	0.0031	(0.0001 , 0.0098)	0.0297	(0.0085 , 0.0725)	0.2357	(0.0761 , 0.3772)
γ_{11}	0.2527	(0.1928 , 0.3032)	0.0830	(0.0425 , 0.1273)	0.1711	(0.0783 , 0.3062)
γ_{22}	0.0761	(0.0569 , 0.1074)	0.0176	(0.0034 , 0.0371)	0.0794	(0.0243 , 0.1549)
γ_{33}	0.0496	(0.0282 , 0.0768)	0.0147	(0.0009 , 0.0383)	0.0375	(0.0033 , 0.0837)
c	0.6280	(0.2165 , 1.2620)	0.5435	(0.1891 , 1.0911)	0.7527	(0.2564 , 1.5120)
k^*	3.8727		3.1701		5.0640	

Note: c is the DP precision parameter and k^* is the average number of nonempty mixture components.

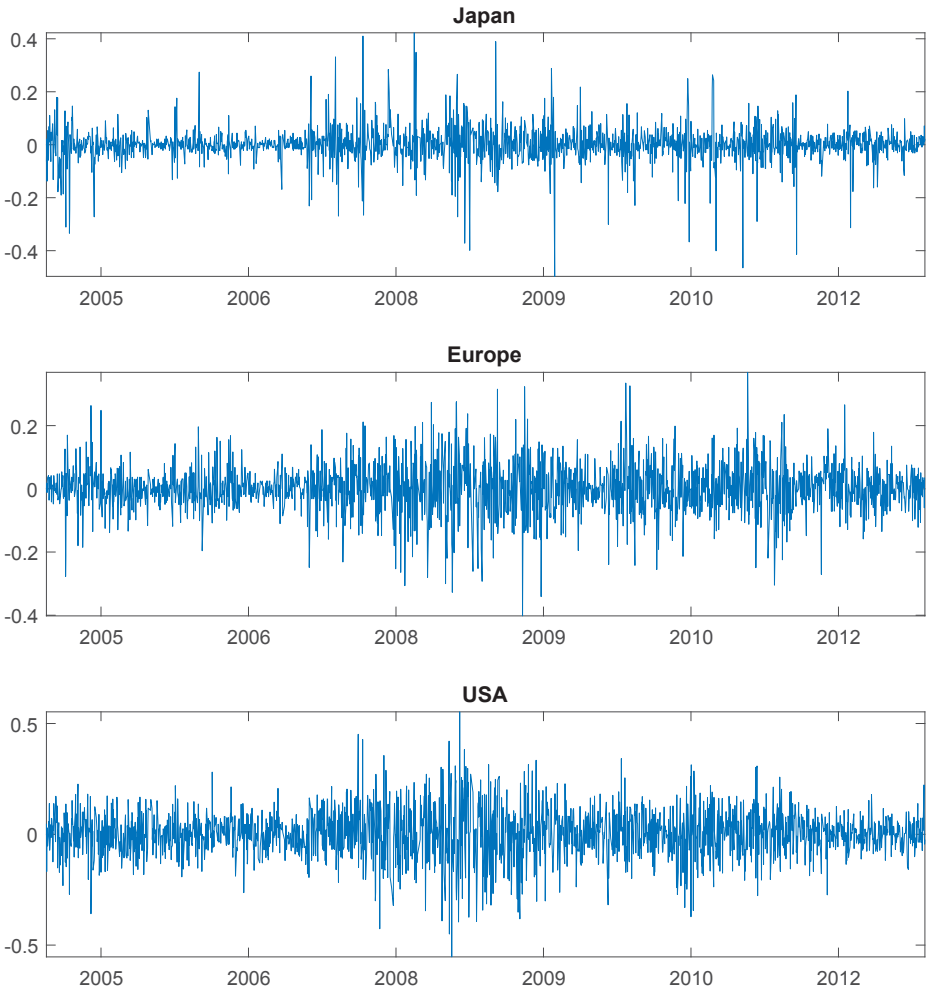


Figure 4.1: Intra-day returns on the bond market multiplied by 100.

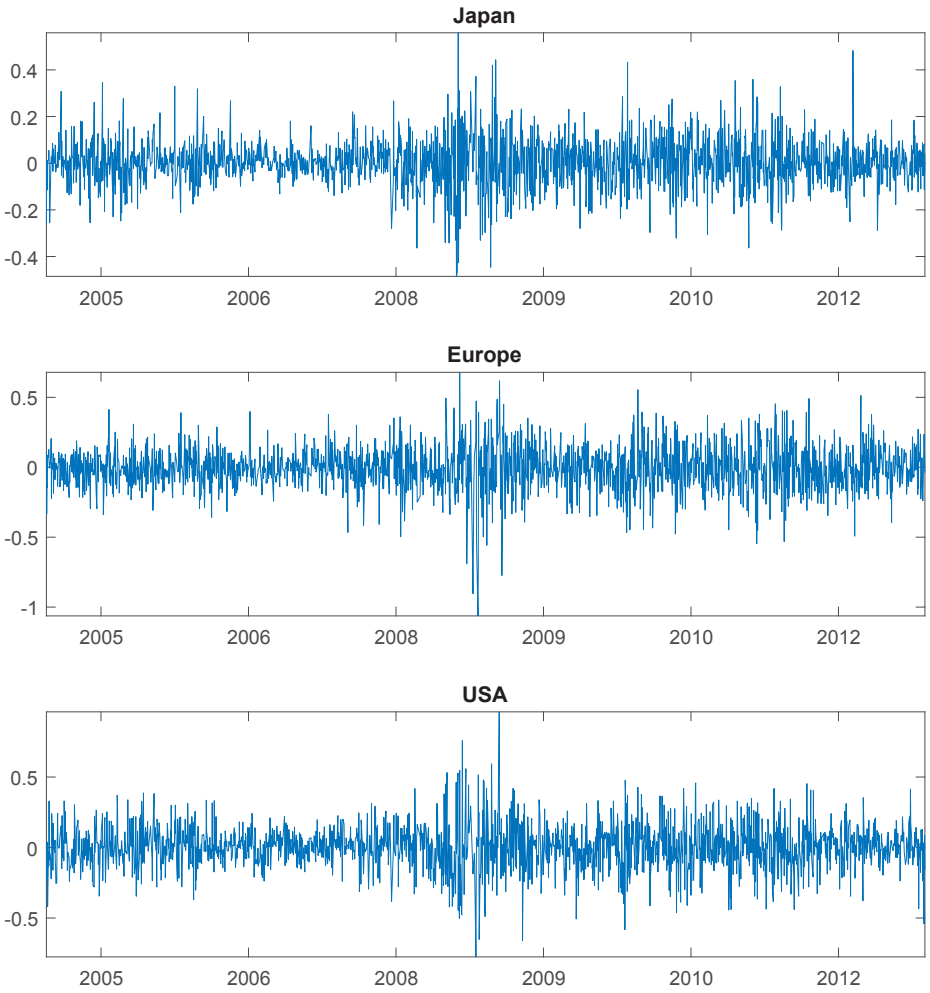


Figure 4.2: Intra-day returns on the foreign exchange market multiplied by 100.

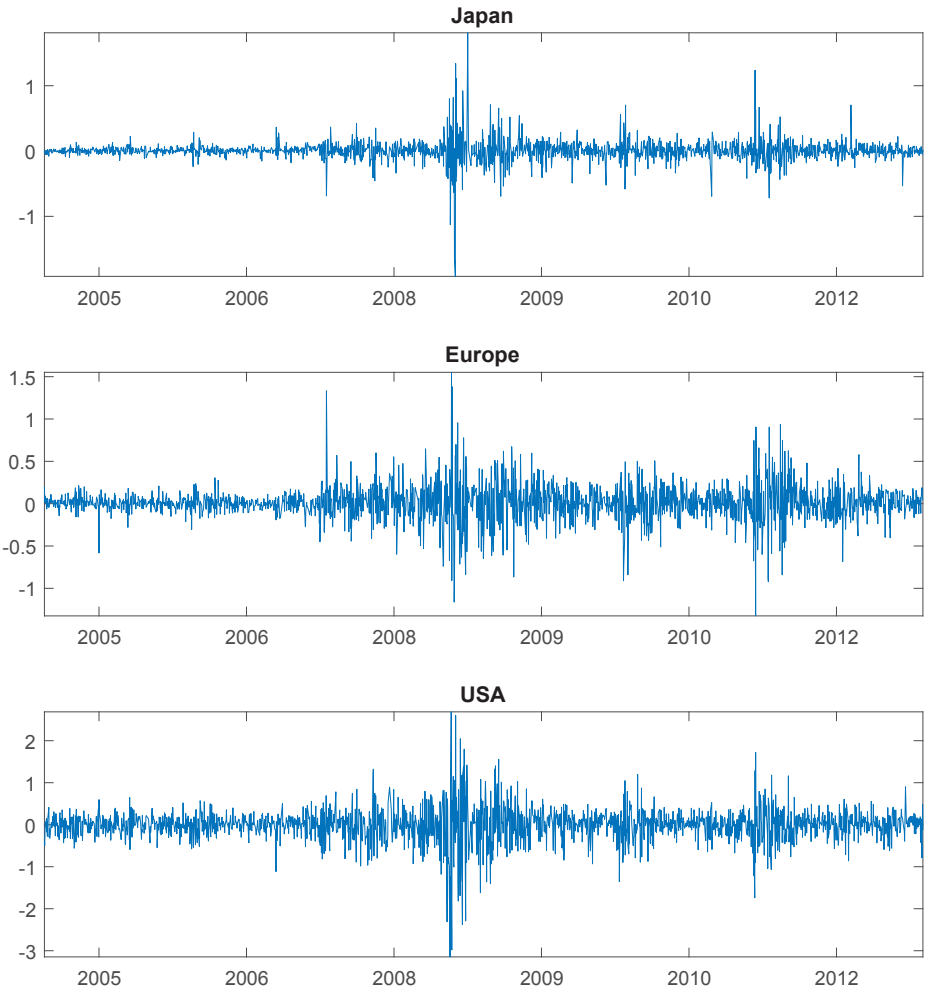


Figure 4.3: Intra-day returns on the equity market multiplied by 100.

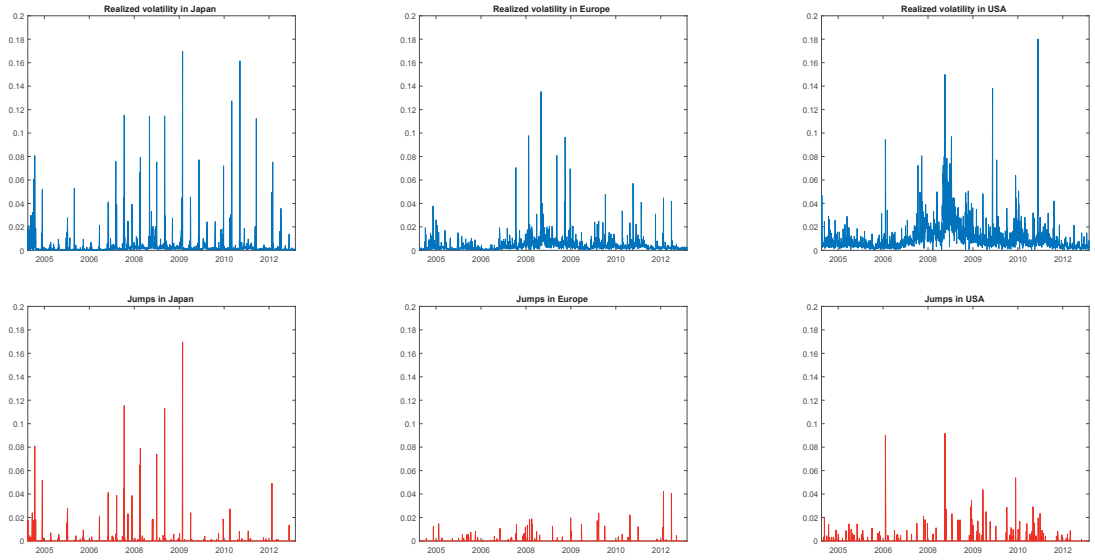


Figure 4.4: Realized volatility and jump estimates (multiplied by 1000) for the bond market for the Japan, Europe and USA trading zones.

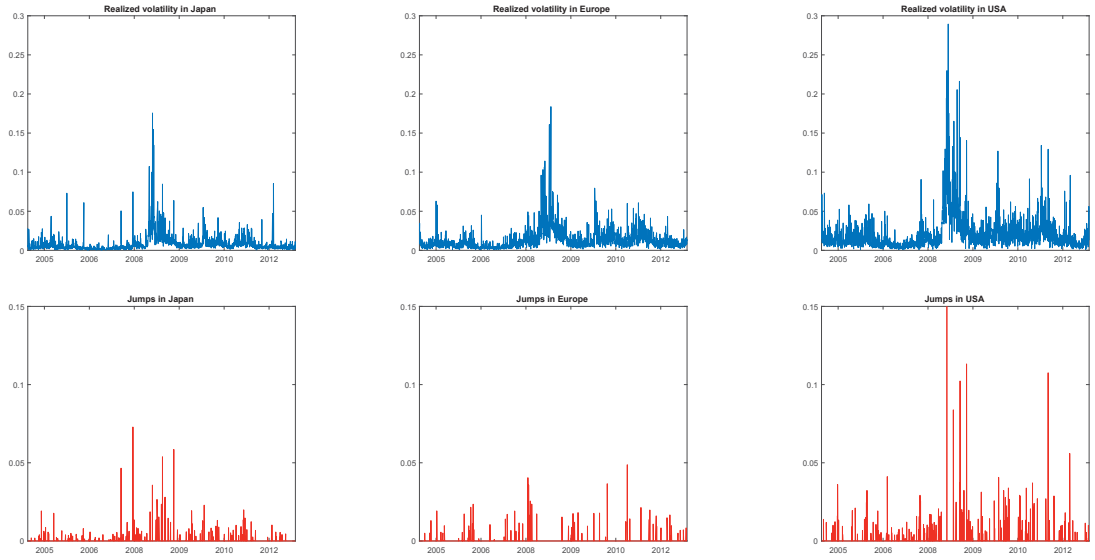


Figure 4.5: Realized volatility and jump estimates (multiplied by 1000) for the foreign exchange market for the Japan, Europe and USA trading zones.

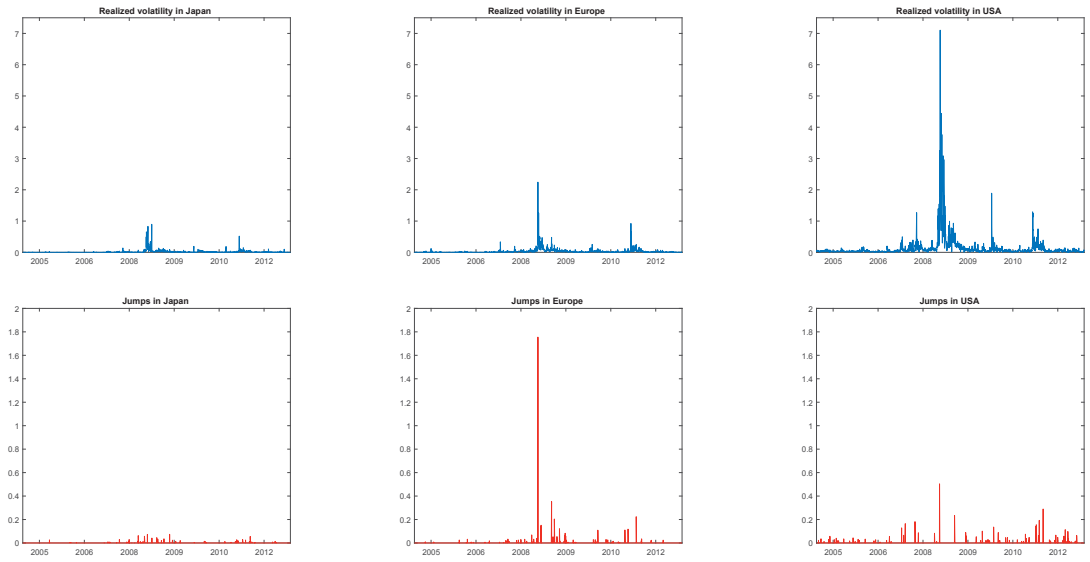


Figure 4.6: Realized volatility and jump estimates (multiplied by 1000) for the equity market for the Japan, Europe and USA trading zones.

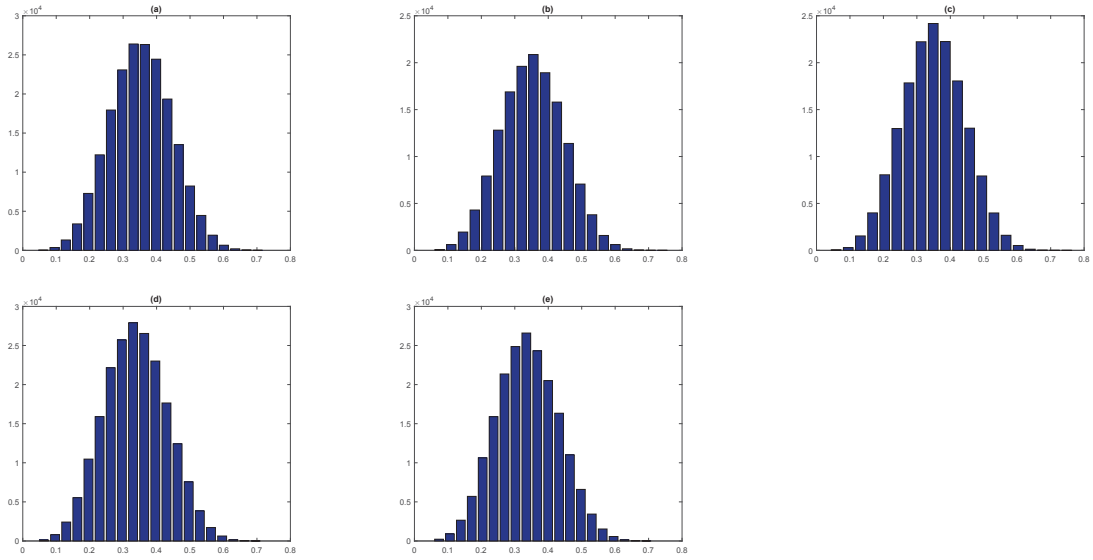


Figure 4.7: Posterior density of the transformed and normalized variable u in the bond market, conditional on models (4.5), (4.6), (4.11), (4.12), (4.13) in panels (a)-(e) respectively.

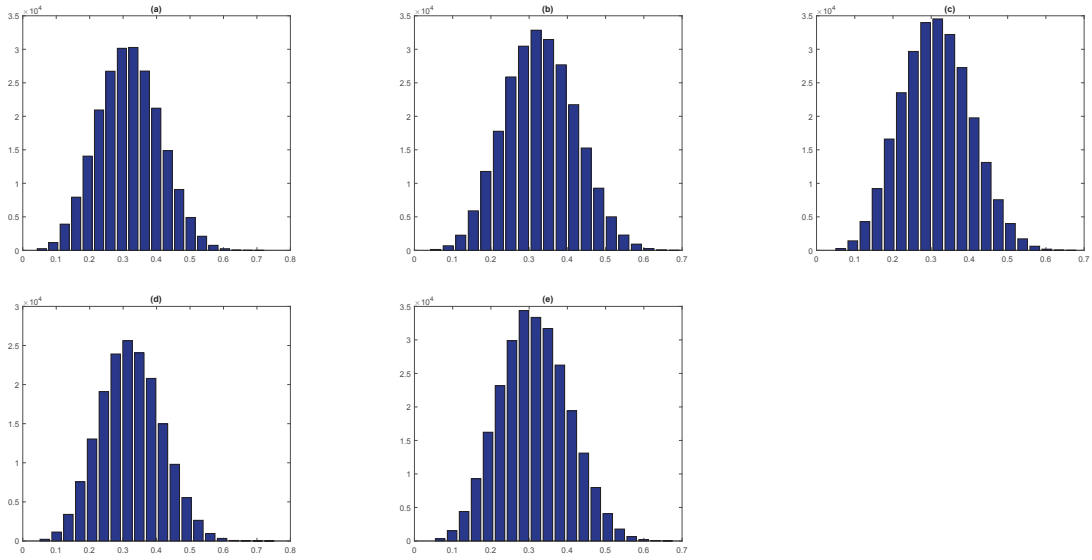


Figure 4.8: Posterior density of the transformed and normalized variable u in the foreign exchange market, conditional on models (4.5), (4.6), (4.11), (4.12), (4.13) in panels (a)-(e) respectively.

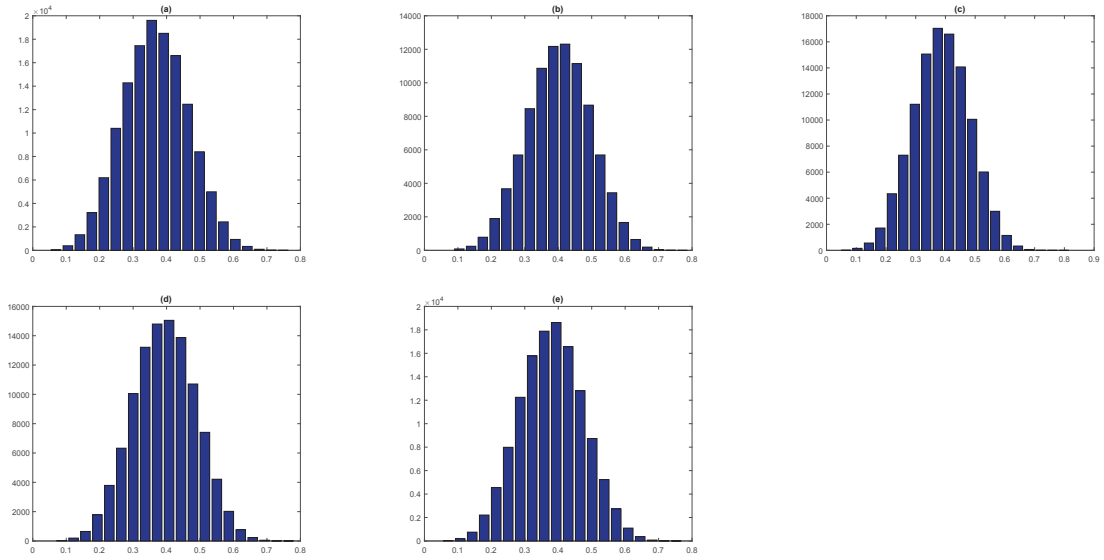


Figure 4.9: Posterior density of the transformed and normalized variable u in the equity market, conditional on models (4.5), (4.6), (4.11), (4.12), (4.13) in panels (a)-(e) respectively.

Chapter 5

Summary and outlook

This thesis provides several computational and modeling extensions to the growing literature on Bayesian nonparametric methods applied to financial econometric models. The thesis is comprised of three distinct essays. First, we present a solution to the challenging task of finding an efficient and adaptive sampling strategy, when dealing with non-linear and/or nonparametric model set-ups. Second, we investigate international financial volatility co-movements by means of a multivariate stochastic volatility framework combined with a highly flexible nonparametric error term. And finally, we study the global volatility transmission by applying a Bayesian nonparametric MGARCH model, proven by a number of studies to be a more suitable tool in comparison to the usual benchmark alternatives. Accordingly, the thesis contributes to the field in various ways, by (i) solving an algorithmic issue, (ii) designing a new model structure, and (iii) conducting an empirical study for the global bond, foreign exchange and equity markets. With respect to aspect (ii), the new nonparametric model structure is accompanied with an appropriate sampling strategy and provides direct implications on international financial market behavior.

In Chapter 2, we propose a new generic and flexible algorithm for estimating non-linear nonparametric state-space models and provide multiple examples, mostly

emphasizing applicability in the context of stochastic volatility estimation in finance. We demonstrate the advantages of the new sampling scheme by means of two examples and by focusing further on the semiparametric stochastic volatility model of Jensen and Maheu (2014). Evidently, the new approach is applicable to models capable of capturing the properties of financial time series and is also readily extendable to more complex econometric models due to its flexibility and generality.

Chapter 3 proposes a new semiparametric Cholesky multivariate stochastic volatility model, in which the dynamics of the multivariate conditional volatility are modeled parametrically, while the error term is modeled as an infinite mixture of normal distributions based on the nonparametric Dirichlet process mixture prior. We establish a fully-fledged Bayesian estimation algorithm and investigate volatility co-movements among five international stock market indices. The empirical analysis has two major findings. (1) Although expecting the markets to move together during turbulent periods, we do not find evidence of this, especially when observing the Asian markets. (2) We find positive co-movements between the US and European markets, but nearly uncorrelated US and Japanese markets.

Our multivariate model has a number of advantages compared to those in the existing literature. First, the combination of the flexible stochastic volatility model with the highly flexible nonparametric prior for the unconditional distribution provides an attractive tool for risk management. Second, the Cholesky structure of the covariance matrix facilitates high-dimensional model estimation without restricting model flexibility. As a result, our Cholesky DPM-MSV framework could also be useful for dealing with high-dimensional portfolio-decision problems.

In Chapter 4, we address GARCH-type specifications, another important class of volatility models. In particular, we are interested in different multivariate GARCH

parameterizations, in order to study the volatility transmission in global bond, foreign exchange and equity markets. In order to establish a suitable modeling framework, which considers the non-Gaussianity in financial returns, we adopt a Bayesian nonparametric MGARCH model based on a DPM prior. This allows us to reexamine the heat wave and meteor shower volatility transmission patterns without depending on a particular distributional assumption. We find that a suitable transformation of the DPM precision parameter, based on the Savage-Dickey density ratio, supports the hypothesis of an infinite mixture model for all three markets.

The empirical results suggest that the pattern of volatility interaction is a combination of effects related to both, (1) volatility in the same region, and (2) volatility in the timely preceding region. In contrast to the benchmark study by Clements et al. (2015), we do not find strong evidence in favor of meteor shower in the equity market. Overall, our results indicate a rather weak meteor shower effect, combined with a distinct heat wave pattern.

Two directions for future research can be identified from these results and implications. First, when dealing with (univariate and multivariate) stochastic volatility models, a natural extension of the framework could be the inclusion of leverage effects, which supports modeling negative skewness in the return distribution. Second, an interesting aspect in the SV literature could be the inclusion of realized volatility measures in an additional state-space equation, in order to improve estimation accuracy.

Finally, extensions to the empirical study from Chapter 4 are conceivable. Recall that the observed trading period in Europe ends at 12:30 pm (GMT) due to the overlap with the US stock market trading. For example, a comparison of the results when subsuming the trading overlap into the European market instead of the US

market (see the structure of the global trading day) could be worth investigating. Here, one could argue that the afternoon-events occurring in Europe might be more relevant to volatility transmission than the US pre-trading period. Including density forecasts to the study is another possible avenue for future research.

References

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169*(1), 75 – 93. Recent Advances in Panel Data, Nonlinear and Nonparametric Models: A Festschrift in Honor of Peter C.B. Phillips.

Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: A review. *Econometric Reviews 25*(2-3), 145–175.

Ausin, M. C., P. Galeano, and P. Ghosh (2014). A semiparametric Bayesian approach to the analysis of financial time series with applications to value at risk estimation. *European Journal of Operational Research 232*(2), 350–358.

Bauwens, L., S. Laurent, and J. V. Rombouts (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics 21*(1), 79–109.

Blackwell, D. and J. B. MacQueen (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, 353–355.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Broto, C. and E. Ruiz (2004). Estimation methods for stochastic volatility models: A survey. *Journal of Economic Surveys 18*(5), 613–649.

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

Chang, C.-L., Y. Li, and M. McAleer (2015). Volatility spillovers between energy and agricultural markets: A critical appraisal of theory and practice. Tinbergen Institute Discussion Paper Series No. TI 15-077/III, Tinbergen Institute.

Chib, S., Y. Omori, and M. Asai (2009). Multivariate stochastic volatility. In *Handbook of Financial Time Series*, pp. 365–400. Springer.

Clements, A. E., A. S. Hurn, and V. V. Volkov (2015). Volatility transmission in global financial markets. *Journal of Empirical Finance 32*, 3–18.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance 1*, 223–236.

Corradi, V., W. Distaso, and M. Fernandes (2012). International market links and volatility transmission. *Journal of Econometrics 170*(1), 117–141.

Delatola, E.-I. and J. E. Griffin (2011). Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Analysis 6*(4), 901–926.

Delatola, E.-I. and J. E. Griffin (2013). A Bayesian semiparametric model for volatility with a leverage effect. *Computational Statistics & Data Analysis 60*(1), 97–110.

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *Annals of Mathematical Statistics 42*(1), 204–223.

Doucet, A., S. Godsill, and C. Andrieu (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing 10*(3), 197–208.

Doucet, A., M. K. Pitt, G. Deligiannidis, and R. Kohn (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika 102*, 295–313.

Dungey, M., L. Fakhrutdinova, and C. Goodhart (2009). After-hours trading in equity futures markets. *Journal of Futures Markets 29*(2), 114–136.

Ehrmann, M., M. Fratzscher, and R. Rigobon (2011). Stocks, bonds, money markets and exchange rates: measuring international financial transmission. *Journal of Applied Econometrics 26*(6), 948–974.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F., T. Ito, and W.-l. Lin (1990). Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica 58 (3)*, 525–542.

Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association 90*(430), 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics 1*(2), 209–230.

Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2005). Estimating dynamic equilibrium economies: Linear versus nonlinear likelihood. *Journal of Applied Econometrics 20*(7), 891–910.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016). Solution and estimation methods for DSGE models. *Handbook of Macroeconomics 2*, 527–724.

Flury, T. and N. Shephard (2011). Bayesian inference based only on simulated likelihood: Particle filter analysis of dynamic economic models. *Econometric Theory 27*(5), 933–956.

Greenberg, E. (2008). *Introduction to Bayesian Econometrics*. Cambridge University Press.

Hammoudeh, S., T. Liu, C.-L. Chang, and M. McAleer (2013). Risk spillovers in oil-related CDS, stock and credit markets. *Energy Economics 36*, 526–535.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate Stochastic Variance Models. *Review of Economic Studies 61*(2), 247–264.

Herbst, E. P. and F. Schorfheide (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press.

Jacquier, E., N. G. Polson, and P. E. Rossi (2002). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics 20*(1), 69–87.

Jacquier, E., N. G. Polson, and P. E. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics 122*(1), 185–212.

Jensen, M. J. and J. M. Maheu (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics 157*(2), 306–316.

Jensen, M. J. and J. M. Maheu (2013). Bayesian semiparametric multivariate GARCH modeling. *Journal of Econometrics 176*(1), 3–17.

Jensen, M. J. and J. M. Maheu (2014). Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture. *Journal of Econometrics 178*, 523–538.

Kalli, M., J. E. Griffin, and S. G. Walker (2011). Slice sampling mixture models. *Statistics and Computing 21*(1), 93–105.

Kalli, M., S. G. Walker, and P. Damien (2013). Modeling the conditional distribution of daily stock index returns: An alternative Bayesian semiparametric model. *Journal of Business & Economic Statistics 31*(4), 371–383.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*(3), 361–393.

Liu, L. Y., A. J. Patton, and K. Sheppard (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics 187*(1), 293–311.

Lopes, H., R. McCulloch, and R. Tsay (2012). Cholesky stochastic volatility models for high-dimensional time series. *Discussion papers*.

Maheu, J. M. and Q. Yang (2016). An infinite hidden Markov model for short-term interest rates. *Journal of Empirical Finance 38*, 202–220.

Nakajima, J. (2017). Bayesian analysis of multivariate stochastic volatility with skew return distribution. *Econometric Reviews 36*(5), 546–562.

Nakajima, J. and T. Watanabe (2011). Bayesian analysis of time-varying parameter vector autoregressive model with the ordering of variables for the Japanese economy and monetary policy. Global COE Hi-Stat Discussion Paper Series gd11-196, Institute of Economic Research, Hitotsubashi University.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics 9*(2), 249–265.

Papaspiliopoulos, O. (2008). A note on posterior sampling from Dirichlet mixture models. Technical report, University of Warwick. Centre for Research in Statistical Methodology.

Papaspiliopoulos, O. and G. O. Roberts (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika 95*(1), 169–186.

Pitt, M. K., R. Silva, P. Giordani, and R. Kohn (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics 171*, 134–151.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies 72*(3), 821–852.

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*, Volume 3. Cambridge University Press.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica 4*, 639–650.

Shirota, S., Y. Omori, H. F. Lopes, and H. Piao (2016). Cholesky realized stochastic volatility model. *Econometrics and Statistics 3*, 34–59.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices 1961-75. *In Anderson, O.D., ed. Time Series Analysis: Theory and Practice 1*, 203–226.

Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester: John Wiley.

Teh, Y. W. (2011). Dirichlet process. In *Encyclopedia of Machine Learning*, pp. 280–287. Springer.

Virbickaitė, A., M. C. Ausín, and P. Galeano (2015). Bayesian inference methods for univariate and multivariate GARCH models: A survey. *Journal of Economic Surveys 29*(1), 76–96.

Virbickaitė, A., M. C. Ausín, and P. Galeano (2016). A Bayesian non-parametric approach to asymmetric dynamic conditional correlation model with application to portfolio selection. *Computational Statistics & Data Analysis 100*, 814–829.

Virbickaitė, A., H. F. Lopes, M. C. Ausín, and P. Galeano (2014). Particle learning for Bayesian non-parametric Markov switching stochastic volatility model. *Discussion papers*.

Volkov, V. V. (2015). *Volatility Transmission in Global Finanial Markets*. Ph. D. thesis, Shool of Eonomis and Finane, Queensland University of Tehnology.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation 36*(1), 45–54.

Appendix A

A.1 Observable state & nonparametric distribution

Consider the model

$$y_t = \epsilon_t, \qquad \epsilon_t \sim \mathcal{G},$$

where $\mathcal{G}$ is a unknown distribution. In this case, we have no latent states and a nonparametric model, and use the present model to illustrate the nonparametric part of the sampler. We simulate a sample of $T = 50$ observations from the following mixture of normals

$$y_t \overset{iid}{\sim} \begin{cases} \mathcal{N}(-20, 1) \text{ with prob. } 0.2 \\ \mathcal{N}(0, 5) \text{ with prob. } 0.5 \\ \mathcal{N}(5, 1) \text{ with prob. } 0.3 \end{cases}.$$

We use Steps A and D, as no filter/smoother is required. Furthermore, the likelihood (conditional on the table assignments) is given in closed form. We scale the random walk proposals to achieve an acceptance ratio of roughly 0.33. We choose non-informative priors, i.e. the base distribution $G_0(\cdot)$ is $\mathcal{N}(0, 3)$ - $\Gamma(1, 1)$, while the concentration parameter for the CRP α has the Gamma prior $\Gamma(1, 1)$. We run the algorithm for 20000 iterations and drop the first 5000 from the calculations. Figure A.1 shows the posterior of the concentration parameter α and the resulting predictive density. The first two panels show the trace plot of α (a), and the Bayes

factors (panel (b)). Lastly, panel (c) compares the data histogram to the posterior predictive density obtained from the DPM. It is evident that the infinite mixture succeeds in identifying the distinct components and provides a flexible forecast, even given the small sample size.

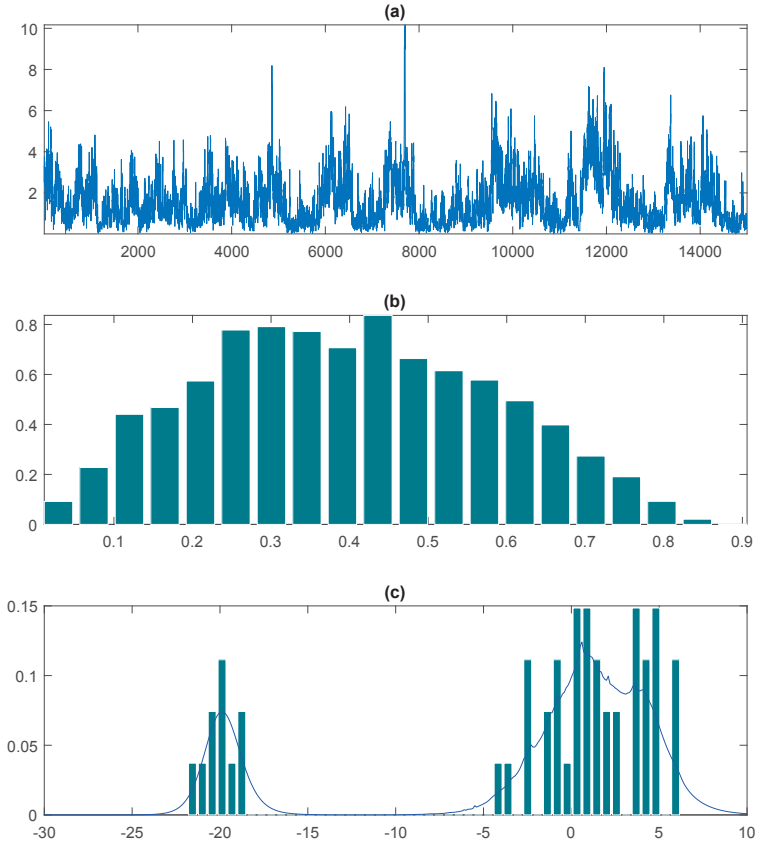


Figure A.1: Density estimation based on Chinese restaurant process (Section 2.2.3): (a) trace plot of α, (b) Bayes factors and (c) data histogram and posterior predictive density (blue line).

A.2 Latent state & parametric distribution

The second example is the stochastic volatility model with parametric error terms.

$$y_t = \exp(s_t/2)\epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma_y^2),$$
$$s_t = \rho s_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2).$$

Our sampling algorithm proceeds as in Flury and Shephard (2011). We use Step C and the particle-filter approximation of the likelihood to sample the parameter set $\boldsymbol{\theta} = \{\rho, \sigma_\eta, \sigma_y\}$. We view this second example as a test of the particle filter's capability to deal with the latent state variable. We set the parameters to $\rho = 0.95$, $\sigma_\eta = 0.2$ and $\sigma_y = 1.2$, simulate 1000 data points from the model and report the marginal posterior densities in Figure A.2. Panel (a) shows the full chain trajectories of ρ (in blue), σ_η (in red) and σ_y (in yellow) followed by the corresponding marginal posterior distributions (panels (b),(c) and (d)), where the blue line indicates the flat prior and the red circles the true values. The last panel (e) shows the posterior predictive density obtained from our estimation. Once more, we used 20000 iterations, where only the last 15000 are used as posterior sample. The posterior means and the 90% Bayesian intervals of the model parameters are reported in Table (A.1).

	True	Post. Mean	CI (90%)
ρ	0.95	0.9334	(*0.8758, 0.9747*)
σ_η^2	0.04	0.0561	(*0.0238, 0.1171*)
σ_y^2	1.44	1.4509	(*1.1121, 1.9055*)

Table A.1: Simulated data: Posterior medians and 90% CI in the parentheses.

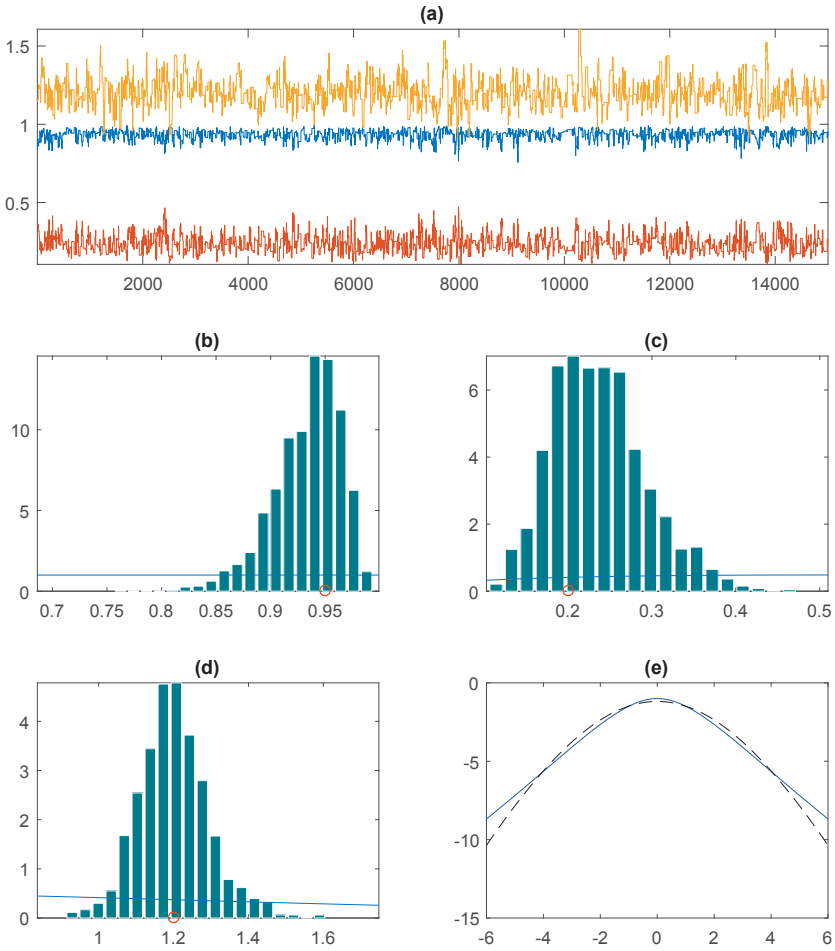


Figure A.2: Classical stochastic volatility model (Section A.2): (a) trace plots of ρ (in blue), σ_η (in red) and σ_y (in yellow), (b), (c) and (d) Marginal posterior of distributions of ρ, σ_η and σ_y (with priors in blue and true values indicated with red circles); (e) posterior log-predictive density (blue line) and true log-predictive density (dashed black line).

A.3 Particle filter

For parsimony, we omit the parameter set from the conditioning set, i.e. we write $p(s_t|y_t) \equiv p(s_t|y_t, \theta)$. Except for minor changes, our notation follows Herbst and Schorfheide (2015).

1. **Initialization**

   Generate a particle swarm $\{\mathbf{s}_0, \mathbf{W}_0\}$ by means of N_p i.i.d. draws from a prior distribution $p(s_0)$ and set the initial weights $\mathbf{W}_0 = \mathbf{1}_{N_p}$, where $\mathbf{1}_{N_p}$ is a $N_p \times 1$ vector of ones.

2. **Recursion**. For $t = 1, \ldots, T$:

   a. **Forecast $\mathbf{s}_t$**

   Iterate $\mathbf{s}_{t-1}$ forward using the state-transition equation

   $$\mathbf{s}_t = f(\mathbf{s}_{t-1}, \boldsymbol{\theta}, \epsilon_s).$$

   The swarm $\{\mathbf{s}_t, \mathbf{W}_{t-1}\}$ approximates the forecast density $p(s_t|\mathbf{y}_{1:t-1})$.

   b. **Forecast $\mathbf{y}_t$**

   The forecast density of y_t is

   $$p(y_t|\mathbf{y}_{1:t-1}) = \int p(y_t|s_t, \mathbf{y}_{1:t-1}) p(s_t|\mathbf{y}_{1:t-1})\, ds_t$$

   with each incremental weight $p(y_t|s_t, \mathbf{y}_{1:t-1}) =: w_t$ computed from the observation equation $f(\cdot)$ and the distribution $\mathcal{F}$. Consequently

   $$\hat{p}(y_t|\mathbf{y}_{1:t-1}) = \frac{1}{N_p} \mathbf{w}_t' \mathbf{W}_{t-1}$$

   is the approximate predictive density.

   c. **Updating**

   Bayes' theorem yields the updated density

   $$p(s_t|\mathbf{y}_{1:t}) = p(s_t|\mathbf{y}_{1:t-1}, y_t) = \frac{p(y_t|s_t, \mathbf{y}_{1:t-1}) p(s_t|\mathbf{y}_{1:t-1})}{p(y_t|\mathbf{y}_{1:t-1})},$$

which is approximated by the swarm $\{\mathbf{s}_t, \tilde{\mathbf{W}}_t := \frac{\mathbf{w}_t \cdot \mathbf{W}_{t-1}}{\hat{p}(y_t|\mathbf{y}_{1:t-1})}\}$.

d. **Resampling**

If the variation of the particles approaches a lower limit defined by the effective sample size

$$\widehat{\mathrm{ESS}_t} = N_p / \left(\frac{\tilde{\mathbf{W}}_t' \tilde{\mathbf{W}}_t}{N_p} \right),$$

all particles $\mathbf{s}_t$ are resampled from a multinomial distribution using weights $\tilde{\mathbf{W}}_t$. In the case of resampling, set $\mathbf{W}_t = \mathbf{1}$, and $\mathbf{W}_t = \tilde{\mathbf{W}}_t$ otherwise.

Appendix B

B.1 Sampling the AR parameters in Cholesky-DPM-MSV

In summary, the priors on the autoregressive parameters are given as

$$c_{\boldsymbol{\alpha}j} \sim N(\alpha_{c_0}, \beta_{c_0}), \tag{B.1}$$

$$\phi_{\boldsymbol{\alpha}j} \sim N(\alpha_{\boldsymbol{\alpha}_0}, \beta_{\boldsymbol{\alpha}_0})\mathbf{1}(|\phi_{\boldsymbol{\alpha}j}| < 1), \tag{B.2}$$

$$\sigma^2_{\mathbf{e}j} \sim \text{IG}(\nu_{\mathbf{e}_0}/2, s_{\mathbf{e}_0}/2), \tag{B.3}$$

$$\phi_{\mathbf{h}i} \sim N(\alpha_{\mathbf{h}_0}, \beta_{\mathbf{h}_0})\mathbf{1}(|\phi_{\mathbf{h}i}| < 1), \tag{B.4}$$

$$\sigma^2_{\boldsymbol{\eta}i} \sim \text{IG}(\nu_{\boldsymbol{\eta}_0}/2, s_{\boldsymbol{\eta}_0}/2), \tag{B.5}$$

for $j = 1, \ldots, p$ and $i = 1, \ldots, m$, and where IG denotes the inverse Gamma distribution.

The full conditional posteriors of the parameters $c_{\boldsymbol{\alpha}j}$ and $\phi_{\boldsymbol{\alpha}j}$ are given by

$$\pi(c_{\boldsymbol{\alpha}j}|\phi_{\boldsymbol{\alpha}j}, \sigma^2_{\mathbf{e}j}, \boldsymbol{\alpha}_j) \propto \pi(c_{\boldsymbol{\alpha}j}) \times \sqrt{1-\phi^2_{\boldsymbol{\alpha}j}} \times \exp\left\{\frac{-(1-\phi^2_{\boldsymbol{\alpha}j})(\alpha_{j1} - \frac{c_{\boldsymbol{\alpha}j}}{1-\phi_{\boldsymbol{\alpha}j}})^2 + 2c_{\boldsymbol{\alpha}j}Q2_j - Tc^2_{\boldsymbol{\alpha}j} - 2c_{\boldsymbol{\alpha}j}\phi_{\boldsymbol{\alpha}j}Q1_j}{2\sigma^2_{\mathbf{e}j}}\right\},$$

$$\pi(\phi_{\boldsymbol{\alpha}j}|c_{\boldsymbol{\alpha}j}, \sigma^2_{\mathbf{e}j}, \boldsymbol{\alpha}_j) \propto \pi(\phi_{\boldsymbol{\alpha}j}) \times \sqrt{1-\phi^2_{\boldsymbol{\alpha}j}} \times \exp\left\{\frac{-(1-\phi^2_{\boldsymbol{\alpha}j})(\alpha_{j1} - \frac{c_{\boldsymbol{\alpha}j}}{1-\phi_{\boldsymbol{\alpha}j}})^2 + 2\phi_{\boldsymbol{\alpha}j}Q4_j - 2c_{\boldsymbol{\alpha}j}\phi_{\boldsymbol{\alpha}j}Q1_j - \phi^2_{\boldsymbol{\alpha}j}Q3_j}{2\sigma^2_{\mathbf{e}j}}\right\},$$

where $\boldsymbol{\alpha}_j = (\alpha_{j1}, \ldots, \alpha_{jT})$, $\pi(c_{\boldsymbol{\alpha}j})$ and $\pi(\phi_{\boldsymbol{\alpha}j})$ are the prior densities (according Eqs. (B.1) and (B.2)), T is the sample length, and

$$Q1_j = \sum_{t=2}^{T} \alpha_{jt-1}, \tag{B.6}$$

$$Q2_j = \sum_{t=2}^{T} \alpha_{jt}, \tag{B.7}$$

$$Q3_j = \sum_{t=2}^{T} \alpha_{jt-1}^2, \tag{B.8}$$

$$Q4_j = \sum_{t=2}^{T} \alpha_{jt-1}\alpha_{jt}. \tag{B.9}$$

The conditional posteriors are of non-standard form and can not be sampled directly. In this case, a Random Walk Metropolis Hastings (RWMH) algorithm with appropriately scaled normal proposals is used.

The AR-parameter $\sigma_{\mathbf{e}j}^2$ can be sampled directly due to the conjugacy of the prior:

$$\sigma_{\mathbf{e}j}^2 | c_{\boldsymbol{\alpha}j}, \phi_{\boldsymbol{\alpha}j}, \boldsymbol{\alpha}_j \sim \mathrm{IG}(\bar{\nu}_{\mathbf{e}}, \bar{s}_{\mathbf{e}j}), \tag{B.10}$$

where $\bar{\nu}_{\mathbf{e}} = (T - 1 + \nu_{\mathbf{e}_0})/2$ and $\bar{s}_{\mathbf{e}j} = (s_{\mathbf{e}_0} + (1 - \phi_{\boldsymbol{\alpha}j}^2)(\alpha_{j1} - \frac{c_{\boldsymbol{\alpha}j}}{1-\phi_{\boldsymbol{\alpha}j}})^2 + \sum_{t=2}^{T}(\alpha_{jt} - c_{\boldsymbol{\alpha}j} - \phi_{\boldsymbol{\alpha}j}\alpha_{jt-1})^2)/2$.

We sample the log-volatility parameters in an analogous way. The full conditional posterior of $\phi_{\mathbf{h}i}$ can be written as

$$\pi(\phi_{\mathbf{h}i} | \sigma_{\boldsymbol{\eta}i}^2, \mathbf{h}_i) \propto \pi(\phi_{\mathbf{h}i}) \times \sqrt{1 - \phi_{\mathbf{h}i}^2}\ exp\left\{ \frac{-(1 - \phi_{\mathbf{h}i}^2)h_{i1}^2 + 2\phi_{\mathbf{h}i}Q4_i - \phi_{\mathbf{h}i}^2 Q3_i}{2\sigma_{\boldsymbol{\eta}i}^2} \right\},$$

where $\mathbf{h}_i = (h_{i1}, \ldots, h_{iT})'$, $\pi(\phi_{\mathbf{h}i})$ is the prior density from Eq. (B.4), $Q3_i = \sum_{t=2}^{T} h_{it-1}^2$, $Q4_i = \sum_{t=2}^{T} h_{it-1}h_{it}$, and sampling is done again by RWMH.

Drawing from the posterior $\sigma_{\boldsymbol{\eta}i}^2 | \phi_{\mathbf{h}i}, \mathbf{h}_i \sim \mathrm{IG}(\bar{\nu}_{\boldsymbol{\eta}}/2, \bar{s}_{\boldsymbol{\eta}i}/2)$ can be executed by directly sampling from the IG with the updated parameters $\bar{\nu}_{\boldsymbol{\eta}} = (T - 1 + \nu_{\boldsymbol{\eta}_0})/2$ and $\bar{s}_{\boldsymbol{\eta}i} = (s_{\boldsymbol{\eta}_0} + (1 - \phi_{\mathbf{h}i}^2)h_{i1}^2 + \sum_{t=2}^{T}(h_{it} - \phi_{\mathbf{h}i}h_{it-1})^2)/2$.