Universität St.Gallen

Modeling Intraday Stochastic Volatility and Conditional Duration Contemporaneously with Regime Shifts

Sebastian Trojan

August 2014 Discussion Paper no. 2014-25

School of Economics and Political Science, Department of Economics

University of St.Gallen

Editor:	Martina Flockerzi University of St.Gallen School of Economics and Political Science Department of Economics Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25 Fax +41 71 224 31 35 Email seps@unisg.ch
Publisher:	School of Economics and Political Science Department of Economics University of St.Gallen Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25
Electronic Publication:	Fax +41 71 224 31 35 http://www.seps.unisg.ch

Modeling Intraday Stochastic Volatility and Conditional Duration Contemporaneously with Regime Shifts[1]

Sebastian Trojan

Author's address: Sebastian Trojan
Institute of Mathematics and Statistics
University of St.Gallen
Bodanstrasse 6
CH-9000 St. Gallen
Email sebastian.trojan@student.unisg.ch
sebastian.trojan@web.de

[1] The author thanks Francesco Audrino, Michael Lechner, Nikolaus Hautsch, and seminar participants at the University of St.Gallen for their helpful comments and suggestions.

Abstract

A high frequency stochastic volatility (SV) model is proposed. 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.

Keywords

Stochastic volatility, stochastic conditional duration, non-Gaussian and nonlinear state space model, tick data, event time, generalized gamma distribution, negative binomial distribution, regime switching, Markov chain Monte Carlo, block sampler, particle filter, adaptive Metropolis.

JEL Classification

C11, C15, C32, C58.

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.[1] 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 $\mathrm{AR}(1)$ 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. Madhaven (2000) for a review of theoretical and empirical literature on market microstructure. Hasbrouck (2007) deals with the topic in a unified framework. Hautsch (2012) presents an overview of the various methodological approaches applied to high frequency data.

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.[2,3] 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

[1] To 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.

[2] The origins of SV trace back to Clark (1973), Taylor (1982, 1986), Rosenberg (1972), Johnson and Shanno (1987), and Hull and White (1987). A review is given by Shephard and Andersen (2008), or in book-length by Shephard (2005).

[3] Engle (1982) proposed ground-breaking 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).

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 2 presents the model. Section 3 proposes a McMC algorithm for posterior computation and Section 4 an associated particle filter. Section 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 6, including parameter and regime inference for model variants, model selection and predictive density tests. Section 7 concludes. The Appendix contains technical details and estimates.

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,

$$d_t \backsim \mathcal{GG}\Big(\exp\left[\psi_{d,t}\right], \zeta, \xi\Big), \tag{1}$$

$$y_t \backsim \mathcal{NB}\Big(\exp\left[\psi_{y,t} + \boldsymbol{\kappa}_y \boldsymbol{J}_y\right], \tau\Big), \qquad t = 1, \ldots, n, \tag{2}$$

$$\boldsymbol{\psi}_t = \boldsymbol{\mu}_{s_t} + \boldsymbol{\Phi}(\boldsymbol{\psi}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}) + \boldsymbol{\eta}_t, \qquad t = 2, \ldots, n, \tag{3}$$

with

$$\boldsymbol{\psi}_t = (\psi_{d,t}, \psi_{y,t})', \quad \boldsymbol{\Phi} = \begin{pmatrix} \phi_{dd} & \phi_{dy} \\ \phi_{yd} & \phi_{yy} \end{pmatrix}, \tag{4}$$

$$\boldsymbol{\mu}_{s_t} = \begin{pmatrix} \mu_{d,s_{d,t}} \\ \mu_{y,s_{y,t}} \end{pmatrix}, \ \ \mu_{k,s_{k,t}} = \sum\nolimits_{i=1}^{M_k} \mu_{k,i} \times \mathrm{I}_{s_{k,t}=i}, \ \ M_k \in \mathbb{N}^+, \ \ k = \{d, y\}, \tag{5}$$

$$\boldsymbol{\eta}_t \backsim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_d^2 & \rho\sigma_d\sigma_y \\ \rho\sigma_y\sigma_d & \sigma_y^2 \end{pmatrix}, \tag{6}$$

$$\boldsymbol{\psi}_1 \backsim \mathcal{N}(\boldsymbol{\mu}_{s_1}, \boldsymbol{\Sigma}_{\text{ini}}), \quad \text{vec}(\boldsymbol{\Sigma}_{\text{ini}}) = (\boldsymbol{I}_{p^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi})^{-1} \text{vec}(\boldsymbol{\Sigma}), \tag{7}$$

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 $\boldsymbol{I}_p$ the identity matrix of dimension p. $\mathrm{I}_{\mathrm{s_{k,t}=i}}$ 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 $\boldsymbol{\psi}_t$, Eq. (3), determining dynamics of price duration and absolute price change contemporaneously. Its persistence matrix $\boldsymbol{\Phi}$, Eq. (4), is fully specified to model the effect of past latent absolute price change on expected duration and vice versa. Moreover, $\boldsymbol{\psi}_t$ is constraint to be stationary (i.e. the eigenvalues of $\boldsymbol{\Phi}$ must have modulus less than one), and initial value $\boldsymbol{\psi}_1$ is assumed to follow the unconditional distribution, Eq. (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 ($\mathcal{GG}$) distribution, having shape parameters ζ and ξ, Eq. (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 $\zeta \to \infty$), enabling a systematic approach 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 ($\mathcal{NB}$) distributed with shape parameter τ, Eq. (2). This distribution is parsimonious but allows for overdispersion, in contrast to the Poisson.

Time-of-day (tod) effects common in intraday data (see 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, $\boldsymbol{J}_y$ is a "selector" matrix such that $\boldsymbol{\kappa}_y \boldsymbol{J}_y$ selects the appropriate time-of-day dummy from $1 \times R$ row vector $\boldsymbol{\kappa}_y$, 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.

Table 1: Observation Densities[†]

Generalized Gamma (price duration):

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

$f_{\mathcal{GG}}(d_t|\psi_{d,t}, \zeta, \xi) = \dfrac{\xi}{\Gamma(\zeta)}\Big(d_t \exp(-\psi_{d,t})a\Big)^{\zeta\xi} d_t^{-1} \exp\Big\{-(d_t \exp(-\psi_{d,t})a)^{\xi}\Big\}$

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

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

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

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

Negative Binomial (absolute price change):

$p \equiv \dfrac{\exp(\psi_{y,t})}{\exp(\psi_{y,t}) + \tau}$

$f_{\mathcal{NB}}(y_t|\psi_{y,t}, \tau) = \dfrac{\Gamma(y_t + \tau)}{\Gamma(y_t + 1)\Gamma(\tau)}(1 - p)^{\tau} p^{y_t}$

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

Discrete support: $y_t \in \{0, 1, 2, \dots\}$, number of failures: $\tau > 0$

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

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

[†] $f(\cdot)$, $F(\cdot)$ denote the density and cumulative distribution functions, respectively.
[††] $\gamma(\cdot)$ denotes the lower incomplete gamma function.
[†††] $I_p(\cdot)$ denotes the normalized incomplete beta function.

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

To account for fast changing nonlinear patterns of market activity over the day, mode parameters $\boldsymbol{\mu}_{s_t}$ of the latent information flow each follow their own regime dependent process, with M_k, $k = \{d, y\}$, the number of states, Eq. (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 $i, j \in \{1, \dots, M_k\}$. The transition probability matrix has the general form

$$\boldsymbol{P}_k = \begin{pmatrix} p_{k,11} & \cdots & p_{k,1M_k} \\ \vdots & \ddots & \vdots \\ p_{k,M_k1} & \cdots & p_{k,M_kM_k} \end{pmatrix}.$$

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 a 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.

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 and Pitt, 1997). For a thorough treatment of the McMC method, consider e.g. Robert and Casella (2004).

Define parameter set $\boldsymbol{\theta} \equiv \{\boldsymbol{\mu}, \boldsymbol{\kappa}_y, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \zeta, \xi, \tau, \boldsymbol{P}_k\}$, with state dependent matrices $\boldsymbol{\mu} = (\boldsymbol{\mu}_d, \boldsymbol{\mu}_y)'$, $\boldsymbol{\mu}_k = (\mu_{k,1}, \dots, \mu_{k,M_k})'$, $k = \{d, y\}$. Further denote by $\boldsymbol{\theta}_{-r}$ all parameters of $\boldsymbol{\theta}$ without r, and let $\boldsymbol{d}_{1:n} = (d_1, \dots, d_n)'$, $\boldsymbol{y}_{1:n} = (y_1, \dots, y_n)'$, $\boldsymbol{\psi}_{k,1:n} = (\boldsymbol{\psi}_{k,1}, \dots, \boldsymbol{\psi}_{k,n})'$, $\boldsymbol{\psi}_{1:n} = (\boldsymbol{\psi}_{d,1:n}, \boldsymbol{\psi}_{y,1:n})$, $\boldsymbol{s}_{k,0:n} = (s_{k,0}, s_{k,1}, \dots, s_{k,n})'$. 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 $\pi(\boldsymbol{\mu}_k)$ are assumed,

$$\boldsymbol{\mu}_k \backsim \mathcal{N}(\boldsymbol{\mu}_{0,k}, \boldsymbol{\Sigma}_{\mu_0,k}),$$

with $\boldsymbol{\Sigma}_{\mu_0,k}$ diagonal. Further assume $\pi(\boldsymbol{\Sigma}^{-1})$ conjugate Wishart,

$$\boldsymbol{\Sigma}^{-1} \backsim \mathcal{W}(\nu_0, \boldsymbol{V}_0),$$

and $\pi(\boldsymbol{P}_{k,i\cdot})$ conjugate Dirichlet,

$$\boldsymbol{P}_{k,i\cdot} \backsim \mathcal{D}(\omega_{k,i1}, \dots, \omega_{k,iM_k}), \quad \omega_{k,ij} \neq 0, \quad k = \{d, y\}, \quad i, j = 1, \dots, M_k,$$

where $\boldsymbol{P}_{k,i\cdot}$ denotes the i^{th} (independent) row of $\boldsymbol{P}_k$.

The McMC algorithm then consists of the following steps:[4]

1. Initialize $\boldsymbol{\theta}$, $\boldsymbol{s}_{d,0:n}$, $\boldsymbol{s}_{y,0:n}$, $\boldsymbol{\psi}_{1:n}$.
2. Generate $\boldsymbol{\psi}_{1:n} \,|\, \boldsymbol{\theta}, \boldsymbol{s}_{d,0:n}, \boldsymbol{s}_{y,0:n}, \boldsymbol{d}_{1:n}, \boldsymbol{y}_{1:n}$.
3. Generate $\boldsymbol{s}_{d,0:n} \,|\, \boldsymbol{\theta}, \boldsymbol{\psi}_{d,1:n}, \boldsymbol{d}_{1:n}$.
4. Generate $\boldsymbol{s}_{y,0:n} \,|\, \boldsymbol{\theta}, \boldsymbol{\psi}_{y,1:n}, \boldsymbol{y}_{1:n}$.
5. Generate $\boldsymbol{P}_d \,|\, \boldsymbol{s}_{d,0:n}$.
6. Generate $\boldsymbol{P}_y \,|\, \boldsymbol{s}_{y,0:n}$.
7. Generate $(\zeta, \xi)' \,|\, \boldsymbol{\theta}_{-(\zeta,\xi)'}, \boldsymbol{\psi}_{d,1:n}, \boldsymbol{s}_{d,0:n}, \boldsymbol{d}_{1:n}$.
8. Generate $\tau \,|\, \boldsymbol{\theta}_{-\tau}, \boldsymbol{\psi}_{y,1:n}, \boldsymbol{s}_{y,0:n}, \boldsymbol{y}_{1:n}$.
9. Generate $\boldsymbol{\kappa}_y \,|\, \boldsymbol{\theta}_{-\kappa_y}, \boldsymbol{\psi}_{y,1:n}, \boldsymbol{s}_{y,0:n}, \boldsymbol{y}_{1:n}$.
10. Generate $\boldsymbol{\mu} \,|\, \boldsymbol{\theta}_{-\mu}, \boldsymbol{s}_{d,0:n}, \boldsymbol{s}_{y,0:n}, \boldsymbol{\psi}_{1:n}$.
11. Generate $\boldsymbol{\Sigma} \,|\, \boldsymbol{\theta}_{-\Sigma}, \boldsymbol{s}_{d,0:n}, \boldsymbol{s}_{y,0:n}, \boldsymbol{\psi}_{1:n}$.
12. Generate $\boldsymbol{\Phi} \,|\, \boldsymbol{\theta}_{-\Phi}, \boldsymbol{s}_{d,0:n}, \boldsymbol{s}_{y,0:n}, \boldsymbol{\psi}_{1:n}$.
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 $\boldsymbol{P}_k$

Step 2. Latent states $\boldsymbol{\psi}_{1:n}$ 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. A.

Step 3. Sampling of discrete states $\boldsymbol{s}_{d,0:n}$ 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 $\Psi_t \equiv \{\boldsymbol{\theta}, \boldsymbol{\psi}_{d,1:t}, \boldsymbol{d}_{1:t}\}$, 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 t^{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.

[4]Implementation (also the particle filter of Sec. 4) is in stand alone C++ code developed by the author using the freely available Scythe statistical library (Pemstein, Quinn, and Martin, 2011).

Forward Filter - Step 2: Once d_t is observed at the end of the t^{th} iteration, the 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. (1), reparameterized as in Sec. 4, Eq. (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,

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

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

Backward Sampling: States $\boldsymbol{s}_{0:n}$ are drawn as a block from the following conditional distribution

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

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 $\boldsymbol{s}_{d,0:n}$, transition probability matrix $\boldsymbol{P}_d$ is independent of all other variables in the model. Further, rows $\boldsymbol{P}_{d,i\cdot}$ of $\boldsymbol{P}_d$ are independent a posteriori and are drawn from the following Dirichlet distribution (omitting subscript d),

$$\boldsymbol{P}_{i\cdot} \backsim \mathcal{D}(\omega_{i1} + n_{i1}(\boldsymbol{s}_{0:n}), \ldots, \omega_{iM} + n_{iM}(\boldsymbol{s}_{0:n})), \quad i = 1, \ldots, M,$$

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

Step 6. In analogy to Step 5.

Generation of distributional shape parameters ζ, ξ, τ

Step 7. No closed form exists for the conditional posterior of generalized gamma shape parameters ζ, ξ, 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. 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. C.

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

Generation of time-of-day dummies $\boldsymbol{\kappa}_y$

Step 9. Time-of-day dummies $\boldsymbol{\kappa}_y$ 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]

Generation of latent process parameters $\boldsymbol{\mu}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Sigma}$

Step 10. Sampling of latent process modes $\boldsymbol{\mu}_k$, $k = \{d, y\}$, 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 $\boldsymbol{\mu}_d|\boldsymbol{\mu}_y, \cdot$ and $\boldsymbol{\mu}_y|\boldsymbol{\mu}_d, \cdot$, i.e. all states simultaneously for duration and absolute price change, respectively. Following this idea, and first considering sampling latent duration modes $\boldsymbol{\mu}_d$, note that the innovation error of the transition equation for duration can be written as

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

with

$$\eta_{y,t} = \bar{\psi}_{y,t} - \phi_{yy}\bar{\psi}_{y,t-1} - \phi_{yd}\bar{\psi}_{d,t-1}, \quad \bar{\boldsymbol{\psi}}_t = \boldsymbol{\psi}_t - \boldsymbol{\mu}_{s_t},$$

and σ_d, σ_y, ρ the components of transition error covariance $\boldsymbol{\Sigma}$, Eq. (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 $\boldsymbol{\mu}_d$ can then be expressed as

$$\begin{aligned}
\pi(\boldsymbol{\mu}_d|\cdot) \propto \pi(\boldsymbol{\mu}_d) \times \exp&\Big\{-\frac{\{\bar{\psi}_{d,1} - \hat{\sigma}_{d,\mathrm{ini}}\eta_{y,1}\}^2}{2\bar{\sigma}^2_{d,\mathrm{ini}}}\Big\}\times \\
&\exp\Big\{-\sum_{t=2}^{n}\frac{\{\bar{\psi}_{d,t} - \phi_{dd}\bar{\psi}_{d,t-1} - \phi_{dy}\bar{\psi}_{y,t-1} - \hat{\sigma}_d\eta_{y,t}\}^2}{2\bar{\sigma}_d^2}\Big\} \\
&= \exp\Big\{-\frac{1}{2}(\boldsymbol{\mu}_d - \boldsymbol{\mu}_{0,d})'\boldsymbol{\Sigma}^{-1}_{\mu_0,d}(\boldsymbol{\mu}_d - \boldsymbol{\mu}_{0,d})\Big\} \times \\
&\exp\Big\{-\frac{1}{2}(\boldsymbol{Y}_{\mu_d} - \boldsymbol{X}_{\mu_d}\boldsymbol{\mu}_d)'(\boldsymbol{Y}_{\mu_d} - \boldsymbol{X}_{\mu_d}\boldsymbol{\mu}_d)\Big\},
\end{aligned}$$

[5]One may sample $\boldsymbol{\kappa}_y$ 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.

with

$$
\begin{aligned}
x_{\mu_{d,t}} &= \bar{\sigma}_d^{-1} \times \\
&\qquad \left[\mathrm{I}_{\mathrm{s_t=1}} - (\phi_{dd} - \hat{\sigma}_d \phi_{yd}) \mathrm{I}_{\mathrm{s_{t-1}=1}} \quad \cdots \quad \mathrm{I}_{\mathrm{s_t=M_d}} - (\phi_{dd} - \hat{\sigma}_d \phi_{yd}) \mathrm{I}_{\mathrm{s_{t-1}=M_d}} \right], \\
y_{\mu_{d,t}} &= \bar{\sigma}_d^{-1} (\psi_{d,t} - \phi_{dd} \psi_{d,t-1} - \phi_{dy} \bar{\psi}_{y,t-1} - \hat{\sigma}_d (\bar{\psi}_{y,t} - \phi_{yy} \bar{\psi}_{y,t-1} - \phi_{yd} \psi_{d,t-1})), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad t = 2, \dots, n, \\
x_{\mu_{d,1}} &= \bar{\sigma}_{d,\mathrm{ini}}^{-1} \left[\mathrm{I}_{\mathrm{s_1=1}} \quad \cdots \quad \mathrm{I}_{\mathrm{s_1=M_d}} \right], \quad y_{\mu_{d,1}} = \bar{\sigma}_{d,\mathrm{ini}}^{-1} (\psi_{d,1} - \hat{\sigma}_{d,\mathrm{ini}} \bar{\psi}_{y,1})
\end{aligned}
$$

the row elements of $\boldsymbol{X}_{\mu_d}$, $\boldsymbol{Y}_{\mu_d}$, and

$$
\bar{\sigma}_d = \sigma_d \sqrt{1 - \rho^2}, \quad \hat{\sigma}_d = \frac{\sigma_d}{\sigma_y} \rho \quad (\bar{\sigma}_{d,\mathrm{ini}},\ \hat{\sigma}_{d,\mathrm{ini}} \text{ accordingly}).
$$

Then we have

$$
\boldsymbol{\mu}_d | \cdot \backsim \mathcal{N}(\boldsymbol{\mu}_{\mu_d}, \boldsymbol{\Sigma}_{\mu_d}),
$$

where

$$
\boldsymbol{\mu}_{\mu_d} = \boldsymbol{\Sigma}_{\mu_d} (\boldsymbol{\Sigma}_{\mu_0,d}^{-1} \boldsymbol{\mu}_{0,d} + \boldsymbol{X}'_{\mu_d} \boldsymbol{Y}_{\mu_d}), \quad \boldsymbol{\Sigma}_{\mu_d} = (\boldsymbol{\Sigma}_{\mu_0,d}^{-1} + \boldsymbol{X}'_{\mu_d} \boldsymbol{X}_{\mu_d})^{-1}.
$$

Sampling of $\boldsymbol{\mu}_y | \boldsymbol{\mu}_d, \cdot$ 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 $\boldsymbol{\mu}$, see Sec. 6.3.

Step 11. Sampling of $\boldsymbol{\Sigma}$ is by the Metropolis-Hastings (MH) algorithm (see e.g Chib and Greenberg, 1995), accounting for the initial condition. The conditional posterior of $\boldsymbol{\Sigma}$ is

$$
\pi(\boldsymbol{\Sigma} | \cdot) \propto \mathrm{c}(\boldsymbol{\Sigma}) \times |\boldsymbol{\Sigma}|^{-\frac{\nu+3}{2}} \exp\Big\{ -\frac{1}{2} \sum_{t=2}^{n} \boldsymbol{v}_t' \boldsymbol{\Sigma}^{-1} \boldsymbol{v}_t \Big\} \tag{8}
$$

$$
= \mathrm{c}(\boldsymbol{\Sigma}) \times |\boldsymbol{\Sigma}|^{-\frac{\nu+3}{2}} \exp\Big\{ -\frac{1}{2} \mathrm{tr}\left(\boldsymbol{V}^{-1} \boldsymbol{\Sigma}^{-1} \right) \Big\}, \tag{9}
$$

where

$$
\mathrm{c}(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}_{\mathrm{ini}}|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2} \bar{\boldsymbol{\psi}}_1' \boldsymbol{\Sigma}_{\mathrm{ini}}^{-1} \bar{\boldsymbol{\psi}}_1 \Big\}, \tag{10}
$$

$$
\boldsymbol{V}^{-1} = \boldsymbol{V}_0^{-1} + \sum_{t=2}^{n} \boldsymbol{v}_t \boldsymbol{v}_t', \quad \boldsymbol{v}_t = \bar{\boldsymbol{\psi}}_t - \boldsymbol{\Phi} \bar{\boldsymbol{\psi}}_{t-1}, \quad \nu = \nu_0 + n - 1. \tag{11}
$$

A proposal $\boldsymbol{\Sigma}^* \backsim [\mathcal{W}(\nu, \boldsymbol{V})]^{-1}$ is accepted with probability $\min\{c(\boldsymbol{\Sigma}^*)/c(\boldsymbol{\Sigma}), 1)\}$, where $\boldsymbol{\Sigma}$ is the current sample.

Step 12. Similar to Step 11, sampling of $\boldsymbol{\Phi}$ is by the MH algorithm. Following linear regression theory, the conditional posterior probability is

$$
\begin{aligned}
\pi(\boldsymbol{\Phi} | \cdot) &\propto \mathrm{c}(\boldsymbol{\Phi}) \times \exp\Big\{ -\frac{1}{2} \sum_{t=2}^{n} \boldsymbol{v}_t' \boldsymbol{\Sigma}^{-1} \boldsymbol{v}_t \Big\} \\
&= \mathrm{c}(\boldsymbol{\Phi}) \times \exp\Big\{ -\frac{1}{2} \mathrm{tr}\left((\boldsymbol{Y}_{\Phi} - \boldsymbol{X}_{\Phi} \boldsymbol{\Phi}')' (\boldsymbol{Y}_{\Phi} - \boldsymbol{X}_{\Phi} \boldsymbol{\Phi}') \boldsymbol{\Sigma}^{-1} \right) \Big\},
\end{aligned}
$$

where $\mathrm{c}(\boldsymbol{\Phi}) = \mathrm{c}(\boldsymbol{\Sigma})$ of Eq. (10), $\boldsymbol{v}_t$ is defined as in Eq. (11), and

$$\boldsymbol{X}_\Phi = \bar{\boldsymbol{\psi}}_{1:n-1}, \quad \boldsymbol{Y}_\Phi = \bar{\boldsymbol{\psi}}_{2:n}.$$

Then propose a candidate

$$\boldsymbol{\Phi}^* \backsim \mathcal{TN}_R(\boldsymbol{\mu}_\Phi, \boldsymbol{\Sigma}_\Phi),$$

with

$$\boldsymbol{\mu}_\Phi = \mathrm{vec}\left((\boldsymbol{X}_\Phi' \boldsymbol{X}_\Phi)^{-1}(\boldsymbol{X}_\Phi' \boldsymbol{Y}_\Phi)\right), \quad \boldsymbol{\Sigma}_\Phi = \boldsymbol{\Sigma} \otimes (\boldsymbol{X}_\Phi' \boldsymbol{X}_\Phi)^{-1},$$

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

4 Particle Filter

A stratified auxiliary particle filter is developed that recursively delivers draws of the unobservable components $\{\boldsymbol{\alpha}_t = \boldsymbol{\psi}_t - \boldsymbol{\mu}_{s_t}, s_{d,t}, s_{y,t}\}$, given parameter values $\boldsymbol{\theta}$ and observables up to time t, $\{\boldsymbol{d}_{1:t}, \boldsymbol{y}_{1:t}\}$. 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 $\boldsymbol{z}_t = (d_t, y_t)'$, $\boldsymbol{z}_{1:t} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t)'$, and define the extended state set $\boldsymbol{x}_t = \{\boldsymbol{\alpha}_t, s_{d,t}, s_{y,t}\}$ (for better readability, time invariant parameter set $\boldsymbol{\theta}$ 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

$$\begin{aligned} f(\boldsymbol{z}_t|\boldsymbol{x}_t) = f(d_t|\boldsymbol{x}_t)f(y_t|\boldsymbol{x}_t) = & \qquad (12) \\ \mathcal{GG}\Big(\exp\left[\alpha_{d,t} + \mu_{d,s_{d,t}}\right], \zeta, \xi\Big) \times \mathcal{NB}\Big(\exp\left[\alpha_{y,t} + \mu_{y,s_{y,t}} + \boldsymbol{\kappa}_y \boldsymbol{J}_y\right], \tau\Big) \end{aligned}$$

and transition densities

$$\begin{aligned} f(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}) &= \mathcal{N}(\boldsymbol{\Phi}\boldsymbol{\alpha}_{t-1}, \boldsymbol{\Sigma}), \\ \Pr(s_{k,t}|s_{k,t-1}) &= \mathrm{Multinomial}(\boldsymbol{p}_{k,s_{t-1}}), \quad k = \{d, y\}, \end{aligned} \qquad (13)$$

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

$$f(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t}) \propto f(\boldsymbol{z}_t|\boldsymbol{x}_t) f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) f(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1}), \qquad (14)$$

with

$$f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \propto f(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}) \Pr(s_{d,t}|s_{d,t-1}) \Pr(s_{y,t}|s_{y,t-1}),$$

where we assume that we have samples (particles) from $f(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1})$ and a discrete uniform approximation $\hat{f}(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1})$ to $f(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1})$.

By including $\boldsymbol{z}_t$ 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. (14) then has the following form,

$$\begin{aligned}
&g(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}^{(i)}|\boldsymbol{z}_{1:t}) \\
&\quad \propto f(\boldsymbol{z}_t|\tilde{\boldsymbol{\alpha}}_t^{(i)}, s_{d,t}, s_{y,t}) f(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}^{(i)}) \mathrm{Pr}(s_{d,t}|s_{d,t-1}^{(i)}) \mathrm{Pr}(s_{y,t}|s_{y,t-1}^{(i)}) \hat{f}(\boldsymbol{x}_{t-1}^{(i)}|\boldsymbol{z}_{1:t-1}) \\
&\quad \propto f(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}^{(i)}) g(\boldsymbol{x}_{t-1}^{(i)}, s_{d,t}, s_{y,t}|\boldsymbol{z}_{1:t}, \tilde{\boldsymbol{\alpha}}_t^{(i)}),
\end{aligned}$$

with

$$\begin{aligned}
&g(\boldsymbol{x}_{t-1}^{(i)}, s_{d,t}, s_{y,t}|\boldsymbol{z}_{1:t}, \tilde{\boldsymbol{\alpha}}_t^{(i)}) \\
&\quad \propto f(\boldsymbol{z}_t|\tilde{\boldsymbol{\alpha}}_t^{(i)}, s_{d,t}, s_{y,t}) \mathrm{Pr}(s_{d,t}|s_{d,t-1}^{(i)}) \mathrm{Pr}(s_{y,t}|s_{y,t-1}^{(i)}) \hat{f}(\boldsymbol{x}_{t-1}^{(i)}|\boldsymbol{z}_{1:t-1})
\end{aligned} \tag{15}$$

and "best" guess

$$\tilde{\boldsymbol{\alpha}}_t^{(i)} = \boldsymbol{\Phi}\boldsymbol{\alpha}_{t-1}^{(i)},$$

where superscripts $i = 1, \ldots, I$ denote particles. Observe that no best guess is made for $s_{d,t}$ and $s_{y,t}$. Instead, $g(\boldsymbol{x}_{t-1}^{(i)}, s_{d,t}, s_{y,t}|\boldsymbol{z}_{1:t}, \tilde{\boldsymbol{\alpha}}_t^{(i)})$, Eq. (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, $\{\boldsymbol{x}_{t-1}, s_{d,t}, s_{y,t}\}$ are drawn jointly from the respective distribution. Accordingly, there is no direct sampling from $\mathrm{Pr}(s_{k,t}|s_{k,t-1}^{(i)})$, $k = \{d, y\}$. The proposed particle filter follows:

1. Initialization, $t = 1$ $(i = 1, \ldots, I,\ k = \{d, y\})$:

   (a) Draw $s_{k,1}^{(i)} \backsim \mathrm{Multinomial}(\boldsymbol{p}_{k,1})$, with $\boldsymbol{p}_{k,1}$ the initial distribution of $\boldsymbol{P}_k$.

   (b) Draw $\boldsymbol{\alpha}_1^{(i)} \backsim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathrm{ini}})$ from the unconditional distribution.

   (c) Compute $w_1^{(i)} = f(\boldsymbol{z}_1|\boldsymbol{x}_1^{(i)})$, and let $\hat{f}(\boldsymbol{x}_1^{(i)}|\boldsymbol{z}_1) = w_1^{(i)} / \sum_{j=1}^{I} w_1^{(j)}$.

2. Iterate, $t = 2, \ldots, n$ $(i, j = 1, \ldots, I)$:

   (a) Generate $\{\boldsymbol{x}_t^{(i)}, \boldsymbol{x}_{t-1}^{(i)}\}$ from $g(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}^{(j)}|\boldsymbol{z}_{1:t})$, with $\boldsymbol{x}_{t-1}^{(j)}$ 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),

   $$w_{t-1}^{(j,l,m)} \propto g(\boldsymbol{x}_{t-1}^{(j)}, s_{d,t} = l, s_{y,t} = m|\boldsymbol{z}_{1:t}, \tilde{\boldsymbol{\alpha}}_t^{(j)}),$$

   and resample $\{\boldsymbol{x}_{t-1}^{(j)}, l, m, w_{t-1}^{(j,l,m)}\}_{(j,l,m)\in\{1,\ldots,I\}\times\{1,\ldots,M_d\}\times\{1,\ldots,M_y\}}$ to obtain $\{\boldsymbol{x}_{t-1}^{(i)}, s_{d,t}^{(i)}, s_{y,t}^{(i)}, I^{-1}\}$. Then generate $\boldsymbol{\alpha}_t^{(i)}$ from $f(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}^{(i)})$, Eq. (13).

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

   $$w_t^{(i)} = \frac{f(\boldsymbol{z}_t|\boldsymbol{x}_t^{(i)})}{f(\boldsymbol{z}_t|\tilde{\boldsymbol{\alpha}}_t^{(i)}, s_{d,t}^{(i)}, s_{y,t}^{(i)})},$$

   and let $\hat{f}(\boldsymbol{x}_t^{(i)}|\boldsymbol{z}_{1:t}) = w_t^{(i)} / \sum_{j=1}^{I} w_t^{(j)}$.

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

1. Resample from $\hat{f}(\boldsymbol{x}_t^{(j)}|\boldsymbol{z}_{1:t})$ after Step 1.(c) and 2.(b) to obtain $\boldsymbol{x}_{t|t}^{(i)}$.
2. Generate $\boldsymbol{\alpha}_{t+1|t}^{(i)}$ from $f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t^{(i)})$, Eq. (13).
3. Estimate the one-step-ahead prediction density by averaging the mixture

$$\hat{f}(\boldsymbol{z}_{t+1}|\boldsymbol{z}_{1:t}, \boldsymbol{\theta}) = \tag{16}$$
$$I^{-1} \sum_{i=1}^{I} \Big(\sum_{l=1}^{M_d} \Pr(s_{d,t+1} = l | s_{d,t}^{(i)}) f(d_{t+1}|\boldsymbol{\alpha}_{t+1|t}^{(i)}, s_{d,t+1} = l) \Big) \times$$
$$\Big(\sum_{m=1}^{M_y} \Pr(s_{y,t+1} = m | s_{y,t}^{(i)}) f(y_{t+1}|\boldsymbol{\alpha}_{t+1|t}^{(i)}, s_{y,t+1} = m) \Big).$$

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_{t+1}^o can be estimated as $F(d_{t+1} \leq d_{t+1}^o|\boldsymbol{z}_{1:t}, \boldsymbol{\theta}) \cong \hat{u}_{d,t+1}$, which is approximated in analogy to Eq. (16). $F(\cdot)$ denotes the cumulative distribution function. Under the null hypothesis of a correctly specified model, observed durations d_t^o are random draws from $f(d_t|\cdot)$, compare Eq. (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 $u_{t+1}^{01} \backsim \mathcal{U}(0, 1)$, and replace Step 3 above $(i = 1, \ldots, I,\ m = 1, \ldots, M_y)$:

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

$$u_{y,t+1,s_{y,t+1}=m}^{\text{up},(i)} = F(y_{t+1} \leq y_{t+1}^o | \alpha_{y,t+1|t}^{(i)}, s_{y,t+1} = m).$$

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

$$u_{y,t+1,s_{y,t+1}=m}^{\text{low},(i)} = F(y_{t+1} \leq y_{t+1}^o - 1 | \alpha_{y,t+1|t}^{(i)}, s_{y,t+1} = m).$$

3.3 Average the mixture

$$F(y_{t+1} \leq y^o_{t+1}|\boldsymbol{z}_{1:t}, \boldsymbol{\theta}) \cong \hat{u}_{y,t+1} = \\ I^{-1} \sum_{i=1}^{I} \sum_{m=1}^{M_y} \Pr(s_{y,t+1} = m|s^{(i)}_{y,t}) \times \\ \Big[u^{\text{low},(i)}_{y,t+1,s_{y,t+1}=m} + u^{01}_{t+1}\big(u^{\text{up},(i)}_{y,t+1,s_{y,t+1}=m} - u^{\text{low},(i)}_{y,t+1,s_{y,t+1}=m}\big)\Big].$$

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 $I \to \infty$.

5 Adaptive Metropolis for the Generalized Gamma Shape Parameters

From a Bayesian perspective, the shape parameters of the generalized gamma ($\mathcal{GG}$) 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 ζ, ξ are compared in this section.

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

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

where independent log-normal priors are assumed,

$$\zeta \backsim \mathcal{LN}(-7, 49), \quad \xi \backsim \mathcal{LN}(-7, 49), \tag{17}$$

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. C for technical details. It must be noted that in order to estimate the $\mathcal{GG}$ 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 ζ, ξ fix. Datasets have empirically relevant size $60{,}000$. In each application of the algorithm, 100 datasets

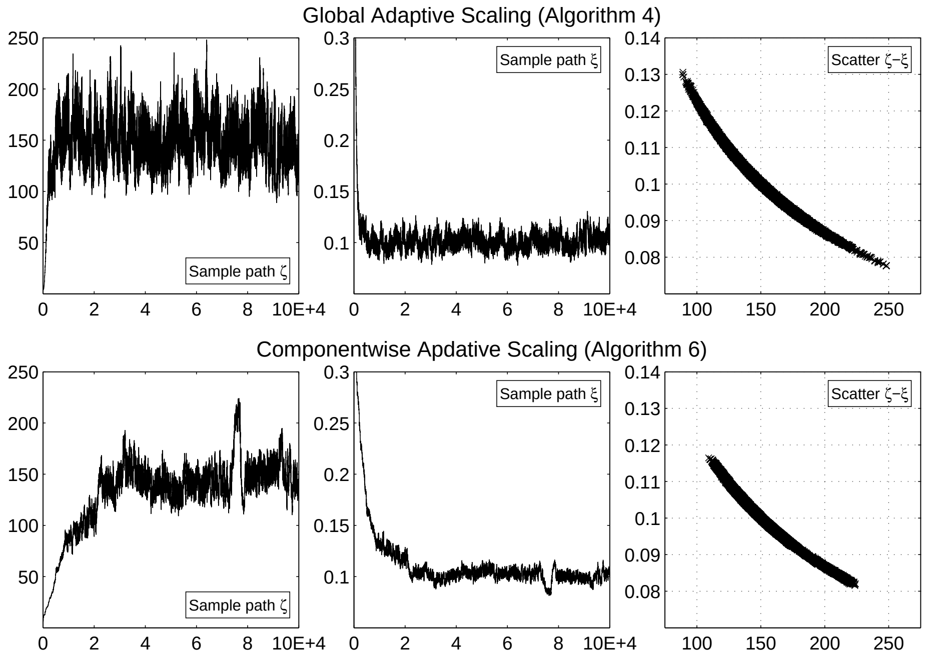

Figure 1: Comparison of AM Algorithms for $\mathcal{GG}$ Parameter Estimation

Table 2: Simulation Results of the Generalized Gamma Shape Parameters

		True	Mean	Stdev	Min	Max	Avg. IF	Avg. CD
Algo 4	ζ	150	152.80	34.581	92.742	249.73	157.9	0.79
	ξ	0.1	0.1019	0.0109	0.0784	0.1280	150.9	0.80
Algo 6	ζ	150	152.00	37.308	92.640	265.68	1130.8	0.49
	ξ	0.1	0.1021	0.0116	0.0765	0.1275	1120.9	0.48

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.

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.

Tab. 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. D 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 suggest.

Fig. 1 illustrates performance of the two algorithms and the parameter identification problem. The first to columns show sample paths of generalized gamma shape parameters ζ, ξ 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.

6 Empirical Application

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 29$^{\text{th}}$ 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. 3 presents descriptive statistics of the dataset. Fig. 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.

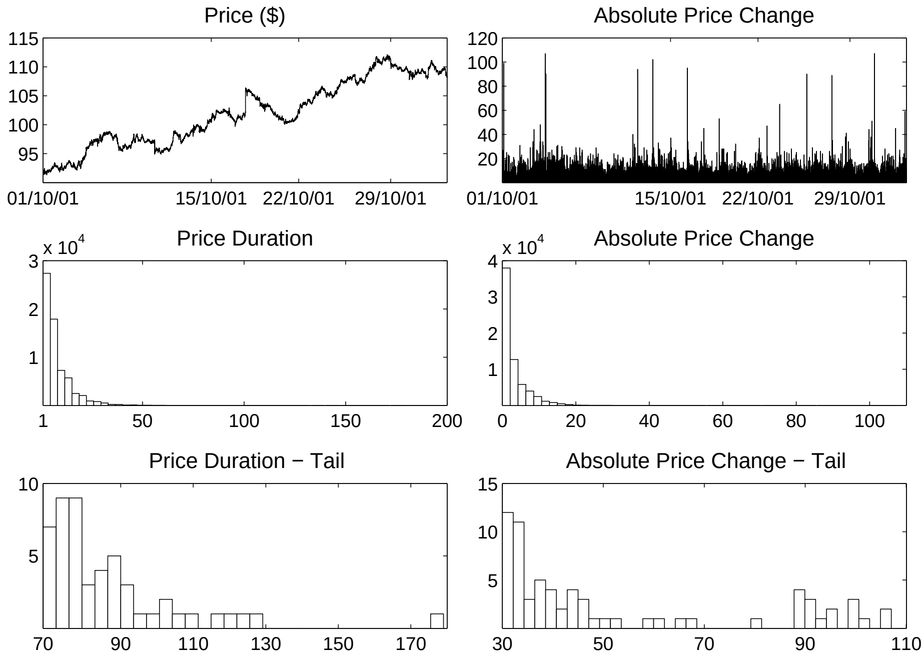

Figure 2: Descriptive Statistics of the IBM stock (NYSE), 10/2001 (65,889 obs.). Price duration in seconds. Abs. price change in half-ticks, threshold $0 = 0.5$ ticks.

Table 3: Descriptive Statistics of the IBM Stock (NYSE), 10/2001 (65,889 obs.)

	Duration	Abs. Price Δ		Duration	Abs. Price Δ
Mean	7.80	4.04	Stdev	7.58	4.04
Mode	3	1	Min	1	1
$\% \leq$ Mode	32.2%	29.0%	Max	179	108

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

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 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. 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. 3. The reason for not pre-adjusting absolute price change is preservation of discreteness, i.e.

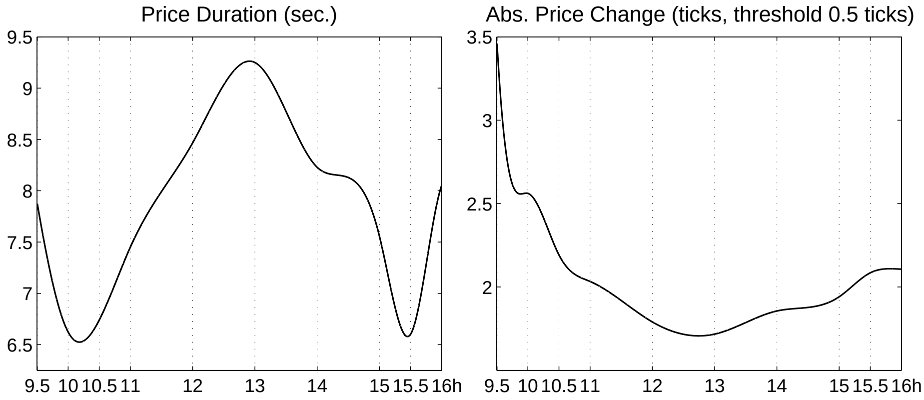

Figure 3: Diurnal Components of the IBM Stock, 10/2001 (cubic spline)

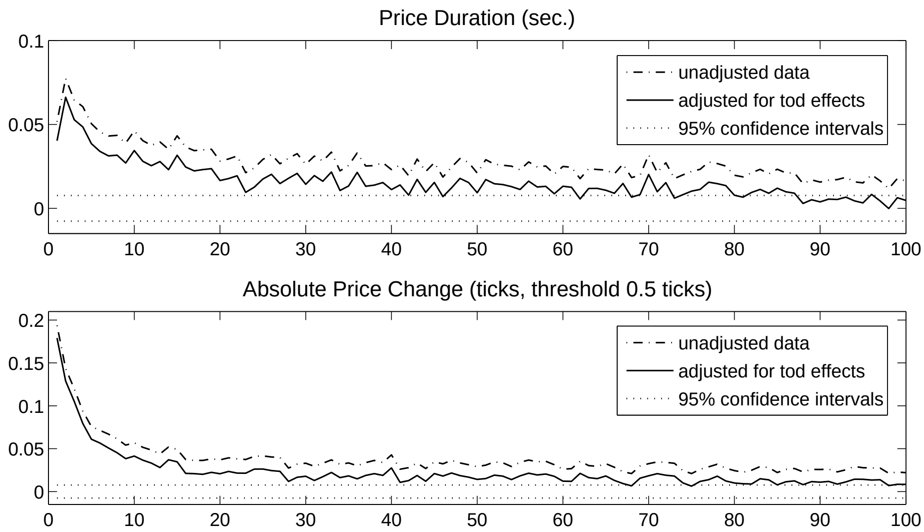

Figure 4: Spearman's ρ for IBM, averaged over trading days 10/2001 (100 lags)

the tick property. Accordingly, diurnal components are estimated jointly with the model for this series.

Finally, Fig. 4 demonstrates the presence of serial dependence in duration and absolute price change using Spearman's ρ (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.

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

$$\mu_d \backsim \mathcal{N}(0, 10^2), \quad \mu_y \backsim \mathcal{N}(1.5, 10^2), \quad \kappa_y^j \backsim \mathcal{N}(0, 10^2), \quad j = \{1, \dots, 8\},$$
$$\frac{\phi_{kk}+1}{2} \backsim \mathcal{B}(20, 1.5), \quad \frac{\phi_{kl}+1}{2} \backsim \mathcal{B}(1, 1), \quad k, l = \{d, y\}, \quad l \neq k,$$
$$\boldsymbol{\Sigma}^{-1} \backsim \mathcal{W}(5, (5\boldsymbol{\Sigma}_0)^{-1}) \text{ with } \boldsymbol{\Sigma}_0 = 0.2^2 \boldsymbol{I}_2,$$
$$\zeta \backsim \mathcal{LN}(-7, 49), \quad \xi \backsim \mathcal{LN}(-7, 49), \quad \tau \backsim \mathcal{LN}(-7, 49),$$

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 μ_d, μ_y are mildly informative reflecting empirical relevant values. Those for time-of-day dummies κ_y^j are conservatively centered around zero. The diagonal components of $\boldsymbol{\Phi}$ 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 $\boldsymbol{\Sigma}^{-1}$ is chosen such that $\mathrm{E}(\boldsymbol{\Sigma}^{-1}) = \boldsymbol{\Sigma}_0^{-1}$, conservatively assuming zero correlation. Distributional shape parameters ζ, ξ, τ 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. D for the deployed formulae). For both 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.

Average AR/MH acceptance rates of latent information flow process $\boldsymbol{\psi}_t$ in the multi-move sampler are high, 99.5% and 99.4%, respectively, as are those for the MH steps sampling $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$, which are 94.3% and 98.3%, respectively. Acceptance rate for the adaptive Metropolis algorithm estimating $\mathcal{GG}$ shape parameters ζ, ξ is with 35.0% close to the theoretical optimal value of $\approx 35\%$ (two dimensions). The same holds for acceptance rates of $\mathcal{NB}$ shape parameter τ and dummy variables $\boldsymbol{\kappa}_y$, 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. 4 reports results. Both information flow processes feature persistence, high for duration, $\phi_{dd} = 0.941$, and more moderate for absolute price change, $\phi_{yy} = 0.893$. Cross effects between the latent processes as captured by $\boldsymbol{\Phi}$ 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

Table 4: Estimation Results of the SCDSV Model

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

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

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.[6] Moreover, ϕ_{yd} is about four times higher in size than ϕ_{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, the concurrent dynamics generated by information signal η_t contrast the intertemporal cross effect forces expressed by ϕ_{dy}, ϕ_{yd}.

Shape parameters ζ, ξ 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 ζ, visualizing the difficulties in estimating the $\mathcal{GG}$ shape parameters. Further support for the $\mathcal{GG}$ distribution comes from Lunde (1999), applying it to the ACD class, and from Bauwens et al. (2004), emphasizing strong predictive performance. Inferred tod dummies $\boldsymbol{\kappa}_y$ clearly reflect the diurnal patterns shown in Fig. 3 and are all significant. Fig. 5 plots observation densities at their unconditional mean, where average time-of-day effect is taken into account. Inefficiency factors are in acceptable

[6] The author also provides further links to traditional market microstructure theory in the context of trade duration/intensity.

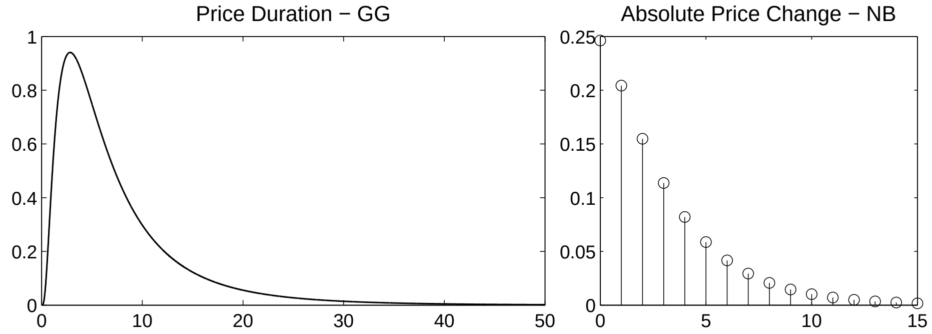

Figure 5: Unconditional Densities of the SCDSV Model. Price duration generalized gamma distributed (sec.), rescaled by average tod effect. Absolute price change negative binomial distributed (half-ticks, threshold $0 = 0.5$ ticks), avg. tod effect.

ranges regarding the number of collected draws. Only a high inefficiency for σ_d stands out, which is linked to the small magnitude of this parameter.

As already mentioned in Sec. 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.

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. 6.5). The deviance information criterion (DIC) (Sec. 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 have been identified from the data during the model fitting process. Priors are assumed as in Sec. 6.2, except as follows,

$$\mu_d \backsim \mathcal{N}((-1.1, -0.1, 0.8)', \boldsymbol{I}_3), \quad \mu_y \backsim \mathcal{N}((-1.8, 0.3, 1.8, 4.2)', \boldsymbol{I}_4),$$
$$\boldsymbol{P}_{k,i\cdot} \backsim \mathcal{D}(1, \dots, 1), \quad i = \{1, \dots, M_k\}, \quad k = \{d, y\}.$$

The priors on latent duration and absolute price change modes $\boldsymbol{\mu}_k$ are empirically relevant and mildly informative to reflect the specific state identification constraints ($\mu_{d,1} < \mu_{d,2} < \mu_{d,3}$, $\mu_{y,1} < \mu_{y,2} < \mu_{y,3} < \mu_{y,4}$), which prevent label switching (see also Sec. 3, Step 8). Priors for the rows of $\boldsymbol{P}_k$ 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. 6.2. AR/MH acceptance rates of the actual run are also comparable.

Table 5: Estimation Results of the SCDSVrs Model

	Mean	Stdev	95% c.b.	IF	CD
$\mu_{d,1}$	-1.0741	0.0247	[-1.1214, -1.0274]	237.1	0.67
$\mu_{d,2}$	-0.0992	0.0252	[-0.1465, -0.0521]	247.0	0.63
$\mu_{d,3}$	0.7797	0.0191	[0.7426, 0.8188]	145.1	0.74
$\mu_{y,1}$	-1.8169	0.1342	[-2.0694, -1.5341]	1,187.1	0.15
$\mu_{y,2}$	0.2727	0.0740	[0.1291, 0.4170]	428.3	0.45
$\mu_{y,3}$	1.7444	0.0335	[1.6779, 1.8088]	109.1	0.82
$\mu_{y,4}$	4.1923	0.1378	[3.9212, 4.4592]	135.5	0.89
ϕ_{dd}	0.9050	0.0098	[0.8860, 0.9241]	346.9	0.90
ϕ_{dy}	-0.0266	0.0033	[-0.0330, -0.0203]	118.4	0.89
ϕ_{yd}	-0.0861	0.0065	[-0.0988, -0.0733]	125.8	0.99
ϕ_{yy}	0.9179	0.0060	[0.9056, 0.9295]	156.6	0.89
σ_d	0.1099	0.0079	[0.0937, 0.1248]	1,936.8	0.92
σ_y	0.1912	0.0094	[0.1726, 0.2106]	1,071.8	0.95
ρ	0.6549	0.0295	[0.5935, 0.7102]	400.0	0.80
ζ	3.9602	0.0912	[3.7888, 4.1427]	321.1	0.81
τ	3.7623	0.1843	[3.4223, 4.1422]	172.3	0.85
$\kappa_{y,2}$	-0.1335	0.0440	[-0.2176, -0.0475]	237.9	0.95
$\kappa_{y,3}$	-0.3251	0.0450	[-0.4149, -0.2379]	212.9	0.98
$\kappa_{y,4}$	-0.4881	0.0396	[-0.5646, -0.4096]	280.2	0.85
$\kappa_{y,5}$	-0.5757	0.0424	[-0.6586, -0.4912]	269.0	0.96
$\kappa_{y,6}$	-0.5503	0.0415	[-0.6318, -0.4681]	278.5	0.91
$\kappa_{y,7}$	-0.4644	0.0404	[-0.5442, -0.3866]	277.1	0.84
$\kappa_{y,8}$	-0.3712	0.0454	[-0.4613, -0.2838]	203.2	0.91
$\kappa_{y,9}$	-0.3216	0.0452	[-0.4103, -0.2329]	202.8	0.93

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor and Geweke's convergence diagnostic (p-value). Transition probabilities in App. B, Tab. 9 and 10.

Estimation Results are reported in Tab. 5. Transition probability matrices are given in Eq. (18) and (19), together with average observations in each state. Complete statistics of the former have been put in App. B, Tab. 9 and 10. Latent duration and absolute price change mode parameters $\boldsymbol{\mu}_k$ differentiate the regimes significantly. Note a distinctively high absolute price change mode $\mu_{y,4}$ reflecting the price jump state. Fig. 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, mode of state 2 is also located at 0.5 ticks, with a significant point mass of 44.5%.

Taking a closer look at the transition probability matrices in Eq. (18) and (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, 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

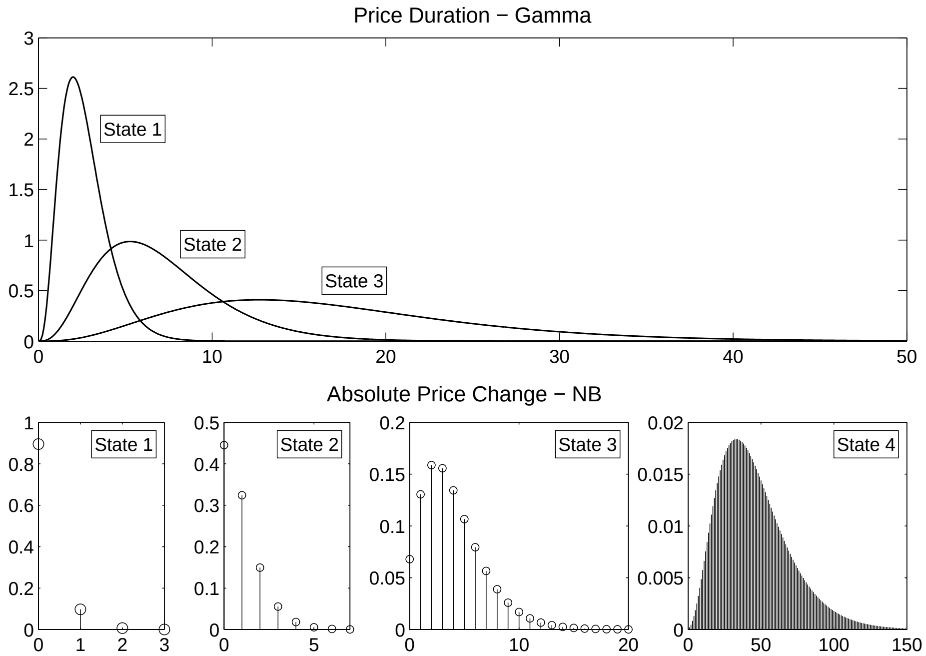

Figure 6: Unconditional Densities of the SCDSVrs Model. Price duration gamma distributed (sec.), rescaled by average tod effect. Absolute price change negative binomial distributed (half-ticks, threshold $0 = 0.5$ ticks), average tod effect.

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.

$$
\begin{array}{cc}
\text{cond. mean} & \begin{array}{ccc} \text{low} & & \text{high} \end{array} \\
 & s_{d,t} \\
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
s_{d,t-1} & \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{pmatrix} \mathbf{0.2315} & 0.5703 & 0.1982 \\ 0.3257 & \mathbf{0.5217} & 0.1525 \\ 0.3147 & 0.4463 & \mathbf{0.2390} \end{pmatrix} \\
 & \text{avg. observations} \\
 & \begin{array}{ccc} 19{,}496.1 & 34{,}422.2 & 11{,}970.7 \\ (29.6\%) & (52.2\%) & (18.2\%) \end{array}
\end{array}
\tag{18}
$$

$$
\begin{array}{cc}
\text{cond. mean} & \begin{array}{cccc} \text{low} & & & \text{high} \end{array} \\
 & s_{y,t} \\
 & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
s_{y,t-1} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} \mathbf{0.2942} & 0.0372 & 0.6680 & 0.00058 \\ 0.0443 & \mathbf{0.7052} & 0.2502 & 0.00028 \\ 0.1558 & 0.0872 & \mathbf{0.7564} & 0.00053 \\ 0.0688 & 0.1212 & 0.3372 & \mathbf{0.4727} \end{pmatrix} \\
 & \text{avg. observations} \\
 & \begin{array}{cccc} 10{,}132.2 & 13{,}708.3 & 41{,}991.9 & 56.6 \\ (15.4\%) & (20.8\%) & (63.7\%) & (0.1\%) \end{array}
\end{array}
\tag{19}
$$

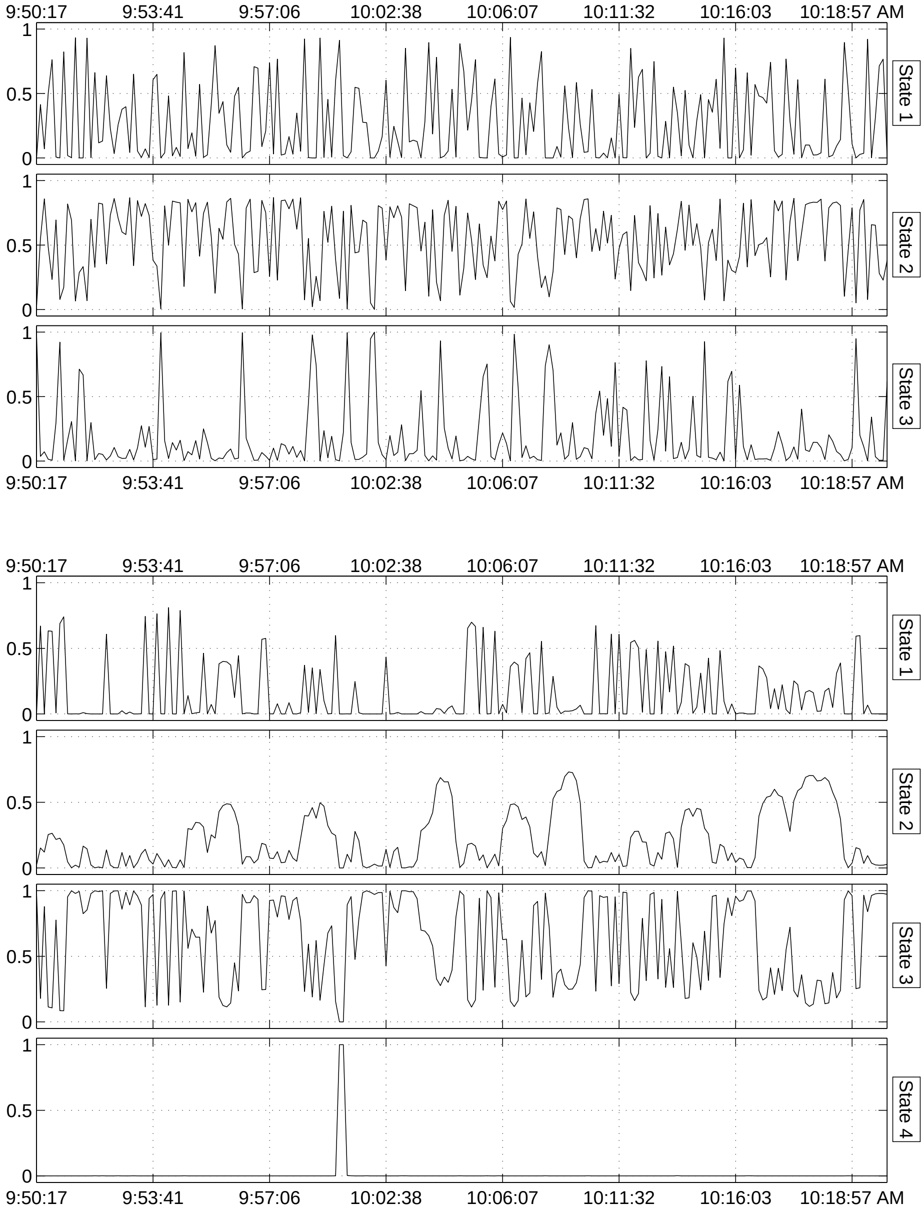

Figure 7: Posterior Probability of States. Upper figure: Duration ($s_{t,d}$). Lower figure: Absolute price change ($s_{t,y}$). 2001/10/01, 9:50-10:20 AM, 220 obs.

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. 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

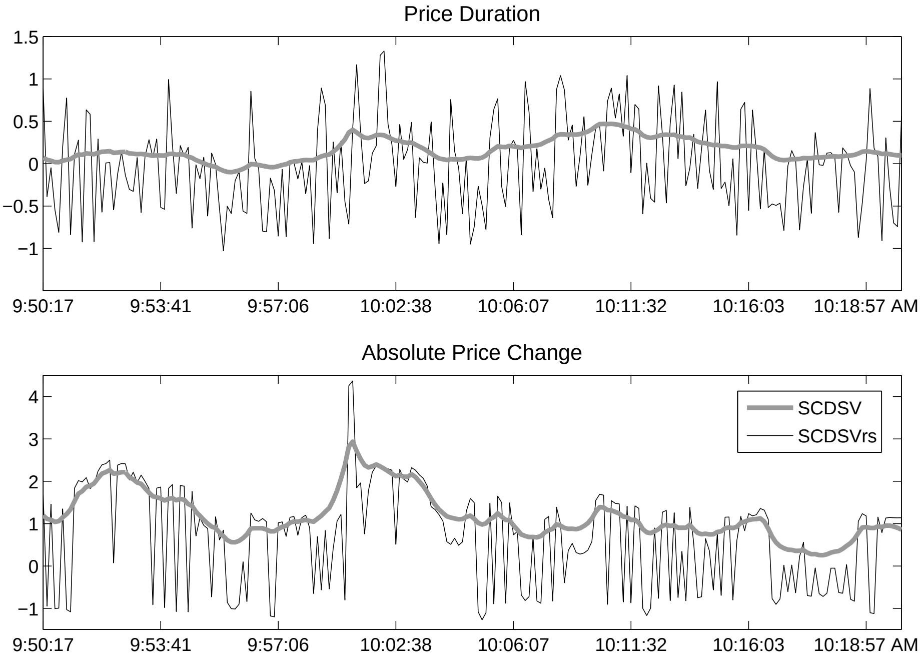

Figure 8: Posterior of Latent Process $\boldsymbol{\psi}_t$ (2001/10/01, 9:50-10:20 AM, 220 obs.)

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 $\boldsymbol{\mu}_y$, 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 τ 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. 8 then shows associated information flow process $\boldsymbol{\psi}_t$ 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 σ_d and high persistence ϕ_{dd}. The SCDSVrs extension, on the contrary, shows an erratic process switching states

Table 6: Model Selection

	DIC	(s.e.)	Rank	p_D	(s.e.)	D_{pred}	(s.e.)	Rank
SCDSV	406,025.4	(92.7)	2	8,532.5	(46.4)	409,975.2	(2.9)	2
SCDSVrs	328,724.2	(1,900.4)	1	34,523.8	(950.2)	407,829.3	(4.5)	1

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

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 σ_y, compared to σ_d, results in a more adaptive flow in the one state case, compared to duration. 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).[7] Note however that the information flow would then be discontinuous, which makes such an extension nontrivial.

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

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

with $D(\boldsymbol{\theta})$ the deviance and $\bar{\boldsymbol{\theta}}$ 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 $D(\boldsymbol{\theta})$ fluctuate more widely around $D(\bar{\boldsymbol{\theta}})$, 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 $D(\bar{\boldsymbol{\theta}})$, the particle filter outlined in Sec. 4 is repeatedly applied 10 times to obtain estimates of log-likelihood ordinate $\log f_{\text{post}}(\boldsymbol{z}_{1:n}|\bar{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \log \hat{f}(\boldsymbol{z}_t|\boldsymbol{z}_{1:t}, \bar{\boldsymbol{\theta}})$. The deviance of the predictive log-likelihood is calculated in a similar way from ordinate $\log f_{\text{prior}}(\boldsymbol{z}_{1:n}|\bar{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \log \hat{f}(\boldsymbol{z}_t|\boldsymbol{z}_{1:t-1}, \bar{\boldsymbol{\theta}})$. The above approach yields 100 different DIC values, from which statistics are calculated.

DIC in Tab. 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

[7] Indeed, 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.

[8] However, as noted already in Sec. 6.3, when interpreting these complexity penalty terms one has to keep in mind the large sample size dealt with.

Table 7: Descriptive Statistics of the IBM Stock 11/2001 (55,652 obs.)

	Duration	Abs. Price Δ		Duration	Abs. Price Δ
mean	8.62	3.76	stdev	9.09	3.74
mode	3	1	min	1	1
$\% \leq$ mode	30.8%	32.2%	max	155	104

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

significant improvement in forecasting ability of the model, as indicated by deviance of the predictive log-likelihood D_{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.

6.5 Predictive Density Tests

Predictive performance of the proposed models is compared using density forecast evaluation techniques (compare Sec. 4 and the references therein). Parameter estimates are those of Sec. 6.2 and 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. 6.1. Descriptive statistics are given in Tab. 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).

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. 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.

Table 8: Predictive Performance: Goodness-of-fit Statistics

		2001-10 Duration	2001-10 Absolute Price Change	2001-11 Duration	2001-11 Absolute Price Change
LB_{15}	Data	1,996.3	11,426.5	2,320.0	7,978.8
	SCDSV	47.6 (0.1)	268.8 (10.1)	49.8 (0.1)	322.0 (9.5)
	SCDSVrs	20.1* (0.1)	33.6 (3.7)	41.5 (0.1)	21.5* (2.4)
LB^2_{15}	Data	247.4	11,104.9	463.0	9,379.9
	SCDSV	157.5 (0.3)	203.9 (18.6)	120.3 (0.3)	273.0 (15.6)
	SCDSVrs	147.9 (0.2)	32.7 (6.3)	110.9 (0.3)	72.3 (12.6)
V	SCDSV	8.83 (0.02)	10.23 (0.29)	11.32 (0.02)	8.37 (0.18)
	SCDSVrs	7.64 (0.02)	1.14* (0.28)	7.99 (0.02)	5.61 (0.10)

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 value for the Kupier statistic V are $2.001/1.747/1.620$ at the 1%/5%/10% level (H_0: $\hat{u}_t \backsim \mathcal{U}(0,1)$). Numerical standard error in parentheses.

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 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. 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. 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. 9. An advantage of the Kupier statistic is that it is equally sensitive 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

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

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

[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.

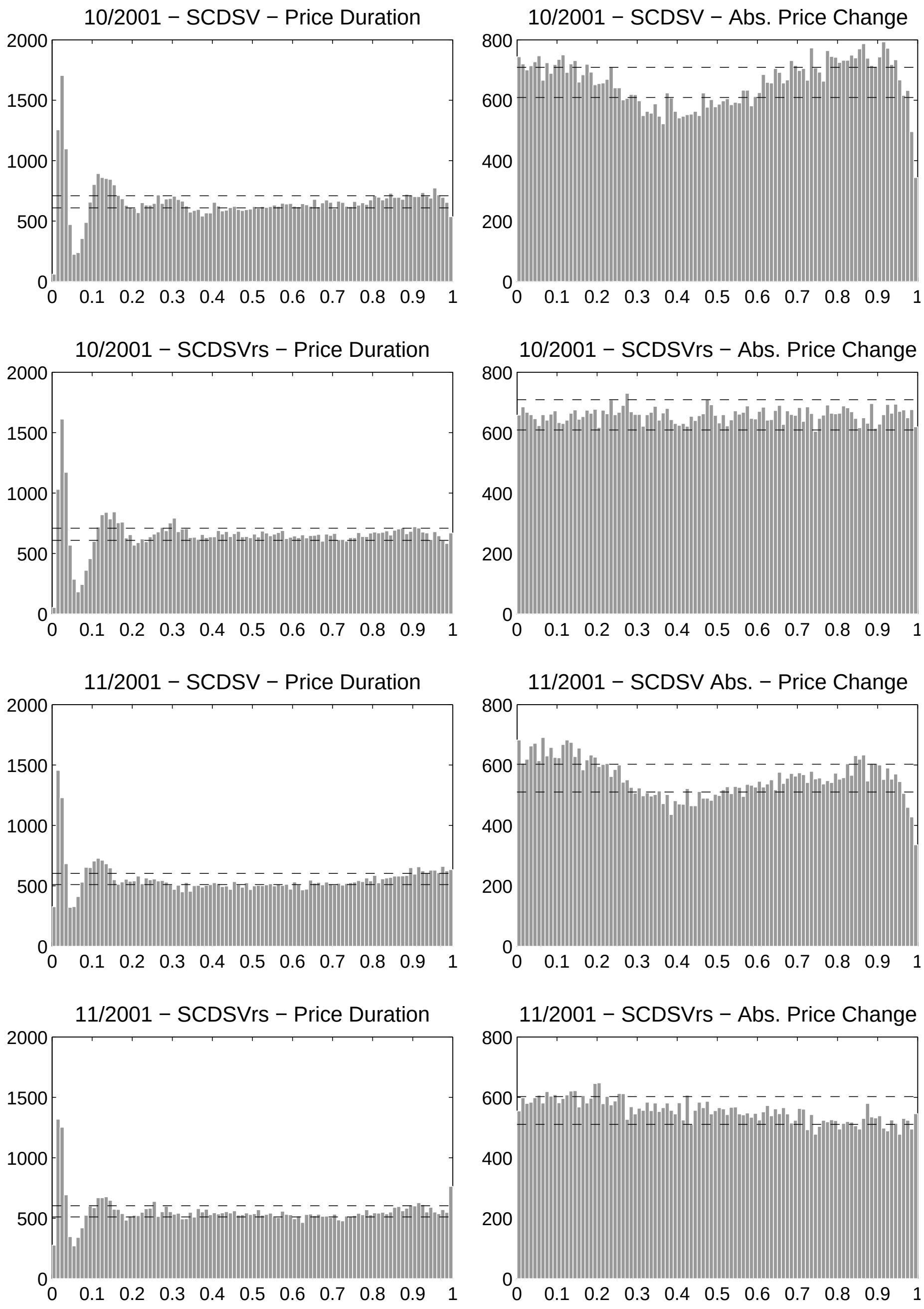

Figure 9: Transformed Predictions (grand averages). Horizontal dashed lines are 95% confidence intervals based on a normal approximation of the binomially distributed bin elements (H_0: $\hat{u}_t \backsim \mathcal{U}(0,1)$).

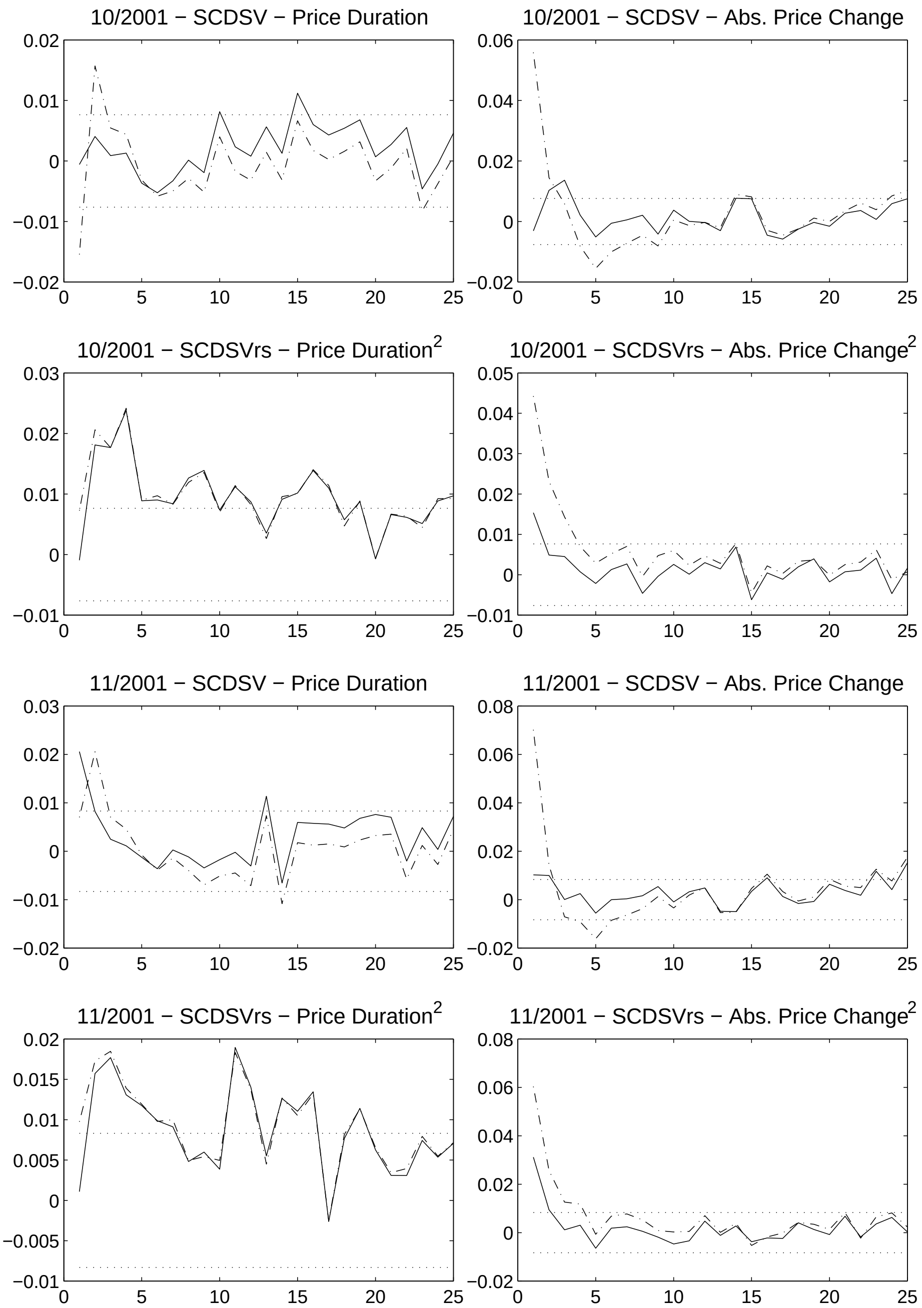

Figure 10: Autocorrelograms of Transformed Predictions $\hat{u}_t$ (1^{st} and 3^{rd} row) and $\hat{u}_t^2$ (2^{nd} and 4^{th} row) (demeaned, grand averages, up to lag 25). Horizontal dashed lines are 95% confidence intervals under the null that $\hat{u}_t$, $\hat{u}_t^2$ are iid.

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_{15}^2 statistic for duration when comparing the raw data and $\hat{u}_{d,t}^2$'s of both models. This once more emphasizes the need to model the second moment of price duration explicitly.

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 $\mathrm{AR}(1)$ persistence matrix. It is interesting that the intertemporal cross effect forces expressed by the latter contrast the concurrent 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. 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.

[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".

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 $\boldsymbol{\psi}_{1:n}$, using a local linear approximation of second order around the mode. By defining $\boldsymbol{\alpha}_t = \boldsymbol{\psi}_t - \boldsymbol{\mu}_{s_t}$, $t = 1, \dots, n$, one obtains the following alternative representation of Eq. (1)-(3) as a state space model w.r.t. $\boldsymbol{\alpha}_{1:n}$,

$$
\begin{aligned}
& d_t \backsim \mathcal{GG}\Big(\exp[\alpha_{d,t} + \mu_{d,s_{d,t}}], \zeta, \xi\Big), \\
& y_t \backsim \mathcal{NB}\Big(\exp[\alpha_{y,t} + \mu_{y,s_{y,t}} + \boldsymbol{\kappa}_y \boldsymbol{J}_y], \tau\Big), && t = 1, \dots, n, \\
& \boldsymbol{\alpha}_t = \boldsymbol{\Phi}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, && t = 2, \dots, n. \qquad (20)
\end{aligned}
$$

To achieve more efficient sampling, $\boldsymbol{\alpha}_{1:n}$ is divided into $K+1$ blocks at random, say $(\boldsymbol{\alpha}_{k_{i-1}+1}, \dots, \boldsymbol{\alpha}_{k_i})'$, $i = 1, \dots, K+1$, 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 $i = 1, \dots, K$ and $k_i - k_{i-1} \geq 1$, with $\mathcal{U}_i$ a random sample from the uniform distribution $\mathcal{U}(0,1)$ (Shephard and Pitt, 1997).

Let $\boldsymbol{x}_t = \boldsymbol{R}_t^{-1}\boldsymbol{\eta}_t$, where $\boldsymbol{R}_t$ denotes a Cholesky decomposition of $\boldsymbol{\Sigma}$ such that $\boldsymbol{\Sigma} = \boldsymbol{R}_t\boldsymbol{R}_t'$, $t > 1$, and of $\boldsymbol{\Sigma}_{\text{ini}}$, $t = 1$, 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 $\boldsymbol{\eta}_t$, block sampling $\underline{\boldsymbol{x}} = (\boldsymbol{x}_{r+1}', \dots, \boldsymbol{x}_{r+c}')'$ instead of $\underline{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}_{r+1}', \dots, \boldsymbol{\alpha}_{r+c}')'$ from the respective full joint conditional posterior density ($r \geq 0$, $c \geq 1$, $r + c \leq n$),

$$
\begin{aligned}
& \pi(\underline{\boldsymbol{x}}|\boldsymbol{\alpha}_{r+c+1}, \boldsymbol{\Theta}) \propto && (21) \\
& \quad \left(\prod_{t=r+1}^{r+c} f_{\mathcal{GG}}(d_t|\alpha_{d,t}, \cdot) f_{\mathcal{NB}}(y_t|\alpha_{y,t}, \cdot)\right)\left(\prod_{t=r+1}^{r+c} f(\boldsymbol{x_t})\right) f(\boldsymbol{\alpha}_{r+c}), \quad r + c < n, \\
& \pi(\underline{\boldsymbol{x}}|\boldsymbol{\Theta}) \propto && (22) \\
& \quad \left(\prod_{t=r+1}^{r+c} f_{\mathcal{GG}}(d_t|\alpha_{d,t}, \cdot) f_{\mathcal{NB}}(y_t|\alpha_{y,t}, \cdot)\right)\left(\prod_{t=r+1}^{r+c} f(\boldsymbol{x_t})\right), \quad r + c = n,
\end{aligned}
$$

with

$$
f(\boldsymbol{\alpha}_{r+c}) = \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}_{r+c+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_{r+c})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha}_{r+c+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_{r+c})\right\}
$$

and $\boldsymbol{\Theta} \equiv \{\boldsymbol{d}_{r+1:r+c}, \boldsymbol{y}_{r+1:r+c}, \boldsymbol{s}_{d,r+1:r+c}, \boldsymbol{s}_{y,r+1:r+c}, \boldsymbol{\theta}\}$. Measurement distributions $f_{\mathcal{GG}}(\cdot)$ and $f_{\mathcal{NB}}(\cdot)$ are given Tab. 1.

To construct a proposal density based on a normal approximation of the posterior density of $\underline{\boldsymbol{x}}$, let $\widehat{\psi}_{y,t} = \alpha_{y,t} + \mu_{y,s_{y,t}} + \boldsymbol{\kappa}_y\boldsymbol{J}_y$, $t = 1, \dots, n$, and start with the logarithm of Eq. (21)-(22) (excluding constant terms), $-\frac{1}{2}\sum_{t=r+1}^{r+c} \boldsymbol{x}_t'\boldsymbol{x}_t + L$, where

$$
\begin{aligned}
L = \sum_{t=r+1}^{r+c} \Bigg\{ & -\zeta\xi\psi_{d,t} - \left(\frac{\Gamma(\zeta + \xi^{-1})}{\Gamma(\zeta)} \frac{d_t}{\exp(\psi_{d,t})}\right)^{\xi} + \\
& + \tau \log\left(\frac{\tau}{\exp(\widehat{\psi}_{y,t}) + \tau}\right) + y_t \log\left(\frac{\exp(\widehat{\psi}_{y,t})}{\exp(\widehat{\psi}_{y,t}) + \tau}\right)\Bigg\} + \log f(\boldsymbol{\alpha}_{r+c})\mathrm{I}_{r+c<n}.
\end{aligned}
$$

Further define

$$
\begin{aligned}
\boldsymbol{\delta} &= (\boldsymbol{\delta}'_{r+1}, \ldots, \boldsymbol{\delta}'_{r+c})', \quad \boldsymbol{\delta}_t = \frac{\partial L}{\partial \boldsymbol{\alpha}_t} = \begin{pmatrix} \delta_{d,t} \\ \delta_{y,t} \end{pmatrix} + j(\boldsymbol{\alpha}_t), \\
\boldsymbol{Q} &= -\mathrm{E}\left(\frac{\partial^2 L}{\partial \underline{\boldsymbol{\alpha}} \partial \underline{\boldsymbol{\alpha}}'}\right) = \mathrm{diag}(\boldsymbol{A}_{r+1}, \boldsymbol{A}_{r+2}, \ldots, \boldsymbol{A}_{r+c}), \\
\boldsymbol{A}_t &= -\mathrm{E}\left(\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \boldsymbol{\alpha}'_t}\right) = \mathrm{diag}(A_{d,t}, A_{y,t}) + j'(\boldsymbol{\alpha}_t), \quad t = r+1, \ldots, r+c.
\end{aligned}
$$

The first derivatives are

$$
\delta_{d,t} = -\zeta\xi + \xi \left(\frac{\Gamma(\zeta + \xi^{-1})}{\Gamma(\zeta)} \frac{d_t}{\exp(\psi_{d,t})}\right)^{\xi}, \quad \delta_{y,t} = \frac{\tau(y_t - \exp(\widehat{\psi}_{y,t}))}{\tau + \exp(\widehat{\psi}_{y,t})},
$$

and

$$
j(\boldsymbol{\alpha}_t) = \boldsymbol{\Phi}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)\mathrm{I}_{t=r+c<n}.
$$

For the second derivatives, take expectations with respect to $(d_t, y_t)'$ multiplied by -1 and obtain

$$
A_{d,t} = \xi^2 \left(\frac{\Gamma(\zeta + \xi^{-1})}{\Gamma(\zeta)} \frac{d_t}{\exp(\psi_{d,t})}\right)^{\xi}, \quad A_{y,t} = \frac{\tau \exp(\widehat{\psi}_{y,t})(y_t + \tau)}{(\tau + \exp(\widehat{\psi}_{y,t}))^2},
$$

and

$$
j'(\boldsymbol{\alpha}_t) = \boldsymbol{\Phi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}\mathrm{I}_{t=r+c<n}.
$$

Applying a second order Taylor expansion to the log of the posterior density around mode $\hat{\underline{\boldsymbol{x}}}$ yields an approximating normal density as follows (excluding constant terms):

$$
\begin{aligned}
&\log \pi(\underline{\boldsymbol{x}}|\cdot) \\
&\approx \hat{L} + \left.\frac{\partial L}{\partial \underline{\boldsymbol{x}}'}\right|_{\underline{\boldsymbol{x}}=\hat{\underline{\boldsymbol{x}}}} (\underline{\boldsymbol{x}} - \hat{\underline{\boldsymbol{x}}}) + \frac{1}{2}(\underline{\boldsymbol{x}} - \hat{\underline{\boldsymbol{x}}})' \,\mathrm{E}\left.\left(\frac{\partial^2 L}{\partial \underline{\boldsymbol{x}} \partial \underline{\boldsymbol{x}}'}\right)\right|_{\underline{\boldsymbol{x}}=\hat{\underline{\boldsymbol{x}}}} (\underline{\boldsymbol{x}} - \hat{\underline{\boldsymbol{x}}}) - \frac{1}{2}\sum_{t=r+1}^{r+c} \boldsymbol{x}'_t \boldsymbol{x}_t \\
&= \hat{L} + \hat{\boldsymbol{\delta}}'(\underline{\boldsymbol{\alpha}} - \hat{\underline{\boldsymbol{\alpha}}}) - \frac{1}{2}(\underline{\boldsymbol{\alpha}} - \hat{\underline{\boldsymbol{\alpha}}})'\hat{\boldsymbol{Q}}(\underline{\boldsymbol{\alpha}} - \hat{\underline{\boldsymbol{\alpha}}}) - \frac{1}{2}\sum_{t=r+1}^{r+c} \boldsymbol{x}'_t \boldsymbol{x}_t \\
&\equiv \log q(\underline{\boldsymbol{x}}|\cdot),
\end{aligned}
$$

where $\hat{L}$, $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{Q}}$ are the values of L, $\boldsymbol{\delta}$ and $\boldsymbol{Q}$ at $\underline{\boldsymbol{\alpha}} = \hat{\underline{\boldsymbol{\alpha}}}$ (or at $\underline{\boldsymbol{x}} = \hat{\underline{\boldsymbol{x}}}$, equivalently). It can be shown that proposal density $q(\underline{\boldsymbol{x}}|\cdot)$ is the posterior of $\underline{\boldsymbol{x}}$ obtained from an ordinary linear state space model given by Eq. (23)-(24) below. Mode $\hat{\underline{\boldsymbol{x}}}$ is then obtained by repeating the following algorithm until it converges (usually 2 to 5 iterations).

Algorithm 1 (Disturbance smoother):

1. Initialize $\hat{\underline{\boldsymbol{x}}}$, and compute $\hat{\underline{\boldsymbol{\alpha}}}$ at $\underline{\boldsymbol{x}} = \hat{\underline{\boldsymbol{x}}}$ using state equation (20) recursively.
2. Evaluate $\hat{\boldsymbol{\delta}}_t$ and $\hat{\boldsymbol{A}}_t$ at $\underline{\boldsymbol{\alpha}} = \hat{\underline{\boldsymbol{\alpha}}}$, $t = r+1, \ldots, r+c$.
3. Define auxiliary variable $\hat{\boldsymbol{z}}_t = \hat{\boldsymbol{\alpha}}_t + \hat{\boldsymbol{A}}_t^{-1}\hat{\boldsymbol{\delta}}_t$, $t = r+1, \ldots, r+c$, and let $\hat{\boldsymbol{K}}_t$ be a Cholesky decomposition of $\hat{\boldsymbol{A}}_t$ such that $\hat{\boldsymbol{A}}_t = \hat{\boldsymbol{K}}_t\hat{\boldsymbol{K}}_t'$.
4. Consider the linear Gaussian state space model formulated by

$$
\begin{aligned}
\hat{\boldsymbol{z}}_t &= \boldsymbol{\alpha}_t + \mathbf{e}_t, && && (23)\\
\boldsymbol{\alpha}_t &= \boldsymbol{\Phi}\boldsymbol{\alpha}_{t-1} + \mathbf{u}_t, && t = r+1, \ldots, r+c, \quad t > 1, \\
\boldsymbol{\alpha}_1 &\backsim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{ini}}), && t = 1, \\
\begin{pmatrix} \mathbf{e}_t \\ \mathbf{u}_t \end{pmatrix} &\backsim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\hat{\boldsymbol{A}}_t^{-1}, \boldsymbol{\Sigma})). && && (24)
\end{aligned}
$$

   Apply the Kalman filter (e.g. Durbin and Koopman, 2008) and disturbance smoother (Koopman, 1993) to this state space model to obtain posterior mode $\hat{\underline{\boldsymbol{x}}}$ or $\hat{\underline{\boldsymbol{\alpha}}}$, equivalently.
5. Go to 2.

In the McMC sampling procedure, the current sample of $\underline{\boldsymbol{x}}$ may be taken as initial value of $\hat{\underline{\boldsymbol{x}}}$ 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, $\underline{\boldsymbol{x}}$ 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 $\underline{\boldsymbol{x}}_0$ denote the current value. Find mode $\hat{\underline{\boldsymbol{x}}}$ using Algorithm 1.
2. Proceed with Steps 2-4 in Algorithm 1 to obtain the approximated linear Gaussian state space model of Eq. (23)-(24).
3. Propose a candidate $\underline{\boldsymbol{x}}^*$ by sampling from $\tilde{q}(\underline{\boldsymbol{x}}^*) \propto \min\{\pi(\underline{\boldsymbol{x}}^*|\cdot), cq(\underline{\boldsymbol{x}}^*|\cdot)\}$ using the AR algorithm as follows:
   (a) Generate $\underline{\boldsymbol{x}}^*$ using the simulation smoother for the approximated state space model of Eq. (23)-(24).
   (b) Accept $\underline{\boldsymbol{x}}^*$ with probability

$$\frac{\min\{\pi(\underline{\boldsymbol{x}}^*|\cdot), cq(\underline{\boldsymbol{x}}^*|\cdot)\}}{cq(\underline{\boldsymbol{x}}^*|\cdot)},$$

   where c is a scaling constant. If it is rejected, go to (a).
4. Conduct the MH algorithm using candidate $\underline{\boldsymbol{x}}^*$, with acceptance probability

$$\min\left\{\frac{\pi(\underline{\boldsymbol{x}}^*|\cdot) \times \min\{\pi(\underline{\boldsymbol{x}}_0|\cdot), cq(\underline{\boldsymbol{x}}_0|\cdot)\}}{\pi(\underline{\boldsymbol{x}}_0|\cdot) \times \min\{\pi(\underline{\boldsymbol{x}}^*|\cdot), cq(\underline{\boldsymbol{x}}^*|\cdot)\}}, 1\right\}.$$

B Transition Probability Estimates

Tab. 9 and Tab. 10 report estimation results of the transition probability matrices $\boldsymbol{P}_d$, $\boldsymbol{P}_y$ for the SCDSVrs model.

Table 9: Results of the SCDSVrs Model - Transition Probability Matrix $\boldsymbol{P}_d$

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

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

Table 10: Results of the SCDSVrs Model - Transition Probability Matrix $\boldsymbol{P}_y$

	Mean	Stdev	95% c.b.	IF	CD
$p_{y,11}$	0.2942	0.0157	[0.2636, 0.3251]	82.1	0.58
$p_{y,12}$	0.0372	0.0181	[0.0052, 0.0742]	343.8	0.72
$p_{y,13}$	0.6680	0.0200	[0.6279, 0.7062]	147.7	0.48
$p_{y,14}$	0.000576	0.000334	[0.000076, 0.001364]	5.1	0.83
$p_{y,21}$	0.0443	0.0147	[0.0178, 0.0749]	250.5	0.45
$p_{y,22}$	0.7052	0.0199	[0.6656, 0.7428]	150.4	0.54
$p_{y,23}$	0.2502	0.0235	[0.2041, 0.2956]	245.6	0.34
$p_{y,24}$	0.000282	0.000253	[0.000009, 0.000941]	8.0	0.93
$p_{y,31}$	0.1558	0.0078	[0.1422, 0.1724]	172.7	0.15
$p_{y,32}$	0.0872	0.0108	[0.0666, 0.1086]	257.8	0.38
$p_{y,33}$	0.7564	0.0101	[0.7363, 0.7758]	147.0	0.90
$p_{y,34}$	0.000530	0.000146	[0.000281, 0.000849]	5.7	0.98
$p_{d,41}$	0.0688	0.0513	[0.0029, 0.1925]	6.0	0.99
$p_{d,42}$	0.1212	0.0774	[0.0090, 0.2987]	9.8	0.97
$p_{d,43}$	0.3372	0.0860	[0.1715, 0.5087]	5.8	0.94
$p_{d,44}$	0.4727	0.0697	[0.3370, 0.6092]	2.2	0.95

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

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)[11]. The latter authors give an overview of the field.

Let $\boldsymbol{X}_i,\, i = 0, 1, \ldots, I$, 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 is not possible in closed form. It requires the definition of a family of proposal distributions $\{q(\boldsymbol{x}, \cdot),\, \boldsymbol{x} \in \boldsymbol{X}\}$ that serves to generate possible transitions for the Markov chain, say from $\boldsymbol{X}$ to $\boldsymbol{Y}$, which are then accepted or rejected according to probability $\alpha(\boldsymbol{X}, \boldsymbol{Y}) = \min\{1, \pi(\boldsymbol{Y})q(\boldsymbol{Y}, \boldsymbol{X})/[\pi(\boldsymbol{X})q(\boldsymbol{X}, \boldsymbol{Y})]\}$.

The univariate normal random walk Metropolis algorithm (RWM) produces symmetric proposals of the form $y \backsim \mathcal{N}(x, \sigma^2)$, and the probability of move reduces to $\alpha(\boldsymbol{X}, \boldsymbol{Y}) = \min\{1, \pi(\boldsymbol{Y})/\pi(\boldsymbol{X})\}$. However, it is well known that its performance deteriorates sharply if σ^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 $I_{\textit{burn-in}}$ (thus ensuring ergodicity of the chain), wherein the optimal acceptance rate α_{opt} (0.44 for one, 0.35 for two dimensions; 0.234 asymptotically) is taken as efficiency criterion. Let $\boldsymbol{X}$ be of dimension $p \times 1$, $\boldsymbol{\Sigma}$ denote the proposal covariance, $\boldsymbol{R}$ the corresponding correlation matrix (used in a variant discussed below), $\boldsymbol{\sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$ a diagonal matrix with distinct proposal standard deviations, and c_1, c_2 scaling parameters. Further, given a $p \times 1$ vector $\boldsymbol{V}$, denote by $\boldsymbol{V}(m)$ the m^{th} component, $m = 1, \ldots, p$, and define by $\boldsymbol{e}_m$ a $p \times 1$ unit vector with the m^{th} component equal to one. The proposed algorithm follows:

1. Initialize, $i = 0$: $\boldsymbol{X}_i$, $\boldsymbol{\Sigma}_i = \boldsymbol{I}_p$, $\boldsymbol{R}_i$ accordingly, and $\boldsymbol{\sigma}_i$.
2. Iterate, $i = 1, \ldots, I$:
   (a) Sample $\boldsymbol{Y}_i \backsim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\sigma}_{i-1}\boldsymbol{\Sigma}_{i-1}\boldsymbol{\sigma}_{i-1})$, and set $\boldsymbol{X}_i = \boldsymbol{X}_{i-1} + \boldsymbol{Y}_i$ with probability $\alpha_i(\boldsymbol{X}_{i-1}, \boldsymbol{X}_{i-1} + \boldsymbol{Y}_i)$, otherwise $\boldsymbol{X}_i = \boldsymbol{X}_{i-1}$.
   (b) If $(i \leq I_{\textit{burn-in}}) \;\wedge\; (i \bmod k = 0)$ then
      i. $\bar{\alpha}_m = \frac{1}{k}\sum_{j=i-k+1}^{i} \alpha_{m,j}(\boldsymbol{X}_{j-1}, \boldsymbol{X}_{j-1} + \boldsymbol{Y}_j(m)\boldsymbol{e}_m), \quad m = 1, \ldots, p.$
      ii. $\log(\sigma^2_{m,i}) = \log(\sigma^2_{m,i-1}) + i^{-c_1}\left[\bar{\alpha}_m - \alpha_{opt}\right], \quad m = 1, \ldots, p.$
      iii. $\bar{\boldsymbol{X}}_i = \frac{1}{k}\sum_{j=i-k+1}^{i} \boldsymbol{X}_j$.
      iv. $\hat{\boldsymbol{\Sigma}}_i = \frac{1}{k-1}\sum_{j=i-k+1}^{i}(\boldsymbol{X}_j - \bar{\boldsymbol{X}}_i)(\boldsymbol{X}_j - \bar{\boldsymbol{X}}_i)'$.
      v. $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i-1} + i^{-c_2}\left[\hat{\boldsymbol{\Sigma}}_i - \boldsymbol{\Sigma}_{i-1}\right]$, calculate $\boldsymbol{R}_i$ from $\boldsymbol{\Sigma}_i$.

      else

      i. $\sigma^2_{m,i} = \sigma^2_{m,i-1}, \quad m = 1, \ldots, p.$
      ii. $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i-1}$, $\boldsymbol{R}_i = \boldsymbol{R}_{i-1}$.

[11] This appendix is taken as an independent unit with parameter notation to be separated from the rest of this work.

Initialization of $\sigma^2_{m,i}$ may be with the asymptotic optimal scaling rate of $2.38^2/m$ for RWM. Componentwise adaptive scaling (Algorithm 6) uses the information contained in the proposal $\boldsymbol{X}_{i-1} + \boldsymbol{Y}_i$ about scalings in the different dimensions. Directional acceptance probabilities $\alpha_{m,i}$ are calculated through "virtual" componentwise updates with increments $\boldsymbol{Y}_i(m)\boldsymbol{e}_m$. For the specific problem at hand, using correlation $\boldsymbol{R}$ instead of covariance $\boldsymbol{\Sigma}$ 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 $\alpha_i(\boldsymbol{X}_{i-1}, \boldsymbol{X}_{i-1} + \boldsymbol{Y}_i)$ instead to adapt a single scaling variable $\log(\sigma^2_i)$ that scales $\boldsymbol{\Sigma}$, steps 2.(a) and 2.(b)i, 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. 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 σ during the burn-in period. The batch size has been set to $k = 100$ throughout.

D 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,

$$IF = 1 + 2\frac{b_w}{b_w - 1}\sum_{i=1}^{b_w} K_{\mathrm{QS}}\left(\frac{i}{b_w}\right)\hat{\rho}_i, \qquad (25)$$

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

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

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

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

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 the 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 θ 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 σ_A and σ_B denote the standard errors of $\bar{\theta}_A$ and $\bar{\theta}_B$, respectively. The convergence diagnostic is then defined as

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

under the assumption that sequence $\theta^{(i)}$ is stationary. For window lengths n_A and n_B Geweke (1992) proposes $0.1N$ and $0.5N$, 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. (25) at hand, variances σ_A^2, σ_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.

References

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance 51*(1), 169–204.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Andrieu, C. and J. Thoms (2008). A tutorial on adaptive McMC. *Statistics and Computing 18*(4), 343–373.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2004). A comparison of financial duration models via density forecasts. *International Journal of Forecasting 20*(4), 589–609.

Bauwens, L. and D. Veredas (2004). The stochastic conditional duration model: A latent variable model for the analysis of financial durations. *Journal of Econometrics 119*(2), 381–412.

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

Bollerslev, T. (2008). Glossary to ARCH (GARCH). CREATES Research Papers 2008-49, School of Economics and Management, University of Aarhus.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton, USA: Princeton University Press.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics 75*(1), 79–97.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *Journal of the American Statistical Association 49*(4), 327–335.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica 41*(1), 135–155.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*(2), 339–350.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*(4), 863–883.

Douc, R., O. Cappé, and E. Moulines (2005). Comparison of resampling schemes for particle filtering. In *4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 64–69.

Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. New York, USA: Springer.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika 84*(3), 669–684.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(1), 3–56.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–615.

Durbin, J. and S. J. Koopman (2008). *Time Series Analysis by State Space Methods*. Oxford, UK: Oxford University Press.

Easley, D. and M. O'Hara (1992). Time and the process of security price adjustment. *The Journal of Finance 47*(2), 577–605.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*(4), 987–1007.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica 68*(1), 1–22.

Engle, R. F. and A. Lunde (2003). Trades and quotes: A bivariate point process. *Journal of Financial Econometrics 1*(2), 159–188.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica 66*(5), 1127–1162.

Engle, R. F. and J. R. Russell (2010). Analysis of high frequency and transaction data. In Y. Aït-Sahalia and L. Hansen (Eds.), *Handbook of Financial Econometrics*, Volume 1, pp. 383–424. Amsterdam, NL: North-Holland.

Feng, D., G. J. Jiang, and P. X.-K. Song (2004). Stochastic conditional duration models with "leverage effect" for financial transaction data. *Journal of Financial Econometrics 2*(3), 390–421.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*(453), 194–209.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York, USA: Springer.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics*, Volume 4, pp. 169–188. New York, USA: Oxford University Press.

Ghysels, E., C. Gouriéroux, and J. Jasiak (2004). Stochastic volatility duration models. *Journal of Econometrics 119*(2), 413–433.

Giot, P. (2001). Time transformations, intraday data and volatility models. *Journal of Computational Finance 4*(2), 31–62.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71–100.

Grammig, J. and K.-O. Maurer (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal 3*(1), 16–38.

Grammig, J. and M. Wellner (2002). Modeling the interdependence of volatility and inter-transaction duration processes. *Journal of Econometrics 106*(2), 369–400.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Hasbrouck, J. (2007). *Empirical Market Microstructure*. Oxford, UK: Oxford University Press.

Hautsch, N. (2002). Modelling intraday trading activity using Box-Cox ACD models. CoFE Discussion paper 02-05, Center of Finance and Econometrics, University of Konstanz.

Hautsch, N. (2012). *Econometrics of Financial High-Frequency Data*. Berlin Heidelberg, Germany: Springer-Verlag.

Hautsch, N., P. Malec, and M. Schienle (2013). Capturing the zero: A new class of zero-augmented distributions and multiplicative error processes. *Journal of Financial Econometrics 12*(1), 89–121.

Hol, J. D., T. B. Schön, and F. Gustafsson (2006). On resampling algorithms for particle filters. In *Nonlinear Statistical Signal Processing Workshop*, Cambridge, UK. IEEE.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance 42*(2), 281–300.

Johnson, H. and D. Shanno (1987). Option pricing when the variance is changing. *The Journal of Financial and Quantitative Analysis 22*(2), 143–151.

Kim, C.-J. and C. R. Nelson (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *The Review of Economics and Statistics 80*(2), 188–201.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching*. Cambridge, USA: MIT Press.

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

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika 80*(1), 117–126.

Liesenfeld, R., I. Nolte, and W. Pohlmeier (2006). Modelling financial transaction price movements: A dynamic integer count data model. *Empirical Economics 30*, 795–825.

Lunde, A. (1999). A Generalized Gamma autoregressive conditional duration model. Working Paper, Department of Economics, Politics and Public Administration, Aalborg Univeristy, Denmark.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets 3*(3), 205–258.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets 8*(4), 377–399.

Nolte, I. (2008). Modeling a multivariate transaction process. *Journal of Financial Econometrics 6*(1), 143–170.

O'Hara, M. (1995). *Market Microstructure Theory*. Oxford, UK: Blackwell.

Pemstein, D., K. M. Quinn, and A. D. Martin (2011). The Scythe statistical library: An open source C++ library for statistical computation. *Journal of Statistical Software 42*(12), 1–26.

Pitt, M. K. and N. Shephard (1999a). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association 94*(446), 590–599.

Rachev, S. T., J. S. J. Hsu, B. S. Bagasheva, and F. J. Fabozzi (2008). *Bayesian Methods in Finance*. New York, USA: John Wiley & Sons.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods*. New York, USA: Springer.

Roberts, G. O. and J. S. Rosenthal (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science 16*(4), 351–367.

Rosenberg, B. (1972). The behavior of random variables with nonstationary variance and the distribution of security prices. Research Program in Finance Working Papers 11, University of California at Berkeley.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23*(3), 470–472.

Russell, J. R. and R. F. Engle (2005). A discrete-state continuous-time model of financial transactions prices and times. *Journal of Business and Economic Statistics 23*(2), 166–180.

Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics 1*(1), 2–25.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford, UK: Oxford University Press.

Shephard, N. and T. G. Andersen (2008). Stochastic volatility: Origins and overview. Discussion paper, 2008-W04, Department of Economics, Nuffield College, University of Oxford.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*(3), 653–667.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 583–639.

Stephens, M. A. (1970). Use of the Kolmogorov-Smirnov, Cramér-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society. Series B (Methodological) 32*(1), 115–122.

Strickland, C. M., C. S. Forbes, and G. M. Martin (2006). Bayesian analysis of the stochastic conditional duration model. *Computational Statistics & Data Analysis 50*(9), 2247–2267.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices, 1961-1979. In O. E. Anderson (Ed.), *Time series analysis: Theory and practice*, pp. 203–226. Amsterdam, NL: North-Holland.

Taylor, S. J. (1986). *Modelling Financial Time Series*. New York, USA: John Wiley & Sons.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics 22*(4), 1701–1728.

Watanabe, T. and Y. Omori (2004). A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997). *Biometrika 91*(1), 246–248.

Whiteley, N. and A. M. Johansen (2011). *Bayesian Time Series Models*. Cambridge, UK: Cambridge University Press.

Zhang, M. Y., J. R. Russell, and R. S. Tsay (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics 104*(1), 179–207.