WORKING PAPER

THE TIME-VARYING ASYMMETRY OF EXCHANGE RATE RETURNS: A STOCHASTIC VOLATILITY – STOCHASTIC SKEWNESS MODEL

Martin Iseringhausen

March 2018
2018/944
The Time-Varying Asymmetry of Exchange Rate Returns:
A Stochastic Volatility - Stochastic Skewness Model

Martin Iseringhausen*
Ghent University

March 2018

Abstract

While the volatility of financial returns has been extensively modelled as time-varying, skewness is usually either assumed constant or neglected by assuming symmetric model innovations. However, it has long been understood that accounting for (time-varying) asymmetry as a measure of crash risk is important for both investors and policy makers. This paper extends a standard stochastic volatility model to account for time-varying skewness. We estimate the model by extensions of traditional Bayesian Markov Chain Monte Carlo (MCMC) methods for stochastic volatility models. When applying this model to the returns of four major exchange rates, skewness is found to vary substantially over time. The results support a potential link between carry trading and crash risk. Finally, investors appear to demand compensation for a negatively skewed return distribution.

JEL classification: C11, C58, F31
Keywords: Bayesian analysis, crash risk, foreign exchange, time variation

*The author would like to thank Gerdie Everaert, Joshua Chan, David Veredas, and the members of the Macroeconomics, Policy, and Econometrics Research Group at Ghent University for helpful comments. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the Flemish Supercomputer Center, funded by Ghent University; the Hercules Foundation; and the Economy, Science, and Innovation Department of the Flemish Government. Martin Iseringhausen gratefully acknowledges financial support from Ghent University’s Special Research Fund (BOF). Correspondence to: Martin Iseringhausen, Ghent University, Faculty of Economics and Business Administration, Campus Tweeckerken, Sint Pietersplein 6, 9000 Ghent, Belgium. E-mail: Martin.Iseringhausen@UGent.be.
1 Introduction

Stochastic volatility models are widely used in order to model time variation in the volatility of financial returns. In addition, previous work suggests that returns are neither symmetrically distributed nor is the degree of asymmetry invariant over time. This paper develops an empirical model to capture time-varying skewness within a stochastic volatility framework.

Among financial time series, the return distributions of exchange rates show particularly pronounced time-varying asymmetry (see e.g. Bakshi et al., 2008; Carr and Wu, 2007; Johnson, 2002). Brunnermeier et al. (2009) suggest that time-varying crash risk is linked to the currency carry trade. This investment strategy relies on borrowing in a low interest rate ('funding') currency and investing in a high interest rate ('investment') currency. According to the uncovered interest rate parity (UIP), this strategy should not be profitable as the interest rate differential is expected to be offset by a depreciation of the investment currency. However, empirically the reverse holds ('forward premium puzzle'), thus making the carry a profitable trading strategy (Fama, 1984). When the carry trade 'unwinds', i.e. investors start to suddenly sell the investment currency, this can lead to extreme exchange rate movements. Brunnermeier et al. (2009) find that high interest rate differentials predict negative skewness both in the cross section and over time. According to the authors, the carry trade 'unwinds', i.e. currencies crash, when speculators face funding constraints.

From a theoretical perspective, the implications of skewness for asset pricing have been studied by Kraus and Litzenberger (1976) and Harvey and Siddique (2000). The authors develop extensions of the traditional Capital Asset Pricing Model (CAPM) where investors make portfolio choices not only within the standard mean-variance framework but also take (co-)skewness into account. In particular, investors demand a risk premium for assets that exhibit negative skewness, i.e. that carry significant crash risk. Recently, Farhi et al. (2015) provide evidence that disaster risk accounts for more than a third of the average carry trade risk premium in G10 currencies.1 Similarly, Burnside et al. (2011) suggest that carry returns are explained by 'Peso problems' and develop a hedged version of the carry trade that protects investors against large losses.

Unlike in the stochastic volatility (SV) literature, time-varying skewness or, more generally, higher moment dynamics, have been extensively modelled in (generalized) autoregressive conditional heteroscedasticity (GARCH) models starting off from Hansen (1994). Numerous papers have followed using different innovation distributions and applications (see e.g. Harvey and Siddique, 1999; Jondeau and Rockinger, 2003; Christoffersen et al., 2006). However, as noted by Feunou and Tedongap (2012), their flexibility is limited as both volatility and skewness remain deterministic and are assumed to undergo the same return shocks. An alternative is to introduce asymmetry in a SV framework, where both volatility and skewness are independent stochastic processes. Only few papers have followed this path. Feunou and Tedongap (2012) develop an affine multivariate latent factor model for returns. In their model the return shocks have a standardized inverse-Gaussian distribution conditional on the factors. While the model offers a great

1Koijen et al. (2018) show that the concept of 'carry' is applicable to any asset. However, the authors note that currencies are unique in the sense that crash risk can explain the currency carry premium while it is unlikely to explain carry premiums of other assets.
deal of flexibility, stochastic volatility and skewness are still generated by the same underlying factors and hence not completely independent stochastic processes. Nakajima (2013) introduces a stochastic volatility model with leverage where the innovations are distributed according to the generalized hyperbolic skew Student t-distribution. Stochastic skewness is modelled by specifying the asymmetry parameter as a first-order Markov switching process. But since ex ante little is known about the dynamic evolution of skewness, a regime switching process with a small number of different regimes, appears rather restrictive.

Our model allows for a flexible evolution of skewness over time. A standard stochastic volatility model is extended to allow for stochastic skewness. To this end, the assumption of Gaussian shocks is replaced by shocks coming from the noncentral t-distribution. This distribution features asymmetry and excess kurtosis both of which have been documented in financial returns. To capture time-varying skewness, the asymmetry parameter of this distribution is specified as an autoregressive process. The resulting stochastic volatility - stochastic skewness (SVSS) model treats volatility and skewness on an equal footing by allowing both to evolve according to independent and flexible stochastic processes. The SVSS model generates a simple instantaneous skewness measure for a single time series. This is different from previous work that has studied skewness of financial returns either by computing it within (overlapping) periods (e.g. Amaya et al., 2015; Brunnermeier et al., 2009) or by relying on more complex option pricing models (e.g. Bakshi et al., 2008; Carr and Wu, 2007). We show that the SVSS model can be estimated by straightforward extensions of standard Bayesian Markov Chain Monte Carlo (MCMC) techniques for stochastic volatility models (Kim et al., 1998; Omori et al., 2007). To speed up computation when applying the model to daily data, where often $T > 10,000$, recently developed fast sparse matrix algorithms are used (Chan and Jeliazkov, 2009; McCausland et al., 2011).

The results of Monte Carlo experiments indicate that the proposed model performs well for samples sizes typically encountered when analysing financial returns at daily frequency provided that asymmetry is not too weak. When applying the SVSS model to daily nominal exchange rate returns of four major currencies relative to the U.S. Dollar over the period 01/01/1977 - 10/31/2017, evidence is found in favour of time-varying asymmetry. First, the model generally fits the data well compared with simpler nested stochastic volatility models. Second, skewness is largely negative in typical 'investment currencies' such as the Australian Dollar and positive in 'funding currencies' such as the Japanese Yen. Third, the inverse link between the interest rate differential and estimated skewness suggests carry trading as at least an amplifier of crash risk in exchange rates. Finally, crash risk is positively related to the excess risk premium asked by investors. In summary, the results point, next to well-known phenomena such as volatility clustering, to time-varying skewness as an important feature of exchange rate returns.

The remainder of the paper is structured as follows: Section 2 presents the SVSS model and discusses estimation. Afterwards, Monte Carlo evidence is shown in Section 3. Section 4 applies the model to exchange rate returns and discusses the results. Finally, Section 5 concludes.

---

2The same distribution has been used by Harvey and Siddique (1999) to introduce conditional skewness in a GARCH framework.

3Throughout, we interpret a lower skewness value as higher crash risk and use the terms interchangeably.
2 A stochastic volatility - stochastic skewness model

In this section, we develop an empirical model to estimate time-varying skewness. First, we consider its main building block, the noncentral t-distribution. Afterwards, the full model specification is described and estimation using Bayesian MCMC methods is discussed.

2.1 The noncentral t-distribution

Since the goal is to statistically model (time-varying) asymmetry, the assumption of normally distributed shocks to the dependent variable is dropped. Instead, a distribution is used that allows for both asymmetric shocks, i.e. nonzero skewness, as well as for higher probabilities of tail events than implied by the normal distribution, i.e. excess kurtosis. While in principal a large number of distributions allows for these features (and has been previously used), a particularly simple choice is the noncentral t-distribution (see Johnson et al., 1995, for an overview). A random variable $X$ is noncentral t-distributed with $\nu$ degrees of freedom and noncentrality parameter $\delta$, i.e. $X \sim \text{NCT}(\nu, \delta)$, if it has the following stochastic representation:

$$X = \sqrt{\lambda}(z + \delta), \quad \text{where} \quad \lambda \sim \text{IG}(\nu/2, \nu/2) \quad \text{and} \quad z \sim \mathcal{N}(0, 1).$$

Conveniently, for $\delta = 0$ the noncentral t-distribution collapses to its symmetric counterpart, the Student t-distribution. If, in addition, $\nu \to \infty$, it simplifies further to the standard normal distribution. All moments of the noncentral t-distribution are jointly determined by the parameters $\nu$ and $\delta$. In particular, the central moments of the noncentral t-distribution can be expressed as polynomials of $\delta$ whose coefficients are functions of $\nu$ (Hogben et al., 1961). Using their formulas, mean, variance, skewness and kurtosis can be straightforwardly computed. To illustrate the shape of the noncentral t-distribution, Figure 1 shows density plots depending on $\nu$ and $\delta$ in comparison with the standard normal distribution.

**Figure 1:** (De-meaned) Noncentral t-distribution vs. standard normal distribution

(a) Degrees of freedom: $\nu = 3$
(b) Degrees of freedom: $\nu = 8$
For $\delta > 0$, the distribution is positively skewed which implies a larger right than left tail whereas $\delta < 0$ causes the distribution to be negatively skewed. The moments of the noncentral t-distribution are strongly linked, i.e. positive (negative) skewness mechanically goes along with a positive (negative) mean. This is an undesirable feature for the purpose of this paper as, later on, variations in $\delta$ are supposed to capture changes in the asymmetry of the distribution rather than changes in the mean. To ensure that $\delta$ does not affect the mean, in what follows we consider the de-meaned version of the noncentral t-distribution.

### 2.2 Model specification

We start from the following univariate stochastic volatility model,

\begin{align}
  y_t &= e^{h_{t/2}} \varepsilon_t, \\
  h_t &= \mu_h + \phi_h (h_{t-1} - \mu_h) + \eta^h_t, \\
  \eta^h_t &\sim \mathcal{N}(0, \sigma^2_h), \quad |\phi_h| < 1,
\end{align}

where $h_t$ is the latent (log-)volatility process assumed to evolve according to a stationary AR(1) process and $\varepsilon_t$ is a zero mean shock term. Depending on the distributional assumption about $\varepsilon_t$, various SV models arise. If $\varepsilon_t$ is, for example, assumed to be standard normal, one obtains the well-known standard normal stochastic volatility model as among others discussed in Kim et al. (1998). In order to allow for deviations from normality, Tsiotas (2012) proposes the noncentral t-distribution as discussed in Section 2.1. We follow this suggestion but additionally allow the noncentrality parameter $\delta_t$ to vary stochastically over time to model time-varying skewness. The error term of the SVSS model is thus assumed to follow a de-meaned noncentral t-distribution with $\nu$ degrees of freedom and time-varying noncentrality parameter $\delta_t$:

\begin{equation}
  \varepsilon_t = u_t - E[u_t], \quad \text{with} \quad u_t \sim \mathcal{NCT}(\nu, \delta_t), \quad \text{and} \quad E[u_t] = c_{11}(\nu)\delta_t, \quad \text{if} \ \nu > 1.
\end{equation}

The exact functional form of the coefficient $c_{11}(\nu)$ is given in Appendix A. The law of motion for the noncentrality parameter $\delta_t$ is, analogous to $h_t$, given by:

\begin{equation}
  \delta_t = \mu_\delta + \phi_\delta (\delta_{t-1} - \mu_\delta) + \eta^\delta_t, \quad \eta^\delta_t &\sim \mathcal{N}(0, \sigma^2_\delta), \quad |\phi_\delta| < 1.
\end{equation}

Hence, the SVSS model is composed of the observation equation obtained by merging Equations (2) and (4) and the state Equations (3) and (5). In order to make this specification operational for Bayesian estimation, the stochastic representation in Equation (1) is explored while taking into account the de-meaning of the error term. The observation equation of the SVSS model is therefore re-written as

\begin{equation}
  y_t = e^{h_{t/2}} \varepsilon_t = e^{h_{t/2}} \left( \sqrt{\lambda_t} (z_t + \delta_t) - c_{11}(\nu)\delta_t \right),
\end{equation}

where again $\lambda_t \sim IG(\nu/2, \nu/2)$ and $z_t \sim \mathcal{N}(0, 1)$. A few remarks on this specification are appropriate: First, exploring the stochastic representation in Equation (6) is preferable to working with the relatively complex probability density function of the noncentral t-distribution. It explores
the principle of data augmentation (Tanner and Wong, 1987) by introducing the latent variable \( \lambda_t \) which facilitates the implementation of a MCMC algorithm in Section 2.4.\(^4\) Second, the choices for the stochastic volatility and skewness processes require motivation. Assuming a stationary AR(1) process for (log-)volatility is common in the literature. The persistence parameter \( \phi_h \) is typically close to 1 indicating strong volatility clustering. The general version of the SVSS model adopts the same stochastic process for the noncentrality state \( \delta_t \). The simulation experiments in Section 3 show that the parameters of the AR(1) process in Equation (5) can be precisely estimated provided that stochastic skewness in the data generating process is not too weak. In contrast, especially the persistence parameter \( \phi_\delta \) turns out to be difficult to pin down when the signal in the time-varying noncentrality parameter \( \delta_t \) is not very strong. For this reason, we also consider a more parsimonious (restricted) version of the SVSS model where the noncentrality parameter \( \delta_t \) is specified as a (driftless) random walk, i.e. \( \mu_\delta = 0 \) and \( \phi_\delta = 1 \). The random walk specification has been a popular choice in time-varying parameter models and recent work has also discussed robustness in case of misspecification (see e.g. Antolin-Diaz et al., 2017, for a discussion). To complete the model specification, we assume the following independent prior distributions for the parameters \( \mu_h, \phi_h, \sigma_h^2, \mu_\delta, \phi_\delta, \sigma_\delta^2, \) and \( \nu \):

\[
\begin{align*}
\mu_h &\sim \mathcal{N}(\mu_{h0}, V_{\mu_h}), & \phi_h &\sim \mathcal{N}(\phi_{h0}, V_{\phi_h}) I(|\phi_h| < 1), & \sigma_h^2 &\sim \mathcal{IG}(c_{h0}, C_{h0}), \\
\mu_\delta &\sim \mathcal{N}(\mu_{\delta0}, V_{\mu_\delta}), & \phi_\delta &\sim \mathcal{N}(\phi_{\delta0}, V_{\phi_\delta}) I(|\phi_\delta| < 1), & \sigma_\delta^2 &\sim \mathcal{IG}(c_{\delta0}, C_{\delta0}), \\
\nu &\sim \mathcal{U}(0, \bar{\nu}).
\end{align*}
\]

Using the expressions for the second and third central moment of the noncentral t-distribution derived by Hogben et al. (1961), the time-varying variance and skewness of the SVSS model are given by

\[
\begin{align*}
\text{Var}[y_t|\delta_t, \nu] &= e^{h_t} \left[ c_{22}(\nu) \delta_t^2 + c_{20}(\nu) \right], & \text{if } \nu > 2, \\
\text{Skew}[y_t|\delta_t, \nu] &= \frac{c_{33}(\nu) \delta_t^3 + c_{31}(\nu) \delta_t}{\left[ c_{22}(\nu) \delta_t^2 + c_{20}(\nu) \right]^{3/2}}, & \text{if } \nu > 3,
\end{align*}
\]

where the functional forms of the coefficients \( c_{20}(\nu), c_{22}(\nu), c_{31}(\nu) \) and \( c_{33}(\nu) \) can be found in Appendix A. Technically, \( \delta_t \) induces time variation in all higher moments. However, when fitting the model to the data, dynamics in the lower order moments dominate the higher order ones. Since the error term of the SVSS model has zero mean and the scale parameter \( h_t \) captures changes in the (log-)second moment, the shape parameter \( \delta_t \) adjusts to reflect changes in the third moment. Finally, for future use, define \( y = (y_1, \ldots, y_T)' \), \( h = (h_1, \ldots, h_T)' \), \( \delta = (\delta_1, \ldots, \delta_T)' \), and \( \lambda = (\lambda_1, \ldots, \lambda_T)' \).

---

\(^4\)This is because conditional on the latent variable \( \lambda_t \), the noncentral t-distribution can be expressed as a location-scale mixture of normal distributions such that standard estimation methods for Gaussian models remain applicable (Tsionas, 2002).
2.3 An extended mixture representation

Before describing the Bayesian estimation approach for the SVSS model presented in Section 2.2, a crucial aspect for estimation is pointed out. In their seminal paper Kim et al. (1998) have developed the so-called auxiliary sampler. This approach to estimate the unobserved (log-)volatility series $h$ has become a widely used tool in the Bayesian estimation of stochastic volatility models. The authors estimate $h$ from the following transformed version of Equation (2),

$$\log(y_t^2 + c) = h_t + \tilde{e}_t,$$  \hspace{1cm} (10)

with $\tilde{e}_t = \log(e_t^2)$ and where $h_t$ enters the model now in a linear manner and could, in principle, be estimated using, for example, the Kalman filter. $c = 0.001$ is an offset constant to ensure numerical stability for small values of $y_t^2$. However, the transformed error term is no longer standard normal but log-$\chi^2$ distributed. Kim et al. (1998) approximate $\tilde{e}_t$ with a seven component mixture of normal distributions. Conditional on the mixture indicators $s = (s_1, ..., s_T)'$, which are sampled together with the other parameters, the model is Gaussian and the Kalman filter becomes applicable. This paper builds on the approach of Kim et al. (1998) but extends their method to deal with the different specification of the error term. In particular, the transformed error term of the SVSS model is

$$\tilde{e}_t = \log \left( \left( \sqrt{\lambda_t(z_t + \delta_t)} - c_{11}(\nu)\delta_t \right) \right),$$ \hspace{1cm} (11)

where $\tilde{e}_t$ is now log-noncentral-t-squared distributed. This implies that a suitable normal mixture approximation depends on the values of $\nu$ and $\delta_t$:

$$f(\tilde{e}_t | \nu, \delta_t) = \sum_{j=1}^{M} q_j(\nu, \delta_t) f_N(\tilde{e}_t | m_j(\nu, \delta_t), v_j^2(\nu, \delta_t)),$$ \hspace{1cm} (12)

where $q_j(\nu, \delta_t)$ is the component probability of a specific normal distribution with mean $m_j(\nu, \delta_t)$ and variance $v_j^2(\nu, \delta_t)$ given a certain parameter combination $[\nu, \delta_t]$. This mixture can equivalently be expressed in terms of component probabilities,

$$\tilde{e}_i | (s_i = j) \sim N \left( m_j(\nu, \delta_t), v_j^2(\nu, \delta_t) \right), \hspace{1cm} Pr(s_i = j) = q_j(\nu, \delta_t).$$ \hspace{1cm} (13)

To implement this extended mixture approximation, samples from the log-noncentral-t-squared distribution in Equation (11) are generated and mixtures of normal distributions are fitted for a large grid of combinations of $\nu$ (3 to 50 with stepsize 0.1) and $\delta$ (-5 to 5 with step size 0.01). We follow Omori et al. (2007) and use $M = 10$ mixture components. Generating the mixtures is a one time computation cost and hence does not affect sampling efficiency. Moreover, the approximation can be made arbitrarily precise by letting the step sizes approach zero and the number of mixture components approach infinity.

$^5$The corresponding values for all mixture components, i.e. means, variances, and component probabilities, can be obtained from the author upon request.
2.4 Bayesian MCMC algorithm

The stochastic volatility - stochastic skewness model, like the standard normal SV model, does not permit to write down the likelihood function in closed form making standard maximum likelihood estimation infeasible. Instead, the SVSS model presented in the previous section is estimated using Bayesian Markov Chain Monte Carlo (MCMC) methods. In particular, we simulate draws from the intractable joint and marginal posterior distributions of the parameters and unobserved states using Gibbs sampling, which only exploits conditional distributions. These are usually easy to derive and belong to well-known distributional families from which samples can be readily obtained. However, some of the conditional distributions are non-standard and sampling can be achieved by implementing a Metropolis-Hastings step. What follows is only a brief overview of the estimation procedure. Details on the conditional distributions and sampling techniques can be found in Appendix B. The SVSS model is split up into the following blocks:

1. Sample $s$ from $p(s|y, h, \delta, \nu)$;
2. Sample $h$ from $p(h|y, s, \delta, \nu, \mu_h, \phi_h, \sigma_h^2)$;
3. Sample $\lambda$ from $p(\lambda|y, h, \delta, \nu)$;
4. Sample $\nu$ from $p(\nu|\lambda)$;
5. Sample $\delta$ from $p(\delta|y, h, \lambda, \nu, \mu_\delta, \phi_\delta, \sigma_\delta^2)$;
6. Sample $\mu_h$ from $p(\mu_h|h, \phi_h, \sigma_h^2)$ and $\mu_\delta$ from $p(\mu_\delta|\delta, \phi_\delta, \sigma_\delta^2)$;
7. Sample $\phi_h$ from $p(\phi_h|h, \mu_h, \sigma_h^2)$ and $\phi_\delta$ from $p(\phi_\delta|\delta, \mu_\delta, \sigma_\delta^2)$;
8. Sample $\sigma_h^2$ from $p(\sigma_h^2|h, \mu_h, \phi_h)$ and $\sigma_\delta^2$ from $p(\sigma_\delta^2|\delta, \mu_\delta, \phi_\delta)$.

Block 1 samples the mixture indicators via the inverse-transform method (Kim et al., 1998). The different mixture components for each period $t$ and each Gibbs iteration $i$ are selected depending on the corresponding (rounded) values of $\nu_i$ and $\delta_{t,i}$. The series of (log-)volatilities $h$ (block 2) and noncentrality parameters $\delta$ (block 5) could in principle be sampled from the corresponding state space models using Kalman filter-based algorithms (e.g. Carter and Kohn, 1994). Instead, this paper relies on recently developed fast sparse matrix algorithms to sample $h$ and $\delta$ which speed up the algorithm significantly (Chan and Hsiao, 2014; Chan and Jeliazkov, 2009). Moreover, the conditional posterior distributions of $\lambda$ (block 3), $\nu$ (block 4) and $\phi = [\phi_h, \phi_\delta]$ (block 7) are non-standard and a Metropolis-Hastings step needs to be included as described in Tsionas (2002), Chan and Hsiao (2014) and Kim et al. (1998), respectively. Finally, the conditional posterior distributions of $\mu = [\mu_h, \mu_\delta]$ (block 6) and $\sigma^2 = [\sigma_h^2, \sigma_\delta^2]$ (block 8) are normal and inverse-gamma such that sampling is standard.

If the model in Equation (6) includes a linear specification for the conditional mean, an additional block to sample the regression coefficients as in Tsionas (2002) is included (see Appendix B). Starting from an arbitrary set of initial values, sampling from these blocks is iterated $J$ times and after a sufficiently long burn-in period $B$, the sequence of draws $(B+1, \ldots, J)$ can be taken as a sample from the joint posterior distribution of interest $f(h, \delta, \lambda, \nu, \phi_h, \mu_h, \sigma_h^2, \phi_\delta, \mu_\delta, \sigma_\delta^2|y)$. 

7
3 Monte Carlo simulations

The goal of this section is to assess the performance of the proposed SVSS model using simulation experiments. To this end, samples of different size are generated from the model given by Equations (2)-(5). We simulate 1,000 datasets for each of the Monte Carlo experiments and estimate the SVSS model with 20,000 Gibbs iterations where 2,000 draws are discarded as burn-in.

Table 1 contains the prior parameters used in the estimations. Given that the large samples considered are supposed to mimic financial returns at the daily frequency, these prior values can be considered almost uninformative. The upper bound of the uniform prior for the degrees of freedom parameter \( \nu \) is based on the consideration that for \( \nu > 50 \) the noncentral t-distribution becomes indistinguishable from the normal distribution.

We start by considering the general version of the SVSS model where \( \delta_t \) is generated and estimated as a stationary AR(1) process. The results are presented in Table 2. The underlying data generating process (DGP) assumes a high degree of volatility persistence (\( \phi_h = 0.99 \)) and asymmetry persistence (\( \phi_\delta = 0.99 \)), identical innovation variances of both (log-)volatility and asymmetry (\( \sigma_h^2 = \sigma_\delta^2 = 0.1^2 \)) as well as moderate fat tails (\( \nu = 10 \)).

As can be seen from Table 2, overall the model estimates the parameters of the (log-)volatility and noncentrality process accurately, even in a relatively small sample of \( T = 1,000 \). Biases for most parameters are either zero or quite small.\(^6\) With respect to the convergence of the Markov chain, the diagnostic of Geweke (1992) indicates convergence for all parameters to be estimated. In order to assess the mixing properties of the chain, the so-called inefficiency factor is reported (see e.g. Chib, 2001). An inefficiency factor of \( m \) indicates that one needs to draw \( m \) times as many MCMC samples as uncorrelated samples. We find that mixing is poor for some of the parameters,\(^7\)

\(^{6}\)Even though bias usually refers to the property of a frequentist estimator, we use the term to describe the average deviation of the estimated posterior mean from the true mean of the parameter’s distribution. The parameters are fixed over Monte Carlo runs. Thus, they can be considered as being drawn from a degenerate distribution.
especially the asymmetry shock variance $\sigma_2^2$. However, the simulation results are robust to changes in the number of Gibbs iterations. Nevertheless, the number of iterations is significantly increased when estimating the model on exchange rate returns in Section 4.

Table 2: Results Monte Carlo simulation: $\delta_t$ specified as AR(1)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Parameter</th>
<th>Mean</th>
<th>SE</th>
<th>2.5%</th>
<th>97.5%</th>
<th>Bias</th>
<th>CD</th>
<th>IF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_h$</td>
<td>0.03</td>
<td>0.25</td>
<td>-0.44</td>
<td>0.50</td>
<td>0.03</td>
<td>0.48</td>
<td>5.37</td>
</tr>
<tr>
<td></td>
<td>$\phi_h$</td>
<td>0.98</td>
<td>0.01</td>
<td>0.96</td>
<td>0.99</td>
<td>-0.01</td>
<td>0.83</td>
<td>80.71</td>
</tr>
<tr>
<td></td>
<td>$\sigma_h^2$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.90</td>
<td>184.44</td>
</tr>
<tr>
<td>$T = 1,000$</td>
<td>$\mu_\delta$</td>
<td>0.00</td>
<td>0.44</td>
<td>-0.86</td>
<td>0.91</td>
<td>0.00</td>
<td>0.87</td>
<td>131.58</td>
</tr>
<tr>
<td></td>
<td>$\phi_\delta$</td>
<td>0.92</td>
<td>0.01</td>
<td>0.90</td>
<td>0.95</td>
<td>-0.07</td>
<td>0.83</td>
<td>60.06</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\delta^2$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.93</td>
<td>402.45</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>18.10</td>
<td>7.70</td>
<td>7.51</td>
<td>33.52</td>
<td>8.10</td>
<td>0.92</td>
<td>231.69</td>
</tr>
<tr>
<td></td>
<td>$\delta_t$</td>
<td>0.02</td>
<td>0.14</td>
<td>-0.26</td>
<td>0.29</td>
<td>0.02</td>
<td>0.47</td>
<td>3.35</td>
</tr>
<tr>
<td></td>
<td>$\phi_h$</td>
<td>0.99</td>
<td>0.00</td>
<td>0.98</td>
<td>0.99</td>
<td>0.00</td>
<td>0.86</td>
<td>106.16</td>
</tr>
<tr>
<td></td>
<td>$\sigma_h^2$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.91</td>
<td>245.49</td>
</tr>
<tr>
<td>$T = 5,000$</td>
<td>$\mu_\delta$</td>
<td>0.00</td>
<td>0.18</td>
<td>-0.36</td>
<td>0.36</td>
<td>0.00</td>
<td>0.82</td>
<td>92.06</td>
</tr>
<tr>
<td></td>
<td>$\phi_\delta$</td>
<td>0.93</td>
<td>0.03</td>
<td>0.90</td>
<td>0.99</td>
<td>-0.06</td>
<td>0.93</td>
<td>307.27</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\delta^2$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.98</td>
<td>1289.48</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>11.12</td>
<td>2.19</td>
<td>8.12</td>
<td>17.07</td>
<td>1.12</td>
<td>0.90</td>
<td>176.21</td>
</tr>
<tr>
<td></td>
<td>$\mu_h$</td>
<td>0.02</td>
<td>0.09</td>
<td>-0.17</td>
<td>0.21</td>
<td>0.02</td>
<td>0.46</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>$\phi_h$</td>
<td>0.99</td>
<td>0.00</td>
<td>0.98</td>
<td>0.99</td>
<td>0.00</td>
<td>0.85</td>
<td>109.58</td>
</tr>
<tr>
<td></td>
<td>$\sigma_h^2$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.90</td>
<td>251.77</td>
</tr>
<tr>
<td>$T = 10,000$</td>
<td>$\mu_\delta$</td>
<td>0.00</td>
<td>0.13</td>
<td>-0.24</td>
<td>0.23</td>
<td>0.00</td>
<td>0.77</td>
<td>71.02</td>
</tr>
<tr>
<td></td>
<td>$\phi_\delta$</td>
<td>0.95</td>
<td>0.03</td>
<td>0.90</td>
<td>0.99</td>
<td>-0.04</td>
<td>0.94</td>
<td>544.74</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\delta^2$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.98</td>
<td>1851.17</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>10.33</td>
<td>1.22</td>
<td>8.50</td>
<td>13.21</td>
<td>0.33</td>
<td>0.89</td>
<td>143.20</td>
</tr>
</tbody>
</table>

True values: $\mu_h = 0$, $\phi_h = 0.99$, $\sigma_h^2 = 0.01$, $\mu_\delta = 0$, $\phi_\delta = 0.99$, $\sigma_\delta^2 = 0.01$, $\nu = 10$. Notes: Mean, SE, 2.5%, 97.5% and bias are the Monte Carlo mean, standard error, 2.5% and 97.5% percentiles and bias, respectively. CD refers to the p-value of the Geweke (1992) convergence diagnostic where the null hypothesis is convergence. IF is the inefficiency factor (Chib, 2001).

The degrees of freedom parameter $\nu$ however, is significantly upward biased for $T = 1,000$. Moreover, there is large uncertainty surrounding the estimate as indicated by the large Monte Carlo standard error. Both bias and uncertainty vanish with increasing sample size. This result is not surprising as pinning down tail risk precisely is a difficult exercise when the number of observations is limited and the tails are not extremely fat (see for example Huisman et al., 2001). In addition, the persistence parameter of the time-varying asymmetry process $\delta_t$, $\phi_\delta$, is downward biased even for $T = 10,000$. While a bias of -0.04 does not appear large at first glance, a slightly lower persistence parameter has severe consequences in large samples resulting in an estimated noncentrality parameter $\delta_t$ (and thus also estimated skewness series) that is too flat and does not
properly capture the true dynamics. However, the problem discussed here is not related to the model itself but reflects a signal in the data that is too weak to allow for a precise estimation of the persistence parameter of $\delta_t$.

**Table 3:** Results Monte Carlo simulation: $\delta_t$ specified as AR(1) and smaller $\nu$

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Parameter</th>
<th>Mean</th>
<th>SE</th>
<th>2.5%</th>
<th>97.5%</th>
<th>Bias</th>
<th>CD</th>
<th>IF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_h$</td>
<td>0.03</td>
<td>0.25</td>
<td>-0.47</td>
<td>0.53</td>
<td>0.03</td>
<td>0.38</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td>$\phi_h$</td>
<td>0.98</td>
<td>0.01</td>
<td>0.95</td>
<td>0.99</td>
<td>-0.01</td>
<td>0.83</td>
<td>89.33</td>
</tr>
<tr>
<td></td>
<td>$\alpha^2_h$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.89</td>
<td>198.85</td>
</tr>
<tr>
<td>$T = 1,000$</td>
<td>$\nu$</td>
<td>6.01</td>
<td>2.00</td>
<td>3.85</td>
<td>11.26</td>
<td>1.01</td>
<td>0.82</td>
<td>78.14</td>
</tr>
<tr>
<td></td>
<td>$\mu_\delta$</td>
<td>0.00</td>
<td>0.34</td>
<td>-0.65</td>
<td>0.73</td>
<td>0.00</td>
<td>0.64</td>
<td>26.00</td>
</tr>
<tr>
<td></td>
<td>$\phi_\delta$</td>
<td>0.94</td>
<td>0.03</td>
<td>0.90</td>
<td>0.99</td>
<td>-0.05</td>
<td>0.83</td>
<td>73.03</td>
</tr>
<tr>
<td></td>
<td>$\alpha^2_\delta$</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.04</td>
<td>0.01</td>
<td>0.95</td>
<td>405.49</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>5.19</td>
<td>0.55</td>
<td>4.42</td>
<td>6.27</td>
<td>0.19</td>
<td>0.81</td>
<td>60.28</td>
</tr>
<tr>
<td>$T = 5,000$</td>
<td>$\mu_h$</td>
<td>0.02</td>
<td>0.14</td>
<td>-0.27</td>
<td>0.29</td>
<td>0.02</td>
<td>0.41</td>
<td>5.21</td>
</tr>
<tr>
<td></td>
<td>$\phi_h$</td>
<td>0.99</td>
<td>0.00</td>
<td>0.98</td>
<td>0.99</td>
<td>0.00</td>
<td>0.87</td>
<td>126.00</td>
</tr>
<tr>
<td></td>
<td>$\alpha^2_h$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.92</td>
<td>272.72</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>5.19</td>
<td>0.55</td>
<td>4.42</td>
<td>6.27</td>
<td>0.19</td>
<td>0.81</td>
<td>60.28</td>
</tr>
<tr>
<td>$T = 10,000$</td>
<td>$\mu_h$</td>
<td>0.01</td>
<td>0.10</td>
<td>-0.17</td>
<td>0.22</td>
<td>0.01</td>
<td>0.42</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>$\phi_h$</td>
<td>0.99</td>
<td>0.00</td>
<td>0.98</td>
<td>0.99</td>
<td>0.00</td>
<td>0.88</td>
<td>129.61</td>
</tr>
<tr>
<td></td>
<td>$\alpha^2_h$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.92</td>
<td>285.71</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>5.07</td>
<td>0.30</td>
<td>4.54</td>
<td>5.71</td>
<td>0.07</td>
<td>0.80</td>
<td>43.50</td>
</tr>
</tbody>
</table>

True values: $\mu_h = 0$, $\phi_h = 0.99$, $\alpha^2_h = 0.01$, $\mu_\delta = 0$, $\phi_\delta = 0.99$, $\alpha^2_\delta = 0.01$, $\nu = 5$. Notes: See Table 2.

To further underpin this point, Table 3 presents results for an experiment that is identical to the previous one except for the degrees of freedom parameter that is now set to $\nu = 5$. First, this results in a more precise estimation of $\nu$ since the tails now contain more information. But second, this also comes along with a more accurate estimation of the asymmetry persistence parameter $\phi_\delta$. When considering the de-meaned noncentral t-distribution, identification of $\delta_t$ depends crucially on $\nu$. In the extreme case of $\nu \to \infty$, $\delta_t$ is not identified. Hence, lower values of $\nu$ lead to more precise estimates of $\delta_t$ and its parameters. Table 3 confirms this as the bias of $\phi_\delta$ is now significantly smaller and even entirely gone for $T = 10,000$. These results highlight that, if $\delta_t$ is to be specified as an AR(1) process, this is in principle possible but demands a strong skewness signal in the underlying DGP.
Since this is not necessarily the case in real-world datasets, an alternative is to apply the more parsimonious random walk specification for $\delta_t$ as done in the following section. Consequently, also simulation results are presented of a scenario where $\delta_t$ is generated and estimated as a (driftless) random walk.\footnote{In this case the DGP of $\delta_t$ is bounded by the interval $[-3; 3]$ which implies a skewness range of $[-0.95; 0.95]$ for $\nu = 10$. This is to avoid unrealistic skewness realizations. For $\nu = 5$, skewness lies within the interval $[-2.5; 2.5]$. Results for the latter case are not reported but available from the author upon request.} The results are presented in Table 4. Not surprisingly, the model parameters of the (log-)volatility and noncentrality process are accurately estimated. However, a bias in the degrees of freedom persists even for a larger sample size.

### Table 4: Results Monte Carlo simulation: $\delta_t$ specified as random walk

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Parameter</th>
<th>Mean</th>
<th>SE</th>
<th>2.5%</th>
<th>97.5%</th>
<th>Bias</th>
<th>CD</th>
<th>IF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_h$</td>
<td>0.01</td>
<td>0.25</td>
<td>-0.46</td>
<td>0.48</td>
<td>0.01</td>
<td>0.47</td>
<td>5.54</td>
</tr>
<tr>
<td>$T = 1,000$</td>
<td>$\phi_h$</td>
<td>0.98</td>
<td>0.01</td>
<td>0.95</td>
<td>0.99</td>
<td>-0.01</td>
<td>0.84</td>
<td>86.08</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_h$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.89</td>
<td>192.11</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_\phi$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.94</td>
<td>337.89</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>21.81</td>
<td>9.77</td>
<td>6.84</td>
<td>39.69</td>
<td>11.81</td>
<td>0.89</td>
<td>206.60</td>
</tr>
<tr>
<td></td>
<td>$\mu_h$</td>
<td>0.02</td>
<td>0.14</td>
<td>-0.25</td>
<td>0.29</td>
<td>0.02</td>
<td>0.55</td>
<td>6.27</td>
</tr>
<tr>
<td>$T = 5,000$</td>
<td>$\phi_h$</td>
<td>0.99</td>
<td>0.00</td>
<td>0.98</td>
<td>0.99</td>
<td>0.00</td>
<td>0.88</td>
<td>120.92</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_h$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.90</td>
<td>256.94</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_\phi$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.97</td>
<td>924.72</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>15.07</td>
<td>5.39</td>
<td>8.44</td>
<td>28.81</td>
<td>5.07</td>
<td>0.93</td>
<td>355.20</td>
</tr>
<tr>
<td></td>
<td>$\mu_h$</td>
<td>0.01</td>
<td>0.10</td>
<td>-0.18</td>
<td>0.20</td>
<td>0.01</td>
<td>0.59</td>
<td>7.16</td>
</tr>
<tr>
<td>$T = 10,000$</td>
<td>$\phi_h$</td>
<td>0.99</td>
<td>0.00</td>
<td>0.98</td>
<td>0.99</td>
<td>0.00</td>
<td>0.87</td>
<td>130.12</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_h$</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.92</td>
<td>273.72</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_\phi$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td>0.97</td>
<td>1207.32</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>13.55</td>
<td>3.16</td>
<td>8.75</td>
<td>20.90</td>
<td>3.55</td>
<td>0.92</td>
<td>482.92</td>
</tr>
</tbody>
</table>

True values: $\mu_h = 0$, $\phi_h = 0.99$, $\sigma^2_h = 0.01$, $\mu_\phi = 0$, $\phi_\phi = 1$, $\sigma^2_\phi = 0.01$, $\nu = 10$. Notes: See Table 2.

To conclude this section, Figure 2 provides some illustrative examples on the flexibility of the random walk specification. The different plots contrast the estimated time-varying noncentrality parameter $\hat{\delta}_t$ with the true parameter for a variety of DGPs and a single Monte Carlo sample. In the first case, the random walk for $\hat{\delta}_t$ is correctly specified and, not surprisingly, the fit is very good. In the second case, the random walk specification is overparametrized as the true underlying distribution is symmetric ($\delta_t = \delta = 0$). This is captured by the random walk specification since the 95% posterior density intervals include zero. The last case can be viewed as a robustness check with respect to misspecification. Here, the true DGP reflects a structural break where $\delta_t = 1$ for $t = 1, \ldots, T/2$ and $\delta_t = -1$ for $t = T/2 + 1, \ldots, T$. Even in this extreme case of misspecification, the random walk provides a reasonable estimate and adjusts quickly to the constant lower level of $\delta_t$.\footnote{In this case the DGP of $\delta_t$ is bounded by the interval $[-3; 3]$ which implies a skewness range of $[-0.95; 0.95]$ for $\nu = 10$. This is to avoid unrealistic skewness realizations. For $\nu = 5$, skewness lies within the interval $[-2.5; 2.5]$. Results for the latter case are not reported but available from the author upon request.}
4 Time-varying asymmetry in exchange rate returns

This section applies the SVSS model to exchange rate returns. The results are compared to those obtained from simpler SV models. In addition, we briefly discuss the potential role of carry trading and the implications of time-varying skewness for asset pricing.

4.1 Data

The dataset contains daily nominal exchange rates of four major currencies relative to the U.S. Dollar (USD) over the period 01/01/1977 - 10/31/2017 ($T = 10,255$). In particular, the Australian Dollar (AUD), the Japanese Yen (JPY), the British Pound (GBP), and the Swiss Franc (CHF) are considered.

Figure 3: Nominal exchange rate returns
The exchange rates $S_t$, which are measured as USD per foreign currency unit, are obtained from the Federal Reserve Economic Data (FRED) database. The nominal returns are calculated as $y_t = \frac{(S_t - S_{t-1})}{S_{t-1}} \times 100$ and displayed in Figure 3. Table 5 contains summary statistics for the four return series under consideration. These numbers, which are unconditional moments over the sample period, point to pronounced non-normality in exchange rate returns. All four return series exhibit unconditional skewness with the USD/AUD and USD/GBP returns being left-tailed while the USD/JPY and USD/CHF returns are right-tailed. Moreover, all four have bigger tails than would be implied by the normal distribution, i.e. a kurtosis greater than three. As expected, the Jarque-Bera test clearly rejects unconditional normality in all four cases.

Table 5: Summary statistics for exchange rate returns

<table>
<thead>
<tr>
<th>Currency pair</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Jarque-Bera</th>
</tr>
</thead>
<tbody>
<tr>
<td>USD/AUD</td>
<td>-0.0010</td>
<td>0.4846</td>
<td>-0.8460</td>
<td>21.0291</td>
<td>0.00</td>
</tr>
<tr>
<td>USD/JPY</td>
<td>0.0115</td>
<td>0.4560</td>
<td>0.4927</td>
<td>7.3560</td>
<td>0.00</td>
</tr>
<tr>
<td>USD/GBP</td>
<td>-0.0005</td>
<td>0.3882</td>
<td>-0.2922</td>
<td>9.3553</td>
<td>0.00</td>
</tr>
<tr>
<td>USD/CHF</td>
<td>0.0115</td>
<td>0.5513</td>
<td>0.6247</td>
<td>19.1881</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: This table contains summary statistics of daily exchange rate returns for four currencies relative to the USD over the period 01/01/1977 - 10/31/2017. The exchange rate is measured as USD per foreign currency unit. The last column contains p-values of the Jarque-Bera test where the null hypothesis is normality.

4.2 Model comparison

In order to assess the empirical relevance of time-varying asymmetry in modelling exchange rate returns, this section compares the estimates obtained from the SVSS model with several alternative SV models, all of which are restricted versions of the SVSS model. The following specification is fitted to the four exchange rate return series where the conditional mean is assumed to evolve according to an autoregressive process:

$$y_t = \beta_0 + \sum_{l=1}^{L} \beta_l y_{t-l} + \epsilon_{t}/\sqrt{\epsilon_t}.$$  \hspace{1cm} (14)

Depending on the distributional assumption about $\epsilon_t$, we distinguish the following four stochastic volatility models: Model 1 is the standard normal stochastic volatility model (SV-sn) as defined in e.g. Kim et al. (1998), i.e. $\epsilon_t \sim \mathcal{N}(0,1)$. Model 2 assumes Student t-distributed innovations (SV-t) as in Chib et al. (2002) and thus allows for fat tails, i.e. $\epsilon_t \sim t(\nu)$. Model 3 allows for both fat tails and a constant degree of asymmetry by imposing a de-meaned noncentral t-distribution (SV-nct), i.e. $\epsilon_t = v_t - E[v_t]$ with $v_t \sim \mathcal{NCT}(\nu, \delta)$. Finally, Model 4 is the previously introduced SVSS model, which allows for time-varying skewness, where $\epsilon_t$ is defined as in Equations (4)-(5). Based on the arguments developed in the previous section, the asymmetry process $\delta_t$ is specified as a random walk, i.e. we set $\mu_\delta = 0$ and $\phi_\delta = 1$ in Equation (5).

\footnote{We set $L = 4$. However, changing the lag length or omitting the conditional mean dynamics as regularly done in the literature on stochastic volatility in financial returns does not affect the remaining estimates.}

43
The prior distributions for the parameters are set as follows: $\beta_l \sim \mathcal{N}(0, 1)$ for $l = 0, \ldots, L$, $\mu_h \sim \mathcal{N}(0, 10)$ and $\phi_h \sim \mathcal{N}(0.95, 1)$. The remaining prior values are as in Table 1. The reported results are based on 50,000 iterations with 10,000 draws being discarded as burn-in. For the purpose of model comparison, we apply the Deviance Information Criterion (DIC) as developed in Spiegelhalter et al. (2002). This measure merges a Bayesian measure of fit with a measure of model complexity where a smaller DIC value indicates a better model. In particular, the conditional version of the DIC, which is based on the likelihood conditional on the unobserved states, is employed.\(^9\) This criterion is easy to compute from MCMC output and has previously been used to compare stochastic volatility models (see e.g. Berg et al., 2004). While it also has been recently criticized, so far alternative approaches involve multiple steps and computationally heavy (model-specific) procedures (see e.g. Li et al., 2015), which limits their feasibility in hierarchical latent variable models and large datasets.\(^{10}\) However, when assessing the model, we do not take the estimated DIC values at face value but also consider the posterior parameter distributions.

Tables 6 - 9 present the estimation results for the four SV models when applied to the exchange rate return series. Overall, the insights gained from the posterior estimates are quite similar across return series. First, the conditional mean is very close to zero as indicated by the small coefficients. This is in line with the literature suggesting that exchange rates evolve closely to a random walk implying near unpredictability (Meese and Rogoff, 1983).

Table 6: USD/AUD exchange rate returns: estimation results

<table>
<thead>
<tr>
<th></th>
<th>SV-sn</th>
<th>SV-t</th>
<th>SV-nct</th>
<th>SVSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coef. Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.009</td>
<td>0.004</td>
<td>0.012</td>
<td>0.004</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.017</td>
<td>0.010</td>
<td>-0.020</td>
<td>0.011</td>
</tr>
<tr>
<td>Mean $\beta_2$</td>
<td>-0.006</td>
<td>0.010</td>
<td>-0.021</td>
<td>0.012</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.020</td>
<td>0.010</td>
<td>-0.029</td>
<td>0.012</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.011</td>
<td>0.010</td>
<td>0.022</td>
<td>0.011</td>
</tr>
<tr>
<td>$\mu_h$</td>
<td>-1.376</td>
<td>0.141</td>
<td>-1.695</td>
<td>0.284</td>
</tr>
<tr>
<td>Variance $\phi_h$</td>
<td>0.987</td>
<td>0.002</td>
<td>0.996</td>
<td>0.001</td>
</tr>
<tr>
<td>$\sigma_h^2$</td>
<td>0.035</td>
<td>0.004</td>
<td>0.009</td>
<td>0.001</td>
</tr>
<tr>
<td>Skewness $\delta$</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Df $\nu$</td>
<td>---</td>
<td>6.297</td>
<td>0.398</td>
<td>6.343</td>
</tr>
<tr>
<td>DIC</td>
<td>16,093</td>
<td>3.61</td>
<td>15,891</td>
<td>0.58</td>
</tr>
</tbody>
</table>

\(^9\)The standard error of the conditional DIC is obtained by a ‘brute force’ approach, i.e. re-estimating the model $R = 100$ times and calculating $SE_{DIC} = \sqrt{\sum_{r=1}^{R}(DIC_r - \bar{DIC})^2}$ (Berg et al., 2004).

\(^{10}\)Chan and Grant (2016) find that the conditional DIC tends to favour overfitted models and Li et al. (2015) argue against its suitability for latent variable models. To provide the reader with an idea about the performance of the conditional DIC in our setting, Appendix C contains a small simulation study.
### Table 7: USD/JPY exchange rate returns: estimation results

<table>
<thead>
<tr>
<th>Coef.</th>
<th>SV-sn</th>
<th>SV-t</th>
<th>SV-nct</th>
<th>SVSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-0.007</td>
<td>0.005</td>
<td>-0.001</td>
<td>0.005</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.000</td>
<td>0.010</td>
<td>0.001</td>
<td>0.010</td>
</tr>
<tr>
<td>Mean</td>
<td>$\beta_2$</td>
<td>-0.009</td>
<td>0.010</td>
<td>-0.001</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.007</td>
<td>0.010</td>
<td>-0.001</td>
<td>0.010</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.019</td>
<td>0.010</td>
<td>0.011</td>
<td>0.009</td>
</tr>
<tr>
<td>Variance</td>
<td>$\mu_h$</td>
<td>-1.123</td>
<td>0.042</td>
<td>-1.403</td>
</tr>
<tr>
<td></td>
<td>$\phi_h$</td>
<td>0.912</td>
<td>0.013</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_h$</td>
<td>0.112</td>
<td>0.018</td>
<td>0.013</td>
</tr>
<tr>
<td>Skewness</td>
<td>$\delta$</td>
<td>- -</td>
<td>- -</td>
<td>0.315</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_\delta$</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>Df</td>
<td>$\nu$</td>
<td>- -</td>
<td>5.809</td>
<td>0.381</td>
</tr>
<tr>
<td>DIC</td>
<td>18.633</td>
<td>2.49</td>
<td>18.842</td>
<td>0.48</td>
</tr>
</tbody>
</table>

### Table 8: USD/GBP exchange rate returns: estimation results

<table>
<thead>
<tr>
<th>Coef.</th>
<th>SV-sn</th>
<th>SV-t</th>
<th>SV-nct</th>
<th>SVSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.010</td>
<td>0.003</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.025</td>
<td>0.010</td>
<td>0.032</td>
<td>0.010</td>
</tr>
<tr>
<td>Mean</td>
<td>$\beta_2$</td>
<td>-0.006</td>
<td>0.010</td>
<td>-0.001</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.014</td>
<td>0.010</td>
<td>-0.016</td>
<td>0.010</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.006</td>
<td>0.010</td>
<td>0.006</td>
<td>0.010</td>
</tr>
<tr>
<td>Variance</td>
<td>$\mu_h$</td>
<td>-1.312</td>
<td>0.119</td>
<td>-1.571</td>
</tr>
<tr>
<td></td>
<td>$\phi_h$</td>
<td>0.987</td>
<td>0.002</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_h$</td>
<td>0.022</td>
<td>0.003</td>
<td>0.010</td>
</tr>
<tr>
<td>Skewness</td>
<td>$\delta$</td>
<td>- -</td>
<td>- -</td>
<td>-0.157</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_\delta$</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>Df</td>
<td>$\nu$</td>
<td>- -</td>
<td>7.728</td>
<td>0.629</td>
</tr>
<tr>
<td>DIC</td>
<td>16,673</td>
<td>0.92</td>
<td>16,373</td>
<td>0.53</td>
</tr>
</tbody>
</table>
Table 9: USD/CHF exchange rate returns: estimation results

<table>
<thead>
<tr>
<th></th>
<th>SV-sn</th>
<th></th>
<th>SV-t</th>
<th></th>
<th>SV-net</th>
<th></th>
<th>SVSS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.002</td>
<td>0.006</td>
<td></td>
<td>0.005</td>
<td>0.006</td>
<td>0.012</td>
<td>0.006</td>
<td>0.012</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.008</td>
<td>0.010</td>
<td></td>
<td>0.004</td>
<td>0.010</td>
<td>0.003</td>
<td>0.010</td>
<td>0.001</td>
</tr>
<tr>
<td>Mean</td>
<td>0.004</td>
<td>0.010</td>
<td></td>
<td>0.002</td>
<td>0.010</td>
<td>0.000</td>
<td>0.010</td>
<td>-0.002</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.005</td>
<td>0.010</td>
<td></td>
<td>0.012</td>
<td>0.010</td>
<td>0.011</td>
<td>0.010</td>
<td>0.011</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.007</td>
<td>0.010</td>
<td></td>
<td>-0.004</td>
<td>0.010</td>
<td>-0.004</td>
<td>0.010</td>
<td>-0.004</td>
</tr>
<tr>
<td>Variance</td>
<td>$\mu_h$</td>
<td>-0.869</td>
<td>0.067</td>
<td>-1.101</td>
<td>0.092</td>
<td>-1.102</td>
<td>0.092</td>
<td>-1.104</td>
</tr>
<tr>
<td></td>
<td>$\phi_h$</td>
<td>0.979</td>
<td>0.004</td>
<td>0.990</td>
<td>0.002</td>
<td>0.990</td>
<td>0.002</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td>$\sigma_h^2$</td>
<td>0.019</td>
<td>0.003</td>
<td>0.008</td>
<td>0.001</td>
<td>0.008</td>
<td>0.001</td>
<td>0.008</td>
</tr>
<tr>
<td>Skewness</td>
<td>$\delta$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.306</td>
<td>0.085</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\delta^2$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.003</td>
</tr>
<tr>
<td>Df</td>
<td>$\nu$</td>
<td>8.451</td>
<td>0.681</td>
<td>8.531</td>
<td>0.704</td>
<td>9.926</td>
<td>0.955</td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td></td>
<td>20,813</td>
<td>1.85</td>
<td>20,759</td>
<td>0.45</td>
<td>20,757</td>
<td>0.31</td>
<td>20,760</td>
</tr>
</tbody>
</table>

Second, (log-)volatility $h_t$ is strongly persistent across all models considered supporting the idea of volatility clustering as first described in Mandelbrot (1963). Third, deviations from the normality assumption are relevant model features that can improve model fit. Exchange rate returns are characterized by fat tails. Posterior means of the degrees of freedom $\nu$ lie in the range of 5 to 10. In addition, the DIC mostly drops sharply once allowing for fat tails, i.e. going from the SV-sn to the SV-t model. The Japanese Yen returns constitute, somewhat surprisingly, an exception since the DIC clearly prefers the SV-sn model even though the posterior mean of $\nu$ is small and its distribution narrow.

Furthermore, exchange rate returns are not symmetric. In the SV-net model, the 95% highest density interval (HDI) of $\delta$ does not include zero for the USD/AUD (left-tailed), the USD/JPY (right-tailed), and the USD/CHF returns (right-tailed). In case of the USD/GBP returns (left-tailed), results are unambiguous. The DIC supports these findings since the SV-net model is characterized by a smaller criterion value compared to the SV-t model in all cases except for the British Pound returns. However, the reductions are smaller than before. It is worth noting that the signs of the estimated noncentrality parameters are in line with the unconditional skewness measure reported in Table 5. Finally, when looking at the SVSS model, the posterior means of the innovation variances of $\delta_t$, $\sigma_t^2$, are of relevant and similar magnitude (even identical when rounded to three digits). In addition, posterior dispersion is fairly small. This can be taken as evidence in favour of time-varying skewness. The DIC ranks the SVSS model first in case of the Australian Dollar and second in case of the Japanese Yen. Differences in the criteria are minor when applying the model to Swiss Franc returns whereas skewness appears weakest in the British Pound returns. In summary, when looking at both the posterior distributions of the parameters and the information criterion, the SVSS model competes well with the remaining (nested) models.
4.3 Time-varying volatility and skewness

The previous section has shown that time-varying asymmetry bears relevance when modelling exchange rate returns. We now turn towards discussing the estimated volatility and, in particular, the skewness series. Figure 4 shows the volatility series obtained from the SVSS model. Alongside, a simple volatility measure, i.e. a centred rolling window variance (window size 300 days), is plotted. All four plots indicate strong volatility clustering. The periods of highest volatility can be found in the USD/AUD and USD/GBP exchange rate returns around the time of the Great Recession. Unexpectedly, the U.S. Dollar appreciated sharply against both currencies during this period causing large return shocks (McCauley and McGuire, 2009).

Figure 4: Estimated variance (volatility) series

![Estimated variance (volatility) series](image)

While time-varying volatility is well-known as a stylized fact of financial returns, Figure 5 presents the key finding of this paper, i.e. time variation in the estimated skewness of exchange rate returns. Again, a rolling window unconditional skewness measure is plotted next to the skewness estimate obtained from the SVSS model. In general, both series move together quite closely. However, the model implied skewness measure does not seem to be strongly affected by return outliers, a problem inherent to standard higher moment estimators. Even though our dataset contains more than 10,000 observations on daily returns, the 95% posterior density intervals are still fairly wide highlighting the challenge of estimating asymmetry precisely.

Overall, the dynamic evolution of skewness across the four exchange rates appears similar and points to the existence of a ‘skewness cycle’, i.e. crash risk alternates between slowly building up and slowly decreasing. To get further insights into cross-return co-movement of volatility and
crash risk, Table 10 presents the correlation matrices for both measures. Since all four bilateral exchange rates are measured relatively to the U.S. Dollar, this 'common factor' naturally induces positive correlation in both volatility and skewness across return series. Nevertheless, some returns move together more closely in terms of variance and skewness than others. The USD/AUD and USD/GBP returns seem to be subject to similar volatility and crash risk shocks. This is possibly due to the historically tight relation between Australia and the United Kingdom.

Figure 5: Estimated skewness series

![Estimated skewness series](image)

The Japanese Yen and the Swiss Franc are both considered typical funding currencies which could explain why the crash risks of both return series show significant co-movement. Finally, strong correlation of both measures in case of USD/GBP and USD/CHF returns might be rooted in the fact that both are important financial centres and similarly exposed to changes in the global economic environment.

Table 10: Correlations of volatility and skewness across return series

<table>
<thead>
<tr>
<th>FX</th>
<th>Volatility</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$/AUD</td>
<td>$/JPY</td>
</tr>
<tr>
<td>$/AUD</td>
<td>1</td>
<td>0.42</td>
</tr>
<tr>
<td>$/JPY</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>$/GBP</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$/CHF</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
4.4 Skewness, carry trades and asset pricing: Some evidence

The focus of this paper has been introducing stochastic skewness into a standard SV model and measuring the degree of time-varying asymmetry in exchange rate returns. Additionally, this section presents suggestive evidence of the link between carry trading and time-varying skewness and well as of the implications of skewness for asset pricing. Figure 6 plots the estimated skewness series (monthly averages) along with the monthly nominal interest rate differential (foreign 3-month government bond yield minus 3-month U.S. Treasury bill yield). Overall, both indicators tend to evolve in an opposite manner. If the interest differential is taken as a proxy for carry trade activity in currency markets, the results are thus in line with the findings of Brunnermeier et al. (2009) and Farhi et al. (2015), i.e. a negative correlation between carry trade activity and crash risk.

Figure 6: Estimated skewness and interest rate differential

The relationship is most pronounced in case of the Australian Dollar, a currency that is known for regularly being in the focus of carry trade speculators. Specifically, crash risk has built up since the end of the 90s and reached its peak, i.e. lowest skewness, around the onset of the Great Recession. At the same time, the interest rate differential has widened while market volatility has been low thus creating an attractive environment for the carry trade (Kohler, 2010; McCauley and McGuire, 2009). As previously mentioned, this paper considers bilateral exchange rates where the domestic currency is always the U.S. Dollar. One would expect the negative relationship between the interest rate differential and skewness to be even more pronounced in bilateral exchange rates including a typical funding currency (e.g. the Japanese Yen) and a typical investment currency (e.g. the Australian Dollar).
Figure 7 visualizes that realized excess returns to carry trading potentially reflect a compensation for the crash risk exposure of speculators. It opposes historical excess returns with crash risk measured as the estimated skewness of the underlying return distribution. Monthly excess returns are composed of the nominal exchange rate return and the interest rate differential,

\[ r_t = \frac{S_t - S_{t-1}}{S_{t-1}} \times 100 + i_{t}^f - i_t, \tag{15} \]

where \( S_t \) is the average spot exchange rate during month \( t \), and \( i_{t}^f \) and \( i_t \) represent the foreign and domestic interest rates (in %), respectively. On average, these excess returns are expected to be zero under the strictest version of the Uncovered Interest Parity condition (UIP).

**Figure 7**: Estimated skewness and excess risk premium

![Graphs showing correlation between skewness and excess returns](image)

(a) USD/AUD returns  
(b) USD/JPY returns  
(c) USD/GBP returns  
(d) USD/CHF returns

In all cases a negative link between excess returns and skewness can be identified. This is in line with models that extend the traditional CAPM to account for skewness (Kraus and Littenberger, 1976; Harvey and Siddique, 2000) and suggests that investors demand a compensation for return distributions that occasionally generate large losses.
5 Conclusion

This paper has presented an econometric approach to estimate time-varying asymmetry. A standard stochastic volatility model is extended to allow for stochastic skewness. Gaussian shocks are replaced by shocks coming from the noncentral t-distribution where the parameter that governs skewness varies stochastically over time. The model can be estimated by straightforward extensions of traditional Bayesian Markov Chain Monte Carlo methods for stochastic volatility models. Monte Carlo simulations suggest that the resulting stochastic volatility - stochastic skewness (SVSS) model performs well for sample sizes typically encountered when analysing daily financial returns.

The model is subsequently used to estimate time variation in the skewness of exchange rate returns of four major currencies relative to the U.S. Dollar over the period 01/01/1977 - 10/31/2017. The following results are obtained: The model fits the data well compared to simpler stochastic volatility models which assume symmetric return shocks or, alternatively, a constant degree of asymmetry. Thus, evidence is found in favour of time-varying skewness in exchange rate returns. Moreover, estimated skewness as a measure of crash risk is negatively correlated with the corresponding interest rate differential. This points towards carry trades as at least an amplifier of crash risk. Finally, crash risk seems to be priced in the sense that realized returns to carry trading are systematically higher when estimated skewness is lower. In summary, the results confirm previous findings in the literature.

Since the introduction of higher moment dynamics, and in particular stochastic skewness, in the stochastic volatility literature is still in its infancy, several avenues for future research arise. This paper has used the noncentral t-distribution to model asymmetric shocks, largely because time-varying skewness can be implemented particularly easy. However, a large number of distributions fulfil the general requirements to model time-varying asymmetry and a thorough comparison of their performance is yet to be done. Moreover, the model can be extended in various ways that have already proven useful in the stochastic volatility literature such as including leverage effects or allowing for stochastic skewness in mean dynamics. Especially the latter approach could prove useful in establishing a closer link with asset pricing theory as it allows for a direct impact of skewness on returns. Finally, possible applications where one would expect to find evidence of time-varying skewness go far beyond exchange rate returns.
References


Appendix A  Moments of the noncentral t-distribution

Following Hogben et al. (1961), the central moments of a noncentral t-distributed random variable $X \sim NCT(\nu, \delta)$ can be written as polynomials of $\delta$ whose coefficients are functions of $\nu$. Specifically, the expected value, variance, and third central moment are given by:

\[
\begin{align*}
E[X] &= c_{11}(\nu)\delta, & \text{if } \nu > 1, \\
E[(X - E[X])^2] &= c_{22}(\nu)\delta^2 + c_{20}(\nu), & \text{if } \nu > 2, \\
E[(X - E[X])^3] &= c_{33}(\nu)\delta^3 + c_{31}(\nu)\delta, & \text{if } \nu > 3.
\end{align*}
\]

The functional forms of the coefficients are:

\[
\begin{align*}
c_{11}(\nu) &= \sqrt{\frac{\Gamma\left(\frac{1}{2}(\nu - 1)\right)}{2\Gamma(\frac{1}{2})}}, \\
c_{22}(\nu) &= \frac{\nu}{\nu - 2} - c_{11}(\nu)^2, \\
c_{20}(\nu) &= \frac{\nu}{\nu - 2}, \\
c_{33}(\nu) &= c_{11}(\nu)\left[\frac{\nu(7 - 2\nu)}{(\nu - 2)(\nu - 3)} + 2c_{11}(\nu)^2\right], \\
c_{31}(\nu) &= \frac{3\nu}{(\nu - 2)(\nu - 3)}c_{11}(\nu).
\end{align*}
\]

Appendix B  Details on the MCMC algorithm

In this appendix, details are given on the blocking scheme of the MCMC algorithm and the conditional posterior distributions of the stochastic volatility - stochastic skewness (SVSS) model introduced in Section 2.

Block 1: Sample the mixture indicators $s$ from $p(s|y, X, h, \delta, \nu, \beta)$

In order to sample the mixture indicators $s$ of the extended mixture representation introduced in Section 2.3, we build on the approach of Kim et al. (1998) but account for the fact that the appropriate mixture components in the SVSS model depend on $\nu$ (which changes over MCMC iterations) and $\delta_t$ (which changes over MCMC iterations and time). $s_t$ is a discrete random variable that follows a ten-point distribution. In particular, each $s_t$ has probability

\[
p(s_t = j|y_t, X_t, h_t, \delta_t, \nu, \beta) = \frac{1}{k_t}q_j(\nu, \delta_t)p_{\mathcal{N}}(\tilde{y}_t; h_t + m_j(\nu, \delta_t), \nu_j^2(\nu, \delta_t)), \tag{A-1}
\]

where $\tilde{y}_t = \log((y_t - X_t)^2 + c)$, $c = 0.001$ is an offset constant and $k_t = \sum_{j=1}^{10}q_j(\nu, \delta_t)p_{\mathcal{N}}(\tilde{y}_t; h_t + m_j(\nu, \delta_t), \nu_j^2(\nu, \delta_t))$ is a normalizing constant. Practical implementation of the indicator sampling is done by using the inverse-transform method as in Chan and Hsiao (2014).11

11See Algorithm 3.2. in Kroese et al. (2013) for a textbook treatment of the inverse-transform method.
Block 2: Sample the (log-)volatility $h$ from $p(h|y, X, s, \delta, \nu, \beta, \mu_h, \phi_h, \sigma^2_h)$

For the purpose of sampling the latent (log-)volatility series $h$, we first specify a general state space model of the following form as given in Durbin and Koopman (2012)

$$w_t = Z_t \kappa_t + e_t, \quad e_t \sim \mathcal{N}(0, H_t),$$

$$\kappa_{t+1} = d_t + T_t \kappa_t + R_t \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q_t),$$

where $w_t$ is an observed data point and $\kappa_t$ the unobserved state. The matrices $Z_t, T_t, H_t, Q_t, R_t$, and $d_t$ are assumed to be known (conditioned upon). The error terms $e_t$ and $\eta_t$ are assumed to be serially uncorrelated and independent of each other at all points in time. Bearing in mind this general form, the specific state space model to be estimated in this block is

$$\begin{align*}
\begin{bmatrix}
\tilde{y}_t - m_{w_t}(\nu, \delta_t) \\
\kappa_{t+1}
\end{bmatrix} 
&= 
\begin{bmatrix}
1 \\
\kappa_t
\end{bmatrix}
\begin{bmatrix}
Z_t & e_t \\
\kappa_{t+1}
\end{bmatrix} 
+ 
\begin{bmatrix}
h_t \\
\mu_h (1 - \phi_h)
\end{bmatrix} + 
\begin{bmatrix}
\phi_h h_t \\
\tau_t
\end{bmatrix}
\begin{bmatrix}
\kappa_t & e_t \\
\kappa_{t+1}
\end{bmatrix} + 
\begin{bmatrix}
1 \\
\kappa_t
\end{bmatrix}
\begin{bmatrix}
\eta_t \\
\eta_t
\end{bmatrix},
\end{align*}$$

where $\tilde{y}_t = \log((y_t - X_t \beta)^2 + \gamma)$, $H_t = \nu^2 (\nu, \delta_t)$ and $Q_t = \sigma^2_h$. As Equations (A-4) and (A-5) constitute a linear Gaussian state space model, the unknown state variable $h_t$ can be filtered using the standard Kalman filter. Sampling $h = (h_1, ..., h_T)$ can then be achieved using the algorithm outlined in Carter and Kohn (1994). More recently, Chan and Jeliazkov (2009) and McCausland et al. (2011) have shown how the unobserved states of a linear Gaussian state space model can be filtered and sampled more efficiently by relying on sparse matrix algorithms. This paper follows Chan and Hsiao (2014) who show how to sample the unobserved (log-)volatilities $h_t$ efficiently using these algorithms. The reader is referred to pp. 5-8 in Chan and Hsiao (2014) for a detailed outline of the so-called precision sampler.

Block 3: Sample the latent state $\lambda$ from $p(\lambda|y, X, h, \delta, \nu, \beta)$

In sampling the latent state variable $\lambda$, this paper follows Tsionas (2002). In particular the conditional distribution of each $\lambda_t$ is

$$p(\lambda_t|y_t, X_t, h_t, \delta_t, \nu, \beta) \propto \lambda_t^{-(\nu+3)/2} \exp \left[ -\frac{u_t^2 / e^{h_t} + \nu}{2\lambda_t} + \delta_t (u_t / e^{h_t} / 2) \lambda_t^{-1/2} \right],$$

where $u_t = y_t - X_t \beta + e^{h_t / 2} c_{11}(\nu) \delta_t$ and the second summand is due to the fact that Tsionas (2002) does not consider the de-meaned version of the noncentral t-distribution. If $\delta_t = \delta = 0$, i.e., the shocks are Student $t$-distributed, $\lambda_t$ is conditionally inverse-gamma distributed and can be straightforwardly sampled as in e.g. Chan and Hsiao (2014). However, in the noncentral case acceptance sampling is required as the conditional distribution is non-standard. To this end, one can make use of the fact that the conditional distribution of $w_t = \lambda_t^{-1/2}$ is log-concave. In
particular, the conditional distribution of each \( w_t \) is
\[
p(w_t|y_t, X_t, h_t, \delta_t, \nu, \beta) \propto w_t^{\nu} \exp \left( -\frac{\delta_t}{2} w_t^2 + \frac{\nu}{\delta_t} w_t \right) .
\] (A-7)

This conditional distribution belongs to a family of distributions with kernel function
\[
f(x) \propto x^{N-1} \exp(-(A/2)x^2 + Bx),
\] (A-8)
where \( N = \nu + 1, A = \delta_t / e^h + \nu \) and \( B = \delta_t / e^h/2 \).

The proposal density for the acceptance sampling is \( g(x) \sim \text{Gamma}(N, \theta^*) \), where \( \theta^* = N/x^* \) and \( x^* \) is the positive root that solves
\[
A_t x^2 - B_t x - N = 0.
\] (A-9)

We then accept the candidate draw \( w_t^* \) with probability
\[
R = \exp(r^* - r),
\] (A-10)
where \( r^* = \log(f(x)/g(x)) \) evaluated at \( w_t^* \) and \( r = \log(f(x)/g(x)) \) evaluated at \( x^* \). Specifically,
\[
r^* = -(A_t/2)w_t^{*2} + (B_t + \theta^*)w_t^* - N \log(\theta^*),
\] (A-11)
\[
r = -(A_t/2)x^{2*} + (B_t + \theta^*)x^* - N \log(\theta^*).
\] (A-12)

After having accepted a candidate draw \( w_t^* \), the original state variable is simply recovered as \( \lambda_t = w_t^{-2} \).

**Block 4: Sample the degrees of freedom \( \nu \) from \( p(\nu|\lambda) \)**

Sampling the degrees of freedom \( \nu \) is identical to the case with (symmetric) Student t-distributed shocks. The description of the sampling approach closely follows Chan and Hsiao (2014). The log-density \( \log p(\nu|\lambda) \) can be derived using the fact that \( \lambda_t \sim IG(\nu/2, \nu/2) \) and the prior distribution \( \nu \sim U(0, \bar{\nu}) \) as
\[
\log p(\nu|\lambda) = \frac{T\nu}{2} \log(\nu/2) - T \log \Gamma(\nu/2) - (\nu/2 + 1) \sum_{t=1}^{T} \log \lambda_t - \frac{\nu}{2} \sum_{t=1}^{T} \lambda_t^{-1} + k,
\] (A-13)
for \( 0 < \nu < \bar{\nu} \) and \( k \) is a normalization constant. The first and second derivative of the log-density with respect to \( \nu \) are then given by
\[
\frac{d \log p(\nu|\lambda)}{d\nu} = \frac{T\nu}{2} \log(\nu/2) + \frac{T}{2} - \frac{T}{2} \Psi'(\nu/2) - \frac{1}{2} \sum_{t=1}^{T} \log \lambda_t - \frac{1}{2} \sum_{t=1}^{T} \lambda_t^{-1},
\] (A-14)
\[
\frac{d^2 \log p(\nu|\lambda)}{d\nu^2} = \frac{T}{2\nu} - \frac{T}{4} \Psi''(\nu/2),
\] (A-15)
where $\Psi(x) = \frac{d}{dx} \log \Gamma(x)$ and $\Psi'(x) = \frac{d}{dx} \Psi(x)$ are the digamma and trigamma function, respectively. Since the first and second derivatives can be evaluated easily, $\log p(\nu | \lambda)$ can be maximized by well-known algorithms (e.g. the Newton-Raphson method). In addition, the mode and the negative Hessian evaluated at the mode, denoted $\hat{\nu}$ and $K_{\nu}$, are obtained. Finally, an independence-chain Metropolis-Hastings step can be implemented with proposal distribution $N(\hat{\nu}, K_{\nu}^{-1})$.

**Block 5: Sample the latent noncentrality parameter $\delta$ from**

$p(\delta | y, X, h, \lambda, \nu, \beta, \mu_\delta, \phi_\delta, \sigma_\delta^2)$

In order to sample the time-varying noncentrality parameter $\delta_t$, we explore the following state space model

$$
\begin{align*}
\bar{y}_t &= \left[ \frac{1}{2} - c_{11}(\nu) \right] \frac{\delta_t}{c_t} + \frac{c_t}{c_t} \\
\delta_{t+1} &= \left[ \frac{\mu_\delta(1 - \phi_\delta)}{d_t} \right] \frac{\delta_t}{c_t} + \frac{\delta_t}{c_t} + \left[ 1 \right] \frac{\eta^\delta}{\eta^t}.
\end{align*}
$$

(A-16)  

(A-17)

where $\bar{y}_t = (y_t - X_t \beta)e^{-h_t/2}$ and with $H_t = \lambda_t$ and $Q_t = \sigma_\delta^2$. Note that the observation Equation (A-16) is obtained by rewriting the SVSS model in Equation (6). Instead of applying the forward-filtering and backward-sampling approach of Carter and Kohn (1994), again the routine developed in Chan and Jeliazkov (2009) is used to obtain a sample of $\delta_t = (\delta_1, \ldots, \delta_T)$. If $\delta_t$ is specified as a (driftless) random walk as in Section 4, we set $\mu_\delta = 0$ and $\phi_\delta = 1$ in the state Equation (A-17).

**Block 6: Sample the constant trend volatility $\mu_h$ from**

$p(\mu_h | h, \phi_h, \sigma_h^2)$

and

**the constant trend asymmetry $\mu_\delta$ from**

$p(\mu_\delta | \delta, \phi_\delta, \sigma_\delta^2)$

The conditional posterior distributions of the constant trend volatility and trend asymmetry are standard and samples can be readily obtained. Following Kim et al. (1998) and the notation of Chan and Hsiao (2014), the conditional distribution of $\mu_t$, where $\tau = (h, \delta)$, is

$$
\mu_t | \tau, \phi_t, \sigma_\tau^2 \sim N(\hat{\mu}_t, D_{\mu_t}),
$$

(A-18)

with

$$
D_{\mu_t} = (V_{\mu_t}^{-1} + X_{\mu_t}' \Sigma_{\tau}^{-1} X_{\mu_t})^{-1},
$$

(A-19)

$$
\hat{\mu}_t = D_{\mu_t}(V_{\mu_t}^{-1} \mu_0 + X_{\mu_t}' \Sigma_{\tau}^{-1} z_{\mu_t}),
$$

(A-20)

where $X_{\mu_t} = (1, 1 - \phi_\tau, \ldots, 1 - \phi_T)'$, $z_{\mu_t} = (\tau_1, \tau_2 - \phi_\tau \tau_1, \ldots, \tau_T - \phi_T \tau_{T-1})'$ and $\Sigma_{\tau} = \text{diag}(\sigma_\tau^2 / (1 - \phi_\tau^2), \sigma_\tau^2, \ldots, \sigma_\tau^2)$. If $\delta_t$ is specified as a (driftless) random walk as in Section 4, sampling $\mu_\delta$ is simply omitted.
Block 7: Sample the volatility AR(1) coefficient $\phi_h$ from $p(\phi_h| h, \mu_h, \sigma^2_h)$ and the asymmetry AR(1) coefficient $\phi_\delta$ from $p(\phi_\delta| \delta, \mu_\delta, \sigma^2_\delta)$

Following Kim et al. (1998) and using the notation of Chan and Hsiao (2014), the conditional posterior distribution of the persistence parameter $\phi_\tau$, where $\tau = (h, \delta)$, is

$$
p(\phi_\tau| \tau, \mu_\tau, \sigma^2_\tau) \propto p(\phi_\tau) g(\phi_\tau) \exp \left( -\frac{1}{2\sigma^2_\tau} \sum_{t=2}^{T} (\tau_t - \mu_\tau - \phi_\tau(\tau_{t-1} - \mu_\tau))^2 \right),
$$

(A-21)

with

$$
g(\phi_\tau) = (1 - \phi^2_\tau)^{1/2} \exp \left( -\frac{1}{2\sigma^2_\tau} (1 - \phi^2_\tau)(\tau_1 - \mu_\tau)^2 \right),
$$

(A-22)

and $p(\phi_\tau)$ is the truncated normal prior defined in Equation (7). Due to the stationarity condition $|\phi_\tau| < 1$, this distribution is non-standard and sampling is achieved using the Metropolis-Hastings algorithm. In particular, the proposal density is $\mathcal{N}(\hat{\phi}_\tau, D_{\phi_\tau})\mathbb{1}(|\phi_\tau| < 1)$ with

$$
D_{\phi_\tau} = (V_{\phi_\tau}^{-1} + X'_{\phi_\tau} X_{\phi_\tau}/\sigma^2_\tau)^{-1},
$$

(A-23)

$$
\hat{\phi}_\tau = D_{\phi_\tau} (V_{\phi_\tau}^{-1} \phi_0 + X'_{\phi_\tau} z_{\phi_\tau}/\sigma^2_\tau),
$$

(A-24)

where $X_{\phi_\tau} = (\tau_1 - \mu_\tau, ..., \tau_{T-1} - \mu_\tau)'$ and $z_{\phi_\tau} = (\tau_2 - \mu_\tau, ..., \tau_T - \mu_\tau)'$ (Chan and Hsiao, 2014). Conditional on the current state $\phi_\tau$, a proposal $\phi^*_\tau$ is accepted with probability $\min(1, g(\phi^*_\tau)/g(\phi_\tau))$.

In case of rejection, the Markov chain remains at the current state $\phi_\tau$. If $\delta_t$ is specified as a (driftless) random walk as in Section 4, sampling $\phi_\delta$ is simply omitted.

Block 8: Sample the shock variances $\sigma^2_h$ from $p(\sigma^2_h| h, \mu_h, \phi_h)$ and $\sigma^2_\delta$ from $p(\sigma^2_\delta| \delta, \mu_\delta, \phi_\delta)$

The shock variances of the (log-)volatility $h_t$ and the noncentrality parameter $\delta_t$ have inverse-gamma conditional posterior distributions (Kim et al., 1998). Specifically, the conditional posterior distribution of $\sigma^2_\tau$, where $\tau = (h, \delta)$, is

$$
\sigma^2_\tau| \tau, \mu_\tau, \phi_\tau \sim \mathcal{IG}(c_{\tau 0} + T/2, C_\tau),
$$

(A-25)

where notation follows Chan and Hsiao (2014) and

$$
C_\tau = C_{\tau 0} + \left[ (1 - \phi^2_\tau)(\tau_1 - \mu_\tau)^2 + \sum_{t=2}^{T} (\tau_t - \mu_\tau - \phi_\tau(\tau_{t-1} - \mu_\tau))^2 \right]/2,
$$

(A-26)

If $\delta_t$ is specified as a (driftless) random walk as in Section 4, we set $\mu_\delta = 0$ and $\phi_\delta = 1$.

Block 9: Sample the regression coefficients $\beta$ from $p(\beta| y, X, h, \delta, \lambda, \nu)$

When the stochastic volatility - stochastic skewness model is augmented by a conditional mean specification as in Section 4.2, the corresponding k-dimensional vector of regression coefficients $\beta$
can be sampled as in Tsionas (2002). The conditional posterior distribution is

\[ \beta | y, X, h, \delta, \lambda, \nu \sim N \left( [X' \Lambda^{-1} X]^{-1} X' \Lambda^{-1} (\tilde{y} - \delta \odot e^{h/2} \odot \lambda^{1/2}), e^{h} [X' \Lambda^{-1} X]^{-1} \right), \]

where \( X \) is a \( T \times k \) matrix of regressors, \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_T) \), \( \tilde{y} = y + e^{h/2} \epsilon_{11}(\nu) \delta \) and \( \odot \) is the element-wise (Hadamard) product of two vectors. The second summand of the transformed dependent variable \( \tilde{y} \) is again due to the fact that Tsionas (2002) does not consider the de-meaned version of the noncentral t-distribution.

**Appendix C  Monte Carlo simulation: Conditional DIC**

This appendix provides Monte Carlo evidence to assess the performance of the conditional Deviance Information Criterion (DIC), as developed by Spiegelhalter et al. (2002), in our setting.

To this end, we simulate 500 datasets from three different stochastic volatility models: The standard normal SV model (see e.g. Kim et al., 1998), the SV model with Student t-distributed innovations (see e.g. Chib et al., 2002) and the restricted (random walk) SVSS model as specified in Section 2.2 and applied in Section 4. The parameter values of the data generating processes are set as follows: \( \mu_h = 0, \phi_h = 0.99, \) and \( \sigma_h^2 = 0.01 \) (SV-sn). For the SV-t model, in addition, the degrees of freedom parameter is \( \nu = 10 \). Finally, for the SVSS model, we set the innovation variance of the asymmetry process \( \delta_t \) to \( \sigma_{\delta}^2 = 0.01 \). For each of the three DGPs, we estimate the same three specifications and compute the conditional DIC for each model and sample. The prior distributions are as in Section 3. Table A-11 presents the results of this simulation experiment. In particular, the percentage shares in which the DIC prefers a certain SV model given the DGP, are reported.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Model</th>
<th>DGP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SV-sn</td>
<td>SV-t</td>
</tr>
<tr>
<td>T = 1,000</td>
<td>99.20</td>
<td>72.80</td>
</tr>
<tr>
<td></td>
<td>SV-t</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>SVSS</td>
<td>0.40</td>
</tr>
<tr>
<td>T = 5,000</td>
<td>100.00</td>
<td>80.00</td>
</tr>
<tr>
<td></td>
<td>SV-t</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>SVSS</td>
<td>0.00</td>
</tr>
<tr>
<td>T = 10,000</td>
<td>100.00</td>
<td>83.20</td>
</tr>
<tr>
<td></td>
<td>SV-t</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>SVSS</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: Reported are shares (in %) of Monte Carlo samples in which the DIC prefers a certain SV model given the DGP.

First, when the DGP is the SV-sn model, the DIC is capable to select the correct model in
almost all the cases independent of the sample size. Unfortunately, problems exist when samples are generated from the SV-t model. In this case, the DIC only picks the correct model in around 20% of the cases while mostly preferring the simpler SV-sn model. Surprisingly, this is at odds with results reported in Chan and Grant (2016), who find that the DIC almost always favours overfitted models.\footnote{We found the DIC to be sensitive with respect to the representation of the conditional likelihood. While we use the t-density directly, Chan and Grant (2016) employ the normal-gamma scale mixture representation to compute the conditional likelihood of the SV-t model (we only use the mixture representation for MCMC estimation). The latter involves an additional latent variable. Given that the DIC is known for being sensitive to the specification of latent variables (Li et al., 2015), this might explain the different outcomes (see also Section 8.2 in Spiegelhalter et al., 2002, for a similar example, where DIC values differ across both distributional representations).} Important for the purpose of this paper, the DIC can properly identify SVSS dynamics in the data. The DIC selects the SVSS model in more than 95% of the samples if and only if it corresponds to the DGP, unless in a small sample where it picks the simple SV model in around 30% of the samples.