Appendix: A Statistical Recurrent Stochastic Volatility Model for Stock Markets

January 10, 2022

A The SR-SV model: graphical representation

Figure A.1 illustrates graphically the RNN model in (7). We follow Goodfellow et al. (2016) and use a black square to indicate the delay of a single time step in the circuit diagram (*left*). The circuit diagram can be interpreted as an unfolded computational graph (*right*), where each node is associated with a particular time step.

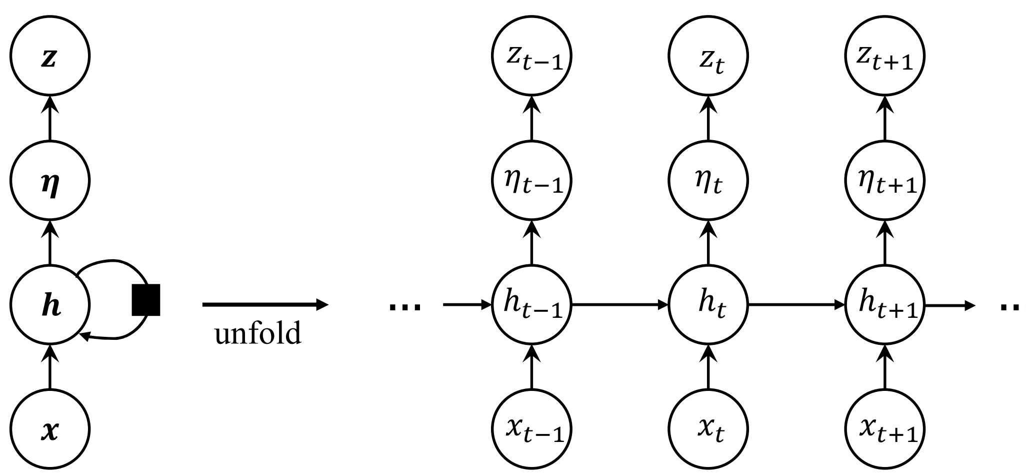

Figure A.1: Graphical representation of the RNN model in (7). The latent state, or hidden unit, h_t are connected through time.

Graphically, the function to compute h_t in (7) can be represented as a Simple Recurrent Neural (SRN) unit, as shown in Figure A.2 (*left*), and hence we refer to it as $h_t = \text{SRN}(x_t, h_{t-1})$, which takes the data x_t at time t and the previous state h_{t-1} as the inputs.

Using the SRN structure, the unfolded graph of the RNN model of Elman (1990), which is usually referred to as the Simple RNN model, can be reinterpreted as the unfolded graph in Figure A.2 (*right*).

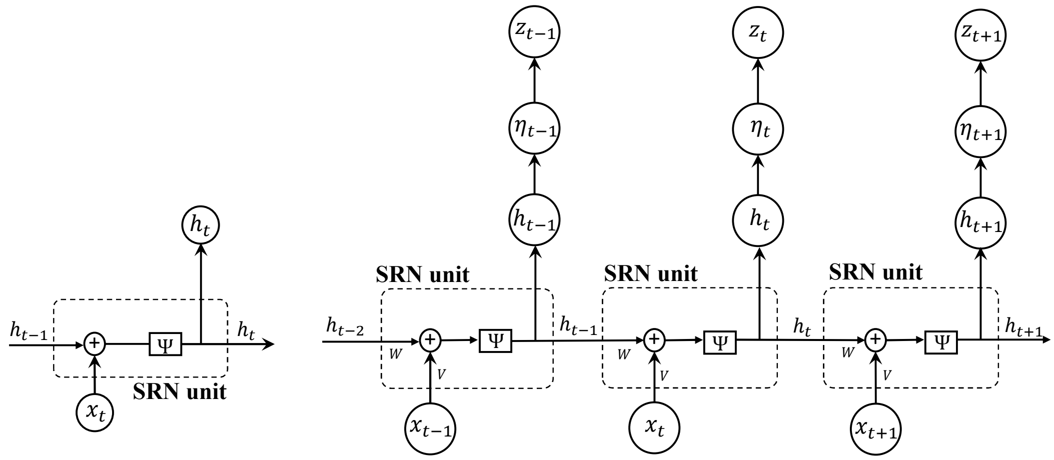

Figure A.2: The structures of the SRN unit (*left*) and the graphical representation of the Simple RNN model (*right*), which uses the SRN unit to compute the latent state, or hidden unit, h_t.

By using SRU units to model the hidden unit h_t, the SR-SV model in Section 2.3 can be presented graphically as

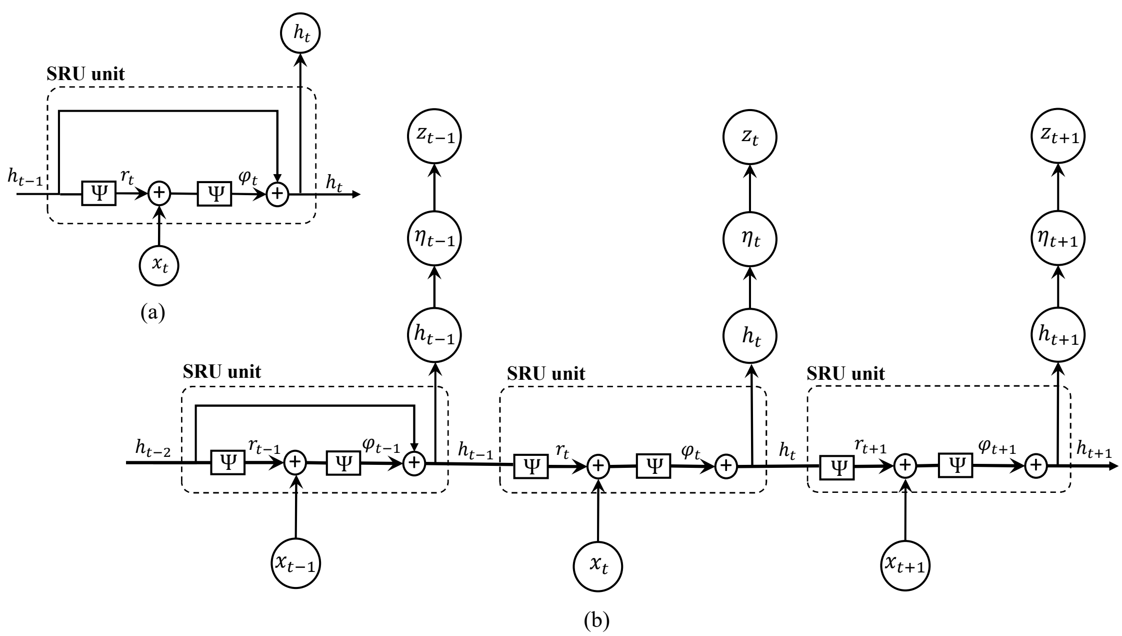

Figure A.3: The structure of the SRU unit (a) and the graphical representation of the SRU model (b), which uses the SRU unit to compute the latent state, or hidden unit, h_t.

B Particle filter and implementation details of the DT-SMC sampler

Algorithm B.1 outlines our DT-SMC method for Bayesian inference. We use a random walk proposal for $q(\theta'|\theta)$. We follow Gunawan et al. (2018) and choose the tempering sequence γ_k adaptively to ensure a sufficient level of particle efficiency by selecting the next value of γ_k such that ESS stays above a threshold.

Algorithm B.2 describes the particle filter for the SR-SV model where $\mathbf{Z}_t = (Z_t^1, ..., Z_t^N)$ denotes the vector of particles at time t. The set of standard normal random numbers U includes two sources of randomness: the set of random numbers $\{U^P_{t,n}, t = 1, ..., T; n = 1, ..., N\}$ used to propose new particles in each time step, and the set of random numbers $\{U^R_{t,n}, t = 1, ..., T-1; n = 1, ..., N\}$ used in the resampling step. For the resampling step, we use multinomial resampling, with sorting, to obtain the vector ancestor indexes $\{A^n_{t-1}, n = 1, ..., N\}$ used to propose particles at time t. The sorting step helps eliminate the discontinuity issues of the selected particles in the ordinary multinomial resampling scheme (Gerber and Chopin, 2014). This sorted resampling scheme allows the selected particles to still be close after being resampled and hence helps to reduce the variability of the likelihood ratio estimator $\widehat{p}(y_{1:T}|\theta', u')/\widehat{p}(y_{1:T}|\theta, u)$ shown in the Algorithm B.1 (Deligiannidis et al., 2018).

Variable	Description	Value
K	Number of annealing levels	10000
M	Number of SMC particles	10000
N	Number of particles in the particle filter	200
ρ	Correlation factor in the CPM algorithm	0.999
c	Constant of the ESS threshold	0.800
N_{CPM}	Number of CPM moves	20

Table B.1: Implementation settings of the DT-SMC sampler.

The multinomial resampling scheme in Steps 2a and 2b generates the ancestor indices $A^n_{t-1}, n = 1, ..., N$, from the multinomial distribution denoted as $\mathcal{F}(\cdot|\mathbf{p}, \mathbf{u})$, with $\mathbf{p}$ the vector of parameters of the multinomial distribution and $\mathbf{u}$ the uniform random numbers used within a multinomial random number generator. We use the standard normal cumulative distribution function $\Phi(\cdot)$ in the resampling step to transform the normal random numbers $U^R_{t-1,n}$ to the uniform random numbers, denoted as $\overline{U}^R_{t-1,n}$.

Algorithm B.1 Density Tempered Sequential Monte Carlo for the SR-SV model

1. Sample $\theta_0^j \sim p(\theta)$, $u_0^j \sim p(u)$ and set $W_0^j = 1/M$ for $j = 1...M$
2. **For** $k = 1, ..., K$,

Step 1: **Reweighting**: Compute the unnormalized weights

$$w_k^j = W_{k-1}^j \frac{\widehat{p}(y_{1:T}|\theta_{k-1}^j, u_{k-1}^j)^{\gamma_k} p(\theta_{k-1}^j)}{\widehat{p}(y_{1:T}|\theta_{k-1}^j, u_{k-1}^j)^{\gamma_{k-1}} p(\theta_{k-1}^j)} = W_{k-1}^j \widehat{p}(y_{1:T}|\theta_{k-1}^j, u_{k-1}^j)^{\gamma_k - \gamma_{k-1}}, \quad j = 1, ..., M \tag{B.1}$$

and set the new normalized weights

$$W_k^j = \frac{w_k^j}{\sum_{s=1}^M w_k^s}, \quad j = 1, ..., M. \tag{B.2}$$

Step 2: Compute the effective sample size (ESS):

$$\text{ESS} = \frac{1}{\sum_{j=1}^M \left(W_k^j\right)^2}. \tag{B.3}$$

If ESS $< cM$ for some $0 < c < 1$, **then**

(i) **Resampling**: Resampling from $\{\theta_{k-1}^j, u_{k-1}^j\}_{j=1}^M$ using the weights $\{W_k^j\}_{j=1}^M$, and then set $W_k^j = 1/M$ for $j = 1...M$, to obtain the new equally-weighted particles $\{\theta_k^j, u_k^j, W_k^j\}_{j=1}^M$.

(ii) **Markov move**: For each $j = 1,...,M$, move the samples θ_k^j, u_k^j according to N_{CPM} steps of the CPM:

(a) Sample $\theta_k^{j\prime}$ from the proposal density $q(\theta_k^{j\prime}|\theta_k^j)$.

(b) Sample $\epsilon^j \sim \mathcal{N}(0_D, I_D)$ and set $u_k^{j\prime} = \rho u_k^j + \sqrt{1-\rho^2}\epsilon^j$ with $\rho \in (-1,1)$ a correlation factor.

(c) Compute the estimated likelihood $\widehat{p}(y_{1:T}|\theta_k^{j\prime}, u_k^{j\prime})$ using a particle filter (see Algorithm B.2 in Appendix B)

(d) Set $\theta_k^j = \theta_k^{j\prime}$ and $u_k^j = u_k^{j\prime}$ with the probability

$$\min\left(1, \frac{\widehat{p}(y_{1:T}|\theta_k^{j\prime}, u_k^{j\prime})^{\gamma_k} p(\theta_k^{j\prime})}{\widehat{p}(y_{1:T}|\theta_k^j, u_k^j)^{\gamma_k} p(\theta_k^j)} \frac{q(\theta_k^j|\theta_k^{j\prime})}{q(\theta_k^{j\prime}|\theta_k^j)}\right), \tag{B.4}$$

otherwise keep θ_k^j, u_k^j unchanged.

end

3. The log of marginal likelihood estimate is

$$\log\widehat{p}(y_{1:T}) = \sum_{t=1}^K \log\left(\sum_{j=1}^M w_k^j\right). \tag{B.5}$$

Algorithm B.2 Particle filter for the SR-SV model

Input: $T, N, y_{1:T}, \theta, U = (U^P_{1,1}, ..., U^P_{T,N}, U^R_{1,1}, ..., U^R_{T-1,N})$

1. At time $t = 1$,

   (a) for $n = 1, ...N$, initialize the particles (H^n_1, η^n_1, Z^n_1), e.g., $H^n_1 = 0$, as the SRU unit initially has no memory, and

   $$\begin{aligned} \eta^n_1 &= \beta_0 + \sigma U^P_{1,n} \\ Z^n_1 &= \eta^n_1 \end{aligned}$$

   (b) compute and normalize the weights

   $$\begin{aligned} w_1(Z^n_1) &= \frac{\mu_\theta(Z^n_1) g_\theta(y_1|Z^n_1)}{q_\theta(Z^n_1|y_1)} = g_\theta(y_1|Z^n_1) \\ W^n_1 &= \frac{w_1(Z^n_1)}{\sum_{m=1}^N w_1(Z^m_1)} \end{aligned}$$

   (c) compute the estimated likelihood $\widehat{p}(y_1|\theta)$ as

   $$\widehat{p}(y_1|\theta, U) = \frac{1}{N} \sum_{k=1}^N w_1(Z^n_1).$$

2. At times $t = 2, ..., T$,

   (a) sort the particle vector $\mathbf{Z}_{t-1}$ in ascending order to obtain the vector of sorted particles $\overline{\mathbf{Z}}_{t-1} = (\overline{Z}^1_{t-1}, ..., \overline{Z}^N_{t-1})$. The sorted index vector associated with $\overline{\mathbf{Z}}_{t-1}$ is denoted as $\mathbf{I}_{t-1} = (I^1_{t-1}, ..., I^N_{t-1})$. In this setting, we have the relation $\overline{Z}^n_{t-1} = Z^{I^n_{t-1}}_{t-1}$ with $n = 1, ..., N$. Use the sorted index vector $\mathbf{I}_{t-1}$ to define the vector of sorted weights $(\overline{W}^1_{t-1}, ..., \overline{W}^N_{t-1})$ such that

   $$\overline{W}^n_{t-1} = W^{I^n_{t-1}}_{t-1}$$

   (b) sample $A^n_{t-1} \sim \mathcal{F}(\cdot|\overline{W}^n_{t-1}, \overline{U}^R_{t-1,n})$ where $\overline{U}^R_{t-1,n} = \Phi(U^R_{t-1,n})$ for $n=1,...,N$.

(c) for $n=1,...N$, generate particles Z_t^n by

$$
\begin{aligned}
x_{t-1} &= [\eta_{t-1}^{A_{t-1}^n}, Z_{t-1}^{A_{t-1}^n}] \\
H_t^n &= \mathrm{SRU}(x_{t-1}, H_{t-1}^{A_{t-1}^n}) \\
\eta_t^n &= \beta_0 + \beta_1 H_t^n + \sigma U_{t,n}^P \\
Z_t^n &= \eta_t^n + \phi Z_{t-1}^{A_{t-1}^n}
\end{aligned}
$$

and set $Z_{1:t}^n = (Z_{1:t-1}^{A_{t-1}^n}, Z_t^n)$.

(d) compute and normalize the weights

$$
\begin{aligned}
w_t(Z_{1:t}^n) &= \frac{f_\theta(Z_t^n|Z_{t-1}^{A_{t-1}^n})g_\theta(y_t|Z_t^n)}{q_\theta(Z_t^n|y_t, Z_{t-1}^{A_{t-1}^n})} = g_\theta(y_1|Z_1^n) \\
W_t^n &= \frac{w_t(Z_{1:t}^n)}{\sum_{m=1}^N w_t(Z_{1:t}^m)}
\end{aligned}
$$

(e) compute the estimated likelihood $\widehat{p}(y_t|y_{1:t-1},\theta)$ as

$$
\widehat{p}(y_t|y_{1:t-1},\theta,U) = \frac{1}{N}\sum_{n=1}^N w_t(Z_{1:t}^n).
$$

Output: Estimate of the likelihood

$$
\widehat{p}(y_{1:T}|\theta,U) = \widehat{p}(y_1|\theta,U)\prod_{t=2}^T \widehat{p}(y_t|y_{1:t-1},\theta,U).
$$

C Bayesian inference and forecasting for the LMSV model

Denote by $x_{1:T} = \{x_t = \log y_t^2,\ t=1,...,T\}$ the series of log squared returns. The LMSV model in (5)-(6) can be transformed to a stationary model with respect to x_t as

$$
\begin{aligned}
(1-B)^d\Phi(B)z_t &= \Theta(B)\eta_t,\ \ \eta_t \sim \mathcal{N}(0,\sigma_\eta^2),\ \ t=2,...,T, && \text{(C.1)} \\
x_t &= c + z_t + \xi_t,\ \ \xi_t \sim (0,\sigma_\xi^2)\ \ t=1,2,...,T, && \text{(C.2)}
\end{aligned}
$$

where $\xi_t = \log\epsilon_t^2 - \mathrm{E}[\log\epsilon_t^2]$ is i.i.d with mean zero and variance σ_ξ^2, $c=\log(\kappa^2)+\mathrm{E}[\log\epsilon_t^2]$. The process x_t is the sum of the long-memory ARFIMA(p,d,q) process z_t and a non-Gaussian noise, with $\mathrm{E}[x_t]=c$ and h order autocovariance

$$
\gamma_x(h) = \mathrm{Cov}(x_t, x_{t+h}) = \gamma(h), h \neq 0, \quad \text{(C.3)}
$$

with $\gamma_x(0)=\gamma(0)+\sigma_\xi^2$.

Breidt et al. (1998) estimate the LMSV model by maximizing the Whittle log-likelihood (Whittle, 1953), defined as

$$\ell_W(\beta_x) := 2\pi T^{-1} \sum_{k=1}^{[T/2]} \left\{ \log f_{\beta_x}(\omega_k) + \frac{J(\omega_k)}{f_{\beta_x}(\omega_k)} \right\}, \tag{C.4}$$

where $[\cdot]$ denotes the integer part, $\beta_x = (d,\phi_1,...,\phi_p,\theta_1,...,\theta_q,\sigma_\eta^2,\sigma_\xi^2,c)$ is the vector of model parameters, $\omega_k{=}2\pi kT^{-1}$ is the kth Fourier frequency, $J(\omega_k)$ is the kth normalized periodogram ordinate

$$J(\omega_k) = \frac{1}{2\pi T}\left(\sum_{t=1}^{T} x_t \cos\omega_k t\right)^2 + \frac{1}{2\pi T}\left(\sum_{t=1}^{T} x_t \sin\omega_k t\right)^2, \tag{C.5}$$

and $f_{\beta_x}(\omega_k)$ is the spectral density of the LMSV model in (C.1) and (C.2)

$$f_{\beta_x}(\omega_k) = \frac{\sigma_\eta^2 \left|\Theta(e^{-i\omega_k})\right|^2}{2\pi \left|1-e^{-i\omega_k}\right|^{2d} \left|\Phi(e^{-i\omega_k})\right|^2} + \frac{\sigma_\xi^2}{2\pi}. \tag{C.6}$$

The Whittle likelihood is an approximation of the time-domain likelihood and is exact if the data are i.i.d. Gaussian.

Let $\pi_W(\beta_x) \propto L_W(\beta_x)\, p(\beta_x)$ be the posterior density based on the Whittle likelihood $L_W(\beta_x){=}\exp(\ell_W(\beta_x))$, given the log of squared return series $x{=}\{x_t,\ t{=}1,...,T\}$. To sample from $\pi_W(\beta_x)$, we use an adaptive random walk MCMC method summarized in Algorithm C.1, with the covariance matrix in the random walk proposal adaptively scaled to target an overall acceptance probability of 25% Garthwaite et al. (2010). We note that the vector of parameters β_x in Algorithm C.1 does not include the constant c in (C.2), which is estimated by the sample mean, i.e., $c{=}\frac{1}{T}\sum_{t=1}^{T} x_t$ (Harvey, 2007).

Algorithm C.1 Markov Chain Monte Carlo with a random walk proposal

---

Sample $\beta_x \sim p(\beta_x)$

For each MCMC iteration:

1. Sample β_x' from the proposal density $q(\beta_x'|\beta_x)$.
2. Compute the Whittle likelihood $L_W(\beta_x'){=}\exp(\mathcal{L}_w(\beta_x'))$ with $\mathcal{L}_w(\cdot)$ defined in (C.4).
3. Accept the proposal β_x' with the probability

$$\min\left\{1, \frac{L_W(\beta_x')}{L_W(\beta_x)} \frac{p(\beta_x')}{p(\beta_x)} \frac{q(\beta_x|\beta_x')}{q(\beta_x'|\beta_x)}\right\}.$$

Given the sample of model parameters generated from the posterior density $\pi_W(\beta_x)$, Harvey (2007) suggests the following convenient way to obtain the estimated values of conditional variance σ_t^2 for the LMSV model. Suppose that z_t is a stationary process and denote by Σ_z and Σ_ξ the covariance matrices of z_t and ξ_t, respectively, then the covariance matrix Σ of the log squared returns series x is $\Sigma{=}\Sigma_z{+}\Sigma_\xi$. The minimum mean square linear estimator of the log volatility $\tilde{z}{=}\{\tilde{z}_t,\ t{=}1,...,T\}$ is calculated as

$$\tilde{z} = (\mathbf{I}_T - \sigma_\xi^2\Sigma^{-1})x' + \sigma_\xi^2\Sigma^{-1}\boldsymbol{\iota}, \tag{C.7}$$

where $\mathbf{I}_T$ is the identity matrix of size T, $\boldsymbol{\iota}$ is a $T\times 1$ column vector of ones. As ξ_t are i.i.d and serially uncorrelated, the covariance matrix Σ_ξ is $\Sigma_\xi = \sigma_\xi^2 \mathbf{I}_T$. We note that for a general ARFIMA(p,d,q) process, there is no closed form for $\gamma(h)$, and hence covariance matrix Σ_z, so approximations of Σ_z are needed, e.g. see Sowell (1992); Doornik and Ooms (2003). However, Hosking (1981) suggests exact ACVF for some simple cases of (p,d,q) such as ARFIMA(0,d,0), ARFIMA(1,d,0) and ARFIMA(0,d,1).

Given the estimates of z_t in (C.7), the conditional variance σ_t^2 is computed as

$$\widetilde{\sigma_t}^2 = \tilde{\kappa}^2 \exp(\widetilde{z_t}),$$

where the scale factor $\tilde{\kappa}^2$ is estimated as

$$\tilde{\kappa}^2 = \frac{1}{T}\sum_{t=1}^{T} \widetilde{y_t}, \tag{C.8}$$

with $\widetilde{y_t} = y_t \exp(-\widetilde{z_t}/2)$ the heteroscredasticity corrected observations.

Denote the $1\times T$ covariance between x_{T+1} and x as $R := [\gamma_x(T),...,\gamma_x(1)]$, the one-step-ahead forecast value of the log squared return is calculated as (Harvey, 2007)

$$\widehat{x}_{T+1} = c + R\Sigma^{-1}(x - c\mathbf{1}_T)$$

The one-step-ahead forecast of the conditional variance is $\widehat{\sigma}_{T+1} = \tilde{\kappa}^2 \exp(\widehat{x}_{T+1} - c)$.

Table C.1 shows the estimation results of the LMSV model using a ARFIMA(1,d,0) process to model the log volatility z_t, parameter vector $\beta_x = [d, \phi_1, \sigma_\eta^2, \sigma_\xi^2]$. We also report the estimates of the scale factor κ in (6) and the constant c in (C.2). The priors for the parameters ϕ_1 and σ_η^2 are similar to those in Table 2 for the SV model. The prior for σ_ξ^2 is the same inverse-Gamma as for σ_η^2. For the fractional integration parameter d, we set the prior $2d \sim$ Beta(20,5). We run $N_{\text{MCMC}} = 100000$ MCMC iterations of Algorithm C.1 and discard the first 10,000 iterations as burnin.

	d	ϕ	σ_η^2	σ_ξ^2	κ	c
DAX	0.442	0.708	0.054	4.933	1.974	−1.616
	(0.026)	(0.082)	(0.024)	(0.170)		
HSI	0.431	0.747	0.052	5.469	2.047	−1.526
	(0.030)	(0.069)	(0.021)	(0.183)		
FCHI	0.434	0.716	0.062	5.241	1.984	−1.562
	(0.029)	(0.083)	(0.029)	(0.182)		
SPX	0.428	0.809	0.040	5.488	2.017	−1.650
	(0.035)	(0.060)	(0.016)	(0.177)		
TSX	0.445	0.714	0.047	4.964	1.918	−1.493
	(0.026)	(0.084)	(0.021)	(0.160)		

Table C.1: Applications: Posterior means of the parameters of the LMSV model with the posterior standard deviations in brackets. We also report the estimation of the scale factor κ and the constant c.

D Additional results

This section provides the additional tables and figures discussed, but not presented, in Section 4.2.

	In-sample Period	Out-of-sample Period	T_{in}	T_{out}
DAX	23 Apr 2004 – 21 Feb 2012	22 Feb 2012 – 05 Feb 2016	2000	1000
HSI	27 Oct 2003 – 28 Nov 2011	29 Nov 2011 – 21 Dec 2015	2000	1000
FCHI	09 Jun 2004 – 22 Mar 2012	23 Mar 2012 – 23 Feb 2016	2000	1000
SPX	27 Feb 2004 – 06 Feb 2012	07 Feb 2012 – 28 Jan 2016	2000	1000
TSX	03 Feb 2004 – 01 Feb 2012	02 Feb 2012 – 27 Jan 2016	2000	1000

Table D.1: Descriptions of the five index datasets. We use the first $T_{\text{in}} = 2000$ returns for in-sample analysis and the rest $T_{\text{out}} = 1000$ for out-of-sample analysis.

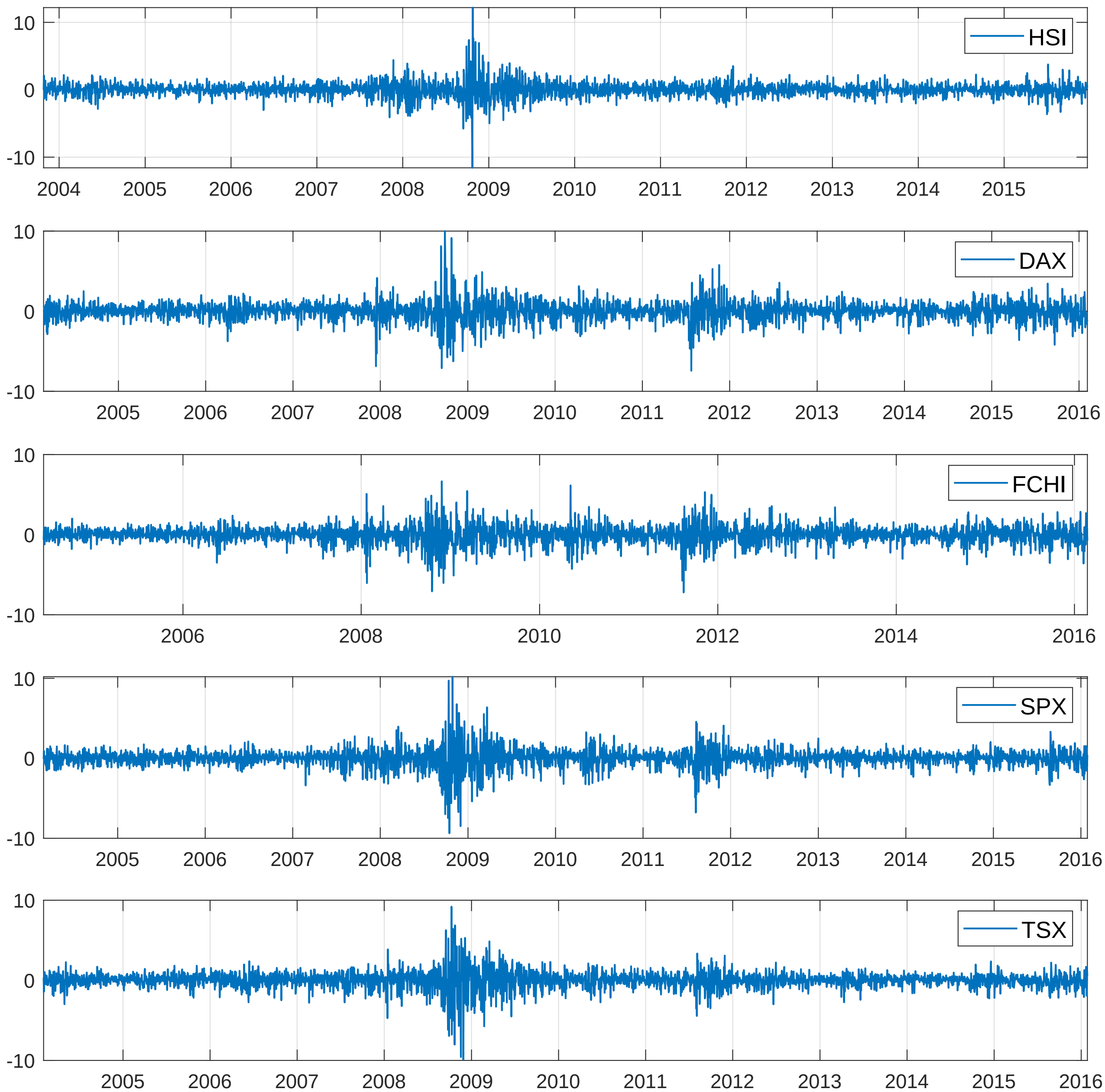

Figure D.1: Applications: Time series plots for the HSI, DAX, FCHI, SPX and TSX return datasets.

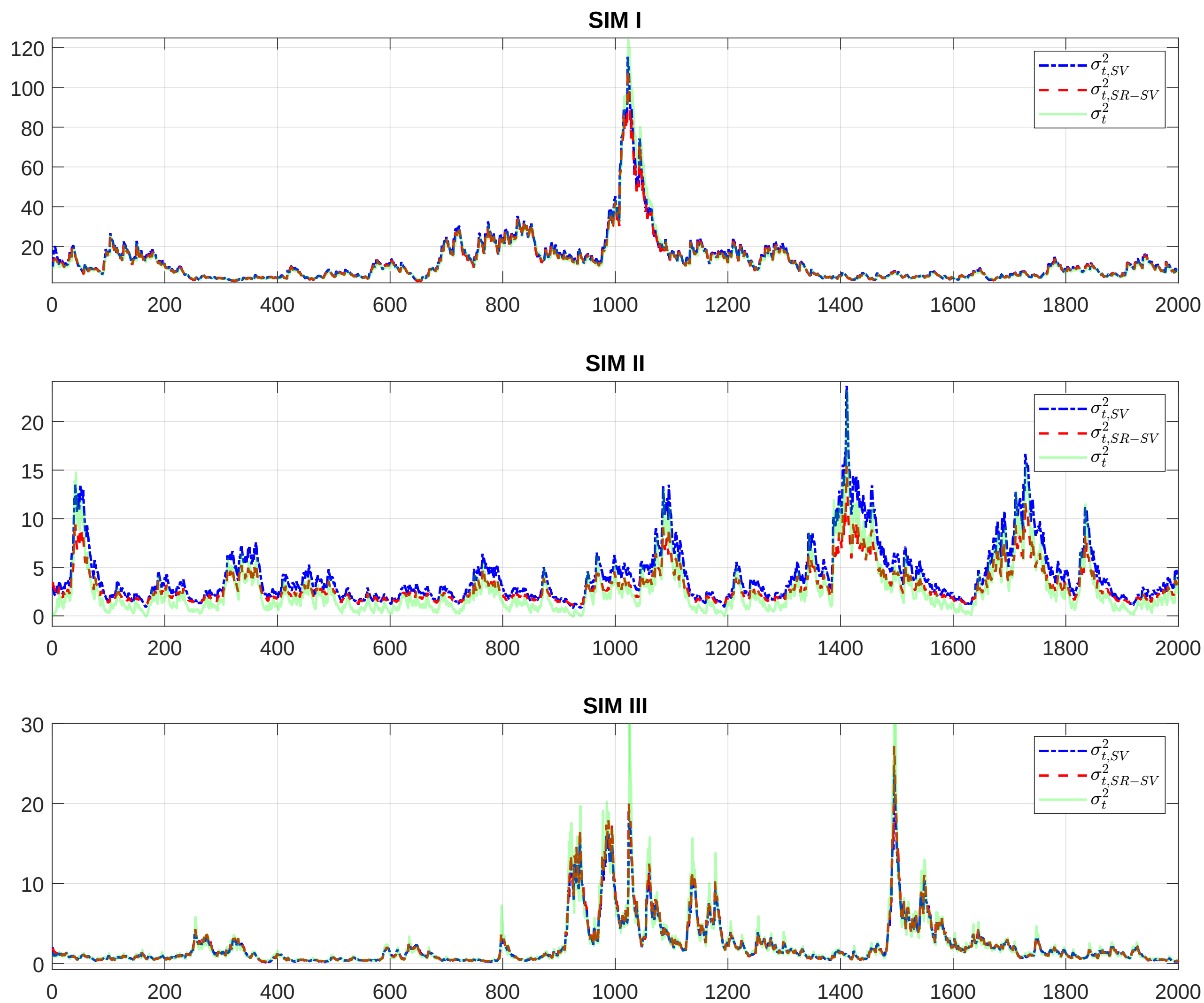

Figure D.2: Simulation: Filtered volatility of the SV and SR-SV models, together with the true volatility, on three simulation datasets. (This is better viewed in colour).

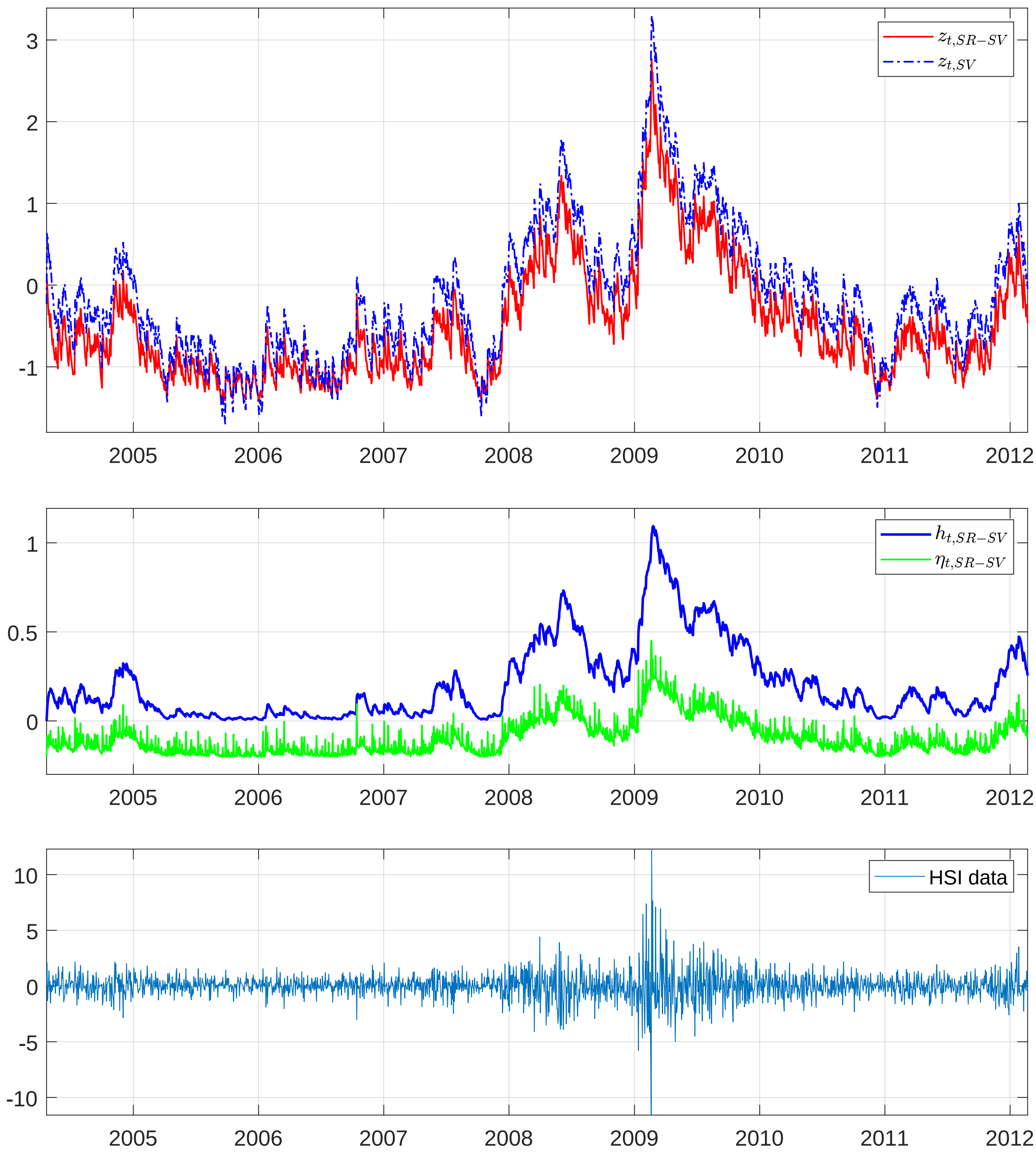

Figure D.3: HSI: (*Top*) The filtered log conditional variance of the SR-SV and SV models. (*Middle*) The filtered values of η_t and h_t of the SR-SV model. (*Bottom*) The in-sample data. (This is better viewed in colour).

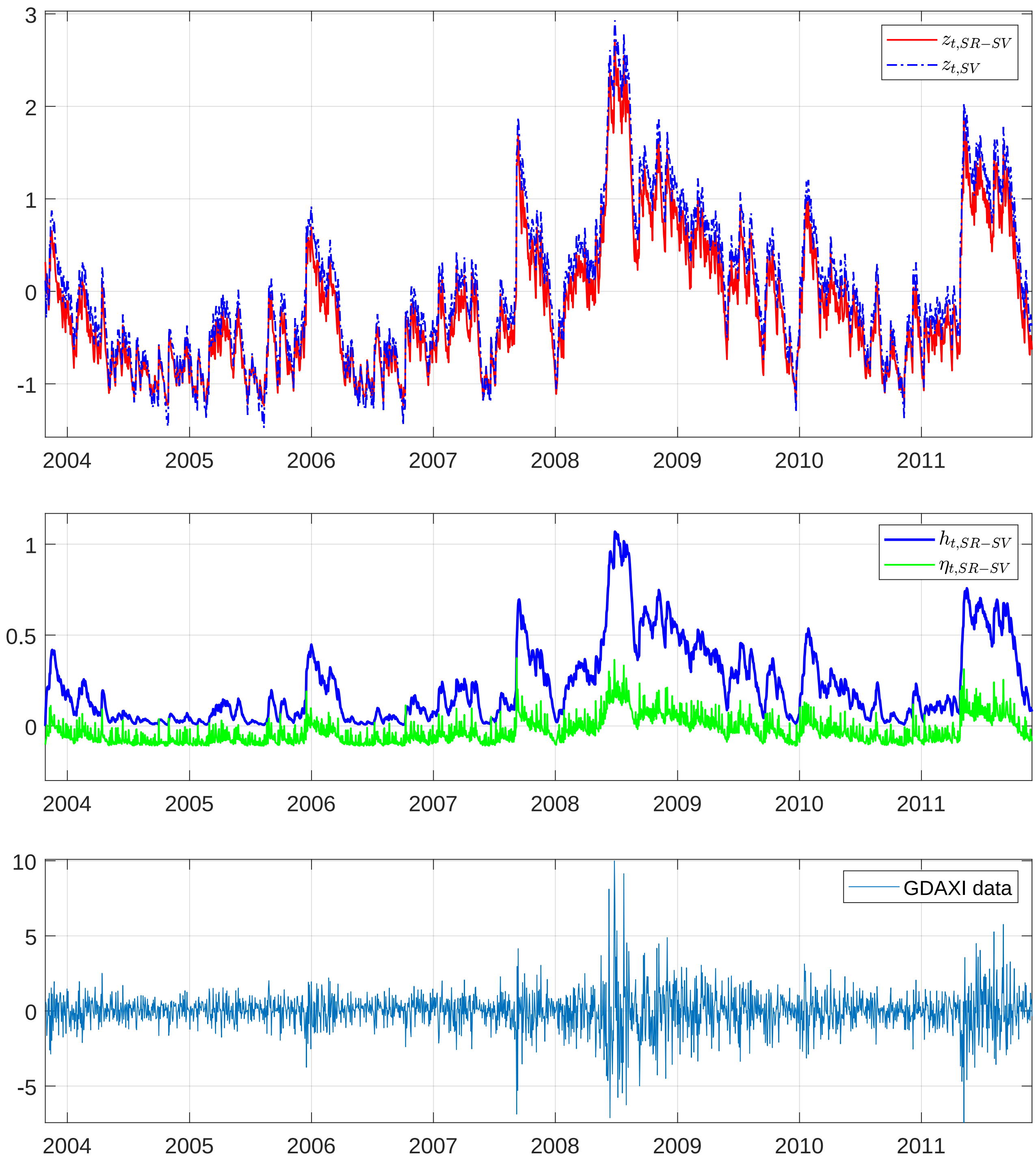

Figure D.4: DAX: (*Top*) The filtered log conditional variance of the SR-SV and SV models. (*Middle*) The filtered values of η_t and h_t of the SR-SV model. (*Bottom*) The in-sample data. (This is better viewed in colour).

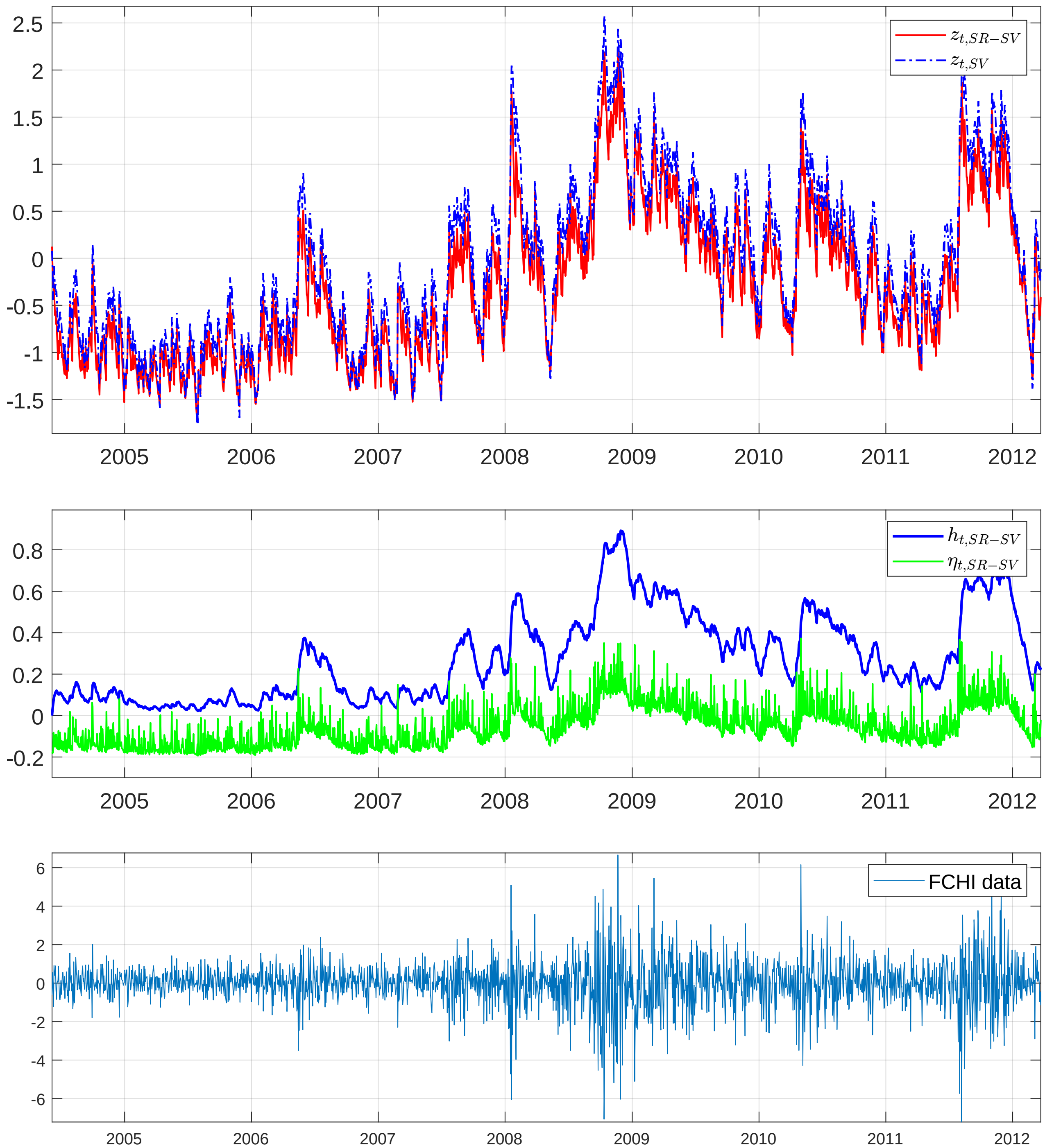

Figure D.5: FCHI: (*Top*) The filtered log conditional variance of the SR-SV and SV models. (*Middle*) The filtered values of η_t and h_t of the SR-SV model. (*Bottom*) The in-sample data. (This is better viewed in colour).

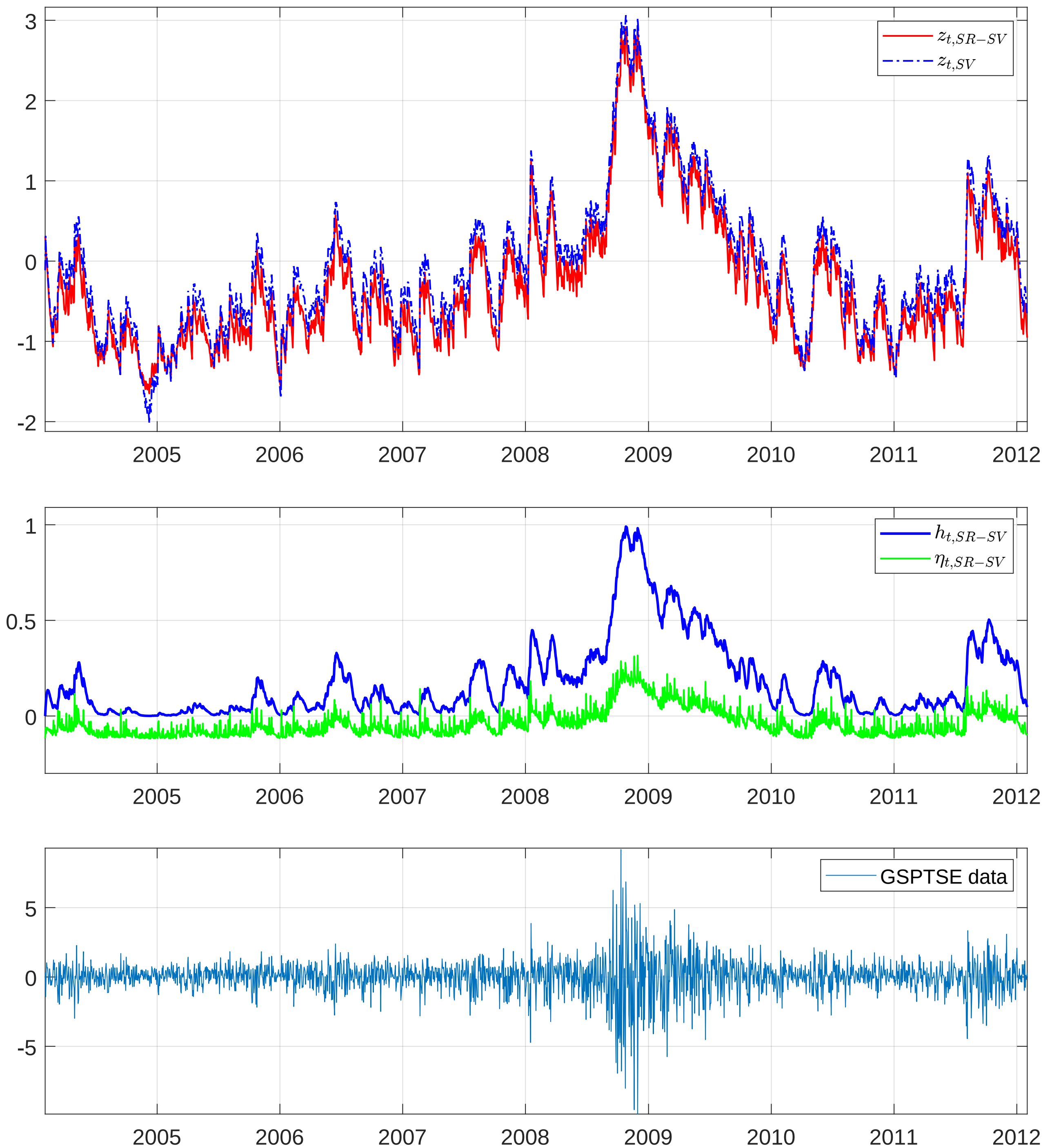

Figure D.6: TSX: (*Top*) The filtered log conditional variance of the SR-SV and SV models. (*Middle*) The filtered values of η_t and h_t of the SR-SV model. (*Bottom*) The in-sample data. (This is better viewed in colour).

Measure		PPS	MSE_1	MSE_2	MAE_1	MAE_2	QLIKE	R^2LOG	Count
	SV	1.368	0.099	0.763	0.234	0.485	0.853	0.423	0
		(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	
BV	N-SV	1.368	0.099	0.769	0.234	0.487	0.850	0.422	0
		(0.000)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	
	LMSV		0.108	0.805	0.245	0.501	0.862	0.435	0
	GP-Vol		0.108	0.857	0.235	0.480	0.907	0.449	0
	SR-SV	**1.365**	**0.090**	**0.745**	**0.220**	**0.452**	**0.847**	**0.386**	7
		(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	
	SV		0.094	0.582	0.237	0.486	0.810	0.443	0
			(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)	
MedRV	N-SV		0.094	0.585	0.238	0.489	0.810	0.447	0
			(0.001)	(0.004)	(0.001)	(0.002)	(0.001)	(0.002)	
	LMSV		0.099	0.612	0.245	0.504	0.825	0.467	0
	GP-Vol		0.111	0.725	0.248	0.503	0.880	0.494	0
	SR-SV		**0.088**	**0.580**	**0.227**	**0.462**	**0.807**	**0.417**	6
			(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	
	SV		0.126	0.994	0.268	0.558	0.858	0.581	0
			(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	
RKV	N-SV		0.126	1.001	0.269	0.559	0.854	0.581	0
			(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)	
	LMSV		0.135	1.023	0.279	0.578	0.878	0.603	0
	GP-Vol		0.132	1.076	0.266	0.536	0.908	0.594	0
	SR-SV		**0.116**	**0.974**	**0.254**	**0.516**	**0.851**	**0.538**	6
			(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	
	SV		0.107	0.909	0.239	0.501	0.859	0.442	0
			(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	
RV	N-SV		0.108	0.915	0.239	0.503	0.855	0.441	0
			(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	
	LMSV		0.115	0.936	0.250	0.527	0.875	0.464	0
	GP-Vol		0.115	1.000	0.237	0.490	0.909	0.460	0
	SR-SV		**0.098**	**0.889**	**0.223**	**0.462**	**0.851**	**0.400**	6
			(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	

Table D.2: DAX data: Forecast performance of the SR-SV and benchmark models using different realized measures. In each panel, the bold numbers indicate the best predictive scores. In each panel, we count the number of times a particular model has lowest (best) predictive scores and list these numbers in the last column; the model with the highest count is preferred.

Measure		PPS	MSE_1	MSE_2	MAE_1	MAE_2	QLIKE	R^2LOG	Count
	SV	1.131	0.069	0.497	0.186	0.319	0.359	0.390	
		(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	
BV	N-SV	1.130	0.067	0.499	0.182	0.313	0.357	0.376	0
		(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	
	LMSV		0.076	0.516	0.205	0.343	0.370	0.423	0
	GP-Vol		**0.057**	**0.451**	0.161	0.272	**0.348**	0.327	3
	SR-SV	**1.127**	0.060	0.504	**0.152**	**0.261**	0.355	**0.294**	4
		(0.000)	(0.00)	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)	
	SV		0.066	0.371	0.191	0.317	0.347	0.435	0
			(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	
MedRV	N-SV		0.065	0.370	0.188	0.312	0.344	0.423	0
			(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	
	LMSV		0.073	0.396	0.207	0.335	0.356	0.469	0
	GP-Vol		**0.058**	**0.336**	0.172	0.280	0.346	0.383	2
	SR-SV		0.059	0.389	**0.164**	**0.272**	**0.341**	**0.338**	4
			(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	
	SV		0.100	0.740	0.230	0.385	0.366	0.665	0
			(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	
RKV	N-SV		0.098	0.741	0.226	0.380	0.364	0.648	0
			(0.000)	(0.002)	(0.001)	(0.001)	(0.000)	(0.002)	
	LMSV		0.112	0.764	0.246	0.405	0.380	0.744	0
	GP-Vol		**0.082**	**0.681**	0.198	0.328	**0.337**	0.554	3
	SR-SV		0.087	0.748	**0.194**	**0.323**	0.360	**0.519**	3
			(0.000)	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)	
	SV		0.069	0.522	0.186	0.318	0.367	0.390	0
			(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	
RV	N-SV		0.068	0.524	0.181	0.311	0.365	0.374	0
			(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	
	LMSV		0.077	0.552	0.204	0.353	0.386	0.419	0
	GP-Vol		**0.057**	**0.475**	0.159	0.269	**0.352**	0.320	3
	SR-SV		0.060	0.530	**0.150**	**0.258**	0.361	**0.291**	3
			(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)	

Table D.3: HSI data: Forecast performance of the SR-SV and benchmark models using different realized measures. In each panel, the bold numbers indicate the best predictive scores. In each panel, we count the number of times a particular model has lowest (best) predictive scores and list these numbers in the last column; the model with the highest count is preferred.

Measure		PPS	MSE_1	MSE_2	MAE_1	MAE_2	QLIKE	R^2LOG	Count
	SV	1.384	0.108	1.076	0.235	0.504	0.863	0.426	0
		(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	
BV	N-SV	1.383	0.108	1.086	0.234	0.507	0.860	0.420	0
		(0.000)	(0.000)	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)	
	LMSV		0.118	1.104	0.246	0.527	0.872	0.469	0
	GP-Vol		0.147	1.616	0.258	0.575	0.959	0.499	0
	SR-SV	**1.381**	**0.095**	**1.057**	**0.210**	**0.448**	**0.856**	**0.354**	7
		(0.000)	(0.00)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	
	SV		0.100	0.670	0.238	0.500	0.833	0.543	0
			(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	
MedRV	N-SV		0.100	0.672	0.237	0.501	0.832	0.538	0
			(0.000)	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)	
	LMSV		0.112	0.695	0.247	0.513	0.849	0.582	0
	GP-Vol		0.147	1.237	0.267	0.583	0.942	0.639	0
	SR-SV		**0.090**	**0.665**	**0.216**	**0.452**	**0.828**	**0.472**	6
			(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	
	SV		0.158	1.271	0.301	0.624	0.901	0.750	0
			(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	
RKV	N-SV		0.159	1.285	0.301	0.628	0.896	0.745	0
			(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)	
	LMSV		0.168	1.332	0.315	0.656	0.908	0.815	0
	GP-Vol		0.177	1.743	0.298	0.645	0.948	0.706	0
	SR-SV		**0.139**	**1.229**	**0.275**	**0.562**	**0.890**	**0.645**	6
			(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	
	SV		0.103	0.908	0.232	0.495	0.877	0.411	0
			(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	
RV	N-SV		0.103	0.920	0.232	0.498	0.873	0.404	0
			(0.000)	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)	
	LMSV		0.113	0.945	0.245	0.521	0.884	0.449	0
	GP-Vol		0.142	1.441	0.258	0.571	0.972	0.486	0
	SR-SV		**0.090**	**0.880**	**0.209**	**0.440**	**0.869**	**0.340**	6
			(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	

Table D.4: FCHI data: Forecast performance of the SR-SV and benchmark models using different realized measures. In each panel, the bold numbers indicate the best predictive scores. In each panel, we count the number of times a particular model has lowest (best) predictive scores and list these numbers in the last column; the model with the highest count is preferred.

Measure		PPS	MSE_1	MSE_2	MAE_1	MAE_2	QLIKE	R^2LOG	Count
	SV	1.004	0.074	0.904	0.192	0.295	0.106	0.542	0
		(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	
BV	N-SV	1.003	0.074	0.902	0.192	0.294	**0.105**	0.543	1
		(0.000)	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	
	LMSV		0.091	0.965	0.217	0.331	0.197	0.676	0
	GP-Vol		0.093	0.966	0.217	0.319	0.222	0.737	0
	SR-SV	**1.001**	**0.069**	**0.900**	**0.181**	**0.273**	0.107	**0.517**	5
		(0.000)	(0.00)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	
	SV		0.074	0.289	0.210	0.310	0.098	1.052	0
			(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
MedRV	N-SV		0.073	0.291	0.210	0.309	0.098	1.044	0
			(0.000)	(0.002)	(0.001)	(0.000)	(0.001)	(0.001)	
	LMSV		0.096	0.360	0.239	0.353	0.209	0.839	0
	GP-Vol		0.097	0.365	0.240	0.342	0.229	0.895	0
	SR-SV		**0.069**	**0.287**	**0.201**	**0.291**	0.098	**0.985**	5
			(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	
	SV		0.096	0.349	0.246	0.357	0.134	0.987	0
			(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	
RKV	N-SV		0.096	0.347	0.246	0.355	0.134	0.991	0
			(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	
	LMSV		0.110	0.392	0.263	0.378	0.220	1.120	0
	GP-Vol		0.105	0.386	0.254	0.355	0.217	1.122	0
	SR-SV		**0.089**	**0.341**	**0.236**	**0.336**	**0.131**	**0.951**	6
			(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	
	SV		0.087	1.370	0.206	0.319	0.118	0.625	0
			(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.002)	
RV	N-SV		0.087	1.368	0.206	0.317	**0.117**	0.627	1
			(0.000)	(0.002)	(0.000)	(0.001)	(0.001)	(0.002)	
	LMSV		0.100	1.418	0.224	0.342	0.195	0.742	0
	GP-Vol		0.099	1.410	0.223	0.329	0.215	0.791	0
	SR-SV		**0.081**	**1.363**	**0.195**	**0.295**	0.119	**0.597**	5
			(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)	

Table D.5: TSX data: Forecast performance of the SR-SV and benchmark models using different realized measures. In each panel, the bold numbers indicate the best predictive scores. In each panel, we count the number of times a particular model has lowest (best) predictive scores and list these numbers in the last column; the model with the highest count is preferred.

References

Breidt, F., Crato, N., and de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83(1):325 – 348.

Deligiannidis, G., Doucet, A., and Pitt, M. K. (2018). The correlated pseudo marginal method. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(5):839–870.

Doornik, J. A. and Ooms, M. (2003). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics and Data Analysis*, 42(3):333 – 348. Computational Ecomometrics.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14:179–21.

Garthwaite, P., Fan, Y., and Sisson, S. (2010). Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process. *Communications in Statistics - Theory and Methods*, 45.

Gerber, M. and Chopin, N. (2014). Sequential Quasi-Monte Carlo. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Gunawan, D., Dang, K., Quiroz, M., Kohn, R., and Tran, M. (2018). Subsampling sequential Monte Carlo for static Bayesian models. *arXiv:1805.03317*.

Harvey, A. C. (2007). 16 - long memory in stochastic volatility. In Knight, J. and Satchell, S., editors, *Forecasting Volatility in the Financial Markets (Third Edition)*, Quantitative Finance, pages 351 – 363. Butterworth-Heinemann, Oxford.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53(1):165 – 188.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.*, 2(5):423–434.