Essays on Multivariate Stochastic Volatility Models

D I S S E R T A T I O N
of the University of St. Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs

to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by

Sebastian Trojan

from

Germany

Approved on the application of

Prof. Francesco Audrino, Ph.D.

and

Prof. Dr. Nikolaus Hautsch

Dissertation no. 4313

(Difo-Druck, Bamberg, 2015)
The University of St. Gallen, School of Management, Economics, Law, Social Sciences and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, September 23, 2014

The President:

Prof. Dr. Thomas Berger
I would like to express my deepest sense of gratitude and appreciation to my supervisor Prof. Dr. Francesco Audrino, who initiated my research efforts in the exciting field of nonlinear non-Gaussian state space systems. His valuable suggestions and deep insights helped to improve this thesis continuously. I would like to thank the chair of the thesis committee, Prof. Dr. Michael Lechner, whose assessment and comments sensibilized me for the required expertise of a scientific researcher. I would also like to thank my co-referee Prof. Dr. Nikolaus Hautsch, who gave me advice and further impetus to excel.

Special thanks go to Prof. Dr. Yasuhiro Omori for highly appreciated feedback on an earlier draft of this work and to Dr. Jouchi Nakajima for timely advice. It is an honor to thank Prof. Dr. Heinz Müller, who encouraged me early on in my PhD. I would like to thank Prof. Dr. Martin Kolmar for backing my endeavors to master complexity and Prof. Dr. Paul Söderlind for support during my research. I also thank Prof. Dr. Reto Föllmi, Prof. Dr. Enrico De Giorgi, and Prof. Dr. Matthias Fengler for their suggestions. And of course, a warm thank you to all seminar participants at the University of St. Gallen for their thoughtful comments.

I would like to thank Prof. Dr. Sven Rady for guidance and support early on in my academic career.

Finally, I would like to say a hearty thank you to my family for their unconditional and enduring support. In particular, I am truly indebted to my mother for helping in whatever way she could during this challenging period. This work is also in memory of Filou, our beloved dog. The recreational English garden strolls with her were a great opportunity to reflect on the subject at hand and formed an important part of my life as a PhD student.

January 2015, Sebastian Trojan
Contents

Summary ix
Zusammenfassung xi
List of Figures xiii
List of Tables xv
Preface xix

A Regime Switching Stochastic Volatility with Skew, Fat Tails and Leverage using Returns and Realized Volatility Contemporaneously 1
A.1 Introduction 2
A.2 General Model Specification 5
A.3 McMC Algorithm 10
A.4 Particle Filter 20
A.5 Simulation Study 24
A.6 Empirical Application 29
  A.6.1 Data 29
  A.6.2 The MS3SVsk Model 30
  A.6.3 The MS4SVskt-RV Model 36
  A.6.4 The MS4SVskt-Range Model 40
  A.6.5 The MS2SVskt-VIX Model 46
  A.6.6 Model Selection 52
### CONTENTS

#### A.6.7 Prior Sensitivity Analysis .......................... 55
#### A.6.8 Out-of-Sample Predictive Density Tests .............. 56
#### A.7 Conclusion .................................................... 61
#### A.A Multi-Move Sampler ........................................ 63
#### A.B The $GH$ skew-$t$ Distribution ....................... 68
#### A.C Inefficiency Factor / Convergence Diagnostic .......... 72
#### A.D Specification Tests ........................................ 73
#### A.E Parameter Estimates ........................................ 77

#### B Modeling Intraday Stochastic Volatility and Conditional Duration Contemporaneously with Regime Shifts 91

B.1 Introduction ...................................................... 92
B.2 The Model ......................................................... 95
B.3 McMC Algorithm .................................................. 98
B.4 Particle Filter ..................................................... 105
B.5 Adaptive Metropolis ................................................ 109
B.6 Empirical Application ............................................. 112
B.6.1 Data ............................................................... 112
B.6.2 The SCDSV Model ............................................. 116
B.6.3 The Regime Switching SCDSVrs Model ................. 119
B.6.4 Model Selection ................................................ 126
B.6.5 Predictive Density Tests ...................................... 127
B.7 Conclusion ......................................................... 132
B.A Multi-Move Sampler ............................................. 134
B.B Transition Probability Estimates ............................ 138
B.C Adaptive Metropolis Algorithm ............................. 139

#### C Multivariate Stochastic Volatility with Dynamic Cross Leverage 143

C.1 Introduction ....................................................... 144
C.2 The Model ........................................................ 146
C.3 McMC Algorithm .................................................. 149
C.4 Particle Filter ...................................................... 154
C.5 Empirical Application .............................................. 158
  C.5.1 Data .................................................................. 158
  C.5.2 The MSVcc Model (Constant Correlation) ............... 160
  C.5.3 The MSVdc Model (Dynamic Correlation) ............... 162
C.6 Model Selection ......................................................... 171
C.7 Conclusion ............................................................... 172
C.A Multi-Move Sampler ................................................ 174
C.B Derivatives $\delta_t, A_t, B_t$ ....................................... 178
C.C Sampling from the Wishart Distribution ....................... 181

Bibliography ........................................................................ 183
Summary

The first essay describes a very general stochastic volatility (SV) model specification with leverage, heavy tails, skew and switching regimes, using realized volatility (RV) as an auxiliary time series to improve inference on latent volatility. The information content of the range and of implied volatility using the VIX index is also analyzed. Database is the S&P 500 index. Asymmetry in the observation error is modeled by the generalized hyperbolic skew Student-$t$ distribution, whose heavy and light tail enable substantial skewness. Resulting number of regimes and dynamics differ dependent on the auxiliary volatility proxy and are investigated in-sample for the financial crash period 2008/09 in more detail. An out-of-sample study comparing predictive ability of various model variants for a calm and a volatile period yields insights about the gains on forecasting performance from different volatility proxies. Results indicate that including RV or the VIX pays off mostly in more volatile market conditions, whereas in calmer environments SV specifications using no auxiliary series outperform. The range as volatility proxy provides a superior in-sample fit, but its predictive performance is found to be weak.

The second essay presents a high frequency stochastic volatility model. Price duration and associated absolute price change in event time are modeled contemporaneously to fully capture volatility on the tick level, combining the SV and stochastic conditional duration (SCD) model. Estimation is with IBM stock intraday data 2001/10 (decimalization completed), taking a minimum midprice threshold of a half tick. Persistent information flow is extracted, featuring a positively correlated innovation term and negative cross effects in the AR(1) persistence matrix. Additionally, regime switching in both duration and absolute price change is introduced to increase nonlinear
capabilities of the model. Thereby, a separate price jump state is identified. Model selection and predictive tests show superiority of the regime switching extension in- and out-of-sample.

The third essay proposes a multivariate stochastic volatility (MSV) model based on a Cholesky-type decomposition of the covariance matrix to model dynamic correlation in the observation and transition error as well as in cross leverage terms. The empirically relevant asymmetric concept of cross leverage is defined as a nonzero correlation between the $i^{th}$ asset return at time $t$ and the $j^{th}$ log-volatility at time $t + 1$. Volatilities and covariances are modeled separately, which makes a direct interpretation of leverage parameters possible. The model is applied on a three-dimensional portfolio consisting of the S&P 500 sector indices Financials, Industrials and Healthcare, spanning the recent financial crisis 2008/09. During and in the aftermath of market turmoil, increased cross leverage effects, higher unconditional kurtosis and stronger correlated information flow are observed. However, there is risk of overfitting and restricting time variation to the elements governing the dynamics of the observation error may be advisable.
Zusammenfassung


eigener Zustand für Preissprünge identifiziert. Modellselektion und Prognosetests zeigen die Überlegenheit der Regimeerweiterung "in-" und "out-of-sample".

List of Figures

A.1 S&P 500 Returns, log of S&P 500 Futures RV, S&P 500 Range, VIX . 31
A.2 MS3SVskt - Selected Posterior Probability Densities . . . . . . . . . . 33
A.3 MS3SVskt - Posterior Probability of States (1990/01/02-2010/08/06) . 35
A.4 SVskt-RV - Selected Posterior Probability Densities . . . . . . . . . . 36
A.5 MS4SVskt-RV - Selected Posterior Probability Densities . . . . . . . . 38
A.6 MS4SVskt-RV - Posterior Probability of States (08/07/01-09/06/30) . 41
A.7 MS4SVskt-Range - Selected Posterior Probability Densities . . . . . . 44
A.8 MS4SVskt-Range - Posterior Probability of States (08/07/01-09/06/30) 46
A.9 SVskt-VIX - Selected Posterior Probability Densities . . . . . . . . . . 47
A.10 MS2SVskt-VIX - Selected Posterior Probability Densities . . . . . . . 49
A.11 MS2SVskt-VIX - Posterior Probability of States (08/07/01-09/06/30) 52
A.12 Partitioning of the Data In-/Out-of-Sample . . . . . . . . . . . . . . . 57
A.13 The $\mathcal{GH}$ skew-$t$ distribution: Tail Behavior and Identification . . . . . . . . . . . . . . . 71
A.14 Transformed Predictions (in-sample, grand averages) . . . . . . . . . . 74

B.1 Comparison of AM Algorithms for $\mathcal{GG}$ Parameter Estimation . . . 112
B.2 Descriptive Statistics of the IBM stock (NYSE), 10/2001 . . . . . . . . 114
B.3 Diurnal Components of the IBM Stock, 10/2001 (cubic spline) . . . . . 114
B.4 Spearman’s $\rho$ for IBM, averaged over trading days 10/2001 . . . . . 115
B.5 Unconditional Densities of the SCDSV Model . . . . . . . . . . . . . . . 118
B.6 Unconditional Densities of the SCDSVrs Model . . . . . . . . . . . . . . 121
B.7 SCDSVrs Model - Posterior Probability of States . . . . . . . . . . . . 123
B.8 Posterior of Latent Process $\psi_t$ (2001/10/01, 9:50-10:20 AM) . . . . . 125
LIST OF FIGURES

B.9 Transformed Predictions (grand averages) .................................. 130
B.10 Autocorrelograms of Transformed Predictions .............................. 131

C.1 Price and Return of S&P 500 Sector Indices .................................. 159
C.2 MSVdc - Volatility with 90% c.b. and Volatility of Volatility ............. 166
C.3 MSVdc - Cholesky Reparameterization Matrix $A_{t:t+1}$ .................... 168
C.4 MSVdc - Correlation Obs. Error $\rho_{\epsilon\epsilon,t}$ and Transition Error $\rho_{\eta\eta,t}$ . . . 169
C.5 MSVdc - Cross Leverage $\rho_{\epsilon\eta}$ .............................................. 170
List of Tables

A.1 Summary Output of Simulated Data from the MS2SVskt-RV Model . 27
A.2 Output of Simulated Data from the MS2SVskt-RV Model, ν Prior #2 . 28
A.3 Descriptive Statistics of the S&P 500 (90/01/02-10/08/06) . . . . 30
A.4 Estimation Results of the SVskt Model . . . . . . . . . . . . . . . . . 32
A.5 Estimation Results of the MS3SVskt Model . . . . . . . . . . . . . . 33
A.6 NBER US Business Cycle Expansions and Contractions . . . . . . 35
A.7 Estimation Results of the SVskt-RV Model . . . . . . . . . . . . . . 37
A.8 Estimation Results of the MS4SVskt-RV Model . . . . . . . . . . . . 39
A.9 Estimation Results of the SVskt-Range Model . . . . . . . . . . . . . 42
A.10 Estimation Results of the MS4SVskt-Range Model . . . . . . . . . 43
A.11 Estimation Results of the SVskt-VIX Model . . . . . . . . . . . . . . 48
A.12 Estimation Results of the MS2SVskt-VIX Model . . . . . . . . . . . 50
A.13 Model Selection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
A.14 Prior Sensitivity of Skew Parameter γ - Model Selection Criteria . . 56
A.15 Descriptive Statistics In-/Out-of-Sample . . . . . . . . . . . . . . . 57
A.16 Forecasting - Predictive Density Ratios . . . . . . . . . . . . . . . 59
A.17 Specification Tests I . . . . . . . . . . . . . . . . . . . . . . . . . . . 75
A.18 Specification Tests II . . . . . . . . . . . . . . . . . . . . . . . . . . . 76
A.19 Estimation Results of the SVt Model . . . . . . . . . . . . . . . . . . 77
A.20 Estimation Results of the SV Model . . . . . . . . . . . . . . . . . . 78
A.21 Estimation Results of the MS3SVskt Model: Transition Probabilities . 78
A.22 Estimation Results of the MS3SVt Model . . . . . . . . . . . . . . . 78
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.23</td>
<td>Estimation Results of the MS3SVt Model: Transition Probabilities</td>
<td>79</td>
</tr>
<tr>
<td>A.24</td>
<td>Estimation Results of the SVt-RV / SV-RV Model</td>
<td>79</td>
</tr>
<tr>
<td>A.25</td>
<td>Estimation Results of the MS4SVskt-RV / MS4SVt-RV Model: Transition</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Probabilities</td>
<td></td>
</tr>
<tr>
<td>A.26</td>
<td>Estimation Results of the MS4SVt-RV Model</td>
<td>81</td>
</tr>
<tr>
<td>A.27</td>
<td>Estimation Results of the MS4SV-RV Model</td>
<td>82</td>
</tr>
<tr>
<td>A.28</td>
<td>Estimation Results of the MS4SV-RV Model: Transition Probabilities</td>
<td>83</td>
</tr>
<tr>
<td>A.29</td>
<td>Estimation Results of the SVt-Range Model</td>
<td>83</td>
</tr>
<tr>
<td>A.30</td>
<td>Estimation Results of the MS4SVskt-Range / MS4SVt-Range Model:</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Transition Probabilities</td>
<td></td>
</tr>
<tr>
<td>A.31</td>
<td>Estimation Results of the MS4SVt-Range Model</td>
<td>85</td>
</tr>
<tr>
<td>A.32</td>
<td>Estimation Results of the MS4SV-Range Model</td>
<td>86</td>
</tr>
<tr>
<td>A.33</td>
<td>Estimation Results of the MS4SV-Range Model</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Transition Probabilities</td>
<td></td>
</tr>
<tr>
<td>A.34</td>
<td>Estimation Results of the SVt-VIX Model</td>
<td>87</td>
</tr>
<tr>
<td>A.35</td>
<td>Estimation Results of the SV-VIX Model</td>
<td>88</td>
</tr>
<tr>
<td>A.36</td>
<td>Results of the MS2SVskt-VIX / MS2SVt-VIX / MS2SV-VIX Model: Transition</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Probabilities</td>
<td></td>
</tr>
<tr>
<td>A.37</td>
<td>Estimation Results of the MS2SVt-VIX / MS2SV-VIX Model</td>
<td>89</td>
</tr>
<tr>
<td>B.1</td>
<td>Observation Densities</td>
<td>96</td>
</tr>
<tr>
<td>B.2</td>
<td>Simulation Results of the Generalized Gamma Shape Parameters</td>
<td>111</td>
</tr>
<tr>
<td>B.3</td>
<td>Descriptive Statistics of the IBM Stock (NYSE), 10/2001</td>
<td>113</td>
</tr>
<tr>
<td>B.4</td>
<td>Estimation Results of the SCDSV Model</td>
<td>117</td>
</tr>
<tr>
<td>B.5</td>
<td>Estimation Results of the SCDSVrs Model</td>
<td>120</td>
</tr>
<tr>
<td>B.6</td>
<td>Model Selection</td>
<td>127</td>
</tr>
<tr>
<td>B.7</td>
<td>Descriptive Statistics of the IBM Stock 11/2001</td>
<td>128</td>
</tr>
<tr>
<td>B.8</td>
<td>Predictive Performance: Goodness-of-fit Statistics</td>
<td>129</td>
</tr>
<tr>
<td>B.9</td>
<td>Results of the SCDSVrs Model - Transition Probability Matrix $P_d$</td>
<td>138</td>
</tr>
<tr>
<td>B.10</td>
<td>Results of the SCDSVrs Model - Transition Probability Matrix $P_y$</td>
<td>139</td>
</tr>
<tr>
<td>C.1</td>
<td>Descriptive Statistics (10/2004-09/2012)</td>
<td>158</td>
</tr>
</tbody>
</table>
LIST OF TABLES

C.2 Estimation Results of the MSVcc Model (constant correlation) . . . . 162
C.3 Estimation Results of the MSVdc Model (dynamic correlation) . . . . 164
C.4 Model Selection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 172
Preface

From a financial markets’ perspective, the concepts of stochastic volatility (SV) and stochastic conditional duration (SCD) can be interpreted as underlying latent information flow that drives market activity (see e.g. Andersen, 1996, in the context of stochastic volatility).\textsuperscript{1,2} The latter is a distinguishing feature and contrasts with the (generalized) autoregressive conditional heteroskedasticity ((G)ARCH) and autoregressive conditional duration (ACD) frameworks, which are deterministic and observable ex-post.\textsuperscript{3,4} The additional stochastic element of the former not only offers increased flexibility in modeling, but is also highly coherent from a broader perspective. As we can not expect to account for every driving force governing information flow, adding probabilistic structure is highly reasonable. However, this rather holistic approach to real world modeling does not come without a cost. The nonlinear latent structure makes maximum likelihood in general infeasible, and simulation based techniques that are computationally more intensive and often challenging to implement haven proven to be most capable. Consequently, a Bayesian approach using Markov chain Monte Carlo (McMC) is employed in all works presented. This thesis offers high level


\textsuperscript{2}The SCD model was introduced by Bauwens and Veredas (2004). Strickland, Forbes, and Martin (2006) offer a Bayesian treatment. See further Pacurar (2008) for an overview.

\textsuperscript{3}Engle (1982) proposed groundbreaking ARCH to model persistence in volatility. The GARCH model introduced by Bollerslev (1986) offers a generally more parsimonious alternative. For a survey, consider e.g. Bollerslev (2008).

\textsuperscript{4}The ACD model was initially developed by Engle and Russell (1998). A survey is given by e.g. Pacurar (2008).
applications of this framework in the field of finance. Its three essays can be read independently.

A basic univariate nonlinear state space system in discrete time, on which the multivariate extensions in this thesis are based upon, can be written as

\[ y_t = x_t' \beta + g(\psi_t) \epsilon_t, \quad t = 1, \ldots, n, \]  
\[ \psi_{t+1} = \mu + z_t' \delta + \phi(\psi_t - \mu) + \eta_{t+1}, \quad t = 1, \ldots, n-1, \]  

where Eq. (1) is the measurement equation having observable \( y_t \) and Eq. (2) the transition or state equation, modeling latent process \( \psi_t \). The latter is commonly assumed to follow a stationary AR(1) process with Gaussian innovation \( \eta_{t+1} \) by imposing \(|\phi| < 1\), and is initialized with the respective unconditional distribution. The state space system is nonlinear in all works presented, deploying \( g(\psi_t) \equiv \exp(\psi_t) \). A further source of nonlinearity is introduced by making parameters of observation and transition equations (1)-(2) regime dependent, see Chapters A and B for applications. Jumps in the transition equation are modeled in this way and can be transient or persistent, according to regime characteristics. Regarding a spurious jump component in the observation equation, one may well choose to use a heavy tailed error like the Student-\( t \) distribution. This is intuitive, as heavy tails and jumps both model outliers. Accordingly, results in the literature using jumps in conjunction with a heavy tailed observation error are mixed and point in the direction of overfitting (see e.g. Chib, Nardari and Shephard, 2002, or Nakajima and Omori, 2009). Consequently, I do not choose to include jumps in the proposed volatility model of Chapter A, which features switching regimes and a skewed heavy tailed observation error.

Generally speaking, the applied framework handles any type of observation error \( \epsilon_t \), whether discrete or continuous, Gaussian or non-Gaussian. Accordingly, Chapter B deploys continuous and discrete distributions with positive support to model financial data on the tick scale. Further, innovations \( \epsilon_t \) and \( \eta_{t+1} \) may be correlated, but this

---

\(^5\)This thesis offers a treatment of SV and SCD in discrete time, motivated by the inherently discrete nature of financial markets. For early work on continuous time SV in the context of option pricing, see Hull and White (1987), Johnson and Shanno (1987), Wiggins (1987), and Black and Scholes (1972). More recent extensions include incorporation of jumps, e.g. Barndorff-Nielsen and Shephard (2001), and Eraker, Johannes, and Polson (2003), or using more than one stochastic process to model volatility, e.g. Chernov, Gallant, Ghysels, and Tauchen (2003), or Barndorff-Nielsen and Shephard (2001).
is only investigated for $\epsilon_t, \eta_{t+1}$ Gaussian in this thesis, see Chapters A and C. Note that only in the special case of $g(\psi_t)$ linear, $\epsilon_t, \eta_{t+1}$ Gaussian, and $E(\epsilon_t \eta_{t+1}) = 0$ an analytical solution for the state space system exists (see e.g Durbin and Koopman, 2008, or Harvey, 1989, for a comprehensive treatment). Finally, covariates $x_t$ and $z_t$ in observation and measurement equations (1) and (2), respectively, are not considered in the current applications. However, these are straightforward extensions and potentially beneficial, especially with respect to forecasting.

The state space system given by Eq. (1)-(2) is a general framework that can be applied in varying context. The current work is concerned with modeling financial volatility and applies the framework to estimating stochastic volatility and stochastic duration. The latter is necessary to give a full description of volatility in event time, see Chapter B. A linear state space model is employed to estimate time varying coefficients in Chapter C.

Estimation of the system given by Eq. (1)-(2) is challenging. The reason is a prohibitive high dimensional integral including latent variables at least of dimension $n$. This renders numerical approaches in general infeasible. Instead, simulation based techniques have proven to be most capable. Moreover, they have become increasingly popular with the advance of computing power. Consequently, the author takes a Bayesian approach and uses the McMC method. Strengths of McMC are e.g. broad applicability, the possibility to draw exact inference, and good performance relative to alternative estimation techniques. Further, the modular structure of McMC makes extensions or adaptations often a straightforward task. For a theoretical thorough treatment of McMC, consider Robert and Casella (2004). Gelman and Rubin (2004) provide an inspiring and comprehensive perspective on the Bayesian approach. More recently, Jacquier and Polson (2011) give an exposition of Bayesian methods in finance. See also Rachev, Hsu, Bagasheva, and Fabozzi (2008) in this regard. Several other estimation techniques exists, which may be differentiated by their degree of efficiency and whether they provide approximate or exact inference. However, a comparison is out of the scope of the current work. For reviews of Bayesian and non-Bayesian approaches to estimate SV models, the interested reader may consider e.g. Broto and Ruiz (2004), Shephard (2005), Asai (2005), or Bos (2012). The latter author focuses on methods that "...have not been clearly surpassed by alternative estimation procedures".
The current work extends the univariate state space system given by Eq. (1)-(2) to the multivariate case along different dimensions. First, one can use more than one observation equation to draw inference on the latent process of Eq. (2). This is investigated in Chapter A, where inference on the SV process is improved by introducing realized volatility (RV) as a second measurement variable. A distinct feature of the paper is a comprehensive analysis of RV, the range, and VIX as auxiliary time series in- and out-of-sample, including an extended regime switching application and analysis around the financial crisis 2008/09. In Chapter B, the focus is still on one asset. However, volatility analysis is now intraday in high frequency on the tick level. Consequently, the concepts of stochastic conditional duration and stochastic volatility are combined to fully capture volatility in event time. Both the observation and latent process are now two dimensional, modeling dynamics of price duration and absolute price change contemporaneously. This is a main contribution of the second paper. Finally, Chapter C presents a multivariate stochastic volatility model applied to a portfolio of assets, each having its own latent volatility process. The main feature is a Cholesky-type decomposition applied to the complete covariance matrix to model dynamic correlation in the observation and transition error as well as in cross leverage terms. All three essays propose an associated McMC algorithm and particle filter for estimation and simulation.

Describing the content in more detail, the first essay presents a very general stochastic volatility model specification with leverage, heavy tails, skew, and switching regimes, using realized volatility as an auxiliary time series to improve inference on latent volatility. The information content of the range, and of implied volatility using the VIX index, is also analyzed. Asymmetry in the observation error is modeled by the generalized hyperbolic skew Student-\(t\) distribution, whose heavy and light tail enable substantial skewness. Up to four regimes are identified from S&P 500 index data using RV as additional time series. Number of regimes and dynamics differ dependent on the deployed auxiliary volatility proxy and are investigated for the financial crash period 2008/09 in more detail. An extensive in-sample model selection follows. Moreover, an out-of-sample study comparing predictive ability of various model variants for a calm and a volatile period yields insights about the gains on forecasting performance that can be expected by incorporating different volatility proxies into the model. Findings
indicate that including RV pays off mostly in more volatile market conditions, whereas in calmer environments SV specifications using no auxiliary series outperform. Results for the VIX as a measure of implied volatility point in a similar direction. The range as volatility proxy provides a superior in-sample fit, but its predictive performance is found to be weak.

In the second essay, the concept of stochastic volatility is applied at the tick level. Absolute changes in the midprice are taken as a proxy for volatility in event time. Volatility, commonly measured over a fixed time interval, is then made up of two components: The duration to observe a price change, and the absolute magnitude of that price change. Consequently, price duration and associated absolute price change in event time are modeled contemporaneously as correlated latent processes, combining the SV and SCD model in high frequency. Estimation is with IBM stock intraday data for the month October, 2001 (decimalization completed), taking a minimum price threshold of 0.5 ticks. Persistent information flow is extracted, featuring a positively correlated innovation term and negative cross effects in the AR(1) persistence matrix. Additionally, regime switching in both duration and absolute price change is introduced to increase nonlinear capabilities of the model. Thereby, a separate price jump state is identified. Model selection and predictive tests show superiority of the regime switching extension in- and out-of-sample. Regarding distributional specification, price duration and absolute price change are modeled using the generalized gamma and negative binomial distribution, respectively. Support for the flexible generalized gamma distribution can be given, as no nested alternative was capable of extracting a persistent information flow process in the regime invariant case. Accordingly, adaptive Metropolis algorithms are explored to efficiently estimate highly correlated shape parameters of the generalized gamma distribution.

The third essay proposes a multivariate stochastic volatility (MSV) model based on a Cholesky-type decomposition of the covariance matrix to model dynamic correlation in the observation and transition error as well as in cross leverage terms. The empirically relevant asymmetric concept of cross leverage is defined as a nonzero correlation between the \( t^{th} \) asset return at time \( t \) and the \( j^{th} \) log-volatility at time \( t + 1 \). Volatilities and covariances are modeled separately, which makes a direct interpretation of leverage parameters possible. The model is applied on a three-dimensional portfolio consisting
of S&P 500 sector indices Financials, Industrials and Healthcare, spanning the recent financial crisis 2008/09. During and in the aftermath of market turmoil, increased cross leverage effects, higher unconditional kurtosis and stronger correlated information flow are observed. To set the model into perspective, the MSV with constant cross leverage proposed by Ishihara and Omori (2012) is estimated as a benchmark. Model comparison demonstrates a superior fit of the proposed model. However, the predictive likelihood criterion indicates risk of overfitting and restricting time variation to the elements governing dynamics of the observation error may be advisable. The latter model variant offers reduced complexity yet still generates nontrivial but intuitive cross leverage and transition error dynamics via the specific structure of the applied Cholesky decomposition.
Chapter A

Regime Switching Stochastic Volatility with Skew, Fat Tails and Leverage using Returns and Realized Volatility Contemporaneously
A.1 Introduction

Since Jacquier, Polson, and Rossi (1994), and Kim, Shephard, and Chib (1998) (KSC) have proposed Bayesian Markov chain Monte Carlo (MCMC) methods for estimating the basic stochastic volatility (SV) model with one factor and Gaussian errors, research in this area has developed considerably. Harvey, Ruiz, and Shephard (1994) propose a Student-\(t\) error based SV model. Subsequently, Chib, Nardari, and Shephard (2002) estimate this type of model by MCMC, using a scale mixture of normals representation to describe excess kurtosis in S&P 500 index data. Abanto-Valle, Bandyopadhyay, Lachos, and Enriquez (2010) examine the slash and variance-gamma distribution. A heavy tailed error results in a more robust underlying stochastic volatility process, as the latter will not increase before a series of large absolute returns is observed. Jacquier, Polson, and Rossi (2004) provide an illustration. They also incorporate the leverage effect into their proposed SV model, albeit defining leverage in a less widespread, contemporaneous way. More commonly, leverage is defined as an increase in future volatility due to a price drop today (see e.g. Omori, Chib, Shephard, and Nakajima, 2007, or Yu, 2005). Leverage is an important stylized fact of especially stock return indices.

Jumps are another tool to model outliers in the data. Chib et al. (2002) include jumps in the observation equation but find that for the specific dataset at hand, a model with Student-\(t\) error outperforms the model with jumps. Moreover, including both jumps and Student-\(t\) error does not yield a significant improvement, probably due to overparameterization. Essentially, fat tails and jumps both model outliers. Once accounted for fat tails of the distribution, there may be too few observations left to estimate the jump parameter reliably. Nakajima and Omori (2009) compare various SV models based on heavy tails and/or jumps, further estimating a candidate with correlated jumps in the return and volatility equation. They arrive at the same conclusion.

Alternatively, tail behavior may be modeled by a second volatility factor. Chernov, Gallant, Ghysels, and Tauchen (2003) test various continuous time model specifications with one and two volatility factors and normal errors in an efficient method of moments (EMM) framework. They find the specification with two factors superior. Molina,
A.1. INTRODUCTION

Han, and Fouque (2010) test one and two factor models with Gaussian errors on various exchange rate series using McMC and conclude that the need for a second factor depends on the dataset at hand. However, as Bouchaud and Potters (2009) note, variance may not be a good description for the tail behavior of a distribution. Results of Durham (2006) support this argument, as he finds the improvement of a second factor to be relatively small and not sufficient to accurately model shape of the conditional return distribution (using simulated maximum likelihood (SML) for estimation). Asai (2008) compares two factor SV models using Gaussian errors with one factor models using the Student-t distribution. His results show empirical superiority of the Student-t specification on various datasets. Liesenfeld and Richard (2003) provide comparable work for the S&P 500, the IBM stock, and the $/DM exchange rate using efficient importance sampling (EIS) for estimation and arrive at similar conclusions. The latter authors also estimate a model with a semi-nonparametric distribution. To this end, Jensen and Maheu (2010) propose to flexibly estimate skewness and kurtosis of the return distribution with nonparametric Bayesian methods.

Yet another flexible approach to model jumps in volatility is by deploying regime switching. A strength of this method is that sudden shifts in volatility can be modeled either as persistent or temporary. For example, during the financial crisis 2008/09, volatility increased significantly for a prolonged period of time. Applications of Markov regime switching to the SV model class can be found in e.g. So, Lam, and Li (1998), Shibata and Watanabe (2005), or Carvalho and Lopes (2007). The latter authors apply sequential Monte Carlo (SMC) for estimation.

In addition to being leptokurtic, return distributions likely exhibit skewness as well. Cappuccio, Lubian, and Raggi (2004) model SV of daily/weekly exchange rates using a skewed generalized error distribution (Skew-GED). They further estimate nested alternatives (neither skew, fat tails, or both), and conclude in favor of the full specification. Abanto-Valle, Lachos, and Dey (2013) use the skew-Student-t (ST) distribution to model SV of NASDAQ daily index returns. In a model selection study similar in structure to that of the former authors they find the skewed heavy tailed specification superior. Nakajima and Omori (2012) model the return innovation as a generalized hyperbolic skew Student-t distribution ($\mathcal{GH}$ skew-t) and show that this specification outperforms both the basic SV model and the model with symmetric
Student-t error for different periods of the S&P 500 and TOPIX daily return index. They include leverage in all their models.

An auxiliary time series potentially improves inference on the latent SV process. In a Bayesian framework, Mahieu and Bauer (1998), Watanabe (2000), Abanto-Valle, Migon, and Lopes (2010), and Abanto-Valle, Dey, and Lachos (2012) incorporate trading volume into the SV model as an additional measurement equation. Their primary interest lies in answering the question of how well an underlying latent process (i.e., the information flow) explains both return and volume. Such a complete specification of the joint dynamic distribution of return and related market variables allows for a more structural approach to stochastic volatility modeling and is generally referred to as the mixture of distributions hypothesis (MDH) (Clark, 1972), and building on the latter, the modified mixture hypothesis (MMM) (Andersen, 1996).

Combining SV models with realized volatility (RV) estimators containing intraday information, or with implied volatility containing future expectations, is another possibility to improve on inference and forecasting of latent volatility. Koopman, Jungbacker, and Hol (2005) include realized and implied volatility measures as explanatory variables in the SV transition equation and compare predictive ability with competing GARCH and RV models for the S&P 100. Alizadeh, Brandt, and Diebold (2002), and Brandt and Jones (2005) use the range to extract latent stochastic volatility of daily exchange rates. Takahashi, Omori, and Watanabe (2009) model a bivariate system of RV and returns to estimate stochastic volatility of the TOPIX, testing several RV estimators in-sample, and proposing a Bayesian McMC estimation algorithm. Dobrev and Szerszen (2010) present a related approach. Takahashi, Watanabe and Omori (2014) extend Takahashi et al. (2009) using the $GH$ skew-t distribution, and conducting volatility and quantile forecasts. Koopman and Scharth (2012) present a two step maximum likelihood estimation method for these kind of models and a detailed empirical study. Jacquier and Miller (2012) (henceforth JM) compare different SV specifications including RV and implied volatility w.r.t. their predictive ability on index returns and foreign exchange rates during the 2008-09 financial crisis, applying sequential Monte Carlo with time varying parameters.

Encouraged by positive results in the literature regarding inference and prediction of latent volatility when including RV into SV models (e.g., JM), this approach is
extended in two directions. First, the asymmetric $\mathcal{GH}$ skew-$t$ distribution, which is able to model substantially skewed data, is used for returns. Second, regime switching is introduced. Moreover, the range and implied volatility as auxiliary time series are also investigated. In fact, four distinct regimes can be identified using realized volatility and return contemporaneously. Regime partitions vary according to the auxiliary measure used and are investigated in-sample during the financial crisis 2008/09 in more detail. Extended in- and out-of-sample model selection and forecasting studies provide further insights.

The paper proceeds as follows. Section A.2 presents the model and Section A.3 details a McMC algorithm to estimate it. Section A.4 contains an associated particle filter. Section A.5 conducts simulation studies. Empirical applications on S&P 500 index data follow in Section A.6, including parameter and regime inference for model variants, model selection, prior sensitivity analysis and out-of-sample predictive density tests. Section A.7 concludes. The Appendix contains technical details, model estimates and additional specification tests.

**A.2 General Model Specification**

A basic stochastic volatility (SV) model with normal observation error and leverage can be written as

$$
y_t = \exp(h_t/2) \epsilon_t, \quad t = 1, \ldots, n, \quad (A.1)
$$

$$
h_{t+1} = \mu + \phi(h_t - \mu) + \eta_{t+1}, \quad t = 1, \ldots, n - 1, \quad (A.2)
$$

$$
h_1 \sim \mathcal{N}(\mu, \sigma_{ini}^2), \quad (A.3)
$$

$$
\begin{pmatrix} \epsilon_t \\ \eta_{t+1} \end{pmatrix} \sim \mathcal{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho \sigma \\ \rho \sigma & \sigma^2 \end{pmatrix}, \quad (A.4)
$$

where $y_t$ is an asset return and $h_t$ the unobserved volatility.\footnote{Strictly speaking, it is the unobserved log-variance, but as in the literature I speak about volatility throughout (similarly for RV related measures). This simplifies terminology and should not cause conceptual confusion as the respective measures are linked by monotonic transformation.} In state space terminology, Eq. (A.1) is called measurement or observation equation and Eq. (A.2) transition...
equation, the latter being unobservable (see e.g Durbin and Koopman, 2008, or Harvey, 1989, for a comprehensive treatment). The volatility process is commonly assumed to follow a stationary AR(1) process by imposing $|\phi| < 1$. The initial value $h_1$ in Eq. (A.3) then follows the unconditional distribution with $\sigma^2_{\text{ini}} = \sigma^2/(1 - \phi^2)$. Parameter $\rho$, Eq. (A.4), measures the correlation between $\epsilon_t$ and $\eta_{t+1}$. We have volatility asymmetry if $\rho \neq 0$ and specifically, if $\rho < 0$, we speak of the leverage effect.

As briefly mentioned in the introduction, there is a timing issue of the leverage effect. Leverage may be modeled in accordance with the traditional definition of leverage as in Black (1976) and Christie (1982). A negative return today will increase volatility tomorrow. Alternatively, the leverage effect may be modeled as an instantaneous relationship between return and volatility, i.e. a negative return today is associated with a concurrent shift in volatility. This introduces an additional source of non-Gaussianity into the model, as the conditional return distribution gets skewed. Jacquier et al. (2004) argue that this may help to model occasionally occurring extreme negative returns (market crashes). Empirically, Durham (2006) concludes positively for the contemporaneous specification, whereas Yu (2005) for the traditional. However, Yu (2005) shows that a contemporaneous correlation between return and volatility implies a nonzero expected return and makes an interpretation of the correlation parameter difficult. Consequently, leverage is implemented using the traditional specification, which is also favored in the literature.\(^2\) Moreover, as asymmetry is modeled explicitly, the additional effect of conditional skewness via a contemporaneous specification is unnecessary.

Asymmetry in the return distribution is incorporated into the model by replacing measurement error $\epsilon_t$ in Eq. (A.4) with a generalized hyperbolic ($\mathcal{GH}$) random variate $w_t$ following Nakajima and Omori (2012). Specifically, as parameters of this distribution are hard to estimate due to a flat likelihood (e.g. Aas and Haff, 2006), a limiting case, the generalized hyperbolic skew Student-$t$ ($\mathcal{GH}$ skew-$t$) distribution, is deployed. It is parsimoniously parameterized, more amenable to estimation, and has a normal mean-variance mixture representation,

$$ w_t = \mu_w + \gamma z_t + \sqrt{z_t} \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 1), \quad z_t \sim \mathcal{IG}(\nu/2, \nu/2), \quad (A.5) $$

\(^2\)Modeling leverage in the traditional definition has an additional feature: The resulting discrete time model is an Euler approximation to the continuous time SV model with leverage effect.
A.2. GENERAL MODEL SPECIFICATION

with $\mu_w$ and $\gamma$ the location and skew parameter, respectively. Scaling variable $z_t$ is inverse gamma distributed. The structure of Eq. (A.5) leans itself well to the construction of a McMC algorithm in the context of Bayesian inference. A distinct feature of the $\mathcal{GH}$ skew-$t$ distribution is its heavy and light tail that enable the modeling of substantial skewness. See App. A.B for a derivation from the $\mathcal{GH}$ distribution and illustrations. To achieve $E(w_t) = 0$, a common assumption of high frequency returns, let $\mu_w = -\gamma \mu_z$ and $\mu_z \equiv E(z_t) = v/(v - 2)$. The additional constraint $v > 4$ ensures existence of the $2^{nd}$ moment, $E(w^2_t) < \infty$.

Realized volatility (RV), which has been independently proposed by Andersen and Bollerslev (1998), and Barndorff-Nielsen and Shephard (2001a), provides a consistent and unbiased estimator of the integrated volatility (IV) of a price process under ideal market assumptions, i.e. continuous and frictionless trading. With $n$ intraday returns during day $t$, \( \{r_{t,i}\}_{i=1}^n \),

\[
RV_t = \sum_{i=1}^n r_{t,i}^2, \quad IV_t = \int_t^{t+1} \sigma^2(s)ds, \quad RV_t \to IV_t, \quad n \to \infty. \quad (A.6)
\]

In real markets, however, RV tends to be a biased estimator of IV due to microstructure effects and non-trading. RV provides empirical content to latent volatility state variable $h_t$ (see e.g. Andersen, 2008, for a treatment in continuous time). Consequently, RV is taken as an observable proxy for latent volatility, on which the focus of modeling and prediction lies in this work. Gains on inference about $h_t$ can be expected, as RV contains valuable intraday information. To give a simple illustration, we may have a near zero return on a specific day, but the market went up and down a lot. Then, using return alone, we would falsely infer low volatility.

Various measures for realized volatility have been proposed in the literature. However, a detailed discussion is out of the scope of the current work. To give examples, Hansen and Lunde (2006) scale RV to account for non-trading hours. Zhang, Mykland, and Aït-Sahalia (2005) propose a two scale RV estimator (TSRV), correcting bias due to microstructure noise by combining two RV estimators from different frequencies. Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) design a realized kernel estimator that explicitly accounts for market frictions while taking advantage of the highest frequency available. For reviews on RV, consider e.g. Andersen,
Bollerslev and Diebold (2010), or McAleer and Medeiros (2008). See further JM, Takahashi et al. (2009), and the references therein for a more detailed discussion of SV in combination with RV estimators and microstructure effects. As in JM, information content of implied volatility and the range as additional volatility proxies is also investigated.

A complete Markov switching stochastic volatility model with $G\mathcal{H}$ skew-$t$ error and RV as auxiliary measurement variable (MSSVskt-RV in the following) is now formulated as

$$ y_{1,t} = \{\gamma_{s_t}(z_t - \mu_z) + \sqrt{z_t} \epsilon_t\} \times \exp(h_t/2), \quad (A.7) $$

$$ y_{2,t} = \zeta + \xi h_t + u_t, \quad t = 1, \ldots, n, \quad (A.8) $$

$$ h_{t+1} = \mu_{s_{t+1}} + \phi_{s_{t+1}}(h_t - \mu_{s_t}) + \eta_{t+1}, \quad t = 1, \ldots, n - 1, \quad (A.9) $$

$$ h_1 \sim \mathcal{N}(\mu_{s_1}, \sigma_{s_1}^2), $$

where

$$ \theta_{s_t} = \sum_{i=1}^{M} \theta_i \times \mathbb{I}_{s_t=i}, \quad M \in \mathbb{N}^+, $$

$$ \theta_i = (\mu_i, \phi_i, \sigma_i, \rho_i, \gamma_i, \sigma_{u,i})^\prime, $$

$$ \left(\begin{array}{c} \epsilon_t \\ \eta_{t+1} \end{array}\right) \sim \mathcal{N}(0, \Sigma_{s_{t+1}}), \quad u_t \sim \mathcal{N}(0, \sigma_{u,s_t}^2), \quad z_t \sim IG(\nu/2, \nu/2), $$

$$ \Sigma_{s_{t+1}} = \begin{pmatrix} 1 & \rho_{s_{t+1}} \sigma_{s_{t+1}}^2 \\ \rho_{s_{t+1}} \sigma_{s_{t+1}} & \sigma_{s_{t+1}}^2 \end{pmatrix}, \quad (A.10) $$

with return $y_{1,t}$, auxiliary time series $y_{2,t}$, and $M$ the number of regimes. As in the basic SV model, Eq. (A.1)-(A.4), stationarity of $h_t$ is imposed, $|\phi_{s_t}| < 1$, and accordingly, $\sigma_{\text{ini}}^2 = \sigma_{s_1}^2/(1 - \phi_{s_1}^2)$. $\mathbb{I}_{s_t=i}$ is an indicator variable that is one if $s_t = i$ (state $i$ prevailing at time $t$) and zero otherwise. Vector $\theta_i, i = 1, \ldots, M$, contains the regime dependent parameters of the $i^{th}$ regime. Note how volatility enters as a linear regression variable into second measurement equation (A.8). If bias correction term $\zeta$ is negative, we may conclude that the effect of non-trading hours dominates market microstructure noise, and vice versa for a positive $\zeta$ (Takahashi et al., 2009). Parameter $\xi$ is a scaling parameter. JM show in a simulation exercise that incorporating RV into the SV model
A.2. GENERAL MODEL SPECIFICATION

via a second measurement equation reduces uncertainty regarding inference of latent volatility in a mean square error sense, both in-sample (smoothing performance) and out-of-sample (predictive ability). The authors further demonstrate that including RV as an exogenous variable directly in Eq. (A.9) has to be considered inferior according to the above criterion.

Model specification is completed by assuming latent state variable $s_t$ to follow a first-order Markov process (see e.g. Hamilton, 1989). Transition probabilities are time invariant, $\Pr(s_t = j | s_{t-1} = i) = p_{ij}$, $i, j \in \{1, \ldots, M\}$, with transition probability matrix having the form

$$P = \begin{pmatrix}
    p_{11} & \cdots & p_{1M} \\
    \vdots & \ddots & \vdots \\
    p_{M1} & \cdots & p_{MM}
\end{pmatrix} \quad (A.11)$$

Generally speaking, in the following empirical applications, as much regimes as possible are extracted out of the employed dataset, using one parameter for state identification, and not imposing rather informative priors.

Almost all parameters in the model are made regime dependent, as ex-ante there is little reason for specific parameters to be assumed equal across regimes. Moreover, extensive simulation has shown that stability of equilibria and associated convergence behavior of the McMC chain often depend on the regime switching variable set. Consequently, a very general model specification is proposed. Specifically, volatility mode $\mu$, volatility persistence $\phi$, volatility of volatility $\sigma$, volatility asymmetry $\rho$, skew parameter $\gamma$, and auxiliary equation measurement noise $\sigma_u$ are made regime dependent. Degrees of freedom parameter $\nu$ is kept constant across regimes, as this parameter is hard to estimate precisely even in large sample sizes. Moreover, computational gains can be made with $\nu$ regime independent, as the inverse gamma distribution has not to be evaluated in the forward filtering backward sampling (FFBS) algorithm for the latent states, see Sec. A.3 Step 3. Note that keeping $\nu$ fix imposes equality on the heavier tail decay rate of the regime dependent distributions. However, skew parameter $\gamma$ may shift large part of the probability mass into the appropriate tail (see App. A.B for an illustration). Regarding a regime switching RV bias parameter $\zeta$, pilot runs extracted a regime with very low volatility, offset by a large positive value for $\zeta$ in the auxiliary
equation. This counteracts our motivation to include intraday information, also pointing
in the direction of overfitting, and consequently auxiliary regression parameters \( \zeta \) and \( \xi \) are set regime independent as well\(^3\). To summarize, the proposed regime switching
specification is quite exhaustive and beyond that of most models in the literature, which
commonly assume a regime dependent volatility mode only. We may miss important
insights if we restrict our analysis solely on a regime switching volatility mode.

The proposed model includes various models of the literature as subcomponents. Dropping regime switching and skew \((\gamma = 0)\) results in the model of JM. Additionally
setting scaling parameters \( z_t = 1, t = 1, \ldots, n \), and \( \xi = 1 \) recovers the model proposed
by Takahashi et al. (2009). Without regime switching and RV as auxiliary measurement
variable, the model reduces to that of Nakajima and Omori (2012). Assuming further a
Gaussian measurement error we obtain the model of Yu (2005), or Omori, Chib, and
Shephard (2007).

A.3 McMC Algorithm

Let \( \theta \equiv \{ \mu, \phi, \sigma, \rho, \gamma, \nu, \zeta, \xi, \sigma_u, P \} \) and \( y_t = (y_{1,t}, y_{2,t})', y_{1:n} = (y_1, \ldots, y_n)', h_{1:n} = (h_1, \ldots, h_n)', z_{1:n} = (z_1, \ldots, z_n)', s_{0:n} = (s_0, s_1, \ldots, s_n)' \). All parameters in \( \theta \)
except \( \nu \) and \( \vartheta \equiv (\zeta, \xi)' \) are regime dependent, e.g. \( \mu = (\mu_1, \ldots, \mu_M)', M \in \mathbb{N}^+ \). Denote by \( \theta_{-r} \) all parameters in \( \theta \) without \( r \). Further, introduce the reparameterization
\( \{ \sigma, \rho \} \mapsto \{ \lambda, a \} \) leading to a Cholesky-type decomposition of \( \Sigma \), Eq. (A.10), to sample
\( \{ \sigma, \rho \} \) in analytical form, see Step 11 below. Prior distributions are assumed to be
independent and generically denoted by \( \pi(\cdot) \), with \( \pi(\mu), \pi(\gamma), \pi(a), \pi(\vartheta) \) conjugate
multivariate normal,

\[
\begin{align*}
\mu & \sim \mathcal{N}(\mu_0, \Sigma_{\mu_0}), & \gamma & \sim \mathcal{N}(\gamma_0, \Sigma_{\gamma_0}), \\
\alpha & \sim \mathcal{N}(\alpha_0, \Sigma_{\alpha_0}), & \vartheta & \sim \mathcal{N}(\vartheta_0, \Sigma_{\vartheta_0}),
\end{align*}
\]

where \( \Sigma_{\mu_0}, \Sigma_{\gamma_0}, \Sigma_{\alpha_0}, \Sigma_{\vartheta_0} \) are diagonal. Priors for \( \lambda, \sigma_u \) are conjugate inverse gamma,

\[
\begin{align*}
\lambda_i & \sim \Gamma_{G}(a_{0,i}/2, b_{0,i}/2), & \sigma_{u,i}^2 & \sim \Gamma_{G}(c_{0,i}/2, d_{0,i}/2), \quad i = 1, \ldots, M,
\end{align*}
\]

\(^3\)However, when using the VIX as auxiliary series, making \( \zeta \) and \( \xi \) regime dependent has improved
convergence properties of these parameters, see Sec. A.6.5.
A.3. McMC ALGORITHM

and those for $P$ conjugate Dirichlet,

$$P_i \sim D(\varrho_{i1}, \ldots, \varrho_{iM}), \quad \varrho_{ij} \neq 0, \quad i, j = 1, \ldots, M,$$

where $P_i$ denotes the $i^{th}$ (independent) row of $P$. Random samples from the full posterior distribution of $\{\theta, h_{1:n}, z_{1:n}, s_{0:n}\}$ given $y_{1:n}$ are then drawn using the McMC method (see e.g. Robert and Casella, 2004). The algorithm for the MSSV$t$-RV model is now structured as follows:\footnote{Implementation is in MATLAB® code developed by the author.}

1. Initialize $\theta, h_{1:n}, z_{1:n}, s_{0:n}$.
2. Generate $h_{1:n} | \theta, z_{1:n}, s_{0:n}, y_{1:n}$.
3. Generate $s_{0:n} | \theta, h_{1:n}, z_{1:n}, y_{1:n}$.
4. Generate $P | s_{0:n}$.
5. Generate $z_{1:n} | \theta, h_{1:n}, s_{0:n}, y_{1,1:n}$.
6. Generate $\nu | \theta_\nu, h_{1:n}, z_{1:n}, s_{0:n}, y_{1,1:n}$.
7. Generate $\gamma | \theta_\gamma, h_{1:n}, z_{1:n}, s_{0:n}, y_{1,1:n}$.
8. Generate $(\zeta, \xi) | \theta_{(\zeta, \xi)}, h_{1:n}, s_{0:n}, y_{2,1:n}$.
9. Generate $\sigma_u | \theta_{-\sigma_u}, h_{1:n}, s_{0:n}, y_{2,1:n}$.
10. Generate $\mu | \theta_{-\mu}, h_{1:n}, z_{1:n}, s_{0:n}, y_{1,1:n}$.
11. Generate $\{\sigma, \rho\} | \theta_{-(\sigma, \rho)}, h_{1:n}, z_{1:n}, s_{0:n}, y_{1,1:n}$.
12. Generate $\phi | \theta_\phi, h_{1:n}, z_{1:n}, s_{0:n}, y_{1,1:n}$.
13. Go to 2.

Note that volatility innovation $\eta_{t+1}$ can be written as

$$\eta_{t+1} = \sigma_{s_{t+1}} \rho_{s_{t+1}} \epsilon_t + \sigma_{s_{t+1}} \sqrt{1 - \rho_{s_{t+1}}^2} \kappa_{t+1}, \quad \begin{pmatrix} \epsilon_t \\ \kappa_{t+1} \end{pmatrix} \sim \mathcal{N}(0, I_2).$$  (A.13)
This is a regression of $\eta_{t+1}$ on $\epsilon_t$, with slope coefficient $\sigma_{s_{t+1}}\rho_{s_{t+1}}$ and error variance $\sigma_{s_{t+1}}^2(1 - \rho_{s_{t+1}}^2)$. Further, define auxiliary variables

$$\bar{h}_t = h_t - \mu_{s_t}, \quad \bar{z}_t = z_t - \mu_z, \quad \text{and} \quad \bar{y}_{1,t} = \sigma_{s_{t+1}}\rho_{s_{t+1}}(y_{1,t}e^{-h_t/2} - \gamma_{s_t}\bar{z}_t)/\sqrt{z_t}.$$ 

In the sequel each updating step of the McMC algorithm is presented in detail.

**Generation of latent states $h_{1:n}$, $s_{0:n}$, and transition probability matrix $P$**

**Step 2.** The multi-move or block sampler proposed by Shepard and Pitt (1997) and modified by Watanabe and Omori (2004), which samples from the true conditional distribution, is applied to extract latent volatility $h_{1:n}$. Another efficient method would be the mixture sampler (e.g. KSC and Omori et al., 2007), which samples from an approximate distribution and error-corrects afterwards. Although computationally less demanding, the required linearization can not be applied to measurement equation (A.7). The single-move sampler proposed by Jacquier, Polson, and Rossi (1994, 2004) or JM is inherently less inefficient but computationally fast. Surveys comparing Bayesian and non-Bayesian methods in the context of SV model estimation are given by Broto and Ruiz (2004), and Bos (2012). The implemented sampler is detailed in App. A.A.

**Step 3.** Sampling of the discrete states is done with a forward filtering backward sampling (FFBS) algorithm as outlined in Chib (1996), or Kim and Nelson (1998, 1999). The forward pass calculates the filtered probabilities recursively by iterating between step 1 and 2, see below. Further define $\Psi_t \equiv \{y_{1:t}, h_{1:t}, z_{1:t}, \theta\}$, the information set available at time $t$ to draw inference on $s_t$.

**Forward Filter - Step 1:** Given $\pi(s_{t-1}|\Psi_{t-1})$, $s_t \in \{1, \ldots, M\}$, at the beginning of the $i^{th}$ iteration, weighting terms are calculated as

$$\pi(s_t, s_{t-1}|\Psi_{t-1}) = \pi(s_t|s_{t-1})\pi(s_{t-1}|\Psi_{t-1}),$$

with $\pi(s_t|s_{t-1})$ the transition probabilities.
Forward Filter - Step 2: Once \( \mathbf{y}_t \) is observed at the end of the \( t \)th iteration, probabilities are updated,

\[
\pi(s_t | \Psi_t) = \frac{f(y_t, \tilde{h}_t | s_t, s_{t-1}, \Psi_t) \pi(s_t, s_{t-1} | \Psi_{t-1})}{f(y_t, \tilde{h}_t | \Psi_t)},
\]

with \( f(y_t, \tilde{h}_t | s_t, s_{t-1}, \Psi_t) \) the joint conditional likelihood and

\[
f(y_t, \tilde{h}_t | \Psi_t) = \sum_{s_t} \sum_{s_{t-1}} f(y_t, \tilde{h}_t | s_t, s_{t-1}, \Psi_t) \pi(s_t, s_{t-1} | \Psi_{t-1}).
\]

The joint conditional likelihood at time \( t \) depends also on time \( t - 1 \), due to the leverage effect. We have

\[
f(y_t, \tilde{h}_t | s_t, s_{t-1}, \Psi_t) \propto \sigma_{u,s_t}^{-1} \exp\left[ -\frac{(y_{2,t} - \zeta - \xi h_t)^2}{2\sigma_{u,s_t}^2} - \frac{h_t}{2} - \frac{(y_{1,t} - \gamma_{s,t} \tilde{z}_t e^{h_t/2})^2}{2z_t e^{h_t}} \right] \\
\times \begin{cases} 
\left(1-\phi_{s_t}^2\right)^{1/2} \sigma_{s_t} \exp\left[ -\frac{(1-\phi_{s_t}^2) \tilde{h}_t^2}{2\sigma_{s_t}^2} \right] & t = 1 \\
\frac{1}{\sigma_{s_t} (1-\rho_{s_t}^2)^{1/2}} \exp\left[ -\frac{(\tilde{h}_t - \phi_{s,t} \tilde{h}_{t-1} - \tilde{y}_{t-1})^2}{2\sigma_{s_t}^2 (1-\rho_{s_t}^2)} \right] & t > 1.
\end{cases}
\]

Initialization of the filter is with the unconditional state probabilities,

\[
\pi(s_0) = (A' A)^{-1} A' \begin{pmatrix} 0_M \\ 1 \end{pmatrix}, \quad A = \begin{pmatrix} I_M - P' \\ i_M' \end{pmatrix},
\]

where \( 0_M \) is a \( M \times 1 \) vector of zeros, \( I_M \) the identity matrix of dimension \( M \), and \( i_M \) a \( M \times 1 \) vector of ones.

Backward Sampling: States \( s_{0:n} \) are drawn as a block from the following conditional distribution

\[
\pi(s_{0:n} | \Psi_n) = \pi(s_n | \Psi_n) \prod_{t=0}^{n-1} \pi(s_t | s_{t+1}, \Psi_t),
\]

where \( \pi(s_t | s_{t+1}, \Psi_t) \propto \pi(s_{t+1} | s_t) \pi(s_t | \Psi_t) \). Thus, we can sample \( s_t \) from posterior mass function

\[
\pi(s_t = i | s_{t+1}, \Psi_t) = \frac{\pi(s_{t+1} | s_t = i) \pi(s_t = i | \Psi_t)}{\sum_{j=1}^{M} \pi(s_{t+1} | s_t = j) \pi(s_t = j | \Psi_t)}, \quad i = 1, \ldots, M.
\]

The backward sampler is initialized with a draw from \( \pi(s_n | \Psi_n) \).
Step 4. Conditional on \( s_{0:n} \), the transition probability matrix \( P \) is independent of all other variables in the model. Further, rows \( P_i \) of \( P \) are independent a posteriori and are drawn from the following Dirichlet distribution,

\[
P_i \sim \mathcal{D}(\varrho_{i1} + n_{i1}(s_{0:n}), \ldots, \varrho_{iM} + n_{iM}(s_{0:n})), \quad i = 1, \ldots, M,
\]

where \( n_{ij}(s_{0:n}) = \# \{ s_{t-1} = i, s_t = j \} \) counts the transitions from \( i \) to \( j \).

**Generation of GH skew-t parameters** \( z_{1:n}, \nu, \gamma \)

Step 5. Latent variables \( z_t, t = 1, \ldots, n \), are sampled one at a time, having conditional posterior

\[
\pi(z_t | \cdot) \propto g(z_t) \times z_t^{-\nu/2} \exp\left\{-\frac{\nu}{2z_t}\right\},
\]

\[
g(z_t) = \exp\left\{-\frac{(y_{1,t} - \gamma_s z_t e^{h_t/2})^2}{2z_t e^{h_t}} - \frac{(\tilde{h}_{t+1} - \phi_{s_{t+1}} \tilde{h}_t - \bar{y}_{1,t})^2}{2\sigma_{s_{t+1}}^2 (1 - \rho^2_{s_{t+1}})} I_{t<n}\right\}.
\]

Using the Metropolis-Hastings (MH) algorithm (see e.g. Chib and Greenberg, 1995), a candidate \( z^*_t \sim IG((\nu + 1)/2, \nu/2) \) is generated and accepted with probability \( \min\{g(z^*_t)/g(z_t), 1\} \), where \( z_t \) is the current draw.

Step 6. The conditional posterior probability density of \( \nu \),

\[
\pi(\nu | \cdot) \propto \pi(\nu) \times \prod_{t=1}^n \left( \frac{\nu/2}{\Gamma(\nu/2)} \right)^{\nu/2} z_t^{-\nu/2} \exp\left\{-\frac{\nu}{2z_t}\right\} \times \exp\left\{-\sum_{t=1}^n \frac{(y_{1,t} - \gamma_s z_t e^{h_t/2})^2}{2z_t e^{h_t}} - \sum_{t=1}^{n-1} \frac{(\tilde{h}_{t+1} - \phi_{s_{t+1}} \tilde{h}_t - \bar{y}_{1,t})^2}{2\sigma_{s_{t+1}}^2 (1 - \rho^2_{s_{t+1}})} I_{t<n}\right\},
\]

does not permit direct sampling. A MH algorithm is deployed, based on a normal approximation of the posterior probability density around the mode, truncated on \( ]4, \infty[ \). Moments are calculated by constraint numerical optimization,

\[
\mu_\nu = \arg \max_\nu \log(\pi(\nu | \cdot)), \quad \sigma^2_\nu = \left\{ -\frac{\partial^2 \log(\pi(\nu | \cdot))}{\partial \nu^2} \right\}^{-1} \bigg|_{\nu = \mu_\nu},
\]
with proposal variance $\sigma_\nu^2$ the negative inverse of the Hessian at $\mu_\nu$. Starting value for optimization is the current draw of $\nu$. Then a proposal $\nu^* \sim \mathcal{N}(\mu_\nu, \sigma_\nu^2)$ is drawn and accepted with probability

$$
\min \left\{ \frac{\pi(\nu^*|\cdot) \times f_N(\nu|\mu_\nu, \sigma_\nu^2)}{\pi(\nu|\cdot) \times f_N(\nu^*|\mu_\nu, \sigma_\nu^2)}, 1 \right\},
$$

where $f_N(x|\mu, \sigma^2)$ denotes the normal density with mean $\mu$ and variance $\sigma^2$.

**Step 7.** Sampling of $\gamma$ is straightforward using results of linear regression theory. The conditional posterior probability density of $\pi(\gamma|\cdot)$ can be written as

$$
\pi(\gamma|\cdot) \propto \pi(\gamma) \times \exp \left\{ -\sum_{t=1}^{n} \left( \frac{(y_{1,t} - \gamma_s, t z_t e^{h_t}/2)^2}{2 z_t e^{h_t}} - \sum_{t=1}^{n-1} \frac{(\hat{h}_{t+1} - \phi_{s,t+1} \hat{h}_t - \bar{y}_{1,t})^2}{2 \sigma_{s,t+1}^2 (1 - \rho_{s,t+1}^2)} \right) \right\}
$$

$$
= \exp \left\{ -\frac{1}{2} \left[ (\gamma - \gamma_0)' \Sigma_0^{-1} (\gamma - \gamma_0) + (Y_{\gamma_1} - X_{\gamma_1} \gamma)'(Y_{\gamma_1} - X_{\gamma_1} \gamma) + (Y_{\gamma_2} - X_{\gamma_2} \gamma)'(Y_{\gamma_2} - X_{\gamma_2} \gamma) \right] \right\},
$$

with

$$
x_{\gamma_{1,t}} = \bar{z}_t \left[ I_{s,t=1} \cdots I_{s,t=M} \right] / \sqrt{z_t e^{h_t}}, \quad y_{\gamma_{1,t}} = y_{1,t} / \sqrt{z_t e^{h_t}}, \quad t = 1, \ldots, n,
$$

$$
x_{\gamma_{2,t}} = -\rho_{s,t+1} \bar{z}_t \left[ I_{s,t=1} \cdots I_{s,t=M} \right] / \sqrt{(1 - \rho_{s,t+1}^2) z_t}, \quad y_{\gamma_{2,t}} = \left( \frac{\hat{h}_{t+1} - \phi_{s,t+1} \hat{h}_t - \bar{y}_{1,t}}{\sigma_{s,t+1}} \right) / \sqrt{1 - \rho_{s,t+1}^2}, \quad t = 1, \ldots, n - 1,
$$

the row elements of $X_{\gamma_1}$, $Y_{\gamma_1}$, $X_{\gamma_2}$, and $Y_{\gamma_2}$, respectively. Hence $\gamma|\cdot \sim \mathcal{N}(\mu_\gamma, \Sigma_\gamma)$, with

$$
\mu_\gamma = \Sigma_\gamma \times (\Sigma_0^{-1} \gamma_0 + X_{\gamma_1}' Y_{\gamma_1} + X_{\gamma_2}' Y_{\gamma_2}), \quad (A.14)
$$

$$
\Sigma_\gamma = (\Sigma_0^{-1} + X_{\gamma_1}' X_{\gamma_1} + X_{\gamma_2}' X_{\gamma_2})^{-1}. \quad (A.15)
$$
Generation of RV regression parameters $\zeta, \xi, \sigma_u$

**Step 8.** Sampling of $\vartheta \equiv (\zeta, \xi)'$ is straightforward using results of linear regression theory. The conditional posterior probability is

$$
\pi(\vartheta | \cdot) \propto \pi(\vartheta) \times \exp\left\{ -\frac{1}{2} (\vartheta - \vartheta_0)' \Sigma^{-1}_{\vartheta_0} (\vartheta - \vartheta_0) - \frac{1}{2} (Y_\vartheta - X_{\vartheta} \vartheta)' (Y_\vartheta - X_{\vartheta} \vartheta) \right\},
$$

with

$$
x_{\vartheta_t} = \sigma_{u,s_t}^{-1} \begin{bmatrix} 1 \\ h_t \end{bmatrix}, \quad y_{\vartheta_t} = \sigma_{u,s_t}^{-1} y_{2,t}, \quad t = 1, \ldots, n,
$$

the row elements of $X_\xi$ and $Y_\xi$, respectively. Hence $\vartheta | \cdot \sim N(\mu_\vartheta, \Sigma_\vartheta)$, where $\mu_\vartheta$ and $\Sigma_\vartheta$ are calculated in analogy to Eq. (A.14)-(A.15).

**Step 9.** Sampling of $\sigma_u^2$ is straightforward using results of linear regression theory. The conditional posterior probability is

$$
\pi(\sigma_u^2 | \cdot) \propto \pi(\sigma_u^2) \times \prod_{i=1}^{M} (\sigma_{u,i}^2)^{-1/2} \exp\left\{ -\frac{1}{2} (y_{2,i} - \zeta - \xi h_t)^2 I_{s_t = i} \right\},
$$

with

$$
c_i = c_{0,i} + n_{s_i}, \quad d_i = d_{0,i} + \sum_{t=1}^{n} (y_{2,t} - \zeta - \xi h_t)^2 I_{s_t = i}, \quad i = 1, \ldots, M,
$$

where $n_{s_i}$ denotes the observations belonging to state $i$. Hence, the elements of $\sigma_u^2$ are independent and sampled state-by-state, $\sigma_{u,i}^2 | \cdot \sim \mathcal{IG}(c_i/2, d_i/2)$. 

A.3. McMC ALGORITHM

Generation of volatility process parameters $\mu, \{\sigma, \rho\}, \phi$

**Step 10.** Sampling of $\mu$ is straightforward using results of linear regression theory. The conditional posterior probability is

$$
\pi(\mu|\cdot) \propto \pi(\mu) \times \exp \left\{ -\frac{(1 - \phi^2_{s_1}) \bar{h}_{t+1}^2}{2\sigma^2_{s_1}} - \sum_{t=1}^{n-1} \left( \frac{(\bar{h}_{t+1} - \phi_{s_{t+1}} \bar{h}_t - \bar{y}_{1,t})^2}{2\sigma^2_{s_{t+1}} (1 - \rho^2_{s_{t+1}})} \right) \right\} 
$$

$$
= \exp \left\{ -\frac{1}{2} (\mu - \mu_0)' \Sigma^{-1}_{\mu_0} (\mu - \mu_0) - \frac{1}{2} (Y_\mu - X_\mu \mu)' (Y_\mu - X_\mu \mu) \right\},
$$

with

$$
x_{\mu_{t+1}} = \left[ I_{s_{t+1}=1} - \phi_{s_{t+1}} I_{s_t=1} \cdots I_{s_{t+1}=M} - \phi_{s_{t+1}} I_{s_t=M} \right] / \sqrt{\sigma^2_{s_{t+1}} (1 - \rho^2_{s_{t+1}})},
$$

$$
y_{\mu_{t+1}} = (h_{t+1} - \phi_{s_{t+1}} h_t - \bar{y}_{1,t}) / \sqrt{\sigma^2_{s_{t+1}} (1 - \rho^2_{s_{t+1}})}, \quad t = 1, \ldots, n-1,
$$

$$
x_{\mu_1} = \begin{bmatrix} I_{s_1=1} \cdots I_{s_1=M} \end{bmatrix} \sqrt{1 - \phi^2_1 \sigma^{-1}_{s_1}}, \quad y_{\mu_1} = h_1 \sqrt{1 - \phi^2_1 \sigma^{-1}_{s_1}},
$$

the row elements of $X_\mu$ and $Y_\mu$, respectively. Hence $\mu|\cdot \sim N(\hat{\mu}, \Sigma_{\mu})$, where $\hat{\mu}$ and $\Sigma_{\mu}$ are calculated in analogy to Eq. (A.14)-(A.15).

Label switching in mixture models may occur when using McMC for estimation, due to the invariance of the likelihood to relabeling the components. This issue has been well investigated (see e.g. Frühwirth-Schnatter, 2001, 2006) and is dealt with by introducing an identification constraint on mode parameter $\mu$, see Sec. A.5 and Sec. A.6.2-A.6.5.

**Step 11.** Sampling of $\{\sigma, \rho\}$ is by Cholesky decomposition as proposed in Chan and Jeliazkov (2009). Let

$$
w_{t:t+1} = \left( \left[ \frac{y_{1,t} e^{-h_t/2} - \gamma_{s_t} \bar{z}_t}{\bar{h}_{t+1} - \phi_{s_{t+1}} \bar{h}_t} \right] / \sqrt{\bar{z}_t} \right) \sim N(0, \Sigma_{s_{t+1}}), \quad t = 1, \ldots, n-1, \quad (A.16)
$$

$$
w_1 = \bar{h}_1 \sqrt{1 - \phi^2_{s_1}} \sim N(0, \sigma^2_{s_1}),
$$

with covariance $\Sigma_{s_{t+1}}$ defined in Eq. (A.10). Applying Cholesky-type decomposition
\[ \Sigma_{s_{t+1}}^{-1} = A'_{s_{t+1}} D_{s_{t+1}}^{-1} A_{s_{t+1}}, \]
rewrite Eq. (A.16) as \[ A_{s_{t+1}} w_{t:t+1} \sim \mathcal{N}(0, D_{s_{t+1}}), \]
with
\[
D_{s_{t+1}} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_{s_{t+1}} \end{pmatrix}, \quad A_{s_{t+1}} = \begin{pmatrix} 1 \\ a_{s_{t+1}} \end{pmatrix}.
\]
The reparameterization in terms of \( \lambda \) and \( a \) allows for a conjugate representation with respect to the multivariate Gaussian likelihood. Define \( \omega_{t:t+1} = A_{s_{t+1}} w_{t:t+1} \), and let \( w = \{w_{1:2}, \ldots, w_{n-1:n}\} \). Observe that \( w_1 \) has been dropped from \( w \). A state specific MH step must be applied due to the initial condition, see below. Note further that \( |\Sigma_{s_{t+1}}^{-1}| = |A'_{s_{t+1}}||D_{s_{t+1}}^{-1}||A_{s_{t+1}}| = |D_{s_{t+1}}|^{-1} = \lambda_{s_{t+1}}^{-1} \). The conditional likelihood without initial condition can now be expressed as
\[
f(w|\cdot) \propto \prod_{t=1}^{n-1} |\Sigma_{s_{t+1}}|^{-1/2} \exp\left\{ -\frac{1}{2} w'_{t:t+1} \Sigma_{s_{t+1}}^{-1} w_{t:t+1} \right\} = \left( \prod_{t=1}^{n-1} \lambda_{s_{t+1}}^{-1/2} \right) \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n-1} \omega'_{t:t+1} D_{s_{t+1}}^{-1} \omega_{t:t+1} \right\}. \tag{A.17}
\]
Rewriting the above expression in terms of \( \lambda_i, i = 1, \ldots, M \) yields
\[
f(w|\cdot) \propto \prod_{i=1}^{M} \lambda_i^{-n_i/2} \exp\left\{ -\frac{1}{2} \text{tr}\left(D_i^{-1} \sum_{t=1}^{n-1} \omega_{t:t+1} \omega'_{t:t+1} I_{s_{t+1}=i}\right) \right\} \propto \prod_{i=1}^{M} \lambda_i^{-n_i/2} \exp\left\{ -\frac{s_i}{2\lambda_i} \right\},
\]
with \( n_i = \#\{s_{t+1} = i, t = 1, \ldots, n-1\} \) and \( s_i = \sum_{t=1}^{n-1} \omega^2_{2,t:t+1} I_{s_{t+1}=i} \), where \( \omega_{j,t:t+1} \) denotes the \( j \)th row of \( \omega_{t:t+1} \). Hence, the conditional posterior of \( \lambda \) is
\[
\pi(\lambda|\cdot) \propto \pi(\lambda) \times \sigma_{s_1}^{-1} \exp\left\{ -\frac{w^2}{2\sigma^2_{s_1}} \right\} \prod_{i=1}^{M} \lambda_i^{-n_i/2} \exp\left\{ -\frac{s_i}{2\lambda_i} \right\}.
\]
Under the inverse gamma prior in Eq. (A.12), sampling of \( \lambda_i \) is state-by-state,
\[
\lambda_i|s_1 = i \sim \mathcal{IG}((a_{0,i} + n_i)/2, (b_{0,i} + s_i)/2).
\]
Note that for state \( i|s_1 = i \) a MH step must be deployed, see below.
A.3. McMC ALGORITHM

Starting with the likelihood of Eq. (A.17), the conditional posterior of $a$ can be written as

$$
\pi(a|\cdot) \propto \pi(a) \times \sigma_{s_1}^{-1} \exp\left\{-\frac{w_1^2}{2\sigma_{s_1}^2} - \frac{n-1}{2} \left(\frac{w_{2,t:t+1} a_{s_{t+1}} + a_{s_{t+1}} w_{1,t:t+1}}{2\lambda_{s_{t+1}}}\right)^2\right\}
$$

$$
= \sigma_{s_1}^{-1} \exp\left\{-\frac{w_1^2}{2\sigma_{s_1}^2}\right\} \times \exp\left\{-\frac{1}{2} (a - a_0)'\Sigma_{a_0}^{-1}(a - a_0) - \frac{1}{2} (Y_a - X_a a)'(Y_a - X_a a)\right\},
$$

with

$$
x_{a_t} = -w_{1,t:t+1}\lambda_{s_{t+1}}^{-1/2}, \quad y_{a_t} = w_{2,t:t+1}\lambda_{s_{t+1}}^{-1/2}, \quad t = 1, \ldots, n - 1,
$$

the row elements of $X_a$ and $Y_a$, respectively. Hence $a|\cdot \sim N(\mu_a, \Sigma_a)$, where $\mu_a$ and $\Sigma_a$ are calculated in analogy to Eq. (A.14)-(A.15). However, as $\Sigma_a$ is diagonal in the current implementation, the elements of $a$ are sampled state-by-state right after the corresponding elements of $\lambda$.

For state $i|s_1 = i$ a MH step must be applied. Corresponding proposal $\{\lambda^*_{i_t}, a^*_{i_t}\}$ is then accepted with probability

$$
\min\left\{\frac{\sigma_{s_1} \exp\{-0.5w_1^2/\sigma_{s_1}^2\}}{\sigma_{s_1}^* \exp\{-0.5w_1^2/\sigma_{s_1}^2\}}, 1\right\},
$$

where $\sigma_{s_1}^2$ is the current state dependent variance prevailing at $t = 1$.

**Step 12.** The conditional posterior probability distribution of $\phi$ can be written as

$$
\pi(\phi|\cdot) \propto \pi(\phi) \times \sqrt{1 - \phi_{s_1}^2} \exp\left\{-\frac{(1 - \phi_{s_1}^2)\bar{h}_{1,1}^2}{2\sigma_{s_1}^2} - \frac{n-1}{2} \left(\frac{\bar{h}_{t+1} - \phi_{s_{t+1}}\bar{h}_t - \bar{y}_{1,t}}{2\sigma_{s_{t+1}}^2 (1 - \rho_{s_{t+1}}^2)}\right)^2\right\}
$$

$$
= \pi(\phi) \times c(\phi) \times \exp\left\{-\frac{1}{2} (Y_\phi - X_\phi \phi)'(Y_\phi - X_\phi \phi)\right\},
$$
with
\[ c(\phi) = \sqrt{1 - \phi_s^2} \exp \left\{ -\frac{(1 - \phi_s^2)\hat{h}_1^2}{2\sigma_s^2} \right\} \]
and
\[ x_{\phi_t+1} = \bar{h}_t \left[ I_{s_{t+1}=1}, \ldots, I_{s_{t+1}=M} \right] / \sqrt{\sigma_{s_{t+1}}^2(1 - \rho_{s_{t+1}}^2)}, \]
\[ y_{\phi_t+1} = (\bar{h}_{t+1} - \bar{y}_{1:t}) / \sqrt{\sigma_{s_{t+1}}^2(1 - \rho_{s_{t+1}}^2)}, \quad t = 1, \ldots, n - 1, \]
the row elements of \( X_\phi \) and \( Y_\phi \), respectively. The MH algorithm is used to sample from this density. A candidate \( \phi^* \sim T N_{(-1,1)}(\mu_\phi, \Sigma_\phi) \) is generated, where \( T N_{(a,b)}(\mu, \Sigma) \) denotes the truncated normal distribution on \((a, b)\), and \( \mu_\phi, \Sigma_\phi \) are calculated in analogy to Eq. (A.14)-(A.15). The proposal is then accepted with probability
\[ \min \left\{ \frac{\pi(\phi^*) c(\phi^*)}{\pi(\phi) c(\phi)}, 1 \right\}, \]
where \( \phi \) is the current draw. Experimentation has shown that when \( \phi_i, i = 1, \ldots, M \), are sampled jointly and vary more significantly in value, the acceptance rate may drop. As priors \( \pi(\phi_t) \) are independent, a more robust state-by-state sampling strategy is implemented in the current work.

### A.4 Particle Filter

A stratified auxiliary particle filter is developed that recursively delivers draws of the latent variables \( \{\alpha_t = h_t - \mu_{s_t}, z_t, s_t\} \) given parameter values \( \theta \) and observables up to time \( t, y_{1:t} \). These draws are needed to evaluate the conditional likelihood, calculate goodness-of-fit statistics, and for forecasting. Pitt and Shephard (1999a) propose the auxiliary particle filter, a sequential Monte Carlo (SMC) technique that increases efficiency in the propagation process by weighting particles according to an importance function dependent on the subsequent observation. Further, multinomial resampling in particle filters is known to increase the Monte Carlo variance of the associated estimators. A stratified approach is able to significantly reduce that variance (see e.g.
A.4. PARTICLE FILTER


Define extended state vector \( x_t \equiv (\alpha_t, s_t, z_t)' \) (for better readability, time invariant parameter set \( \theta \) may be dropped from the conditioning arguments in this section). The proposed model is a nonlinear and non-Gaussian state space model with measurement density

\[
f(y_{1:t}|x_t) = N(y_{1:t}|\gamma_s, z_t \exp[(\mu_s + \alpha_t)/2], z_t \exp[\mu_s + \alpha_t]) \times
N(y_{2:t}|\zeta + \xi(\mu_s + \alpha_t), \sigma_u^2),
\]

transition densities

\[
f(\alpha_{t+1}|s_{t+1}, x_t, y_{1:t}) = N(\phi_{s_{t+1}} \alpha_t + \bar{y}_{1:t}, \sigma_{s_{t+1}}^2 (1 - \rho_{s_{t+1}}^2)),
\]
\[
\Pr(s_{t+1}|s_t) = \text{Multinomial}(p_{s_t}),
\]

and scaling variable

\[
f(z_t) = IG(v/2,v/2),
\]

where \( p_{s_t|i} = (p_{i1}, \ldots, p_{iM})' \), \( i \in M \). Applying Bayes Theorem we obtain the target posterior density

\[
f(x_{t+1}, x_t | y_{1:t+1}) \propto f(y_{t+1}|x_{t+1}) f(x_{t+1}|x_t, y_{1:t}) f(x_t | y_{1:t}),
\]

with

\[
f(x_{t+1}|x_t, y_{1:t}) \propto f(z_{t+1}) f(\alpha_{t+1}|s_{t+1}, x_t, y_{1:t}) \Pr(s_{t+1}|s_t),
\]

where we assume that we have samples (particles) from \( f(x_t | y_{1:t}) \) and a discrete uniform approximation \( \hat{f}(x_t | y_{1:t}) \) to \( f(x_t | y_{1:t}) \).
Samples from Eq. (A.20) are generated with the help of the following importance probability density function involving $y_{t+1}$,

$$
g(x_{t+1}, \tilde{x}_t^{(i)} | y_{1:t+1})
\propto f(y_{t+1} | \hat{\alpha}_{t+1}^{(i,k)}, s_{t+1}, \tilde{z}) f(z_{t+1}) f(\alpha_{t+1} | s_{t+1}, x_t^{(i)}, y_{1:t}) \Pr(s_{t+1} | s_t^{(i)}) \hat{f}(x_t^{(i)} | y_{1:t})$$

\[\text{with}
\]

$$
g(x_t^{(i)}, s_{t+1} | y_{1:t+1}, \hat{\alpha}_{t+1}^{(i,k)}, \tilde{z})
\propto f(y_{t+1} | \hat{\alpha}_{t+1}^{(i,k)}, s_{t+1}, \tilde{z}) \Pr(s_{t+1} | s_t^{(i)}) \hat{f}(x_t^{(i)} | y_{1:t})$$

and "best" guesses

$$
\hat{\alpha}_{t+1}^{(i,k)} = \phi_{s_{t+1}} \alpha_t^{(i)} + \tilde{y}_{1,t}^{(i,k)}, \quad \tilde{y}_{1,t}^{(i,k)} = \tilde{y}_{1,t} | \alpha_t = \alpha_t^{(i)}, s_t = s_t^{(i)}, s_{t+1} = k,
$$

$$\tilde{z} = 1,$$

where superscripts $i = 1, \ldots, I$ index particles and $k = 1, \ldots, M$ states. By including $y_{t+1}$ in the importance function more weight is given to particles with larger predictive values. Note that no best guess is made for $s_{t+1}$. Instead, $g(x_t^{(i)}, s_{t+1} | y_{1:t+1}, \hat{\alpha}_{t+1}^{(i,k)}, \tilde{z})$, Eq. (A.21), is evaluated for every possible value of $s_{t+1}$ (strata), resulting in a collection of $I \times M$ weighted sample points. Then, $\{x_t, s_{t+1}\}$ are drawn jointly from the respective distribution. Accordingly, there is no direct sampling from $\Pr(s_{t+1} | s_t^{(i)})$. This leads to the following particle filter:

1. Initialization, $t = 1 (i = 1, \ldots, I)$:

   a. Generate $s_1^{(i)} \sim \text{Multinomial}(p_1)$, with $p_1$ the initial distribution of $P$.
   b. Generate $\alpha_1^{(i)} \sim \mathcal{N}(0, \sigma_{s_1^{(i)}}^2 / (1 - \phi_{s_1^{(i)}}^2))$ from the initial distribution.
   c. Generate $z_1^{(i)} \sim IG(\nu/2, \nu/2)$.
   d. Compute $w_1^{(i)} = f(y_1 | x_1^{(i)})$, and let $\hat{f}(x_1^{(i)} | y_1) = w_1^{(i)} / \sum_{j=1}^{I} w_1^{(j)}$.

---

*A implementation is in stand alone C++ code developed by the author using the Scythe statistical library (Pemstein, Quinn, and Martin, 2011).*
2. Iterate, \( t = 1, \ldots, n - 1 \) \((i, j = 1, \ldots, I)\):

(a) Generate \( \{x_{t+1}^{(i)}, x_t^{(i)}\} \) from \( g(x_{t+1}, x_t^{(j)} | y_{1:t+1}) \) as follows, with \( x_t^{(j)} \) the current particle set at time \( t \). First, evaluate importance weights for all possible values of \( s_{t+1} = 1, \ldots, M \) (pre-weighting),

\[
u_t^{(j,k)} \propto g(x_t^{(j)}), s_{t+1} = k | y_{1:t+1}, \tilde{z}_{t+1}^{(j,k)}, \tilde{z},
\]

and resample \( \{x_t^{(j), k}, w_{t+1}^{(j,k)}\}_{(j,k) \in \{1, \ldots, I\} \times \{1, \ldots, M\}} \) to get equally weighted \( \{x_t^{(i)}, s_t^{(i)}, I^{-1}\} \). Then generate \( \alpha_t^{(i)} \) and \( z_{t+1}^{(i)} \) from the densities given in Eq. (A.18) and (A.19), respectively.

(b) Compute weights (correcting for the pre-weighting)

\[
w_{t+1}^{(i)} = \frac{f(y_{t+1} | x_t^{(i)}, y_{1:t+1})}{f(y_{t+1} | \tilde{a}_t^{(i)}, s_t^{(i)}, \tilde{z})},
\]

and let \( \hat{f}(x_t^{(i)} | y_{1:t+1}) = w_{t+1}^{(i)} / \sum_{j=1}^{I} w_{t+1}^{(j)} \).

Draws \( x_t^{(i)} \) obtained from Step 1 below are used to calculate posterior log-likelihood ordinate \( \log f_{\text{post}}(y_{1:n} | \theta) = \sum_{t=1}^{n} \log \hat{f}(y_t | y_{1:t}, \theta) \). Draws \( \{\alpha_t^{(i,k)}, z_{t+1}^{(i)}\} \) are deployed for simulating one-step-ahead prediction density \( f(y_{t+1} | y_{1:t}, \theta) \), which is needed for calculation of predictive ordinate \( \log f_{\text{prior}}(y_{1:n} | \theta) = \sum_{t=1}^{n} \log \hat{f}(y_t | y_{1:t-1}, \theta) \), forecasting, and model diagnostics. Simulation of the one-step-ahead prediction density is now summarized by the following steps \((i, j = 1, \ldots, I)\):

1. Resample from \( \hat{f}(x_t^{(j)} | y_{1:t}) \) after Step 1.(d) and 2.(b) to obtain \( x_t^{(i)} \).

2. Draw a state independent random variate \( N(0,1) \) to calculate \( \{\alpha_t^{(i,k)}\}_{k=1,\ldots,M} \), see Eq. (A.18).

3. Draw \( z_{t+1}^{(i)} \sim IG(\nu/2,\nu/2) \).

4. Estimate the one-step-ahead prediction density by the mixture

\[
\hat{f}(y_{t+1} | y_{1:t}) = I^{-1} \sum_{i=1}^{I} \sum_{k=1}^{M} \Pr(s_{t+1} = k | s_t^{(i)}) f(y_{t+1} | \alpha_{t+1|t}^{(i,k)}, z_{t+1}^{(i)}; s_{t+1} = k).
\]
Diebold, Gunther, and Tay (1998) show that the correct predictive density is weakly superior to all other forecasts, i.e. will be preferred, in terms of expected loss, by all forecast users regardless of their loss function. In the current context, however, the predictive ability of the model w.r.t. realized volatility is not of interest. Instead, RV is treated as an auxiliary variable that is supposed to improve prediction of the return distribution, on which the focus in modeling lies.

Testing whether the forecasting densities are correct can be done by exploiting the properties of the probability integral transform. The probability that \( y_{1,t+1}^2 \) will be less than observed \( y_{1,t+1}^o \) conditional on \( y_{1:t} \) can be estimated as

\[
\Pr(y_{1,t+1}^2 \leq y_{1,t+1}^o | y_{1:t}) \equiv \hat{u}_{t+1} = \int \sum_{i=1}^{I} \sum_{k=1}^{M} \Pr(s_{t+1} = k | s_t^{(i)}) \Pr(y_{1,t+1}^2 \leq y_{1,t+1}^o | \alpha_{t+1}^{(i,k)}, z_{t+1}^{(i)}, s_{t+1} = k).
\]

Under the null hypothesis of a correctly specified model, observed returns \( y_{1,t}^o \) are random draws from observation density \( \mathcal{N}(y_{1,t} | \cdot) \), compare Eq. (A.7), and \( \hat{u}_t \) converges in distribution to independent and identically distributed uniform random variables as \( I \to \infty \) (Rosenblatt 1952). These variables can be mapped into the normal distribution, by using the inverse of the normal distribution function, to give a standard sequence \( \hat{n}_t \) of independently and identically distributed normal random variables, which are then transformed one-step-ahead forecasts normed by their correct standard errors. This provides a valid basis for diagnostic checking and was popularized by KSC, and in another context by Diebold et al. (1998), among others.

**A.5 Simulation Study**

Regime switching data is simulated using parameters reflecting empirical relevant values to show that the proposed sampler works well. Moreover, sensitivity of skewness and kurtosis estimates regarding prior choice for degrees of freedom parameter \( \nu \) is investigated. Specifically, two regimes are simulated, one showing higher volatility and
a negative skew, the other lower volatility and a positive skew. True parameter values are

\[
\mu = (1, -1)', \quad \phi = 0.95, \quad \sigma = 0.15, \quad \rho = -0.5, \quad 
\gamma = (-1, 1)', \quad \nu = 30, 
\zeta = 0, \quad \xi = 1, \quad \sigma_u = 0.4, 
\]

\[
P = \begin{pmatrix} 0.50 & 0.50 \\ 0.20 & 0.80 \end{pmatrix}. 
\]

Regime switching is restricted to volatility mode \( \mu_i \) and skew parameter \( \gamma_i, i = \{1, 2\} \), for simplicity. Regime 1 is rather transitory, whereas regime 2 shows moderate persistence, resulting in approximately one third of the data belonging to regime 1 and two third to regime 2. The underlying volatility process exhibits common high persistence and leverage is present. A rather extreme choice of the different volatility modes (they are modeled on the log-scale) keeps false state assignment of observations low, demonstrating most clearly well-functioning of the sampler. Of course, less distinct regime characteristic parameters make a correct state assignment of observations increasingly difficult and hence, inference on true parameter values. Consider in this context Lo and Müller (2010) for a reflective discussion on limits of system identification and uncertainty in general.

The following independent prior distributions are assumed

\[
\mu \sim N((0.5, -0.5)', I_2), \quad \frac{\phi + 1}{2} \sim B(20, 1.5), 
\lambda \sim IG \left( \frac{5}{2}, \frac{0.05}{2} \right), \quad \alpha \sim N(0, 10^{12}), 
\gamma \sim N(0, 10^{12} I_2), \quad \nu \sim G(16, 0.8)_{\nu > 4}, 
\]

\[
(\zeta, \xi)' \sim N((0, 1)', \text{diag}(10^2, 1)), \quad \sigma_u^2 \sim IG \left( \frac{5}{2}, \frac{0.3}{2} \right), 
\]

\[
P_i \sim D(1, 1), 
\]

where \( B(\cdot) \) denotes the beta and \( G(\cdot) \) the gamma distribution. Operator \( \text{diag}(\cdot) \) creates a diagonal matrix. The mildly informative priors on \( \mu \) reflect the regime identification constraint into high and low volatility state \( (\mu_1 > \mu_2) \) to prevent label switching (see Sec. A.3, Step 10). The priors for \( \phi \) and \( \lambda \) are identical to those used by KSC, the latter
in the case of zero correlation. Then, volatility of volatility $\sigma$ has implied mean 0.12 and standard deviation 0.05 (moments obtained by MC simulation). Accordingly, the prior on Cholesky reparameterization parameter $a$ implies a diffuse belief in zero correlation. The prior on $\phi$ mirrors a belief in moderate volatility persistence with mean 0.86 and standard deviation 0.11. For skew related parameter vector $\gamma$ a diffuse prior centered around zero is chosen. A possible motivation for such an uninformative prior would be the desire to "let the data speak for itself". Moreover, as will be demonstrated below, prior choice of degrees of freedom parameter $\nu$ influences higher moment estimates. Hence the choice of a very weak prior for $\gamma$ to isolate this effect. Then, the prior for degrees of freedom parameter $\nu$ is mildly informative, having mean 20.0 and standard deviation 5.0, reflecting a conservative belief in the existence of tail risk. Priors on $\zeta$ and $\xi$ assume no bias correction for RV and unity scaling, respectively. The prior on $\sigma_u^2$ implies a $\sigma_u$ with mean 0.29 and standard deviation 0.12 (moments obtained by MC simulation), reflecting empirical relevant values. Finally, priors for each row of $P$ are uniform over the unit simplex, expressing no specific prior belief in regime persistence.

50 datasets are simulated, each containing 2,000 observations. McMC is iterated for 15,000 iterations, discarding the first 5,000 as burn-in, and collecting the next 10,000 draws for summary statistics. In the multi-move sampler, average block size is set to 100 and number of iterations to achieve convergence to 3.

Tab. A.1 reports grand averages, standard deviations, minima, and maxima of the posterior mean estimates, and average inefficiency factors (IF) (see App. A.C for the deployed formulae). Except for higher moment parameters $\gamma$ and $\nu$, grand averages are very close to their true values. Observe a higher standard deviation of volatility mode $\mu_1$ compared to $\mu_2$ and especially of skew parameter $\gamma_1$ to $\gamma_2$. Of course these are a due to less observations in state 1 (on average, 569.1/1,430.9 data points are assigned to regime 1/2), but it is clearly visible how an increase in observations leads to a stronger improvement in precision for more difficult to estimate skew parameter $\gamma$. Note also that the maximum estimated value of $\gamma_1$ is actually positive. More importantly, however, higher moment parameters $\gamma$ and $\nu$ are clearly biased downwards in absolute value.

Prior sensitivity of degrees of freedom parameter $\nu$ for the SV model with Student-$t$ error is known, see e.g. Nakajima and Omori (2009). For the SV model with
Table A.1: Summary Output of Simulated Data from the MS2SVskt-RV Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True</th>
<th>Mean</th>
<th>Stdev</th>
<th>Min</th>
<th>Max</th>
<th>Avg. IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>1</td>
<td>0.9731</td>
<td>0.1010</td>
<td>0.7346</td>
<td>1.1865</td>
<td>16.9</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-1</td>
<td>-1.0045</td>
<td>0.0908</td>
<td>-1.2300</td>
<td>-0.7400</td>
<td>8.8</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.95</td>
<td>0.9474</td>
<td>0.0084</td>
<td>0.9287</td>
<td>0.9620</td>
<td>7.3</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.15</td>
<td>0.1490</td>
<td>0.0107</td>
<td>0.1232</td>
<td>0.1680</td>
<td>70.5</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.5</td>
<td>-0.4822</td>
<td>0.0564</td>
<td>-0.6038</td>
<td>-0.3333</td>
<td>21.4</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-1</td>
<td>-0.7353</td>
<td>0.3873</td>
<td>-1.5266</td>
<td>0.1868</td>
<td>54.2</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>1</td>
<td>0.7498</td>
<td>0.2873</td>
<td>0.1150</td>
<td>1.3941</td>
<td>73.4</td>
</tr>
<tr>
<td>$\nu$</td>
<td>30</td>
<td>23.825</td>
<td>1.9319</td>
<td>18.786</td>
<td>27.844</td>
<td>141.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True</th>
<th>Mean</th>
<th>Stdev</th>
<th>Min</th>
<th>Max</th>
<th>Avg. IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>skewness$_1$</td>
<td>-0.2374</td>
<td>-0.2346</td>
<td>0.1231</td>
<td>-0.4662</td>
<td>0.0555</td>
<td>40.2</td>
</tr>
<tr>
<td>skewness$_2$</td>
<td>0.2374</td>
<td>0.2352</td>
<td>0.0832</td>
<td>0.0517</td>
<td>0.4587</td>
<td>43.4</td>
</tr>
<tr>
<td>kurtosis$_1$</td>
<td>0.3545</td>
<td>0.5206</td>
<td>0.1413</td>
<td>0.3179</td>
<td>0.8434</td>
<td>52.0</td>
</tr>
<tr>
<td>kurtosis$_2$</td>
<td>0.3545</td>
<td>0.4834</td>
<td>0.0905</td>
<td>0.3216</td>
<td>0.8445</td>
<td>64.9</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>0</td>
<td>0.0247</td>
<td>0.0596</td>
<td>-0.1159</td>
<td>0.1425</td>
<td>91.9</td>
</tr>
<tr>
<td>$\xi$</td>
<td>1</td>
<td>1.0152</td>
<td>0.0438</td>
<td>0.9366</td>
<td>1.1033</td>
<td>135.7</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.4</td>
<td>0.4007</td>
<td>0.0089</td>
<td>0.3814</td>
<td>0.4213</td>
<td>4.5</td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>0.5</td>
<td>0.5027</td>
<td>0.0208</td>
<td>0.4633</td>
<td>0.5479</td>
<td>1.2</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.5</td>
<td>0.4973</td>
<td>0.0208</td>
<td>0.4521</td>
<td>0.5367</td>
<td>1.2</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>0.2</td>
<td>0.1981</td>
<td>0.0098</td>
<td>0.1772</td>
<td>0.2221</td>
<td>1.2</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.8</td>
<td>0.8019</td>
<td>0.0098</td>
<td>0.7779</td>
<td>0.8228</td>
<td>1.2</td>
</tr>
</tbody>
</table>

True value, mean, standard deviation, minimum, maximum of the posterior mean estimates, and average inefficiency factor across 50 replications.

$\mathcal{GH}$ skew-$t$ distributed error, Nakajima and Omori (2012) show that skew parameter $\gamma$ is also affected by the prior choice of $\nu$. As skewness and kurtosis are determined jointly by parameters $\gamma$ and $\nu$, if the posterior mean of $\nu$ decreases, then $\gamma$ has to decrease in absolute value as to maintain skew and heavy tailedness of the empirical return distribution. This is also observed here. Moreover, if we consider the average implied skewness and kurtosis of the estimated distribution, we infer that whereas true skewness is precisely extracted, kurtosis is overestimated. This justifies the conservative perception of the prior choice for $\nu$ in this stylized example. To investigate the point further, an alternative prior for $\nu$ is chosen,

Prior #2: $\nu \sim \mathcal{G}(18, 0.6) I_{\nu>4},$

which is centered at the true value having mean 30.0 and standard deviation 7.1. Again 50 datasets are generated, as above. Results are reported in Tab. A.2. Both skew and
Table A.2: Output of Simulated Data from the MS2SVskt-RV Model, $\nu$ Prior #2

<table>
<thead>
<tr>
<th></th>
<th>True Mean</th>
<th>Stdev</th>
<th>Min</th>
<th>Max</th>
<th>Avg. IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>-1</td>
<td>-1.0315</td>
<td>0.4967</td>
<td>-2.2501</td>
<td>0.2560</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>1</td>
<td>1.0424</td>
<td>0.4052</td>
<td>0.1100</td>
<td>1.8217</td>
</tr>
<tr>
<td>$\nu$</td>
<td>30</td>
<td>33.031</td>
<td>2.7021</td>
<td>26.000</td>
<td>39.284</td>
</tr>
<tr>
<td>skewness$_1$</td>
<td>-0.2374</td>
<td>-0.2185</td>
<td>0.1024</td>
<td>-0.4241</td>
<td>0.0514</td>
</tr>
<tr>
<td>skewness$_2$</td>
<td>0.2374</td>
<td>0.2188</td>
<td>0.0786</td>
<td>0.0314</td>
<td>0.3905</td>
</tr>
<tr>
<td>kurtosis$_1$</td>
<td>0.3545</td>
<td>0.3726</td>
<td>0.0977</td>
<td>0.2340</td>
<td>0.6027</td>
</tr>
<tr>
<td>kurtosis$_2$</td>
<td>0.3545</td>
<td>0.3508</td>
<td>0.0722</td>
<td>0.2236</td>
<td>0.5594</td>
</tr>
</tbody>
</table>

True value, mean, standard deviation, minimum, maximum of the posterior mean estimates, and average inefficiency factor across 50 replications.

Kurtosis parameters are now located satisfactorily close to their true values. Specifically, kurtosis is recovered with adequate precision.

To conclude, in real world applications, some subjectivity remains as we have to impose a belief on what true kurtosis is. Handling this additional freedom in a responsible way, the researcher may conservatively set the prior on $\nu$ sufficiently low to guarantee some excess kurtosis in the data.

Referring back to the results in Tab. A.1, average inefficiencies ($IF$) are highest for degrees of freedom parameter $\nu$ and auxiliary equation bias and scale parameters $\xi$ and $\zeta$. Average AR/MH acceptance rates in the multi-move sampler are high regarding the rather large block size used in this study, 90.8% and 90.5%, respectively. This can be attributed to a positive effect of the auxiliary time series for inference on latent volatility. Acceptance probabilities of the MH steps sampling volatility persistence $\phi$, Cholesky reparameterization parameters {$\lambda, a$}, degrees of freedom parameter $\nu$, and scaling variable $z_t$ are on average 95.6%, 99.1%, 98.0% and 86.4%, respectively. The above results suggest good mixing properties of the proposed MCMC algorithm.

---

6For the SV model with symmetric Student-t error and leverage, Jacquier et al. (2004) sample $\nu$ over a discrete grid, applying an uniform prior. They draw $\pi(z_{1:n}, \nu|\cdot)$ as a block, $\pi(z_{1:n}|\nu, \cdot)\pi(\nu|\cdot)$, rather than with a Gibbs cycle $[\pi(z_{1:n}|\nu, \cdot), \pi(\nu|z_{1:n}, \cdot)]$, increasing efficiency. However, a marginalization over $z_{1:n}$ is not applicable in the current asymmetric context.
A.6 Empirical Application

A.6.1 Data

Proposed models are differentiated along the attributes fat tails, skew, regimes and auxiliary observation variable. S&P 500 daily return data from Thomson Reuters Datastream is used, spanning the period from Jan-02-1990 to Aug-06-2010, a total of 5,164 observations or roughly 20 years. As auxiliary series, realized volatility, the range, and the VIX are deployed.\footnote{The CBOE Volatility Index\textsuperscript{©} measures market expectations of near-term volatility conveyed by S&P 500 stock index option prices. For more information, see www.cboe.com.} The latter dictates the start date of the sample and has been downloaded from the CBOE (Chicago Board of Exchange) web page. Realized volatility is calculated from tick-by-tick prices of S&P 500 futures traded "open outcry" on the CBOE. I gratefully acknowledge provision of the dataset by Francesco Audrino, Faculty of Mathematics and Statistics, University of St. Gallen.

Returns are continuously compounded, $100 \times (\log P_t - \log P_{t-1})$, with $P_t$ the closing price on day $t$, and demeaned. Logarithmized realized volatility is calculated following Eq. (A.6), using 5-min returns of the contract closest to expiry minimum 1 month, applying the previous tick rule, and discarding observations prior to the open (8.30 AM) and after the close (15.15 PM).\footnote{The basic 5-min RV estimator can be taken as a baseline, against which the range and VIX are measured, having in mind that more elaborated estimators like e.g. TSRV exist.} The logarithmized range is calculated as $y_{2,t} = \log(P_{\text{high},t}/P_{\text{low},t-1})$.\footnote{A cut-off of $10^{-3}$ is applied before logarithmizing to prevent a negative numerical overflow. However, this constraint is never binding for the current dataset.} Finally, the VIX is scaled by $252^{-1/2}$ to approximately express expected daily movement of the S&P 500 in % over the next 30-day period and logarithmized.

Descriptive statistics of the datasets are reported in Tab. A.3. Considering returns, we observe common negative skew and fat tails. Note that the largest absolute return in the analyzed period is actually positive. RV, the range and the VIX all exhibit positive skew and large kurtosis, with RV the most pronounced. After logarithmizing, however, these series get much closer to normal. RV fluctuates the most and the VIX the least, as indicated by respective standard deviations, minimum and maximum values. By inspection of Fig A.1, the range appears to be the least persistent (or most erratic)
Table A.3: Descriptive Statistics of the S&P 500 (90/01-02-10/08/06, 5,164 obs.)

<table>
<thead>
<tr>
<th></th>
<th>Return</th>
<th>Return^2</th>
<th>RV</th>
<th>log-RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0210</td>
<td>1.3880</td>
<td>1.0545</td>
<td>-0.5323</td>
</tr>
<tr>
<td>Stdev</td>
<td>1.1781</td>
<td>4.5657</td>
<td>2.1640</td>
<td>0.9713</td>
</tr>
<tr>
<td>Skew</td>
<td>-0.1970</td>
<td>12.728</td>
<td>10.904</td>
<td>0.5662</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>11.847</td>
<td>239.89</td>
<td>198.59</td>
<td>3.6168</td>
</tr>
<tr>
<td>Min</td>
<td>-9.4695</td>
<td>0.0000</td>
<td>0.0331</td>
<td>-3.4087</td>
</tr>
<tr>
<td>Max</td>
<td>10.957</td>
<td>120.06</td>
<td>62.460</td>
<td>4.1345</td>
</tr>
<tr>
<td>LB(15)†</td>
<td>78.7</td>
<td>5,637.1</td>
<td>20,065.6</td>
<td>37,842.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Range</th>
<th>log-Range</th>
<th>VIX</th>
<th>log-VIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.3518</td>
<td>0.1103</td>
<td>1.2837</td>
<td>0.1818</td>
</tr>
<tr>
<td>Stdev</td>
<td>1.0028</td>
<td>0.5999</td>
<td>0.5248</td>
<td>0.3570</td>
</tr>
<tr>
<td>Skew</td>
<td>3.3743</td>
<td>0.2792</td>
<td>2.0115</td>
<td>0.5372</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>23.067</td>
<td>3.1681</td>
<td>10.104</td>
<td>3.2175</td>
</tr>
<tr>
<td>Min</td>
<td>0.1776</td>
<td>-1.7283</td>
<td>0.5865</td>
<td>-0.5336</td>
</tr>
<tr>
<td>Max</td>
<td>11.5208</td>
<td>2.4442</td>
<td>5.0937</td>
<td>1.6280</td>
</tr>
<tr>
<td>LB(15)†</td>
<td>22,816.8</td>
<td>20,462.9</td>
<td>66,370.9</td>
<td>67,826.0</td>
</tr>
</tbody>
</table>

† Critical value for the LB(15) test is 30.6 at the 1% significance level.

and the VIX the smoothest measure, with RV in-between. The period of significantly increased volatility during the financial crisis 2008/09 is remarkable.

### A.6.2 The MS3SVskt Model

This section covers the SVskt model with skewed observation error but no auxiliary measurement equation and extends it to a regime switching MS3SVskt model featuring three states. Variants with Student-\( t \) and Gaussian error are reported in App. A.E, as for the models treated in Sec. A.6.3-A.6.5. Priors are those of Sec. A.5, except the following. Regarding the volatility mode of the SVskt model, \( \mu_0 = -0.5, \sigma_{\mu_0} = 1 \). For the regime switching MS3SVskt variant, a mildly informative prior reflects the specific state identification constraint \( \mu_1 > \mu_2 > \mu_3 \),

\[
\mu_0 = (1.5, 0, -1)', \quad \Sigma_{\mu_0} = I_3.
\]

Regarding skew parameter \( \gamma \) more centered priors are chosen,

\[
\gamma \sim \mathcal{N}(0, 1), \quad \gamma \sim \mathcal{N}(0, I_3),
\]

(A.22)

the reason being better predictive ability of the models, see Sec. A.6.7 for a prior
sensitivity analysis. Further, a moderate belief of state persistence is imposed on the transition probability matrix via the priors

\[ P_1 \sim \mathcal{D}(30, 1.5, 1.5), \quad P_2 \sim \mathcal{D}(1.5, 30, 1.5), \quad P_3 \sim \mathcal{D}(1.5, 1.5, 30), \]

implying a probability of about 90.9% to remain in the current state and of 4.5% to switch to another state.

For both models, the first 5,000 draws are discarded as burn-in, collecting the following 20,000 draws for parameter inference. In the multi-move sampler, average block size as a tuning parameter is set to 25 and number of iterations to achieve convergence to 5. Posterior means, standard deviations, 95% credible bounds, inefficiency factors (\(IF\)), and p-values of Geweke’s (1992) convergence diagnostic (\(CD\)) (see App. A.C for the deployed formulae) are reported.
Table A.4: Estimation Results of the SVskt Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>-0.3959</td>
<td>0.1198</td>
<td>[-0.6341, -0.1624]</td>
<td>2.5</td>
<td>0.86</td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.9872</td>
<td>0.0021</td>
<td>[0.9827, 0.9910]</td>
<td>11.9</td>
<td>0.67</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.1553</td>
<td>0.0111</td>
<td>[0.1347, 0.1785]</td>
<td>163.6</td>
<td>0.44</td>
</tr>
<tr>
<td>( \rho )</td>
<td>-0.7771</td>
<td>0.0306</td>
<td>[-0.8333, -0.7136]</td>
<td>101.3</td>
<td>0.97</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>-0.6207</td>
<td>0.1752</td>
<td>[-1.0145, -0.3286]</td>
<td>161.0</td>
<td>0.50</td>
</tr>
<tr>
<td>( \nu )</td>
<td>20.734</td>
<td>3.3102</td>
<td>[14.980, 27.734]</td>
<td>280.2</td>
<td>0.61</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.2349</td>
<td>0.0488</td>
<td>[-0.3309, -0.1404]</td>
<td>68.8</td>
<td>0.75</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.5088</td>
<td>0.1000</td>
<td>[0.3368, 0.7302]</td>
<td>126.4</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke’s convergence diagnostic (p-value). Lower panel: Implied higher moments of \( \epsilon_t \).

Analyzing results of the SVskt model in Tab. A.4 first, we observe common high persistence of specifically \( \phi = 0.987 \) and a strong leverage effect \( \rho = -0.777 \). Skew parameter \( \gamma \) is significantly negative, resulting in a skewness of \(-0.235 \) and excess kurtosis of \( 0.510 \) for observation error \( \epsilon_t \). The null of the CD statistic (“chain has converged”) can not be rejected for all parameters at conventional significance levels, as it can not for all parameters of the remaining models in this work.

The SVskt is extended to the MS3SVskt model, containing three regimes, with state 1 the rare event regime. However, stabilizing the latter was rather difficult (in the sense of containing a stable portion of observations during the McMC run, gauged by visual inspection), consequently leaving volatility persistence \( \phi \), volatility of volatility \( \sigma \), and leverage \( \rho \) constant across states. Motivated by similar concerns, only one skew parameter \( \gamma_{12} \) is estimated for states 1 and 2 with higher volatility. Moreover, relatively few observations in the rare event state (\( \approx 216 \) obs., see Eq. (A.23)) would make it difficult to obtain a more precise estimate of higher moment parameter \( \gamma_1 \). Note further that a Gaussian MS3SV version of the model could not be estimated without imposing a rather strong informative prior on the volatility mode of the rare event state, preventing a state collapse in that way.\(^{10}\) Estimates are not reported for this variant.

Inspecting results in Tab. A.5, volatility modes of the regimes are significantly different, see also Fig. A.2 for a graphical visualization of selected posterior probability

\(^{10}\)Consider in this context e.g. Bulla (2011), who observes that a fat tailed distribution like the Student-\( t \) can lead to an increased persistence of states.
### A.6. EMPIRICAL APPLICATION

#### Table A.5: Estimation Results of the MS3SVskt Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>1.5935</td>
<td>0.2731</td>
<td>[1.0189, 2.1045]</td>
<td>57.7</td>
<td>0.91</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.0940</td>
<td>0.0776</td>
<td>[-0.0632, 0.2371]</td>
<td>22.4</td>
<td>0.88</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-1.0971</td>
<td>0.0815</td>
<td>[-1.2628, -0.9411]</td>
<td>19.5</td>
<td>1.00</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9532</td>
<td>0.0080</td>
<td>[0.9357, 0.9673]</td>
<td>62.4</td>
<td>0.85</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.1903</td>
<td>0.0160</td>
<td>[0.1625, 0.2224]</td>
<td>295.7</td>
<td>0.82</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.8005</td>
<td>0.0312</td>
<td>[-0.8609, -0.7369]</td>
<td>133.9</td>
<td>0.81</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>-0.7706</td>
<td>0.2827</td>
<td>[-1.3934, -0.2900]</td>
<td>160.8</td>
<td>0.98</td>
</tr>
<tr>
<td>$\nu$</td>
<td>22.621</td>
<td>4.0860</td>
<td>[16.284, 31.638]</td>
<td>322.2</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke’s convergence diagnostic (p-value). Lower panel: Implied higher moments of $\epsilon_t$. Transition probabilities in App. A.E, Tab. A.21.

Volatility persistence $\phi = 0.953$ is significantly lower than in the SVskt model. Interestingly, skew parameter $\gamma_3 = -0.771$ of the low volatility state is considerably more negative than $\gamma_{12} = -0.475$ of the high volatility states. Although not significantly different from a classical viewpoint, modes of the respective posterior probability densities appear rather well separated. One may thus conjecture that more data supports different skewness across regimes.

![Figure A.2: MS3SVskt - Selected Posterior Probability Densities](image-url)
Persistence is high in all regimes:

\[
\begin{array}{ccc}
\text{Vol} & \text{high} & \text{low} \\
\hline
s_t & 1 & 2 & 3 \\
1 & 0.9766 & 0.0155 & 0.0079 \\
s_{t-1} & 2 & 0.0014 & 0.9970 & 0.0016 \\
3 & 0.0009 & 0.0020 & 0.9971 \\
\end{array}
\]

\text{(A.23)}

An observed high persistence of 97.7% for the rare event state is rather uncommon but readily explained by Fig. A.3, which shows the states’ posterior probability for the complete sample period. Division into the different states appears almost deterministic and most notably, the meltdown of the recent financial crisis 2008/09 is modeled by its own prolonged rare event state. This emphasizes the distinct nature of that event in recent history.

To put the analysis in a macroeconomic context, NBER US business cycle turning points are additionally graphed. Related dates are reported in Tab. A.6. Several regime shifts match turning points (first trading day of month is taken) rather close. Most notably the trough in June 2009, where the extremely high volatility period ends, but also the peak in July 1990 and subsequent trough in March 1991, which encapsulate a high volatility period. The December 2007 peak is also reached quite soon after a shift into high volatility has been observed. Only the short recession from March to November 2001 remains undetected by the model, but it is consistently located in a high volatility period. These overlaps are remarkably, as the current analysis is purely technical but the NBER bases its decision on multiple fundamental criteria. Moreover, announcement dates of the July 1990 peak and November 2001 trough, but also of the December 2007 peak coincide with or are rather close to volatility regime shifts in the S&P 500. Especially the coincidence of the trough announcement in July 2003 with the shift to a low volatility period after more than 6 years is striking. One may suggest a significant influence of the NBER announcement on market sentiment in this case.
Figure A.3: MS3SVskt - Posterior Probability of States (1990/01/02-2010/08/06). Background: S&P 500 (log-scale, right). Vertical lines relate to NBER US business cycles (solid: Turning points; dashed: Announcement dates; see Tab. A.6 for details)

Table A.6: NBER US Business Cycle Expansions and Contractions

<table>
<thead>
<tr>
<th>Turning Point Date</th>
<th>Peak or Trough</th>
<th>Duration (months)</th>
<th>Announcement Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 2009</td>
<td>Trough</td>
<td>18</td>
<td>September 20, 2010</td>
</tr>
<tr>
<td>December 2007</td>
<td>Peak</td>
<td>73</td>
<td>December 1, 2008</td>
</tr>
<tr>
<td>November 2001</td>
<td>Trough</td>
<td>8</td>
<td>July 17, 2003</td>
</tr>
<tr>
<td>March 2001</td>
<td>Peak</td>
<td>120</td>
<td>November 26, 2001</td>
</tr>
<tr>
<td>March 1991</td>
<td>Trough</td>
<td>8</td>
<td>December 22, 1992</td>
</tr>
<tr>
<td>July 1990</td>
<td>Peak</td>
<td>92</td>
<td>April 25, 1991</td>
</tr>
</tbody>
</table>

A.6.3 The MS4SVskt-RV Model

This section covers the SVskt-RV model with realized volatility as auxiliary measurement variable and a regime switching four state MS4SVskt-RV variant. Priors are as in Sec. A.5 and Eq. (A.22), except for the following. Regarding the volatility mode of the SVskt-RV model, $\mu_0 = -0.5$, $\sigma_{\mu_0} = 1$. For the regime switching MS4SVskt-RV extension, a mildly informative prior reflects the specific state identification constraint $\mu_1 > \mu_2 > \mu_3 > \mu_4$,

$$\mu_0 = (1, 0, -0.5, -1.5)', \quad \Sigma_{\mu_0} = I_4.$$ 

Burn-in is 5,000 draws for the SVskt-RV and 10,000 for the MS4SVskt-RV model. The next 20,000 draws are recorded for parameter inference. Average block size and number of iterations to achieve convergence in the multi-move sampler are as in Sec. A.6.2.

Consider first the SVskt-RV model, Tab. A.7. Relative to the SVskt model with no auxiliary equation, we observe lower persistence $\phi = 0.970$, higher volatility of volatility $\sigma = 0.221$, and lower leverage $\rho = -0.466$, all differences significant. Moreover, error term $\epsilon_t$ is much less negatively skewed. Although the 95% credible bounds of skew parameter $\gamma = -0.262$ contain zero, most of the posterior probability mass is located in the negative domain, consider Fig. A.4. Bias correction term $\zeta$ is negative, indicating a dominance of non-trading hours over market microstructure effects. Scaling parameter $\xi$ lies in the vicinity of unity, as would be expected theoretically. The important positive constant $\sigma/\sigma_u = 54.1\%$ is the signal-to-noise

![Skew Parameter $\gamma$](image)

**Figure A.4:** SVskt-RV - Selected Posterior Probability Densities
### A.6. EMPIRICAL APPLICATION

Table A.7: Estimation Results of the SVskt-RV Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.2698</td>
<td>0.0928</td>
<td>[-0.4514, -0.0864]</td>
<td>1.5</td>
<td>0.91</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9694</td>
<td>0.0034</td>
<td>[0.9625, 0.9757]</td>
<td>4.5</td>
<td>0.94</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.2207</td>
<td>0.0096</td>
<td>[0.2025, 0.2402]</td>
<td>54.0</td>
<td>1.00</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.4661</td>
<td>0.0292</td>
<td>[-0.5228, -0.4080]</td>
<td>22.8</td>
<td>0.96</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.2623</td>
<td>0.1857</td>
<td>[-0.6465, 0.0820]</td>
<td>53.3</td>
<td>0.88</td>
</tr>
<tr>
<td>$\nu$</td>
<td>28.771</td>
<td>5.0516</td>
<td>[20.132, 40.138]</td>
<td>215.1</td>
<td>0.96</td>
</tr>
<tr>
<td>skewness</td>
<td>-0.0663</td>
<td>0.0442</td>
<td>[-0.1514, 0.0207]</td>
<td>45.9</td>
<td>0.90</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.2667</td>
<td>0.0554</td>
<td>[0.1749, 0.3913]</td>
<td>180.7</td>
<td>0.97</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>-0.2159</td>
<td>0.0233</td>
<td>[-0.2616, -0.1710]</td>
<td>50.4</td>
<td>0.86</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.9040</td>
<td>0.0213</td>
<td>[0.8634, 0.9478]</td>
<td>41.5</td>
<td>0.99</td>
</tr>
<tr>
<td>$\sigma_\nu$</td>
<td>0.4083</td>
<td>0.0059</td>
<td>[0.3969, 0.4200]</td>
<td>7.7</td>
<td>0.99</td>
</tr>
<tr>
<td>SNR</td>
<td>0.5407</td>
<td>0.0276</td>
<td>[0.4889, 0.5973]</td>
<td>47.4</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke’s convergence diagnostic (p-value). Second panel: Implied higher moments of $\epsilon_t$. Third panel: Parameters auxiliary measurement equation. Lower panel: Implied signal-to-noise ratio.

The signal-to-noise ratio (SNR) known from engineering, measuring the sustained system variance relative to the ephemeral observation variance (see e.g. West and Harrison, 1997).

Turning to the four state MS4SVskt-RV variant, some experimentation was necessary to find an appropriate parameter configuration, as initial runs with the full regime switching parameter set resulted in a quite unstable state partition (especially regarding the state with lowest volatility, gauged by visual inspection). The first measure was then to make leverage $\rho$ regime independent, as posterior distributions of this parameter overlapped in large part.\textsuperscript{11} Moreover, the skew centered around zero for some states, which appeared to destabilize the system when switching continuously between positive and negative values during the McMC run. Consequently, several variants were investigated in the search for parameter configurations yielding stable equilibria, i.e. merging the skew parameter of different states or setting it to zero. Further, setting volatility persistence and volatility of volatility regime invariant. Then

\textsuperscript{11}Step 11 of the McMC algorithm has to be modified, as the combination $\{\sigma, \rho\}$ can not be sampled using the proposed Cholesky decomposition. Instead, a MH algorithm similar to Step 6 is applied, sampling $\{\sigma, \rho\}$ jointly.
the *DIC* (see Sec. A.6.6) was deployed as decision criterion to come up with the proposed parameter configuration.

Analyzing estimation results of Tab. A.8, the identified four states show significantly different volatility modes, see also Fig. A.5. Persistence levels appear distinct, with high and low volatility states 1/4 (which make up only for 5.6% of the data, see Eq. (A.24)) featuring lower persistence. Roughly speaking, volatility of volatility is separated between high 1/2 and low 3/4 volatility states. Observing a larger transition error for the high volatility states is rather intuitive. Observation error $\sigma_u$ of realized volatility is distinctively high for state 1. The signal-to-noise ratios indicate that RV as an observable proxy for latent volatility is most informative in state 2, which is the second most frequent state with 17.5% of the observations. Moreover, compared to the SNR of the SVskt-RV model, states 1/3/4 show a lower and state 2 a higher SNR. Interestingly, skew of medium volatility states 2/3 is now significantly positive as compared to the SVskt-RV model.

It is important to note that skewness and excess kurtosis of the observation error are lowest for the RV class compared to all other models estimated in this work.
### Table A.8: Estimation Results of the MS4SVskt-RV Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>0.8775</td>
<td>0.1681</td>
<td>[0.5515, 1.2142]</td>
<td>24.6</td>
<td>0.99</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.2082</td>
<td>0.1351</td>
<td>[-0.0464, 0.4866]</td>
<td>11.6</td>
<td>0.80</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.4356</td>
<td>0.1250</td>
<td>[-0.6711, -0.1805]</td>
<td>8.3</td>
<td>0.98</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>-1.5771</td>
<td>0.1549</td>
<td>[-1.8855, -1.2675]</td>
<td>21.9</td>
<td>0.68</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.9353</td>
<td>0.0361</td>
<td>[0.8542, 0.9911]</td>
<td>15.5</td>
<td>0.90</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.9839</td>
<td>0.0102</td>
<td>[0.9597, 0.9982]</td>
<td>13.2</td>
<td>0.78</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.9860</td>
<td>0.0032</td>
<td>[0.9797, 0.9924]</td>
<td>12.0</td>
<td>0.96</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.9078</td>
<td>0.0422</td>
<td>[0.8174, 0.9819]</td>
<td>50.6</td>
<td>0.88</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.2659</td>
<td>0.0606</td>
<td>[0.1500, 0.3858]</td>
<td>69.4</td>
<td>0.94</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.2300</td>
<td>0.0271</td>
<td>[0.1791, 0.2858]</td>
<td>90.6</td>
<td>0.43</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.1281</td>
<td>0.0100</td>
<td>[0.1093, 0.1488]</td>
<td>132.9</td>
<td>0.92</td>
</tr>
<tr>
<td>$\sigma_4$</td>
<td>0.1433</td>
<td>0.0425</td>
<td>[0.0783, 0.2461]</td>
<td>53.8</td>
<td>0.73</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.5149</td>
<td>0.0323</td>
<td>[-0.5760, -0.4507]</td>
<td>37.9</td>
<td>0.96</td>
</tr>
<tr>
<td>$\gamma_{23}$</td>
<td>0.4498</td>
<td>0.2353</td>
<td>[0.0145, 0.9535]</td>
<td>76.7</td>
<td>0.94</td>
</tr>
<tr>
<td>$\nu$</td>
<td>34.160</td>
<td>5.1186</td>
<td>[25.535, 45.217]</td>
<td>151.8</td>
<td>0.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Moment</th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness $\gamma_{23}$</td>
<td>0.0920</td>
<td>0.0453</td>
<td>[0.0031, 0.1845]</td>
<td>70.7</td>
<td>0.98</td>
</tr>
<tr>
<td>Kurtosis $\sigma_{23}$</td>
<td>0.2047</td>
<td>0.0347</td>
<td>[0.1455, 0.2786]</td>
<td>142.4</td>
<td>0.95</td>
</tr>
<tr>
<td>Kurtosis $\nu$</td>
<td>0.2277</td>
<td>0.0406</td>
<td>[0.1582, 0.3174]</td>
<td>115.6</td>
<td>0.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\zeta$</td>
<td>-0.1940</td>
<td>0.0254</td>
<td>[-0.2427, -0.1442]</td>
<td>85.2</td>
<td>0.99</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.8931</td>
<td>0.0188</td>
<td>[0.8561, 0.9306]</td>
<td>53.3</td>
<td>0.99</td>
</tr>
<tr>
<td>$\sigma_{u,1}$</td>
<td>0.6432</td>
<td>0.0640</td>
<td>[0.5282, 0.7803]</td>
<td>11.4</td>
<td>0.98</td>
</tr>
<tr>
<td>$\sigma_{u,2}$</td>
<td>0.3107</td>
<td>0.0296</td>
<td>[0.2543, 0.3712]</td>
<td>58.5</td>
<td>0.48</td>
</tr>
<tr>
<td>$\sigma_{u,3}$</td>
<td>0.2992</td>
<td>0.0091</td>
<td>[0.2807, 0.3162]</td>
<td>37.8</td>
<td>0.85</td>
</tr>
<tr>
<td>$\sigma_{u,4}$</td>
<td>0.3521</td>
<td>0.0461</td>
<td>[0.2660, 0.4461]</td>
<td>15.9</td>
<td>0.69</td>
</tr>
<tr>
<td>SNR$_1$</td>
<td>0.4184</td>
<td>0.1080</td>
<td>[0.2217, 0.6420]</td>
<td>47.6</td>
<td>0.95</td>
</tr>
<tr>
<td>SNR$_2$</td>
<td>0.7493</td>
<td>0.1286</td>
<td>[0.5274, 1.0298]</td>
<td>84.9</td>
<td>0.32</td>
</tr>
<tr>
<td>SNR$_3$</td>
<td>0.4285</td>
<td>0.0363</td>
<td>[0.3598, 0.5038]</td>
<td>108.8</td>
<td>1.00</td>
</tr>
<tr>
<td>SNR$_4$</td>
<td>0.4143</td>
<td>0.1382</td>
<td>[0.2138, 0.7552]</td>
<td>44.6</td>
<td>0.93</td>
</tr>
</tbody>
</table>


This implies that more dynamics are explained by the AR(1) volatility process, which appears to be a desirable feature especially regarding forecasting ability.
Persistence of regimes is considerably less than in the MS3SVskt Model:

\[
\begin{array}{cccc}
\text{Vol} & \text{high} & \text{low} \\
1 & 0.3512 & 0.3118 & 0.3218 & 0.0153 \\
2 & 0.0318 & 0.5479 & 0.4135 & 0.0068 \\
3 & 0.0198 & 0.0845 & 0.8747 & 0.0210 \\
4 & 0.0448 & 0.1424 & 0.5833 & 0.2294 \\
\end{array}
\]

\( s_{t-1} \) \( s_t \) \( s_{t-1} \) \( s_t \)

avg. observations

\[
\begin{array}{cccc}
170.6 & 901.2 & 3,976.0 & 116.2 \\
(3.3\%) & (17.5\%) & (77.0\%) & (2.2\%) \\
\end{array}
\]

Posterior probability of the states are visualized for the meltdown period during the financial crisis 2008/09, see Fig. A.6. First observe a distinct accumulation of probability mass in high volatility state 1 at the top of the cascade. Note also that this is the state where the information signal contained in the RV volatility proxy is buried under the largest amount of noise. As the downturn gains momentum, state 2 starts to occur more frequently. Actually, state 2 appears to dominate large part of the first half of the downturn. Signal intensity is almost twice that of the other states, mirroring a heavy information flow during the meltdown. Interestingly, in the approximately last third of the downturn, including the time when the market touches bottom, most frequent state 3 has clearly taken over again. Finally, there are a few relatively inactive periods during the second half of the bear market, located on what one may recognize as an intermediate plateau.

### A.6.4 The MS4SVskt-Range Model

Attractiveness of the range stems from its broad availability and a probably condensed information set contained in the market’s intraday top and bottom. Accordingly, this section sheds light on the information content of the range as a proxy for latent volatility, analyzing the SVskt-Range model and a four state MS4SVskt-Range variant. Priors
are as in Sec. A.5 and Eq. (A.22), except for the following. Regarding the volatility mode of the SVskt-Range model, $\mu_0 = -0.5$, $\sigma_{\mu_0} = 1$. For the regime switching MS4SVskt-Range extension, a mildly informative prior is chosen that reflects the specific state identification constraint $\mu_1 > \mu_2 > \mu_3 > \mu_4$,

$$\mu_0 = (0.5, 0, -1, -2)', \quad \Sigma_{\mu_0} = I_4.$$
Table A.9: Estimation Results of the SVskt-Range Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.6279</td>
<td>0.1102</td>
<td>[-0.8525, -0.4187]</td>
<td>13.1</td>
<td>0.97</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9577</td>
<td>0.0063</td>
<td>[0.9441, 0.9690]</td>
<td>38.8</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.3049</td>
<td>0.0238</td>
<td>[0.2620, 0.3545]</td>
<td>264.2</td>
<td>0.95</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.6795</td>
<td>0.0424</td>
<td>[-0.7677, -0.6036]</td>
<td>185.3</td>
<td>0.99</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-1.9463</td>
<td>0.4226</td>
<td>[-2.8464, -1.2076]</td>
<td>463.7</td>
<td>0.94</td>
</tr>
<tr>
<td>$\nu$</td>
<td>32.135</td>
<td>4.7584</td>
<td>[23.632, 42.010]</td>
<td>333.9</td>
<td>0.91</td>
</tr>
<tr>
<td>skewness</td>
<td>-0.4142</td>
<td>0.0547</td>
<td>[-0.5263, -0.3112]</td>
<td>139.8</td>
<td>0.95</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.5915</td>
<td>0.1125</td>
<td>[0.3938, 0.8384]</td>
<td>131.3</td>
<td>0.90</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>0.3943</td>
<td>0.0315</td>
<td>[0.3422, 0.4653]</td>
<td>176.5</td>
<td>1.00</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.4504</td>
<td>0.0109</td>
<td>[0.4297, 0.4723]</td>
<td>22.2</td>
<td>0.96</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.3641</td>
<td>0.0051</td>
<td>[0.3538, 0.3740]</td>
<td>11.7</td>
<td>0.97</td>
</tr>
<tr>
<td>SNR</td>
<td>0.8383</td>
<td>0.0734</td>
<td>[0.7081, 0.9938]</td>
<td>213.3</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke’s convergence diagnostic (p-value). Second panel: Implied higher moments of $\epsilon_t$. Third panel: Parameters auxiliary measurement equation. Lower panel: Implied signal-to-noise ratio.

Gaussian error to extract a persistent volatility process ($\phi = 0.555$). Results for the latter are not reported. Average block size and number of iterations to achieve convergence in the multi-move sampler are as in Sec. A.6.2.

Estimation results for the SVskt-Range model are reported in Tab. A.9. We observe a lower volatility mode $\mu$ compared to the SVskt-RV model, but the difference is barely significant. Moreover, volatility of volatility $\sigma$ (or more technically, signal strength) is significantly higher. The leverage effect can be considered still strong, $\rho = -0.680$, and is significantly higher in absolute magnitude than in the SVskt-RV model. Observation error $\epsilon_t$ shows a pronounced negative skewness of $-0.414$, almost double in size compared to the SVskt model. Analyzing the parameters of the auxiliary measurement equation, we observe that the SNR is with 83.8% significantly higher than in the SVskt-RV model, mainly due to the more intense signal. Interestingly, bias correction term $\zeta$ is now positive. However, an explanation in terms of dominating non-trading hours or market microstructure effects as in the SVskt-RV model appears inadequate.
### Table A.10: Estimation Results of the MS4SVsk-$t$-Range Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>0.7627</td>
<td>0.1331</td>
<td>[0.5038, 1.0242]</td>
<td>12.0</td>
<td>0.93</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-0.1007</td>
<td>0.1166</td>
<td>[-0.3284, 0.1310]</td>
<td>6.7</td>
<td>0.99</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.9723</td>
<td>0.1174</td>
<td>[-1.1979, -0.7371]</td>
<td>6.5</td>
<td>0.92</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>-1.8596</td>
<td>0.1198</td>
<td>[-2.0911, -1.6200]</td>
<td>7.6</td>
<td>0.80</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.9195</td>
<td>0.0255</td>
<td>[0.8688, 0.9688]</td>
<td>19.2</td>
<td>0.99</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.9633</td>
<td>0.0124</td>
<td>[0.9380, 0.9865]</td>
<td>34.1</td>
<td>0.92</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.9767</td>
<td>0.0110</td>
<td>[0.9537, 0.9956]</td>
<td>20.6</td>
<td>0.94</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.9701</td>
<td>0.0159</td>
<td>[0.9351, 0.9959]</td>
<td>33.3</td>
<td>0.91</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.3975</td>
<td>0.0400</td>
<td>[0.3167, 0.4747]</td>
<td>51.2</td>
<td>0.87</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.2391</td>
<td>0.0240</td>
<td>[0.1938, 0.2873]</td>
<td>112.0</td>
<td>0.83</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.3017</td>
<td>0.0213</td>
<td>[0.2615, 0.3451]</td>
<td>71.2</td>
<td>0.98</td>
</tr>
<tr>
<td>$\sigma_4$</td>
<td>0.2473</td>
<td>0.0254</td>
<td>[0.1984, 0.2978]</td>
<td>69.7</td>
<td>0.96</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>-0.3725</td>
<td>0.0747</td>
<td>[-0.5143, -0.2201]</td>
<td>17.8</td>
<td>0.97</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>-0.4616</td>
<td>0.0590</td>
<td>[-0.5722, -0.3419]</td>
<td>34.7</td>
<td>0.90</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>-0.5986</td>
<td>0.0450</td>
<td>[-0.6811, -0.5041]</td>
<td>34.7</td>
<td>0.95</td>
</tr>
<tr>
<td>$\rho_4$</td>
<td>-0.6268</td>
<td>0.0555</td>
<td>[-0.7294, -0.5111]</td>
<td>27.4</td>
<td>0.92</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>2.6680</td>
<td>0.4337</td>
<td>[1.7559, 3.4741]</td>
<td>37.8</td>
<td>0.97</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>2.1983</td>
<td>0.3588</td>
<td>[1.4758, 2.8821]</td>
<td>61.3</td>
<td>0.92</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.8000</td>
<td>0.3805</td>
<td>[-1.4968, 0.0204]</td>
<td>46.9</td>
<td>0.94</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>-0.9752</td>
<td>0.3948</td>
<td>[-1.7371, -0.1678]</td>
<td>28.4</td>
<td>0.87</td>
</tr>
<tr>
<td>$\nu$</td>
<td>50.564</td>
<td>6.3158</td>
<td>[39.069, 64.236]</td>
<td>121.1</td>
<td>0.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
</tr>
</thead>
<tbody>
<tr>
<td>skewness$_1$</td>
<td>0.3378</td>
<td>0.0598</td>
<td>[0.2196, 0.4561]</td>
</tr>
<tr>
<td>skewness$_2$</td>
<td>0.2835</td>
<td>0.0505</td>
<td>[0.1854, 0.3819]</td>
</tr>
<tr>
<td>skewness$_3$</td>
<td>-0.1070</td>
<td>0.0532</td>
<td>[-0.2075, 0.0027]</td>
</tr>
<tr>
<td>skewness$_4$</td>
<td>-0.1300</td>
<td>0.0553</td>
<td>[-0.2397, -0.0212]</td>
</tr>
<tr>
<td>kurtosis$_1$</td>
<td>0.3706</td>
<td>0.0941</td>
<td>[0.2133, 0.5831]</td>
</tr>
<tr>
<td>kurtosis$_2$</td>
<td>0.3018</td>
<td>0.0715</td>
<td>[0.1835, 0.4583]</td>
</tr>
<tr>
<td>kurtosis$_3$</td>
<td>0.1614</td>
<td>0.0356</td>
<td>[0.1083, 0.2451]</td>
</tr>
<tr>
<td>kurtosis$_4$</td>
<td>0.1732</td>
<td>0.0423</td>
<td>[0.1119, 0.2751]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\zeta$</td>
<td>0.4866</td>
<td>0.0183</td>
<td>[0.4524, 0.5246]</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.4274</td>
<td>0.0083</td>
<td>[0.4120, 0.4447]</td>
</tr>
<tr>
<td>$\sigma_{u,1}$</td>
<td>0.1643</td>
<td>0.0128</td>
<td>[0.1396, 0.1895]</td>
</tr>
<tr>
<td>$\sigma_{u,2}$</td>
<td>0.1039</td>
<td>0.0078</td>
<td>[0.0889, 0.1192]</td>
</tr>
<tr>
<td>$\sigma_{u,3}$</td>
<td>0.1148</td>
<td>0.0086</td>
<td>[0.0981, 0.1317]</td>
</tr>
<tr>
<td>$\sigma_{u,4}$</td>
<td>0.1762</td>
<td>0.0097</td>
<td>[0.1574, 0.1953]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR$_1$</td>
<td>2.4424</td>
<td>0.3678</td>
<td>[1.7772, 3.2280]</td>
</tr>
<tr>
<td>SNR$_2$</td>
<td>2.3185</td>
<td>0.3239</td>
<td>[1.7280, 2.9978]</td>
</tr>
<tr>
<td>SNR$_3$</td>
<td>2.6464</td>
<td>0.3120</td>
<td>[2.1166, 3.3328]</td>
</tr>
<tr>
<td>SNR$_4$</td>
<td>1.4087</td>
<td>0.1743</td>
<td>[1.0851, 1.7780]</td>
</tr>
</tbody>
</table>

As already mentioned above, it has been relatively difficult to extract a persistent latent volatility process for the SVskr-Range model. However, the MS4skr-Range model with four regimes fits the data rather naturally. Equilibrium is stable upfront without a need to make certain parameters regime invariant. Results are reported in Tab. A.10, and selected posterior probability densities of the parameters are shown in Fig. A.7. Extracted volatility modes are significantly different. Moreover, some interesting patterns are visible inspecting posterior density plots of the remaining parameters. Persistence $\phi$ and volatility of volatility $\sigma$ appear to be distinctively lower respective higher for high volatility state 1. Leverage $\rho$ is higher the lower volatility is (taking mode $\mu$ as proxy for volatility), which is in accordance with the findings of the
MS3SVskt model, and appears to be grouped into states 1/2 and 3/4. This separation is even more obvious for skew parameter $\gamma$, indicating a strong positive skew for high volatility regimes 1/2 and a negative one for low volatility regimes 3/4. Regime switching range noise $\sigma_u$ shows distinctive behavior as well, discriminating between more observation noise in the high/low volatility regimes and a less contaminated signal for the states featuring moderate volatility. Importantly, when we compare the signal-to-noise ratios with that of the SVskt-Range model, we recognize that they are significantly higher. This stems in large part from lower observation errors $\sigma_u$, indicating that the model extracts far more information out of the range measure than its regime invariant counterpart.

Comparing the time series of the logarithmized range and RV in Fig. A.1, a more diffuse nature of the former may let suggest heavier noise contamination. This can be made more precise now. Taking the estimation results of the MS4SVskt-RV and MS4SVskt-Range models as reference, one can say that RV features a less erratic information flow, meaning the received signal is subject to smaller shocks. Contamination by ephemeral observation error, on the contrary, can be relatively large. For the range the argument goes the other way around, explaining the higher signal-to-noise ratios and counteracting hereby its perception as a noisy measure. A more intense signal, however, is all else equal not advantageous for prediction, as will be made concrete in Sec. A.6.6 and A.6.8.

Relative to the MS4SVskt-RV model, the states show an even more transient behavior, and observations are more equally divided between the regimes:

<table>
<thead>
<tr>
<th>Vol</th>
<th>high</th>
<th>$s_t$</th>
<th>low</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>0.0952</td>
<td>0.2412</td>
<td>0.3985</td>
</tr>
<tr>
<td>2</td>
<td>0.0835</td>
<td>0.2908</td>
<td>0.3937</td>
</tr>
<tr>
<td>3</td>
<td>0.1485</td>
<td>0.3621</td>
<td>0.3003</td>
</tr>
<tr>
<td>4</td>
<td>0.1552</td>
<td>0.3671</td>
<td>0.3527</td>
</tr>
</tbody>
</table>

avg. observations

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>628.2</td>
<td>1,680.0</td>
<td>1,826.4</td>
<td>1,029.4</td>
</tr>
<tr>
<td>(12.2%)</td>
<td>(32.5%)</td>
<td>(35.4%)</td>
<td>(19.9%)</td>
</tr>
</tbody>
</table>
Posterior probability of the states are visualized for the meltdown period during the financial crisis 2008/09, see Fig. A.8. Compared to the posterior state probability of the MS4SVsk\(_t\)-RV model for the same period, Fig. A.6, we observe a more erratic regime process. However, observations appear rather well divided between the states. Specifically, note some distinctive occurrences of the positively skewed high volatility state.

### A.6.5 The MS2SVsk\(_t\)-VIX Model

This section presents the SVsk\(_t\)-VIX model and a two state MS2SVsk\(_t\)-VIX extension, using the VIX as additional measurement series. In contrast to RV and the range, which measure ex-post volatility, the VIX as volatility proxy measures future expectations. It is also a much smoother measure, clearly visible from Fig. A.1. However, the latter
A.6. EMPIRICAL APPLICATION

trait makes it more difficult to extract regimes and the approach is adopted in the following way. Specifically, the volatility mode has been shown inappropriate for regime identification using pilot runs and is left constant across states. Instead, states are identified by signal intensity $\sigma$. Further, auxiliary series bias $\zeta$ and scale $\xi$ are made regime dependent. There is a convergence issue regarding these two parameters, probably a result of the aforementioned smooth characteristic of the VIX. Making them regime dependent has been found to partly alleviate this issue. Moreover, experimentation with more than two states has shown notable potential. However, convergence issues prevail, equilibria are very difficult to extract and often unstable. Consequently, a robust two state version is presented.

Discussing the SVskt-VIX model first, priors are as in Sec. A.5 and Eq. (A.22), except for the following,

$$
\mu \sim \mathcal{N}(-0.5, 1),
$$

$$(\zeta, \xi)^\prime \sim \mathcal{N}((0, 0.5)^\prime, \text{diag}(10^2, 1)), \quad \sigma_u^2 \sim \mathcal{IG}(5/2, 0.001/2).
$$

The prior on the VIX measurement noise variance implies a $\sigma_u$ with mean 0.017 and standard deviation 0.007 (moments obtained by MC simulation), reflecting the smooth nature of the series. The first 10,000 draws are discarded as burn-in, collecting the following 100,000 draws for parameter inference. The larger number of draws reflects higher inefficiencies incurred when sampling bias correction $\zeta$ and scale $\xi$. In the multi-move sampler, average block size is set to 50 and number of iterations to achieve convergence to 3. The higher block size and lower number of iterations are motivated

Figure A.9: SVskt-VIX - Selected Posterior Probability Densities
Table A.11: Estimation Results of the SVskt-VIX Model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>µ</td>
<td>-0.3348</td>
<td>0.1724</td>
<td>[-0.6751, 0.0036]</td>
<td>1.1</td>
<td>0.87</td>
</tr>
<tr>
<td>φ</td>
<td>0.9884</td>
<td>0.0021</td>
<td>[0.9842, 0.9925]</td>
<td>1.8</td>
<td>0.97</td>
</tr>
<tr>
<td>σ</td>
<td>0.1390</td>
<td>0.0051</td>
<td>[0.1291, 0.1491]</td>
<td>143.1</td>
<td>0.70</td>
</tr>
<tr>
<td>ρ</td>
<td>-0.0155</td>
<td>0.0212</td>
<td>[-0.0582, 0.0247]</td>
<td>17.1</td>
<td>0.89</td>
</tr>
<tr>
<td>γ</td>
<td>0.1359</td>
<td>0.0975</td>
<td>[-0.0455, 0.3383]</td>
<td>33.5</td>
<td>0.93</td>
</tr>
<tr>
<td>ν</td>
<td>16.816</td>
<td>2.7303</td>
<td>[12.269, 22.856]</td>
<td>378.0</td>
<td>0.71</td>
</tr>
<tr>
<td>skewness</td>
<td>0.0675</td>
<td>0.0462</td>
<td>[-0.0246, 0.1572]</td>
<td>25.5</td>
<td>0.82</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.5064</td>
<td>0.1071</td>
<td>[0.3311, 0.7475]</td>
<td>298.1</td>
<td>0.66</td>
</tr>
<tr>
<td>ζ</td>
<td>0.3131</td>
<td>0.0101</td>
<td>[0.2967, 0.3345]</td>
<td>1,854.1</td>
<td>0.43</td>
</tr>
<tr>
<td>ξ</td>
<td>0.3871</td>
<td>0.0102</td>
<td>[0.3692, 0.4080]</td>
<td>2,383.3</td>
<td>0.50</td>
</tr>
<tr>
<td>σ_u</td>
<td>0.0181</td>
<td>0.0020</td>
<td>[0.0138, 0.0219]</td>
<td>279.0</td>
<td>0.79</td>
</tr>
<tr>
<td>SNR</td>
<td>7.7989</td>
<td>1.1403</td>
<td>[6.0130, 10.459]</td>
<td>306.4</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke’s convergence diagnostic (p-value). Second panel: Implied higher moments of \( \epsilon_t \). Third panel: Parameters auxiliary measurement equation. Lower panel: Implied signal-to-noise ratio.

by computational cost savings. Practically, inference on latent volatility is very efficient due to only minor noise in the VIX series, keeping acceptance rates high in the block sampler also for larger block sizes.

Tab. A.11 reports results. Latent process parameters \( \mu, \phi, \sigma \) are not significantly different from the SVskt model, whereupon volatility persistence is even slightly higher and volatility of volatility lower. This is not surprising, as the VIX is a very clean proxy for volatility. However, leverage and skew differ significantly from the former. Respective posterior densities are visualized in Fig. A.9. Evidence for leverage is very weak and is not significant at the 95% level. This fits well with results of JM, who report estimates close to zero. The former authors give a possible explanation, arguing that the VIX estimator is based on future expectations of volatility and may thus incorporate the information content of the leverage parameter. Skew parameter \( \gamma \) is also not significantly different from zero, but most of its probability mass lies in the positive domain. This is in contrast to all other regime invariant models analyzed, which feature negative skewness. However, kurtosis equals that of the SVskt model. Analyzing auxiliary measurement equation related parameters, bias correction \( \zeta \) is positive as in
the range class of models. Mentioned already, the inefficiency factors for $\zeta$ and $\xi$ are very high, but the range around which these parameters fluctuate is rather small, as can be inferred from corresponding standard deviations and 95% credible bounds. Finally, the SNR is very high, a result of low VIX measurement noise $\sigma_u = 0.018$.

Turning to the proposed two state MS2Vskt-VIX variant, priors are as in Sec. A.5 and Eq. (A.22), except for the following,

$$
\mu \sim \mathcal{N}(-1,1),
$$

$$
\lambda_1 \sim \mathcal{IG}(5/2,0.2/2), \quad \lambda_2 \sim \mathcal{IG}(5/2,0.05/2),
$$

$$
\zeta \sim \mathcal{N}((0.5,0.5)', I_2), \quad \xi \sim \mathcal{N}((0.4,0.4)', I_2),
$$

$$
\sigma_{u,i}^2 \sim \mathcal{IG}(5/2,0.001/2), \quad i = 1,2.
$$

The priors on $\lambda_1, \lambda_2$ reflect state identification constraint $\sigma_1 > \sigma_2$, implying means of 0.24, 0.12 and standard deviations of 0.10, 0.05 for $\sigma_1, \sigma_2$, respectively, in the case of zero correlation (moments obtained by MC simulation). Inefficiencies encountered when drawing inference on auxiliary measurement equation parameters $\zeta, \xi$ increase in
Table A.12: Estimation Results of the MS2SVskt-VIX Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-1.1072</td>
<td>0.1658</td>
<td>[-1.4544, -0.7978]</td>
<td>4.9</td>
<td>0.90</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9898</td>
<td>0.0019</td>
<td>[0.9860, 0.9936]</td>
<td>6.6</td>
<td>0.97</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.2129</td>
<td>0.0084</td>
<td>[0.1971, 0.2301]</td>
<td>41.5</td>
<td>0.79</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.0798</td>
<td>0.0032</td>
<td>[0.0736, 0.0862]</td>
<td>125.2</td>
<td>0.78</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>-0.0522</td>
<td>0.0412</td>
<td>[-0.1333, 0.0281]</td>
<td>5.5</td>
<td>0.98</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.0124</td>
<td>0.0247</td>
<td>[-0.0365, 0.0604]</td>
<td>8.0</td>
<td>0.95</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-7.3662</td>
<td>0.4298</td>
<td>[-8.2482, -6.5468]</td>
<td>136.2</td>
<td>0.88</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.4561</td>
<td>0.2300</td>
<td>[0.0134, 0.9186]</td>
<td>71.7</td>
<td>0.86</td>
</tr>
<tr>
<td>$\nu$</td>
<td>34.4165</td>
<td>3.8903</td>
<td>[27.3706, 42.6165]</td>
<td>211.3</td>
<td>0.97</td>
</tr>
<tr>
<td>skewness$_1$</td>
<td>-0.9235</td>
<td>0.0729</td>
<td>[-1.0791, -0.7923]</td>
<td>181.9</td>
<td>1.00</td>
</tr>
<tr>
<td>skewness$_2$</td>
<td>0.0931</td>
<td>0.0462</td>
<td>[0.0028, 0.1841]</td>
<td>73.4</td>
<td>0.88</td>
</tr>
<tr>
<td>kurtosis$_1$</td>
<td>1.8361</td>
<td>0.3042</td>
<td>[1.3293, 2.5239]</td>
<td>188.5</td>
<td>0.99</td>
</tr>
<tr>
<td>kurtosis$_2$</td>
<td>0.2241</td>
<td>0.0346</td>
<td>[0.1664, 0.3011]</td>
<td>151.8</td>
<td>0.95</td>
</tr>
<tr>
<td>$\zeta_1$</td>
<td>0.5296</td>
<td>0.0168</td>
<td>[0.4981, 0.5692]</td>
<td>2,822.7</td>
<td>0.70</td>
</tr>
<tr>
<td>$\zeta_2$</td>
<td>0.5271</td>
<td>0.0169</td>
<td>[0.4957, 0.5670]</td>
<td>11,638.5</td>
<td>0.44</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>0.4203</td>
<td>0.0103</td>
<td>[0.4015, 0.4408]</td>
<td>1,529.9</td>
<td>0.83</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.4237</td>
<td>0.0103</td>
<td>[0.4051, 0.4441]</td>
<td>9,189.9</td>
<td>0.69</td>
</tr>
<tr>
<td>$\sigma_{u,1}$</td>
<td>0.0316</td>
<td>0.0025</td>
<td>[0.0265, 0.0365]</td>
<td>55.5</td>
<td>0.99</td>
</tr>
<tr>
<td>$\sigma_{u,2}$</td>
<td>0.0090</td>
<td>0.0012</td>
<td>[0.0069, 0.0114]</td>
<td>261.0</td>
<td>0.95</td>
</tr>
<tr>
<td>SNR$_1$</td>
<td>6.7822</td>
<td>0.6493</td>
<td>[5.6830, 8.2300]</td>
<td>50.8</td>
<td>0.89</td>
</tr>
<tr>
<td>SNR$_2$</td>
<td>9.0434</td>
<td>1.3416</td>
<td>[6.7309, 11.9585]</td>
<td>262.2</td>
<td>0.87</td>
</tr>
</tbody>
</table>


the regime switching case even more. To account for this, 300,000 draws are collected, after discarding the first 30,000 as burn-in. In the multi-move sampler, average block size is set to 100 and number of iterations to achieve convergence to 3.

Tab. A.12 reports results and selected posterior probability densities are shown in Fig. A.10. Two regimes are clearly identified by signal intensity $\sigma$. Leverage is barely present in both states, albeit for intense signal state 1 there is a considerable portion of probability mass located in the negative domain. Volatility mode $\mu$ is significantly lower than in the SVskt-VIX model, as more volatility dynamics are shifted into the tails. Strong tail characteristics are especially present in state 1, with measurement error $\epsilon_t$.
A.6. **EMPIRICAL APPLICATION**

featuring a negative skewness of $-0.924$ and excess kurtosis of 1.836. About one fourth of the observations fall into this extremely negative skewed state, see Eq. (A.25). Inefficiency factors for $\zeta$, $\xi$ are high, but corresponding standard deviations are low and 95% credible bounds narrow, indicating only minor fluctuations of these parameters. Note that $\zeta$, $\xi$ are almost equal across states; however, as already mentioned, merging the respective states would result in even higher inefficiency factors.\textsuperscript{12} Observation noise $\sigma_u$ is more than three times higher for high signal intensity state 1. Moreover, the SNR is larger in state 2, albeit not significantly, despite state 1 exhibiting a more intense signal.

Transition probabilities and observations per state are summarized:

$$
\begin{array}{ccc}
\text{Vol} & \text{high} & \text{low} \\
\text{high} & & \\
1 & s_t & \\
& 1 & 2 \\
2 & 1 (0.2701 & 0.7299) \\
& 2 (0.2471 & 0.7529) \\
\text{avg. observations} & 1,304.1 & 3,859.9 \\
& (25.3\%) & (74.7\%) \\
\end{array}
$$

State 1 is transitory, with the probability to remain in this state only 27.0%. Persistence of state 2 is with 75.3% rather moderate too small. Interestingly, the model variants with Gaussian or Student-$t$ error exhibit higher persistent states, around 88% for state 1 and 97% for state 2 (estimates are reported in App. A.E). From Fig. A.11, which plots state posterior probability for the meltdown period during the financial crisis 2008/09, an increased occurrence of state 1 during the downturn around Q4-08 is clearly visible. This fosters the perception of this state modeling negative unexpected news flow.

\textsuperscript{12}Moreover, for model variants with Student-$t$ and Gaussian error, scale parameters $\xi_1$ and $\xi_2$ have found to be significantly different, see App. A.E.
A.6.6 Model Selection

Models are compared using the deviance information criterion ($DIC$) (Spiegelhalter, Best, Carlin, and van der Linde, 2002). In addition, the deviance of the in-sample predictive log-likelihood is reported to give an indication of out-of-sample forecasting performance. As outlined in Sec. A.4, in-sample fit and predictive ability are evaluated for the return distribution only, which is the object of interest in this work. The $DIC$ is defined by

$$DIC = \bar{D} + p_d,$$

where

$$p_d = \bar{D} - D(\bar{\theta}), \quad \bar{D} = E_{\theta|y}[D(\theta)], \quad D(\theta) = -2 \log f(y_{1:n}|\theta),$$

with $D(\theta)$ the deviance and $\bar{\theta}$ the posterior mean. It is a Bayesian measure of model fit, attaining smaller values for better models, with $p_d$ a penalty term for model complexity. Intuitively, higher parameter uncertainty lets $D(\theta)$ fluctuate more

---

Figure A.11: MS2SVskt-VIX - Posterior Probability of States (08/07/01-09/06/30). Background: S&P 500 (log-scale, right).
widely around $D(\bar{\theta})$, yielding larger $p_D$ values. Numerical standard errors of the estimates are obtained the following way. Posterior expectation of the deviance $\bar{D}$, readily available as a byproduct of the McMC run, is calculated by dividing the sample in 10 equally sized batches. Regarding deviance of the posterior mean $D(\bar{\theta})$, the particle filter outlined in Sec. A.4 is repeatedly applied 10 times to obtain estimates of log-likelihood ordinate $\log f_{\text{post}}(y_{1:n} | \bar{\theta}) = \sum_{t=1}^{n} \log \hat{f}(y_{1:t} | y_{1:t}, \bar{\theta})$. The deviance of the predictive log-likelihood is calculated in a similar way from ordinate $\log f_{\text{prior}}(y_{1:n} | \bar{\theta}) = \sum_{t=1}^{n} \log \hat{f}(y_{1:t} | y_{1:t-1}, \bar{\theta})$. The above approach yields 100 different DIC values, from which statistics are calculated.

Tab. A.13 reports DIC values, parameter penalty terms and the deviances of the predictive likelihoods. The regime switching range class of models, but also the SVskt-Range and the MS2SVskt-VIX model show distinctively low DIC values, indicating a superior in-sample fit. However, their predictive performance (in-sample) is among the worst. Specifically, the MS4SVskt-Range model has the best DIC but the worst predictive likelihood. This discrepancy can in large part be attributed to the regime switching dynamics and tail behavior. First, especially in case of the range class of models, rather transient regime dynamics increase the difference between mixture forecast and a posterior realized state. Second, the above models all feature a pronounced tail behavior. A considerable part of the volatility dynamics has been shifted into the tails. As these are unconditional within regimes, forecasting performance likely deteriorates, but a posterior fit increases. On the other side, the basic SV model has the worst DIC but performs considerably better according to the predictive criterion. The more sophisticated Markov switching and fat tailed versions in this class improve on predictive ability and in-sample fit, with the MS3SVskt the best according to both criteria. The RV class of models shows best predictive performance, most notably the MS4SV$\tau$-RV variant, and has a good to medium in-sample fit. As for the models with no auxiliary measurement equation, modeling tail behavior and regime dynamics overall improves both in-sample fit and predictive ability. However, introducing skewness in addition to regime switching and fat tails does not yield an improvement in predictive performance. Performance of the VIX class of models is mixed under both criteria. Only the MS2SVskt-VIX shows a superior in-sample fit,
Table A.13: Model Selection

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC (s.e.)</th>
<th>Rank</th>
<th>PD (s.e.)</th>
<th>$D_{pred}$ (s.e.)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV</td>
<td>13,761.7 (8.3)</td>
<td>22</td>
<td>311.2 (4.3)</td>
<td>13,799.7 (0.6)</td>
<td>14</td>
</tr>
<tr>
<td>SVt</td>
<td>13,541.2 (33.5)</td>
<td>18</td>
<td>363.5 (16.8)</td>
<td>13,770.3 (1.3)</td>
<td>12</td>
</tr>
<tr>
<td>SVskt</td>
<td>13,434.8 (32.9)</td>
<td>17</td>
<td>486.9 (16.8)</td>
<td>13,747.5 (2.1)</td>
<td>8</td>
</tr>
<tr>
<td>MS3SVt</td>
<td>13,547.9 (35.2)</td>
<td>19</td>
<td>400.9 (17.7)</td>
<td>13,731.9 (1.7)</td>
<td>7</td>
</tr>
<tr>
<td>MS3SVskt</td>
<td>13,413.2 (55.9)</td>
<td>15</td>
<td>479.5 (28.4)</td>
<td>13,712.1 (2.9)</td>
<td>5</td>
</tr>
<tr>
<td>SV-RV</td>
<td>13,337.4 (3.5)</td>
<td>12</td>
<td>125.8 (2.0)</td>
<td>13,749.9 (0.5)</td>
<td>9</td>
</tr>
<tr>
<td>SVt-RV</td>
<td>13,268.7 (12.4)</td>
<td>10</td>
<td>211.0 (6.3)</td>
<td>13,716.1 (0.6)</td>
<td>6</td>
</tr>
<tr>
<td>SVskt-RV</td>
<td>13,235.4 (16.1)</td>
<td>9</td>
<td>217.4 (8.2)</td>
<td>13,697.8 (0.9)</td>
<td>4</td>
</tr>
<tr>
<td>MS4SV-RV</td>
<td>13,169.2 (4.2)</td>
<td>8</td>
<td>110.9 (2.1)</td>
<td>13,679.8 (0.4)</td>
<td>2</td>
</tr>
<tr>
<td>MS4SVt-RV</td>
<td>13,128.6 (11.8)</td>
<td>7</td>
<td>185.0 (5.9)</td>
<td>13,674.1 (0.5)</td>
<td>1</td>
</tr>
<tr>
<td>MS4SVskt-RV</td>
<td>13,081.9 (29.6)</td>
<td>6</td>
<td>226.1 (14.8)</td>
<td>13,696.5 (0.6)</td>
<td>3</td>
</tr>
<tr>
<td>SVt-Range</td>
<td>13,285.2 (26.9)</td>
<td>11</td>
<td>480.9 (13.4)</td>
<td>13,893.3 (0.6)</td>
<td>19</td>
</tr>
<tr>
<td>SVskt-Range</td>
<td>12,545.1 (277.4)</td>
<td>5</td>
<td>994.9 (138.7)</td>
<td>13,872.7 (1.8)</td>
<td>18</td>
</tr>
<tr>
<td>MS4SV-Range</td>
<td>10,899.7 (5.2)</td>
<td>2</td>
<td>96.2 (2.7)</td>
<td>14,032.6 (1.9)</td>
<td>20</td>
</tr>
<tr>
<td>MS4SVt-Range</td>
<td>10,942.5 (5.3)</td>
<td>3</td>
<td>124.3 (2.7)</td>
<td>14,035.5 (1.9)</td>
<td>21</td>
</tr>
<tr>
<td>MS4SVskt-Range</td>
<td>10,708.6 (110.3)</td>
<td>1</td>
<td>531.6 (55.2)</td>
<td>14,073.8 (2.3)</td>
<td>22</td>
</tr>
<tr>
<td>SV-VIX</td>
<td>13,619.1 (3.1)</td>
<td>21</td>
<td>1.4 (2.9)</td>
<td>13,849.8 (0.8)</td>
<td>17</td>
</tr>
<tr>
<td>SVt-VIX</td>
<td>13,384.8 (42.6)</td>
<td>14</td>
<td>160.2 (22.3)</td>
<td>13,751.6 (1.1)</td>
<td>10</td>
</tr>
<tr>
<td>SVskt-VIX</td>
<td>13,374.3 (41.7)</td>
<td>13</td>
<td>148.9 (22.8)</td>
<td>13,776.9 (0.9)</td>
<td>13</td>
</tr>
<tr>
<td>MS2SV-VIX</td>
<td>13,596.5 (6.0)</td>
<td>20</td>
<td>-2.5 (4.6)</td>
<td>13,831.2 (2.8)</td>
<td>16</td>
</tr>
<tr>
<td>MS2SVt-VIX</td>
<td>13,427.4 (39.9)</td>
<td>16</td>
<td>193.8 (20.9)</td>
<td>13,755.0 (3.5)</td>
<td>11</td>
</tr>
<tr>
<td>MS2SVskt-VIX</td>
<td>11,917.3 (219.5)</td>
<td>4</td>
<td>1408.8 (110.7)</td>
<td>13,809.9 (1.0)</td>
<td>15</td>
</tr>
</tbody>
</table>


but its predictive likelihood is rather low. The fat tailed but symmetric SVt-VIX and MS2SVt-VIX models show best predictive performance in this category.

Especially a Student-$t$ distributed error term has a positive impact on forecasting ability. Except in the range class, these model variants perform uniformly better than their Gaussian counterparts. However, whereas the additional effect of asymmetric tails on in-sample fit is overall positive, impact on forecasting ability is mixed. For the models using no auxiliary series, introducing skewness yields an improvement. When RV or the range as additional volatility proxy is used, this is so only for the regime invariant model variants. Modeling asymmetry in the error term for the VIX class
yields an improvement relative to the Gaussian version, but the variants with Student-t error dominate significantly.

Considering complexity penalty term $p_D$, three points are to mention. First, a positive effect on parameter uncertainty by incorporating an additional proxy for latent volatility is clearly visible from in general higher $p_D$ values for the models using no auxiliary equation, comparing e.g. the SV and SV-RV model. Second, modeling skewness in the error distribution substantially increases model complexity, which is especially striking for the SVskt-Range and MS2SVskt-VIX model. Third, using the smooth VIX series as additional time series may substantially improve parameter inference, which is most pronounced in the Gaussian case, where the $p_D$ term even turns slightly negative for the MS2SV-VIX model. Further specification tests are provided in App. A.D for the interested reader.

### A.6.7 Prior Sensitivity Analysis

Sensitivity of skewness and kurtosis regarding prior selection for degrees of freedom parameter $\nu$ has been investigated in Sec. A.5. Now, prior selection for skew parameter $\gamma$ and resulting consequences for in-sample fit and predictive ability are analyzed. As already mentioned, a diffuse prior for $\gamma$ can be motivated by the desire "to let the data speak for itself". On the contrary and as in the current implementation, having an inherently time varying character of skewness in mind, one may argue for a prior more centered around zero to constrain fluctuations and potentially increasing predictive ability as a result. To investigate this point in more detail, following Prior #2 is additionally proposed (as employed in the simulation study of Sec. A.5),

Prior #1 : $\gamma \sim \mathcal{N}(0,1)$, \quad Prior #2 : $\gamma \sim \mathcal{N}(0,10^{12})$.

Model selection criteria along the lines of Sec. A.6.6 are provided for regime switching models MS3SVskt and MS4SVskt-Range, which feature pronounced skewness.\(^{13}\)

Tab. A.14 reports results. Not very surprising, diffuse Prior #2 provides a better in-sample fit as indicated by posterior likelihood measure $\hat{D}$. However, the difference

\(^{13}\)The MS2SVskt-VIX model has not been chosen, as the already high inefficiency factors for bias and scale parameters $\zeta$, $\xi$ increased further using diffuse Prior #2.
### Table A.14: Prior Sensitivity of Skew Parameter $\gamma$ - Model Selection Criteria

<table>
<thead>
<tr>
<th>Prior</th>
<th>DIC (s.e.)</th>
<th>$p_D$ (s.e.)</th>
<th>$\bar{D}$ (s.e.)</th>
<th>$D_{\text{pred}}$ (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>13,413.2 (55.9)</td>
<td>479.5 (28.4)</td>
<td>12,933.8 (27.8)</td>
<td>13,712.1 (2.9)</td>
</tr>
<tr>
<td>#2</td>
<td>13,421.7 (68.0)</td>
<td>495.7 (34.2)</td>
<td>12,926.0 (33.9)</td>
<td>13,711.4 (3.0)</td>
</tr>
<tr>
<td>#1</td>
<td>10,708.6 (110.3)</td>
<td>531.6 (55.2)</td>
<td>10,177.0 (55.2)</td>
<td>14,073.8 (2.3)</td>
</tr>
<tr>
<td>#2</td>
<td>10,561.8 (180.8)</td>
<td>711.6 (90.4)</td>
<td>9,850.3 (90.4)</td>
<td>14,091.4 (1.9)</td>
</tr>
</tbody>
</table>

Deviance information criterion, complexity penalty term, posterior mean of the deviance, and deviance of the predictive log-likelihood. Numerical standard error in parentheses.

is only significant for the MS4SVskt-Range model, featuring a more pronounced skewed error term. A positive effect of less diffuse Prior #1 on parameter fluctuation is indicated by lower parameter complexity penalty $p_D$, the effect again significant only for the MS4SVskt-Range model. The lower complexity penalty leads the $DIC$ criterion to favor the MS3SVskt model under Prior #1, albeit not significantly, despite a worse in-sample fit. In contrast, lower complexity penalty can not outweigh worse in-sample fit in case of the MS4SVskt-Range model, and more diffuse Prior #2 is favored significantly according to the $DIC$. Inspecting (in-sample) predictive deviance criterion $D_{\text{pred}}$, more centered Prior #1 yields significantly better results for the MS4SVskt-Range model. In case of the less heavy skewed MS3SVskt model differences are not significant. In accordance with these positive indications regarding predictive ability, common Prior #1 is also employed for all models in the out-of-sample forecasting exercise that follows.

### A.6.8 Out-of-Sample Predictive Density Tests

This section investigates out-of-sample performance. Models are estimated on a subsample period, and forecasts are generated for two distinct out-of-sample periods spanning approximately 4 years each, see Tab. A.15 for descriptive statistics and Fig. A.12 for a graphical illustration. The in-sample period contains the dotcom
Table A.15: Descriptive Statistics In-/Out-of-Sample

<table>
<thead>
<tr>
<th>Period</th>
<th>In-sample</th>
<th>Out-of-sample Period I</th>
<th>Out-of-sample Period II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs.</td>
<td>90/01/02 - 04/03/31</td>
<td>04/04/01 - 08/03/31</td>
<td>06/08/01 - 10/08/06</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0305</td>
<td>0.0158</td>
<td>-0.0125</td>
</tr>
<tr>
<td>Stdev</td>
<td>1.0509</td>
<td>0.8302</td>
<td>1.7109</td>
</tr>
<tr>
<td>Skew</td>
<td>-0.1001</td>
<td>-0.2293</td>
<td>-0.2112</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>6.5780</td>
<td>5.4358</td>
<td>10.2863</td>
</tr>
<tr>
<td>Min</td>
<td>-7.1127</td>
<td>-3.5343</td>
<td>-9.4695</td>
</tr>
<tr>
<td>Max</td>
<td>5.573</td>
<td>4.1535</td>
<td>10.9572</td>
</tr>
<tr>
<td>$LB^2(15)$</td>
<td>1162.4</td>
<td>377.7</td>
<td>1210.4</td>
</tr>
</tbody>
</table>

The $LB^2(15)$ test is on squared return $y^2_{1,t}$, critical value is 30.6 at the 1% significance level.

Figure A.12: Partitioning of the Data In-/Out-of-Sample

peak and subsequent trough, 1990/01/02 - 2004/03/31, spanning 3,567 observations (14¼ years). Out-of-sample periods are chosen as to reflect a quiet period on the one side, 2004/04/01 - 2008/03/31 (1,004 obs., 4 yrs) and the volatile recent financial crisis on the other, 2006/08/01 - 2010/08/06 (1,011 obs., ≈ 4 yrs). Distinctive characteristics of the latter period are reflected in much higher volatility and kurtosis. Note that skew is negative for all samples and about the same magnitude for both out-of-sample periods. Moreover, the Ljung-Box statistic indicates far lower autocorrelation in squared returns for the more quiet period.

When generating out-of-sample forecasts, the particle filter of Sec. A.4 is started 50 observations prior to the start of the respective periods to avoid any deficiencies.
due to unconditional initialization. Further, the out-of-sample data is demeaned over
the complete period.\textsuperscript{14} Some model variants have been fitted to the data with slight
modifications or have been dropped, for the following reasons. In the MS3SVsk\(t\) class
with no auxiliary series, difficulties were encountered when extracting a stable third
rare event regime, and thus a more precise prior \(\sigma_{\mu_{0,1}} = 0.1\) on the volatility mode of
state 1 has been imposed, all else equal. Consequently, the two state versions MS2SV,
MS2SV\(t\), and MS2SVsk\(t\) have been estimated in addition, employing the usual mildly
informative priors. Moreover, the SVsk\(t\)-RV variant is dropped, as the probability mass
of skew parameter \(\gamma\) centers around zero and can not be considered significant. Further,
no persistent volatility process could be extracted for the one regime SVsk\(t\)-Range
class, irrespective of modeling the tails. These variants are also dropped. Finally, to
make the selection of models more comprehensible from a forecaster’s point of view,
model variants using RV or the range with less than four regimes have been estimated.
Thereby, respective parameter configurations were retained, and a symmetric heavy
tailed error was deployed. However, these model variants have to be considered inferior
relative to their four state versions according to the \textit{DIC} criterion, both regarding the
complete likelihood and the return distribution only (in-sample).

Tab. A.16 reports log predictive density ratios relative to the basic SV model for
the two out-of-sample periods for estimated models, with higher values indicating
better fit. Analyzing calm forecasting period I first, performance is rather well divided
between model classes. The models using no auxiliary time series perform uniformly
better, with the basic SV model the least strong and the MS2SVsk\(t\) variant on top.
Clearly, modeling regimes and tail behavior pays off, with the asymmetric versions the
best. However, a third state does not appear to deliver significant improvements, as
the respective two state variants all show higher LPDRs. Recall that the third state has
been enforced by a very informative prior. The RV class of models follows, with the
MS4SV\(t\)-RV variant located most closely behind the baseline SV model. Interestingly,
both the most basic SV-RV and sophisticated MS4SVsk\(t\) model are the worst performer
in this class. Moreover, it appears crucial to model symmetric fat tails especially
for the one regime variant. Note a strong negative effect of modeled skewness on

\textsuperscript{14}This actually induces a forward-looking bias. However, mean dynamics are generally assumed
negligible when working with daily volatility.
### Table A.16: Forecasting - Predictive Density Ratios

<table>
<thead>
<tr>
<th>Model</th>
<th>Period I (calm)</th>
<th>Period II (volatile)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LPDR (s.e.)</td>
<td>Rank</td>
</tr>
<tr>
<td>SV</td>
<td>0 (0.2)</td>
<td>9</td>
</tr>
<tr>
<td>SV_t</td>
<td>3.7 (0.2)</td>
<td>8</td>
</tr>
<tr>
<td>SV_{skt}</td>
<td>7.8 (0.2)</td>
<td>4</td>
</tr>
<tr>
<td>MS2SV</td>
<td>5.8 (0.2)</td>
<td>6</td>
</tr>
<tr>
<td>MS2SV_{t}</td>
<td>9.1 (0.1)</td>
<td>2</td>
</tr>
<tr>
<td>MS2SV_{skt}</td>
<td>11.9 (0.3)</td>
<td>1</td>
</tr>
<tr>
<td>MS3SV</td>
<td>4.7 (0.1)</td>
<td>7</td>
</tr>
<tr>
<td>MS3SV_{t}</td>
<td>6.7 (0.1)</td>
<td>5</td>
</tr>
<tr>
<td>MS3SV_{skt}</td>
<td>8.1 (0.1)</td>
<td>3</td>
</tr>
<tr>
<td>SV-RV</td>
<td>-11.9 (0.4)</td>
<td>15</td>
</tr>
<tr>
<td>SV_{t}-RV</td>
<td>-5.6 (0.6)</td>
<td>13</td>
</tr>
<tr>
<td>MS4SV-RV</td>
<td>-2.1 (0.1)</td>
<td>11</td>
</tr>
<tr>
<td>MS4SV_{t}-RV</td>
<td>-0.8 (0.2)</td>
<td>10</td>
</tr>
<tr>
<td>MS4SV_{skt}-RV</td>
<td>-10.8 (0.1)</td>
<td>14</td>
</tr>
<tr>
<td>MS4SV-Range</td>
<td>-29.0 (0.3)</td>
<td>21</td>
</tr>
<tr>
<td>MS4SV_{t}-Range</td>
<td>-29.2 (0.2)</td>
<td>22</td>
</tr>
<tr>
<td>MS4SV_{skt}-Range</td>
<td>-31.2 (0.3)</td>
<td>23</td>
</tr>
<tr>
<td>SV-VIX</td>
<td>-17.8 (0.5)</td>
<td>18</td>
</tr>
<tr>
<td>SV_{t}-VIX</td>
<td>-5.5 (0.4)</td>
<td>12</td>
</tr>
<tr>
<td>SV_{skt}-VIX</td>
<td>-20.2 (0.4)</td>
<td>19</td>
</tr>
<tr>
<td>MS2SV-VIX</td>
<td>-20.3 (0.5)</td>
<td>20</td>
</tr>
<tr>
<td>MS2SV_{t}-VIX</td>
<td>-11.9 (1.1)</td>
<td>16</td>
</tr>
<tr>
<td>MS2SV_{skt}-VIX</td>
<td>-15.6 (0.2)</td>
<td>17</td>
</tr>
</tbody>
</table>

Cumulative log predictive density ratio. Numerical standard error in parentheses.

Forecasting performance for the regime switching case, qualitatively comparable to the respective in-sample results of Sec. A.6.6. The range variants are located on the lowest ranks, reinforcing the in-sample results for the range as a bad predictor. LPDRs of these models are all distinctively low. The VIX class also shows an unsatisfying forecasting ability for calm period I. Only the SV_{t}-VIX model relatively stands out, again congruent with the respective in-sample findings.

Regarding volatile forecasting period II, performance can not be anymore differentiated between model classes as clearly as for period I. High performers are found in all categories, except again in the range class. There is a considerable
improvement compared to period I for the Gaussian and Student-$t$ version of the latter, but performance remains unsatisfying. The overall best performer is the relatively simple SVsk$t$ model. This is perhaps surprising, but can be related to its relatively parsimonious specification especially in conjunction with no re-estimation. It is followed by the MS4SV$t$-RV model on rank 2. Notably, ranks 3 to 5 are now occupied by models using the VIX, with the SV$t$-VIX variant the best. This is in stark contrast to the overall inferior predictive performance of this model class found for the calmer period. JM investigate forecasting ability of SV models using RV and the VIX as additional measurement variables during the crash period 2008/09, using a sequential Monte Carlo sampler with time varying parameters. They come to qualitatively comparable results, supporting seizable information content w.r.t. volatility forecasting contained in the VIX and RV.\textsuperscript{15} Having four competitors on the first five ranks that use an auxiliary variable and/or regime switching also puts the success of the top performing SVsk$t$ model into a broader perspective. Further, note quite heavy differences in the rankings especially within the RV and VIX classes. Similar patterns have already been observed for calmer period I, but less pronounced. In case of the VIX, modeling the tails appears essential. Concerning skewness, the big difference in performance between the SVsk$t$-VIX and the regime switching MS2SVsk$t$-VIX model indicates a crucial necessity to model the strong negatively skewed second state, mirroring bad unexpected news.

Regarding both out-of-sample periods and analogous to the in-sample results of Sec. A.6.6, a Student-$t$ error in general increases predictive ability. Interestingly, a skewed error term improves upon its symmetric counterpart only for the models featuring no auxiliary series in this out-of-sample study. However, the MS2SVsk$t$-VIX model capturing extreme skewness is competitive in this respect. Clearly, introducing switching regimes can increase predictive ability, but more frequent and timely re-estimation is necessary to get a clearer picture. However, the superior performance of the models using no auxiliary times series during less volatile markets is rather obvious. Moreover, results suggest that gains w.r.t. to volatility forecasting using RV or the VIX as additional time series can be expected in more turbulent market periods.

\textsuperscript{15}JM further estimate a model using the VIX and RV simultaneously as two auxiliary equations. Considering the findings of the current paper, this clearly is an attractive option.
A very general SV model specification featuring leverage, heavy asymmetric tails, switching regimes and an auxiliary time series to improve inference on latent volatility is provided. The model encompasses a wide variety of models encountered in the literature. A McMC sampler and associated particle filter are proposed and outlined in detail.

Regime switching applications of model variants using RV, the range and the VIX as additional volatility proxy are presented, identifying up to four regimes from the data. Depending on the model, significantly skewed and heavy tailed errors are found. Regime switching dynamics are analyzed around the crash period 2008/09 in more detail.

In-sample model selection and out-of-sample predictive density tests are conducted. The range has been found to provide the best in-sample fit, but its predictive ability is weak. On the other side, RV and the VIX do provide seizable content with respect to volatility forecasting. This has also been found by JM in a comparable context. However, current findings suggest that the potential additional utility of RV and especially the VIX appears to pay off mostly in more turbulent market conditions. In stable environments, SV specifications using no auxiliary time series may be preferred. An overall assessment of (asymmetric) tail modeling and regime switching regarding predictive ability yields the following insights. Whereas heavy tails in general attribute positively, an asymmetric error term improves forecasting ability only for models using no auxiliary volatility proxy. An exception appears to be the skewed two state model variant using the VIX, which shows competitive performance for the volatile period. Thereby, the extremely negative skewed second state models unexpected bad news flow. Finally, the findings support the concept of regime switching as a useful tool to improve model fit and predictive performance.

Extensions are imaginable. Using various volatility measures simultaneously to draw inference on latent volatility is certainly useful. However, how to implement this in a regime switching scenario is not a straightforward question to answer. Going a step further in this direction, and considering the higher utility of RV and the VIX during
turbulent market phases, making the inclusion of these measures into the model regime dependent could be another option. Then, a more elaborated estimator for realized volatility, as e.g. two scale RV (JM), can be used. Finally, one may drop the Gaussian error assumption of the auxiliary series and introduce foremost skewness to even better separate persistent information flow from spurious noise in the data.
A.A Multi-Move Sampler

Nakajima and Omori (2012) extend the multi-move (or block) sampler with leverage described by Omori and Watanabe (2008) to handle the Generalized Hyperbolic skew-$t$ distribution. Takahashi et al. (2009) introduce realized volatility as additional measurement equation. By defining $\alpha_t = h_t - \mu_{s_t}, t = 1, \ldots, n$, and $\kappa_{s_t} = \exp(\mu_{s_t}/2)$, the following alternative representation of Eq. (A.7)-(A.9) as a state space model with respect to $\alpha_{1:n}$ is obtained,

$$
y_{1,t} = (\gamma_{s_t}\bar{z}_t + \sqrt{\bar{z}_t} \epsilon_t) \exp(\alpha_t/2)\kappa_{s_t}, \quad (A.26)
$$

$$
y_{2,t} = \zeta + \xi(\mu_{s_t} + \alpha_t) + u_t, \quad t = 1, \ldots, n,
$$

$$
\alpha_{t+1} = \phi_{s_{t+1}} \alpha_t + \eta_{t+1}, \quad t = 1, \ldots, n - 1. \quad (A.27)
$$

To achieve more efficient sampling, $\alpha_{1:n}$ is divided into $K + 1$ blocks at random, say $(\alpha_{k_{i-1}}^{\prime}, \ldots, \alpha_{k_i})$ for $i = 1, \ldots, K + 1$ with $k_0 = 0$ and $k_{K+1} = n$, where $K$ is a tuning parameter. I use stochastic knots given by $k_i = \text{int}[n(i + \mathcal{U}_i)/(K + 2)]$, for $i = 1, \ldots, K$ and $k_i - k_{i-1} \geq 2$, with $\mathcal{U}_i$ a random sample from the uniform distribution $\mathcal{U}(0,1)$ (Shephard and Pitt, 1997).

Suppose that $k_{i-1} = r$ and $k_i = r + d$ for the $i^{th}$ block. The sampler then exploits the independence of disturbances $\eta_t$, block sampling $\eta = (\eta_{r+1}, \ldots, \eta_{r+d})$’ instead of $\underline{\alpha} = (\alpha_{r+1}, \ldots, \alpha_{r+d})$’ from the respective full joint conditional posterior density ($r \geq 0, d \geq 2, r + d \leq n$),

$$
\pi(\eta | \alpha_{r+d+1}, s_{r+d+1}, \Theta) \propto \left( \prod_{t=r+1}^{r+d} f(y_t | \alpha_t, \alpha_{t+1}) \right) \left( \prod_{t=r+1}^{r+d} f(\eta_t) \right) f(\alpha_{r+d}), \quad r = 0, \quad (A.28)
$$

$$
\pi(\eta | \alpha_r, \alpha_{r+d+1}, s_r, s_{r+d+1}, y_{1:r}, z_r, \Theta) \propto f(y_{1:r} | \alpha_r, \alpha_{r+1}) \left( \prod_{t=r+1}^{r+d} f(y_t | \alpha_t, \alpha_{t+1}) \right) \left( \prod_{t=r+1}^{r+d} f(\eta_t) \right) f(\alpha_{r+d}), \quad r + d < n, \quad (A.29)
$$

$$
\pi(\eta | \alpha_r, s_r, y_{1:r}, z_r, \Theta) \propto f(y_{1:r} | \alpha_r, \alpha_{r+1}) \left( \prod_{t=r+1}^{r+d-1} f(y_t | \alpha_t, \alpha_{t+1}) \right) f(y_n | \alpha_n) \left( \prod_{t=r+1}^{r+d} f(\eta_t) \right), \quad r + d = n,
$$

where $\underline{\alpha} = (\alpha_{r+1}, \ldots, \alpha_{r+d})$ and $s_r = (s_{r+1}, \ldots, s_{r+d})$.
where \( y_t = (y_{1,t}, y_{2,t})' \) and \( \Theta \equiv \{ y_{r+1:r+d}, s_{r+1:r+d}, z_{r+1:r+d}, \theta \} \). The logarithm of \( f(y_{1,r}|\alpha_r, \alpha_{r+1}), f(y_t|\alpha_t, \alpha_{t+1}), f(y_n|\alpha_n) \) in Eq. (A.28)-(A.29) is (excluding constant terms)
\[
l_t = -\frac{\alpha_t}{2} - \frac{(y_{1,t} - \tilde{\mu}_t)^2}{2\tilde{\sigma}_t^2} - \frac{(\tilde{y}_{2,t} - \xi \alpha_t)^2}{2\sigma_u^2} & I_{t>r},
\]
with
\[
\tilde{\mu}_t = \begin{cases} 
[\gamma_{s,t} \bar{z}_t + \rho_{s_{t+1}} \sqrt{\bar{z}_t} (\alpha_{t+1} - \phi s_{t+1} \alpha_t)] / \sigma_{s_{t+1}} \exp(\alpha_t/2) \kappa_{s_t} & t < n \\
\gamma_{s,n} \bar{z}_n \exp(\alpha_n/2) \kappa_{s_n} & t = n,
\end{cases}
\]
\[
\tilde{\sigma}_t^2 = \begin{cases} 
(1 - \rho_{s_{t+1}}^2) \bar{z}_t \exp(\alpha_t) \kappa_{s_t}^2 & t < n \\
\bar{z}_n \exp(\alpha_n) \kappa_{s_n}^2 & t = n,
\end{cases}
\]
\[
\tilde{y}_{2,t} = y_{2,t} - \zeta - \xi \mu_{s_t}.
\]
Then the logarithm of Eq. (A.28)-(A.29) can be written as
\[
L - \sum_{t=r+1}^{r+d} \eta_t^2/(2\sigma_{s_t}^2), \text{ where (excluding constant terms)}
\]
\[
L = \log f(\alpha_{r+d}) I_{r+d<n} + \begin{cases} 
\sum_{r+1}^{r+d} l_t & r = 0 \\
\sum_{r}^{r+d} l_t & r > 0,
\end{cases}
\]
\[
\log f(\alpha_{r+d}) = -\frac{(\alpha_{r+d+1} - \phi s_{r+d+1} \alpha_{r+d})^2}{2\sigma_{s_{r+d+1}}^2}.
\]

To construct a proposal density based on a normal approximation of the posterior density of \( \eta \), define further
\[
\delta = (\delta_{r+1}, \ldots, \delta_{r+d})', \quad \delta_t = \frac{\partial L}{\partial \alpha_t},
\]
\[
Q = -E \left( \frac{\partial^2 L}{\partial \alpha \partial \alpha'} \right) = \begin{pmatrix} A_{r+1} & B_{r+2} & 0 & \cdots & 0 \\
B_{r+2} & A_{r+2} & B_{r+3} & \cdots & 0 \\
0 & B_{r+3} & A_{r+3} & \cdots & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots \\
0 & \cdots & 0 & B_{r+d} & A_{r+d} \end{pmatrix},
\]
\[
A_t = -E \left( \frac{\partial^2 L}{\partial \alpha_t^2} \right), \quad t = r + 1, \ldots, r + d,
\]
\[
B_t = -E \left( \frac{\partial^2 L}{\partial \alpha_t \partial \alpha_{t-1}} \right), \quad t = r + 2, \ldots, r + d, \quad B_{r+1} = 0.
\]
The first derivatives are
\[
\delta_t = -\frac{1}{2} + \frac{(y_{1,t} - \tilde{\mu}_t)^2}{2\tilde{\sigma}_t^2} + \frac{y_{1,t} - \hat{\mu}_t}{\tilde{\sigma}_t^2} \cdot \frac{\partial \hat{\mu}_t}{\partial \alpha_t} + \frac{y_{t-1} - \tilde{\mu}_{t-1}}{\tilde{\sigma}_{t-1}^2} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t} + \frac{\xi (\tilde{y}_{2,t} - \xi \alpha_t)}{\sigma_u^2} + j(\alpha_t),
\]
where
\[
\frac{\partial \hat{\mu}_t}{\partial \alpha_t} = \begin{cases} 
\gamma_{s,t} \tilde{z}_t + \rho_{s,t} \sqrt{z_t} \left(-\phi_{s,t+1} + \frac{\alpha_{t+1} - \phi_{s,t+1} \alpha_t}{2}\right)/\sigma_{s,t+1} \times \exp(\alpha_t/2)\kappa_{s,t} & t < n \\
0 & t = n,
\end{cases}
\]
\[
\frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t} = \begin{cases} 
0 & t = 1 \\
\rho_{s,t} \sqrt{z_{t-1}} \exp(\alpha_{t-1}/2)\kappa_{s,t-1}/\sigma_{s,t} & t > 1,
\end{cases}
\]
\[
\frac{\phi_{s,t+1} (\alpha_{t+1} - \phi_{s,t+1} \alpha_t)}{\sigma_{s,t+1}^2} & t = r + d < n \\
0 & \text{else}.
\]

For the second derivatives, take expectations with respect to \(y_t\) multiplied by \(-1\) and obtain
\[
A_t = \frac{1}{2} + \tilde{\sigma}_t^{-2} \left(\frac{\partial \hat{\mu}_t}{\partial \alpha_t}\right)^2 + \tilde{\sigma}_{t-1}^{-2} \left(\frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t}\right)^2 + \frac{\xi^2}{\sigma_u^2} + j'(\alpha_t),
\]
\[
B_t = \tilde{\sigma}_{t-1}^{-1} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_{t-1}} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t},
\]
where
\[
\frac{\phi_{s,t+1}^2 / \sigma_{s,t+1}^2}{\sigma_{s,t+1}^2} & t = r + d < n \\
0 & \text{else}.
\]
Applying a second order Taylor expansion to the log of the posterior density around mode $\eta_\hat{}$ yields an approximating normal density as follows (excluding constant terms):

$$
\log \pi(\eta | \cdot) \\
\approx \hat{L} + \frac{\partial L}{\partial \eta'}|_{\eta = \hat{\eta}} (\eta - \hat{\eta}) + \frac{1}{2} (\eta - \hat{\eta})' \left( \frac{\partial^2 L}{\partial \eta \partial \eta'} \right)|_{\eta = \hat{\eta}} (\eta - \hat{\eta}) - \frac{1}{2} \sum_{t=r+1}^{r+d} \frac{\eta^2_t}{\sigma^2_{st}} \\
= \hat{L} + \hat{\delta}'(\alpha - \hat{\alpha}) - \frac{1}{2} (\alpha - \hat{\alpha})' \hat{Q}(\alpha - \hat{\alpha}) - \frac{1}{2} \sum_{t=r+1}^{r+d} \frac{\eta^2_t}{\sigma^2_{st}} \\
\equiv \log q(\eta | \cdot),
$$

where $\hat{L}$, $\hat{\delta}$ and $\hat{Q}$ are the values of $L$, $\delta$ and $Q$ at $\alpha = \hat{\alpha}$ (or at $\eta = \hat{\eta}$, equivalently). It can be shown that proposal density $q(\eta | \cdot)$ is the posterior of $\eta$ obtained from an ordinary linear state space model given by Eq. (A.30)-(A.31) below (see Omori and Watanabe, 2007, in a related context). Mode $\eta_\hat{}$ can be obtained by repeating the following algorithm until it converges (usually 2 to 5 iterations).

**Algorithm 1 (Disturbance smoother):**

1. Initialize $\eta_\hat{}$, and compute $\hat{\alpha}$ at $\eta = \hat{\eta}$ using state equation (A.27) recursively.

2. Evaluate $\hat{\delta}_t$, $\hat{A}_t$, and $\hat{B}_t$ at $\alpha = \hat{\alpha}, t = r + 1, \ldots, r + d$.

3. Let $\hat{D}_{r+1} = \hat{A}_{r+1}$ and $\hat{b}_{r+1} = \hat{\delta}_{r+1}$. Compute the following variables recursively for $t = r + 2, \ldots, r + d$:

   $$
   \hat{D}_t = \hat{A}_t - \hat{D}_{t-1}^{-1}\hat{B}_t^2, \quad \hat{K}_t = \sqrt{\hat{D}_t}, \\
   \hat{b}_t = \hat{\delta}_t - \hat{B}_t \hat{D}_{t-1}^{-1}\hat{b}_{t-1}.
   $$

4. Define auxiliary variable $\hat{y}_t = \hat{\varphi}_t + \hat{D}_t^{-1}\hat{b}_t$, where $\hat{\varphi}_t = \hat{\alpha}_t + \hat{D}_t^{-1}\hat{B}_{t+1}\hat{\alpha}_{t+1}, t = r + 1, \ldots, r + d$, and set $\hat{B}_{d+r+1} = 0$, $\hat{\alpha}_{r+d+1} = \alpha_{r+d+1}$.
5. Consider the linear Gaussian state space model formulated by

$$
\hat{y}_t = Z_t \alpha_t + G_t e_t, \quad t = r + 1, \ldots, r + d,
$$

(A.30)

$$
\alpha_{t+1} = \phi_{s_{t+1}} \alpha_t + H_{t+1} u_{t+1}, \quad t = r, \ldots, r + d - 1, \quad t > 0,
$$

$$
\alpha_1 = \mathcal{N}(0, \sigma_{s_1}/\sqrt{1 - \phi_{s_1}}), \quad t = 0,
$$

\begin{pmatrix}
e_t \\
u_{t+1}
\end{pmatrix} \sim \mathcal{N}(0, I_2),

(A.31)

where

$$
Z_t = 1 + \phi_{s_{t+1}} \hat{D}_t^{-1} \hat{B}_{t+1}, \quad G_t = (\hat{K}_t^{-1}, \hat{D}_t^{-1} \hat{B}_{t+1} \sigma_{s_{t+1}}), \quad H_{t+1} = (0, \sigma_{s_{t+1}}).
$$

Apply the Kalman filter (e.g. Durbin and Koopman, 2008) and disturbance smoother (Koopman, 1993) to this state space model to obtain posterior mode $\hat{\eta}$ or $\hat{\alpha}$, equivalently.

6. Go to 2.

In the McMC sampling procedure, the current sample of $\eta$ may be taken as initial value of $\hat{\eta}$ in Step 1. Further note that the above steps are equivalent to the method of scoring used to maximize the conditional posterior density. After convergence of Algorithm 1, $\eta$ is sampled from the conditional posterior density by conducting an AR (Accept - Reject) - MH algorithm (Tierney, 1994) using the simulation smoother (de Jong and Shephard, 1995, or Durbin and Koopman, 2002).

**Algorithm 2 (AR-MH step and simulation smoother):**

1. Let $\eta_0$ denote the current value. Find mode $\hat{\eta}$ using Algorithm 1.

2. Proceed with Steps 2-4 in Algorithm 1 to obtain the approximated linear Gaussian state space model of Eq. (A.30)-(A.31).

3. Propose a candidate $\eta^*$ by sampling from $\tilde{q}(\eta^*) \propto \min\{\pi(\eta^*|\cdot), cq(\eta^*|\cdot)\}$ using the AR algorithm as follows:
(a) Generate $\eta^*$ using the simulation smoother for the approximated state space model of Eq. (A.30)-(A.31).

(b) Accept $\eta^*$ with probability

$$\frac{\min \{\pi(\eta^*|\cdot), cq(\eta^*|\cdot)\}}{cq(\eta^*|\cdot)}$$

where $c$ is a scaling constant. If it is rejected, go to (a).

4. Conduct the MH algorithm using candidate $\eta^*$, with acceptance probability

$$\min \left\{ \frac{\pi(\eta^*|\cdot) \times \min \{\pi(\eta_0|\cdot), cq(\eta_0|\cdot)\}}{\pi(\eta_0|\cdot) \times \min \{\pi(\eta^*|\cdot), cq(\eta^*|\cdot)\}}, 1 \right\}.$$ 

A.B The $\mathcal{GH}$ Skew Student-$t$ Distribution

This appendix includes some important properties of the generalized hyperbolic skew Student-$t$ distribution ($\mathcal{GH}$ skew-$t$). For a more complete treatment, see Aas and Haff (2006). Consider further Scott, Würtz, Dong, and Tran (2011), Hu (2009), and Prause (1999). The $\mathcal{GH}$ skew-$t$ distribution is a limiting case of the $\mathcal{GH}$ distribution, which was introduced in Barndorff-Nielsen (1977). Starting with the latter, a possible parameterization is as follows,

$$f_{\mathcal{GH}}(x) = \frac{(\beta^2 - \gamma^2)^{1/2} K_{\lambda-1/2} \left( \beta \sqrt{\delta^2 + (x - \mu)^2} \right) \exp(\gamma(x - \mu))}{\sqrt{2\pi} \beta^{1/2} \delta^\lambda K_{\lambda} \left( \delta \sqrt{\beta^2 - \gamma^2} \right) \left( \delta + (x - \mu)^2 \right)^{1/2-\lambda}}, \quad (A.32)$$

where $K_j$ is the modified Bessel function of the third kind of order $j$ (Abramowitz and Stegun, 1972). Parameters must fulfill the conditions

$$\delta \geq 0, \ |\gamma| < \beta \quad \text{if} \ \lambda > 0,$$

$$\delta > 0, \ |\gamma| < \beta \quad \text{if} \ \lambda = 0,$$

$$\delta > 0, \ |\gamma| \leq \beta \quad \text{if} \ \lambda < 0.$$
The tails of the generalized hyperbolic (GH) distribution behave as

\[ f_{GH}(x) \sim \text{const}|x|^\lambda \exp(-\beta |x| + \gamma x) \quad \text{as} \quad x \to \pm \infty \forall \lambda. \]  

(A.33)

It follows that as long as $|\gamma| \neq \beta$, both tails are semi-heavy.

The GH distribution can be represented as a normal mean-variance mixture with generalized inverse Gaussian (GIG) distributed mixture variable (Barndorff-Nielsen and Blæsøid, 1981),

\[ X = \mu + \gamma Z + \sqrt{Z}Y, \quad Y \sim N(0,1), \quad Z \sim GIG(\lambda,\delta,\psi), \]  

(A.34)

and $\psi = \sqrt{\beta^2 - \gamma^2}$. It follows from Eq. (A.34) that $X|Z = z \sim N(\mu + \gamma z, z)$. However, parameters are difficult to estimate due to flatness of the likelihood function (see e.g. Aas and Haff, 2006). A parsimonious representation that is more amenable to estimation can be obtained by setting $\lambda = -\nu/2 (\nu > 0)$, $\delta = \sqrt{\nu}$, and $\beta \to |\gamma|$ in Eq. (A.32). Then $Z \sim GIG(-\nu/2, \sqrt{\nu}, 0)$ or equivalently $Z \sim IG(\nu/2, \nu/2)$. This limiting case of the GH distribution is referred to as the GH skew-$t$ distribution (e.g. Barndorff-Nielsen and Shephard, 2001b), with skew parameter $\gamma$, degrees of freedom parameter $\nu$ and mixture representation

\[ X = \mu + \gamma Z + \sqrt{Z}Y, \quad Y \sim N(0,1), \quad Z \sim IG(\nu/2, \nu/2). \]

Its closed form density is given by

\[ f_{GH_{skew}}(x) = 2^{\frac{\nu+1}{2}} \frac{\Gamma\left(\frac{\nu}{2}\right)}{\sqrt{\pi \nu \Gamma\left(\frac{\nu+1}{2}\right)}} K_{\nu+1}\left(\sqrt{\gamma^2(\nu + (x - \mu)^2)}\right) \exp(\gamma(x - \mu)), \quad \gamma \neq 0, \]

and

\[ f_{GH_{t}}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi \nu \Gamma\left(\frac{\nu}{2}\right)}} \left[1 + \frac{(x - \mu)^2}{\nu}\right]^{-\frac{(\nu+1)/2}{\nu}}, \quad \gamma = 0. \]

The latter density is commonly referred to as the non-central Student-$t$ distribution with $\nu$ degrees of freedom, expectation $\mu$, and variance $\nu/(\nu - 2)$. 

\[ A.B. \ \text{THE } G'H \ \text{skew-}t \ \text{DISTRIBUTION} \]
The first four moments of a $\mathcal{GH}$ skew-$t$ distributed random variate $X$ are (Aas and Haff, 2006)

\[
\begin{align*}
E(X) &= \mu + \frac{\gamma \nu}{\nu - 2}, \\
\text{var}(X) &= \frac{2\gamma^2\nu^2}{(\nu - 2)^2(\nu - 4)} + \frac{\nu}{\nu - 2}, \\
s(X) &= \frac{2\gamma[\nu(\nu - 4)]^{1/2}}{[2\gamma^2\nu + (\nu - 2)(\nu - 4)]^{3/2}} \left[ 3(\nu - 2) + \frac{8\gamma^2\nu}{\nu - 6} \right], \\
k(X) &= \frac{6}{[2\gamma^2\nu + (\nu - 2)(\nu - 4)]^2} \times \\
&\quad \left[ (\nu - 2)^2(\nu - 4) + \frac{16\gamma^2\nu(\nu - 2)(\nu - 4)}{\nu - 6} + \frac{8\gamma^4\nu^2(5\nu - 22)}{(\nu - 6)(\nu - 8)} \right].
\end{align*}
\]

We observe that for the mean and variance to exist, $\nu > 2$ and $\nu > 4$, respectively. Skewness and excess kurtosis are defined only if $\nu > 6$ and $\nu > 8$, respectively.

From Eq. (A.33) it follows that the tails of the $\mathcal{GH}$ skew-$t$ distribution are given by

\[
\begin{align*}
\mathcal{GH}_{\text{skew}}(x) &\sim \text{const}|x|^{-\nu/2-1} \exp(-|\gamma x| + \gamma x) \quad \text{as} \; x \to \pm \infty.
\end{align*}
\]

Thus we have a heavy tail decaying as

\[
\mathcal{GH}_{\text{skew}}(x) \sim \text{const}|x|^{-\nu/2-1} \quad \text{if} \quad \begin{cases} 
\gamma < 0 \text{ and } x \to -\infty \\
\gamma > 0 \text{ and } x \to +\infty,
\end{cases}
\]

and a light tail decaying as

\[
\mathcal{GH}_{\text{skew}}(x) \sim \text{const}|x|^{-\nu/2-1} \exp(-2|\gamma x|) \quad \text{if} \quad \begin{cases} 
\gamma < 0 \text{ and } x \to +\infty \\
\gamma > 0 \text{ and } x \to -\infty.
\end{cases}
\]

The heavy tail shows polynomial and the light tail exponential behavior. Thus the $\mathcal{GH}$ skew-$t$ distribution is able to model substantially skewed and heavy tailed data, as found in e.g. financial markets. This feature distinguishes it well from alternative skew Student-$t$ distributions in the literature (for a more detailed discussion, see Aas
Figure A.13: The $G\mathcal{H}$ skew-$t$ distribution: Tail Behavior and Identification

and Haff, 2006). The tails of the $G\mathcal{H}$ skew-$t$ distribution are characterized uniquely by parameters $\gamma$ and $\nu$, which determine jointly the degree of skewness and heavy tailedness. Fig. A.13 demonstrates how both parameters shift probability mass into the tails of the distribution (upper two plots). Observe that a specific skewness and kurtosis of the distribution can be obtained by different combinations of higher moment parameters $\gamma$ and $\nu$ (lower two plots).

Finally, note that the heavy tail of the $G\mathcal{H}$ skew-$t$ distribution is heavier than the tails of the symmetric Student-$t$ distribution, which decay as

$$f_{G\mathcal{H}t}(x) \sim \text{const}|x|^{-\nu-1} \quad \text{as} \quad x \to \pm \infty.$$
A.C Inefficiency Factor / Convergence Diagnostic

The inefficiency factor is the estimated ratio of the numerical variance of the posterior sample mean to the variance of the hypothetical sample mean from uncorrelated draws. With an inefficiency factor of \( m \), we need to draw McMC samples \( m \) times as many as uncorrelated samples. The following formula is used to calculate the inefficiency factor,

\[
IF = 1 + 2 \frac{b_w}{b_w - 1} \sum_{i=1}^{b_w} K_{QS} \left( \frac{i}{b_w} \right) \hat{\rho}_i, \tag{A.35}
\]

where \( \hat{\rho}_i \) is the sample autocorrelation at lag \( i \), \( K_{QS} \) the Quadratic Spectral (QS) kernel and \( b_w \) the bandwidth parameter (the correction term \( b_w/(b_w - 1) \) dropped if \( b_w = 1 \)). Andrews (1991) finds the QS kernel to be superior in terms of an asymptotic truncated mean squared error criterion, relative to other kernels. The QS kernel is defined as

\[
K_{QS}(x) = \frac{25}{12\pi^2x^2} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right).
\]

The bandwidth \( b_w \) is determined automatically as a function of the data, following Andrews (1991). Specifically,

\[
\hat{b}_w = 1.3221(\hat{\alpha}(2)N)^{1/5},
\]

where \( N \) is the number of collected draws and

\[
\hat{\alpha}(2) = \frac{4\hat{\rho}_a^2\hat{\sigma}_a^4}{(1 - \hat{\rho}_a)^8} \left[ \frac{\hat{\sigma}_a^4}{(1 - \hat{\rho}_a)^4} \right],
\]

with \( \{\hat{\rho}_a, \hat{\sigma}_a\} \) the autoregressive and variance innovation parameter, respectively, obtained by running a first-order autoregressive linear regression on the demeaned draws of the Markov chain. As Andrews (1991) notes, there can be sensitivity of the estimator to the choice of the bandwidth parameter. In the current context, this would suggest obtaining multiple bandwidth values by varying \( \hat{\rho}_a \) by one or two standard deviations, \( \hat{\rho}_a \pm 1/ \sqrt{N} \) or \( \hat{\rho}_a \pm 2/ \sqrt{N} \), respectively.
Geweke (1992) proposes a widely used convergence diagnostic (CD) to test if a Markov chain actually has converged. The basic idea is to compare values sampled early in the sequence (subsample $A$) with those late in the sequence (subsample $B$). Denote by $\theta^{(i)}$ the $i$th draw of parameter $\theta$ in the recorded $N$ draws, and calculate $\bar{\theta}_A = \frac{1}{n_A} \sum_{i=1}^{n_A} \theta^{(i)}$, $\bar{\theta}_B = \frac{1}{n_B} \sum_{i=N-n_B+1}^{N} \theta^{(i)}$. Further, let $\sigma_A$ and $\sigma_B$ denote the standard errors of $\bar{\theta}_A$ and $\bar{\theta}_B$, respectively. The convergence diagnostic is then defined as

$$CD = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\sigma_A^2 + \sigma_B^2}} \xrightarrow{d} N(0, 1),$$

under the assumption that sequence $\theta^{(i)}$ is stationary. For window lengths $n_A$ and $n_B$, Geweke (1992) proposes 0.1$N$ and 0.5$N$, respectively, which have become standard in the literature and are also employed in this work. With a corresponding estimate of the inefficiency factor in Eq. (A.35) at hand, variances $\sigma_A^2, \sigma_B^2$ are calculated as $\sigma_A^2 = s_A^2(b_{w,A} + 1)^{-1}IF_A$, $\sigma_B^2 = s_B^2(b_{w,B} + 1)^{-1}IF_B$, where $s_A^2$ and $s_B^2$ are the variances of subsamples $A$ and $B$, respectively.

### A.D Specification Tests

Tab. A.17 reports the Kupier statistic of probability integral transforms $\hat{u}_t$ and skewness and kurtosis estimates of transformed forecasts $\hat{n}_t$. Standard errors are provided by repeatedly applying the associated particle filter ten times, see Sec. A.4. The Kupier statistic tests for uniformity and is equally sensitive for all values of $\hat{u}_t$. In contrast, the more well-known Kolmogorov-Smirnov statistic is most sensitive around the median. The Kupier statistic $V$ is defined as

$$D^+ = \max_{1 \leq t \leq n} [t/n - u_t], \quad D^- = \max_{1 \leq t \leq n} [u_t - (t - 1)/n], \quad V = D^+ + D^-,$$

under the null of uniformly distributed $u_t$. For all models using no auxiliary equation or RV the Kupier statistic is rather close to the critical value of 2.001 for the 1% significance level\(^{16}\). Especially the

\(^{16}\)The asymptotic distribution of the statistic is known, but Stephens (1970) provides a simplified finite-sample statistic with negligible error to its asymptotic distribution.
Figure A.14: Transformed Predictions $\hat{u}_t$ (in-sample, grand averages). Horizontal dashed lines are 95% confidence intervals based on a normal approximation of the binomially distributed bin elements ($H_0: \hat{u}_t \sim U(0, 1)$). Each bin contains 5% of the data. Y-axis in hundreds.
Table A.17: Specification Tests I

<table>
<thead>
<tr>
<th></th>
<th>Kupier (s.e.×10)</th>
<th>Skewness (s.e.×10^2)</th>
<th>Kurtosis (s.e.×10^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV</td>
<td>3.06 (0.2)</td>
<td>0.089 (0.1)</td>
<td>-0.062 (0.2)</td>
</tr>
<tr>
<td>SV_t</td>
<td>2.55 (0.3)</td>
<td>0.021 (0.1)</td>
<td>-0.134 (0.2)</td>
</tr>
<tr>
<td>SVskt</td>
<td>2.63 (0.3)</td>
<td>0.019 (0.1)</td>
<td>-0.153 (0.2)</td>
</tr>
<tr>
<td>MS3SV_t</td>
<td>2.38 (0.2)</td>
<td>0.003 (0.9)</td>
<td>-0.133 (3.5)</td>
</tr>
<tr>
<td>MS3SVskt</td>
<td>2.43 (0.4)</td>
<td>0.014 (0.5)</td>
<td>-0.105 (1.8)</td>
</tr>
<tr>
<td>SV-RV</td>
<td>2.88 (0.2)</td>
<td>0.128 (0.0)</td>
<td>-0.008 (0.1)</td>
</tr>
<tr>
<td>SV_t-RV</td>
<td>2.61 (0.2)</td>
<td>0.077 (0.6)</td>
<td>-0.066 (2.3)</td>
</tr>
<tr>
<td>SVskt-RV</td>
<td>2.62 (0.1)</td>
<td>0.082 (0.5)</td>
<td>-0.048 (2.1)</td>
</tr>
<tr>
<td>MS4SV-RV</td>
<td>2.62 (0.2)</td>
<td>0.053 (0.0)</td>
<td>-0.128 (0.1)</td>
</tr>
<tr>
<td>MS4SV_t-RV</td>
<td>2.39 (0.1)</td>
<td>0.023 (0.1)</td>
<td>-0.126 (0.3)</td>
</tr>
<tr>
<td>MS4SVskt-RV</td>
<td>2.46 (0.2)</td>
<td>0.021 (0.0)</td>
<td>-0.130 (0.2)</td>
</tr>
<tr>
<td>SVt-Range</td>
<td>3.18 (0.2)</td>
<td>0.074 (0.1)</td>
<td>-0.114 (0.3)</td>
</tr>
<tr>
<td>SVskt-Range</td>
<td>3.31 (0.3)</td>
<td>0.036 (0.4)</td>
<td>-0.200 (1.6)</td>
</tr>
<tr>
<td>MS4SV-Range</td>
<td>7.04 (0.3)</td>
<td>-0.238 (0.1)</td>
<td>-0.133 (0.5)</td>
</tr>
<tr>
<td>MS4SV_t-Range</td>
<td>7.04 (0.3)</td>
<td>-0.253 (0.1)</td>
<td>-0.137 (0.2)</td>
</tr>
<tr>
<td>MS4SVskt-Range</td>
<td>7.45 (0.5)</td>
<td>-0.290 (0.1)</td>
<td>-0.143 (0.2)</td>
</tr>
<tr>
<td>SV-VIX</td>
<td>4.00 (0.2)</td>
<td>0.173 (0.0)</td>
<td>0.005 (0.1)</td>
</tr>
<tr>
<td>SV_t-VIX</td>
<td>3.02 (0.2)</td>
<td>0.077 (0.1)</td>
<td>-0.113 (0.3)</td>
</tr>
<tr>
<td>SVskt-VIX</td>
<td>3.07 (0.1)</td>
<td>0.086 (0.1)</td>
<td>-0.097 (0.7)</td>
</tr>
<tr>
<td>MS2SV-VIX</td>
<td>3.92 (0.3)</td>
<td>0.163 (0.1)</td>
<td>-0.016 (0.4)</td>
</tr>
<tr>
<td>MS2SV_t-VIX</td>
<td>3.07 (0.5)</td>
<td>0.068 (0.1)</td>
<td>-0.135 (0.3)</td>
</tr>
<tr>
<td>MS2SVskt-VIX</td>
<td>3.16 (0.2)</td>
<td>-0.135 (0.1)</td>
<td>-0.304 (0.2)</td>
</tr>
</tbody>
</table>

Kupier statistic of \( \hat{u}_t \), skewness, and kurtosis of transformed forecasts \( \hat{u}_t \). Critical value for \( V \) is 2.001 at the 1% level (\( H_0: \hat{u}_t \sim \mathcal{U}(0,1) \)). Numerical standard error in parentheses.

regime switching range variants, but also the SV-VIX and MS2SV-VIX perform worse, mirroring the respective predictive likelihoods in Tab. A.13 or equally, the LPDRs in Tab. A.18. A graphical analysis of histograms of \( \hat{u}_t \) together with approximate confidence intervals in the spirit of Bauwens, Giot, Grammig, and Veredas (2004) visualizes these deficits in predictive ability, see Fig. A.14. The remaining skewness in the transformed one-step-ahead forecasts appears in general small and is positive in almost all models, indicating that the negative outliers have been appropriately modeled. However, similarly to the Kupier statistic, the regime switching range variants negatively stand out. Interestingly, there is some remaining negative skewness for the MS2SVskt-VIX model, despite its second state modeling extreme negative
<table>
<thead>
<tr>
<th>Model</th>
<th>LPDR (s.e.)</th>
<th>Rank</th>
<th>$LB^2_{15}$ (s.e.)</th>
<th>$LB^3_{15}$ (s.e.)</th>
<th>$LB^4_{15}$ (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$ (return)</td>
<td>5653.5</td>
<td>877.7</td>
<td>2183.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV</td>
<td>0 (0.3)</td>
<td>14</td>
<td>14.2 (0.1)</td>
<td>40.0 (0.1)</td>
<td>14.7 (0.1)</td>
</tr>
<tr>
<td>$SV_t$</td>
<td>14.7 (0.7)</td>
<td>12</td>
<td>15.6 (0.2)</td>
<td>38.6 (0.2)</td>
<td>15.8 (0.2)</td>
</tr>
<tr>
<td>$SV_{skt}$</td>
<td>26.1 (1.0)</td>
<td>8</td>
<td>16.1 (0.1)</td>
<td>39.4 (0.2)</td>
<td>16.1 (0.2)</td>
</tr>
<tr>
<td>MS3$SV_t$</td>
<td>33.9 (0.9)</td>
<td>7</td>
<td>15.6 (0.2)</td>
<td>45.7 (0.2)</td>
<td>16.4 (0.2)</td>
</tr>
<tr>
<td>MS3$SV_{skt}$</td>
<td>43.8 (1.5)</td>
<td>5</td>
<td>16.5 (0.3)</td>
<td>45.6 (0.3)</td>
<td>17.3 (0.3)</td>
</tr>
<tr>
<td>SV-RV</td>
<td>24.9 (0.3)</td>
<td>9</td>
<td>15.5 (0.0)</td>
<td>50.1 (0.1)</td>
<td>13.6 (0.0)</td>
</tr>
<tr>
<td>$SV_t$-RV</td>
<td>41.8 (0.3)</td>
<td>6</td>
<td>15.2 (0.0)</td>
<td>47.6 (0.0)</td>
<td>12.9 (0.0)</td>
</tr>
<tr>
<td>$SV_{skt}$-RV</td>
<td>51.0 (0.5)</td>
<td>4</td>
<td>15.1 (0.1)</td>
<td>46.6 (0.1)</td>
<td>12.8 (0.0)</td>
</tr>
<tr>
<td>MS4$SV$-RV</td>
<td>59.9 (0.2)</td>
<td>2</td>
<td>16.2 (0.1)</td>
<td>47.6 (0.1)</td>
<td>13.1 (0.0)</td>
</tr>
<tr>
<td>MS4$SV_t$-RV</td>
<td>62.8 (0.2)</td>
<td>1</td>
<td>16.2 (0.0)</td>
<td>46.1 (0.1)</td>
<td>13.2 (0.1)</td>
</tr>
<tr>
<td>MS4$SV_{skt}$-RV</td>
<td>51.6 (0.3)</td>
<td>3</td>
<td>16.3 (0.0)</td>
<td>46.7 (0.1)</td>
<td>13.2 (0.0)</td>
</tr>
<tr>
<td>$SV_t$-Range</td>
<td>-46.8 (0.3)</td>
<td>19</td>
<td>18.3 (0.1)</td>
<td>100.4 (0.1)</td>
<td>17.6 (0.1)</td>
</tr>
<tr>
<td>$SV_{skt}$-Range</td>
<td>-36.5 (0.9)</td>
<td>18</td>
<td>18.2 (0.1)</td>
<td>93.5 (0.3)</td>
<td>17.0 (0.1)</td>
</tr>
<tr>
<td>MS4$SV$-Range</td>
<td>-116.5 (1.0)</td>
<td>20</td>
<td>20.1 (0.3)</td>
<td>31.3 (0.3)</td>
<td>24.2 (0.3)</td>
</tr>
<tr>
<td>MS4$SV_t$-Range</td>
<td>-117.9 (1.0)</td>
<td>21</td>
<td>19.9 (0.4)</td>
<td>31.0 (0.3)</td>
<td>24.2 (0.3)</td>
</tr>
<tr>
<td>MS4$SV_{skt}$-Range</td>
<td>-137.1 (1.1)</td>
<td>22</td>
<td>19.8 (0.2)</td>
<td>28.9 (0.2)</td>
<td>24.3 (0.2)</td>
</tr>
<tr>
<td>SV-VIX</td>
<td>-25.0 (0.4)</td>
<td>17</td>
<td>30.1 (0.1)</td>
<td>49.3 (0.1)</td>
<td>36.4 (0.2)</td>
</tr>
<tr>
<td>$SV_t$-VIX</td>
<td>24.0 (0.5)</td>
<td>10</td>
<td>28.8 (0.1)</td>
<td>46.0 (0.1)</td>
<td>31.2 (0.1)</td>
</tr>
<tr>
<td>$SV_{skt}$-VIX</td>
<td>11.4 (0.5)</td>
<td>13</td>
<td>28.8 (0.1)</td>
<td>46.3 (0.1)</td>
<td>31.7 (0.1)</td>
</tr>
<tr>
<td>MS2$SV$-VIX</td>
<td>-15.8 (1.4)</td>
<td>16</td>
<td>30.4 (0.6)</td>
<td>47.2 (0.5)</td>
<td>35.3 (1.0)</td>
</tr>
<tr>
<td>MS2$SV_t$-VIX</td>
<td>22.3 (1.7)</td>
<td>11</td>
<td>30.0 (0.3)</td>
<td>44.6 (0.4)</td>
<td>31.1 (0.3)</td>
</tr>
<tr>
<td>MS2$SV_{skt}$-VIX</td>
<td>-5.1 (0.5)</td>
<td>15</td>
<td>24.2 (0.1)</td>
<td>47.1 (0.1)</td>
<td>22.3 (0.1)</td>
</tr>
</tbody>
</table>

Cumulative log predictive density ratio. Ljung-Box statistic of demeaned $\tilde{u}_t^2$, $\tilde{u}_t^3$, $\tilde{u}_t^4$ and $y_{1,t}^2$, $\tilde{y}_{1,t}^3$, $y_{1,t}^4$. Critical values for the $LB(15)$ test are 30.6/25.0/22.3 at the 1%/5%/10% significance level. Numerical standard error in parentheses.

Skewness explicitly. However, as already outlined in the main part, transient regime dynamics may have weaker positive impact on forecasting performance, more so in a time invariant transition probability context. Remaining excess kurtosis in the
data is overall negative and small in absolute value. This is rather satisfying under a conservative viewpoint.

Tab. A.18 reports log predictive density ratios relative to the basic SV model and Ljung-Box statistics at lag 15 for $\hat{u}_t^2$, $\hat{u}_t^3$, $\hat{u}_t^4$ and $y_1$. The LPDRs make the better potential predictive ability of models with no additional time series, RV or that of some VIX variants clearly visible. The $LB(15)$ statistics support this picture, but are satisfying for all models in general, especially for dynamics in the second and fourth moment. The relatively high statistics for the third moment indicate some remaining dynamics, but note that the AR(1) effect, which is common in stock index data, has not been removed. Moreover, Nakajima and Omori (2012) estimate the SVsk$\tau$ model on different sample periods of the S&P 500 and find changing skewness of the empirical return distribution.

## A.E Parameter Estimates

This appendix contains estimation results for the transition probabilities of the presented models. The results for estimated nested models (i.e. models with Gaussian or Student-$t$ error) are also reported. All estimates are based on the full sample 1990/01/02-2010/08/06. If not stated otherwise, priors and McMC parameters are those of the main text.

### SVsk$\tau$ Model Variants

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>$IF$</th>
<th>$CD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.2822</td>
<td>0.1173</td>
<td>[-0.5101, -0.0470]</td>
<td>1.6</td>
<td>0.96</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9871</td>
<td>0.0022</td>
<td>[0.9825, 0.9910]</td>
<td>12.0</td>
<td>0.97</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.1484</td>
<td>0.0106</td>
<td>[0.1292, 0.1696]</td>
<td>186.4</td>
<td>0.99</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.7294</td>
<td>0.0356</td>
<td>[-0.7933, -0.6580]</td>
<td>94.8</td>
<td>0.97</td>
</tr>
<tr>
<td>$\nu$</td>
<td>17.901</td>
<td>3.2304</td>
<td>[12.430, 24.700]</td>
<td>254.7</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value). Bottom panel: Implied higher moments of $\epsilon_t$. 

Table A.19: Estimation Results of the SV$\tau$ Model
### Table A.20: Estimation Results of the SV Model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.2026</td>
<td>0.1143</td>
<td>[-0.4283, 0.0242]</td>
<td>1.2</td>
<td>0.98</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9850</td>
<td>0.0025</td>
<td>[0.9798, 0.9896]</td>
<td>12.5</td>
<td>0.87</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.1590</td>
<td>0.0113</td>
<td>[0.1384, 0.1833]</td>
<td>148.3</td>
<td>0.82</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.6845</td>
<td>0.0360</td>
<td>[-0.7499, -0.6116]</td>
<td>76.6</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).

### MS3SVskt Model Variants

#### Table A.21: Estimation Results of the MS3SVskt Model: Transition Probabilities

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.9766</td>
<td>0.0112</td>
<td>[0.9502, 0.9933]</td>
<td>3.0</td>
<td>0.97</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.0155</td>
<td>0.0093</td>
<td>[0.0028, 0.0378]</td>
<td>3.1</td>
<td>0.91</td>
</tr>
<tr>
<td>$p_{13}$</td>
<td>0.0079</td>
<td>0.0065</td>
<td>[0.0006, 0.0250]</td>
<td>2.9</td>
<td>0.92</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>0.0014</td>
<td>0.0010</td>
<td>[0.0002, 0.0038]</td>
<td>6.3</td>
<td>0.98</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.9970</td>
<td>0.0013</td>
<td>[0.9939, 0.9989]</td>
<td>4.6</td>
<td>0.97</td>
</tr>
<tr>
<td>$p_{23}$</td>
<td>0.0016</td>
<td>0.0009</td>
<td>[0.0004, 0.0037]</td>
<td>2.7</td>
<td>0.98</td>
</tr>
<tr>
<td>$p_{31}$</td>
<td>0.0009</td>
<td>0.0007</td>
<td>[0.0001, 0.0027]</td>
<td>3.9</td>
<td>0.80</td>
</tr>
<tr>
<td>$p_{32}$</td>
<td>0.0020</td>
<td>0.0011</td>
<td>[0.0005, 0.0046]</td>
<td>2.6</td>
<td>0.98</td>
</tr>
<tr>
<td>$p_{33}$</td>
<td>0.9971</td>
<td>0.0012</td>
<td>[0.9943, 0.9990]</td>
<td>2.4</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).

#### Table A.22: Estimation Results of the MS3SVt Model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>1.5364</td>
<td>0.2725</td>
<td>[0.9521, 2.0312]</td>
<td>67.9</td>
<td>0.59</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.1013</td>
<td>0.0776</td>
<td>[-0.0602, 0.2449]</td>
<td>21.1</td>
<td>0.66</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-1.0333</td>
<td>0.0812</td>
<td>[-1.2010, -0.8777]</td>
<td>16.0</td>
<td>0.78</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9563</td>
<td>0.0078</td>
<td>[0.9407, 0.9717]</td>
<td>59.0</td>
<td>0.73</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.1768</td>
<td>0.0129</td>
<td>[0.1520, 0.2037]</td>
<td>210.0</td>
<td>0.84</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.7756</td>
<td>0.0324</td>
<td>[-0.8364, -0.7097]</td>
<td>103.9</td>
<td>0.88</td>
</tr>
<tr>
<td>$\nu$</td>
<td>18.480</td>
<td>3.1091</td>
<td>[13.122, 25.668]</td>
<td>234.1</td>
<td>0.80</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.4338</td>
<td>0.0945</td>
<td>[0.2769, 0.6576]</td>
<td>238.0</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value). Bottom panel: Implied higher moments of $\epsilon_t$. 

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).
Table A.23: Estimation Results of the MS3SVt Model: Transition Probabilities

<table>
<thead>
<tr>
<th>Transition</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{11}$</td>
<td>0.9745</td>
<td>0.0135</td>
<td>[0.9428, 0.9928]</td>
<td>7.1</td>
<td>0.70</td>
</tr>
<tr>
<td>$P_{12}$</td>
<td>0.0168</td>
<td>0.0106</td>
<td>[0.0030, 0.0425]</td>
<td>5.6</td>
<td>0.68</td>
</tr>
<tr>
<td>$P_{13}$</td>
<td>0.0087</td>
<td>0.0077</td>
<td>[0.0006, 0.0283]</td>
<td>5.5</td>
<td>0.88</td>
</tr>
<tr>
<td>$P_{21}$</td>
<td>0.0015</td>
<td>0.0011</td>
<td>[0.0002, 0.0042]</td>
<td>10.5</td>
<td>0.72</td>
</tr>
<tr>
<td>$P_{22}$</td>
<td>0.9970</td>
<td>0.0013</td>
<td>[0.9938, 0.9989]</td>
<td>6.2</td>
<td>0.80</td>
</tr>
<tr>
<td>$P_{23}$</td>
<td>0.0015</td>
<td>0.0009</td>
<td>[0.0003, 0.0036]</td>
<td>3.1</td>
<td>0.95</td>
</tr>
<tr>
<td>$P_{31}$</td>
<td>0.0010</td>
<td>0.0009</td>
<td>[0.0001, 0.0033]</td>
<td>5.1</td>
<td>0.63</td>
</tr>
<tr>
<td>$P_{32}$</td>
<td>0.0020</td>
<td>0.0011</td>
<td>[0.0004, 0.0046]</td>
<td>3.5</td>
<td>0.77</td>
</tr>
<tr>
<td>$P_{33}$</td>
<td>0.9970</td>
<td>0.0013</td>
<td>[0.9939, 0.9990]</td>
<td>3.4</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).

### SVskt-RV Model Variants

Table A.24: Results of the SVt-RV (upper block) / SV-RV Model (lower block)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.2589</td>
<td>0.0931</td>
<td>[-0.4378, -0.0749]</td>
<td>1.5</td>
<td>1.00</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9690</td>
<td>0.0035</td>
<td>[0.9620, 0.9755]</td>
<td>5.2</td>
<td>0.86</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.2218</td>
<td>0.0096</td>
<td>[0.2039, 0.2414]</td>
<td>54.0</td>
<td>0.76</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.4552</td>
<td>0.0291</td>
<td>[-0.5112, -0.3970]</td>
<td>23.1</td>
<td>0.84</td>
</tr>
<tr>
<td>$\nu$</td>
<td>28.581</td>
<td>4.5704</td>
<td>[20.862, 38.382]</td>
<td>202.6</td>
<td>0.92</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.2526</td>
<td>0.0472</td>
<td>[0.1745, 0.3557]</td>
<td>202.5</td>
<td>0.92</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>-0.2224</td>
<td>0.0229</td>
<td>[-0.2679, -0.1783]</td>
<td>42.9</td>
<td>0.98</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.9047</td>
<td>0.0204</td>
<td>[0.8665, 0.9456]</td>
<td>35.0</td>
<td>0.82</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.4072</td>
<td>0.0059</td>
<td>[0.3955, 0.4188]</td>
<td>8.5</td>
<td>0.87</td>
</tr>
<tr>
<td>SNR</td>
<td>0.5449</td>
<td>0.0281</td>
<td>[0.4933, 0.6029]</td>
<td>50.1</td>
<td>0.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.1971</td>
<td>0.0918</td>
<td>[-0.3760, -0.0143]</td>
<td>1.3</td>
<td>0.98</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9683</td>
<td>0.0035</td>
<td>[0.9612, 0.9749]</td>
<td>4.9</td>
<td>0.99</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.2252</td>
<td>0.0095</td>
<td>[0.2066, 0.2440]</td>
<td>51.0</td>
<td>0.97</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.4422</td>
<td>0.0283</td>
<td>[-0.4982, -0.3874]</td>
<td>19.1</td>
<td>0.94</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>-0.2814</td>
<td>0.0191</td>
<td>[-0.3195, -0.2448]</td>
<td>24.4</td>
<td>0.95</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.9029</td>
<td>0.0197</td>
<td>[0.8657, 0.9427]</td>
<td>33.0</td>
<td>1.00</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.4056</td>
<td>0.0058</td>
<td>[0.3942, 0.4169]</td>
<td>7.6</td>
<td>0.97</td>
</tr>
<tr>
<td>SNR</td>
<td>0.5556</td>
<td>0.0280</td>
<td>[0.5023, 0.6118]</td>
<td>45.2</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).

Second panel (SVt-RV only): Implied higher moments of $\epsilon_t$. Lower two panels: Parameters auxiliary measurement equation and implied signal-to-noise ratio.
Table A.25: Estimation Results of the MS4SVskt-RV (upper block) / MS4SVt-RV Model (lower block): Transition Probabilities

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.3512</td>
<td>0.1030</td>
<td>[0.1443, 0.5492]</td>
<td>35.6</td>
<td>0.94</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.3118</td>
<td>0.1232</td>
<td>[0.0920, 0.5637]</td>
<td>42.0</td>
<td>0.84</td>
</tr>
<tr>
<td>$p_{13}$</td>
<td>0.3218</td>
<td>0.0986</td>
<td>[0.1304, 0.5139]</td>
<td>28.4</td>
<td>0.72</td>
</tr>
<tr>
<td>$p_{14}$</td>
<td>0.0153</td>
<td>0.0148</td>
<td>[0.0004, 0.0545]</td>
<td>5.5</td>
<td>0.85</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>0.0318</td>
<td>0.0210</td>
<td>[0.0003, 0.0846]</td>
<td>41.6</td>
<td>0.88</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.5479</td>
<td>0.0713</td>
<td>[0.4104, 0.6922]</td>
<td>82.6</td>
<td>0.70</td>
</tr>
<tr>
<td>$p_{23}$</td>
<td>0.4135</td>
<td>0.0676</td>
<td>[0.2757, 0.5415]</td>
<td>77.7</td>
<td>0.52</td>
</tr>
<tr>
<td>$p_{24}$</td>
<td>0.0068</td>
<td>0.0059</td>
<td>[0.0002, 0.0221]</td>
<td>15.0</td>
<td>0.80</td>
</tr>
<tr>
<td>$p_{31}$</td>
<td>0.0198</td>
<td>0.0054</td>
<td>[0.0106, 0.0318]</td>
<td>27.9</td>
<td>0.98</td>
</tr>
<tr>
<td>$p_{32}$</td>
<td>0.0845</td>
<td>0.0202</td>
<td>[0.0475, 0.1260]</td>
<td>87.7</td>
<td>0.95</td>
</tr>
<tr>
<td>$p_{33}$</td>
<td>0.8747</td>
<td>0.0211</td>
<td>[0.8287, 0.9117]</td>
<td>82.4</td>
<td>0.86</td>
</tr>
<tr>
<td>$p_{34}$</td>
<td>0.0210</td>
<td>0.0050</td>
<td>[0.0127, 0.0321]</td>
<td>28.3</td>
<td>0.61</td>
</tr>
<tr>
<td>$p_{41}$</td>
<td>0.0448</td>
<td>0.0368</td>
<td>[0.0018, 0.1391]</td>
<td>12.3</td>
<td>0.91</td>
</tr>
<tr>
<td>$p_{42}$</td>
<td>0.1424</td>
<td>0.0761</td>
<td>[0.0165, 0.3067]</td>
<td>20.9</td>
<td>0.64</td>
</tr>
<tr>
<td>$p_{43}$</td>
<td>0.5833</td>
<td>0.0924</td>
<td>[0.3963, 0.7555]</td>
<td>18.2</td>
<td>0.62</td>
</tr>
<tr>
<td>$p_{44}$</td>
<td>0.2294</td>
<td>0.0642</td>
<td>[0.1130, 0.3636]</td>
<td>12.2</td>
<td>0.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.3476</td>
<td>0.1183</td>
<td>[0.1234, 0.5874]</td>
<td>51.5</td>
<td>0.98</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.3225</td>
<td>0.1317</td>
<td>[0.0895, 0.6000]</td>
<td>51.0</td>
<td>0.82</td>
</tr>
<tr>
<td>$p_{13}$</td>
<td>0.3145</td>
<td>0.1044</td>
<td>[0.1182, 0.5207]</td>
<td>34.3</td>
<td>0.82</td>
</tr>
<tr>
<td>$p_{14}$</td>
<td>0.0153</td>
<td>0.0153</td>
<td>[0.0004, 0.0566]</td>
<td>6.0</td>
<td>0.86</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>0.0338</td>
<td>0.0238</td>
<td>[0.0032, 0.0904]</td>
<td>55.1</td>
<td>0.63</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.5641</td>
<td>0.0743</td>
<td>[0.4194, 0.7031]</td>
<td>80.6</td>
<td>0.70</td>
</tr>
<tr>
<td>$p_{23}$</td>
<td>0.3956</td>
<td>0.0720</td>
<td>[0.2596, 0.5378]</td>
<td>86.7</td>
<td>0.86</td>
</tr>
<tr>
<td>$p_{24}$</td>
<td>0.0065</td>
<td>0.0059</td>
<td>[0.0002, 0.0219]</td>
<td>14.6</td>
<td>0.85</td>
</tr>
<tr>
<td>$p_{31}$</td>
<td>0.0195</td>
<td>0.0056</td>
<td>[0.0097, 0.0318]</td>
<td>34.8</td>
<td>0.65</td>
</tr>
<tr>
<td>$p_{32}$</td>
<td>0.0762</td>
<td>0.0215</td>
<td>[0.0410, 0.1273]</td>
<td>116.2</td>
<td>0.73</td>
</tr>
<tr>
<td>$p_{33}$</td>
<td>0.8847</td>
<td>0.0226</td>
<td>[0.8264, 0.9198]</td>
<td>104.9</td>
<td>0.85</td>
</tr>
<tr>
<td>$p_{34}$</td>
<td>0.0196</td>
<td>0.0047</td>
<td>[0.0116, 0.0298]</td>
<td>27.6</td>
<td>0.91</td>
</tr>
<tr>
<td>$p_{41}$</td>
<td>0.0450</td>
<td>0.0372</td>
<td>[0.0017, 0.1392]</td>
<td>10.0</td>
<td>0.96</td>
</tr>
<tr>
<td>$p_{42}$</td>
<td>0.1317</td>
<td>0.0770</td>
<td>[0.0106, 0.3025]</td>
<td>21.6</td>
<td>0.98</td>
</tr>
<tr>
<td>$p_{43}$</td>
<td>0.5935</td>
<td>0.0933</td>
<td>[0.4042, 0.7648]</td>
<td>18.6</td>
<td>0.96</td>
</tr>
<tr>
<td>$p_{44}$</td>
<td>0.2298</td>
<td>0.0646</td>
<td>[0.1136, 0.3648]</td>
<td>11.9</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).
Table A.26: Estimation Results of the MS4SVt-RV Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>0.8792</td>
<td>0.1833</td>
<td>[0.5301, 1.2430]</td>
<td>29.4</td>
<td>0.63</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.2381</td>
<td>0.1414</td>
<td>[-0.0347, 0.5214]</td>
<td>12.8</td>
<td>0.77</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.4185</td>
<td>0.1273</td>
<td>[-0.6619, -0.1612]</td>
<td>8.2</td>
<td>0.80</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>-1.5878</td>
<td>0.1556</td>
<td>[-1.8926, -1.2798]</td>
<td>20.5</td>
<td>0.73</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.9346</td>
<td>0.0373</td>
<td>[0.8502, 0.9911]</td>
<td>16.1</td>
<td>0.92</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.9836</td>
<td>0.0104</td>
<td>[0.9590, 0.9982]</td>
<td>14.5</td>
<td>1.00</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.9863</td>
<td>0.0032</td>
<td>[0.9800, 0.9927]</td>
<td>13.1</td>
<td>0.94</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.9072</td>
<td>0.0433</td>
<td>[0.8134, 0.9812]</td>
<td>44.0</td>
<td>0.81</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.2730</td>
<td>0.0622</td>
<td>[0.1534, 0.4002]</td>
<td>66.6</td>
<td>0.74</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.2267</td>
<td>0.0312</td>
<td>[0.1644, 0.2903]</td>
<td>121.1</td>
<td>0.83</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.1286</td>
<td>0.0113</td>
<td>[0.1064, 0.1516]</td>
<td>195.4</td>
<td>0.52</td>
</tr>
<tr>
<td>$\sigma_4$</td>
<td>0.1450</td>
<td>0.0428</td>
<td>[0.0788, 0.2425]</td>
<td>67.1</td>
<td>0.67</td>
</tr>
<tr>
<td>$\chi$</td>
<td>-0.5167</td>
<td>0.0321</td>
<td>[-0.5789, -0.4554]</td>
<td>37.9</td>
<td>0.98</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>33.329</td>
<td>5.0549</td>
<td>[24.455, 44.635]</td>
<td>159.3</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).

Second panel: Implied higher moments of $\epsilon_t$. Lower two panels: Parameters auxiliary measurement equation and implied signal-to-noise ratios.
### Table A.27: Estimation Results of the MS4SV-RV Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>0.8959</td>
<td>0.1654</td>
<td>[0.5829, 1.2283]</td>
<td>23.9</td>
<td>0.97</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.2823</td>
<td>0.1393</td>
<td>[0.0148, 0.5652]</td>
<td>13.1</td>
<td>0.96</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.3698</td>
<td>0.1259</td>
<td>[-0.6117, -0.1110]</td>
<td>8.7</td>
<td>0.93</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>-1.5391</td>
<td>0.1526</td>
<td>[-1.8402, -1.2393]</td>
<td>20.6</td>
<td>0.94</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.9336</td>
<td>0.0378</td>
<td>[0.8478, 0.9910]</td>
<td>18.1</td>
<td>0.92</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.9841</td>
<td>0.0102</td>
<td>[0.9596, 0.9981]</td>
<td>12.7</td>
<td>0.97</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.9864</td>
<td>0.0032</td>
<td>[0.9800, 0.9926]</td>
<td>12.8</td>
<td>0.93</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.9034</td>
<td>0.0426</td>
<td>[0.8171, 0.9804]</td>
<td>45.3</td>
<td>0.89</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.2684</td>
<td>0.0611</td>
<td>[0.1526, 0.3947]</td>
<td>66.7</td>
<td>0.52</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.2279</td>
<td>0.0304</td>
<td>[0.1657, 0.2907]</td>
<td>117.9</td>
<td>0.69</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.1265</td>
<td>0.0103</td>
<td>[0.1074, 0.1473]</td>
<td>180.2</td>
<td>0.98</td>
</tr>
<tr>
<td>$\sigma_4$</td>
<td>0.1458</td>
<td>0.0423</td>
<td>[0.0805, 0.2435]</td>
<td>59.0</td>
<td>0.94</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.5146</td>
<td>0.0324</td>
<td>[-0.5769, -0.4504]</td>
<td>39.5</td>
<td>0.96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\zeta$</td>
<td>-0.2568</td>
<td>0.0197</td>
<td>[-0.2954, -0.2181]</td>
<td>41.9</td>
<td>0.88</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.8952</td>
<td>0.0192</td>
<td>[0.8595, 0.9347]</td>
<td>52.1</td>
<td>0.70</td>
</tr>
<tr>
<td>$\sigma_{u,1}$</td>
<td>0.6539</td>
<td>0.0643</td>
<td>[0.5410, 0.7935]</td>
<td>11.8</td>
<td>0.97</td>
</tr>
<tr>
<td>$\sigma_{u,2}$</td>
<td>0.3183</td>
<td>0.0301</td>
<td>[0.2573, 0.3752]</td>
<td>56.8</td>
<td>0.86</td>
</tr>
<tr>
<td>$\sigma_{u,3}$</td>
<td>0.3040</td>
<td>0.0083</td>
<td>[0.2879, 0.3199]</td>
<td>29.1</td>
<td>0.99</td>
</tr>
<tr>
<td>$\sigma_{u,4}$</td>
<td>0.3494</td>
<td>0.0458</td>
<td>[0.2657, 0.4445]</td>
<td>14.1</td>
<td>0.99</td>
</tr>
<tr>
<td>$\text{SNR}_1$</td>
<td>0.4151</td>
<td>0.1057</td>
<td>[0.2234, 0.6376]</td>
<td>48.1</td>
<td>0.64</td>
</tr>
<tr>
<td>$\text{SNR}_2$</td>
<td>0.7255</td>
<td>0.1368</td>
<td>[0.4839, 1.0255]</td>
<td>104.4</td>
<td>0.73</td>
</tr>
<tr>
<td>$\text{SNR}_3$</td>
<td>0.4165</td>
<td>0.0367</td>
<td>[0.3494, 0.4904]</td>
<td>131.9</td>
<td>0.98</td>
</tr>
<tr>
<td>$\text{SNR}_4$</td>
<td>0.4246</td>
<td>0.1378</td>
<td>[0.2209, 0.7539]</td>
<td>43.7</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value). Middle panel: Parameters auxiliary measurement equation. Bottom panel: Implied signal-to-noise ratios.
### A.E. PARAMETER ESTIMATES

#### Table A.28: Results of the MS4SV-RV Model: Transition Probabilities

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.3307</td>
<td>0.1106</td>
<td>[0.1147, 0.5403]</td>
<td>41.7</td>
<td>0.96</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.3350</td>
<td>0.1298</td>
<td>[0.1019, 0.6014]</td>
<td>44.7</td>
<td>0.96</td>
</tr>
<tr>
<td>$p_{13}$</td>
<td>0.3197</td>
<td>0.0982</td>
<td>[0.1311, 0.5127]</td>
<td>29.0</td>
<td>0.99</td>
</tr>
<tr>
<td>$p_{14}$</td>
<td>0.0147</td>
<td>0.0144</td>
<td>[0.0004, 0.0528]</td>
<td>6.2</td>
<td>0.98</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>0.0378</td>
<td>0.0277</td>
<td>[0.0037, 0.1118]</td>
<td>57.8</td>
<td>0.90</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.5640</td>
<td>0.0762</td>
<td>[0.4078, 0.7071]</td>
<td>86.8</td>
<td>0.82</td>
</tr>
<tr>
<td>$p_{23}$</td>
<td>0.3920</td>
<td>0.0701</td>
<td>[0.2556, 0.5259]</td>
<td>80.0</td>
<td>0.83</td>
</tr>
<tr>
<td>$p_{24}$</td>
<td>0.0063</td>
<td>0.0054</td>
<td>[0.0002, 0.0202]</td>
<td>12.3</td>
<td>0.99</td>
</tr>
<tr>
<td>$p_{31}$</td>
<td>0.0207</td>
<td>0.0057</td>
<td>[0.0108, 0.0333]</td>
<td>31.6</td>
<td>0.97</td>
</tr>
<tr>
<td>$p_{32}$</td>
<td>0.0744</td>
<td>0.0176</td>
<td>[0.0422, 0.1115]</td>
<td>75.4</td>
<td>0.86</td>
</tr>
<tr>
<td>$p_{33}$</td>
<td>0.8854</td>
<td>0.0181</td>
<td>[0.8463, 0.9171]</td>
<td>66.9</td>
<td>0.88</td>
</tr>
<tr>
<td>$p_{34}$</td>
<td>0.0194</td>
<td>0.0043</td>
<td>[0.0120, 0.0288]</td>
<td>22.2</td>
<td>0.95</td>
</tr>
<tr>
<td>$p_{41}$</td>
<td>0.0483</td>
<td>0.0393</td>
<td>[0.0017, 0.1473]</td>
<td>12.1</td>
<td>0.87</td>
</tr>
<tr>
<td>$p_{42}$</td>
<td>0.1316</td>
<td>0.0765</td>
<td>[0.0129, 0.3037]</td>
<td>21.9</td>
<td>0.87</td>
</tr>
<tr>
<td>$p_{43}$</td>
<td>0.5892</td>
<td>0.0921</td>
<td>[0.4022, 0.7591]</td>
<td>17.8</td>
<td>0.87</td>
</tr>
<tr>
<td>$p_{44}$</td>
<td>0.2309</td>
<td>0.0653</td>
<td>[0.1122, 0.3684]</td>
<td>12.7</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).

#### SVsk-t-Range Model Variants

#### Table A.29: Estimation Results of the SVt-Range Model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.3344</td>
<td>0.0871</td>
<td>[-0.5031, -0.1618]</td>
<td>1.4</td>
<td>0.96</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9601</td>
<td>0.0060</td>
<td>[0.9475, 0.9705]</td>
<td>25.6</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.2838</td>
<td>0.0222</td>
<td>[0.2451, 0.3291]</td>
<td>164.4</td>
<td>0.97</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.5378</td>
<td>0.0347</td>
<td>[-0.6054, -0.4702]</td>
<td>57.1</td>
<td>0.88</td>
</tr>
<tr>
<td>$\nu$</td>
<td>30.973</td>
<td>5.0957</td>
<td>[21.310, 41.563]</td>
<td>218.0</td>
<td>0.81</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.2309</td>
<td>0.0466</td>
<td>[0.1597, 0.3464]</td>
<td>247.2</td>
<td>0.84</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>0.2865</td>
<td>0.0120</td>
<td>[0.2631, 0.3102]</td>
<td>15.6</td>
<td>0.97</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.4522</td>
<td>0.0112</td>
<td>[0.4310, 0.4745]</td>
<td>20.8</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.3673</td>
<td>0.0051</td>
<td>[0.3574, 0.3771]</td>
<td>8.9</td>
<td>0.92</td>
</tr>
<tr>
<td>SNR</td>
<td>0.7732</td>
<td>0.0675</td>
<td>[0.6573, 0.9127]</td>
<td>148.5</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value). Second panel: Implied higher moments of $\epsilon_t$. Lower two panels: Parameters auxiliary measurement equation and implied signal-to-noise ratio.
# MS4SV\textit{skt}-Range Model Variants

Table A.30: Estimation Results of the MS4SV\textit{skt}-Range (upper block) / MS4SV\textit{rt}-Range Model (lower block): Transition Probabilities

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.0952</td>
<td>0.0218</td>
<td>[0.0552, 0.1413]</td>
<td>16.7</td>
<td>0.91</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.2412</td>
<td>0.0374</td>
<td>[0.1672, 0.3138]</td>
<td>26.4</td>
<td>1.00</td>
</tr>
<tr>
<td>$p_{13}$</td>
<td>0.3985</td>
<td>0.0490</td>
<td>[0.3016, 0.4919]</td>
<td>35.4</td>
<td>0.83</td>
</tr>
<tr>
<td>$p_{14}$</td>
<td>0.2651</td>
<td>0.0456</td>
<td>[0.1833, 0.3598]</td>
<td>43.4</td>
<td>0.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{21}$</td>
<td>0.0835</td>
<td>0.0195</td>
<td>[0.0495, 0.1252]</td>
<td>54.7</td>
<td>0.90</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.2908</td>
<td>0.0246</td>
<td>[0.2407, 0.3372]</td>
<td>28.8</td>
<td>0.97</td>
</tr>
<tr>
<td>$p_{23}$</td>
<td>0.3937</td>
<td>0.0320</td>
<td>[0.3296, 0.4538]</td>
<td>44.1</td>
<td>0.76</td>
</tr>
<tr>
<td>$p_{24}$</td>
<td>0.2320</td>
<td>0.0371</td>
<td>[0.1665, 0.3132]</td>
<td>87.5</td>
<td>0.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{31}$</td>
<td>0.1485</td>
<td>0.0228</td>
<td>[0.1068, 0.1944]</td>
<td>47.6</td>
<td>0.82</td>
</tr>
<tr>
<td>$p_{32}$</td>
<td>0.3621</td>
<td>0.0294</td>
<td>[0.3016, 0.4180]</td>
<td>41.1</td>
<td>0.94</td>
</tr>
<tr>
<td>$p_{33}$</td>
<td>0.3003</td>
<td>0.0294</td>
<td>[0.2396, 0.3571]</td>
<td>42.9</td>
<td>0.79</td>
</tr>
<tr>
<td>$p_{34}$</td>
<td>0.1892</td>
<td>0.0309</td>
<td>[0.1359, 0.2568]</td>
<td>76.5</td>
<td>0.87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{41}$</td>
<td>0.1552</td>
<td>0.0264</td>
<td>[0.1073, 0.2103]</td>
<td>32.8</td>
<td>0.84</td>
</tr>
<tr>
<td>$p_{42}$</td>
<td>0.3671</td>
<td>0.0335</td>
<td>[0.3011, 0.4336]</td>
<td>27.3</td>
<td>0.87</td>
</tr>
<tr>
<td>$p_{43}$</td>
<td>0.3527</td>
<td>0.0321</td>
<td>[0.2906, 0.4154]</td>
<td>24.6</td>
<td>0.92</td>
</tr>
<tr>
<td>$p_{44}$</td>
<td>0.1250</td>
<td>0.0306</td>
<td>[0.0693, 0.1878]</td>
<td>55.2</td>
<td>0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.0912</td>
<td>0.0199</td>
<td>[0.0544, 0.1321]</td>
<td>13.1</td>
<td>0.91</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.2759</td>
<td>0.0369</td>
<td>[0.2033, 0.3477]</td>
<td>22.7</td>
<td>0.91</td>
</tr>
<tr>
<td>$p_{13}$</td>
<td>0.3999</td>
<td>0.0482</td>
<td>[0.3070, 0.4954]</td>
<td>33.7</td>
<td>0.90</td>
</tr>
<tr>
<td>$p_{14}$</td>
<td>0.2331</td>
<td>0.0419</td>
<td>[0.1569, 0.3192]</td>
<td>38.8</td>
<td>0.79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{21}$</td>
<td>0.0866</td>
<td>0.0168</td>
<td>[0.0553, 0.1210]</td>
<td>41.7</td>
<td>0.77</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.3079</td>
<td>0.0234</td>
<td>[0.2619, 0.3530]</td>
<td>27.3</td>
<td>0.97</td>
</tr>
<tr>
<td>$p_{23}$</td>
<td>0.3983</td>
<td>0.0301</td>
<td>[0.3373, 0.4563]</td>
<td>42.2</td>
<td>0.87</td>
</tr>
<tr>
<td>$p_{24}$</td>
<td>0.2072</td>
<td>0.0316</td>
<td>[0.1510, 0.2711]</td>
<td>74.0</td>
<td>0.79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{31}$</td>
<td>0.1454</td>
<td>0.0219</td>
<td>[0.1010, 0.1883]</td>
<td>45.5</td>
<td>0.72</td>
</tr>
<tr>
<td>$p_{32}$</td>
<td>0.3890</td>
<td>0.0278</td>
<td>[0.3335, 0.4412]</td>
<td>35.3</td>
<td>0.85</td>
</tr>
<tr>
<td>$p_{33}$</td>
<td>0.3015</td>
<td>0.0284</td>
<td>[0.2401, 0.3532]</td>
<td>40.4</td>
<td>0.75</td>
</tr>
<tr>
<td>$p_{34}$</td>
<td>0.1642</td>
<td>0.0280</td>
<td>[0.1156, 0.2218]</td>
<td>72.9</td>
<td>0.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{41}$</td>
<td>0.1495</td>
<td>0.0254</td>
<td>[0.1003, 0.2005]</td>
<td>27.1</td>
<td>0.80</td>
</tr>
<tr>
<td>$p_{42}$</td>
<td>0.3924</td>
<td>0.0322</td>
<td>[0.3297, 0.4573]</td>
<td>19.7</td>
<td>0.91</td>
</tr>
<tr>
<td>$p_{43}$</td>
<td>0.3487</td>
<td>0.0332</td>
<td>[0.2844, 0.4149]</td>
<td>23.4</td>
<td>0.89</td>
</tr>
<tr>
<td>$p_{44}$</td>
<td>0.1094</td>
<td>0.0266</td>
<td>[0.0604, 0.1636]</td>
<td>38.4</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>0.9816</td>
<td>0.1258</td>
<td>[0.7464, 1.2367]</td>
<td>7.8</td>
<td>0.85</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.0287</td>
<td>0.1156</td>
<td>[-0.1921, 0.2640]</td>
<td>5.1</td>
<td>0.86</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.9221</td>
<td>0.1177</td>
<td>[-1.1469, -0.6887]</td>
<td>5.3</td>
<td>0.85</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>-1.8424</td>
<td>0.1225</td>
<td>[-2.0758, -1.6004]</td>
<td>7.2</td>
<td>0.78</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.9252</td>
<td>0.0260</td>
<td>[0.8731, 0.9742]</td>
<td>21.6</td>
<td>0.92</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.9658</td>
<td>0.0117</td>
<td>[0.9417, 0.9875]</td>
<td>31.8</td>
<td>0.71</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.9758</td>
<td>0.0114</td>
<td>[0.9516, 0.9958]</td>
<td>23.4</td>
<td>0.75</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.9677</td>
<td>0.0179</td>
<td>[0.9284, 0.9958]</td>
<td>38.2</td>
<td>0.86</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.4034</td>
<td>0.0403</td>
<td>[0.3247, 0.4849]</td>
<td>49.0</td>
<td>0.93</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.2554</td>
<td>0.0235</td>
<td>[0.2093, 0.3020]</td>
<td>91.0</td>
<td>0.65</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.3054</td>
<td>0.0204</td>
<td>[0.2647, 0.3444]</td>
<td>62.4</td>
<td>0.71</td>
</tr>
<tr>
<td>$\sigma_4$</td>
<td>0.2359</td>
<td>0.0290</td>
<td>[0.1795, 0.2932]</td>
<td>97.0</td>
<td>0.80</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>-0.3468</td>
<td>0.0764</td>
<td>[-0.4959, -0.1940]</td>
<td>18.5</td>
<td>0.92</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>-0.4463</td>
<td>0.0559</td>
<td>[-0.5527, -0.3350]</td>
<td>33.2</td>
<td>0.78</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>-0.6049</td>
<td>0.0415</td>
<td>[-0.6820, -0.5200]</td>
<td>33.9</td>
<td>0.96</td>
</tr>
<tr>
<td>$\rho_4$</td>
<td>-0.6339</td>
<td>0.0634</td>
<td>[-0.7476, -0.5008]</td>
<td>33.3</td>
<td>0.93</td>
</tr>
<tr>
<td>$\nu$</td>
<td>53.860</td>
<td>6.1432</td>
<td>[42.640, 66.448]</td>
<td>72.0</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table A.31: Estimation Results of the MS4SV-$r$-Range Model

- kurtosis: 0.1222, 0.0152, [0.0961, 0.1553], 73.9, 0.97
- kurtosis: 0.4220, 0.0102, [0.4023, 0.4423], 62.3, 0.93

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value). Second panel: Implied higher moments of $\epsilon_t$. Lower two panels: Parameters auxiliary measurement equation and implied signal-to-noise ratios.
Table A.32: Estimation Results of the MS4SV-Range Model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>0.9967</td>
<td>0.1199</td>
<td>[0.7662, 1.2368]</td>
<td>6.2</td>
<td>0.84</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.0458</td>
<td>0.1108</td>
<td>[-0.1664, 0.2677]</td>
<td>3.9</td>
<td>0.89</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.8959</td>
<td>0.1170</td>
<td>[-1.1202, -0.6601]</td>
<td>5.4</td>
<td>0.97</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>-1.8118</td>
<td>0.1249</td>
<td>[-2.0585, -1.5671]</td>
<td>8.4</td>
<td>0.87</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.9239</td>
<td>0.0260</td>
<td>[0.8718, 0.9731]</td>
<td>20.7</td>
<td>0.85</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.9646</td>
<td>0.0122</td>
<td>[0.9402, 0.9873]</td>
<td>34.7</td>
<td>0.92</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.9763</td>
<td>0.0111</td>
<td>[0.9527, 0.9956]</td>
<td>22.2</td>
<td>0.93</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.9680</td>
<td>0.0174</td>
<td>[0.9300, 0.9958]</td>
<td>37.0</td>
<td>0.85</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.4028</td>
<td>0.0389</td>
<td>[0.3287, 0.4787]</td>
<td>46.1</td>
<td>0.97</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.2548</td>
<td>0.0235</td>
<td>[0.2084, 0.3005]</td>
<td>93.0</td>
<td>0.76</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.3074</td>
<td>0.0213</td>
<td>[0.2663, 0.3502]</td>
<td>61.4</td>
<td>0.99</td>
</tr>
<tr>
<td>$\sigma_4$</td>
<td>0.2402</td>
<td>0.0281</td>
<td>[0.1850, 0.2944]</td>
<td>80.7</td>
<td>0.90</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>-0.3444</td>
<td>0.0730</td>
<td>[-0.4879, -0.1988]</td>
<td>16.0</td>
<td>0.89</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>-0.4469</td>
<td>0.0542</td>
<td>[-0.5502, -0.3391]</td>
<td>29.4</td>
<td>0.86</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>-0.5957</td>
<td>0.0427</td>
<td>[-0.6735, -0.5074]</td>
<td>33.5</td>
<td>0.88</td>
</tr>
<tr>
<td>$\rho_4$</td>
<td>-0.6230</td>
<td>0.0635</td>
<td>[-0.7339, -0.4832]</td>
<td>34.7</td>
<td>0.95</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.4165</td>
<td>0.0093</td>
<td>[0.3980, 0.4353]</td>
<td>49.0</td>
<td>0.90</td>
</tr>
<tr>
<td>$\tilde{\xi}$</td>
<td>0.4113</td>
<td>0.0062</td>
<td>[0.3990, 0.4231]</td>
<td>63.0</td>
<td>0.97</td>
</tr>
<tr>
<td>$\sigma_{u,1}$</td>
<td>0.1607</td>
<td>0.0120</td>
<td>[0.1373, 0.1847]</td>
<td>24.4</td>
<td>0.94</td>
</tr>
<tr>
<td>$\sigma_{u,2}$</td>
<td>0.1054</td>
<td>0.0073</td>
<td>[0.0917, 0.1201]</td>
<td>55.6</td>
<td>0.88</td>
</tr>
<tr>
<td>$\sigma_{u,3}$</td>
<td>0.1156</td>
<td>0.0088</td>
<td>[0.0989, 0.1334]</td>
<td>64.3</td>
<td>0.89</td>
</tr>
<tr>
<td>$\sigma_{u,4}$</td>
<td>0.1779</td>
<td>0.0098</td>
<td>[0.1584, 0.1969]</td>
<td>23.1</td>
<td>0.90</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR$_1$</td>
<td>2.5268</td>
<td>0.3542</td>
<td>[1.8875, 3.2643]</td>
<td>40.1</td>
<td>0.98</td>
</tr>
<tr>
<td>SNR$_2$</td>
<td>2.4336</td>
<td>0.3270</td>
<td>[1.8367, 3.1218]</td>
<td>91.1</td>
<td>0.90</td>
</tr>
<tr>
<td>SNR$_3$</td>
<td>2.6792</td>
<td>0.3115</td>
<td>[2.1262, 3.3505]</td>
<td>62.4</td>
<td>0.92</td>
</tr>
<tr>
<td>SNR$_4$</td>
<td>1.3557</td>
<td>0.1827</td>
<td>[1.0078, 1.7351]</td>
<td>63.0</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value). Lower two panels: Parameters auxiliary measurement equation and implied signal-to-noise ratios.
Table A.33: Results of the MS4SV-Range Model: Transition Probabilities

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>p11</td>
<td>0.0902</td>
<td>0.0197</td>
<td>[0.0538, 0.1304]</td>
<td>12.8</td>
<td>0.84</td>
</tr>
<tr>
<td>p12</td>
<td>0.2747</td>
<td>0.0359</td>
<td>[0.2028, 0.3453]</td>
<td>20.8</td>
<td>0.97</td>
</tr>
<tr>
<td>p13</td>
<td>0.3949</td>
<td>0.0492</td>
<td>[0.2980, 0.4904]</td>
<td>34.1</td>
<td>0.77</td>
</tr>
<tr>
<td>p14</td>
<td>0.2402</td>
<td>0.0467</td>
<td>[0.1539, 0.3343]</td>
<td>48.7</td>
<td>0.81</td>
</tr>
<tr>
<td>p21</td>
<td>0.0841</td>
<td>0.0162</td>
<td>[0.0545, 0.1188]</td>
<td>38.5</td>
<td>0.81</td>
</tr>
<tr>
<td>p22</td>
<td>0.3049</td>
<td>0.0234</td>
<td>[0.2587, 0.3497]</td>
<td>26.9</td>
<td>0.82</td>
</tr>
<tr>
<td>p23</td>
<td>0.3962</td>
<td>0.0314</td>
<td>[0.3354, 0.4581]</td>
<td>45.9</td>
<td>0.95</td>
</tr>
<tr>
<td>p24</td>
<td>0.2148</td>
<td>0.0373</td>
<td>[0.1452, 0.2910]</td>
<td>99.9</td>
<td>0.94</td>
</tr>
<tr>
<td>p31</td>
<td>0.1437</td>
<td>0.0202</td>
<td>[0.1058, 0.1850]</td>
<td>35.6</td>
<td>0.82</td>
</tr>
<tr>
<td>p32</td>
<td>0.3863</td>
<td>0.0283</td>
<td>[0.3298, 0.4408]</td>
<td>38.5</td>
<td>0.85</td>
</tr>
<tr>
<td>p33</td>
<td>0.3004</td>
<td>0.0261</td>
<td>[0.2465, 0.3494]</td>
<td>31.5</td>
<td>0.82</td>
</tr>
<tr>
<td>p34</td>
<td>0.1696</td>
<td>0.0308</td>
<td>[0.1135, 0.2319]</td>
<td>85.1</td>
<td>0.86</td>
</tr>
<tr>
<td>p41</td>
<td>0.1475</td>
<td>0.0230</td>
<td>[0.1047, 0.1945]</td>
<td>21.1</td>
<td>0.80</td>
</tr>
<tr>
<td>p42</td>
<td>0.3902</td>
<td>0.0324</td>
<td>[0.3297, 0.4562]</td>
<td>21.1</td>
<td>0.85</td>
</tr>
<tr>
<td>p43</td>
<td>0.3475</td>
<td>0.0348</td>
<td>[0.2793, 0.4148]</td>
<td>28.0</td>
<td>0.86</td>
</tr>
<tr>
<td>p44</td>
<td>0.1148</td>
<td>0.0288</td>
<td>[0.0606, 0.1718]</td>
<td>46.1</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).

SVsk t-VIX Model Variants

Table A.34: Estimation Results of the SV t-VIX Model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>μ</td>
<td>-0.3421</td>
<td>0.1741</td>
<td>[-0.6859, 0.0019]</td>
<td>1.2</td>
<td>0.77</td>
</tr>
<tr>
<td>ϕ</td>
<td>0.9885</td>
<td>0.0021</td>
<td>[0.9843, 0.9926]</td>
<td>1.8</td>
<td>0.93</td>
</tr>
<tr>
<td>σ</td>
<td>0.1383</td>
<td>0.0045</td>
<td>[0.1298, 0.1472]</td>
<td>108.8</td>
<td>0.88</td>
</tr>
<tr>
<td>ρ</td>
<td>-0.0199</td>
<td>0.0210</td>
<td>[-0.0622, 0.0204]</td>
<td>15.8</td>
<td>0.92</td>
</tr>
<tr>
<td>ν</td>
<td>15.512</td>
<td>2.7446</td>
<td>[11.263, 21.724]</td>
<td>492.5</td>
<td>0.53</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.5503</td>
<td>0.1284</td>
<td>[0.3385, 0.8261]</td>
<td>453.0</td>
<td>0.52</td>
</tr>
<tr>
<td>ζ</td>
<td>0.3170</td>
<td>0.0122</td>
<td>[0.2946, 0.3386]</td>
<td>3,214.6</td>
<td>0.17</td>
</tr>
<tr>
<td>ξ</td>
<td>0.3868</td>
<td>0.0082</td>
<td>[0.3719, 0.4029]</td>
<td>1,513.4</td>
<td>0.53</td>
</tr>
<tr>
<td>σu</td>
<td>0.0187</td>
<td>0.0019</td>
<td>[0.0147, 0.0223]</td>
<td>220.4</td>
<td>0.86</td>
</tr>
<tr>
<td>SNR</td>
<td>7.5075</td>
<td>0.9942</td>
<td>[5.9333, 9.844]</td>
<td>241.6</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value). Second panel: Implied higher moments of ε t. Lower two panels: Parameters auxiliary measurement equation and implied signal-to-noise ratio.
Table A.35: Estimation Results of the SV-VIX Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.2032</td>
<td>0.1729</td>
<td>[-0.5472, 0.1370]</td>
<td>1.1</td>
<td>0.98</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9885</td>
<td>0.0021</td>
<td>[0.9843, 0.9926]</td>
<td>1.8</td>
<td>0.93</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.1392</td>
<td>0.0048</td>
<td>[0.1299, 0.1486]</td>
<td>126.0</td>
<td>0.91</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.0121</td>
<td>0.0213</td>
<td>[-0.0547, 0.0285]</td>
<td>17.6</td>
<td>0.86</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>0.2604</td>
<td>0.0068</td>
<td>[0.2461, 0.2733]</td>
<td>1,071.4</td>
<td>0.70</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.3871</td>
<td>0.0085</td>
<td>[0.3726, 0.4046]</td>
<td>1,955.3</td>
<td>0.93</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.0180</td>
<td>0.0022</td>
<td>[0.0133, 0.0219]</td>
<td>312.7</td>
<td>0.70</td>
</tr>
<tr>
<td>SNR</td>
<td>7.8803</td>
<td>1.2670</td>
<td>[6.0212, 10.969]</td>
<td>357.8</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value). Lower two panels: Parameters auxiliary measurement equation and implied signal-to-noise ratio.

MS2Vskt-VIX Model Variants

Table A.36: Estimation Results of the MS2Vskt-VIX (upper block) / MS2Vt-VIX (middle block) / MS2SV-VIX Model (lower block): Transition Probabilities

<table>
<thead>
<tr>
<th>Transition</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.2701</td>
<td>0.0356</td>
<td>[0.2027, 0.3421]</td>
<td>39.0</td>
<td>0.96</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.7299</td>
<td>0.0356</td>
<td>[0.6579, 0.7973]</td>
<td>39.0</td>
<td>0.96</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>0.2471</td>
<td>0.0235</td>
<td>[0.2031, 0.2944]</td>
<td>100.9</td>
<td>0.87</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.7529</td>
<td>0.0235</td>
<td>[0.7056, 0.7969]</td>
<td>100.9</td>
<td>0.87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transition</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.8832</td>
<td>0.0187</td>
<td>[0.8430, 0.9161]</td>
<td>40.1</td>
<td>0.88</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.1168</td>
<td>0.0187</td>
<td>[0.0839, 0.1570]</td>
<td>40.1</td>
<td>0.88</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>0.0302</td>
<td>0.0049</td>
<td>[0.0216, 0.0407]</td>
<td>36.2</td>
<td>0.90</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.9698</td>
<td>0.0049</td>
<td>[0.9593, 0.9784]</td>
<td>36.2</td>
<td>0.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transition</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.8812</td>
<td>0.0176</td>
<td>[0.8442, 0.9129]</td>
<td>32.6</td>
<td>0.94</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.1188</td>
<td>0.0176</td>
<td>[0.0871, 0.1558]</td>
<td>32.6</td>
<td>0.94</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>0.0306</td>
<td>0.0048</td>
<td>[0.0222, 0.0409]</td>
<td>34.0</td>
<td>0.72</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.9694</td>
<td>0.0048</td>
<td>[0.9591, 0.9778]</td>
<td>34.0</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).
### Table A.37: The MS2SVt-VIX (upper block) / MS2SV-VIX (lower block) Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.8697</td>
<td>0.1444</td>
<td>[-1.1763, -0.6088]</td>
<td>4.1</td>
<td>0.94</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9869</td>
<td>0.0020</td>
<td>[0.9831, 0.9908]</td>
<td>5.5</td>
<td>0.78</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.2532</td>
<td>0.0160</td>
<td>[0.2210, 0.2838]</td>
<td>103.1</td>
<td>0.91</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.1006</td>
<td>0.0036</td>
<td>[0.0939, 0.1080]</td>
<td>140.4</td>
<td>0.76</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>-0.0055</td>
<td>0.0450</td>
<td>[-0.1040, 0.0725]</td>
<td>28.8</td>
<td>0.93</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.0238</td>
<td>0.0214</td>
<td>[-0.0191, 0.0650]</td>
<td>9.6</td>
<td>0.93</td>
</tr>
<tr>
<td>$\nu$</td>
<td>15.8231</td>
<td>2.6897</td>
<td>[11.4959, 21.8092]</td>
<td>541.7</td>
<td>0.73</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.5336</td>
<td>0.1203</td>
<td>[0.3369, 0.8004]</td>
<td>511.8</td>
<td>0.69</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.3191</td>
<td>0.0138</td>
<td>[0.2948, 0.3486]</td>
<td>1,562.7</td>
<td>0.56</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>0.3239</td>
<td>0.0128</td>
<td>[0.2976, 0.3492]</td>
<td>11,124.0</td>
<td>0.55</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.3465</td>
<td>0.0108</td>
<td>[0.3272, 0.3713]</td>
<td>1,211.3</td>
<td>0.71</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.4075</td>
<td>0.0107</td>
<td>[0.3848, 0.4266]</td>
<td>6,868.8</td>
<td>0.67</td>
</tr>
<tr>
<td>$\sigma_{u,1}$</td>
<td>0.0250</td>
<td>0.0077</td>
<td>[0.0113, 0.0399]</td>
<td>587.3</td>
<td>0.90</td>
</tr>
<tr>
<td>$\sigma_{u,2}$</td>
<td>0.0101</td>
<td>0.0014</td>
<td>[0.0074, 0.0130]</td>
<td>306.9</td>
<td>0.83</td>
</tr>
<tr>
<td>SNR$_1$</td>
<td>11.4948</td>
<td>4.8244</td>
<td>[5.7237, 23.8847]</td>
<td>562.7</td>
<td>0.95</td>
</tr>
<tr>
<td>SNR$_2$</td>
<td>10.2058</td>
<td>1.6612</td>
<td>[7.4699, 13.9673]</td>
<td>305.2</td>
<td>0.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stddev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.7461</td>
<td>0.1417</td>
<td>[-1.0491, -0.4918]</td>
<td>3.7</td>
<td>0.94</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.9867</td>
<td>0.0019</td>
<td>[0.9830, 0.9905]</td>
<td>4.8</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.2584</td>
<td>0.0137</td>
<td>[0.2306, 0.2846]</td>
<td>73.0</td>
<td>0.72</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.0996</td>
<td>0.0034</td>
<td>[0.0930, 0.1062]</td>
<td>125.1</td>
<td>0.88</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.0150</td>
<td>0.0375</td>
<td>[-0.0649, 0.0827]</td>
<td>16.2</td>
<td>0.90</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.0224</td>
<td>0.0216</td>
<td>[-0.0207, 0.0637]</td>
<td>10.1</td>
<td>0.85</td>
</tr>
<tr>
<td>Mean</td>
<td>0.2617</td>
<td>0.0105</td>
<td>[0.2412, 0.2835]</td>
<td>1,033.6</td>
<td>0.89</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>0.2729</td>
<td>0.0089</td>
<td>[0.2568, 0.2915]</td>
<td>6,228.4</td>
<td>0.39</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>0.3493</td>
<td>0.0086</td>
<td>[0.3327, 0.3669]</td>
<td>1,028.2</td>
<td>0.65</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.4122</td>
<td>0.0092</td>
<td>[0.3950, 0.4301]</td>
<td>3,703.1</td>
<td>0.89</td>
</tr>
<tr>
<td>$\sigma_{u,1}$</td>
<td>0.0214</td>
<td>0.0067</td>
<td>[0.0105, 0.0351]</td>
<td>638.2</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma_{u,2}$</td>
<td>0.0102</td>
<td>0.0015</td>
<td>[0.0076, 0.0132]</td>
<td>356.6</td>
<td>0.90</td>
</tr>
<tr>
<td>SNR$_1$</td>
<td>13.5754</td>
<td>5.1022</td>
<td>[6.7958, 25.7449]</td>
<td>546.1</td>
<td>0.97</td>
</tr>
<tr>
<td>SNR$_2$</td>
<td>10.0094</td>
<td>1.6225</td>
<td>[7.2926, 13.5356]</td>
<td>331.7</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value).

Second panel (MS2SVt only): Implied higher moments of $\epsilon_t$. Lower two panels: Parameters auxiliary measurement equation and implied signal-to-noise ratios.

Altered priors: MS2SVt-VIX: $\zeta \sim N((0.30, 0.30)', I_2)$, $\xi \sim N((0.35, 0.40)', I_2)$.

MS2SV-VIX: $\zeta \sim N((0.25, 0.25)', I_2)$, $\xi \sim N((0.35, 0.40)', I_2)$. 

---

A.E. PARAMETER ESTIMATES

89
Chapter B

Modeling Intraday Stochastic Volatility and Conditional Duration Contemporaneously with Regime Shifts
CHAPTER B. SCDSVrs

B.1 Introduction

Price activity can be modeled as the result of an inherently unobservable information flow (see e.g. Andersen, 1996, in the context of stochastic volatility). State space models are a framework that deal with such latent, stochastic processes. Nonlinearity and non-Gaussianity prevent an analytical solution, but with the advance of computing power these systems can be approached by simulation. One main feature of intraday tick-by-tick data is irregular spacing. Models operate in event time and the related statistical tool are marked point processes (for a compact exposition to the field, see e.g. Engle and Russell, 2010).

The current work proposes an approach to model stochastic volatility on the tick level. Volatility (measured over a fixed time unit) can be high either because price duration (the time to observe a certain price change) is low or price change is large or both. Accordingly, duration and associated price change above a threshold (the absolute midprice change is taken as an observable proxy for volatility) are necessary for a complete description of volatility in event time. Consequently, the concepts of stochastic conditional price duration (SCD) and stochastic volatility (SV) are merged to propose the SCDSV model, which models SV in event time as a bivariate regime switching correlated latent process.

The concept of asymmetric information in market microstructure theory (see Glosten and Milgrom, 1985, and Easley and O'Hara, 1992; for a survey of this field consider O’Hara, 1995) implies periods of high volatility as informed traders try to capitalize on their vanishing superior knowledge (informational epochs). The proposed model is able to model associated nonlinearities through various channels. A fully specified AR(1) persistence matrix captures the effect of past latent duration on expected absolute price change and vice versa, and transition innovation terms are correlated. Moreover, both latent price duration and absolute price change can be regime dependent, following their own Markov process. These features help to model complex and fast changing patterns of market activity as observed in practice. Consider e.g. Madhavan (2000) for a review of theoretical and empirical literature on

---

1To clarify things, in the sequel, when it is spoken about volatility and do not stated otherwise, the usual definition over a fixed time unit is referred; else it is directly spoken about absolute price change.
B.1. INTRODUCTION


The stochastic conditional duration model treats duration as a latent stochastic variable and was introduced by Bauwens and Veredas (2004). Its deterministic counterpart is the autoregressive conditional duration (ACD) model (Engle and Russell, 1998), and both together mirror the SV/(G)ARCH dichotomy. The additional stochastic component in the SCD and SV model is an increased source of flexibility that may prove advantageous especially in turbulent market conditions. In fact, it is often concluded in favor of the SCD (e.g. Bauwens and Veredas, 2004, or Strickland, Forbes, and Martin, 2006) and the SV model (e.g. Kim, Shephard, and Chib, 1998). Ghysels, Gouriéroux, and Jasiak (2004) (first appearing in draft form in 1997) introduce the stochastic volatility duration (SVD) model, a two factor version of the SCD model where both mean and volatility of duration are stochastic. Feng, Jiang, and Song (2004) incorporate the leverage effect into the SCD model.

Joint models of price and duration may be divided into whether they treat price as a discrete or continuous variable. Starting with the latter, Engle (2000) presents an ultra-high-frequency (UHF) GARCH model. He decomposes the joint density of duration and associated marks into marginal and conditionals, a specification often used in subsequent work. This enables him to apply a two-step estimation procedure. First, expected trade durations based on an ACD model are estimated. Then the obtained estimates are fed into a GARCH model to yield conditional volatility per unit of time. Further work along this line is found in e.g. Manganelli (2005), who includes volume into the decomposition. However, as Grammig and Wellner (2002) argue, for a full quantification of asymmetric information effects as described by market microstructure theory, taking account of a possible effect of volatility on duration is necessary. Consequently, the feature of their interdependent trade duration-volatility (IDV) model

---


is an additional dependence of present trade duration on past conditional volatility. Sequential estimation is no longer applicable and they develop a GMM procedure to estimate their model. Engle and Lunde (2003) formulate a bivariate ACD point process to jointly analyze trade and quote updates. In their approach, they recognize the possibility that an intervening trade censors the time between a trade and the following quote update. As the UHF-GARCH, it is recursively estimated.

Discrete prices are one main feature of high frequency data. Accordingly, Russell and Engle (2005) present an autoregressive conditional multinomial (ACM) ACD model, where discrete price change evolves conditional on contemporaneous duration. However, the multinomial approach lets the model’s parameter space increase quickly with the number of different price states considered. Rydberg and Shephard (2003) model price change of assets recorded trade-by-trade as a joint distribution of activity, direction and size. They apply a decomposition into conditionals and marginal, estimating each part separately. Liesenfeld, Nolte, and Pohlmeier (2006) merge the binary activity and direction variables of the above approach into one trinomial ACM model increasing parsimony. Nolte (2008) proposes a model for the joint transaction process of price change, volume, bid-ask spread and intertrade duration using a copula approach. As in the current paper, he does not apply a decomposition into conditional and marginal densities, instead modeling a contemporaneous relationship between the variables directly.

The paper proceeds as follows: Section B.2 presents the model. Section B.3 proposes a McMC algorithm for posterior computation and Section B.4 an associated particle filter. Section B.5 contains a simulation study, where adaptive Metropolis algorithms for efficient estimation of shape parameters of the generalized gamma distribution are investigated. Empirical applications on quote data of the IBM stock follow in Section B.6, including parameter and regime inference for model variants, model selection and predictive density tests. Section B.7 concludes. The Appendix contains technical details and estimates.
B.2 The Model

Working in event time one can define durations for variables like transactions, price or volume (see e.g. Giot, 2001, Bauwens and Veredas, 2004). We are concerned with price duration, i.e. the time to observe an absolute cumulative price change of at least $|c_p|$. More formally, let $t_i$ be the time of the $i$th cumulative price update $y_i$, where $y_i \geq |c_p|$ (event). Then the corresponding duration is defined as $d_i = t_i - t_{i-1}$. Now the SCDSVrs model reads as follows,

$$d_t \sim \mathcal{G}(\exp[\psi_{d,t}], \xi, \xi), \quad (B.1)$$
$$y_t \sim \mathcal{NB}(\exp[\psi_{y,t} + \kappa_y J_y], \tau), \quad t = 1, \ldots, n, \quad (B.2)$$
$$\psi_t = \mu_{s_t} + \Phi(\psi_{t-1} - \mu_{s_{t-1}}) + \eta_t, \quad t = 2, \ldots, n, \quad (B.3)$$

with

$$\psi_t = (\psi_{d,t}, \psi_{y,t})', \quad \Phi = \begin{pmatrix} \phi_{dd} & \phi_{dy} \\ \phi_{yd} & \phi_{yy} \end{pmatrix}, \quad (B.4)$$
$$\mu_{s_t} = \begin{pmatrix} \mu_{d,s_{d,t}} \\ \mu_{y,s_{y,t}} \end{pmatrix}, \quad \mu_{k,s_{k,t}} = \sum_{i=1}^{M_k} \mu_{k,i} \times I_{s_k,t=i}, \quad M_k \in \mathbb{N}^+, \quad k = \{d, y\}, \quad (B.5)$$
$$\eta_t \sim \mathcal{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_d^2 & \rho \sigma_d \sigma_y \\ \rho \sigma_y \sigma_d & \sigma_y^2 \end{pmatrix}, \quad (B.6)$$
$$\psi_1 \sim \mathcal{N}(\mu_{s_1}, \Sigma_{ini}), \quad \vec{\Sigma}_{ini} = (I_p^2 - \Phi \otimes \Phi)^{-1} \vec{\Sigma}, \quad (B.7)$$

where $d_t$ denotes price duration and $y_t$ absolute price change. Operator $\text{diag}(\cdot)$ creates a diagonal matrix, $\text{vec}(\cdot)$ stacks a matrix column-wise, $\otimes$ denotes the Kronecker product and $I_p$ the identity matrix of dimension $p$. $I_{s_k,t=i}$ is an indicator variable that is one if $s_k,t = i$ (state $i$ prevailing at time $t$) and zero otherwise. A feature of the model is the bivariate, correlated latent process $\psi_t$, Eq. (B.3), determining dynamics of price duration and absolute price change contemporaneously. Its persistence matrix $\Phi$, Eq. (B.4), is fully specified to model the effect of past latent absolute price change on expected duration and vice versa. Moreover, $\psi_t$ is constraint to be stationary (i.e. the eigenvalues of $\Phi$ must have modulus less than one), and initial value $\psi_1$ is assumed to
Generalized Gamma (price duration):

\[ a \equiv \Gamma(\zeta + \xi^{-1})\Gamma(\zeta)^{-1} \]

\[ f_{\mathcal{G}\mathcal{G}}(d_t|\psi_{d,t},\zeta,\xi) = \frac{\xi}{\Gamma(\zeta)} \left( d_t \exp(-\psi_{d,t})a \right)^{\xi d_t^{-1}} \exp\left\{ -(d_t \exp(-\psi_{d,t})a)^\xi \right\} \]

\[ F_{\mathcal{G}\mathcal{G}}(d_t|\psi_{d,t},\zeta,\xi) = \gamma(\zeta,(d_t \exp(-\psi_{d,t})a)^\xi)\Gamma(\zeta)^{-1} \]

Continuous support: \( d_t \in (0; +\infty) \), scale: \( \exp(\psi_{d,t})a^{-1} \), shape: \( \zeta,\xi > 0 \)

\( \zeta = 1 \): Weibull, \( \xi = 1 \): Gamma, \( \zeta = \xi = 1 \): Exponential

\( E(d_t) = \exp(\psi_{d,t}) \), \( \text{Var}(d_t) = \exp(\psi_{d,t})^2 \left( \frac{\Gamma(\zeta + 2\xi^{-1})}{\Gamma(\zeta + \xi^{-1})}a^{-1} - 1 \right) \)

Mode\( (d_t) = \exp(\psi_{d,t})a^{-1}(\zeta - \xi^{-1})^{1/\xi} \) for \( \zeta \xi > 1 \)

Negative Binomial (absolute price change):

\[ p \equiv \frac{\exp(\psi_{y,t})}{\exp(\psi_{y,t}) + \tau} \]

\[ f_{\mathcal{N}\mathcal{B}}(y_t|\psi_{y,t},\tau) = \frac{\Gamma(y_t + \tau)}{\Gamma(y_t + 1)\Gamma(\tau)}(1 - p)^\tau p^{y_t} \]

\[ F_{\mathcal{N}\mathcal{B}}(y_t|\psi_{y,t},\tau) = 1 - I_p(y_t + 1, \tau) \]

Discrete support: \( y_t \in \{0,1,2,\ldots\} \), number of failures: \( \tau > 0 \)

\( E(y_t) = \exp(\psi_{y,t}) \), \( \text{Var}(y_t) = \exp(\psi_{y,t}) + \exp(\psi_{y,t})^2 \tau^{-1} \)

Mode\( (y_t) = \exp(1 - \tau^{-1}) \) for \( \tau > 1 \) else 0

\( \dagger \) \( f(\cdot) \), \( F(\cdot) \) denote the density and cumulative distribution function, respectively.

\( \dagger \dagger \gamma(\cdot) \) denotes the lower incomplete gamma function.

\( \dagger \dagger \dagger I_p(\cdot) \) denotes the normalized incomplete beta function.

follow the unconditional distribution, Eq. (B.7). Note that each trading day the model is initialized anew, starting with the first observed price duration.

Duration is modeled in continuous time using the generalized gamma \( (\mathcal{G}\mathcal{G}) \) distribution, having shape parameters \( \zeta \) and \( \xi \), Eq. (B.1). The generalized gamma distribution nests the Weibull \( (\zeta = 1) \), gamma \( (\xi = 1) \), and exponential \( (\zeta = \xi = 1) \) as special cases (further the log-normal as \( \zeta \rightarrow \infty \)), enabling a systematic approach
B.2. THE MODEL

to model fitting. The Burr would be another flexible distribution, containing the Weibull, Log-Logistic, and exponential distribution as special cases (see e.g. Grammig and Maurer, 2000, for an application to the ACD class). Hautsch (2002) provides different specifications of the ACD model based on the F-distribution, which includes the generalized gamma and Log-Logistic distributions.

Price movements on the tick scale are discrete. Absolute price change is taken as observable proxy for volatility in event time and is assumed negative binomial (NB) distributed with shape parameter \( \tau \), Eq. (B.2). This distribution is parsimonious but allows for overdispersion, in contrast to the Poisson.

Time-of-day (tod) effects common in intraday data (e.g. Engle and Russell, 2010, or Engle and Russell, 1998, and references therein) are explicitly modeled for absolute price change \( y_t \) to preserve discreteness (tick property). Specifically, \( J_y \) is a "selector" matrix such that \( \kappa_y J_y \) selects the appropriate time-of-day dummy from \( 1 \times R \) row vector \( \kappa_y \), where \( R \in \mathbb{N}^+ | R \geq 2 \), is the number of dummies. For identification the first dummy of the day must be set to zero.

Both \( \mathcal{GG} \) and \( \mathcal{NB} \) distributions are normalized to have \( E(d_t | F_t) = \exp(\psi_{d,t}) \) and \( E(y_t | F_t) = \exp(\psi_{y,t} + \kappa_y J_y) \), respectively, with \( F_t \) the information set at time \( t \). A description of the distributions is given in Tab. B.1 (dropping dummy term \( \kappa_y J_y \) for absolute price change).

To account for fast changing nonlinear patterns of market activity over the day, mode parameters \( \mu_{s,t} \) of the latent information flow each follow their own regime dependent process, with \( M_k \), \( k = \{d, y\} \), the number of states, Eq. (B.5). This allows for a characterization of volatility regimes according to different duration/absolute price change patterns. Zhang, Russell, and Tsay (2001) propose a regime switching ACD model (threshold ACD). Model specification is completed by assuming latent state variables \( s_{k,t} \) to follow a first-order Markov process (see e.g. Hamilton, 1989) with time invariant transition probabilities \( \Pr(s_{k,t} = j | s_{k,t-1} = i) = p_{k,ij}, k = \{d, y\} \), and \( i, j \in \{1, \ldots, M_k\} \). The transition probability matrix has the general form

\[
P_k = \begin{pmatrix}
p_{k,11} & \cdots & p_{k,1M_k} \\
\vdots & \ddots & \vdots \\
p_{k,M_k1} & \cdots & p_{k,M_kM_k}
\end{pmatrix}.
\]
Finally, the proposed model may be gauged in relation to the UHF-GARCH model of Engle (2000), or the IDV model of Grammig and Wellner (2002), as referred in the introduction. These models deliver high frequency measures of volatility, and the volatility measure in event time as produced by the SCDSVrs model can readily be converted to common fixed time units, if desired. Moreover, Engle and Russell (1998) describe a measure for instantaneous volatility in event time based on price duration and derive conditional volatility per second in the next instant as a limiting case. However, this measure is downward biased, as the observed price change will either hit or exceed the threshold. This is especially true for small, sensitive thresholds as the one applied in this work. The proposed model may thus also be interpreted as an extension of the latter approach in discrete time, providing a fully specified absolute price change process. The next section describes the associated McMC algorithm for posterior computation.

B.3 McMC Algorithm

The proposed SCDSV model is set up in state space form. In contrast to the observation driven ACD/(G)ARCH class of models it is parameter driven, by augmenting the state space with a set of unknown latent variables. Deriving an analytical solution of the likelihood is in general not possible and a numerical solution of the high dimensional integral consisting of unobservables and parameters infeasible. However, simulation based estimation methods have been proven capable of estimating these kind of models. Consequently, a Markov chain Monte Carlo (McMC) algorithm is presented, building on an adaption of the block sampler first proposed by Shephard and Pitt (1997) and modified by Watanabe and Omori (2004). Strickland et al. (2006) develop a McMC algorithm to estimate their SCD model along the lines of Durbin and Koopman (2003, 2008), which is conceptually comparable to the proposed algorithm here. They compare performance with the quasi maximum likelihood (QML) approach of Bauwens and Veredas (2004) and find McMC superior. Feng et al. (2004) apply Monte Carlo maximum likelihood (MCML) (Durbin and Koopman, 1997, Shephard
B.3. McMC ALGORITHM


Define parameter set $\theta \equiv \{\mu, \kappa, \Phi, \xi, \xi, \tau, P_k\}$, with state dependent matrices $\mu = (\mu_d, \mu_y)'$, $\mu_k = (\mu_{d,k}, \mu_{y,k})'$, $k = \{d, y\}$. Further denote by $\theta_{-r}$ all parameters of $\theta$ without $r$, and let $d_1:n = (d_1, \ldots, d_n)'$, $y_1:n = (y_1, \ldots, y_n)'$, $\psi_{k,1:n} = (\psi_{k,1}, \ldots, \psi_{k,n})'$, $\psi_{1:n} = (\psi_{d,1:n}, \psi_{y,1:n})$, $s_{k,0:n} = (s_{k,0}, s_{k,1}, \ldots, s_{k,n})'$. Prior distributions are assumed to be independent and generically denoted by $\pi(\cdot)$. Specifically, for the latent process specific state dependent unconditional modes conjugate multivariate normal priors $\pi(\mu_k)$ are assumed,

$$\mu_k \sim N(\mu_{0,k}, \Sigma_{\mu_{0,k}}),$$

with $\Sigma_{\mu_{0,k}}$ diagonal. Further assume $\pi(\Sigma^{-1})$ conjugate Wishart,

$$\Sigma^{-1} \sim W(\nu_0, V_0),$$

and $\pi(P_{k,i})$ conjugate Dirichlet,

$$P_{k,i} \sim D(\omega_{k,i1}, \ldots, \omega_{k,iM_k}), \quad \omega_{k,ij} \neq 0, \quad k = \{d, y\}, \quad i, j = 1, \ldots, M_k,$$

where $P_{k,i}$ denotes the $i^{th}$ (independent) row of $P_k$.

The McMC algorithm then consists of the following steps:\footnote{Implementation (also the particle filter of Sec. B.4) is in stand alone C++ code developed by the author using the freely available Scythe statistical library (Pemstein, Quinn, and Martin, 2011).}

1. Initialize $\theta, s_{d,0:n}, s_{y,0:n}, \psi_{1:n}$.
2. Generate $\psi_{1:n} | \theta, s_{d,0:n}, s_{y,0:n}, d_{1:n}, y_{1:n}$.
3. Generate $s_{d,0:n} | \theta, \psi_{d,1:n}, d_{1:n}$.
4. Generate $s_{y,0:n} | \theta, \psi_{y,1:n}, y_{1:n}$.
5. Generate $P_d | s_{d,0:n}$.
6. Generate $P_y | s_{y,0:n}$.
7. Generate \((\zeta', \xi') | \theta_{-(\zeta', \xi')}, \psi_{d,1:n}, s_{d,0:n}, d_{1:n}\).

8. Generate \(\tau | \theta_{-\tau}, \psi_{y,1:n}, s_{y,0:n}, y_{1:n}\).

9. Generate \(\kappa_y | \theta_{-\kappa_y}, \psi_{y,1:n}, s_{y,0:n}, y_{1:n}\).

10. Generate \(\mu | \theta_{-\mu}, s_{d,0:n}, s_{y,0:n}, \psi_{1:n}\).

11. Generate \(\Sigma | \theta_{-\Sigma}, s_{d,0:n}, s_{y,0:n}, \psi_{1:n}\).

12. Generate \(\Phi | \theta_{-\Phi}, s_{d,0:n}, s_{y,0:n}, \psi_{1:n}\).

13. Go to 2.

In the sequel each updating step of the McMC algorithm is presented in detail.

**Generation of states \(\psi_{1:n}, s_{k,0:n}, \) and transition probability matrices \(P_k\)**

**Step 2.** Latent states \(\psi_{1:n}\) are sampled efficiently in blocks from the true distribution, adapting the multi-move sampler proposed by Shephard and Pitt (1997) and modified by Watanabe and Omori (2004). It is detailed in App. B.A.

**Step 3.** Sampling of discrete states \(s_{d,0:n}\) is done with a forward filtering backward sampling (FFBS) algorithm as outlined in Chib (1996), or Kim and Nelson (1998, 1999). The forward pass calculates the filtered probabilities recursively by iterating between step 1 and 2, see below. Further define \(\Psi_t \equiv \{\theta, \psi_{d,1:t}, d_{1:t}\}\), the information set available at time \(t\) to draw inference on \(s_{d,t}\) (dropping subscript \(d\) in the following for notational clarity).

**Forward Filter - Step 1:** Given \(\pi(s_{t-1}|\Psi_{t-1})\), \(s_t \in \{1, \ldots, M\}\), at the beginning of the \(i^{th}\) iteration, weighting terms are calculated as

\[
\pi(s_t, s_{t-1}|\Psi_{t-1}) = \pi(s_t|s_{t-1})\pi(s_{t-1}|\Psi_{t-1}),
\]

with \(\pi(s_t|s_{t-1})\) the transition probabilities.
Forward Filter - Step 2: Once $d_t$ is observed at the end of the $t^{th}$ iteration, probabilities are updated,

$$\pi(s_t | \Psi_t) = \sum_{s_{t-1}} \frac{f(d_t | s_t, \Psi_t) \pi(s_t, s_{t-1} | \Psi_{t-1})}{f(d_t | \Psi_t)},$$

with $f(d_t | s_t, \Psi_t)$ the conditional likelihood of Eq. (B.1), reparameterized as in Sec. B.4, Eq. (B.12), and

$$f(d_t | \Psi_t) = \sum_{s_t} \sum_{s_{t-1}} f(d_t | s_t, \Psi_t) \pi(s_t, s_{t-1} | \Psi_{t-1}).$$

The filter is initialized with the unconditional probabilities,

$$\pi(s_0) = (A' A)^{-1} A' \begin{pmatrix} 0_M \\ 1 \end{pmatrix}, \quad A = \begin{pmatrix} I_M - P' \\ i'_M \end{pmatrix},$$

where $0_M$ is a $M \times 1$ vector of zeros and $i'_M$ a $M \times 1$ vector of ones.

Backward Sampling: States $s_{0:n}$ are drawn as a block from the following conditional distribution

$$\pi(s_{0:n} | \Psi_n) = \pi(s_n | \Psi_n) \prod_{t=0}^{n-1} \pi(s_t | s_{t+1}, \Psi_t),$$

where

$$\pi(s_t | s_{t+1}, \Psi_t) \propto \pi(s_{t+1} | s_t) \pi(s_t | \Psi_t).$$

Then, we can sample $s_t$ from posterior mass function

$$\pi(s_t = i | s_{t+1}, \Psi_t) = \frac{\pi(s_{t+1} | s_t = i) \pi(s_t = i | \Psi_t)}{\sum_j^M \pi(s_{t+1} | s_t = j) \pi(s_t = j | \Psi_t)}, \quad i = 1, \ldots, M.$$

The backward sampler is initialized with a draw from $\pi(s_n | \Psi_n)$.

**Step 4.** In analogy to Step 3.
Step 5. Conditional on \( s_{d,0:n} \), transition probability matrix \( P_d \) is independent of all other variables in the model. Further, rows \( P_{d,i} \) of \( P_d \) are independent a posteriori and are drawn from the following Dirichlet distribution (omitting subscript \( d \)),

\[
P_i \sim \mathcal{D}(\omega_i + n_{i1}(s_{0:n}), \ldots, \omega_i + n_{iM}(s_{0:n})), \quad i = 1, \ldots, M,
\]

where \( n_{ij}(s_{0:n}) = \#\{s_{t-1} = i, s_t = j\} \) counts the transitions from \( i \) to \( j \).

Step 6. In analogy to Step 5.

**Generation of distributional shape parameters \( \zeta, \xi, \tau \)**

Step 7. No closed form exists for the conditional posterior of generalized gamma shape parameters \( \zeta, \xi, \) and a symmetric random walk Metropolis algorithm (RWM) is applied. Both shape parameters are sampled jointly, an advisable strategy considering their high correlation, see Sec. B.5. To improve upon the often slow convergence of RWM, an adaptive Metropolis (AM) algorithm with global scaling as in Andrieu and Thoms (2008) is implemented. The algorithm is outlined in App. B.C.

Step 8. Similar to step 7, no closed form exists for the conditional posterior of negative binomial shape parameter \( \tau \), and sampling is by RWM. Since draws are univariate, the algorithm in App. B.C simplifies accordingly.

**Generation of time-of-day dummies \( \kappa_y \)**

Step 9. Time-of-day dummies \( \kappa_y \) are independent of each other. Sampling them one-by-one applying RWM in analogy to Step 5 is a natural strategy, as the conditional posterior does not have closed form.\(^5\)

\(^5\)One may sample \( \kappa_y \) in closed form along the lines of Step 8. However, experiments have shown that estimating dummies in the transition equation requires much more data for a precise extraction, else estimates appear biased towards zero.
### B.3. McMC ALGORITHM

**Generation of latent process parameters $\mu, \Phi, \Sigma$**

**Step 10.** Sampling of latent process modes $\mu_k$, $k = \{d, y\}$, could be done by specifying a complete multivariate linear regression system (see Step 12, further Rachev, Hsu, Bagasheva, and Fabozzi, 2008, Ch. 4). A more sensible strategy that exploits the specific context would be to sample the conditionals $\mu_d | \mu_y, \cdot$ and $\mu_y | \mu_d, \cdot$, i.e. all states simultaneously for duration and absolute price change, respectively. Following this idea, and first considering sampling latent duration modes $\mu_d$, note that the innovation error of the transition equation for duration can be written as

$$
\eta_{d,t} = \frac{\sigma_d}{\sigma_y} \rho \eta_{y,t} + \sigma_d \sqrt{1 - \rho^2} \kappa_t, \quad \begin{pmatrix} \kappa_t \\ \eta_{y,t} \end{pmatrix} \sim \mathcal{N}(0, \text{diag}(1, \sigma_y^2)),
$$

with

$$
\eta_{y,t} = \tilde{\psi}_{y,t} - \phi_{yy} \tilde{\psi}_{y,t-1} - \phi_{yd} \tilde{\psi}_{d,t-1}, \quad \tilde{\psi}_t = \psi_t - \mu_{s_t},
$$

and $\sigma_d, \sigma_y, \rho$ the components of transition error covariance $\Sigma$, Eq. (B.6). This is a regression of $\eta_{d,t}$ on $\eta_{y,t}$, with slope coefficient $\sigma_d \sigma_y^{-1} \rho$ and error variance $\sigma_d^2 (1 - \rho^2)$. The conditional posterior probability of $\mu_d$ can then be expressed as

$$
\pi(\mu_d | \cdot) \propto \pi(\mu_d) \times \exp \left\{ - \frac{\{ \tilde{\psi}_{d,1} - \hat{\sigma}_{d,ini} \eta_{y,1} \}_1^2}{2 \hat{\sigma}_{d,ini}^2} \right\} \times
$$

$$
\exp \left\{ - \frac{1}{2} (\mu_d - \mu_{0,d})' \Sigma_{0,d}^{-1} (\mu_d - \mu_{0,d}) \right\} \times
$$

$$
\exp \left\{ - \frac{1}{2} (Y_{\mu_d} - X_{\mu_d} \mu_d)' (Y_{\mu_d} - X_{\mu_d} \mu_d) \right\},
$$

with

$$
x_{\mu_d,t} = \hat{\sigma}_d^{-1} \times
$$

$$
\begin{bmatrix}
I_{s_1=1} - (\phi_{dd} - \hat{\sigma}_d \phi_{yd}) I_{s_{t-1}=1} & \ldots & I_{s_1=M_d} - (\phi_{dd} - \hat{\sigma}_d \phi_{yd}) I_{s_{t-1}=M_d}
\end{bmatrix},
$$

$$
y_{\mu_d,t} = \hat{\sigma}_d^{-1} (\tilde{\psi}_{d,t} - \phi_{dd} \tilde{\psi}_{d,t-1} - \phi_{dy} \tilde{\psi}_{y,t-1} - \hat{\sigma}_d (\tilde{\psi}_{y,t} - \phi_{yy} \tilde{\psi}_{y,t-1} - \phi_{yd} \tilde{\psi}_{d,t-1})),
$$

$$
t = 2, \ldots, n,
$$

$$
x_{\mu_d,1} = \hat{\sigma}_{d,ini}^{-1} \begin{bmatrix}
I_{s_1=1} & \ldots & I_{s_1=M_d}
\end{bmatrix}, \quad y_{\mu_d,1} = \hat{\sigma}_{d,ini}^{-1} (\psi_{d,1} - \hat{\sigma}_{d,ini} \tilde{\psi}_{y,1})\]
the row elements of $X_{\mu_d}, Y_{\mu_d}$, and
\[
\hat{\sigma}_d = \sigma_d \sqrt{1 - \rho^2}, \quad \hat{\sigma}_d = \frac{\sigma_d}{\sigma_y} \rho \quad (\hat{\sigma}_{d, \text{ini}}, \hat{\sigma}_{d, \text{ini}} \text{ accordingly}).
\]

Then we have
\[
\mu_d | \cdot \sim N(\mu_{\mu_d}, \Sigma_{\mu_d}),
\]
where
\[
\mu_{\mu_d} = \Sigma_{\mu_d} (\Sigma_{\mu_0,d}^{-1} \mu_{0,d} + X'_{\mu_d} Y_{\mu_d}), \quad \Sigma_{\mu_d} = (\Sigma_{\mu_0,d}^{-1} + X'_{\mu_d} X_{\mu_d})^{-1}.
\]

Sampling of $\mu_{y|\mu_d, \cdot}$ proceeds in analogy.

Label switching in mixture models may occur when using McMC for estimation, due to the invariance of the likelihood to relabeling the components. This issue has been well investigated (see e.g. Frühwirth-Schnatter, 2001, 2006) and is dealt with by introducing appropriate constraints on mode parameters $\mu$, see Sec. B.6.3.

Step 11. Sampling of $\Sigma$ is by the Metropolis-Hastings (MH) algorithm (see e.g Chib and Greenberg, 1995), accounting for the initial condition. The conditional posterior of $\Sigma$ is
\[
\pi(\Sigma | \cdot) \propto c(\Sigma) \times |\Sigma|^{-\frac{\nu + 3}{2}} \exp\left\{-\frac{1}{2} \sum_{t=2}^{n} v_t' \Sigma^{-1} v_t \right\} \tag{B.8}
\]
\[
= c(\Sigma) \times |\Sigma|^{-\frac{\nu + 3}{2}} \exp\left\{-\frac{1}{2} \text{tr} \left( V^{-1} \Sigma^{-1} \right) \right\}, \tag{B.9}
\]
where
\[
c(\Sigma) = |\Sigma_{\text{ini}}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \bar{\psi}_1' \Sigma_{\text{ini}}^{-1} \bar{\psi}_1 \right\}, \tag{B.10}
\]
\[
V^{-1} = V_0^{-1} + \sum_{t=2}^{n} v_t v_t', \quad v_t = \bar{\psi}_t - \Phi \bar{\psi}_{t-1}, \quad \nu = \nu_0 + n - 1. \tag{B.11}
\]

A proposal $\Sigma^* \sim [W(\nu, V)]^{-1}$ is accepted with probability $\min\{c(\Sigma^*)/c(\Sigma), 1\}$, where $\Sigma$ is the current sample.
B.4. PARTICLE FILTER

Step 12. Similar to Step 11, sampling of \( \Phi \) is by the MH algorithm. Following linear regression theory, the conditional posterior probability is

\[
\pi(\Phi|\cdot) \propto c(\Phi) \times \exp\left\{ -\frac{1}{2} \sum_{t=2}^{n} v_t' \Sigma^{-1} v_t \right\} = c(\Phi) \times \exp\left\{ -\frac{1}{2} \text{tr} \left\{ (Y_\Phi - X_\Phi \Phi')(Y_\Phi - X_\Phi \Phi')\Sigma^{-1} \right\} \right\},
\]

where \( c(\Phi) = c(\Sigma) \) of Eq. (B.10), \( v_t \) is defined as in Eq. (B.11), and

\[
X_\Phi = \tilde{\psi}_{1:n-1}, \quad Y_\Phi = \tilde{\psi}_{2:n}.
\]

Then propose a candidate \( \Phi^* \sim T \mathcal{N}_R(\mu_\Phi, \Sigma_\Phi) \), with

\[
\mu_\Phi = \text{vec} \left( (X_\Phi' X_\Phi)^{-1} (X_\Phi' Y_\Phi) \right), \quad \Sigma_\Phi = \Sigma \otimes (X_\Phi' X_\Phi)^{-1},
\]

and \( T \mathcal{N}_R(\cdot) \) denotes a multivariate normal distribution truncated over region \( R \) where all eigenvalues of \( \Phi^* \) have modulus less than one. Accept the draw with probability \( \min \{ c(\Phi^*)/c(\Phi), 1 \} \), where \( \Phi \) is the current sample.

B.4 Particle Filter

A stratified auxiliary particle filter is developed that recursively delivers draws of the unobservable components \( \{ \alpha_t = \psi_t - \mu_{s_t}, s_{d_t}, s_{y_t} \} \), given parameter values \( \theta \) and observables up to time \( t \), \( \{ d_{1:t}, y_{1:t} \} \). These are necessary for evaluating the conditional likelihood, calculating goodness-of-fit statistics, and forecasting. Pitt and Shephard (1999a) propose the auxiliary particle filter, a sequential Monte Carlo (SMC) technique that increases efficiency in the propagation process by weighting particles according to an importance function dependent on the subsequent observation. Further, multinomial resampling in particle filters is known to increase the Monte Carlo variance of the associated estimators. A stratified approach is able to significantly reduce that variance (Douc, Cappé, and Moulines, 2005, or Hol, Schön, and Gustafsson, 2006). For more information on particle filtering, consider Whiteley and Johansen (2011), or Doucet, de Freitas, and Gordon (2001).
Let \( z_t = (d_t, y_t)' \), \( z_{1:t} = (z_1, \ldots, z_t)' \), and define the extended state set \( x_t = \{ \alpha_t, s_{d,t}, s_{y,t} \} \) (for better readability, time invariant parameter set \( \theta \) may be dropped from the conditioning arguments in this section). The SCDSVrs model is a nonlinear and non-Gaussian state space model with measurement density

\[
f(z_t|x_t) = f(d_t|x_t) f(y_t|x_t) = \mathcal{G}\mathcal{G}(\exp[\alpha_{d,t} + \mu_{d,s_{d,t}}], \zeta, \xi) \times \mathcal{N}\mathcal{B}(\exp[\alpha_{y,t} + \mu_{y,s_{y,t}} + \kappa_y J_y], \tau)
\]

and transition densities

\[
f(\alpha_t|\alpha_{t-1}) = \mathcal{N}(\Phi \alpha_{t-1}, \Sigma),
\]

\[
\Pr(s_{k,t}|s_{k,t-1}) = \text{Multinomial}(p_{k,s_{t-1}}), \quad k = \{d, y\},
\]

where \( p_{k,s_{t-1}} = p_{k,i1}, \ldots, p_{k,im_k} \)' \( i \in M_k \). Applying Bayes Theorem we obtain the target posterior density,

\[
f(x_t, x_{t-1}|z_{1:t}) \propto f(z_t|x_t)f(x_t|x_{t-1})f(x_{t-1}|z_{1:t-1}),
\]

with

\[
f(x_t|x_{t-1}) \propto f(\alpha_t|\alpha_{t-1})\Pr(s_{d,t}|s_{d,t-1})\Pr(s_{y,t}|s_{y,t-1}),
\]

where we assume that we have samples (particles) from \( f(x_{t-1}|z_{1:t-1}) \) and a discrete uniform approximation \( \hat{f}(x_{t-1}|z_{1:t-1}) \) to \( f(x_{t-1}|z_{1:t-1}) \).

By including \( z_t \) in the importance probability density function more weight is given to particles with larger predictive values. The importance function that is used to sample from Eq. (B.14) then has the following form,

\[
g(x_t, x_{t-1}^{(i)}|z_{1:t})
\]

\[
\propto f(z_t|\tilde{\alpha}_{t}^{(i)}, s_{d,t}, s_{y,t}) f(\alpha_t|\alpha_{t-1}^{(i)})\Pr(s_{d,t}|s_{d,t-1}^{(i)})\Pr(s_{y,t}|s_{y,t-1}^{(i)})\hat{f}(x_{t-1}^{(i)}|z_{1:t-1})
\]

\[
\propto f(\alpha_t|\alpha_{t-1}^{(i)})g(x_{t-1}^{(i)}, s_{d,t}, s_{y,t}|z_{1:t}, \tilde{\alpha}_{t}^{(i)}),
\]

with

\[
g(x_{t-1}^{(i)}, s_{d,t}, s_{y,t}|z_{1:t}, \tilde{\alpha}_{t}^{(i)})
\]

\[
\propto f(z_t|\tilde{\alpha}_{t}^{(i)}, s_{d,t}, s_{y,t})\Pr(s_{d,t}|s_{d,t-1}^{(i)})\Pr(s_{y,t}|s_{y,t-1}^{(i)})\hat{f}(x_{t-1}^{(i)}|z_{1:t-1})
\]
and "best" guess
\[ \hat{\alpha}_t^{(i)} = \Phi \alpha_{t-1}^{(i)}, \]
where superscripts \( i = 1, \ldots, I \) denote particles. Observe that no best guess is made for \( s_{d,t} \) and \( s_{y,t} \). Instead, \( g(x_{t-1}^{(i)}, s_{d,t}, s_{y,t} | z_{1:t}, \hat{\alpha}_t^{(i)}) \), Eq. (B.15), is evaluated for every possible \((s_{d,t}, s_{y,t})\) (strata), resulting in a collection of \( I \times M_d \times M_y \) weighted sample points. Then, \( \{x_{t-1}, s_{d,t}, s_{y,t}\} \) are drawn jointly from the respective distribution. Accordingly, there is no direct sampling from \( \Pr(s_{k,t}|s_{k,t-1}^{(i)}), k \in \{d, y\} \). The proposed particle filter follows:

1. Initialization, \( t = 1 (i = 1, \ldots, I, k \in \{d, y\}) \):
   
   (a) Draw \( s_{k,1}^{(i)} \sim \text{Multinomial}(p_{k,1}) \), with \( p_{k,1} \) the initial distribution of \( P_k \).
   
   (b) Draw \( \alpha_1^{(i)} \sim \mathcal{N}(0, \Sigma_{\text{ini}}) \) from the unconditional distribution.
   
   (c) Compute \( w_1^{(i)} = f(z_1|x_1^{(i)}) \), and let \( \hat{f}(x_1^{(i)}|z_1) = w_1^{(i)}/\sum_{j=1}^{I} w_1^{(j)} \).

2. Iterate, \( t = 2, \ldots, n (i, j = 1, \ldots, I) \):
   
   (a) Generate \( \{x_t^{(i)}, x_{t-1}^{(i)}\} \) from \( g(x_t, x_{t-1}^{(j)} | z_{1:t}) \), with \( x_{t-1} \) the current particle set at time \( t - 1 \), as follows. First, evaluate importance weights on \( \{(s_{d,t}, s_{y,t})|s_{d,t} = 1, \ldots, M_d, s_{y,t} = 1, \ldots, M_y\} \) (pre-weighting),
   
   \[ w_{t-1}^{(j, l, m)} \propto g(x_{t-1}^{(j)}, s_{d,t} = l, s_{y,t} = m | z_{1:t}, \hat{\alpha}_t^{(j)}) \]
   
   and resample \( \{x_{t-1}^{(j)}, l, m, w_{t-1}^{(j, l, m)}\}_{(j, l, m) \in \{1, \ldots, I\} \times \{1, \ldots, M_d\} \times \{1, \ldots, M_y\}} \) to obtain \( \{x_t^{(i)}, s_d^{(i)}, s_y^{(i)}, I^{-1}\} \). Then generate \( \alpha_t^{(i)} \) from \( f(\alpha_t|\alpha_{t-1}^{(i)}) \), Eq. (B.13).
   
   (b) Compute weights (correcting for the pre-weighting)
   
   \[ w_t^{(i)} = \frac{f(z_t|x_t^{(i)})}{f(z_t|\hat{\alpha}_t^{(i)}, s_d^{(i)}, s_y^{(i)})}, \]
   
   and let \( \hat{f}(x_t^{(i)}|z_{1:t}) = w_t^{(i)}/\sum_{j=1}^{I} w_t^{(j)} \).

Draws \( x_{t|t}^{(i)} \) obtained from Step 1 below are used to calculate posterior log-likelihood ordinate \( \log f_{\text{post}}(z_{1:n} | \theta) = \sum_{t=1}^{n} \log \hat{f}(z_t | z_{1:t}, \theta) \). Draws \( \alpha_{t+1|t}^{(i)} \) are used to simulate
one-step-ahead prediction density \( f(z_{t+1}|z_{1:t}, \theta) \), which is needed for calculation of predictive log-likelihood ordinate \( \log f_{\text{prior}}(z_{1:n}|\theta) = \sum_{t=1}^{n} \log \hat{f}(z_{t}|z_{1:t-1}, \theta) \), forecasting, and model diagnostics. Simulation of the one-step-ahead prediction density is summarized by the following steps \((i, j = 1, \ldots, I)\):

1. Resample from \( \hat{f}(x^{(j)}_{t}|z_{1:t}) \) after Step 1.(c) and 2.(b) to obtain \( x^{(i)}_{t}|t \).

2. Generate \( \alpha_{t+1,t}^{(i)} \) from \( f(\alpha_{t+1}|\alpha^{(i)}_{t}, \theta) \), Eq. (B.13).

3. Estimate the one-step-ahead prediction density by averaging the mixture

\[
\hat{f}(z_{t+1}|z_{1:t}, \theta) = I^{-1} \sum_{i=1}^{I} \left( \sum_{l=1}^{M_d} \Pr(s_{d,t+1} = l|s^{(i)}_{d,t}) f(d_{t+1}|\alpha^{(i)}_{t+1,t}, s_{d,t+1} = l) \right) \times \left( \sum_{m=1}^{M_y} \Pr(s_{y,t+1} = m|s^{(i)}_{y,t}) f(y_{t+1}|\alpha^{(i)}_{t+1,t}, s_{y,t+1} = m) \right).
\]

Diebold, Gunther, and Tay (1998) show that the correct predictive density is weakly superior to all other forecasts, i.e. will be preferred, in terms of expected loss, by all forecast users regardless of their loss function. Testing whether the forecasting densities are correct can be done by exploiting properties of the probability integral transform. To this end, and first considering the predictive distribution of duration, the probability that \( d_{t+1} \) will be less than observed \( d_{o,t+1} \) can be estimated as \( F(d_{t+1} \leq d_{o,t+1}|z_{1:t}, \theta) \approx \hat{u}_{d,t+1} \), which is approximated in analogy to Eq. (B.16). \( F(\cdot) \) denotes the cumulative distribution function. Under the null hypothesis of a correctly specified model, observed durations \( d_{o,t} \) are random draws from \( f(d_{t}|\cdot) \), compare Eq. (B.12), and \( \hat{u}_{d,t} \ (t = 1, \ldots, n) \) converges in distribution to independent and identically distributed uniform random variables as \( I \to \infty \) (Rosenblatt 1952). This provides a valid basis for diagnostic checking and was popularized in different contexts by Kim et al. (1998) and Diebold et al. (1998), among others. Bauwens, Giot, Grammig, and Veredas (2004) apply the technique on intraday financial duration data, comparing several ACD/SCD specifications.

Regarding the predictive distribution of absolute price change, the above method is not directly applicable as the underlying key assumption of a continuous cumulative distribution function is violated. However, a continuization of \( y_{t} \) can be achieved by
adding random noise (Liesenfeld et al., 2006). The proposed modified method works as follows. Draw a state independent uniform random variate $u_{t+1}^{01} \sim U(0,1)$, and replace Step 3 above ($i = 1, \ldots, I, m = 1, \ldots, M_y$):

3.1 Calculate $u^{up,(i)}_{y,t+1,s_{y,t+1}=m}$ as the probability that $y_{t+1} \leq y_{t+1}^{0} | s_{y,t+1} = m$,

$$u^{up,(i)}_{y,t+1,s_{y,t+1}=m} = F(y_{t+1} \leq y_{t+1}^{0} | \alpha^{(i)}_{y,t+1} | t, s_{y,t+1} = m).$$

3.2 Calculate $u^{low,(i)}_{y,t+1,s_{y,t+1}=m}$ as the probability that $y_{t+1} \leq y_{t+1}^{0} - 1 | s_{y,t+1} = m$,

$$u^{low,(i)}_{y,t+1,s_{y,t+1}=m} = F(y_{t+1} \leq y_{t+1}^{0} - 1 | \alpha^{(i)}_{y,t+1} | t, s_{y,t+1} = m).$$

3.3 Average the mixture

$$F(y_{t+1} \leq y_{t+1}^{0} | z_{1:t}, \theta) \equiv \hat{u}_{y,t+1} = I^{-1} \sum_{i=1}^{I} \sum_{m=1}^{M_y} \Pr(s_{y,t+1} = m | s_{y,t}^{(i)}) \times \left[ u^{low,(i)}_{y,t+1,s_{y,t+1}=m} + u^{01}_{t+1}(u^{up,(i)}_{y,t+1,s_{y,t+1}=m} - u^{low,(i)}_{y,t+1,s_{y,t+1}=m}) \right].$$

If the model is correctly specified, $\hat{u}_{y,t}$ ($t = 1, \ldots, n$) converges in distribution to independent and identically distributed random variables as $I \to \infty$.

**B.5 Adaptive Metropolis for the Generalized Gamma Shape Parameters**

From a Bayesian perspective, the shape parameters of the generalized gamma (GG) distribution are difficult to estimate due to their high correlation. Further, when estimating SCD models in a maximum likelihood setting, Bauwens and Veredas (2004) point to the possibility of overparameterization. However, for the specific dataset at hand, the smooth and persistent duration information flow process found in the one regime case was only possible to extract using the generalized gamma distribution. The gamma and Weibull distributions failed in this case. In a modeling context, the
flexible generalized gamma distribution presents a natural starting point as it can readily be inferred if simpler, nested alternatives are justified. This motivates the search for an efficient estimation strategy and consequently, the performance of two different adaptive Metropolis (AM) algorithms to jointly estimate shape parameters $\zeta$, $\xi$ are compared in this section.

Datasets are generated with parameters that mirror empirical values found for the regime invariant SCDSV model of Sec. B.6.2. True $GG$ shape parameters are set as follows,

$$\zeta = 150, \quad \xi = 0.1,$$

where independent log-normal priors are assumed,

$$\zeta \sim LN(-7, 49), \quad \xi \sim LN(-7, 49),$$

(B.17)

with mean 1 and standard deviation 1096.6. These priors reflect a very weak belief in the exponential distribution. Investigated adaptive Metropolis algorithms are variants of Andrieu and Thoms (2008), namely the AM with componentwise adaptive scaling (Algorithm 6), and a less sophisticated variant, the AM with global adaptive scaling (Algorithm 4). The key idea of these algorithms is to increase efficiency in the sampling process by adapting the proposal covariance during the McMC run, using information from the previous chain. See App. B.C for technical details. It must be noted that in order to estimate the $GG$ shape parameters efficiently, not to stop adaption after the burn-in has proven advantageous. Accordingly, the concept of diminishing adaption is applied (consider e.g. Andrieu and Thoms, 2008). However, ergodicity of the Markov chain may be an issue. This concern is addressed with a rather extensive simulation study to demonstrate convergence of the algorithms to the true values.

Results are from reduced runs, keeping all parameters but $\zeta$, $\xi$ fix. Datasets have empirically relevant size 60,000. In each application of the algorithm, 100 datasets are simulated. McMC is run for 100,000 iterations, discarding the initial 50,000 draws. Convergence of the AM with componentwise adaptive scaling (Algorithm 6) has shown to be rather slow, explaining the large burn-in.
Table B.2: Simulation Results of the Generalized Gamma Shape Parameters

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>Mean</th>
<th>Stdev</th>
<th>Min</th>
<th>Max</th>
<th>Avg. IF</th>
<th>Avg. CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algo 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\zeta$</td>
<td>150</td>
<td>152.80</td>
<td>34.581</td>
<td>92.742</td>
<td>249.73</td>
<td>157.9</td>
<td>0.79</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.1</td>
<td>0.1019</td>
<td>0.0109</td>
<td>0.0784</td>
<td>0.1280</td>
<td>150.9</td>
<td>0.80</td>
</tr>
<tr>
<td>Algo 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\zeta$</td>
<td>150</td>
<td>152.00</td>
<td>37.308</td>
<td>92.640</td>
<td>265.68</td>
<td>1130.8</td>
<td>0.49</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.1</td>
<td>0.1021</td>
<td>0.0116</td>
<td>0.0765</td>
<td>0.1275</td>
<td>1120.9</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Reduced runs. Mean, standard deviation, minimum, maximum of the posterior mean estimates, average inefficiency factor, and average Geweke’s convergence diagnostic (p-value) from 100 replications of the SCDSV model each.

Tab. B.2 reports results. Setting standard deviation of posterior mean estimates and minimum/maximum values in relation to exactness of posterior mean estimates, then convergence is confidently indicated. Average p-values of Geweke’s (1992) convergence diagnostic (CD) (see App. A.C for the deployed formulae) support the claim. Analyzing performance of both algorithms, less sophisticated global scaling Algorithm 4 features far lower average inefficiency factors (IF) and higher average CD p-values. Further, average acceptance rate for Algorithm 4 is 35.0%, which is very close to the theoretical optimal rate of $\approx 35\%$ (in two dimensions), guaranteeing sufficient efficiency (Roberts and Rosenthal, 2001). There is no direct control on the acceptance rate of Algorithm 6, which is with 88.5% probably too high, as the inefficiency factors let suggest.

Fig. B.1 illustrates performance of the two algorithms and the parameter identification problem. The first to columns show sample paths of generalized gamma shape parameters $\zeta$, $\xi$ from a selected simulated dataset. Fast convergence and an efficient exploration of the parameter space are clearly visible for Algorithm 4. On the contrary, deviations from equilibrium values are more persistent and can be highly pronounced for Algorithm 6. Scatter plots in the last column show the better exploration of state space by Algorithm 4, visualizing the extremely high correlation between shape parameters of the $\mathcal{GG}$ distribution ($-0.9865$ and $-0.9887$ on average for Algorithm 4 and 6, respectively). The above exercise illustrates the difficulties in estimating the shape parameters of the generalized gamma distribution and highlights the importance of choosing an efficient sampling strategy.
Figure B.1: Comparison of AM Algorithms for \( \mathcal{G} \mathcal{G} \) Parameter Estimation

### B.6 Empirical Application

#### B.6.1 Data

Models are estimated using quote data of the IBM stock traded on the New York Stock Exchange (NYSE). Source is the TAQ database. The period analyzed is from 2001-10-01 to 2001-10-31, a total of 22 trading days or 140,881 observations after deleting Columbus Day (2001-10-08), dropping zero/erroneous bid-ask quotes/sizes, and not considering quotes recorded before 9:30 AM (official open)/after 16:00 PM (official close). Price duration is calculated using the midprice of the most recently posted quote in each second. A price threshold of \(|c_p| = 0.5\) ticks is introduced, considering the smallest possible price change of 0.5 cent (decimalization was completed on the 29th January 2001 on the NYSE). Defining larger thresholds would introduce some
hidden aggregation of price changes, which increasingly distorts the joint innovation error of latent duration and absolute price change. The latter measure thins the sample to 46.8% of the original dataset, or 65,889 observations.

One may suggest that midprice moves of 0.5 ticks are largely transitory and a result of market microstructure effects (for a discussion of the bid-ask bounce and further microstructure issues, consider e.g. Campbell, Lo, and MacKinlay, 1997, Ch. 3). Consequently, an own data generating process may be assumed, modeling 0.5 ticks separately (see Hautsch, 2013, in a related context). However, it is found that the proposed regime switching model specification captures the predictive distribution of absolute price change rather well.

Tab. B.3 presents descriptive statistics of the dataset. Fig. B.2 plots price, absolute price change, and histograms of duration and absolute price change for the IBM stock in October 2001. There appears to be an upward trend during the period (upper left plot), and several price jumps are visible (upper right plot). Looking at the histograms, we infer that for both duration and absolute price change most of the probability mass is centered around small values (middle plots). Specifically, the mode for duration is just 3 seconds and that for absolute price change 0.5 ticks. Interestingly, for both series roughly 30% of the data is smaller or equal to the mode. Probably more important, there is a significant amount of probability mass in the right tail of the empirical distributions (lower plots), especially regarding absolute price change. For the latter this is of particular interest, as outliers reflect price jumps (or rare events). Moreover, such irregular distributed data encourages the search for different regimes, and in fact a jump state is consistently identified.

Intraday data generally features seasonal effects, arising from a systematic variation of market activity during the day (e.g. Engle and Russell, 1998, Bauwens and
Figure B.2: Descriptive Statistics of the IBM stock (NYSE), 10/2001 (65,889 obs.). Price duration in seconds. Abs. price change in half-ticks, threshold 0 = 0.5 ticks.

Figure B.3: Diurnal Components of the IBM Stock, 10/2001 (cubic spline)

Veredas, 2004). In a modeling context, we may think of duration and absolute price change as consisting of two parts: A stochastic component to be explained by the
model, and a deterministic part, namely the diurnal component. Consider Fig. B.3 for an illustration of time-of-day seasonal patterns for duration and absolute price change, extracted using a cubic spline. Regarding duration, we have the characteristic peak around lunch time, with most intense trading around half an hour after the open and before the close. For absolute price change, we have a clear positive effect around the open, then declining until bottom is touched around lunchtime, and rising again slightly afterwards. The latter is an important observation as it implies that the higher volatility observed around the open and close is a result of both more intense and larger price updates (especially in the morning).

A common practice in the literature is to pre-adjust intraday data for seasonal effects, and to use the adjusted data for model fitting. In the current application, however, this is done only for price duration using the cubic spline of Fig. B.3. The reason for not pre-adjusting absolute price change is preservation of discreteness, i.e. the tick property. Accordingly, diurnal components are estimated jointly with the model for this series.

Finally, Fig. B.4 demonstrates the presence of serial dependence in duration and absolute price change using Spearman’s $\rho$ (autocorrelation coefficients are misleading
for irregular spaced data). Comparing raw and adjusted data we conclude that accounting for seasonal effects removes only part of the serial dependence in the series.

B.6.2 The SCDSV Model

First, the SCDSV model without regime switching is fitted to the data. The subsequent section covers the regime switching case. As prior distributions are assumed

\[
\begin{align*}
\mu_d &\sim N(0, 10^2), \quad \mu_y \sim N(1.5, 10^2), \quad k_j^y \sim N(0, 10^2), \quad j = \{1, \ldots, 8\}, \\
\frac{\phi_{kk} + 1}{2} &\sim \mathcal{B}(20, 1.5), \quad \frac{\phi_{kl} + 1}{2} \sim \mathcal{B}(1, 1), \quad k, l = \{d, y\}, \quad l \neq k, \\
\Sigma^{-1} &\sim \mathcal{W}(5, (5\Sigma_0)^{-1}) \text{ with } \Sigma_0 = 0.2^2 I_2, \\
\zeta &\sim \mathcal{LN}(-7, 49), \quad \xi \sim \mathcal{LN}(-7, 49), \quad \tau \sim \mathcal{LN}(-7, 49),
\end{align*}
\]

where \(\mathcal{B}(\cdot)\) denotes the beta and \(\mathcal{LN}(\cdot)\) the log-normal distribution. The priors on latent duration and absolute price change modes \(\mu_d, \mu_y\) are mildly informative reflecting empirical relevant values. Those for time-of-day dummies \(k_j^y\) are conservatively centered around zero. The diagonal components of \(\Phi\) have priors implying a modest belief in a persistent information flow with mean 0.86 and standard deviation 0.11. The off-diagonals modeling cross effects receive uninformative priors with equal probability on the domain \((-1, 1)\). The prior on \(\Sigma^{-1}\) is chosen such that \(\mathbb{E}(\Sigma^{-1}) = \Sigma_0^{-1}\), conservatively assuming zero correlation. Distributional shape parameters \(\zeta, \xi, \tau\) receive very diffuse priors, having mean 1 and a standard deviation of 1096.6. In the case of the generalized gamma distribution, this prior mirrors a weak belief in the exponential distribution.

The first 10,000 draws are discarded, and the following 100,000 draws are collected for inference. In the multi-move sampler, average block size as a tuning parameter is set to 7 and number of iterations to achieve convergence to 3. Posterior means, standard deviations, 95% credible bounds, inefficiency factors (IF), and p-values of Geweke’s (1992) convergence diagnostic (CD) are reported (see App. A.C for the deployed formulae). For both models estimated in this work, the null of the CD
Table B.4: Estimation Results of the SCDSV Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_d$</td>
<td>-0.0183</td>
<td>0.0067</td>
<td>[-0.0315, -0.0052]</td>
<td>7.0</td>
<td>0.98</td>
</tr>
<tr>
<td>$\mu_y$</td>
<td>1.3852</td>
<td>0.0346</td>
<td>[1.3195, 1.4549]</td>
<td>108.8</td>
<td>0.78</td>
</tr>
<tr>
<td>$\phi_{dd}$</td>
<td>0.9408</td>
<td>0.0052</td>
<td>[0.9299, 0.9501]</td>
<td>165.0</td>
<td>0.99</td>
</tr>
<tr>
<td>$\phi_{dy}$</td>
<td>-0.0186</td>
<td>0.0019</td>
<td>[-0.0223, -0.0150]</td>
<td>99.1</td>
<td>0.92</td>
</tr>
<tr>
<td>$\phi_{yd}$</td>
<td>-0.0768</td>
<td>0.0049</td>
<td>[-0.0862, -0.0670]</td>
<td>37.0</td>
<td>0.98</td>
</tr>
<tr>
<td>$\phi_{yy}$</td>
<td>0.8929</td>
<td>0.0049</td>
<td>[0.8829, 0.9021]</td>
<td>73.8</td>
<td>1.00</td>
</tr>
<tr>
<td>$\sigma_d$</td>
<td>0.0779</td>
<td>0.0043</td>
<td>[0.0700, 0.0870]</td>
<td>775.8</td>
<td>0.96</td>
</tr>
<tr>
<td>$\sigma_y$</td>
<td>0.2502</td>
<td>0.0073</td>
<td>[0.2365, 0.2647]</td>
<td>231.5</td>
<td>1.00</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.5889</td>
<td>0.0299</td>
<td>[0.5261, 0.6446]</td>
<td>305.8</td>
<td>0.78</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>156.53</td>
<td>34.267</td>
<td>[107.05, 241.46]</td>
<td>387.6</td>
<td>0.77</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.0974</td>
<td>0.0098</td>
<td>[0.0771, 0.1160]</td>
<td>319.9</td>
<td>0.76</td>
</tr>
<tr>
<td>$\tau$</td>
<td>1.2089</td>
<td>0.0149</td>
<td>[1.1799, 1.2386]</td>
<td>24.6</td>
<td>0.99</td>
</tr>
<tr>
<td>$\kappa_{y,2}$</td>
<td>-0.1563</td>
<td>0.0464</td>
<td>[-0.2467, -0.0663]</td>
<td>166.4</td>
<td>0.83</td>
</tr>
<tr>
<td>$\kappa_{y,3}$</td>
<td>-0.3563</td>
<td>0.0487</td>
<td>[-0.4520, -0.2610]</td>
<td>167.3</td>
<td>0.77</td>
</tr>
<tr>
<td>$\kappa_{y,4}$</td>
<td>-0.5342</td>
<td>0.0425</td>
<td>[-0.6190, -0.4517]</td>
<td>218.9</td>
<td>0.74</td>
</tr>
<tr>
<td>$\kappa_{y,5}$</td>
<td>-0.6244</td>
<td>0.0436</td>
<td>[-0.7109, -0.5406]</td>
<td>193.2</td>
<td>0.81</td>
</tr>
<tr>
<td>$\kappa_{y,6}$</td>
<td>-0.6100</td>
<td>0.0441</td>
<td>[-0.6978, -0.5227]</td>
<td>208.3</td>
<td>0.80</td>
</tr>
<tr>
<td>$\kappa_{y,7}$</td>
<td>-0.5087</td>
<td>0.0425</td>
<td>[-0.5943, -0.4286]</td>
<td>216.6</td>
<td>0.73</td>
</tr>
<tr>
<td>$\kappa_{y,8}$</td>
<td>-0.4102</td>
<td>0.0482</td>
<td>[-0.5064, -0.3154]</td>
<td>157.9</td>
<td>0.89</td>
</tr>
<tr>
<td>$\kappa_{y,9}$</td>
<td>-0.3534</td>
<td>0.0483</td>
<td>[-0.4500, -0.2602]</td>
<td>159.9</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor and Geweke’s convergence diagnostic (p-value).

statistic ("chain has converged") can not be rejected at conventional significance levels for all parameters.

Average AR/MH acceptance rates of latent information flow process $\psi_t$ in the multi-move sampler are high, 99.5% and 99.4%, respectively, as are those for the MH steps sampling $\Phi$ and $\Sigma$, which are 94.3% and 98.3%, respectively. Acceptance rate for the adaptive Metropolis algorithm estimating $GG$ shape parameters $\zeta$, $\xi$ is with 35.0% close to the theoretical optimal value of $\approx 35\%$ (two dimensions). The same holds for acceptance rates of $NB$ shape parameter $\tau$ and dummy variables $\kappa_y$, which range between 42.2% and 44.6% (optimal value is $\approx 44\%$ for one dimension). The above results suggest good mixing properties of the proposed McMC algorithm.

Tab. B.4 reports results. Both information flow processes feature persistence, high for duration, $\phi_{dd} = 0.941$, and more moderate for absolute price change, $\phi_{yy} = 0.893$. 

B.6. EMPIRICAL APPLICATION
Cross effects between the latent processes as captured by $\Phi$ have been found, although they appear rather small in absolute magnitude. Concretely, past latent absolute price change has a negative impact on expected duration, $\phi_{dy} = -0.0186$. Said differently, large past latent price change leads to shorter expected duration. This is intuitive, as large price changes are mainly triggered by unexpected news, which probably cause further adjacent price updates. Congruently, past latent duration negatively influences expected absolute price change, $\phi_{yd} = -0.0768$. In other words, lower past latent duration leads to higher expected absolute price change. This is again intuitive, as shorter durations probably indicate the arrival of new information, which makes larger future price changes more likely. Engle (2000) analyzes the impact of past trade duration on volatility and obtains qualitatively equal results.\textsuperscript{6} Moreover, $\phi_{yd}$ is about four times higher in size than $\phi_{dy}$. Volatility of the latent duration process is remarkably low, $\sigma_d = 0.0779$, whereas its absolute price change counterpart is more than three times higher, $\sigma_y = 0.250$, reflecting a more diffuse process for the latter. Innovation error of latent information flow is rather strong positively correlated, $\rho = 0.590$. The more time passes between successive price changes, the higher the price change. A possible interpretation could be a continuous hidden accumulation of information, causing larger price moves as market participants finally decide to trade. Interestingly,

\textsuperscript{6}The author also provides further links to traditional market microstructure theory in the context of trade duration/intensity.
the concurrent dynamics generated by information signal $\eta_t$ contrast the intertemporal cross effect forces expressed by $\phi_{dy}, \phi_{yd}$.

Shape parameters $\zeta, \xi$ of the generalized gamma distribution and their respective 95% credible bounds clearly indicate that neither the gamma, Weibull or exponential are supported by the data. Observe also the wide credible bounds for $\zeta$, visualizing the difficulties in estimating the $GG$ shape parameters. Further support for the $GG$ distribution comes from Lunde (1999), applying it to the ACD class, and from Bauwens et al. (2004), emphasizing strong predictive performance. Inferred tod dummies $\kappa_y$ clearly reflect the diurnal patterns shown in Fig. B.3 and are all significant. Fig. B.5 plots observation densities at their unconditional mean, where average time-of-day effect is taken into account. Inefficiency factors are in acceptable ranges regarding the number of collected draws. Only a high inefficiency for $\sigma_d$ stands out, which is linked to the small magnitude of this parameter.

As already mentioned in Sec. B.5, for the specific dataset at hand, a persistent duration process in the regime invariant case could not be extracted using the gamma or Weibull distribution. This strongly supports the use of a flexible, general-to-specific structured distribution like the generalized gamma as a starting point for model fitting.

### B.6.3 The Regime Switching SCDSVrs Model

Regime switching for latent duration and absolute price change modes is introduced to better capture the complex dynamics of intraday data. Thereby, each latent mode follows its own independent state process. Three duration and four absolute price change regimes are extracted, mirroring fast changing market patterns during the day. One regime models rare events, or price jumps. The number of states has been determined by multiple criteria, most importantly autocorrelation in probability integral transforms $\hat{u}_t$ and $\hat{u}_{t}^2$, but also the Kupier statistic and a visual inspection of histograms based on $\hat{u}_t$ (compare Sec. B.6.5). The deviance information criterion ($DIC$) (Sec. B.6.4) always favored the higher parameterized specification, making it less helpful in the decision process. The reason is a relatively low penalty for model complexity, a direct result of the large sample size. In fact, up to nine duration regimes
Table B.5: Estimation Results of the SCDSVrs Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{d,1}$</td>
<td>-1.0741</td>
<td>0.0247</td>
<td>[-1.1214, -1.0274]</td>
<td>237.1</td>
<td>0.67</td>
</tr>
<tr>
<td>$\mu_{d,2}$</td>
<td>-0.0992</td>
<td>0.0252</td>
<td>[-0.1465, -0.0521]</td>
<td>247.0</td>
<td>0.63</td>
</tr>
<tr>
<td>$\mu_{d,3}$</td>
<td>0.7797</td>
<td>0.0191</td>
<td>[0.7426, 0.8188]</td>
<td>145.1</td>
<td>0.74</td>
</tr>
<tr>
<td>$\mu_{y,1}$</td>
<td>-1.8169</td>
<td>0.1342</td>
<td>[-2.0694, -1.5341]</td>
<td>1,187.1</td>
<td>0.15</td>
</tr>
<tr>
<td>$\mu_{y,2}$</td>
<td>0.2727</td>
<td>0.0740</td>
<td>[0.1291, 0.4170]</td>
<td>428.3</td>
<td>0.45</td>
</tr>
<tr>
<td>$\mu_{y,3}$</td>
<td>1.7444</td>
<td>0.0335</td>
<td>[1.6779, 1.8088]</td>
<td>109.1</td>
<td>0.82</td>
</tr>
<tr>
<td>$\mu_{y,4}$</td>
<td>4.1923</td>
<td>0.1378</td>
<td>[3.9212, 4.4592]</td>
<td>135.5</td>
<td>0.89</td>
</tr>
<tr>
<td>$\phi_{dd}$</td>
<td>0.9050</td>
<td>0.0098</td>
<td>[0.8860, 0.9241]</td>
<td>346.9</td>
<td>0.90</td>
</tr>
<tr>
<td>$\phi_{dy}$</td>
<td>-0.0266</td>
<td>0.0033</td>
<td>[-0.0330, -0.0203]</td>
<td>118.4</td>
<td>0.89</td>
</tr>
<tr>
<td>$\phi_{yd}$</td>
<td>-0.0861</td>
<td>0.0065</td>
<td>[-0.0988, -0.0733]</td>
<td>125.8</td>
<td>0.99</td>
</tr>
<tr>
<td>$\phi_{yy}$</td>
<td>0.9179</td>
<td>0.0060</td>
<td>[0.9056, 0.9295]</td>
<td>156.6</td>
<td>0.96</td>
</tr>
<tr>
<td>$\sigma_{d}$</td>
<td>0.1099</td>
<td>0.0079</td>
<td>[0.0937, 0.1248]</td>
<td>1,936.8</td>
<td>0.93</td>
</tr>
<tr>
<td>$\sigma_{y}$</td>
<td>0.1912</td>
<td>0.0094</td>
<td>[0.1726, 0.2106]</td>
<td>1,071.8</td>
<td>0.95</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.6549</td>
<td>0.0295</td>
<td>[0.5935, 0.7102]</td>
<td>400.0</td>
<td>0.80</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>3.9602</td>
<td>0.0912</td>
<td>[3.7888, 4.1427]</td>
<td>321.1</td>
<td>0.81</td>
</tr>
<tr>
<td>$\tau$</td>
<td>3.7623</td>
<td>0.1843</td>
<td>[3.4223, 4.1422]</td>
<td>172.3</td>
<td>0.85</td>
</tr>
<tr>
<td>$\kappa_{y,2}$</td>
<td>-0.1335</td>
<td>0.0440</td>
<td>[-0.2176, -0.0475]</td>
<td>237.9</td>
<td>0.95</td>
</tr>
<tr>
<td>$\kappa_{y,3}$</td>
<td>-0.3251</td>
<td>0.0450</td>
<td>[-0.4149, -0.2379]</td>
<td>212.9</td>
<td>0.98</td>
</tr>
<tr>
<td>$\kappa_{y,4}$</td>
<td>-0.4881</td>
<td>0.0396</td>
<td>[-0.5646, -0.4096]</td>
<td>280.2</td>
<td>0.85</td>
</tr>
<tr>
<td>$\kappa_{y,5}$</td>
<td>-0.5757</td>
<td>0.0424</td>
<td>[-0.6586, -0.4912]</td>
<td>269.0</td>
<td>0.96</td>
</tr>
<tr>
<td>$\kappa_{y,6}$</td>
<td>-0.5503</td>
<td>0.0415</td>
<td>[-0.6318, -0.4681]</td>
<td>278.5</td>
<td>0.91</td>
</tr>
<tr>
<td>$\kappa_{y,7}$</td>
<td>-0.4644</td>
<td>0.0404</td>
<td>[-0.5442, -0.3866]</td>
<td>277.1</td>
<td>0.84</td>
</tr>
<tr>
<td>$\kappa_{y,8}$</td>
<td>-0.3712</td>
<td>0.0454</td>
<td>[-0.4613, -0.2838]</td>
<td>203.2</td>
<td>0.91</td>
</tr>
<tr>
<td>$\kappa_{y,9}$</td>
<td>-0.3216</td>
<td>0.0452</td>
<td>[-0.4103, -0.2329]</td>
<td>202.8</td>
<td>0.93</td>
</tr>
</tbody>
</table>


have been identified from the data during the model fitting process. Priors are assumed as in Sec. B.6.2, except as follows,

$$
\mu_d \sim N((-1.1, -0.1, 0.8)', I_3), \quad \mu_y \sim N((-1.8, 0.3, 1.8, 4.2)', I_4), \\
P_{k,i.} \sim D(1, \ldots, 1), \quad i = \{1, \ldots, M_k\}, \quad k = \{d, y\}.
$$

The priors on latent duration and absolute price change modes $\mu_k$ are empirically relevant and mildly informative to reflect the specific state identification constraints.
(μ_{d,1} < μ_{d,2} < μ_{d,3}, \mu_{y,1} < μ_{y,2} < μ_{y,3} < μ_{y,4}), which prevent label switching (see also Sec. B.3, Step 8). Priors for the rows of \( P_k \) are uniform over the unit simplex.

Similar to the SCDSV model, McMC discards the first 10,000 draws as burn-in and collects the next 100,000 for parameter inference. Average block size in the multi-move sampler and number of iterations to achieve convergence are as in Sec. B.6.2. AR/MH acceptance rates of the actual run are also comparable.

Estimation Results are reported in Tab. B.5. Transition probability matrices are given in Eq. (B.18) and (B.19), together with average observations in each state. Complete statistics of the former have been put in App. B.B, Tab. B.9 and B.10. Latent duration and absolute price change mode parameters \( \mu_k \) differentiate the regimes significantly. Note a distinctively high absolute price change mode \( \mu_{y,4} \) reflecting the price jump state. Fig. B.6 then plots unconditional observation densities of the different
states, where average time-of-day effect is taken into account. Remarkably, absolute price change state 1 puts almost all probability mass (89.5%) on moves of 0.5 ticks, suggesting an interpretation as market microstructure state. Moreover, the mode of state 2 is also located at 0.5 ticks, with a significant point mass of 44.5%.

\[
\begin{array}{ccc}
\text{cond. mean} & \text{low} & \text{high} \\
\hline
s_{d,t} \\
1 & 0.2315 & 0.5703 & 0.1982 \\
2 & 0.3257 & 0.5217 & 0.1525 \\
3 & 0.3147 & 0.4463 & 0.2390 \\
\end{array}
\]

\[
\text{avg. observations} \\
19,496.1 & 34,422.2 & 11,970.7
\]

Taking a closer look at the transition probability matrices in Eq. (B.18) and (B.19), state transitions of duration show a rather transient character. Only state 2 features a slight persistence of 52.2%, containing also 52.2% of the observations. State transitions of absolute price change, however, are more persistent in nature. Specifically,
Figure B.7: SCDSVrs - Posterior Probability of States. Upper figure: Duration ($s_{t,d}$). Lower figure: Absolute price change ($s_{t,y}$). 2001/10/01, 9:50-10:20 AM, 220 obs.
states 2 and 3 are the most persistent, with 70.5% and 75.6% probability to remain in the respective state. State 3 featuring medium to high price changes contains the most observations, about 63.7% of the data. Interestingly, after having observed a price jump, the chance to remain in the rare event state is as high as 47.3%. This indicates that higher uncertainty in the market regarding price change will likely not be resolved within one larger adjustment. Assuming state 1 to model mainly market microstructure effects, we would expect a rather transient character, and indeed we observe a probability of only 29.4% to remain in this state.

Posterior probability of the states for a half hour window in the morning of 2001/10/01, from 9:50 AM to 10:20 AM, is plotted in Fig. B.7. Analyzing duration first, the transient character is clearly visible. Note the higher persistence of state 2 and pronounced occurrences of state 3 modeling periods of price inactivity. Regarding absolute price change, the higher persistence for medium states 2 and 3 is visualized. State 1 shows frequent and distinct occurrences, which adds to the interpretation of this state as modeling largely market microstructure effects. Moreover, prolonged periods of zero or near zero probability are observed in this state, indicating persistent periods of market activity. Finally, observe a price jump just before 10 o’clock (which actually reverses immediately after) (state 4).

Regarding the remaining parameters, information flow of duration has become less persistent, with $\phi_{dd} = 0.905$ lower and $\sigma_d = 0.110$ higher than in the SCDSV model. The latent absolute price change process, on the contrary, has become more persistent, with $\phi_{yy} = 0.918$ higher and $\sigma_y = 0.191$ lower. It can be argued that innovation of the absolute price change process is now in part modeled by regime switching parameters $\mu_y$, reducing the transition error in that way. Moreover, the effect of past latent absolute price change on expected duration is now stronger, $\phi_{dy} = -0.0266$. Correlation between the innovation terms of latent duration and absolute price change is with $\rho = 0.655$ comparable in magnitude to that of the SCDSV model. Shape parameter $\tau$ of the negative binomial distribution is significantly higher relative to the no regime case and implies all else equal a lower variance of the observation error. Again one may conclude that this is a result of the highly different absolute price change levels identified by the regime switching approach. Dummy variables are of equal magnitude.
Inefficiency factors are overall higher than for the SCDSV model, attributable in large part to the regime switching approach. Specifically, note high inefficiencies encountered when sampling mode parameters $\mu_{y,1}$, $\mu_{y,2}$. The associated states put a significant probability mass on 0.5 tick moves, which appears to impede extraction of the modes. This is also reflected in low $CD$ statistics for the parameters, and relatively large credible bounds, especially for $\mu_{y,1}$.

Fig. B.8 then plots associated information flow process $\psi_t$ for 2001/10/01, 9:50-10:20 AM. Analyzing duration first, latent flow in the SCDSV model is very smooth, reflecting low innovation volatility $\sigma_d$ and high persistence $\phi_{dd}$. The SCDSVrs extension, on the contrary, shows an erratic process switching states frequently. The same holds for the absolute price change process, albeit to a lesser degree as state transitions show a more persistent trait. A larger innovation volatility $\sigma_y$, compared to $\sigma_d$, results in a more adaptive flow in the one state case, compared to duration.
CHAPTER B. SCDSVrs

As expected, the SCDSVrs variant is far more flexible when it comes to model high price changes, consider especially a distinct difference in the vicinity of the price jump around 10 o’clock. Moreover, it is visible from the lower plot that very small price changes appear to interrupt the information flow. Taking into account that absolute price change state 1 almost exclusively models 0.5 ticks, further the high inefficiency and low $CD$ statistic encountered when estimating this state, then assuming a separate data generating process for half-ticks becomes an obvious extension (see Hautsch, 2013). Note however that the information flow would then be discontinuous, which makes such an extension nontrivial.

B.6.4 Model Selection

Models are compared using the deviance information criterion ($DIC$) (Spiegelhalter, Best, Carlin, and van der Linde, 2002). It is defined by

$$DIC = \bar{D} + p_d,$$

where

$$p_d = \bar{D} - D(\bar{\theta}), \quad \bar{D} = E_{\theta|y}[D(\theta)], \quad D(\theta) = -2 \log f(y_{1:n}|\theta),$$

with $D(\theta)$ the deviance and $\bar{\theta}$ the posterior mean. It is a Bayesian measure of model fit, attaining smaller values for better models, with $p_d$ a penalty term for model complexity. Intuitively, higher parameter uncertainty lets $D(\theta)$ fluctuate more widely around $D(\bar{\theta})$, yielding larger $p_d$ values. Numerical standard errors of the estimates are obtained the following way. Posterior expectation of the deviance $\bar{D}$, readily available as a byproduct of the McMC run, is calculated by dividing the sample in 10 equally sized batches. Regarding deviance of the posterior mean $D(\bar{\theta})$, the particle filter outlined in Sec. B.4 is repeatedly applied 10 times to obtain estimates of log-likelihood ordinate $\log f_{\text{post}}(z_{1:n}|\bar{\theta}) = \sum_{t=1}^{n} \log \hat{f}(z_t|z_{1:t},\bar{\theta})$. The deviance of the pre-

7Indeed, descriptive statistics of the data show that, after having observed an absolute price move of 0.5 ticks, the price sign reverses in 83.2% of the cases. This transient feature supports the assumption of a different data generating process.
B.6. EMPIRICAL APPLICATION

Table B.6: Model Selection

<table>
<thead>
<tr>
<th></th>
<th>DIC</th>
<th>(s.e.)</th>
<th>Rank</th>
<th>pD</th>
<th>(s.e.)</th>
<th>Dpred</th>
<th>(s.e.)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCDSV</td>
<td>406,025.4</td>
<td>(92.7)</td>
<td>2</td>
<td>8,532.5</td>
<td>(46.4)</td>
<td>409,975.2</td>
<td>(2.9)</td>
<td>2</td>
</tr>
<tr>
<td>SCDSVrs</td>
<td>328,724.2</td>
<td>(1,900.4)</td>
<td>1</td>
<td>34,523.8</td>
<td>(950.2)</td>
<td>407,829.3</td>
<td>(4.5)</td>
<td>1</td>
</tr>
</tbody>
</table>


deviative log-likelihood is calculated in a similar way from ordinate log \( f_{\text{prior}}(z_{1:n} | \tilde{\theta}) = \sum_{t=1}^{n} \log \hat{f}(z_t | z_{1:t-1}, \tilde{\theta}) \). The above approach yields 100 different DIC values, from which statistics are calculated.

DIC in Tab. B.6 clearly indicates that the regime switching SCDSVrs model is preferred by the data, despite higher complexity penalty \( p_D \).\(^8\) There is also a significant improvement in forecasting ability of the model, as indicated by deviance of the predictive log-likelihood \( D_{\text{pred}} \). However, the difference is much smaller than the DIC suggests. This is in large part due to the DIC using the posterior probability of states, and the predictive likelihood using the conditional, but static and rather transient transition probabilities to calculate mixture forecasts. The next section investigates improvement of the predictive distribution when deploying the more sophisticated SCDSVrs model.

B.6.5 Predictive Density Tests

Predictive performance of the proposed models is compared using density forecast evaluation techniques (compare Sec. B.4 and the references therein). Parameter estimates are those of Sec. B.6.2 and B.6.3. Tests are both in-sample and out-of-sample for the following month, November 2001. This additional dataset contains 55,652 observations after pre-processing the data as in Sec. B.6.1. Descriptive statistics are given in Tab. B.7. Average duration is higher and average absolute price change lower, indicating less price activity (however, the \( \% < \text{Mode} \) is approximately equal). The lower price activity leads to more pronounced time-of-day effects for duration (plots omitted).

\(^8\)However, as noted already in Sec. B.6.3, when interpreting these complexity penalty terms one has to keep in mind the large sample size dealt with.
Table B.7: Descriptive Statistics of the IBM Stock 11/2001 (55,652 obs.)

<table>
<thead>
<tr>
<th></th>
<th>Duration</th>
<th>Abs. Price Δ</th>
<th></th>
<th>Duration</th>
<th>Abs. Price Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>8.62</td>
<td>3.76</td>
<td>stdev</td>
<td>9.09</td>
<td>3.74</td>
</tr>
<tr>
<td>mode</td>
<td>3</td>
<td>1</td>
<td>min</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>% ≤ mode</td>
<td>30.8%</td>
<td>32.2%</td>
<td>max</td>
<td>155</td>
<td>104</td>
</tr>
</tbody>
</table>

Price duration in seconds. Absolute price change in half-ticks, threshold 0.5 ticks.

To obtain predictive densities and associated numerical standard errors for the new dataset, the particle filter is repeatedly applied 10 times, as for the in-sample dataset. Hypothesized i.i.d. uniform \( \hat{u}_{d,t}, \hat{u}_{y,t} \) \( t = 1, \ldots, n \) are obtained by applying the probability integral transform on draws of the predictive density. To avoid a forward looking bias regarding price duration, each day of the 11/2001 dataset is cleaned by tod effects calculated from the past 10 trading days.

Calculating predictive deviance criterion first, the SCDSVrs model is also significantly favored out-of-sample, with a deviance of 339,249.2 (3.0) compared to 341,224.6 (3.6) of the SCDSV model (standard error in parentheses).

Fig. B.9 visualizes predictive density fit using histograms of \( \hat{u}_{d,t}, \hat{u}_{y,t} \). Analyzing in-sample results first (upper four plots), better performance of the SCDSVrs model is clearly visible. Especially for absolute price change the improvement in modeling the predictive density is remarkably (note that confidence intervals are narrow as the sample size is large). Regarding duration, there is a significant improvement in forecasting the middle and right part of the distribution. However, both models have difficulties in predicting the left part correctly. A distinctive peak somehow mirrors the unconditional empirical distribution (with a mode at 3 seconds). Moreover, increasing the number of regimes did not resolve the issue (in fact up to nine regimes have been identified). A possible remedy could be to model duration by a discrete distribution, optionally introducing more complex time varying transition dynamics for very small durations. This is subject to current research.

Out-of-sample results are qualitatively equal. Specifically, there is again better forecasting performance deploying the SCDSVrs model. However, too much observations fall in the top percentile of the predictive regime switching duration distribution. This is probably a result of the distinct data characteristics featuring larger durations on average.
## B.6. EMPIRICAL APPLICATION

### Table B.8: Predictive Performance: Goodness-of-fit Statistics

<table>
<thead>
<tr>
<th></th>
<th>2001-10 Duration</th>
<th>Absolute Price Change</th>
<th>2001-11 Duration</th>
<th>Absolute Price Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LB_{15}$ Data</td>
<td>1996.3</td>
<td>11,426.5</td>
<td>2320.0</td>
<td>7978.8</td>
</tr>
<tr>
<td>SCDSV</td>
<td>47.6 (0.1)</td>
<td>268.8 (10.1)</td>
<td>49.8 (0.1)</td>
<td>322.0 (9.5)</td>
</tr>
<tr>
<td>SCDSVrs</td>
<td>20.1 * (0.1)</td>
<td>33.6 (3.7)</td>
<td>41.5 (0.1)</td>
<td>21.5 * (2.4)</td>
</tr>
<tr>
<td>$LB_{15}^2$ Data</td>
<td>247.4</td>
<td>11,104.9</td>
<td>463.0</td>
<td>9,379.9</td>
</tr>
<tr>
<td>SCDSV</td>
<td>157.5 (0.3)</td>
<td>203.9 (18.6)</td>
<td>120.3 (0.3)</td>
<td>273.0 (15.6)</td>
</tr>
<tr>
<td>SCDSVrs</td>
<td>147.9 (0.2)</td>
<td>32.7 (6.3)</td>
<td>110.9 (0.3)</td>
<td>72.3 (12.6)</td>
</tr>
<tr>
<td>$V$</td>
<td>SCDSV</td>
<td>8.83 (0.02)</td>
<td>10.23 (0.29)</td>
<td>11.32 (0.02)</td>
</tr>
<tr>
<td>SCDSVrs</td>
<td>7.64 (0.02)</td>
<td>1.14 * (0.28)</td>
<td>7.99 (0.02)</td>
<td>5.61 (0.10)</td>
</tr>
</tbody>
</table>

Ljung-Box ($LB$) test at lag 15 for demeaned probability integral transforms $\hat{u}_t$ and $\hat{u}_t^2$. Data is raw data adjusted for tod effects by cubic spline and demeaned. Critical values for the $LB(15)$ test are 30.58/25.00/22.31 for the 1%/5%/10% significance level. Critical values for the Kupier statistic $V$ are 2.001/1.747/1.620 at the 1%/5%/10% level ($H_0: \hat{u}_t \sim U(0,1)$). Numerical standard error in parentheses.

in combination with the higher parameterized SCDSVrs model fitting in-sample data better. The predictive regime switching density of absolute price change puts to much probability in the right part of the distribution, but the imbalance is barely significant. Further, it should be noted that more frequent re-estimation potentially resolves the above issues.

Remaining serial dependence in transformed predictions $\hat{u}_t$ and $\hat{u}_t^2$ is visualized by autocorrelograms in Fig. B.10. Overall we again see better performance of the regime switching SCDSVrs model variant, especially at smaller lags. However, there is no real improvement over the SCDSV model when cleaning mean dynamics of duration out-of-sample. This is probably linked to the aforementioned need to re-estimate more frequently to better model the nonstationarities inherent in financial data. Moreover, both in- and out-of-sample higher moment dynamics $\hat{u}_{d,t}^2$, of duration are not appropriately captured. In a related context, Ghysels et al. (2004) refer to the second moment of trade duration as liquidity risk and propose to model it explicitly.

Finally, Tab. B.8 reports Ljung-Box statistics at lag 15 and the Kupier statistic of series $\hat{u}_{d,t}$, $\hat{u}_{y,t}$. The latter tests for uniformity and complements the graphical discussion of Fig. B.9. An advantage of the Kupier statistic is that it is equally sensitive
Figure B.9: Transformed Predictions $\hat{u}_t$ (grand averages). Horizontal dashed lines are 95% confidence intervals based on a normal approximation of the binomially distributed bin elements ($H_0: \hat{u}_t \sim \mathcal{U}(0, 1)$).
Figure B.10: Autocorrelograms of Transformed Predictions $\hat{u}_t$ (1st and 3rd row) and $\hat{u}_t^2$ (2nd and 4th row) (demeaned, grand averages, up to lag 25). Horizontal dashed lines are 95% confidence intervals under the null that $\hat{u}_t$, $\hat{u}_t^2$ are iid.
for all values of $\hat{u}_t$.\(^9\) The more well-known Kolmogorov-Smirnov statistic, in contrast, is most sensitive around the median. Kupier statistic $V$ is defined as

$$
D^+ = \max_{1 \leq t \leq n} \left[ \frac{t}{n} - u_t \right], \quad D^- = \max_{1 \leq t \leq n} \left[ u_t - \frac{(t - 1)}{n} \right], \quad V = D^+ + D^-.
$$

The respective null hypothesis of uniformly distributed $\hat{u}_{d,t}, \hat{u}_{y,t}$ can be rejected for all models and all samples at the 1% level, except for absolute price change in the SCDSVrs model in-sample. However, the superiority of the regime switching variant is overall obvious. This is also indicated by the Ljung-Box test, which indicates major improvements for all models and all samples with respect to the raw data. Notably, for the SCDSVrs model in-sample the null hypothesis of no serial correlation in $\hat{u}_{d,t}$ can not be rejected at usual significance levels, further for $\hat{u}_{y,t}$ out-of-sample.\(^10\) Moreover, observe a relatively small improvement in the $LB^2_{15}$ statistic for duration when comparing the raw data and $\hat{u}^2_{d,t}$'s of both models. This once more emphasizes the need to model the second moment of price duration explicitly.

### B.7 Conclusion

To fully capture volatility tick-by-tick the concepts of stochastic conditional duration (SCD) and stochastic volatility (SV) are combined to propose the SCDSV model in event time. Volatility as commonly measured over a fixed interval is then made up of two components, the duration to observe a price change and its absolute magnitude. Consequently, duration and associated absolute price change in event time are modeled contemporaneously as correlated latent processes. For price change mid prices are deployed, applying the smallest possible threshold of 0.5 ticks. The proposed model is applied to IBM stock intraday data.

Persistent information flow is extracted, featuring a positively correlated innovation term and negative cross effects in the AR(1) persistence matrix. It is interesting that the intertemporal cross effect forces expressed by the latter contrast the concurrent

\(^9\)The asymptotic distribution of the statistic is known, but Stephens (1970) provides a simplified finite-sample statistic with negligible error to its asymptotic distribution.

\(^10\)When dealing with sample sizes as large as in the current work, Engle (2000) states in a similar context that "...it is not clear how seriously to take these p-values".
dynamics generated by the information signal. Regime switching is introduced, identifying three duration and four absolute price change states each following their own Markov process. A rare event or price jump state is identified. Model selection and predictive tests confirm the superiority of the regime switching extension in- and out-of-sample. Regarding duration modeling support for the flexible generalized gamma distribution can be given, as no nested alternative was capable of extracting a persistent information flow process in the regime invariant case. Estimating the highly correlated shape parameters of the generalized gamma distribution efficiently is difficult, and adaptive Metropolis algorithms are explored in a simulation study. A McMC algorithm for model estimation and an associated particle filter are proposed and outlined in detail.

Several extensions are imaginable. First of all, no exogenous microstructure variables, foremost transaction intensity, have been included yet. Impact on predictive performance would be especially interesting. In this regard, experimentation with time varying transition probabilities would be tempting. This could also be a possible route to alleviate the inflexibility in the predictive distribution regarding very small price durations. Further, information flow persistence and innovation can be assumed to follow a regime switching process as well. Introducing some feedback mechanism (leverage) from the observation to the transition equation would also be of interest. Then, assuming an own data generating process for half ticks, which are mainly due to microstructure noise, appears to be a sensible choice. Finally, one may decide to give up the price jump state in order to work with significantly smaller datasets. Adding to this point, introducing volume thresholds would be a sensible choice.
B.A Multi-Move Sampler

The multi-move sampler (or block sampler) proposed by Shepard and Pitt (1997) and modified by Watanabe and Omori (2004) delivers draws from the true conditional posterior distribution of $\psi_{1:n}$, using a local linear approximation of second order around the mode. By defining $\alpha_t = \psi_t - \mu_{st}, t = 1, \ldots, n$, one obtains the following alternative representation of Eq. (B.1)-(B.3) as a state space model w.r.t. $\alpha_{1:n}$,

$$d_t \sim GG(\exp[\alpha_{d,t} + \mu_{d,s,t}], \zeta, \xi),$$

$$y_t \sim NB(\exp[\alpha_{y,t} + \mu_{y,s,t} + \kappa_y J_y], \tau), \quad t = 1, \ldots, n,$$

$$\alpha_t = \Phi \alpha_{t-1} + \eta_t, \quad t = 2, \ldots, n. \quad (B.20)$$

To achieve more efficient sampling, $\alpha_{1:n}$ is divided into $K + 1$ blocks at random, say $(\alpha_{k_{i-1}+1}, \ldots, \alpha_{k_i})', i = 1, \ldots, K + 1$, with $k_0 = 0$ and $k_{K+1} = n$, where $K$ is a tuning parameter. I use stochastic knots given by $k_i = \text{int}[n(i + \mathcal{U}_i)/(K + 2)],$ for $i = 1, \ldots, K$ and $k_i - k_{i-1} \geq 1$, with $\mathcal{U}_i$ a random sample from the uniform distribution $\mathcal{U}(0,1)$ (Shepard and Pitt, 1997).

Let $x_t = R_t^{-1} \eta_t$, where $R_t$ denotes a Cholesky decomposition of $\Sigma$ such that $\Sigma = R_t R_t', t > 1$, and of $\Sigma_{ini}, t = 1$, accordingly. Further suppose that $k_{i-1} = r$ and $k_i = r + c$ for the $i^{th}$ block. The sampler then exploits the independence of disturbances $\eta_t$, block sampling $x = (x'_{r+1}, \ldots, x'_{r+c})'$ instead of $\alpha = (\alpha'_{r+1}, \ldots, \alpha'_{r+c})'$ from the respective full joint conditional posterior density $(r \geq 0, c \geq 1, r + c \leq n)$,

$$\pi(x|\alpha_{r+c+1}, \Theta) \propto$$

$$\left( \prod_{t=r+1}^{r+c} f_{GG}(d_t|\alpha_{d,t}, \cdot) f_{NB}(y_t|\alpha_{y,t}, \cdot) \right) \left( \prod_{t=r+1}^{r+c} f(x_t) f(\alpha_{r+c}), \quad r + c < n, \right)$$

$$\pi(x|\Theta) \propto$$

$$\left( \prod_{t=r+1}^{r+c} f_{GG}(d_t|\alpha_{d,t}, \cdot) f_{NB}(y_t|\alpha_{y,t}, \cdot) \right) \left( \prod_{t=r+1}^{r+c} f(x_t) \right), \quad r + c = n. \quad (B.21)$$
with
\[ f(\alpha_{r+c}) = \exp \left\{ -\frac{1}{2} (\alpha_{r+c+1} - \Phi \alpha_{r+c})' \Sigma^{-1} (\alpha_{r+c+1} - \Phi \alpha_{r+c}) \right\} \]

and \( \Theta \equiv \{ d_{r+1:r+c}, y_{r+1:r+c}, s_{d,r+1:r+c}, s_{y,r+1:r+c}, \theta \} \). Measurement distributions \( f_{G\theta}(\cdot) \) and \( f_{\mathcal{N}\mathcal{B}}(\cdot) \) are given Tab. B.1.

To construct a proposal density based on a normal approximation of the posterior density of \( x \), let \( \hat{\psi}_{y,t} = \alpha_{y,t} + \mu_{y,s,y,t} + \kappa_{y} J_{y}, t = 1, \ldots, n \), and start with the logarithm of Eq. (B.21)-(B.22) (excluding constant terms), 
\[-\frac{1}{2} \sum_{t=r+1}^{r+c} x'_t x_t + L, \]
where
\[ L = \sum_{t=r+1}^{r+c} \left\{ -\zeta \xi \psi_{d,t} - \left( \frac{\Gamma(\zeta + \xi^{-1})}{\Gamma(\zeta)} \frac{d_t}{\exp(\psi_{d,t})} \right)^\xi + \right. \]
\[ + \left. \tau \log \left( \frac{\exp(\hat{\psi}_{y,t})}{\exp(\hat{\psi}_{y,t}) + \tau} \right) + y_t \log \left( \frac{\exp(\hat{\psi}_{y,t})}{\exp(\hat{\psi}_{y,t}) + \tau} \right) \right\} + \log f(\alpha_{r+c}) I_{r+c < n}. \]

Further define
\[ \delta = (\delta'_{r+1}, \ldots, \delta'_{r+c})', \quad \delta_t = \frac{\partial L}{\partial \alpha_t} = \left( \delta_{d,t}', \delta_{y,t}' \right) + j(\alpha_t), \]
\[ Q = -E \left( \frac{\partial^2 L}{\partial \alpha \partial \alpha'} \right) = \text{diag}(A_{r+1}, A_{r+2}, \ldots, A_{r+c}), \]
\[ A_t = -E \left( \frac{\partial^2 L}{\partial \alpha_t \alpha'_t} \right) = \text{diag}(A_{d,t}, A_{y,t}) + j'(\alpha_t), \quad t = r + 1, \ldots, r + c. \]
The first derivatives are
\[ \delta_{d,t} = -\zeta \xi + \xi \left( \frac{\Gamma(\zeta + \xi^{-1})}{\Gamma(\zeta)} \frac{d_t}{\exp(\psi_{d,t})} \right)^\xi, \quad \delta_{y,t} = \frac{\tau(y_t - \exp(\hat{\psi}_{y,t}))}{\tau + \exp(\hat{\psi}_{y,t})}, \]
and
\[ j(\alpha_t) = \Phi' \Sigma^{-1} (\alpha_{t+1} - \Phi \alpha_t) I_{t=r+c < n}. \]
For the second derivatives, take expectations with respect to \((d_t, y_t)\)' multiplied by \(-1\) and obtain
\[
A_{d,t} = \xi^2 \left( \Gamma(\zeta + \xi^{-1}) \frac{d_t}{\Gamma(\zeta)} \exp(\psi_{d,t}) \right) \xi, \quad A_{y,t} = \frac{\tau \exp(\hat{\psi}_{y,t})(y_t + \tau)}{(\tau + \exp(\hat{\psi}_{y,t}))^2},
\]
and
\[
j'(\alpha_t) = \Phi'\Sigma^{-1}\Phi_{t=r+c<n}.
\]

Applying a second order Taylor expansion to the log of the posterior density around mode \(\hat{x}\) yields an approximating normal density as follows (excluding constant terms):
\[
\log \pi(x|\cdot) \approx \hat{L} + \frac{\partial L}{\partial \alpha}|_{x = \hat{x}} (x - \hat{x}) + \frac{1}{2} (x - \hat{x})' E \left( \frac{\partial^2 L}{\partial x \partial x'} \right) |_{x = \hat{x}} (x - \hat{x}) - \frac{1}{2} \sum_{t=r+1}^{r+c} x'_t x_t
\]
\[
= \hat{L} + \hat{\delta}'(\alpha - \hat{\alpha}) - \frac{1}{2} (\alpha - \hat{\alpha})' \hat{Q}(\alpha - \hat{\alpha}) - \frac{1}{2} \sum_{t=r+1}^{r+c} x'_t x_t
\]
\[
\equiv \log q(x|\cdot),
\]
where \(\hat{L}, \hat{\delta}\) and \(\hat{Q}\) are the values of \(L, \delta\) and \(Q\) at \(\alpha = \hat{\alpha}\) (or at \(x = \hat{x}\), equivalently). It can be shown that proposal density \(q(x|\cdot)\) is the posterior of \(x\) obtained from an ordinary linear state space model given by Eq. (B.23)-(B.24) below. Mode \(\hat{x}\) is then obtained by repeating the following algorithm until it converges (usually 2 to 5 iterations).

**Algorithm 1 (Disturbance smoother):**

1. Initialize \(\hat{x}\), and compute \(\hat{\alpha}\) at \(x = \hat{x}\) using state equation (B.20) recursively.

2. Evaluate \(\hat{\delta}_t\) and \(\hat{A}_t\) at \(\alpha = \hat{\alpha}, t = r + 1, \ldots, r + c\).

3. Define auxiliary variable \(\hat{z}_t = \hat{\alpha}_t + \hat{A}_t^{-1}\hat{\delta}_t, t = r + 1, \ldots, r + c\), and let \(\hat{K}_t\) be a Cholesky decomposition of \(\hat{A}_t\) such that \(\hat{A}_t = \hat{K}_t\hat{K}'_t\).
4. Consider the linear Gaussian state space model formulated by

\[
\hat{z}_t = \alpha_t + \epsilon_t, \quad (B.23)
\]
\[
\alpha_t = \Phi \alpha_{t-1} + u_t, \quad t = r + 1, \ldots, r + c, \quad t > 1,
\]
\[
\alpha_1 \sim N(0, \Sigma_{ini}), \quad t = 1,
\]
\[
\begin{pmatrix} \epsilon_t \\ u_t \end{pmatrix} \sim N(0, \text{diag}(\hat{A}_t^{-1}, \Sigma)). \quad (B.24)
\]

Apply the Kalman filter (e.g. Durbin and Koopman, 2008) and disturbance smoother (Koopman, 1993) to this state space model to obtain posterior mode \( \hat{x} \) or \( \hat{\alpha} \), equivalently.

5. Go to 2.

In the McMC sampling procedure, the current sample of \( \alpha \) may be taken as initial value of \( \hat{\alpha} \) in Step 1. Further note that the above steps are equivalent to the method of scoring used to maximize the conditional posterior density. After convergence of Algorithm 1, \( \alpha \) is sampled from the conditional posterior density by conducting an AR (Accept - Reject) - MH algorithm (Tierney, 1994) using the simulation smoother (de Jong and Shephard, 1995, or Durbin and Koopman, 2002).

**Algorithm 2 (AR-MH step and simulation smoother):**

1. Let \( \alpha_0 \) denote the current value. Find mode \( \hat{\alpha} \) using Algorithm 1.

2. Proceed with Steps 2-4 in Algorithm 1 to obtain the approximated linear Gaussian state space model of Eq. (B.23)-(B.24).

3. Propose a candidate \( \alpha^* \) by sampling from \( \bar{q}(\alpha^*) \propto \min \{ \pi(\alpha^* | \cdot), cq(\alpha^* | \cdot) \} \) using the AR algorithm as follows:

   (a) Generate \( \alpha^* \) using the simulation smoother for the approximated state space model of Eq. (B.23)-(B.24).
(b) Accept $x^*$ with probability
\[
\frac{\min \{\pi(x^*|\cdot), cq(x^*|\cdot)\}}{cq(x^*|\cdot)},
\]
where $c$ is a scaling constant. If it is rejected, go to (a).

4. Conduct the MH algorithm using candidate $x^*$, with acceptance probability
\[
\min \left\{ \frac{\pi(x^*|\cdot) \times \min \{\pi(x_0|\cdot), cq(x_0|\cdot)\}}{\pi(x_0|\cdot) \times \min \{\pi(x^*|\cdot), cq(x^*|\cdot)\}}, 1 \right\}.
\]

### B.B Transition Probability Estimates

Tab. B.9 and Tab. B.10 report estimation results of the transition probability matrices $P_d, P_y$ for the SCDSVrs model.

#### Table B.9: Results of the SCDSVrs Model - Transition Probability Matrix $P_d$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{d,11}$</td>
<td>0.2315</td>
<td>0.0174</td>
<td>[0.1980, 0.2640]</td>
<td>268.2</td>
<td>0.67</td>
</tr>
<tr>
<td>$P_{d,12}$</td>
<td>0.5703</td>
<td>0.0114</td>
<td>[0.5477, 0.5925]</td>
<td>43.8</td>
<td>0.91</td>
</tr>
<tr>
<td>$P_{d,13}$</td>
<td>0.1982</td>
<td>0.0147</td>
<td>[0.1703, 0.2270]</td>
<td>228.4</td>
<td>0.66</td>
</tr>
<tr>
<td>$P_{d,21}$</td>
<td>0.3257</td>
<td>0.0138</td>
<td>[0.2993, 0.3519]</td>
<td>245.2</td>
<td>0.63</td>
</tr>
<tr>
<td>$P_{d,22}$</td>
<td>0.5217</td>
<td>0.0101</td>
<td>[0.5015, 0.5411]</td>
<td>73.9</td>
<td>0.77</td>
</tr>
<tr>
<td>$P_{d,23}$</td>
<td>0.1525</td>
<td>0.0098</td>
<td>[0.1334, 0.1716]</td>
<td>195.2</td>
<td>0.66</td>
</tr>
<tr>
<td>$P_{d,31}$</td>
<td>0.3147</td>
<td>0.0136</td>
<td>[0.2891, 0.3426]</td>
<td>78.3</td>
<td>0.64</td>
</tr>
<tr>
<td>$P_{d,32}$</td>
<td>0.4463</td>
<td>0.0150</td>
<td>[0.4162, 0.4750]</td>
<td>45.5</td>
<td>0.94</td>
</tr>
<tr>
<td>$P_{d,33}$</td>
<td>0.2390</td>
<td>0.0128</td>
<td>[0.2138, 0.2638]</td>
<td>67.4</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke’s convergence diagnostic (p-value).
Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke’s convergence diagnostic (p-value).

### Table B.10: Results of the SCDSVrs Model - Transition Probability Matrix $P_y$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{y,11}$</td>
<td>0.2942</td>
<td>0.0157</td>
<td>[0.2636, 0.3251]</td>
<td>82.1</td>
<td>0.58</td>
</tr>
<tr>
<td>$p_{y,12}$</td>
<td>0.0372</td>
<td>0.0181</td>
<td>[0.0052, 0.0742]</td>
<td>343.8</td>
<td>0.72</td>
</tr>
<tr>
<td>$p_{y,13}$</td>
<td>0.6680</td>
<td>0.0200</td>
<td>[0.6279, 0.7062]</td>
<td>147.7</td>
<td>0.48</td>
</tr>
<tr>
<td>$p_{y,14}$</td>
<td>0.000576</td>
<td>0.000334</td>
<td>[0.000076, 0.001364]</td>
<td>5.1</td>
<td>0.83</td>
</tr>
<tr>
<td>$p_{y,21}$</td>
<td>0.0443</td>
<td>0.0147</td>
<td>[0.0178, 0.0749]</td>
<td>250.5</td>
<td>0.45</td>
</tr>
<tr>
<td>$p_{y,22}$</td>
<td>0.7052</td>
<td>0.0199</td>
<td>[0.6656, 0.7428]</td>
<td>150.4</td>
<td>0.54</td>
</tr>
<tr>
<td>$p_{y,23}$</td>
<td>0.2502</td>
<td>0.0235</td>
<td>[0.2041, 0.2956]</td>
<td>245.6</td>
<td>0.34</td>
</tr>
<tr>
<td>$p_{y,24}$</td>
<td>0.000282</td>
<td>0.000253</td>
<td>[0.000090, 0.000941]</td>
<td>8.0</td>
<td>0.93</td>
</tr>
<tr>
<td>$p_{y,31}$</td>
<td>0.1558</td>
<td>0.0078</td>
<td>[0.1422, 0.1724]</td>
<td>172.7</td>
<td>0.15</td>
</tr>
<tr>
<td>$p_{y,32}$</td>
<td>0.0872</td>
<td>0.0108</td>
<td>[0.0666, 0.1086]</td>
<td>257.8</td>
<td>0.38</td>
</tr>
<tr>
<td>$p_{y,33}$</td>
<td>0.7564</td>
<td>0.0101</td>
<td>[0.7363, 0.7758]</td>
<td>147.0</td>
<td>0.90</td>
</tr>
<tr>
<td>$p_{y,34}$</td>
<td>0.000530</td>
<td>0.000146</td>
<td>[0.000281, 0.000849]</td>
<td>5.7</td>
<td>0.98</td>
</tr>
<tr>
<td>$p_{d,41}$</td>
<td>0.0688</td>
<td>0.0513</td>
<td>[0.0029, 0.1925]</td>
<td>6.0</td>
<td>0.99</td>
</tr>
<tr>
<td>$p_{d,42}$</td>
<td>0.1212</td>
<td>0.0774</td>
<td>[0.0090, 0.2987]</td>
<td>9.8</td>
<td>0.97</td>
</tr>
<tr>
<td>$p_{d,43}$</td>
<td>0.3372</td>
<td>0.0860</td>
<td>[0.1715, 0.5087]</td>
<td>5.8</td>
<td>0.94</td>
</tr>
<tr>
<td>$p_{d,44}$</td>
<td>0.4727</td>
<td>0.0697</td>
<td>[0.3370, 0.6092]</td>
<td>2.2</td>
<td>0.95</td>
</tr>
</tbody>
</table>

**B.C Adaptive Metropolis Algorithm**

The proposed algorithms are variants of Algorithm 4 and 6 (adaptive Metropolis (AM) with global/component-wise adaptive scaling) in Andrieu and Thoms (2008). The latter authors give an overview of the field.

Let $X_i$, $i = 0, 1, \ldots, I$, denote a sampled parameter vector from distribution $\pi(\cdot)$ under consideration at the $i^{th}$ iteration of the Markov chain. The Metropolis-Hastings algorithm is a common choice in McMC estimation if sampling from the posterior

---

\[\text{This appendix is taken as an independent unit with parameter notation to be separated from the rest of this work.}\]
is not possible in closed form. It requires the definition of a family of proposal distributions \( \{q(x, \cdot), x \in X\} \) that serves to generate possible transitions for the Markov chain, say from \( X \) to \( Y \), which are then accepted or rejected according to probability \( \alpha(X, Y) = \min\{1, \pi(Y)q(Y, X)/[\pi(X)q(X, Y)]\} \).

The univariate normal random walk Metropolis algorithm (RWM) produces symmetric proposals of the form \( y \sim N(x, \sigma^2) \), and the probability of move reduces to \( \alpha(X, Y) = \min\{1, \pi(Y)/\pi(X)\} \). However, it is well known that its performance deteriorates sharply if \( \sigma^2 \) is chosen either too small or too large. Problems increase in the multivariate case. Asymptotic results regarding the optimal variance and the corresponding optimal acceptance rate are known (Roberts and Rosenthal, 2001), but initializing the algorithm with the optimal variance often results in slow convergence to optimal asymptotic efficiency.

The basic idea is to adapt the covariance each \( k \)th iteration during a burn-in period \( I_{\text{burn-in}} \) (thus ensuring ergodicity of the chain), wherein the optimal acceptance rate \( \alpha_{\text{opt}} \) (0.44 for one, 0.35 for two dimensions; 0.234 asymptotically) is taken as efficiency criterion. Let \( X \) be of dimension \( p \times 1 \), \( \Sigma \) denote the proposal covariance, \( R \) the corresponding correlation matrix (used in a variant discussed below), \( \sigma = \text{diag}(\sigma_1, \ldots, \sigma_p) \) a diagonal matrix with distinct proposal standard deviations, and \( c_1, c_2 \) scaling parameters. Further, given a \( p \times 1 \) vector \( V \), denote by \( V(m) \) the \( m \)th component, \( m = 1, \ldots, p \), and define by \( e_m \) a \( p \times 1 \) unit vector with the \( m \)th component equal to one. The proposed algorithm follows:

1. Initialize, \( i = 0: X_i, \Sigma_i = I_p, R_i \) accordingly, and \( \sigma_i \).

2. Iterate, \( i = 1, \ldots, I \):

   (a) Sample \( Y_i \sim N(0, \sigma_{i-1}\Sigma_{i-1}\sigma_{i-1}) \), and set \( X_i = X_{i-1} + Y_i \) with probability \( \alpha_i(X_{i-1}, X_{i-1} + Y_i) \), otherwise \( X_i = X_{i-1} \).

   (b) If \( (i \leq I_{\text{burn-in}}) \land (i \mod k = 0) \) then

      i. \( \bar{\alpha}_m = \frac{1}{k} \sum_{j=i-k+1}^i \alpha_{m,j}(X_{j-1}, X_{j-1} + Y_j(m)e_m), \ m = 1, \ldots, p \).

      ii. \( \log(\sigma_{m,i}^2) = \log(\sigma_{m,i-1}^2) + i^{-c_1} \left[ \bar{\alpha}_m - \alpha_{\text{opt}} \right], \ m = 1, \ldots, p \).

      iii. \( \bar{X}_i = \frac{1}{k} \sum_{j=i-k+1}^i X_j \).
iv. $\hat{\Sigma}_i = \frac{1}{k-1} \sum_{j=i-k+1}^{i} (X_j - \bar{X}_i)(X_j - \bar{X}_i)'$.

v. $\Sigma_i = \Sigma_{i-1} + i^{-c_2} [\hat{\Sigma}_i - \Sigma_{i-1}]$, calculate $R_i$ from $\Sigma_i$.

else

i. $\sigma_{m,i}^2 = \sigma_{m,i-1}^2$, $m = 1, \ldots, p$.

ii. $\Sigma_i = \Sigma_{i-1}$, $R_i = R_{i-1}$.

Initialization of $\sigma_{m,i}^2$ may be with the asymptotic optimal scaling rate of $2.38^2/p$ for RWM. Componentwise adaptive scaling (Algorithm 6) uses the information contained in the proposal $X_{i-1} + Y_i$ about scalings in the different dimensions. Directional acceptance probabilities $\alpha_{m,i}$ are calculated through "virtual" componentwise updates with increments $Y_i(m)e_m$. For the specific problem at hand, using correlation $R$ instead of covariance $\Sigma$ in Step 2.(a) of the algorithm shows faster and smoother convergence. This is also an intuitive strategy, taking only the correlation structure from recently accepted draws; scaling is then by the proposal standard deviations dependent on the directional acceptance probabilities. Global adaptive scaling (Algorithm 4) uses $\alpha_i(X_{i-1}, X_{i-1} + Y_i)$ instead to adapt a single scaling variable $\log(\sigma_i^2)$ that scales $\Sigma$, steps 2.(a) and 2.(b)ii modify accordingly.

Note that for the specific application at hand, the generalized gamma shape parameters are rather pronounced, and continuing adaption has been shown to result in more efficient estimation, see further Sec. B.5. In that case, scaling parameter $c_1 = 0.1$ and covariance adaption related $c_2 = 0.001$. In all univariate applications, scaling parameter $c_1 = 0$, resulting in an aggressive adaption of $\sigma$ during the burn-in period. The batch size has been set to $k = 100$ throughout.
Chapter C

Multivariate Stochastic Volatility
with Dynamic Cross Leverage
C.1 Introduction

Multivariate Stochastic Volatility (MSV) has more recently become an active and exiting research field. Reviews are given by e.g. Chib, Omori, and Asai, 2009, or Asai, McAleer, and Yu, 2006. Thereby, the asymmetric concept of cross leverage is empirically relevant but relatively less explored. It is defined as a nonzero correlation between the $i^{th}$ asset return at time $t$ and the $j^{th}$ log-volatility at time $t + 1$. The current paper proposes a MSV model based on a Cholesky-type decomposition of the covariance matrix to model dynamic correlation in the observation and transition error as well as in cross leverage terms. Volatilities and covariances are modeled separately, which makes an interpretation of the leverage parameters straightforward. A latent and nonlinear nature of stochastic volatility prevents direct computation of the likelihood in general. Accordingly, simulation based techniques, among those Markov chain Monte Carlo (McMC) certainly one of the more efficient and flexible, are commonly employed for estimation.

Three main strands of MSV models may be identified in the literature. The first, generally referred to as the basic MSV model, is a direct extension of the univariate SV model by introducing correlation in the observation and transition errors and dates back to Harvey, Ruiz, and Shephard (1994), who employ (inefficient) quasi maximum likelihood (QML) for estimation. Asai and McAleer (2006) extend this class of models to incorporate leverage effects, which is empirically important at least for stock indices (see e.g. Yu, 2005, or Omori, Shephard, and Nakajima, 2007). They additionally consider size effects and use Monte Carlo likelihood (MCL) for estimation. Chan, Kohn, and Kirby (2006), and Ishihara and Omori (2012) estimate a complete specification of the covariance matrix using McMC, including cross leverage effects, but the former specify leverage in a less common, contemporaneous way. The latter further introduce a heavy tailed observation error into the model and find evidence of both in a five-dimensional time series of S&P 500 sector index returns. However, a major drawback of the basic MSV model is that, although the covariance of returns is time varying due to the volatility dynamics, the implied conditional correlation matrix is constant.
Volatility mean factor models (e.g. Pitt and Shephard, 1999c, and Aguilar and West, 2000) offer a possible remedy to the afore mentioned drawback. In this class of models, volatility comovements are determined by a combination of underlying latent factors. Thereby, factor and idiosyncratic observation errors evolve according to univariate stochastic volatility dynamics. Chib, Nardari, and Shephard (2006) generalize this approach to allow for fat tails and jumps in the observation equation. Han (2006) models factor evolution as an AR(1) process. Lopes and Carvalho (2007) introduce time varying factor loadings and Markov switching in the common factor volatilities. Philipov and Glickman (2006b) let the factor covariance matrix evolve according to an inverse Wishart autoregressive process (WAR). All preceding models are estimated by McMC. A great advantage of factor models, if specified appropriately, is their scalability in terms of factors and assets. However, as Chan et al. (2006) note, interpretation of a possible leverage effect would be difficult due to now two sources of volatility.

A further approach to generate time varying correlation is by modeling correlation (or functions of correlation) directly. Yu and Meyer (2006) model correlation in a bivariate SV model using the fisher transformation. Asai and McAleer (2009) propose two dynamic correlation MSV models involving the Wishart distribution, building on ideas of Engle (2002). An advantage of both works is that volatility and correlation dynamics are modeled separately. However, extending the former to higher dimensions is difficult due to positive definiteness constraints on the correlation matrix. A feature of the latter, instead, is its parsimonious parameterization, making it potentially capable to handle a larger number of assets. Philipov and Glickman (2006a) let the conditional covariance matrix follow an inverse Wishart distribution with autoregressive evolution. Ishihara, Omori, and Asai (2014) model dynamics in the covariance matrix of returns using the matrix exponential transformation. A feature of the latter work is the inclusion of (time invariant) cross leverage effects into the model. Tsay (2005) introduces a Cholesky decomposition to model correlation dynamics. Lopes, McCulloch, and Tsay (2013) present a remarkably large scale application of this method, fitting a multivariate series of 94 members of the S&P 100 index. In a different development, Gouriéroux, Jasak, and Sufana (2009), and Gouriéroux (2006) propose the Wishart
autoregressive process. Except the latter two, all works presented in the current paragraph apply again McMC for model estimation.

The approach taken in this work is an application of the Cholesky decomposition method, whereby the complete covariance matrix of returns and log-volatilities is decomposed. In this way one is not only able to model the covariance dynamics of returns, but also dynamics in the (cross) leverage effects and transition error. Estimation is by McMC, where the block sampler proposed by Ishihara and Omori (2012) is extended to handle time varying covariances.


C.2 The Model

Let $y_t$ denote an asset return at time $t$. The basic univariate stochastic volatility (SV) model with normal error and leverage can then be specified as follows,

\[
y_t = \exp(\alpha_t/2)\epsilon_t, \quad t = 1, \ldots, n, \tag{C.1}
\]

\[
\alpha_{t+1} = \phi\alpha_t + \eta_{t+1}, \quad t = 1, \ldots, n - 1, \tag{C.2}
\]

\[
\alpha_1 \sim \mathcal{N}(0, \sigma^2_{\text{ini}}), \tag{C.3}
\]

\[
\begin{pmatrix} \epsilon_t \\ \eta_{t+1} \end{pmatrix} \sim \mathcal{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma^2_\epsilon & \rho\sigma_\epsilon\sigma_\eta \\ \rho\sigma_\eta\sigma_\epsilon & \sigma^2_\eta \end{pmatrix}, \tag{C.4}
\]

where $\alpha_t$ is the unobserved volatility.\footnote{Strictly speaking, it is the unobserved log-variance, but as in the literature I speak about volatility throughout. This simplifies terminology and should not cause conceptual confusion as the respective measures are linked by monotonic transformation.} The expected value of the volatility evolution process, Eq. (C.2), must be set to 0 for identification.\footnote{Pitt and Shephard (1999b) show that estimating the volatility mode of SV models in the transition rather than observation equation is superior in terms of efficiency. However, a reparameterization is needed} Further, Eq. (C.1)-(C.4)
C.2. THE MODEL

constitute a state space model with measurement equation (C.1) and transition equation (C.2) (see e.g Durbin and Koopman, 2008, or Harvey, 1989, for a comprehensive treatment). Stationarity of the volatility process is imposed (|φ| < 1), and the initial value α₁ in Eq. (C.3) is assumed to follow the respective unconditional distribution, with σ²_{ini} = σ²/(1 − φ²). Parameter ρ measures the correlation between ϵₜ and ηₜ₊₁. We have volatility asymmetry if ρ ≠ 0 and specifically, if ρ < 0, we speak of the leverage effect.

The basic SV model of Eq. (C.1)-(C.4) is now extended to the multivariate case with time varying covariance matrix Σₜ:ₜ₊₁. Let yₜ = (y₁,ₜ,..., yₚ,ₜ)' denote a p dimensional vector of asset returns and αₜ = (α₁,ₜ,...,αₚ,ₜ)' the corresponding stochastic volatility. The proposed MSVdc model is then given by

\[ yₜ = Vₜ^{1/2} ϵₜ, \quad t = 1,\ldots,n, \]  \hfill (C.5)

\[ αₜ₊₁ = Φαₜ + ηₜ₊₁, \quad t = 1,\ldots,n-1, \]  \hfill (C.6)

\[ α₁ ∼ N(0, Σ_{ini}), \]

with

\[ Vₜ = \text{diag}(\exp(α₁,ₜ),\ldots,\exp(αₚ,ₜ)), \]

\[ Φ = \text{diag}(φ₁,\ldots,φₚ), \]

\[ \left( \begin{array}{c} ϵₜ \\ ηₜ₊₁ \end{array} \right) ∼ N_{p\times2} (0, Σₜ:ₜ₊₁), \quad Σₜ:ₜ₊₁ = \left( \begin{array}{cc} Σ_{εε,ₜ:ₜ₊₁} & Σ_{εη,ₜ:ₜ₊₁} \\ Σ_{ηε,ₜ:ₜ₊₁} & Σ_{ηη,ₜ:ₜ₊₁} \end{array} \right), \]  \hfill (C.7)

where operator diag(·) creates a diagonal matrix. Again, the underlying latent volatility process is assumed stationary, |φᵢ| < 1, i = 1,...,p, with Σ_{ini} the unconditional distribution, vec(Σ_{ini}) = (I_p² − Φ ⊗ Φ)^⁻¹vec(Σ_{ηη,1}), where vec(·) stacks a matrix column-wise, ⊗ denotes the Kronecker product, and I_p is the identity matrix of dimension p. To obtain Σ_{ηη,1}, it is assumed that Σ_{0:1} = Σ_{1:2}, a reasonable strategy for initialization. Observe that volatilities and covariances are modeled separately, which makes an interpretation of leverage straightforward. Note further that leaving Σₜ:ₜ₊₁ time invariant recovers the multivariate stochastic volatility model with cross leverage proposed by Ishihara and Omori (2012).

that would result in a constraint covariance matrix. Chan and Jeliazkov (2009) propose McMC techniques, but an extension to time varying covariance matrices appears rather infeasible.
CHAPTER C. MSVdc

Time varying covariance matrix $\Sigma_{t:t+1}$ is modeled by Cholesky-type decomposition following Chan and Jeliazkov (2009), which ensures positive semi-definiteness and is amenable to Bayesian estimation,

$$\Sigma_{t:t+1}^{-1} = A'_{t:t+1} H^{-1} A_{t:t+1},$$

where $A_{t:t+1}$ is a lower triangular matrix with ones on the diagonal and $H$ a diagonal matrix with positive elements,

$$A_{t:t+1} = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
a_{21,t} & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
a_{q1,t:t+1} & \cdots & a_{q,q-1,t+1} & 1
\end{pmatrix}, \quad H = \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & \lambda_q
\end{pmatrix}, \quad (C.8)$$

with $q = 2p$ the dimension of $\Sigma_{t:t+1}$. This reparameterization is in fact an orthogonal transformation, and the dynamics specified in Eq. (C.5)-(C.6) can now be formulated as

$$A_{t:t+1} z_{t:t+1} \sim \mathcal{N}(0, H), \quad z_{t:t+1} = \begin{pmatrix}
V_t^{-1/2} y_t \\
\alpha_{t+1} - \Phi \alpha_t
\end{pmatrix}. \quad (C.9)$$

Rewriting Eq. (C.9) leads to a linear Gaussian state space model as follows,

$$z_{t:t+1} = X_{t:t+1} a_{t:t+1} + H^{1/2} e_{t:t+1}, \quad e_{t:t+1} \sim \mathcal{N}(0, I_q), \quad (C.10)$$

$$X_{t:t+1} = \begin{pmatrix}
0 & \cdots & 0 \\
z_1 & 0 & 0 & \cdots \\
0 & z_1 & z_2 & \cdots \\
0 & 0 & 0 & z_1 & \cdots \\
\vdots & \ddots & 0 & \cdots & 0 \\
0 & \cdots & 0 & z_1 & \cdots & z_{q-1}
\end{pmatrix},$$

$$a_{t:t+1} = \begin{pmatrix}
a_{21} & a_{31} & \cdots & a_{q,q-1}
\end{pmatrix}'.$$

with $X_{t:t+1}$ a $q \times q(q - 1)/2$ matrix containing elements $z_i$ of $z_{t:t+1}$, and $a_{t:t+1}$ a $q(q - 1)/2 \times 1$ vector having the $a_{ij}$'s of $A_{t:t+1}$ stacked row-wise (time subscripts dropped for clarity). The recursive conditional (or triangular) regressions in Eq. (C.10)
constitute the observation equations and \( a_{t:t+1} \) the state variables. Model specification is then completed by specifying simple independent random walk dynamics for the state transitions,

\[
a_{ij,t+1} = a_{ij,t} + u_{ij,t+1}, \quad u_{ij,t+1} \sim \mathcal{N}(0, \sigma_{ij}^2),
\]

with \( a_{ij,ini} \sim \mathcal{N}(0, 100^2), i \leq p \), and \( a_{ij,ini} \sim \mathcal{N}(0, 1), i > p \), set.\(^3\) The initial variances can be regarded as diffuse considering empirical relevant values for \( a_{t:t+1} \).

In the current application, the elements of \( A_{t:t+1} \) corresponding to dynamics of transition error \( \eta_{t+1} \) are kept constant over time, introducing some parsimony. More formally, \( a_{ij} \) fix for \( i, j > p \). Such a stepwise approach to complexity is a sensible strategy, regarding that these dynamics are completely second order latent and inference can be expected difficult. Note however that \( \Sigma_{\eta \eta,t+1} \) continues to be time varying due to the specific structure of the Cholesky decomposition. Similarly, we obtain time varying volatility modes for \( i = 2, \ldots, p \).

### C.3 McMC Algorithm

Define parameters \( \theta \equiv \{ \phi, \lambda, \sigma \} \), with \( \phi = (\phi_1, \ldots, \phi_p)' \), \( \lambda = (\lambda_1, \ldots, \lambda_q)' \), and \( q(q-1)/2 \times 1 \) vector \( \sigma = (\sigma_{21}, \ldots, \sigma_{q,q-1})' \). Further let \( y_{1:n} = (y_1, \ldots, y_n)' \), \( \alpha_{1:n} = (\alpha_1, \ldots, \alpha_n)' \), and \( a_{1:n} = (a_1, \ldots, a_n)' \), the latter a \( n \times q(q-1)/2 \) matrix containing the elements of \( a_{t:t+1} \) in Eq. (C.11), \( t = 1, \ldots, n \), stacked appropriately. Denote by \( \theta_{-r} \) all parameters in \( \theta \) without \( r \). Prior distributions are assumed to be independent and generically denoted by \( \pi(\cdot) \), with

\[
\lambda_i \sim IG(v_{0,i}/2, \delta_{0,i}/2), \quad i = 1, \ldots, q,
\]

\[
\sigma_{ij}^2 \sim IG(b_{0,ij}/2, c_{0,ij}/2), \quad i = 2, \ldots, q, \quad j < i,
\]

where \( IG(\cdot) \) is the inverse gamma distribution. In the case of \( a_{ij} \) fix,

\[
a_{ij} \sim \mathcal{N}(a_{0,ij}, \sigma_{a0,ij}^2).
\]

\(^3\)One may consider such an approach to initialization ad hoc and decide to implement exact state filtering and smoothing algorithms as e.g. proposed by de Jong (1991a), and outlined in Durbin and Koopman (2008).
The McMC algorithm then consists of the following steps:\footnote{Implementation (also the particle filter in Sec. C.4) is in stand alone C++ code developed by the author using the freely available Scythe statistical library (Pemstein, Quinn, and Martin, 2011).}

1. Initialize \( \theta, \alpha_{1:n}, a_{1:n} \).
2. Generate \( \alpha_{1:n} \mid \theta, a_{1:n}, y_{1:n} \).
3. Generate \( a_{1:n} \mid \theta, \alpha_{1:n}, y_{1:n} \).
4. Generate \( \lambda \mid \theta - \lambda, \alpha_{1:n}, a_{1:n}, y_{1:n} \).
5. Generate \( \sigma \mid \theta - \sigma, \alpha_{1:n}, a_{1:n}, y_{1:n} \).
6. Generate \( \phi \mid \theta - \phi, \alpha_{1:n}, a_{1:n}, y_{1:n} \).
7. Go to 2.

In the sequel each updating step of the McMC algorithm is presented in detail.

**Generation of latent volatilities \( \alpha_{1:n} \)**

**Step 2.** The multi-move or block sampler proposed by Shepard and Pitt (1997) and modified by Watanabe and Omori (2004), which samples from the true conditional distribution, is applied to extract latent volatility \( \alpha_{1:n} \). Ishihara and Omori (2012) propose a multi-move sampler for the efficient estimation of a multivariate stochastic volatility model with cross leverage. Their sampler is extended to handle time varying covariances. It is detailed in App. C.A.

**Generation of Cholesky decomposition matrices \( A_{1:t+1} \) and \( H \)**

Due to the initial condition, a final Metropolis-Hastings (MH) step (see e.g. Chib and Greenberg, 1995) must be added to sample \( A_{1:2} \) and \( H \), except for \( a_{ij,1}, i \leq p \). The latter do not appear in the calculation of \( \Sigma_{ini} \) and can be sampled directly.

**Step 3.1** Sampling of \( a_{ij,1:n}, i \leq p \), and \( a_{ij,3:n}, i > p \), is a standard application of forward filtering, backward sampling (FFBS) as Eq. (C.10) and Eq. (C.12) constitute
an ordinary multivariate linear Gaussian state space system. However, sampling the elements of $\alpha_{t:t+1}$ row-wise soon starts to slow down the sampler with an increasing number of assets $p$. An one-by-one strategy avoids computationally costly multivariate matrix operations and is applied here. Moreover, no significant deterioration of sampler performance appeared using the latter approach.

**Step 3.2** Proposals for $a_{ij,1:2}$ time varying, $i > p$, form $a_{ij,1:2}^* \sim N(\mu_{a_{ij}},\sigma_{a_{ij}}^2)$, with

$$
\mu_{a_{ij}} = \sigma_{a_{ij}}^2 \lambda_i^{-1} z_{j,1:2}(z_{i,1:2} - X_{i\setminus j,1:2} a_{\setminus ij,1:2}) + \sigma_{a_{ij}}^{-2} a_{ij,2:3},
$$

$$
\sigma_{a_{ij}}^{-2} = \lambda_i^{-1} z_{j,1:2}^2 + \sigma_{ij,ini}^{-2} + \sigma_{ij}^{-2},
$$

where $z_{j,1:2}$ denotes the $j$th element of $z_{1:2}$, $X_{i\setminus j,1:2}$ the $i$th row of $X_{1:2}$ without $z_{j,1:2}$, and $a_{\setminus ij,1:2}$ denotes $a_{1:2}$ without $a_{ij,1:2}$. Further, current proposals are used in the sequential calculations.

**Step 3.3** Proposals for $a_{ij}$ time invariant, $i > p$, is a routine application of linear regression theory. The conditional posterior of $a_{ij}$ is

$$
\pi(a_{ij} | \cdot) \propto \pi(a_{ij}) \times \exp\left\{ -\frac{1}{2\lambda_i} (Y_{a_{ij}} - X_{a_{ij}} a_{ij})' (Y_{a_{ij}} - X_{a_{ij}} a_{ij}) \right\},
$$

with

$$
X_{a_{ij}} = (z_{j,1:2} \cdots z_{j,n-1:n})',
$$

$$
Y_{a_{ij}} = (z_{i,1:2} - X_{i\setminus j,1:2} a_{\setminus ij,1:2} \cdots z_{i,n-1:n} - X_{i\setminus j,n-1:n} a_{\setminus ij,n-1:n} )'.
$$

Hence $a_{ij}^* | \cdot \sim N(\mu_{a_{ij}},\sigma_{a_{ij}}^2)$, with

$$
\mu_{a_{ij}} = \sigma_{a_{ij}}^2 (\sigma_{a_{0,ij}}^{-2} a_{0,ij} + \lambda_i^{-1} X_{a_{ij}}' Y_{a_{ij}}), \quad \sigma_{a_{ij}}^{-2} = \sigma_{a_{0,ij}}^{-2} + \lambda_i^{-1} X_{a_{ij}}' X_{a_{ij}}.
$$

---

**Step 4.** To sample $\lambda^*$, let $w_{t:t+1} = A_{t:t+1} z_{t:t+1}$ and $w = \{w_{1:2}, \ldots, w_{n-1:n}, w_n\}$. Further observe that $|\Sigma_{t:t+1}^{-1}| = |A_{t:t+1}'||H^{-1}||A_{t:t+1}| = |H|^{-1} = \prod_{i=1}^q \lambda_i^{-1}$. The conditional likelihood without initial condition can then be expressed as

$$
 f(w|\cdot) \propto \left(\prod_{t=1}^{n} |\Sigma_{t:t+1}|^{-1/2}\right) \exp\left\{-\frac{1}{2} \sum_{t=1}^{n-1} z_{t:t+1}' \Sigma_{t:t+1}^{-1} z_{t:t+1}\right\} \times \\
 |\Sigma_n|^{-1/2} \exp\left\{-\frac{1}{2} z_n' \Sigma_n^{-1} z_n\right\} \\
 \propto \left(\prod_{i=1}^{q} \lambda_i^{-n_i/2}\right) \exp\left\{-\frac{1}{2} \left[\sum_{t=1}^{n-1} w_{t:t+1}' H^{-1} w_{t:t+1} + w_n' H_n^{-1} w_n\right]\right\},
$$

where $n_i = n$ if $1 \leq i \leq p$ else $n_i = n-1$, and $H_n = \text{diag}(\lambda_1, \ldots, \lambda_p)$. Rewriting the above expression in terms of $\lambda_i$, $i = 1, \ldots, q$ yields

$$
 f(w|\cdot) \propto \left(\prod_{i=1}^{q} \lambda_i^{-n_i/2}\right) \exp\left\{-\frac{1}{2} \left[\text{tr}(H^{-1} \sum_{t=1}^{n-1} w_{t:t+1} w_{t:t+1}')\right]\right\} + \\
 \text{tr}(H_n^{-1} w_n w_n') \right\} = \left[\prod_{i=1}^{q} \lambda_i^{-n_i/2}\right] \exp\left\{-\frac{r_i + s_i}{2\lambda_i}\right\},
$$

where $r_i$ denotes the $(i,i)$-element of $\sum_{t=1}^{n-1} w_{t:t+1} w_{t:t+1}'$, and $s_i$ the $(i,i)$-element of $w_n w_n'$ if $1 \leq i \leq p$ else $s_i = 0$. Hence, the conditional posterior of $\lambda$ is

$$
 \pi(\lambda|\cdot) \propto \pi(\lambda) \times |\Sigma_{ini}|^{-1/2} \exp\left\{-\frac{1}{2} \alpha_1' \Sigma_{ini}^{-1} \alpha_1\right\} \times \left[\prod_{i=1}^{q} \lambda_i^{-n_i/2}\right] \exp\left\{-\frac{r_i + s_i}{2\lambda_i}\right\}.
$$

Under the inverse gamma prior of Eq. (C.13), proposals $\lambda_i^*$ are sampled independently,

$$
 \lambda_i^*|\cdot \sim IG\left(\frac{\nu_0, i + n_i}{2}, \frac{\delta_0, i + r_i + s_i}{2}\right), \quad i = 1, \ldots, q.
$$

**Step 5.** A final MH step must be applied. Proposals are $\{(a_{ij,1:2}^* \lor a_{ij}^*)|i > p, \lambda^*\}$. These are accepted with probability

$$
 \min \left\{ \frac{|\Sigma_{ini}|^{-1/2} \exp\left\{-\frac{1}{2} \alpha_1' \Sigma_{ini}^{-1} \alpha_1\right\}}{|\Sigma_{ini}|^{-1/2} \exp\left\{-\frac{1}{2} \alpha_1' \Sigma_{ini}^{-1} \alpha_1\right\}, 1}\right\},
$$

where $\Sigma_{ini}$ is the current unconditional variance of $\alpha_1$. 
C.3. McMC ALGORITHM

**Generation of $\sigma^2$ governing the random walk dynamics of $a_{1:n}$**

**Step 6.** Random walk variances $\sigma_{ij}^2 \in \sigma^2$ are independent and sampled one-by-one using standard results of linear regression theory. The conditional posterior probability of $\sigma_{ij}^2$ is

$$
\pi(\sigma_{ij}^2 | \cdot) \propto \pi(\sigma_{ij}^2) \times (\sigma_{ij}^2)^{-n_i/2} \exp\left\{-\sum_{t=k_i}^{n-1} \frac{(a_{ij,t+1} - a_{ij,t})^2}{2\sigma_{ij}^2}\right\}
$$

$$
= (\sigma_{ij}^2)^{-b_{ij}/2-1} \exp\left\{-\frac{c_{ij}}{2\sigma_{ij}^2}\right\},
$$

with

$$
b_{ij} = b_{0,ij} + n_i, \quad c_{ij} = c_{0,ij} + \sum_{t=k_i}^{n-1} (a_{ij,t+1} - a_{ij,t})^2,
$$

where $n_i = n - 1 \land k_i = 1$ if $1 \leq i \leq p$ else $n_i = n - 2 \land k_i = 2$, and $\sigma_{ij,1} = \sigma_{ij,ini}$ known. Hence, $\sigma_{ij}^2 | \cdot \sim IG(b_{ij}/2, c_{ij}/2)$.

**Generation of volatility persistence $\Phi$**

**Step 7.** The conditional posterior probability of $\Phi$ does not have closed form, and sampling is by the MH algorithm. Let $\Sigma_{i:j+1}^{ij}$ be a $p \times p$ matrix that denotes the $(i,j)$-block of $\Sigma_{i:j+1}^{-1}$. The posterior of $\Phi$ can then be written as

$$
\pi(\Phi | \cdot) \propto g(\Phi) \times \exp\left\{-\frac{1}{2} (\Phi - \mu_\Phi)' \Sigma_\Phi^{-1} (\Phi - \mu_\Phi)\right\},
$$

with

$$
\mu_\Phi = \Sigma_\Phi \times \text{diag} \left(\sum_{t=1}^{n-1} \left\{\alpha_t y_i' V_t^{-1/2} \Sigma_{t:t+1}^{22} + \alpha_t' \alpha_{t+1}' \Sigma_{t:t+1}^{22}\right\}\right),
$$

$$
\Sigma_\Phi^{-1} = \sum_{t=1}^{n-1} \Sigma_{t:t+1}^{22} \odot \alpha_t \alpha_t',
$$

$$
g(\Phi) = \pi(\Phi) \times |\Sigma_{ini}|^{-1/2} \exp\left\{-\frac{1}{2} \alpha_1' \Sigma_{ini}^{-1} \alpha_1\right\},
$$

where $\text{diag}(\cdot)$ returns the diagonal as column vector, and $\odot$ denotes the Hadamard product. Then propose a candidate from the multivariate normal distribution truncated over the region $R = \{\Phi : |\phi_i| < 1, i = 1, \ldots, p\}$, $\Phi^* \sim T N_R(\mu_\Phi, \Sigma_\Phi)$, and accept it with probability $\min \{g(\Phi^*)/g(\Phi), 1\}$, where $\Phi$ is the current sample.
C.4 Particle Filter

An auxiliary particle filter (Pitt and Shephard, 1999a) is developed that recursively delivers draws of the unobservable volatilities $\alpha_t$, given parameter values $\theta$ and observables up to time $t$, $y_{1:t}$. These can be utilized for likelihood evaluation, goodness-of-fit statistics, and forecasting. The auxiliary particle filter is a sequential Monte Carlo (SMC) technique that increases efficiency in the propagation process by weighting particles according to an importance function dependent on the subsequent observation. A key characteristic of the proposed filter is that, given draws of latent volatilities $\alpha_t$, latent covariance dynamics $\alpha_t$ can be extracted in analytical form using the standard Kalman Filter prediction and updating equations. For more information on particle filtering, consider e.g. Doucet, de Freitas, and Gordon (2001), or Whiteley and Johansen (2011).

Observe that the model as defined in Eq. (C.5)-(C.7) can be reformulated having measurement density

$$f(y_t | \alpha_t) = \mathcal{N}(0, V_t^{1/2} \Sigma_{\epsilon \epsilon, t} V_t^{1/2})$$

and transition density

$$f(\alpha_{t+1} | \Sigma_{t:t+1}, \alpha_t, y_t) = \mathcal{N}(\mu_{\alpha, t+1}, \Sigma_{\alpha, t+1})$$

(C.14)

where

$$\mu_{\alpha, t+1} = \Phi \alpha_t + \Sigma_{\eta \epsilon, t:t+1} \Sigma_{\epsilon \epsilon, t}^{-1} V_t^{-1/2} y_t,$$

$$\Sigma_{\alpha, t+1} = \Sigma_{\eta \eta, t+1} - \Sigma_{\eta \epsilon, t:t+1} \Sigma_{\epsilon \epsilon, t}^{-1} \Sigma_{\epsilon \eta, t:t+1}$$

(for clarity, time invariant parameter vector $\theta$ may be dropped from conditioning arguments in this section). Applying Bayes Theorem we obtain the target posterior density,

$$f(\alpha_{t+1}, \alpha_t | y_{1:t+1}) \propto f(y_{t+1} | \alpha_{t+1}) f(\alpha_{t+1} | \alpha_t, y_t) f(\alpha_t | y_{1:t}),$$

(C.15)
where we assume that we have samples (particles) from $f(\alpha_t|y_{1:t})$ and a discrete uniform approximation $\hat{f}(\alpha_t|y_{1:t})$ to $f(\alpha_t|y_{1:t})$. Latent covariance dynamics $\alpha_{t:t+1}$ appear in linear Gaussian state space form and have been integrated out of Eq. (C.15) analytically.

More precisely, measurement equation (C.10) is treated row-wise, having density

$$f(z_{i,t+1}|a_{i,t+1}) = \mathcal{N}(X'_{i,t+1}a_{i,t+1}, \lambda_i), \quad i = 2, \ldots, q,$$

with $a_{i,t+1}$ a $i - 1 \times 1$ column vector containing the $i^{th}$ row of the lower triangular of matrix $A_{t+1}$, and $X'_{i,t+1}$ a column vector of same dimension stacking all nonzero elements of the $i^{th}$ row of $X_{t+1}$. The corresponding sequential learning process has priors (prediction)

$$\hat{a}_{i,t+1} = a_{i,t-1}, \quad \hat{P}_{i,t+1} = P_{i,t-1} + Q_i,$$

with $P_{i,t-1}$ the error covariance of $E(a_{i,t-1}|z_{i,t-1})$, and diagonal process error covariance matrix $Q = \text{diag}(\sigma^2_{i1}, \ldots, \sigma^2_{i,i-1})$. Posteriors (correction) are

$$a_{i,t+1} = \hat{a}_{i,t+1} + K_{i,t+1}(z_{i,t+1} - X'_{i,t+1}\hat{a}_{i,t+1}), \quad P_{i,t+1} = (I - K_{i,t+1}X'_{i,t+1})\hat{P}_{i,t+1},$$

where $K_{i,t+1}$ is the Kalman gain,

$$K_{i,t+1} = \hat{P}_{i,t+1}X'_{i,t+1}(X'_{i,t+1}\hat{P}_{i,t+1}X_{i,t+1} + \lambda_i)^{-1}.$$

Note that Eq. (C.16), (C.17) and (C.18), (C.19) are optimal in the mean square error sense. Further, if elements of $a_{i,t+1}$ are kept constant over time, the Kalman filter operates on the associated lower dimensional system only.

By including $y_{t+1}$ in the importance probability density function more weight is given to particles with stronger predictive power. The importance function that is used
to sample from Eq. (C.15) then has the following form (superscripts \( k = 1, \ldots, K \) denote particles),

\[
g(\alpha_{t+1}, \alpha^{(k)}_t | y_{1:t+1}) \propto f(y_{t+1} | \tilde{\mu}^{(k)}_{\alpha_{t+1}}, \tilde{a}^{(k)}_{t+1}) \times f(\alpha_{t+1} | \alpha^{(k)}_t, y_t) \tilde{f}(\alpha^{(k)}_t | y_{1:t}) \tag{C.20}
\]

with

\[
\tilde{V}^{(k)}_{t+1} = V_{t+1} | _{\alpha_{t+1} = \tilde{\mu}^{(k)}_{\alpha_{t+1}}},
\]

\[
\Sigma^{(k)}_{ee, t+1} = \Sigma_{ee, t+1} | _{a_{i,t+1} = \tilde{a}^{(k)}_{i,t+1}}, i = 2, \ldots, p,
\]

and

\[
\tilde{\mu}^{(k)}_{\alpha_{t+1}} = \Phi \alpha^{(k)}_t + \sum_{\eta \in \epsilon, : : + 1} \Sigma^{(k)}_{\eta \in \epsilon, t} V^{(k)}_{t+1} \frac{1}{2} y_t,
\]

\[
\Sigma^{(k)}_{\alpha_{t+1}} = \Sigma^{(k)}_{\eta \in \epsilon, t+1} - \sum_{\eta \in \epsilon, t+1} \Sigma^{(k)}_{\eta \in \epsilon, t} \Sigma^{(k)}_{ee, t},
\]

where

\[
V^{(k)}_t = V_t | _{\alpha_t = \alpha^{(k)}_t},
\]

\[
\Sigma^{(k)}_{ee, t} = \Sigma_{ee, t} | _{a_{i,t} = \tilde{a}^{(k)}_{i,t}}, i = 2, \ldots, p,
\]

\[
\tilde{\Sigma}^{(k)}_{\epsilon \eta, t+1} = \Sigma_{\epsilon \eta, t+1} | _{a_{i,t} = \tilde{a}^{(k)}_{i,t}}, i = 2, \ldots, p, a_{i,1:t+1} = \tilde{a}^{(k)}_{i,1:t+1}, i = p+1, \ldots, q,
\]

\[
\tilde{\Sigma}^{(k)}_{\epsilon \eta, t+1}, \tilde{\Sigma}^{(k)}_{\eta \epsilon, t+1} \text{ in analogy to } \Sigma^{(k)}_{\epsilon \eta, t+1}, \text{ and}
\]

\[
\tilde{a}^{(k)}_{i, t+1} = a^{(k)}_{i, t}, \quad a^{(k)}_{i, t} = a_{i,t} | _{\alpha_t = \alpha^{(k)}_t}, \tilde{a}^{(k)}_{i, t}, i = 2, \ldots, p,
\]

\[
\tilde{a}^{(k)}_{i, : : + 1} = a^{(k)}_{i, : : + 1}, \quad a^{(k)}_{i, : : + 1} = a_{i, t+1} | _{\alpha_t = \alpha^{(k)}_t}, \tilde{a}^{(k)}_{i, t+1}, i = p + 1, \ldots, q.
\]

Note that in Eq. (C.20) and in the sequel, as there is a one-to-one relationship between \( \alpha^{(k)}_t \) and \( \tilde{a}^{(k)}_{i, t} \), \( i = 2, \ldots, q \), the latter may be omitted in the arguments to not clutter notation. Moreover, note that as the initial states of the random walk dynamics \( a_t \) are centered at zero, an appropriate convergence period for the particle filter must be chosen.
C.4. PARTICLE FILTER

The proposed implementation follows:

1. Initialization, \( t = 1 \) (\( k = 1, \ldots, K \)):

   (a) Calculate \( \Sigma_{0:1} \) by setting random walk elements \( a_{ij:0:1} = 0, j \leq p \), and compute corresponding \( \Sigma_{ini} \). Then generate \( \alpha_{1}^{(k)} \sim \mathcal{N}(0, \Sigma_{ini}) \).

   (b) Initialize \( \bar{\alpha}_{i,1}^{(k)} = 0 \), \( \bar{P}_{i,1} = \text{diag}(I_{i-1}) \), and update
       \[
       \alpha_{i,1}^{(k)} = \bar{\alpha}_{i,1}^{(k)} \mid \alpha_{1}^{(k)}, \bar{a}_{i,1}^{(k)}, i = 2, \ldots, p.
       \]

   (c) Compute \( w_{1}^{(k)} = f(y_{1} \mid \alpha_{1}^{(k)}) \), and let \( \hat{f}(\alpha_{1}^{(k)} \mid y_{1}) = w_{1}^{(k)} / \sum_{l=1}^{K} w_{1}^{(l)} \).

   (d) Initialize \( \bar{\alpha}_{i,1:2}^{(k)} = 0 \), \( \bar{P}_{i,1:2} = \text{diag}(0.1^{2} I_{i-1}) \), \( i = p + 1, \ldots, q \).

2. Iterate, \( t = 1, \ldots, n - 1 \) (\( k, l = 1, \ldots, K \)):

   (a) Generate \( \{ \alpha_{t+1}^{(k)}, \alpha_{t}^{(k)} \} \) from \( g(\alpha_{t+1}^{(k)}, \alpha_{t}^{(l)} \mid y_{1:t+1}) \), with \( \alpha_{t}^{(l)} \) the current particle set at time \( t \), as follows. First, evaluate importance weights (pre-weighting)
       \[
       w_{t}^{(l)} \propto f(y_{t+1} \mid \bar{\alpha}_{t+1}^{(l)}, \bar{\alpha}_{t}^{(l)}, \bar{a}_{t+1}^{(l)}, \bar{a}_{t}^{(l)}) \hat{f}(\alpha_{t}^{(l)} \mid y_{1:t}),
       \]
       and resample \( \{ \alpha_{t}^{(l)}, w_{t}^{(l)} \} \) to get \( \{ \alpha_{t}^{(k)}, K^{-1} \} \). Then generate \( \alpha_{t+1}^{(k)} \) from \( f(\alpha_{t+1}^{(k)} \mid \alpha_{t}^{(k)}, y_{t}) \) given in Eq. (C.14), and update
       \[
       \alpha_{i,t+1}^{(k)} = \alpha_{i,t+1}^{(k)} \mid \alpha_{t+1}^{(k)}, \bar{a}_{i,t+1}^{(k)}, i = 2, \ldots, p,
       \]
       \[
       \alpha_{i,t+1}^{(k)} = \alpha_{i,t+1}^{(k)} \mid \alpha_{t+1}^{(k)}, \bar{a}_{i,t+1}^{(k)}, i = p + 1, \ldots, q.
       \]

   (b) Compute weights (correcting for the pre-weighting)
       \[
       w_{t+1}^{(k)} = \frac{f(y_{t+1} \mid \alpha_{t+1}^{(k)}, \alpha_{t+1}^{(k)}, \alpha_{t+1}^{(k)})}{f(y_{t+1} \mid \bar{\mu}_{t+1}^{(k)}, \bar{a}_{t+1}^{(k)}, \bar{a}_{t+1}^{(k)})},
       \]
       and let \( \hat{f}(\alpha_{t+1}^{(k)} \mid y_{t+1}) = w_{t+1}^{(k)} / \sum_{l=1}^{K} w_{t+1}^{(l)} \).

Draws \( \alpha_{t+1}^{(k)} \), which are utilized to calculate posterior log-likelihood ordinate
\[
\log f_{\text{post}}(y_{1:n} \mid \theta) = \sum_{t=1}^{n} \log \hat{f}(y_{t} \mid y_{1:t}, \theta),
\]
are obtained by resampling from \( \hat{f}(\alpha_{t}^{(k)} \mid y_{t}) \) after step 1.(c) and 2.(b). These particles are propagated forward in time to obtain \( \alpha_{t+1}^{(k)} \), by first calculating priors as given by Eq. (C.16) and (C.17), and then
sampling from Eq. (C.14), to obtain an estimate of the one-step-ahead prediction density \( f(y_t|y_{1:t-1}, \theta) \), which in turn is needed for forecasting and calculation of predictive log-likelihood ordinate \( \log f_{\text{prior}}(y_{1:n}|\theta) = \sum_{t=1}^{n} \log \hat{f}(y_t|y_{1:t-1}, \theta) \).

**C.5 Empirical Application**

**C.5.1 Data**

Descriptive statistics of the dataset are presented. In Sec. C.5.2 and C.5.3 estimation results of the MSV model with constant correlation and dynamic correlation are discussed, respectively. The dataset consists of three S&P 500 sector indices, namely Financials (series 1), Industrials (series 2) and Healthcare (series 3), obtained from Thomson Reuters Datastream. The choice was motivated by Industrials and Financials being main pillars of the economy, supposed to exhibit strong interdependence, and Healthcare having distinct characteristics due to being countercyclical. The period analyzed covers 8 years from 2004/10/01 to 2012/09/28, a total of 2,015 observations. Market holidays were excluded. Returns are calculated as \((P_{t+1}/P_t - 1) \times 100\), where \(P_t\) denotes the asset price at time \(t\). Further, the data is demeaned and standardized by its sample standard deviation. Working on a common scale facilitates prior choice, especially for the dynamic correlation model using Cholesky decomposition, however it is not necessary.

Descriptive statistics are given in Tab. C.1, including the S&P 500 index as reference. Financials clearly is the most volatile sector, while Healthcare is the least volatile. Notably, the largest return in absolute value is positive for all sector indices.
considered and the S&P 500, despite the period under analysis containing the financial crisis 2008/09. A quite strong observed positive skew for Financials and less so for Healthcare complements the picture. As usual, all time series show kurtosis larger than 3, with Financials and Healthcare having the heaviest tails, the latter influenced largely by one extraordinary positive outlier 2008/10/13. Excluding this outlier would actually turn the sample skewness of Healthcare the most negative. Fig. C.1 visualizes distinct volatility characteristics of the series. Comovement of volatility among the sector indices is clearly visible and most pronounced during the 2008/09 financial crisis with significantly increased volatility in all three sectors. Further, volatility surges following the rating downgrade of U.S. Treasury bonds from their AAA status by Standard & Poor’s in early August 2011. The latter extended period of financial turmoil may be explained with deepening global macro dislocations like a continuing European
debt crisis and the prospect of a slowing global economy. Note also that Financials did not recover to previous pricing levels, in contrast to Industrials and Healthcare.

### C.5.2 The MSVcc Model (Constant Correlation)

This section reports estimation results of the MSV model with constant correlation and cross leverage proposed by Ishihara and Omori (2012) (henceforth MSVcc), which serves as a baseline for the more complex MSVdc model with dynamic correlation and cross leverage. The afore mentioned authors apply the model to a five-dimensional portfolio of S&P 500 sector indices covering a slightly different time span, but estimation results may be compared with those obtained here. The MSVcc model is recovered from the MSVdc model as defined in Eq. (C.5)-(C.12) by replacing time varying covariance matrix $\Sigma_{t:t+1}$ in Eq. (C.7) with time invariant

$$
\Sigma_{\text{MSVcc}} = \begin{pmatrix}
\Sigma_{ee} & \Sigma_{e\eta} \\
\Sigma_{\eta e} & \Sigma_{\eta\eta}
\end{pmatrix}.
$$

Accordingly, the proposed McMC algorithm of Sec. C.3 simplifies. Specifically, Steps 3 to 5 are replaced by an appropriate sampling step as outlined in Sec. C.C of the Appendix. Then, the following priors are assumed,

$$
\frac{\phi_i + 1}{2} \sim B(20, 1.5), \quad i = 1, \ldots, p,
$$

$$
\Sigma^{-1} \sim W(10, (10\Sigma_0)^{-1}), \quad \Sigma_0 = \begin{pmatrix} 0.4^2 I_p & 0 \\ 0 & 0.15^2 I_p \end{pmatrix},
$$

where $B(\cdot)$ denotes the beta and $W(\cdot)$ the Wishart distribution. The prior on $\phi_i$ implies a belief in moderate persistence with mean 0.86 and standard deviation 0.11. The prior on $\Sigma^{-1}$ conservatively assumes zero correlation and is chosen such that $E(\Sigma^{-1}) = \Sigma_0^{-1}$. Beliefs in volatility of volatility of the latent processes, expressed by the lower block diagonal of $\Sigma_0$, are rather standard. However, the specific prior choice for the volatility modes in the upper diagonal block matrix of $\Sigma_0$ demands some explanation. A unit prior for the standardized data would be an obvious choice, but did not yield satisfactory estimation results for the specific dataset at hand.
Concretely, a unit prior led to overestimation of the volatility modes, offset by latent volatility processes with unconditional posterior sample averages considerably below their model theoretic value of zero. Hence the choice of a smaller prior to pull the sample averages of $\alpha_{1:n}$ closer to zero, where "close" is of course subject to the researcher’s opinion, say a threshold of approximately $10^{-1}$. One may choose a larger sample size to alleviate such model fitting problems and associated prior tuning, but this would increase the computational burden. Moreover, regarding forecasting, one may feel more comfortable working with a dataset closer to the actual state. The proposed approach is coherent and systematic using pilot runs.

The first 10,000 draws are discarded as burn-in, collecting the next 40,000 draws for parameter inference. In the multi-move sampler, average block size as a tuning parameter is set to 20 and number of iterations to achieve convergence to 5. Posterior means, standard deviations, 95% credible bounds, p-values of Geweke’s (1992) convergence diagnostic ($CD$), and inefficiency factors ($IF$) (see App. A.C for the deployed formulae) are reported. For all models estimated in this work, the null of the $CD$ statistic ("chain has converged") can not be rejected at conventional significance levels for all parameters.

Turning to results in Tab. C.2, persistence $\phi_i$ is high for all processes and highest for Financials. Higher volatility persistence for Financials is in accordance with this sector being the most volatile.\textsuperscript{6} Interestingly, volatility mode $\sigma_{\epsilon,1}$ of Financials features the lowest value, but the data is standardized. Moreover, the high persistence of Financials’ latent volatility makes it more difficult to extract the volatility mode, as mirrored by a higher inefficiency factor relative to Industrials and Healthcare.

Ephemeral observation error $\epsilon_t$ and transition error $\eta_{t+1}$ both feature high positive correlations $\rho_{\epsilon\epsilon}$ and $\rho_{\eta\eta}$, respectively. Leverage and cross leverage effects are significant and moderate in size. Note that cross leverage effects to Healthcare, the least volatile asset, are highest, albeit not significantly. Ishihara and Omori (2012) further estimate univariate versions of the model and report stronger leverage $\rho_{\epsilon\eta,i}$.

\textsuperscript{6}Jacquier and Miller (2012) estimate latent volatility of the S&P 500 index around the financial crisis using a sequential Monte Carlo (SMC) approach. They make volatility persistence parameter $\phi$ time varying and find increased persistence during the turmoil, high volatility period.
arguing that in case of the multivariate model, part of the variation of one series may be explained by those of other series through high correlations among $\epsilon_t$ and $\eta_{t+1}$.

### C.5.3 The MSVdc Model (Dynamic Correlation)

This section fits the MSVdc model, featuring dynamic correlation and cross leverage. Database remains the same, namely the S&P 500 sector indices Financials (series 1),
C.5. EMPIRICAL APPLICATION

Industrials (series 2) and Healthcare (series 3). With 3 assets there are 3 latent volatility processes \( \alpha_t \) and 12 latent processes governing the covariance dynamics of \( \Sigma_{t:t+1} \) (the elements of \( a_{t:t+1} \) modeling the covariance dynamics of \( \Sigma_{\eta\eta,t+1} \) are kept constant for parsimony). With such many latent parameters, prior selection inevitably becomes important in the model fitting procedure. Through a careful choice of latent process variances \( \sigma^2_{ij} \), Eq. (C.12), we are able to control the smoothness of \( a_{t:t+1} \) and consequently, that of \( \Sigma_{t:t+1} \). If we choose \( \Sigma_{t:t+1} \) too smooth, we risk to ignore potentially important dynamics. On the other hand, a \( \Sigma_{t:t+1} \) that is too adaptive may chase spurious noise in the data. Moreover, too diffuse dynamics for \( \Sigma_{\eta\eta,t+1} \) have the undesirable side effect of decreasing persistence \( \Phi \) of latent volatility, overfitting the model further. Such a sensibility to prior choice may be perceived as a disadvantage. However, with a careful and systematic tuning approach, these interdependencies allow the researcher to improve predictive performance.

The proposed approach to prior selection is explained in more detail. First, priors on the diagonal elements of \( \Phi \) are as in Sec. C.5.2. For remaining parameters, the following priors are assumed,

\[
\begin{align*}
\lambda_i &\sim \text{IG}(5/2, 3/2), \quad i = 1, \ldots, p, \quad \lambda_i &\sim \text{IG}(5/2, 0.05/2), \quad i = p + 1, \ldots, q, \\
\sigma^2_{ij} &\sim \text{IG}(5/2, 0.005/2), \quad i \leq p, \quad \sigma^2_{ij} &\sim \text{IG}(5/2, 0.000001/2), \quad i > p,
\end{align*}
\]

and for \( a_{ij} \) fix \( (i > p, j > p) \)

\[
a_{ij} \sim \mathcal{N}(0, 10^{12}).
\]

The priors for \( \lambda_i, i = 1, \ldots, p \), have mean 1.00 and standard deviation 1.41, making them a direct choice for the standardized data as they imply unit variance for the observation error in the case of no correlation. Those for transition error related \( \lambda_i, i = p + 1, \ldots, q \), imply volatilities with mean 0.12 and standard deviation 0.05 in the case of zero correlation, and are rather standard in the literature (moments obtained by MC simulation). Time invariant elements \( a_{ij} \) of the lower triangular matrix \( A_{t:t+1} \) receive noninformative priors, centered around zero.

---

Note that ordering of the assets matters when deploying Cholesky-type decomposition and is according to unconditional sample volatility in decreasing order. Reverse ordering has been tested, but yielded a worse in-sample fit according to the deviance information criterion (DIC).
### Table C.3: Estimation Results of the MSVdc Model (dynamic correlation)

<table>
<thead>
<tr>
<th>Series $i/i_j$</th>
<th>Mean</th>
<th>Stdev</th>
<th>95% c.b.</th>
<th>IF</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.9810</td>
<td>0.0036</td>
<td>[0.9735, 0.9877]</td>
<td>22.6</td>
<td>0.89</td>
</tr>
<tr>
<td>2</td>
<td>0.9728</td>
<td>0.0050</td>
<td>[0.9622, 0.9818]</td>
<td>32.2</td>
<td>0.94</td>
</tr>
<tr>
<td>3</td>
<td>0.9625</td>
<td>0.0074</td>
<td>[0.9466, 0.9755]</td>
<td>36.2</td>
<td>0.90</td>
</tr>
<tr>
<td>$\lambda_{\epsilon,i}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.2625</td>
<td>0.0521</td>
<td>[0.1763, 0.3780]</td>
<td>244.6</td>
<td>0.91</td>
</tr>
<tr>
<td>2</td>
<td>0.1558</td>
<td>0.0182</td>
<td>[0.1234, 0.1940]</td>
<td>111.6</td>
<td>0.93</td>
</tr>
<tr>
<td>3</td>
<td>0.2097</td>
<td>0.0215</td>
<td>[0.1711, 0.2560]</td>
<td>74.2</td>
<td>1.00</td>
</tr>
<tr>
<td>$\lambda_{\eta,i}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0412</td>
<td>0.0079</td>
<td>[0.0278, 0.0586]</td>
<td>114.3</td>
<td>0.85</td>
</tr>
<tr>
<td>2</td>
<td>0.0062</td>
<td>0.0015</td>
<td>[0.0038, 0.0098]</td>
<td>167.1</td>
<td>0.96</td>
</tr>
<tr>
<td>3</td>
<td>0.0090</td>
<td>0.0024</td>
<td>[0.0052, 0.0146]</td>
<td>195.5</td>
<td>0.76</td>
</tr>
<tr>
<td>$a_{\eta\eta,ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>-0.7546</td>
<td>0.0671</td>
<td>[-0.8920, -0.6275]</td>
<td>325.8</td>
<td>0.90</td>
</tr>
<tr>
<td>31</td>
<td>0.3446</td>
<td>0.1244</td>
<td>[0.1097, 0.5912]</td>
<td>755.3</td>
<td>0.70</td>
</tr>
<tr>
<td>32</td>
<td>-1.2499</td>
<td>0.1694</td>
<td>[-1.5621, -0.9524]</td>
<td>1563.8</td>
<td>0.66</td>
</tr>
<tr>
<td>$\sigma_{\epsilon\epsilon,ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.0265</td>
<td>0.0053</td>
<td>[0.0178, 0.0383]</td>
<td>321.9</td>
<td>0.94</td>
</tr>
<tr>
<td>31</td>
<td>0.0241</td>
<td>0.0050</td>
<td>[0.0163, 0.0356]</td>
<td>340.1</td>
<td>0.62</td>
</tr>
<tr>
<td>32</td>
<td>0.0258</td>
<td>0.0048</td>
<td>[0.0180, 0.0365]</td>
<td>263.7</td>
<td>0.88</td>
</tr>
<tr>
<td>$\sigma_{\eta\epsilon,ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.000123</td>
<td>0.000063</td>
<td>[0.000030, 0.000266]</td>
<td>57.7</td>
<td>1.00</td>
</tr>
<tr>
<td>12</td>
<td>0.000570</td>
<td>0.000283</td>
<td>[0.000287, 0.001237]</td>
<td>996.3</td>
<td>0.97</td>
</tr>
<tr>
<td>13</td>
<td>0.000538</td>
<td>0.000231</td>
<td>[0.000270, 0.001179]</td>
<td>1,245.3</td>
<td>0.90</td>
</tr>
<tr>
<td>21</td>
<td>0.000521</td>
<td>0.000200</td>
<td>[0.000275, 0.001034]</td>
<td>1,217.0</td>
<td>0.86</td>
</tr>
<tr>
<td>22</td>
<td>0.000503</td>
<td>0.000176</td>
<td>[0.000277, 0.000947]</td>
<td>831.6</td>
<td>0.58</td>
</tr>
<tr>
<td>23</td>
<td>0.000545</td>
<td>0.000211</td>
<td>[0.000292, 0.001129]</td>
<td>1,109.6</td>
<td>0.91</td>
</tr>
<tr>
<td>31</td>
<td>0.000494</td>
<td>0.000170</td>
<td>[0.000273, 0.000925]</td>
<td>689.0</td>
<td>0.82</td>
</tr>
<tr>
<td>32</td>
<td>0.000506</td>
<td>0.000200</td>
<td>[0.000268, 0.001026]</td>
<td>951.7</td>
<td>0.97</td>
</tr>
<tr>
<td>33</td>
<td>0.000516</td>
<td>0.000229</td>
<td>[0.000275, 0.001108]</td>
<td>1,297.8</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke’s convergence diagnostic (p-value). $a_{\eta\eta,ij}, \sigma_{\epsilon\epsilon,ij}, \sigma_{\eta\epsilon,ij}$ relate to lower triangular of $A_{t:t+1}$.

The priors for $\sigma^2_{ij}$, $i \leq p$, imply volatilities with mean 0.038 and standard deviation 0.016. Those for $\sigma^2_{ij}$, $i > p$, a mean of 0.00053 and standard deviation of 0.00022 (all moments obtained by MC simulation). They control the smoothness of $\Sigma_{t:t+1}$ and are regarded as tuning parameters. Accordingly, we may want $\sigma^2_{ij}$ small in general. Thereby, choosing larger values for $\sigma^2_{ij}$, $i = 2, \ldots, p$, and smaller values for $\sigma^2_{ij}$, $i = p + 1, \ldots, q$, is rather intuitive considering the specific structure of the triangular regressions in Eq. (C.10). More concretely, scale parameters for the inverse gamma priors in Eq. (C.21) have been found by conducting pilot runs, selecting...
combinations over the set \{0.05, 0.01, 0.005, \ldots, 0.0001, 0.00005\} for \(\sigma^2_i, i = 2, \ldots, p\), and \{0.0005, 0.0001, 0.00005, \ldots, 0.0000005, 0.00000001\} for \(\sigma^2_i, i = p + 1, \ldots, q\). Multiple selection criteria have been applied. Specifically, a sample average of latent volatility process \(\alpha_{1:n}\) close to zero, and in-sample fit as measured by the deviance information criterion (DIC), accounting for model complexity. Further, the predictive log-likelihood (in-sample), which may be seen as an indicator for out-of-sample performance. Comparability of resulting parameter estimates with the baseline MSVcc model has also been taken into account, inspecting volatility persistence \(\Phi\) and unconditional correlation and volatility sample averages as derived from the full covariance matrix including \(\alpha_t\). The proposed approach is certainly not optimal, but systematic and only moderately time consuming. Moreover, as the data is standardized, it should be generally applicable.

The first 10,000 draws are discarded as burn-in, collecting the next 200,000 draws for parameter inference. Average block size and number of iterations to achieve convergence in the multi-move sampler are similar to Sec. C.5.2. As the parameter estimates are mostly unintuitive due to the applied Cholesky decomposition, more emphasis is put on a graphical discussion.

Tab. C.3 reports estimation results. Volatility persistence \(\Phi\) closely mirrors that of the MSVcc model. Moreover, time invariant elements \(a_{\eta \eta, ij}\) associated with \(\Sigma_{\eta \eta, t+1}\) are all significant. Note also that except for \(\sigma_{\eta \epsilon, 11}\) (leverage Financials) variances governing the random walks related to cross leverage dynamics are mainly determined by their priors. The applied Cholesky-type decomposition reparameterizes the covariance matrix as a series of regressions, modeling the joint distribution as a marginal and then a series of conditionals. Accordingly, the latter series do not appear to carry significant new information.

Importantly, the MSVdc model tends to produce lower cross leverage effects compared to the MSVcc model, which are compensated by more pronounced transition error dynamics, featuring higher standard deviations and correlations. One may choose to equate parameters with the MSVcc model rather closely through a more and more

---

8 Inaccurate priors may provoke numerical instabilities when block sampling the volatilities, especially for very extreme choices. Such problems can be alleviated using square root filtering and related techniques (de Jong, 1991b; for a more recent review, consider e.g. Grewal and Andrews, 2001). However, it is clearly preferable to avoid such problems in the first place through an adequate prior choice.
tedious prior choice. However, at some point this is considered impractical. Instead, this shift of dynamics is attributed to the general structure of the applied Cholesky-type decomposition. Above all, referring to both models, dynamics are latent and in part to a significant degree prior dependent.

Fig. C.2 plots latent volatility as calculated from the full covariance matrix including $\alpha_t$, with associated 90% credible bounds.\(^9\) For all three assets there is a pronounced increase in volatility during the financial crisis 2008/09. A more extended heavily volatile period for Financials during the crash period is visible. Interestingly, Healthcare shows the most distinct volatility spike, but volatility also tapers off fastest for this countercyclical sector. The high volatility period starting in early August 2011 and following the rating downgrade of U.S. Treasury bonds can also be clearly inferred.

\(^9\)Credible bounds have been calculated from a thinned sample of the posterior output, collecting every 100\(^{th}\) draw.
As already mentioned, the current restricted model still produces dynamic volatility of volatility, and thus time varying unconditional kurtosis (for theoretical properties of SV models, see e.g. Harvey, 1994, or Taylor, 1994). The model generates increased unconditional tail risk during periods of market turmoil, which is intuitive. Interestingly, this is most pronounced for the period of increased macroeconomic uncertainty following the rating downgrade of U.S. Treasury bonds. Again, outlier risk is most distinctive for the Healthcare sector. Moreover, tail risk is overall higher after the crisis 2008/09, which is most pronounced for Industrials and Healthcare.

Fig. C.3 plots dynamics of latent processes $a_{t:t+1}$ governing the covariance evolution of $\Sigma_{t:t+1}$ (the lower triangular of $A_{t:t+1}$ excluding $a_{ij}, i > p, j > p$, is shown). The upper two rows responsible for observation error dynamics show strong time varying patterns. However, the elements of $A_{t:t+1}$ in the lower three rows responsible for cross leverage dynamics feature some pronounced patterns as well. Observe e.g. a distinct spike of $a_{11,\eta\epsilon}$, corresponding to the leverage effect of Financials, or of $a_{32,\eta\epsilon}$ and $a_{33,\eta\epsilon}$, associated with cross leverage from Industrials to Healthcare and leverage of Healthcare, respectively, around the crisis.

Fig. C.4 and Fig. C.5 then visualize resulting correlations in $\Sigma_{t:t+1}$. Inspecting observation error correlations $\rho_{\epsilon\epsilon,t}$ in Fig. C.4 first, we infer overall stronger interdependencies between Financials and Industrials than between Healthcare and the former two. Moreover, correlation between Financials and Industrials rises steadily during the examined period and shows a more stable pattern compared to the dependency structure between the latter two and Healthcare, which in turn look very similar. Correlation of Healthcare with both Financials and Industrials drops in the aftermath of the financial crisis 2008/09 and the increased volatility period following the debt downgrade. Prior to that it has reached local maxima, probably evidence of the well know contagion effect during turmoil periods. Generally note substantial uncertainty in the estimates, as indicated by 50% credible bounds. Correlation $\rho_{\eta\eta,t}$ in the transition error is not modeled directly, but the generated dynamics are intuitive, with periods of increased uncertainty showing local correlation peaks in the information flow.

Turning to cross leverage dynamics visualized in Fig. C.5, the conjecture that these are dominated by the series connected to the observation error and to some extent by the first cross leverage element is confirmed. Accordingly, dynamics show large
commonalities, but differences exist. Overall, there is a substantial increase in cross leverage around periods of market turmoil as the financial crisis 2008/09 and the high volatility period following the debt downgrade, which is intuitive. For example, it can be argued that investors are more alert during and shortly after market turmoil, rebalancing their market-wide portfolios significantly as a large price drop occurs in one sector. This would lead to the observed volatility increase in other sectors. Further and as in the MSVcc model, cross leverage effects to Healthcare appear stronger, albeit not significantly. Values also exhibit the largest fluctuations. Moreover, patterns in
cross leverage dynamics may be distinguished between spillovers to the respective sectors each.

The ways in which dynamics feed through in the Cholesky-type approach are nontrivial, and the intuitive content of the generated cross leverage and transition error patterns reconfirms the usefulness of this technique. That point remains valid, even though, or indeed because the bulk of the dynamics in cross leverage effects is a result of the observation error dynamics being the dominant force in this factor-like approach.

Figure C.4: MSVdc - Correlation Observation Error $\rho_{\epsilon\epsilon,t}$ (left col.) and Transition Error $\rho_{\eta\eta,t}$ (right col.), with 50% c.b. Average, min., max., and range reported.
Figure C.5: MSVdc - Cross Leverage $\rho_{e\eta}$. Average, minimum, maximum, and range reported. 50% c.b.
C.6 Model Selection

Models are compared using the deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, and van der Linde, 2002). It is defined by

\[
DIC = \bar{D} + p_d,
\]

where

\[
p_d = \bar{D} - D(\bar{\theta}), \quad \bar{D} = E_{\theta|y}[D(\theta)], \quad D(\theta) = -2 \log f(y_{1:n}|\theta),
\]

with \(D(\theta)\) the deviance and \(\bar{\theta}\) the posterior mean. It is a Bayesian measure of model fit, attaining smaller values for better models, with \(p_d\) a penalty term for model complexity. Intuitively, higher parameter uncertainty lets \(D(\theta)\) fluctuate more widely around \(D(\bar{\theta})\), yielding larger \(p_D\) values. Numerical standard errors of the estimates are obtained the following way. Posterior expectation of the deviance \(\bar{D}\), readily available as a byproduct of the McMC run, is calculated by dividing the sample in 10 equally sized batches. Regarding deviance of the posterior mean \(D(\bar{\theta})\), the particle filter outlined in Sec. C.4 is repeatedly applied 10 times to obtain estimates of log-likelihood ordinate \(\log f_{\text{post}}(y_{1:n}|\bar{\theta}) = \sum_{t=1}^n \log \hat{f}(y_t|y_{1:t},\bar{\theta})\). The deviance of the predictive log-likelihood is calculated in a similar way from ordinate \(\log f_{\text{prior}}(y_{1:n}|\bar{\theta}) = \sum_{t=1}^n \log \hat{f}(y_t|y_{1:t-1},\bar{\theta})\). The above approach yields 100 different DIC values, from which statistics are calculated.

In addition to the dynamic correlation model of Sec. C.5.3, a MSVdc #2 model is estimated, deploying larger priors for the latent random walks variances governing cross leverage dynamics,

\[
\sigma_{ij}^2 \sim IG(5/2, 0.00001/2), \quad i > p, \quad j \leq p,
\]

allowing for more diffuse dynamics. Another candidate is the MSVdc \(\epsilon\epsilon\) model, with only the \(a_{ij,t}\) associated with covariance dynamics \(\Sigma_{\epsilon\epsilon,t}\) time variant. Its performance is also investigated. For the initial convergence period of the particle filter a time frame of 500 observations prior to the dataset is chosen for all models, which is standardized by means and standard deviations calculated from a backward rolling window having size equal to the dataset.

Tab. C.4 reports results, with the MSVcc as baseline model, and larger values indicating better fit. The DIC clearly favors the models featuring dynamic correlation.
Table C.4: Model Selection

<table>
<thead>
<tr>
<th></th>
<th>DIC (s.e.)</th>
<th>Rank</th>
<th>pD (s.e.)</th>
<th>D (s.e.)</th>
<th>Rank</th>
<th>D_{pred} (s.e.)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSVcc</td>
<td>0 (6.4)</td>
<td>4</td>
<td>0 (3.4)</td>
<td>0 (3.1)</td>
<td>4</td>
<td>0 (1.4)</td>
<td>4</td>
</tr>
<tr>
<td>MSVdc $\epsilon\epsilon$</td>
<td>93.7 (5.2)</td>
<td>3</td>
<td>93.2 (2.9)</td>
<td>186.9 (2.5)</td>
<td>3</td>
<td>55.0 (1.2)</td>
<td>1</td>
</tr>
<tr>
<td>MSVdc</td>
<td>95.9 (7.0)</td>
<td>1</td>
<td>101.1 (4.2)</td>
<td>196.9 (3.2)</td>
<td>2</td>
<td>18.5 (6.0)</td>
<td>2</td>
</tr>
<tr>
<td>MSVdc #2</td>
<td>94.1 (7.7)</td>
<td>2</td>
<td>110.3 (4.2)</td>
<td>204.3 (3.7)</td>
<td>1</td>
<td>17.1 (3.6)</td>
<td>3</td>
</tr>
</tbody>
</table>

Deviance information criterion, complexity penalty, posterior expectation of the deviance, and deviance of the predictive log-likelihood. Numerical standard error in parentheses.

and cross leverage over the MSVcc with time invariant correlation, with the presented MSVdc model the best. However, the dynamic correlation variants do not show significant differences among each other according to this criterion. Complexity penalty terms $pD$ are intuitive, with the MSVcc model having the lowest and the MSVdc #2 with most diffuse dynamics the highest value. Inspecting posterior expectation of the deviance $\bar{D}$, we can infer a significant but relatively small improvement in in-sample fit for the MSVdc variants modeling dynamics in $\Sigma_{\epsilon\eta,t:t+1}$ explicitly over the MSVdc $\epsilon\epsilon$ model. Not surprisingly then, the MSVdc #2 variant fits the data best. Regarding forecasting ability, however, deviance of the predictive log-likelihood $D_{\text{pred}}$ favors the MSVdc $\epsilon\epsilon$ variant, indicating improved predictive performance by making $\Sigma$ time varying but also risk of overfitting when time dependency in cross leverage dynamics $a_{ij,t}$, $i > p$, is modeled explicitly.

C.7 Conclusion

A multivariate stochastic volatility model with dynamic correlation and cross leverage (MSVdc) is proposed, using a Cholesky-type decomposition. Modeling volatilities and covariances separately makes a direct interpretation of leverage possible. A McMC algorithm to efficiently estimate the model and an associated particle filter are proposed and outlined in detail. Parsimony is introduced by not modeling dynamics in the transition error explicitly. An empirical application based on three main S&P 500 sector indices illustrates features of the model. Cross leverage effects peak during and in the aftermath of financial turmoil. Moreover, the indirectly generated dynamics of
volatility of volatility are intuitive. Specifically, increased unconditional tail risk and stronger correlated transition shocks are observed during distress periods.

Model selection indicates a superior fit of the proposed MSVdc model, benchmarked against the MSV model with constant correlation and cross leverage of Ishihara and Omori (2012). However, there is risk of overfitting when modeling the drivers of cross leverage explicitly, as the observation error plays a dominant role in generating the cross leverage and transition error dynamics. Consequently, a less complex model variant making only the Cholesky elements responsible for the observation error correlation dynamics time varying has shown to be competitive. This model variant also features dynamic cross leverage, due to the specific structure of the applied Cholesky decomposition. Moreover, the mechanism through which these dynamics are generated is nontrivial.

Dynamic sparsity could be introduced by deploying a latent threshold technique on elements of the Cholesky-type decomposed covariance matrix, as proposed by Nakajima (2012), and Nakajima and West (2013), especially on those elements associated with cross leverage dynamics. Another way to improve on predictive performance could be to re-estimate daily, conducting one-step-ahead forecasts during the McMC run. An obvious advantage of such a strategy is the incorporation of parameter uncertainty into the forecasts. The smoothing approach of McMC estimation, which uses all available information of the sample, may also be beneficial in this regard. In contrast, the sequential particle filter conditions only on the most recent observation. Although computationally costly, the latter strategy is feasible, but not investigated in this work. Further, especially the drivers underlying cross leverage can be modeled as stationary processes, stabilizing estimates around the mean.

It would be interesting to compare performance of the proposed model with other related models in literature, especially the matrix exponential stochastic volatility model with cross leverage proposed by Ishihara et al. (2014), or the MSV model of Nakajima (2012b) with leverage and dynamic observation error. The former model features dynamics in the observation error with constant cross leverage effects. As such, a comparison with the MSVdc $\epsilon\epsilon$ model would deliver insights about the benefits of dynamics for $\Sigma_{\epsilon\eta}$ generated indirectly via Cholesky-type decomposition. However, this is left for future research.
The multi-move sampler described below is an extension of Ishihara and Omori (2012) to handle time varying covariances and cross leverage effects. First divide \( \alpha_{1:n} \) into \( K + 1 \) blocks at random, say \( (\alpha_{k_i-1}, \ldots, \alpha_{k_i})', i = 1, \ldots, K + 1 \), with \( k_0 = 0 \) and \( k_{K+1} = n \). Knots are stochastic as proposed by Shephard and Pitt (1997) and are given by

\[
k_i = \text{int}[n(i + U_i)/(K + 2)], \quad i = 1, \ldots, K, \quad k_i - k_{i-1} \geq 2,
\]

where \( U_i \) is a random sample from the uniform distribution \( U(0, 1) \). This makes sampling more efficient as the points of conditioning (knots) change over the MCMC iterations, with \( K \) as a tuning parameter.

Let \( x_t = R_t^{-1}\eta_t \), where matrix \( R_t \) denotes a Cholesky decomposition of \( \Sigma_{\eta\eta,t} \) such that \( \Sigma_{\eta\eta,t} = R_t R'_t \), \( t > 1 \), and \( \Sigma_{\eta\eta,\text{ini}} = R_1 R'_1 \). Further suppose that \( k_{i-1} = r \) and \( k_i = r + d \) for the \( i \)th block. The sampler then exploits the independence of disturbances \( \eta_t \), block sampling \( x \equiv (x'_{r+1}, \ldots, x'_{r+d})' \) instead of \( \alpha \equiv (\alpha'_{r+1}, \ldots, \alpha'_{r+d})' \) from its full joint conditional posterior density \( (r \geq 0, d \geq 2, r + d \leq n) \),

\[
\pi(x|\alpha_{r+d+1}, y_{r+1}, \ldots, y_{r+d}, \theta) \propto \prod_{t=r+1}^{r+d} f(y_t|\alpha_t, \alpha_{t+1}) \times \prod_{t=r+1}^{r+d} f(x_t) \times f(\alpha_{r+d}), \quad r = 0, \quad (C.22)
\]

\[
\pi(x|\alpha_r, \alpha_{r+d+1}, y_r, \ldots, y_{r+d}, \theta) \propto \prod_{t=r}^{r+d} f(y_t|\alpha_t, \alpha_{t+1}) \times \prod_{t=r+1}^{r+d} f(x_t) \times f(\alpha_{r+d}), \quad r + d < n,
\]

\[
\pi(x|\alpha_r, y_r, \ldots, y_{r+d}, \theta) \propto \prod_{t=r}^{r+d-1} f(y_t|\alpha_t, \alpha_{t+1}) \times f(y_n|\alpha_n) \times \prod_{t=r+1}^{r+d} f(x_t), \quad r + d = n. \quad (C.23)
\]
The logarithm of \( f(y_t|\alpha_t, \alpha_{t+1}) \), \( f(y_n|\alpha_n) \) in Eq. (C.22)-(C.23) can be written as (excluding constant terms)

\[
l_t = -\frac{1}{2} 1_p' \alpha_t - \frac{1}{2} (z_t - m_t)' S_t^{-1} (z_t - m_t),
\]

where \( 1_p \) is a \( p \times 1 \) vector of ones, and

\[
z_t = V_t^{-1/2} y_t, \quad m_t = \sum_{\eta,t:t+1} \Sigma_{\eta,t+1}^{-1} (\alpha_{t+1} - \Phi \alpha_t) I_{t<n} + 0,
\]

\[
S_t = \begin{cases} 
\Sigma_{\epsilon,t} - \sum_{\eta,t:t+1} \Sigma_{\eta,t+1}^{-1} \Sigma_{\eta,t+1} & t < n \\
\Sigma_{\epsilon,t} & t = n.
\end{cases}
\]

Then the logarithmized posterior density given by Eq. (C.22)-(C.23) can be expressed as

\[
L = \log f(\alpha_{r+d}) I_{r+d<n} + \begin{cases} 
\sum_{r+1}^{r+d} l_t & r = 0 \\
\sum_{r}^{r+d} l_t & r > 0,
\end{cases}
\]

\[
(C.24)
\]

\[
\log f(\alpha_{r+d}) = -\frac{1}{2} (\alpha_{r+d+1} - \Phi \alpha_{r+d})' \Sigma_{\eta,r+1}^{-1} (\alpha_{r+d+1} - \Phi \alpha_{r+d}).
\]

To construct a proposal density based on a normal approximation of the posterior density of \( \alpha \), define further (for first and second derivatives \( \delta_t, A_t, B_t \), see App. C.B)

\[
\delta = (\delta'_{r+1}, \ldots, \delta'_{r+d})', \quad \delta_t = \frac{\partial L}{\partial \alpha_t},
\]

\[
Q = -E \left( \frac{\partial^2 L}{\partial \alpha \partial \alpha'} \right) = \begin{pmatrix}
A_{r+1} & B_{r+2} & 0 & \cdots & 0 \\
B_{r+2} & A_{r+2} & B_{r+3} & \cdots & 0 \\
0 & B_{r+3} & A_{r+3} & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & B_{r+d} & A_{r+d}
\end{pmatrix},
\]

\[
A_t = -E \left( \frac{\partial^2 L}{\partial \alpha_t \partial \alpha_t'} \right), \quad t = r + 1, \ldots, r + d,
\]

\[
B_t = -E \left( \frac{\partial^2 L}{\partial \alpha_t \partial \alpha_{t-1}'} \right), \quad t = r + 2, \ldots, r + d, \quad B_{r+1} = 0.
\]
Applying a second order Taylor expansion to the log of the posterior density in Eq. (C.24) around mode \(\hat{x}\) yields an approximating normal density as follows (excluding constant terms):

\[
\log \pi(x|\cdot) \\
\approx \hat{L} + \frac{\partial L}{\partial x'} \bigg|_{x=\hat{x}} (x - \hat{x}) + \frac{1}{2} (x - \hat{x})' E \left( \frac{\partial^2 L}{\partial x \partial x'} \right) \bigg|_{x=\hat{x}} (x - \hat{x}) - \frac{1}{2} \sum_{t=r+1}^{r+d} x'_t x_t \\
= \hat{L} + \delta'(\alpha - \hat{\alpha}) - \frac{1}{2} (\alpha - \hat{\alpha})' \hat{Q} (\alpha - \hat{\alpha}) - \frac{1}{2} \sum_{t=r+1}^{r+d} x'_t x_t \\
\equiv \log q(x|\cdot),
\]

where \(\hat{L}, \hat{\delta}\) and \(\hat{Q}\) are the values of \(L, \delta\) and \(Q\) at \(\alpha = \hat{\alpha}\) (or at \(x = \hat{x}\), equivalently). It can be shown that proposal density \(q(x|\cdot)\) is the posterior of \(x\) obtained from an ordinary linear state space model given by Eq. (C.25)-(C.26) below. Mode \(\hat{x}\) is found by repeating the following algorithm until convergence (usually 2 to 5 iterations).

**Algorithm 1 (Disturbance smoother):**

1. Initialize \(\hat{x}\), and compute \(\hat{\alpha}\) at \(x = \hat{x}\) using state equation Eq. (C.6) recursively.

2. Evaluate \(\hat{\delta}_t, \hat{A}_t, \text{ and } \hat{B}_t\) at \(\alpha = \hat{\alpha}, t = r + 1, \ldots, r + d\).

3. Set \(\hat{D}_{r+1} = \hat{A}_{r+1}\) and \(\hat{b}_{r+1} = \hat{\delta}_{r+1}\). Compute the following variables recursively for \(t = r + 2, \ldots, r + d\):

\[
\hat{D}_t = \hat{A}_t - \hat{B}_t \hat{D}_{t-1}^{-1} \hat{B}_t', \quad \hat{K}_t = \text{chol}(\hat{D}_t), \\
\hat{b}_t = \hat{\delta}_t - \hat{B}_t \hat{D}_{t-1}^{-1} \hat{b}_{t-1},
\]

with \(\text{chol}(\hat{D}_t)\) a Cholesky decomposition of \(\hat{D}_t\) such that \(\hat{D}_t = \hat{K}_t \hat{K}_t'\).

4. Define auxiliary vector \(\hat{y}_t = \hat{\varphi}_t + \hat{D}_t^{-1} \hat{b}_t\), where \(\hat{\varphi}_t = \hat{\alpha}_t + \hat{D}_t^{-1} \hat{B}_t \hat{\alpha}_{t+1}, t = r + 1, \ldots, r + d\), and set \(\hat{B}_{d+r+1} = 0, \hat{\alpha}_{r+d+1} = \alpha_{r+d+1}\).
5. Consider the linear Gaussian state space model formulated by
\[
\hat{y}_t = Z_t \alpha_t + G_t \kappa_t, \quad t = r + 1, \ldots, r + d, \tag{C.25}
\]
\[
\alpha_{t+1} = \Phi \alpha_t + H_{t+1} \omega_{t+1}, \quad t = r, \ldots, r + d - 1,
\]
\[
(\kappa_t, \omega_{t+1})' \sim N(0, I_{2p}), \tag{C.26}
\]
with
\[
Z_t = I_p + \hat{D}_t^{-1} \hat{B}_t' \Phi, \quad G_t = (\hat{K}_t^{-1}, \hat{D}_t^{-1} \hat{B}_t' \hat{R}_{t+1}), \quad H_{t+1} = (0, \hat{R}_{t+1}),
\]
and \(\alpha_0 = 0\). Apply the Kalman filter (e.g. Durbin and Koopman, 2008) and disturbance smoother (Koopman, 1993) to this state space model to obtain posterior mode \(\hat{x}\) or \(\hat{\alpha}\), equivalently.

6. Go to 2.

In the McMC sampling procedure, the current sample of \(x\) may be taken as initial value of \(\hat{x}\) in Step 1. Further note that the above steps are equivalent to the method of scoring used to maximize the conditional posterior density. After convergence of Algorithm 1, \(x\) is sampled from the conditional posterior density by conducting an AR (Accept - Reject) - MH algorithm (Tierney, 1994) using the simulation smoother (de Jong and Shephard, 1995, or Durbin and Koopman, 2002).

**Algorithm 2 (AR-MH step and simulation smoother):**

1. Let \(x_0\) denote the current value. Find mode \(\hat{x}\) using Algorithm 1.

2. Proceed with Steps 2-4 in Algorithm 1 to obtain the approximated linear Gaussian state space model of Eq. (C.25)-(C.26).

3. Propose a candidate \(x^*\) by sampling from \(\tilde{q}(x^*) \propto \min\{\pi(x^*|\cdot), cq(x^*|\cdot)\}\) using the AR algorithm as follows:

   (a) Generate \(x^*\) using the simulation smoother for the approximated state space of model Eq. (C.25)-(C.26).

   (b) Accept \(x^*\) with probability
\[
\frac{\min\{\pi(x^*|\cdot), cq(x^*|\cdot)\}}{cq(x^*|\cdot)},
\]
where \(c\) is a scaling constant. If it is rejected, go to (a).
4. Conduct the MH algorithm using candidate \( \mathbf{x}^* \), with acceptance probability
\[
\min \left\{ \frac{\pi(\mathbf{x}^* | \cdot)}{\pi(\mathbf{x}_0 | \cdot)} \times \min \{ \pi(\mathbf{x}_0^* | \cdot), cq(\mathbf{x}_0^* | \cdot) \}, 1 \right\}.
\]

### C.B Derivatives \( \delta_t, A_t, B_t \)

Before calculating derivatives \( \delta_t, A_t \) and \( B_t \), definitions of the first and second derivative of a matrix and the product and chain rule as in Magnus and Abadir (2007) are provided. Let \( F \) be a twice differentiable \( m \times p \) matrix function of an \( n \times q \) matrix \( \mathbf{X} \). Then the first derivative (Jacobian) of \( F \) at \( \mathbf{X} \) is defined by the \( mp \times nq \) matrix
\[
\mathbf{D}F(\mathbf{X}) = \frac{\partial F(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \text{vec}(F(\mathbf{X}))}{\partial \text{vec}(\mathbf{X})'},
\]
and the second derivative (Hessian) of \( F \) at \( \mathbf{X} \) is defined by the \( mnpq \times nq \) matrix
\[
\mathbf{H}F(\mathbf{X}) = \mathbf{D}((\mathbf{D}F(\mathbf{X}))') = \frac{\partial}{\partial \text{vec}(\mathbf{X})'} \text{vec}\left( \left( \frac{\partial \text{vec}(F(\mathbf{X}))}{\partial \text{vec}(\mathbf{X})'} \right)' \right).
\]

**Product Rule:** Let \( S \) be a subset of \( \mathbb{R}^{n \times q} \), and assume that \( F : S \to \mathbb{R}^{m \times p} \) and \( G : S \to \mathbb{R}^{p \times r} \) are differentiable at an interior point \( C \) of \( S \). Then
\[
\frac{\partial \text{vec}(FG)}{\partial \text{vec}(\mathbf{X})'} = (G' \otimes \mathbf{I}_m) \frac{\partial \text{vec}(F)}{\partial \text{vec}(\mathbf{X})'} + (\mathbf{I}_r \otimes F) \frac{\partial \text{vec}(G)}{\partial \text{vec}(\mathbf{X})'}.
\]  
(C.27)

**Chain Rule:** Let \( S \) be a subset of \( \mathbb{R}^{n \times q} \), and assume that \( F : S \to \mathbb{R}^{m \times p} \) is differentiable at an interior point \( C \) of \( S \). Let \( T \) be a subset of \( \mathbb{R}^{m \times p} \) such that \( F(\mathbf{X}) \in T \) for all \( \mathbf{X} \in S \), and assume that \( G : T \to \mathbb{R}^{r \times s} \) is differentiable at an interior point \( B = F(C) \) of \( T \). Then the composite function \( H : S \to \mathbb{R}^{r \times s} \) defined by \( H(\mathbf{X}) = G(F(\mathbf{X})) \) is differentiable at \( C \), and
\[
\mathbf{D}H(\mathbf{X}) = (\mathbf{D}G(F(\mathbf{X}))(\mathbf{D}F(\mathbf{X})) = \frac{\partial \text{vec}(G(F(\mathbf{X}))))}{\partial \text{vec}(F(\mathbf{X}'))} \frac{\partial \text{vec}(F(\mathbf{X}))}{\partial \text{vec}(\mathbf{X})'}.
\]

If \( q = 1 \) then \( \mathbf{x} \in \mathbb{R}^{n \times 1}, f : \mathbb{R}^{n \times 1} \to \mathbb{R}^{m \times p}, g : \mathbb{R}^{m \times p} \to \mathbb{R}^{r \times s} \), and
\[
\frac{\partial g(f(\mathbf{x}))}{\partial \mathbf{x}'} = \frac{\partial \text{vec}(g(f(\mathbf{x}))))}{\partial \text{vec}(f(\mathbf{x}'))} \frac{\partial \text{vec}(f(\mathbf{x})))}{\partial \text{vec}(\mathbf{x})'}.
\]  
(C.28)
With $L$ as given in Eq. (C.24) first derivatives are

$$\delta_t = \frac{\partial L}{\partial \alpha_t} = \frac{\partial l_t}{\partial \alpha_t} + \frac{\partial l_{t-1}}{\partial \alpha_t} + \Phi \Sigma^{-1}_{\eta_i,t+1}(\alpha_{t+1} - \Phi \alpha_t)I_{t=r+d<n},$$

where, applying the chain rule of Eq. (C.28),

$$\frac{\partial l_t}{\partial \alpha'_t} = -\frac{1}{2} \mathbf{1}_p - (z_t - m_t)' S^{-1}_t \frac{\partial (z_t - m_t)}{\partial \alpha'_t}$$

$$= -\frac{1}{2} \mathbf{1}_p + \frac{1}{2} (z_t - m_t)' S^{-1}_t \left[ \text{diag}(z_t) - 2\Sigma_{\eta_i,t+1} \Sigma^{-1}_{\eta_i,t+1} \Phi I_{t<n} \right],$$

$$\frac{\partial l_{t-1}}{\partial \alpha'_t} = (z_{t-1} - m_{t-1})' S^{-1}_{t-1} \frac{\partial m_{t-1}}{\partial \alpha'_t}$$

$$= (z_{t-1} - m_{t-1})' S^{-1}_{t-1} \Sigma_{\eta_i,t-1:t} \Sigma^{-1}_{\eta_i,t} I_{t>1}.$$ 

Thus we have

$$\delta_t = -\frac{1}{2} \mathbf{1}_p + \frac{1}{2} \left[ \text{diag}(z_t) - 2\Phi \Sigma^{-1}_{\eta_i,t+1} \Sigma_{\eta_i,t+1} I_{t<n} \right] S^{-1}_t (z_t - m_t)$$

$$+ \Sigma_{\eta_i,t} \Sigma_{\eta_i,t-1:t} S^{-1}_{t-1} (z_{t-1} - m_{t-1}) I_{t>1} + \Phi \Sigma^{-1}_{\eta_i,t+1} (\alpha_{t+1} - \Phi \alpha_t) I_{t=r+d<n}.$$

To compute second derivative $A_t$ start by calculating the Hessian matrix

$$\frac{\partial^2 L}{\partial \alpha_t \partial \alpha'_t} = \frac{\partial^2 l_t}{\partial \alpha_t \partial \alpha'_t} + \frac{\partial^2 l_{t-1}}{\partial \alpha_t \partial \alpha'_t} - \Phi \Sigma^{-1}_{\eta_i,t+1} \Phi I_{t=r+d<n},$$

where, applying the product rule of Eq. (C.27),

$$\frac{\partial^2 l_t}{\partial \alpha_t \partial \alpha'_t} = \frac{1}{2} \left[ (z_t - m_t)' S^{-1}_t \otimes I_p \right] \frac{\partial \text{vec}(\text{diag}(z_t))}{\partial \alpha'_t}$$

$$- \frac{1}{4} \left[ \text{diag}(z_t) - 2\Phi \Sigma^{-1}_{\eta_i,t+1} \Sigma_{\eta_i,t+1} I_{t<n} \right]$$

$$\times S^{-1}_t \left[ \text{diag}(z_t) - 2\Sigma_{\eta_i,t+1} \Sigma^{-1}_{\eta_i,t+1} \Phi I_{t<n} \right],$$

$$\frac{\partial^2 l_{t-1}}{\partial \alpha_t \partial \alpha'_t} = -\Sigma_{\eta_i,t} \Sigma_{\eta_i,t-1:t} S^{-1}_{t-1} \Sigma_{\eta_i,t-1:t} \Sigma^{-1}_{\eta_i,t} I_{t>1}. $$
Now observe that
\[
\frac{\partial \text{vec}(\text{diag}(z_t))}{\partial \alpha_i} = -\frac{1}{2} \begin{pmatrix}
    z_{1,t} e_1 e_1' \\
    \vdots \\
    z_{p,t} e_p e_p'
\end{pmatrix}, \quad E(z_{i,t}(z_t - m_t)) = S_t e_i,
\]
where \(e_i\) is a \(p \times 1\) unit vector with the \(i\)th component equal to 1. Then the expected value of the first term in Eq. (C.29) is
\[
E\left(\left((z_t - m_t)' S_t^{-1} \otimes I_p\right) \frac{\partial \text{vec}(\text{diag}(z_t))}{\partial \alpha_i}\right) = -\frac{1}{2} \sum_{i=1}^{p} (e_i' S_t^{-1})(S_t e_i) e_i' = -\frac{1}{2} I_p.
\]
Noting that \(\text{diag}(z_t) S_t^{-1} \text{diag}(z_t) = S_t^{-1} \otimes (z_t z_t')\), the expected value of the second term in Eq. (C.29) is
\[
E\left(\left[\text{diag}(z_t) - 2 \Phi \Sigma_{\eta \eta, t+1} \Sigma_{\eta \epsilon, t:t+1} I_{t<n}\right] S_t^{-1} \left[\text{diag}(z_t) - 2 \Sigma_{\eta \epsilon, t:t+1} \Sigma_{\eta \eta, t+1} \Phi I_{t<n}\right]\right)
\]
\[
= S_t^{-1} \otimes (S_t + m_t m_t') + 4 \Phi \Sigma_{\eta \eta, t+1} \Sigma_{\eta \epsilon, t:t+1} S_t^{-1} \Sigma_{\eta \epsilon, t:t+1} \Sigma_{\eta \eta, t+1} \Phi I_{t<n}
\]
\[
- 2 \left[\Phi \Sigma_{\eta \eta, t+1} \Sigma_{\eta \epsilon, t:t+1} S_t^{-1} \text{diag}(m_t) + \text{diag}(m_t) S_t^{-1} \Sigma_{\eta \epsilon, t:t+1} \Sigma_{\eta \eta, t+1} \Phi\right] I_{t<n},
\]
and finally we have
\[
A_t = \frac{1}{4} \left[I_p + S_t^{-1} \otimes (S_t + m_t m_t')\right] + \Phi \Sigma_{\eta \eta, t+1} \Sigma_{\eta \epsilon, t:t+1} S_t^{-1} \Sigma_{\eta \epsilon, t:t+1} \Sigma_{\eta \eta, t+1} \Phi I_{t<n}
\]
\[
- \frac{1}{2} \left[\Phi \Sigma_{\eta \eta, t+1} \Sigma_{\eta \epsilon, t:t+1} S_t^{-1} \text{diag}(m_t) + \text{diag}(m_t) S_t^{-1} \Sigma_{\eta \epsilon, t:t+1} \Sigma_{\eta \eta, t+1} \Phi\right] I_{t<n}
\]
\[
+ \Sigma_{\eta \eta, t} \Sigma_{\eta \epsilon, t-1:t} S_{t-1}^{-1} \Sigma_{\eta \eta, t-1:t} I_{t-1} + \Phi \Sigma_{\eta \eta, t+1} \Phi I_{t=r+d<n}.
\]
In a similar fashion it is straightforward to show that
\[
B_t = -E\left(\frac{\partial^2 l_{t-1}}{\partial \alpha_t \partial \alpha_{i-1}'}\right) = \frac{1}{2} \Sigma_{\eta \eta, t} \Sigma_{\eta \epsilon, t-1:t} S_{t-1}^{-1} \left[\text{diag}(m_{t-1}) - 2 \Sigma_{\eta \epsilon, t-1:t} \Sigma_{\eta \eta, i} \Phi\right].
\]
C.C  
**Sampling from the Wishart Distribution**

Sampling of $\Sigma$ time invariant is done by the MH algorithm. Assuming a conjugate Wishart prior of the form $\Sigma^{-1} \sim \mathcal{W}(\nu_0, V_0)$, the conditional posterior of $\Sigma$ reads

$$\pi(\Sigma|\cdot) \propto |\Sigma|^{-\nu+2p+1/2} \exp\left\{ -\frac{1}{2} \text{tr}(V^{-1}\Sigma^{-1}) \right\} \times c(\Sigma),$$

where

$$c(\Sigma) = |\Sigma_{\text{ini}}|^{1/2} |\Sigma_{ee}|^{-1/2} \exp\left\{ -\frac{1}{2} \left( \alpha'_{1} \Sigma_{\text{ini}}^{-1} \alpha_1 + y_n' V_n^{-1/2} \Sigma_{ee}^{-1} V_n^{-1/2} y_n \right) \right\},$$

$$V^{-1} = V_0^{-1} + \sum_{t=1}^{n-1} v_t v_t', \quad v_t = \left( V_t^{-1/2} y_t \right), \quad v = \nu_0 + n - 1.$$

A proposal $\Sigma^* \sim [\mathcal{W}(\nu, V)]^{-1}$ is accepted with probability $\min\{c(\Sigma^*)/c(\Sigma), 1\}$, where $\Sigma$ is the current sample.
**Bibliography**


CURRICULUM VITAE

Sebastian Trojan
Munich, Germany
Email: sebastian.trojan@web.de

Education

2008 – 2014 PhD in Economics and Finance,
University of St. Gallen (HSG, CH).
2001 – 2007 Diplom-Volkswirt Univ. (M.Sc. Econ) (with distinction),
Ludwig Maximilian University of Munich (LMU, DE).

Work Experience

2008 – 2009 Quantitative Analyst, portfolio management,
cominvest Asset Management GmbH (Frankfurt/DE),
Equities / Quantitative Products (buy side).
2007 – 2007 Associate, IT consultancy,
BUR Gesellschaft für EDV-Anwendungsberatung und
Realisierung m.b.H. (Munich/DE).
2004 – 2005 Working Student, controlling,
AXA Assistance Deutschland GmbH (Munich/DE).
Assistance insurance.

Further Accomplishments

1999 – 2000 Overland by car to India.
1998 – 1999 Civil service (Munich/DE).