FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

WORKING PAPER

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

Martin Iseringhausen

March 2018 (revised June 2020)
2018/944

GHENT UNIVERSITY

Faculty of Economics and Business Administration
Campus Tweekerken, Tweekerkenstraat 2, 9000 Ghent - BELGIUM

D/2018/7012/02

The Time-Varying Asymmetry of Exchange Rate Returns: A Stochastic Volatility - Stochastic Skewness Model

Martin Iseringhausen*

Ghent University

June 2020

Abstract

While the time-varying volatility of financial returns has been extensively modelled, most existing stochastic volatility models either assume a constant degree of return shock asymmetry or impose symmetric model innovations. However, 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 allow for time-varying skewness of the return innovations. We estimate the model by extensions of traditional 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. In addition, stochastic skewness can help to improve forecasts of risk measures. Finally, the results support a potential link between carry trading and crash risk.

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, Koen Inghelbrecht, Dimitris Korobilis, Matteo Ciccarelli, Kris Boudt, the members of the Macroeconomics, Policy, and Econometrics Research Group at Ghent University and the participants of the International Association for Applied Econometrics 6$^{\text{th}}$ Annual Conference 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, Department of Economics, Campus Tweekerken, 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. (2008) suggest that time-varying crash (or downside) 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. The uncovered interest rate parity (UIP) implies that this strategy should not be profitable as the interest rate differential is expected to be offset by a depreciation of the investment currency. 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. (2008) find that high interest rate differentials predict negative skewness across currencies and over time. Following the authors, the carry trade 'unwinds', i.e. currencies crash, when speculators face funding constraints.

From a theoretical perspective, previous work has shown that investors prefer positively skewed return distributions and that an asset's own skewness can be priced (Arditti, 1967; Scott and Horvath, 1980; Brunnermeier and Parker, 2005; Barberis and Huang, 2008). In principle, excess carry trade returns can thus be rationalized if they are viewed as investors' compensation for holding an asset that exhibits negative skewness, i.e. that carries significant crash risk. Following the same idea, Farhi et al. (2015) develop a structural exchange rate model that accounts for disaster risk where a disaster is defined as a hypothetical large increase in the stochastic discount factor. When fitting the model using monthly data they provide evidence that disaster risk accounts for more than a third of the average carry trade risk premium in G10 currencies. Burnside et al. (2011) compare an unhedged carry trade with its hedged counterpart that protects investors against large losses. Since the latter has a smaller payoff than the former they conclude that excess carry returns are explained by 'Peso problems'.

Unlike in the stochastic volatility (SV) literature, time-varying skewness or, more generally, higher moment dynamics of return innovations, have been extensively modelled in (generalized) autoregressive conditional heteroscedasticity (GARCH) models starting off from Hansen (1994). The author extends the ARCH model of Engle (1982) so that not only the conditional variance but also the shape parameters of the standardized innovation distribution evolve as a function of conditioning information. When modelling the U.S. term structure and the USD/CHF exchange rate, he finds that models with (skewed) Student-t innovations and conditional shape parameters have a significantly improved in-sample fit compared to restricted model versions. 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).[1] Overall, it is found

[1]The class of jump models, possibly with time-varying jump intensities, is a popular alternative (instead of skewed innovation distributions) to model skewness dynamics, especially in option pricing. Examples of such models

that allowing for conditional skewness is beneficial for modelling stock returns, exchange rate returns and for option pricing. However, as noted by Feunou and Tedongap (2012), these GARCH approaches have limited flexibility 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.[2] 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 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. When estimating the model using daily Euro and Yen exchange rates versus the U.S. Dollar over the period 01/01/2001–30/12/2008, the author finds the corresponding skewness regime probabilities to vary over time but to generally evolve slowly, suggesting that the underlying skewness process is fairly persistent.

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.[3] 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., 2008) 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 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 generally well for sample sizes typically encountered when analysing financial returns at daily frequency. When applying the SVSS model to daily nominal exchange rate returns of four major currencies relative to the U.S. Dollar over the period 01/01/1977–31/10/2017, evidence is found in favour of time-varying asymmetry. First, in-sample the model outperforms alternative (nested) SV models for currencies that are heavily 'carry-traded' whereas a jump specification dominates for currencies

including applications to currencies can be found in, e.g. Carr and Wu (2007) and Chernov et al. (2018).

[2] Univariate SV models with skewed error distributions but constant asymmetry have been discussed, for example, in Cappuccio et al. (2004), Nakajima and Omori (2012), Tsiotas (2012) and Abanto-Valle et al. (2015).

[3] Harvey and Siddique (1999) use the same distribution to introduce conditional skewness in a GARCH framework.

that are less involved in carry trading. Second, the SVSS model forecasts risk measures, on average, more precisely than the alternative specifications. Third, skewness is largely negative in typical 'investment currencies' such as the Australian Dollar and positive in 'funding currencies' such as the Japanese Yen.[4] Fourth, interest rate differentials can predict future skewness suggesting that carry trading acts at least as an amplifier of crash risk in exchange rates. Finally, the results support the notion that skewness is priced.

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.

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 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 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 for computational implementation is the noncentral t-distribution (see Johnson et al., 1995, for an overview). A random variable X is noncentral t-distributed with ν degrees of freedom and noncentrality parameter δ, i.e. $X \sim \mathcal{NCT}(\nu, \delta)$, if it has the following stochastic representation:

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

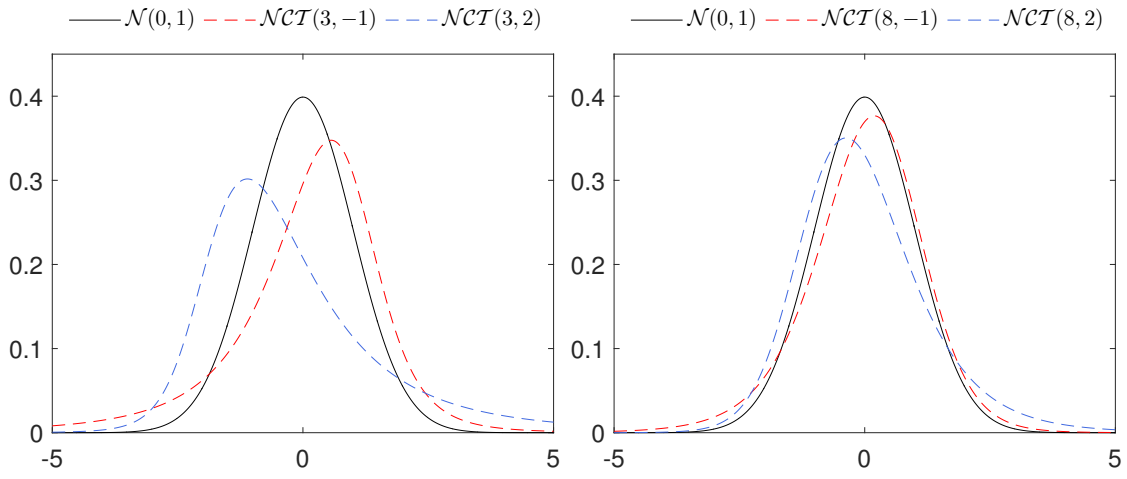
where $\mathcal{IG}(\alpha, \zeta)$ denotes the inverse-gamma distribution with shape parameter α and scale parameter ζ. For $\delta = 0$ the noncentral t-distribution collapses to its symmetric counterpart, the Student t-distribution. If, in addition, $\nu \to \infty$, it simplifies further to the standard normal distribution. All moments of the noncentral t-distribution are jointly determined by the parameters ν and δ. In particular, the central moments of the noncentral t-distribution can be expressed as polynomials of δ whose coefficients are functions of ν (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 ν and δ in comparison with the standard normal distribution. 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 this distribution are strongly linked, i.e. positive (negative) skewness mechanically goes along with a positive (negative) mean. This is an undesirable feature

[4] Throughout, we interpret a *lower skewness* value as *higher crash risk* and use the terms interchangeably.

for the purpose of this paper as, later on, variations in δ are supposed to capture changes in the asymmetry of the distribution rather than changes in the mean. To ensure that δ does not affect the mean, in what follows we consider the de-meaned version of the noncentral t-distribution.

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



(a) Degrees of freedom: $\nu = 3$

(b) Degrees of freedom: $\nu = 8$

2.2 Model specification

We start from the following univariate stochastic volatility model,

$$y_t = e^{h_t/2}\varepsilon_t, \qquad t = 1, ..., T, \tag{2}$$

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \eta_t^h, \qquad \eta_t^h \sim \mathcal{N}(0, \sigma_h^2), \quad |\phi_h| < 1, \tag{3}$$

where h_t is the latent (log-)volatility process assumed to evolve according to a stationary AR(1) process and ε_t is a zero mean shock term. Depending on the distributional assumption about ε_t, various SV models arise. If ε_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 δ 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 ν degrees of freedom and time-varying noncentrality parameter δ_t:

$$\varepsilon_t = u_t - \mathbb{E}[u_t], \quad \text{with} \quad u_t \sim \mathcal{NCT}(\nu, \delta_t), \quad \text{and} \quad \mathbb{E}[u_t] = c_{11}(\nu)\delta_t, \quad \text{if } \nu > 1. \tag{4}$$

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

$$\delta_t = \mu_\delta + \phi_\delta(\delta_{t-1} - \mu_\delta) + \eta_t^\delta, \qquad \eta_t^\delta \sim \mathcal{N}(0, \sigma_\delta^2), \qquad |\phi_\delta| < 1. \tag{5}$$

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

$$y_t = e^{h_t/2}\varepsilon_t = e^{h_t/2}\left(\sqrt{\lambda_t}(z_t + \delta_t) - c_{11}(\nu)\delta_t\right), \tag{6}$$

where again $\lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$ 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 λ_t which facilitates the implementation of a MCMC algorithm in Section 2.4.[5] 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 ϕ_h is typically close to one indicating strong volatility clustering. The general version of the SVSS model adopts the same stochastic process for the noncentrality state δ_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 ϕ_δ turns out to be difficult to pin down when the signal in the time-varying noncentrality parameter δ_t is not very strong. For this reason, we also consider a more parsimonious (restricted) version of the SVSS model where the noncentrality parameter δ_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 μ_h, ϕ_h, σ_h^2, μ_δ, ϕ_δ, σ_δ^2, and ν:

$$\begin{aligned}
&\mu_h \sim \mathcal{N}(\mu_{h0}, V_{\mu_h}), && \phi_h \sim \mathcal{TN}_{(-1,1)}(\phi_{h0}, V_{\phi_h}), && \sigma_h^2 \sim \mathcal{IG}(c_{h0}, C_{h0}), \\
&\mu_\delta \sim \mathcal{N}(\mu_{\delta 0}, V_{\mu_\delta}), && \phi_\delta \sim \mathcal{TN}_{(-1,1)}(\phi_{\delta 0}, V_{\phi_\delta}), && \sigma_\delta^2 \sim \mathcal{IG}(c_{\delta 0}, C_{\delta 0}), \\
&\nu \sim \mathcal{U}(0, \bar{\nu}), &&&&
\end{aligned} \tag{7}$$

where $\mathcal{TN}_{(a,b)}(\mu, \sigma^2)$ denotes the truncated normal distribution with support over the interval (a, b). 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

[5] This is because conditional on the latent variable λ_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).

model are given by

$$Var[y_t|h_t, \delta_t, \nu] = e^{h_t}\left[c_{22}(\nu)\delta_t^2 + c_{20}(\nu)\right], \qquad \text{if } \nu > 2, \tag{8}$$

$$Skew[y_t|\delta_t, \nu] = \frac{c_{33}(\nu)\delta_t^3 + c_{31}(\nu)\delta_t}{[c_{22}(\nu)\delta_t^2 + c_{20}(\nu)]^{3/2}}, \qquad \text{if } \nu > 3, \tag{9}$$

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.[6] Technically, δ_t induces time variation in all higher moments. In particular, given a certain value for ν, an increase in the absolute size of δ results in a larger kurtosis of the noncentral t-distribution thus reinforcing our crash risk interpretation.[7] 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 δ_t adjusts to reflect changes in the third moment. Finally, for future use, define $y = (y_1, ..., y_T)'$, $h = (h_1, ..., h_T)'$, $\delta = (\delta_1, ..., \delta_T)'$, and $\lambda = (\lambda_1, ..., \lambda_T)'$.

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 assume that $\varepsilon_t \sim \mathcal{N}(0, 1)$ and apply the following transformation to Equation (2),

$$log(y_t^2 + c) = h_t + \widetilde{\varepsilon}_t, \tag{10}$$

with $\widetilde{\varepsilon}_t = log(\varepsilon_t^2)$ 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-χ^2 distributed. Kim et al. (1998) approximate $\widetilde{\varepsilon}_t$ 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 in Equation (6) is

$$\widetilde{\varepsilon}_t = log\left[\left(\sqrt{\lambda_t}(z_t + \delta_t) - c_{11}(\nu)\delta_t\right)^2\right], \tag{11}$$

where $\widetilde{\varepsilon}_t$ is now log-noncentral-t-squared distributed. This implies that a single mixture of normal distributions is no longer sufficient but that the specific approximation to be used depends on the

[6] The error term in Equation (4) could, in principle, be standardized to have unit variance, which would simplify the expression of the conditional variance to e^{h_t}. However, this complicates the sampling of the asymmetry parameter δ within the MCMC algorithm since deriving a state space model that is linear in δ is no longer straightforward.
[7] See Appendix A for the expression of the fourth central moment.

values of ν and δ_t:

$$f(\widetilde{\varepsilon}_t|\nu,\delta_t) = \sum_{j=1}^{M} q_j(\nu,\delta_t) f_{\mathcal{N}}(\widetilde{\varepsilon}_t|m_j(\nu,\delta_t), v_j^2(\nu,\delta_t)), \tag{12}$$

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,

$$\widetilde{\varepsilon}_t|(s_t=j) \sim \mathcal{N}\left(m_j(\nu,\delta_t), v_j^2(\nu,\delta_t)\right), \qquad Pr(s_t=j) = q_j(\nu,\delta_t). \tag{13}$$

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 ν (3 to 50 with step size 0.1) and δ (-5 to 5 with step size 0.01).[8] We follow Omori et al. (2007) and use $M = 10$ mixture components. Experimenting with different numbers of mixture components suggests that $M = 10$ approximates the log-noncentral-t-squared distribution sufficiently well across the grid of values for ν and δ. However, if desired, the approximation could be made more precise by choosing a larger M. Importantly, generating the mixtures is a one time computation cost and hence does not affect sampling efficiency.

2.4 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 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 λ from $p(\lambda|y,h,\delta,\nu)$;
4. Sample v from $p(\nu|\lambda)$;
5. Sample δ from $p(\delta|y,h,\lambda,\nu,\mu_\delta,\phi_\delta,\sigma_\delta^2)$;
6. Sample μ_h from $p(\mu_h|h,\phi_h,\sigma_h^2)$ and μ_δ from $p(\mu_\delta|\delta,\phi_\delta,\sigma_\delta^2)$;
7. Sample ϕ_h from $p(\phi_h|h,\mu_h,\sigma_h^2)$ and ϕ_δ from $p(\phi_\delta|\delta,\mu_\delta,\sigma_\delta^2)$;

[8] This leads to a total of around 470,000 mixture approximations each of which has ten Gaussian components.

8. Sample σ_h^2 from $p(\sigma_h^2|h, \mu_h, \phi_h)$ and σ_δ^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 ν_i and $\delta_{t,i}$. The series of (log-)volatilities h (block 2) and noncentrality parameters δ (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 δ which speed up the algorithm significantly (Chan and Hsiao, 2014; Chan and Jeliazkov, 2009). Moreover, the conditional posterior distributions of λ (block 3), ν (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 $(B+1, ..., J)$ can be taken as a sample from the joint posterior distribution of interest $f(h, \delta, \lambda, \nu, \phi_h, \mu_h, \sigma_h^2, \phi_\delta, \mu_\delta, \sigma_\delta^2|y)$. A MATLAB program implementing this algorithm and replicating parts of the results can be downloaded from `https://sites.google.com/view/martiniseringhausen`.

3 Monte Carlo simulations

This section assesses the SVSS model and the proposed estimation technique using simulation experiments. The goal is to check the MCMC algorithm's ability to accurately estimate the underlying model dynamics while also highlighting potential obstacles when estimating time-varying skewness. 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 100,000 Gibbs iterations where 50,000 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 ν is based on the consideration that for $\nu > 50$ the noncentral t-distribution becomes indistinguishable from the normal distribution. Most of these relatively flat prior distributions are centered around the true values so that any potential imprecision must originate from a too small signal in the data or problems with the algorithm presented in Section 2.4. Appendix D contains additional simulation results when using prior distributions that are not centered around the true values.

Table 2 presents the simulation results of the general version of the SVSS model where δ_t is generated and estimated as a stationary AR(1) process. The data generating process (DGP) assumes a high degree of volatility and asymmetry persistence ($\phi_h = \phi_\delta = 0.99$), identical innovation

variances of volatility and asymmetry ($\sigma_h^2 = \sigma_\delta^2 = 0.1^2$) and moderately fat-tailed shocks ($\nu = 10$).

Table 1: Prior distributions: Monte Carlo simulation

Name	Description	Density	Specification	
			a_0	$\sqrt{A_0}$
μ_h	Intercept volatility	$\mathcal{N}(a_0, A_0)$	0.0	1.0
μ_δ	Intercept asymmetry	$\mathcal{N}(a_0, A_0)$	0.0	1.0
ϕ_h	AR parameter volatility	$\mathcal{TN}_{(-1,1)}(a_0, A_0)$	0.99	0.1
ϕ_δ	AR parameter asymmetry	$\mathcal{TN}_{(-1,1)}(a_0, A_0)$	0.99	0.1
			c_0	C_0
σ_h^2	Variance of volatility shocks	$\mathcal{IG}(c_0, C_0)$	2.5	0.025
σ_δ^2	Variance of asymmetry shocks	$\mathcal{IG}(c_0, C_0)$	2.5	0.025
			$\underline{\nu}$	$\bar{\nu}$
ν	Degrees of freedom	$\mathcal{U}(\underline{\nu}, \bar{\nu})$	0	50

Note: The parameters of the inverse-gamma prior distributions are taken from Kim et al. (1998) and imply a prior expectation of 0.017 and a prior standard deviation of 0.024.

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.[9] With respect to the convergence of the Markov chains, the diagnostic of Geweke (1992) indicates that the chains of all parameters have converged. 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, especially the asymmetry shock variance σ_δ^2. However, the simulation results are robust to changes in the number of Gibbs iterations.[10]

The degrees of freedom parameter ν 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 δ_t, ϕ_δ, 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 δ_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

[9] 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.
[10] Alternatively, the approach outlined in Yu and Meng (2011) could be used to increase sampling efficiency.

estimation of the persistence parameter of δ_t.

Table 2: Monte Carlo simulation results: δ_t specified as AR(1)

Sample size	Parameter	Mean	SE	2.5%	97.5%	Bias	CD	IF
	μ_h	0.02	0.27	−0.51	0.55	0.02	0.49	8.50
	ϕ_h	0.98	0.01	0.96	0.99	−0.01	0.48	78.48
	σ_h^2	0.01	0.00	0.01	0.02	0.00	0.47	167.61
T = 1,000	μ_δ	−0.02	0.46	−0.93	0.96	−0.02	0.46	116.96
	ϕ_δ	0.92	0.01	0.90	0.96	−0.07	0.47	56.88
	σ_δ^2	0.01	0.00	0.01	0.02	0.00	0.42	378.38
	ν	17.67	7.55	7.18	33.27	7.67	0.47	205.05
	μ_h	0.01	0.13	−0.23	0.27	0.01	0.47	6.48
	ϕ_h	0.99	0.00	0.98	0.99	0.00	0.48	95.59
	σ_h^2	0.01	0.00	0.01	0.02	0.00	0.47	212.23
T = 5,000	μ_δ	−0.01	0.19	−0.38	0.37	−0.01	0.46	83.28
	ϕ_δ	0.93	0.02	0.90	0.99	−0.06	0.44	274.06
	σ_δ^2	0.01	0.01	0.01	0.03	0.00	0.30	1037.04
	ν	11.01	2.14	8.04	16.00	1.01	0.47	161.67
	μ_h	0.02	0.10	−0.18	0.22	0.02	0.42	6.39
	ϕ_h	0.99	0.00	0.98	0.99	0.00	0.45	97.62
	σ_h^2	0.01	0.00	0.01	0.01	0.00	0.45	219.19
T = 10,000	μ_δ	0.00	0.13	−0.25	0.25	0.00	0.45	62.75
	ϕ_δ	0.95	0.03	0.90	0.99	−0.04	0.38	482.12
	σ_δ^2	0.01	0.01	0.01	0.03	0.00	0.22	1294.10
	ν	10.36	1.18	8.56	13.06	0.36	0.44	140.63

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).

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 ν since the tails now contain more information. But second, this also comes along with a more accurate estimation of the asymmetry persistence parameter ϕ_δ. When considering the de-meaned noncentral t-distribution, identification of δ_t depends crucially on ν. In the extreme case of $\nu \to \infty$, δ_t is not identified. Hence, lower values of ν lead to more precise estimates of δ_t and its parameters. Table 3 confirms this as the bias of ϕ_δ is now significantly smaller and even entirely gone for $T = 10,000$. These results highlight that, if δ_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.

Table 3: Monte Carlo simulation results: δ_t specified as AR(1) and smaller ν

Sample size	Parameter	Mean	SE	2.5%	97.5%	Bias	CD	IF
T = 1,000	μ_h	0.02	0.27	−0.53	0.52	0.02	0.48	7.17
	ϕ_h	0.98	0.01	0.95	0.99	−0.01	0.47	88.78
	σ_h^2	0.01	0.00	0.01	0.02	0.00	0.47	182.94
	μ_δ	0.01	0.32	−0.61	0.64	0.01	0.48	25.57
	ϕ_δ	0.94	0.03	0.90	0.99	−0.05	0.48	68.04
	σ_δ^2	0.02	0.02	0.01	0.05	0.01	0.43	373.97
	ν	6.05	2.06	3.93	11.55	1.05	0.46	81.60
T = 5,000	μ_h	0.01	0.13	−0.26	0.26	0.01	0.41	6.93
	ϕ_h	0.99	0.00	0.98	0.99	0.00	0.47	112.70
	σ_h^2	0.01	0.00	0.01	0.02	0.00	0.47	236.88
	μ_δ	0.00	0.15	−0.27	0.30	0.00	0.46	8.09
	ϕ_δ	0.98	0.01	0.94	0.99	−0.01	0.36	371.70
	σ_δ^2	0.02	0.03	0.01	0.03	0.01	0.34	849.95
	ν	5.16	0.48	4.41	6.18	0.16	0.48	59.12
T = 10,000	μ_h	0.01	0.10	−0.18	0.20	0.01	0.39	5.77
	ϕ_h	0.99	0.00	0.98	0.99	0.00	0.47	113.82
	σ_h^2	0.01	0.00	0.01	0.01	0.00	0.47	242.69
	μ_δ	0.00	0.11	−0.20	0.22	0.00	0.47	4.59
	ϕ_δ	0.99	0.01	0.97	0.99	0.00	0.33	545.95
	σ_δ^2	0.01	0.00	0.01	0.02	0.00	0.33	968.20
	ν	5.09	0.30	4.56	5.67	0.09	0.47	46.60

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

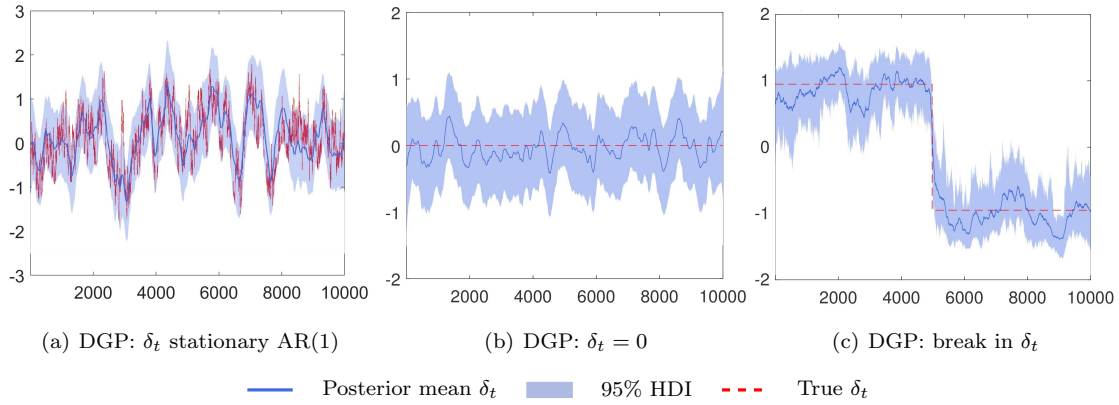
Since this is not necessarily the case in real-world datasets, an alternative is to apply the more parsimonious random walk specification for δ_t. The random walk specification can provide an accurate approximation even of dynamics that can theoretically not be non-stationary. We found this parsimonious approach very useful both when fitting the SVSS model in-sample and when using it to forecast risk measures. Figure 2 provides some illustrative examples on the flexibility of the random walk specification. The different plots contrast the estimated noncentrality parameter δ_t with the true parameter for a variety of DGPs and a single Monte Carlo sample.

In the first case, δ_t is generated as an AR(1) process (with $\phi_\delta = 0.99$). While here δ_t is misspecified when modelled as a random walk, the estimated evolution of the noncentrality parameter is a relatively accurate approximation of the true data generating process. 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.[11] The last case can be viewed as a robustness check with respect to extreme

[11] Appendix D presents additional results of a large scale Monte Carlo simulation to assess the model's performance if the DGP is symmetric. These confirm that the model can, on average, clearly identify if skewness is non-existing.

misspecification. Here, the true DGP reflects a structural break where $\delta_t = 1$ for $t = 1, ..., T/2$ and $\delta_t = -1$ for $t = T/2 + 1, ..., T$. Even in this case, the random walk provides a reasonable estimate and adjusts quickly to the constant lower level of δ_t. Lastly, Appendix D compares the SVSS model with a GARCH-type conditional skewness model (Hansen, 1994) using simulated data.

Figure 2: True vs. estimated (random walk) noncentrality parameter δ_t for various DGPs



(a) DGP: δ_t stationary AR(1) (b) DGP: $\delta_t = 0$ (c) DGP: break in δ_t

Posterior mean δ_t 95% HDI True δ_t

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 alternative SV models both in-sample and when forecasting out-of-sample. Finally, we briefly discuss the potential role of carry trading in explaining time-varying skewness.

4.1 Data

The dataset contains daily nominal exchange rates of four major currencies relative to the U.S. Dollar (USD) over the period 01/01/1977–31/10/2017 ($T = 10,255$). In particular, the Australian Dollar (AUD), the Japanese Yen (JPY), the British Pound (GBP), and the Swiss Franc (CHF) are considered. 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 $y_t = (S_t - S_{t-1})/S_{t-1} \times 100$. The data series are displayed in Appendix C. Table 4 contains summary statistics of the four return series under consideration.

Table 4: Summary statistics of exchange rate returns

Currency pair	Mean	Variance	Skewness	Kurtosis	Jarque-Bera
USD/AUD	−0.0010	0.4846	−0.8460	21.0291	0.00
USD/JPY	0.0115	0.4560	0.4927	7.3560	0.00
USD/GBP	−0.0005	0.3882	−0.2922	9.3553	0.00
USD/CHF	0.0115	0.5513	0.6247	19.1881	0.00

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

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 fatter 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.

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 to several alternative SV models, some 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:[12]

$$y_t = \beta_0 + \sum_{l=1}^{L} \beta_l y_{t-l} + e^{h_t/2}\varepsilon_t. \tag{14}$$

Depending on the distributional assumption about ε_t, we distinguish the following three stochastic volatility models: The first model (SV-t) assumes Student t-distributed innovations as in Chib et al. (2002) and thus allows for fat-tailed standardized return shocks, i.e. $\varepsilon_t \sim t(\nu)$. The second model (SV-nct) allows for both fat-tailed shocks and a constant degree of asymmetry by imposing a de-meaned noncentral t-distribution, i.e. $\varepsilon_t = v_t - \mathbb{E}[v_t]$ with $v_t \sim \mathcal{NCT}(\nu, \delta)$. The third model is the previously introduced SVSS model, which allows for time-varying skewness, where ε_t is defined as in Equations (4)-(5). Based on the arguments developed in the previous section, the asymmetry process δ_t is specified as a random walk, i.e. we set $\mu_\delta = 0$ and $\phi_\delta = 1$ in Equation (5). Finally, we also compare the results from the SVSS model to a SV specification that models skewness through occasional jumps in the return equation (SVJ-t). For this model, Equation (14) is augmented as follows

$$y_t = \beta_0 + \sum_{l=1}^{L} \beta_l y_{t-l} + k_t q_t + e^{h_t/2}\varepsilon_t, \qquad \varepsilon_t \sim t(\nu), \tag{15}$$

where $q_t \in [0, 1]$ is a binary variable that takes the value one with probability ω and k_t are the jump sizes modelled as $k_t \sim \mathcal{N}(\mu_k, \sigma_k^2)$. The prior distributions for the parameters are set as follows: $\beta_l \sim \mathcal{N}(0, 1)$ for $l = 0, ..., L$, $\mu_h \sim \mathcal{N}(0, 10)$ and $\phi_h \sim \mathcal{TN}_{(-1,1)}(0.95, 1)$. The remaining prior values are as in Table 1. The prior choice for the SVJ-t model follows Chan and Grant (2016a). The jump intensity ω is assumed to follow a uniform distribution, $\omega \sim \mathcal{U}(0, 0.1)$. The average jump size and the log-jump variance $\gamma = (\mu_k, \log(\sigma_k^2))$ are assumed to be distributed as $\gamma \sim \mathcal{N}(\gamma_0, V_\gamma)$. The hyperparameters are $\gamma_0 = (0, \log(10))'$ and $V_\gamma = \text{diag}(10, 1)$.[13]

[12] 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.

[13] These values imply an expected jump probability of 5% (around 13 jumps per year with daily data). The jump size is centered around zero with a standard deviation of around 3pp.

The reported results are based on 100,000 iterations with 50,000 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. 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., 2020), which limits their feasibility in hierarchical latent variable models and large datasets.[14] However, when assessing the models, we do not take the estimated DIC values at face value but also consider the posterior parameter distributions.

Tables 5 - 8 present the estimation results for the four SV models applied to the 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 β 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). Second, (log-)volatility h_t is strongly persistent across all models considered supporting the idea of volatility clustering (Mandelbrot, 1963). Third, standardized exchange rate returns exhibit fat tails. Posterior means of the degrees of freedom ν lie in the range of 5 to 10 with the posterior standard deviations being small.

Table 5: USD/AUD exchange rate returns: estimation results

		SV-t		SVJ-t		SV-nct		SVSS	
	Coef.	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mean	β_0	0.009	0.004	0.009	0.004	0.003	0.004	0.006	0.004
	β_1	0.007	0.009	0.007	0.009	0.005	0.009	0.001	0.009
	β_2	−0.016	0.009	−0.016	0.009	−0.019	0.009	−0.022	0.009
	β_3	−0.019	0.009	−0.019	0.009	−0.019	0.009	−0.021	0.009
	β_4	0.009	0.010	0.009	0.010	0.007	0.009	0.005	0.009
Variance	μ_h	−1.695	0.282	−1.696	0.285	−1.707	0.293	−1.727	0.279
	ϕ_h	0.996	0.001	0.996	0.001	0.996	0.001	0.996	0.001
	σ_h^2	0.009	0.001	0.010	0.002	0.009	0.001	0.010	0.001
Jumps	κ	—	—	0.0004	0.0008	—	—	—	—
	μ_k	—	—	0.131	0.043	—	—	—	—
	σ_k^2	—	—	9.775	12.120	—	—	—	—
Skewness	δ	—	—	—	—	−0.327	0.069	—	—
	σ_δ^2	—	—	—	—	—	—	0.003	0.001
Df	ν	6.324	0.415	6.353	0.420	6.320	0.398	6.722	0.433
DIC		15,872		15,854		15,856		15,823	

[14] Chan and Grant (2016b) find that the conditional DIC tends to favour overfitted models and Li et al. (2020) argue against its suitability for latent variable models.

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

		SV-t		SVJ-t		SV-nct		SVSS	
	Coef.	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mean	β_0	−0.007	0.005	−0.007	0.005	0.003	0.006	0.002	0.006
	β_1	−0.007	0.009	−0.007	0.009	−0.008	0.009	−0.008	0.009
	β_2	−0.006	0.009	−0.006	0.009	−0.008	0.009	−0.011	0.009
	β_3	0.004	0.009	0.004	0.009	0.003	0.009	0.002	0.009
	β_4	0.016	0.009	0.016	0.009	0.015	0.009	0.015	0.009
Variance	μ_h	−1.404	0.075	−1.403	0.077	−1.401	0.076	−1.412	0.076
	ϕ_h	0.984	0.003	0.984	0.003	0.984	0.003	0.984	0.003
	σ_h^2	0.013	0.002	0.013	0.003	0.013	0.002	0.013	0.003
Jumps	κ	—	—	0.0004	0.0007	—	—	—	—
	μ_k	—	—	−0.387	0.096	—	—	—	—
	σ_k^2	—	—	9.034	10.481	—	—	—	—
Skewness	δ	—	—	—	—	0.275	0.065	—	—
	σ_δ^2	—	—	—	—	—	—	0.002	0.001
Df	ν	5.798	0.366	5.810	0.376	5.937	0.387	6.308	0.429
DIC		18,837		18,818		18,821		18,815	

Table 7: USD/GBP exchange rate returns: estimation results

		SV-t		SVJ-t		SV-nct		SVSS	
	Coef.	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mean	β_0	0.008	0.003	0.008	0.003	0.007	0.003	0.007	0.003
	β_1	0.016	0.009	0.017	0.009	0.016	0.009	0.016	0.009
	β_2	−0.008	0.010	−0.008	0.009	−0.009	0.009	−0.010	0.010
	β_3	−0.014	0.010	−0.014	0.010	−0.013	0.010	−0.014	0.010
	β_4	0.004	0.009	0.004	0.009	0.004	0.009	0.004	0.009
Variance	μ_h	−1.575	0.179	−1.576	0.180	−1.580	0.186	−1.589	0.181
	ϕ_h	0.994	0.001	0.994	0.001	0.994	0.001	0.994	0.001
	σ_h^2	0.010	0.002	0.010	0.001	0.010	0.001	0.010	0.001
Jumps	κ	—	—	0.0009	0.0024	—	—	—	—
	μ_k	—	—	−0.035	0.174	—	—	—	—
	σ_k^2	—	—	8.920	10.705	—	—	—	—
Skewness	δ	—	—	—	—	−0.113	0.076	—	—
	σ_δ^2	—	—	—	—	—	—	0.003	0.001
Df	ν	7.598	0.575	7.654	0.609	7.507	0.548	8.037	0.622
DIC		16,365		16,328		16,368		16,373	

Table 8: USD/CHF exchange rate returns: estimation results

		SV-t		SVJ-t		SV-nct		SVSS	
	Coef.	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mean	β_0	−0.001	0.006	−0.002	0.006	0.004	0.006	0.004	0.006
	β_1	0.000	0.010	0.000	0.010	−0.001	0.010	−0.003	0.010
	β_2	0.002	0.010	0.002	0.010	0.001	0.010	−0.001	0.010
	β_3	0.006	0.010	0.006	0.010	0.005	0.010	0.005	0.010
	β_4	0.003	0.010	0.003	0.010	0.002	0.010	0.002	0.010
Variance	μ_h	−1.100	0.091	−1.101	0.091	−1.105	0.093	−1.107	0.091
	ϕ_h	0.989	0.002	0.989	0.002	0.990	0.002	0.990	0.002
	σ_h^2	0.008	0.001	0.008	0.001	0.008	0.001	0.008	0.001
Jumps	κ	—	—	0.0025	0.0050	—	—	—	—
	μ_k	—	—	0.632	0.108	—	—	—	—
	σ_k^2	—	—	6.289	6.892	—	—	—	—
Skewness	δ	—	—	—	—	0.272	0.085	—	—
	σ_δ^2	—	—	—	—	—	—	0.003	0.001
Df	ν	8.469	0.735	8.674	0.891	8.422	0.714	9.781	0.946
DIC		20,756		20,639		20,754		20,755	

Furthermore, exchange rate returns are not symmetric. In the SV-nct model, the 95% highest density interval (HDI) of δ 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 ambiguous. The DIC supports these findings since the SV-nct model is characterized by a smaller criterion value compared to the SV-t model in all cases except for the British Pound returns. It is worth noting that the signs of the estimated noncentrality parameters are in line with the unconditional skewness measures reported in Table 4.

The SVJ-t model outperforms both the SV-t and the SV-nct model in all four cases whereas the reduction in the DIC is particularly strong in case of the British Pound and the Swiss Franc. Especially for the Swiss Franc the SVJ-t model seems to capture the positively skewed return distribution very well. In our daily dataset the jump probability of 0.25% implies, on average, around one jump per year and the positive average jump size μ_k indicates that these jumps mostly reflect appreciations of the Swiss currency.

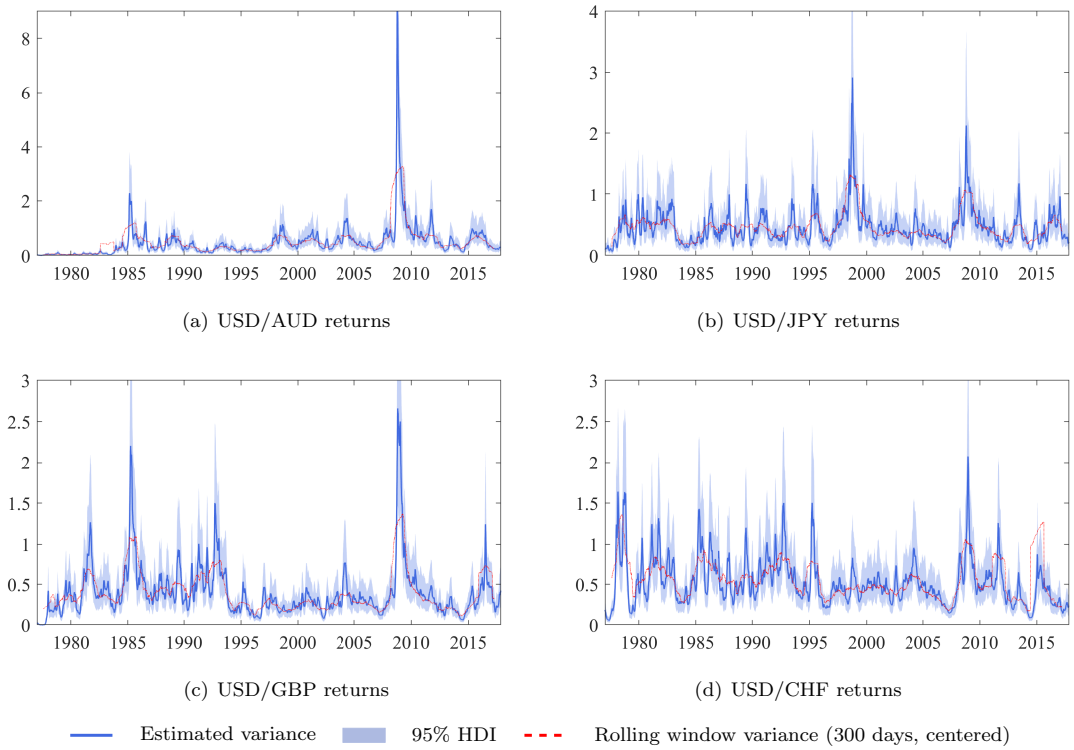
When looking at the SVSS model, the posterior means of the innovation variances of δ_t, σ_δ^2, are of relevant and similar magnitude (even almost identical when rounded to three digits). In addition, posterior dispersion is fairly small. This can be taken as evidence in favour of time-varying skewness. The DIC ranks the SVSS model first in case of the Australian Dollar and the Japanese Yen while in case of the British Pound and the Swiss Franc the jump specification seems to fit the data better. In summary, when looking at both the posterior distributions of the parameters and the information criterion, the SVSS model seems to be the most suitable modelling

approach for the currencies (AUD and JPY) that are heavily involved in carry trading and that thus presumably show a particularly skewed return distribution with a changing degree of skewness over time. Finally, Appendix D presents the results of additional analyses. First, it contains an extensive section on comparing the empirical results of the SVSS model with a GARCH-type model that features conditional skewness. Second, since the sample used for the estimations spans a relatively long period, another section discusses the aspect of parameter stability by means of a small subsample analysis.

4.3 Time-varying volatility and skewness

We now turn towards discussing the estimated volatility and, in particular, the skewness series. Figure 3 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 3: Estimated variance (volatility) series

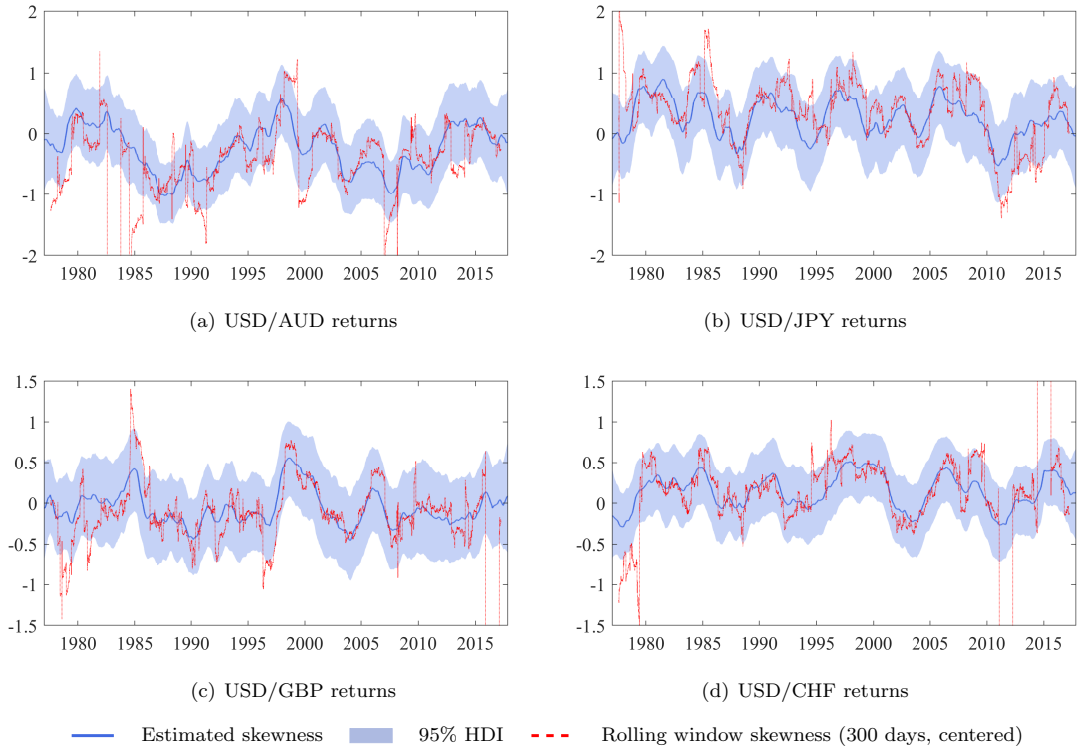


While time-varying volatility is a well-known stylized fact of financial returns, Figure 4 presents the key finding of this paper, i.e. time variation in the estimated skewness of exchange rate returns. Again, a rolling window unconditional skewness measure is plotted next to the skewness estimate

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 the cross-currency co-movement of volatility and crash risk, Table 9 presents the correlation matrices for both measures. Since all four bilateral exchange rates are measured relative to the U.S. Dollar, this 'common factor' naturally induces positive correlation in both volatility and skewness across return series. Nevertheless, some returns co-move stronger 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.

Figure 4: Estimated skewness series



This is possibly due to the historically tight relationship between Australia and the United Kingdom. The Japanese Yen and the Swiss Franc are both considered funding currencies which could explain why their skewness shows a high degree of 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 global economic shocks.

Table 9: Correlations of variance (volatility) and skewness across return series

	Volatility				Skewness			
FX	\$/AUD	\$/JPY	\$/GBP	\$/CHF	\$/AUD	\$/JPY	\$/GBP	\$/CHF
\$/AUD	1.00	0.42	0.64	0.31	1.00	0.27	0.41	0.28
\$/JPY	—	1.00	0.31	0.38	—	1.00	0.20	0.64
\$/GBP	—	—	1.00	0.64	—	—	1.00	0.60
\$/CHF	—	—	—	1.00	—	—	—	1.00

Note: All p-values of tests for statistical significance of these correlation coefficients are virtually 0.

4.4 Forecasting risk measures

This section assesses the potential value of the SVSS model relative to the alternative specifications for forecasting risk. Specifically, we focus on forecasting Value-at-Risk (VaR), which equals a certain quantile of the return distribution and expected shortfall (ES) defined as the conditional expectation beyond the VaR level.[15] Both measures are regularly used by practitioners.

For the forecasting exercise, the sample is split at the of 2006, so that we forecast risk measures for each day over the period 01/01/2007–31/10/2017. The out-of-sample period includes both tranquil market periods and several extreme events that had strong direct and indirect effects on our set of currencies, such as the 2011 earthquake in Japan, the European debt crisis, the Great Recession and the 'Brexit' vote. For each day t of the out-of-sample period we generate VaR and ES forecasts using the model estimated given data up until day $t-1$. Through this recursive one-day ahead forecasting exercise we obtain a set of $N = 2,722$ VaR and ES forecasts for each model. We forecast VaR and ES at the left (right) tail of the return distribution to measure the risk of an investor that takes a long (short) position in the foreign exchange market. The forecasts are estimated and evaluated at levels $\alpha \in \{0.5\%, 1\%, 5\%, 95\%, 99\%, 99.5\%\}$. Specifically, to generate one-day ahead VaR forecasts, in each MCMC iteration we obtain the one-day ahead forecasts of the latent variables and generate a draw from the corresponding predictive density using the observation equation. The VaR^{α} is then estimated as the empirical α-quantile of the draws from the predictive density whereas ES^{α} is estimated as the mean of the draws that lie below (above) the estimated VaR for the left (right) tail of the predictive distribution.

In order to evaluate the VaR forecasts, we apply two commonly used testing procedures. Kupiec (1995) develops a likelihood ratio test for unconditional coverage (UC) to check whether the empirical rate of violations equals α. Define a VaR$^{\alpha}$ violation for the left (right) tail as $V_t^{\alpha} = 1$ if $y_t < \widehat{VaR}_t^{\alpha}$ ($y_t > \widehat{VaR}_t^{\alpha}$) and 0 otherwise and the total number of violations as $n_0 = \sum_{t=1}^{N} V_t^{\alpha}$ and $n_1 = N - n_0$. Under the null hypothesis of unconditional coverage, the test statistic

$$LR_{UC} = -2\, log\left(\frac{(1-\alpha)^{n_1}\alpha^{n_0}}{(1-\widehat{\alpha})^{n_1}\widehat{\alpha}^{n_0}}\right) \tag{16}$$

is distributed as $\chi^2(1)$ where $\widehat{\alpha} = \frac{n_0}{n_0+n_1}$ is the empirical coverage rate. Moreover, to test whether

[15] The design of the forecasting exercise is similar to Aas and Haff (2006) and Nakajima (2013).

the violations are independent, we apply the testing procedure proposed by Christoffersen (1998). Under the null hypothesis of independence, the test statistic

$$LR_{IND} = -2\,log\left(\frac{(1-\pi)^{n_{00}+n_{10}}\pi^{n_{01}+n_{11}}}{(1-\pi_0)^{n_{00}}\pi_0^{n_{01}}(1-\pi_0)^{n_{10}}\pi_1^{n_{11}}}\right) \tag{17}$$

is distributed as $\chi^2(1)$. Let n_{ij} the number of times where outcome j, i.e. a violation or no violation, occurred on a day conditional on outcome i occurring on the previous day and where $\pi_0 = \frac{n_{01}}{n_{00}+n_{01}}$, $\pi_1 = \frac{n_{11}}{n_{10}+n_{11}}$ and $\pi = \frac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}}$. To jointly test for unconditional coverage and independence, Christoffersen (1998) defines the test statistic for conditional coverage (CC) as

$$LR_{CC} = LR_{UC} + LR_{IND}, \tag{18}$$

which, under the null hypothesis, is distributed as $\chi^2(2)$.

To evaluate the expected shortfall forecasts we follow, among others, Aas and Haff (2006) and Nakajima (2013) and use the measure developed by Embrechts et al. (2005). Given the one-day ahead forecast of expected shortfall, $\widehat{ES}_t^{\alpha}$, define $D_t^{\alpha} = y_t - \widehat{ES}_t^{\alpha}$. Moreover, $\widetilde{D}^{\alpha}$ denotes the α-quantile of $\{D_t^{\alpha}\}_{t=1}^N$, $J_t^{\alpha} = \mathbb{1}\{D_t^{\alpha} < \widetilde{D}^{\alpha}\}$ for the left tail and $J_t^{\alpha} = \mathbb{1}\{D_t^{\alpha} > \widetilde{D}^{\alpha}\}$ for the right tail. Finally, let $n_\tau = \sum_{t=1}^N J_t^{\alpha}$. The measure of expected shortfall forecast precision is then defined as

$$V^{ES} = \frac{|V_1^{ES}| + |V_2^{ES}|}{2}, \quad \text{where} \quad V_1^{ES} = \frac{1}{n_0}\sum_{V_t^{\alpha}=1} D_t^{\alpha}, \quad \text{and} \quad V_2^{ES} = \frac{1}{n_\tau}\sum_{J_t^{\alpha}=1} D_t^{\alpha}. \tag{19}$$

V_1^{ES} is a natural measure of forecast accuracy as it captures the average difference between the realized returns and the one-period ahead ES forecasts conditional on the returns being smaller (larger) than the predicted VaR for the left (right) tail. V_2^{ES} can be seen as a term correcting for the fact that V_1^{ES} strongly depends on the precision of the VaR forecast. A smaller value of V^{ES} reflects a more accurate expected shortfall forecast.

Table 10 presents the results of the VaR forecast evaluation for the four models under consideration.[16] The performance of the models in forecasting risk varies across exchange rates. In case of the Australian Dollar, conditional coverage is rejected for certain α-levels for all models except the SVSS specification. Even though independence is rejected for the SVSS model at the 5%- and 95%-quantile, overall the model seems to capture VaR better than the alternative models, especially at the extreme levels. In case of the Yen, all models seem to do well in forecasting VaR. Interestingly, while all models fail in capturing VaR at $\alpha = 99\%$ for the British pound rate, the SVJ-t model, which was preferred in the in-sample analysis, also fails at $\alpha = 99.5\%$. Finally, in case of the Swiss Franc, the models perform rather similar and forecast VaR overall accurately.

When evaluating the expected shortfall forecasts, we observe larger differences across models with the SVSS model performing, on average, better than the alternative specifications (Table 11). In case of the Australian Dollar and the Japanese Yen, the SVSS model provides the most accurate ES forecasts for the majority of levels.

[16] Again, in case of the SVSS model, δ_t is assumed to follow a random walk. We found this parsimonious specification to generally forecast more accurately than the stationary AR(1) alternative.

Table 10: Likelihood ratio tests for unconditional coverage, independence and conditional coverage of the Value-at-Risk violations

	USD/AUD				USD/JPY				USD/GBP				USD/CHF			
α	SV-t	SVJ-t	SV-nct	SVSS	SV-t	SVJ-t	SV-nct	SVSS	SV-t	SVJ-t	SV-nct	SVSS	SV-t	SVJ-t	SV-nct	SVSS
(a) Number of Value-at-Risk violations																
0.5%	11	11	11	11	9	10	12	12	11	11	11	11	9	7	11	12
1%	21	21	17	19	20	21	25	26	24	24	23	22	18	18	20	24
5%	156	158	143	146	125	123	133	133	162	161	156	156	119	119	130	128
0.95%	116	114	125	130	138	139	119	133	121	120	121	126	127	126	121	126
99%	14	14	20	19	22	23	17	21	11	12	12	12	30	29	25	27
99.5%	1	1	4	7	11	12	10	11	5	5	6	6	14	13	11	12
(b) p-values for LR-UC test																
0.5%	0.46	0.46	0.46	0.46	0.18	0.30	0.66	0.66	0.46	0.46	0.46	0.46	0.18	**0.05**	0.46	0.66
1%	0.21	0.21	**0.03**	0.09	0.14	0.21	0.66	0.81	0.53	0.53	0.40	0.30	0.06	0.06	0.14	0.53
5%	0.09	0.06	0.55	0.39	0.32	0.24	0.78	0.78	**0.03**	**0.03**	0.09	0.09	0.12	0.12	0.59	0.47
95%	0.07	**0.05**	0.32	0.59	0.87	0.80	0.12	0.78	0.18	0.15	0.18	0.37	0.42	0.37	0.18	0.37
99%	**0.00**	**0.00**	0.14	0.09	0.30	0.40	**0.03**	0.21	**0.00**	**0.00**	**0.00**	**0.00**	0.60	0.73	0.66	0.97
99.5%	**0.00**	**0.00**	**0.00**	**0.05**	0.46	0.66	0.30	0.46	**0.01**	**0.01**	**0.02**	**0.02**	0.92	0.87	0.46	0.66
(c) p-values for LR-IND test																
0.5%	0.77	0.77	0.77	0.77	0.81	0.79	0.74	0.74	0.77	0.77	0.77	0.77	0.81	0.85	0.77	0.74
1%	0.57	0.57	0.64	0.61	0.59	0.57	0.50	0.48	0.21	0.21	0.19	0.17	0.62	0.62	0.59	0.51
5%	0.05	0.11	0.05	**0.04**	0.91	0.85	0.83	0.83	0.90	0.87	0.73	0.73	0.57	0.28	0.93	0.99
95%	0.12	0.14	0.06	**0.04**	0.68	0.66	0.57	0.27	0.25	0.27	0.25	0.18	0.68	0.71	0.86	0.71
99%	0.70	0.70	0.59	0.61	0.55	0.53	0.64	0.57	0.77	0.74	0.74	0.74	0.34	0.32	0.23	0.27
99.5%	0.98	0.98	0.91	0.85	0.77	0.74	0.79	0.77	0.89	0.89	0.87	0.87	0.06	0.05	**0.03**	**0.04**
(d) p-values for LR-CC test																
0.5%	0.73	0.73	0.73	0.73	0.40	0.57	0.86	0.86	0.73	0.73	0.73	0.73	0.40	0.14	0.73	0.86
1%	0.39	0.39	0.10	0.22	0.30	0.39	0.72	0.76	0.37	0.37	0.30	0.23	0.15	0.15	0.30	0.66
5%	**0.03**	**0.05**	0.12	0.08	0.61	0.49	0.94	0.94	0.09	0.10	0.22	0.22	0.26	0.17	0.86	0.77
95%	0.06	**0.05**	0.11	0.11	0.91	0.88	0.26	0.52	0.21	0.19	0.21	0.28	0.66	0.62	0.39	0.62
99%	**0.02**	**0.02**	0.30	0.22	0.49	0.58	0.10	0.39	**0.00**	**0.00**	**0.00**	**0.00**	0.55	0.57	0.44	0.55
99.5%	**0.00**	**0.00**	**0.01**	0.14	0.73	0.86	0.57	0.73	**0.03**	**0.03**	0.07	0.07	0.17	0.15	0.08	0.11

Note: This table contains the number of VaR violations based on the out-of-sample forecasting exercise for each exchange rate and each model (a) as well as the p-values corresponding to the likelihood ratio (LR) test for unconditonal coverage following Kupiec (1995) (b) and the LR tests for independence of the violations (c) and conditional coverage (d) following Christoffersen (1998). Bold numbers indicate p-values < 0.05.

Notably, in cases where the SVSS model is preferred, the gains in precision are often large, whereas in the remaining cases the differences compared to the preferred specification are mostly small. In case of the British Pound, the SVSS model performs somewhat better than the remaining models in forecasting expected shortfall at the higher levels. Stochastic skewness seems to be least relevant when forecasting ES in case of the Swiss Franc. In addition, no model dominates the alternatives for this currency.

Table 11: Precision of expected shortfall forecasts

	USD/AUD				USD/JPY			
α	SV-t	SVJ-t	SV-nct	SVSS	SV-t	SVJ-t	SV-nct	SVSS
0.5%	0.331	0.272	0.072	**0.048**	0.147	0.109	0.189	**0.042**
1%	0.140	0.114	0.121	**0.080**	0.077	0.057	0.083	**0.004**
5%	**0.054**	0.058	0.066	0.060	0.037	0.036	0.030	**0.018**
0.95%	0.151	0.150	0.075	**0.060**	0.030	0.026	0.027	**0.016**
99%	0.462	0.460	0.316	**0.243**	**0.066**	0.067	0.144	0.074
99.5%	0.820	0.731	0.642	**0.501**	0.130	0.130	0.111	**0.059**
	USD/GBP				USD/CHF			
α	SV-t	SVJ-t	SV-nct	SVSS	SV-t	SVJ-t	SV-nct	SVSS
0.5%	0.230	0.223	**0.207**	0.265	**0.314**	0.455	0.378	0.388
1%	**0.053**	0.058	0.069	0.121	0.135	**0.123**	0.212	0.196
5%	0.038	0.035	**0.033**	0.035	0.031	0.028	**0.005**	0.025
0.95%	0.117	0.121	0.104	**0.087**	0.142	0.132	**0.112**	0.128
99%	0.145	0.151	0.116	**0.103**	0.558	**0.537**	0.567	0.607
99.5%	**0.209**	0.257	0.243	0.222	**1.078**	1.079	1.169	1.223

Note: This table contains the expected shortfall accuracy measure of Embrechts et al. (2005). Bold numbers indicate the lowest value for each quantile.

Overall, the results of this forecasting exercise suggest that allowing the return innovations to feature stochastic skewness can help to improve Value-at-Risk predictions and, especially, expected shortfall forecasts. Interestingly, the SVJ-t model, which has been the strongest competitor of the SVSS model in the in-sample analysis, predicts both Value-at-Risk and expected shortfall, on average, worse than the SVSS model. In line with the previous section, time-varying asymmetry seems to be particularly relevant when modelling the return series of the Australian Dollar and Japanese Yen.

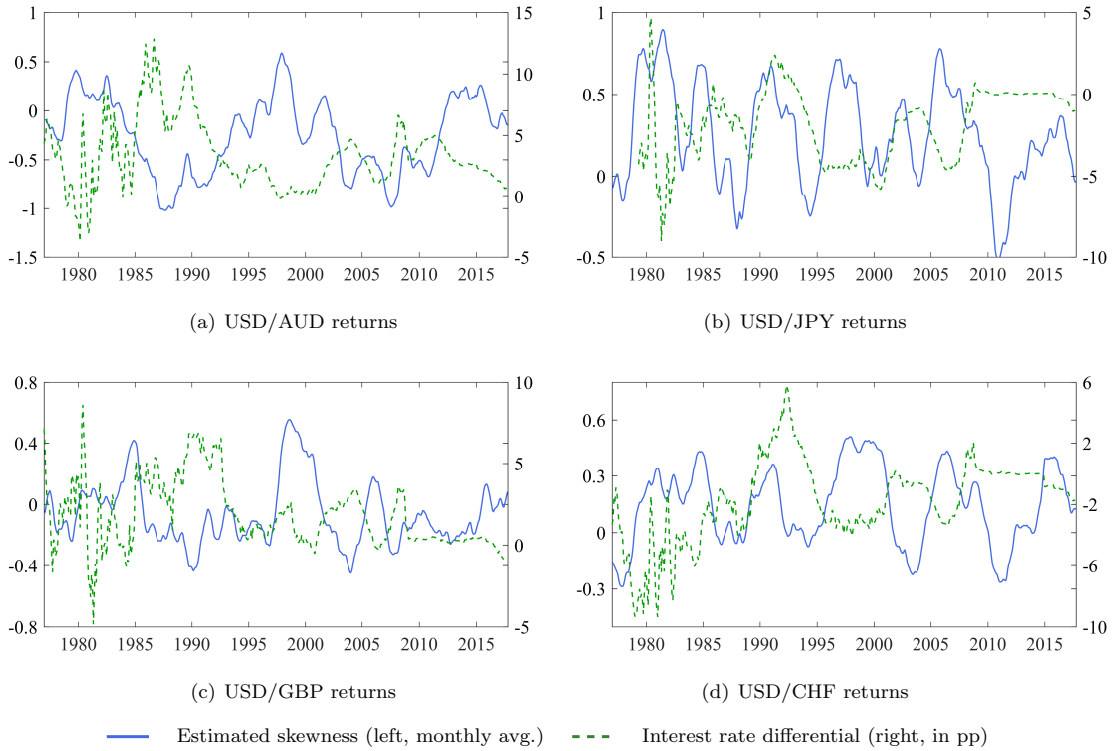
4.5 Skewness, carry trade and return predictability

The focus of this paper has been to introduce stochastic skewness into a standard SV model, measuring the degree of time-varying asymmetry in exchange rate returns and assessing its potential to improve risk forecasts. This section returns to the initial motivation for modelling time-varying asymmetry in exchange rate returns. It presents evidence of the link between carry trading and

skewness and some support for the notion that excess returns potentially reflect investors' compensation for an increased crash risk exposure.

Figure 5 plots the estimated skewness series (monthly averages) along with the monthly nominal interest rate differentials (foreign 3-months government bond yield minus 3-months U.S. Treasury-bill yield).[17] 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. (2008) and Farhi et al. (2015), i.e. a negative correlation between carry trade activity and crash risk. 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). Once the carry trade unwinded, this resulted in a crash of the USD/AUD exchange rate from around 1.0 USD to 0.6 USD between July and October 2008.

Figure 5: Estimated skewness and interest rate differential



Note: The monthly nominal interest rate differential is defined as the foreign 3-months government bond yield minus the 3-months U.S. Treasury-bill yield. Data availability for Japanese rates only starts in 05/1979.

An illustrative example in the case of Japan is the rising interest rate differential during 2005 that created incentives for investors to increase their carry trade positions. Since the Japanese Yen

[17] Appendix C includes a time series plot showing the raw interest rates over the sample period.

has acted as a funding currency, skewness has increased over this period. As noted by Han and Westelius (2019), the unwinding of carry trade positions due to increased risk aversion starting in 2006 led to an appreciation of the Yen. Notably, while skewness dropped significantly, the interest rate differential stayed almost constant during this period. 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 5 also indicates that the negative correlation has essentially vanished after the global financial crisis in all cases except for the Australian Dollar since most countries hit the zero lower bound of short-term nominal interest rates.

To get further insights into the dynamic relationship between the interest rate differential and skewness, Table 12 applies a similar analysis as in Brunnermeier et al. (2008) and considers predictive panel regressions at different monthly horizons of skewness on the interest rate differential, where i_t^* and i_t represent the foreign and domestic interest rates (in %), the nominal exchange rate return y_t and the skewness realization in month t. The first row suggests that the interest rate differential has significant predictive power for future crash risk even when considering a horizon of two years.[18] While the direct effect of a change in the interest rate differential increases for larger horizons, the long-run effect, which accounts for the persistence of skewness, decreases. Moreover, past returns help to predict skewness up to six months ahead. This could be the result of "currency gains leading to larger speculator positions and larger future crash risk" (Brunnermeier et al., 2008, p. 329).

Table 12: Predictive regressions with estimated skewness (monthly avg.) as dependent variable

	Skewness_{t+1}	Skewness_{t+6}	Skewness_{t+12}	Skewness_{t+24}
$i_t^* - i_t$	-0.001^{**}	-0.009^{***}	-0.021^{***}	-0.034^{***}
	(0.0005)	(0.0028)	(0.0057)	(0.0077)
y_t	-0.027^{***}	-0.044^{**}	-0.031	0.021
	(0.0052)	(0.0182)	(0.0348)	(0.0695)
Skewness_t	0.988^{***}	0.826^{***}	0.543^{***}	0.123
	(0.0021)	(0.0246)	(0.0651)	(0.1439)
Adj. R^2	0.99	0.76	0.43	0.13

Note: The table reports the estimated coefficients (standard errors) of unbalanced panel regressions with currency fixed effects and monthly data. The standard errors are adjusted for serial correlation using a Newey-West covariance matrix and *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 13 sheds light on the question whether skewness is priced by looking into its predictive power for future exchange rate returns. Specifically, the table presents for each daily return series short-horizon ($h = 1$) and longer-horizon ($h > 1$) predictive regressions where the predictors are

[18] Moreover, when excluding the recent zero lower bound period, which is in most cases nearly uninformative with respect to the link between interest rate differentials and skewness, the results are even more pronounced.

the estimated volatility and skewness series obtained from the SVSS model. Since both of these predictors are highly persistent, we apply the IVX approach developed by Magdalinos and Phillips (2009), and extended in Kostakis et al. (2015), to ensure reliable inference. While the evidence for the predictive ability of volatility across different horizons appears strongest in case of the Japanese Yen, the results are mixed when looking at the remaining currencies. In contrast, skewness has significant predictive power for future returns across the four currencies and across horizons. Overall, these results confirm previous findings in the literature suggesting that skewness is priced while the consistently negative signs of the estimated coefficients indicate investors' demand for compensation when accepting crash risk exposure. Related recent empirical contributions showing that skewness is indeed priced include Rafferty (2012), who identifies a global currency skewness factor, and Broll (2016), who finds evidence of a skewness risk premium by comparing currency options' implied skewness and future realized skewness.

Table 13: Predictive regressions with daily exchange rate returns as dependent variable

	$h=1$		$h=10$		$h=30$		$h=60$	
USD/AUD	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald
Volatility_t	-0.047	17.930^{***}	-0.034	9.119^{***}	-0.015	1.583	0.001	0.011
Skewness_t	-0.057	9.952^{***}	-0.054	8.810^{***}	-0.048	7.102^{***}	-0.042	5.470^{**}
USD/JPY	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald
Volatility_t	0.141	32.239^{***}	0.115	20.872^{***}	0.093	11.936^{***}	0.075	6.078^{**}
Skewness_t	-0.065	8.791^{***}	-0.062	8.038^{***}	-0.057	6.759^{***}	-0.050	5.193^{**}
USD/GBP	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald
Volatility_t	-0.059	8.754^{***}	-0.048	5.773^{**}	-0.034	2.699	-0.016	0.521
Skewness_t	-0.073	6.042^{**}	-0.071	5.793^{**}	-0.066	4.985^{**}	-0.060	3.974^{**}
USD/CHF	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald	$\hat{\beta}_{IVX}$	Wald
Volatility_t	0.050	3.504^{*}	0.030	1.224	0.003	0.014	-0.010	0.092
Skewness_t	-0.136	12.802^{***}	-0.135	12.636^{***}	-0.131	11.897^{***}	-0.123	10.389^{***}

Note: The table reports the estimated coefficients of predictive regressions with daily nominal exchange rate returns as the dependent variable using the IVX approach (Magdalinos and Phillips, 2009; Kostakis et al., 2015). *Wald* refers to the Wald statistic under the null hypothesis that the corresponding coefficient is equal to zero and *, **, and *** denote significance at the 10%, 5%, and 1% level. For the long-horizon predictive regressions ($h > 1$), the dependent variable is the h-period cumulative nominal exchange rate return (see Kostakis et al., 2015).

Clearly, to make definite statements about whether and how strongly skewness is priced requires a more formal approach. One possibility would be to apply the SVSS model to a much larger panel of exchange rate returns and to use the estimated skewness series as risk factors in the procedure suggested by Fama and MacBeth (1973). While such an analysis is beyond the scope of this paper, it would help determine the actual skewness risk premium.

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 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 01/01/1977–31/10/2017. The following results are obtained: For currencies that are frequently subject to carry trading, the model outperforms alternative SV models that assume either symmetric shocks, a constant degree of asymmetry or that include jumps in the return equation, both in-sample and when forecasting risk measures out-of-sample. Thus, evidence is found in favour of time-varying skewness in certain exchange rate returns. In addition, interest rate differentials can help to predict future skewness. This points towards the carry trade as at least an amplifier of crash risk. Lastly, the results support the existence of a skewness risk premium in exchange rate markets.

From the perspective of policy makers and investors, these results have important implications. If monetary policy uses the policy rate to target inflation, there might be undesired side effects in case this leads to larger interest rate differentials relative to other countries. These growing interest rate differentials increase carry trading activity resulting in higher currency crash risk. If this risk eventually materializes in a sudden depreciation of the domestic currency, this could in turn give rise to higher inflation since imports become more expensive. The proposed model could serve as a tool to monitor exchange rate vulnerability. The potential benefits of the results for investors arise naturally from the forecasting exercise. Since the SVSS model can help to improve forecasts of risk associated with FX positions, which are required input for the process of portfolio optimization, the quality of investors' risk management can improve.

We identify several avenues for future research. This paper has used the noncentral t-distribution to model asymmetric shocks, largely because implementing time-varying skewness is relatively straightforward. However, a large number of distributions fulfills the general requirements to model time-varying asymmetry and a thorough comparison of their performance could be fruitful. 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 be beneficial in establishing a closer link with asset pricing theory as it would allow for a direct impact of skewness on returns.

References

Aas, K. and Haff, I. H. (2006). The Generalized Hyperbolic Skew Students t-Distribution. *Journal of Financial Econometrics*, 4(2):275–309.

Abanto-Valle, C., Lachos, V., and Dey, D. K. (2015). Bayesian Estimation of a Skew-Student-t Stochastic Volatility Model. *Methodology and Computing in Applied Probability*, 17(3):721.

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118(1):135–167.

Antolin-Diaz, J., Drechsel, T., and Petrella, I. (2017). Tracking the Slowdown in Long-Run GDP Growth. *The Review of Economics and Statistics*, 99(2):343–356.

Arditti, F. D. (1967). Risk and the Required Return on Equity. *The Journal of Finance*, 22(1):19–36.

Bakshi, G., Carr, P., and Wu, L. (2008). Stochastic risk premiums, stochastic skewness in currency options, and stochastic discount factors in international economies. *Journal of Financial Economics*, 87(1):132–156.

Barberis, N. and Huang, M. (2008). Stocks as Lotteries: The Implications of Probability Weighting for Security Prices. *American Economic Review*, 98(5):2066–2100.

Berg, A., Meyer, R., and Yu, J. (2004). Deviance Information Criterion for Comparing Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 22(1):107–120.

Broll, M. (2016). The skewness risk premium in currency markets. *Economic Modelling*, 58:494–511.

Brunnermeier, M. K., Nagel, S., and Pedersen, L. H. (2008). Carry Trades and Currency Crashes. *NBER Macroeconomics Annual*, 23(1):313–348.

Brunnermeier, M. K. and Parker, J. A. (2005). Optimal Expectations. *American Economic Review*, 95(4):1092–1118.

Burnside, C., Eichenbaum, M., Kleshchelski, I., and Rebelo, S. (2011). Do Peso Problems Explain the Returns to the Carry Trade? *The Review of Financial Studies*, 24(3):853–891.

Cappuccio, N., Lubian, D., and Raggi, D. (2004). MCMC Bayesian Estimation of a Skew-GED Stochastic Volatility Model. *Studies in Nonlinear Dynamics & Econometrics*, 8(2):1–31.

Carr, P. and Wu, L. (2007). Stochastic skew in currency options. *Journal of Financial Economics*, 86(1):213–247.

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

Chan, J. C. and Grant, A. L. (2016a). Modeling energy price dynamics: GARCH versus stochastic volatility. *Energy Economics*, 54:182–189.

Chan, J. C. and Grant, A. L. (2016b). On the Observed-Data Deviance Information Criterion for Volatility Modeling. *Journal of Financial Econometrics*, 14(4):772–802.

Chan, J. C. and Hsiao, C. Y. (2014). Estimation of Stochastic Volatility Models with Heavy Tails and Serial Dependence. In Jeliazkov, I. and Yang, X.-S., editors, *Bayesian Inference in the Social Sciences*, pages 155–176. Wiley-Blackwell.

Chan, J. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Chernov, M., Graveline, J., and Zviadadze, I. (2018). Crash Risk in Currency Returns. *Journal of Financial and Quantitative Analysis*, 53(1):137–170.

Chib, S. (2001). Markov Chain Monte Carlo Methods: Computation and Inference. *Handbook of Econometrics*, 5:3569–3649.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316.

Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4):841–862.

Christoffersen, P., Heston, S., and Jacobs, K. (2006). Option valuation with conditional skewness. *Journal of Econometrics*, 131(1):253–284.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford University Press.

Embrechts, P., Kaufmann, R., and Patie, P. (2005). Strategic Long-Term Financial Risks: Single Risk Factors. *Computational Optimization and Applications*, 32(1-2):61–90.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007.

Fama, E. F. (1984). Forward and spot exchange rates. *Journal of Monetary Economics*, 14(3):319–338.

Fama, E. F. and MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81(3):607–636.

Farhi, E., Fraiberger, S., Gabaix, X., Ranciere, R., and Verdelhan, A. (2015). Crash Risk in Currency Markets. *Unpublished working paper*.

Fernández, C. and Steel, M. F. (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Feunou, B. and Tedongap, R. (2012). A Stochastic Volatility Model With Conditional Skewness. *Journal of Business & Economic Statistics*, 30(4):576–591.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. In Bernardo, J. M., Berger, J., Dawid, A. P., and Smith, J. F. M., editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press.

Ghalanos, A. (2014). racd: Autoregressive Conditional Density Models (R Package Version 1.0-5). *Unpublished manual*.

Han, F. and Westelius, N. J. (2019). Anatomy of Sudden Yen Appreciations. *IMF Working Paper. International Monetary Fund*.

Hansen, B. E. (1994). Autoregressive Conditional Density Estimation. *International Economic Review*, 35(3):705–730.

Harvey, C. R. and Siddique, A. (1999). Autoregressive Conditional Skewness. *Journal of Financial and Quantitative Analysis*, 34(4):465–487.

Hogben, D., Pinkham, R., and Wilk, M. (1961). The Moments of the Non-Central t-Distribution. *Biometrika*, 48(3/4):465–468.

Huisman, R., Koedijk, K. G., Kool, C. J. M., and Palm, F. (2001). Tail-Index Estimates in Small Samples. *Journal of Business & Economic Statistics*, 19(2):208–216.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). Continuous Univariate Distributions, Vol. 2, *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*.

Johnson, T. C. (2002). Volatility, Momentum, and Time-Varying Skewness in Foreign Exchange Returns. *Journal of Business & Economic Statistics*, 20(3):390–411.

Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699–1737.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65(3):361–393.

Kohler, M. (2010). Exchange rates during financial crises. *BIS Quarterly Review*.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust Econometric Inference for Stock Return Predictability. *The Review of Financial Studies*, 28(5):1506–1553.

Kroese, D. P., Taimre, T., and Botev, Z. I. (2013). *Handbook of Monte Carlo Methods*. John Wiley & Sons.

Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives*, 3(2):73–84.

Li, Y., Yu, J., and Zeng, T. (2020). Deviance information criterion for latent variable models and misspecified models. *Journal of Econometrics*, 216(2):450–493.

Magdalinos, T. and Phillips, P. C. (2009). Econometric Inference in the Vicinity of Unity. *Unpublished working paper*.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, 36:394–419.

McCauley, R. N. and McGuire, P. (2009). Dollar appreciation in 2008: safe haven, carry trades, dollar shortage and overhedging. *BIS Quarterly Review*.

McCausland, W. J., Miller, S., and Pelletier, D. (2011). Simulation smoothing for state–space models: A computational efficiency analysis. *Computational Statistics & Data Analysis*, 55(1):199–212.

Meese, R. A. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14(1-2):3–24.

Nakajima, J. (2013). Stochastic volatility model with regime-switching skewness in heavy-tailed errors for exchange rate returns. *Studies in Nonlinear Dynamics & Econometrics*, 17(5):499–520.

Nakajima, J. and Omori, Y. (2012). Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Students t-distribution. *Computational Statistics & Data Analysis*, 56(11):3690–3704.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449.

Rafferty, B. (2012). Currency Returns, Skewness and Crash Risk. *Unpublished working paper*.

Scott, R. C. and Horvath, P. A. (1980). On the Direction of Preference for Moments of Higher Order than the Variance. *The Journal of Finance*, 35(4):915–919.

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

Tanner, M. A. and Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Tsionas, E. G. (2002). Bayesian Inference in the Noncentral Student-t Model. *Journal of Computational and Graphical Statistics*, 11(1):208–221.

Tsiotas, G. (2012). On generalised asymmetric stochastic volatility models. *Computational Statistics & Data Analysis*, 56(1):151–172.

Yu, Y. and Meng, X.-L. (2011). To Center or Not to Center: That Is Not the Question–An Ancillarity–Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570.

Appendix A Moments of the noncentral t-distribution

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

$$\begin{aligned}
\mathbb{E}[X] &= c_{11}(\nu)\delta, && \text{if } \nu > 1,\\
\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] &= c_{22}(\nu)\delta^2 + c_{20}(\nu), && \text{if } \nu > 2,\\
\mathbb{E}\left[(X - \mathbb{E}[X])^3\right] &= c_{33}(\nu)\delta^3 + c_{31}(\nu)\delta, && \text{if } \nu > 3,\\
\mathbb{E}\left[(X - \mathbb{E}[X])^4\right] &= c_{44}(\nu)\delta^4 + c_{42}(\nu)\delta^2 + c_{40}, && \text{if } \nu > 4.
\end{aligned}$$

The functional forms of the coefficients are:

$$c_{11}(\nu) = \sqrt{\frac{1}{2}\nu}\frac{\Gamma\left[\frac{1}{2}(\nu-1)\right]}{\Gamma(\frac{1}{2}\nu)}, \qquad c_{22}(\nu) = \frac{\nu}{\nu-2} - c_{11}(\nu)^2, \qquad c_{20}(\nu) = \frac{\nu}{\nu-2},$$

$$c_{33}(\nu) = c_{11}(\nu)\left[\frac{\nu(7-2\nu)}{(\nu-2)(\nu-3)} + 2c_{11}(\nu)^2\right], \qquad c_{31}(\nu) = \frac{3\nu}{(\nu-2)(\nu-3)}c_{11}(\nu),$$

$$c_{44}(\nu) = \frac{\nu^2}{(\nu-2)(\nu-4)} - \frac{2\nu(5-\nu)c_{11}(\nu)^2}{(\nu-2)(\nu-3)} - 3c_{11}(\nu)^4,$$

$$c_{42}(\nu) = \frac{6\nu}{\nu-2}\left[\frac{\nu}{\nu-4} - \frac{(\nu-1)c_{11}(\nu)^2}{\nu-3}\right], \qquad c_{40}(\nu) = \frac{3\nu^2}{(\nu-2)(\nu-4)}.$$

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 ν (which changes over MCMC iterations) and δ_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

$$p(s_t = j|y_t, X_t, h_t, \delta_t, \nu, \beta) = \frac{1}{k_t}q_j(\nu, \delta_t)p_{\mathcal{N}}\left(\widetilde{y}_t; h_t + m_j(\nu, \delta_t), v_j^2(\nu, \delta_t)\right), \tag{B-1}$$

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

Hsiao (2014).[19]

Block 2: Sample the (log-)volatility h from $p(h|y, X, s, \delta, \nu, \beta, \mu_h, \phi_h, \sigma_h^2)$

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)

$$w_t = Z_t\kappa_t + e_t, \qquad e_t \sim \mathcal{N}(0, H_t), \tag{B-2}$$

$$\kappa_{t+1} = d_t + T_t\kappa_t + R_t\eta_t, \qquad \eta_t \sim \mathcal{N}(0, Q_t), \tag{B-3}$$

where w_t is an observed data point and κ_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 η_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

$$\underbrace{\widetilde{y}_t - m_{s_t}(\nu, \delta_t)}_{w_t} = \underbrace{\left[1\right]}_{Z_t}\underbrace{h_t}_{\kappa_t} + \underbrace{\epsilon_t}_{e_t}, \tag{B-4}$$

$$\underbrace{h_{t+1}}_{\kappa_{t+1}} = \underbrace{\left[\mu_h(1-\phi_h)\right]}_{d_t} + \underbrace{\left[\phi_h\right]}_{T_t}\underbrace{h_t}_{\kappa_t} + \underbrace{\left[1\right]}_{R_t}\underbrace{\eta_t^h}_{\eta_t}, \tag{B-5}$$

where $\widetilde{y}_t = \log((y_t - X_t\beta)^2 + c)$, $H_t = v_{s_t}^2(\nu, \delta_t)$ and $Q_t = \sigma_h^2$. As Equations (B-4) and (B-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 efficiently sample the unobserved (log-)volatilities h_t 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 λ from $p(\lambda|y, X, h, \delta, \nu, \beta)$

In sampling the latent state variable λ, this paper follows Tsionas (2002). In particular, the conditional distribution of each λ_t is

$$p(\lambda_t|y_t, X_t, h_t, \delta_t, \nu, \beta) \propto \lambda_t^{-(\nu+3)/2}\exp\left[-\frac{u_t^2/e^{h_t} + \nu}{2\lambda_t} + \delta_t(u_t/e^{h_t/2})\lambda_t^{-1/2}\right], \tag{B-6}$$

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, λ_t is conditionally inverse-gamma distributed and can be straightforwardly sampled as in e.g. Chan and Hsiao (2014). However, in the noncentral case

[19] See Algorithm 3.2. in Kroese et al. (2013) for a textbook treatment of the inverse-transform method.

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

$$p(w_t|y_t, X_t, h_t, \delta_t, \nu, \beta) \propto w_t^{\nu} \exp\left(-\frac{u_t^2/e^{h_t} + \nu}{2} w_t^2 + \frac{\delta_t u_t}{e^{h_t/2}} w_t\right). \tag{B-7}$$

This conditional distribution belongs to a family of distributions with kernel function

$$f(x) \propto x^{N-1} \exp(-(A/2)x^2 + Bx), \tag{B-8}$$

where $N = \nu + 1$, $A = u^2/e^h + \nu$ and $B = \delta u/e^{h/2}$.

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

$$A_t x^2 - B_t x - N = 0. \tag{B-9}$$

We then accept the candidate draw w_t^* with probability

$$R = \exp(r^* - r), \tag{B-10}$$

where $r^* = \log(f(x)/g(x))$ evaluated at w_t^* and $r = \log(f(x)/g(x))$ evaluated at x^*. Specifically,

$$r^* = -(A_t/2){w_t^*}^2 + (B_t + \theta^*)w_t^* - N \log(\theta^*), \tag{B-11}$$

$$r = -(A_t/2){x^*}^2 + (B_t + \theta^*)x^* - N \log(\theta^*). \tag{B-12}$$

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

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

Sampling the degrees of freedom ν 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 $\lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$ and the prior distribution $\nu \sim \mathcal{U}(0, \bar{\nu})$ as

$$\log p(\nu|\lambda) = \frac{T\nu}{2} \log(\nu/2) - T \log \Gamma(\nu/2) - (\nu/2 + 1) \sum_{t=1}^{T} \log \lambda_t - \frac{\nu}{2} \sum_{t=1}^{T} \lambda_t^{-1} + k, \tag{B-13}$$

for $0 < \nu < \bar{\nu}$ and k is a normalization constant. The first and second derivative of the log-density with respect to ν are then given by

$$\frac{d \log p(\nu|\lambda)}{d\nu} = \frac{T}{2} \log(\nu/2) + \frac{T}{2} - \frac{T}{2}\Psi(\nu/2) - \frac{1}{2} \sum_{t=1}^{T} \log \lambda_t - \frac{1}{2} \sum_{t=1}^{T} \lambda_t^{-1}, \tag{B-14}$$

$$\frac{d^2 \log p(\nu|\lambda)}{d\nu^2} = \frac{T}{2\nu} - \frac{T}{4}\Psi'(\nu/2), \tag{B-15}$$

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_ν, are obtained. Finally, an independence-chain Metropolis-Hastings step can be implemented with proposal distribution $\mathcal{N}(\hat{\nu}, K_\nu^{-1})$.

Block 5: Sample the latent noncentrality parameter δ 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 δ_t, we explore the following state space model

$$\underbrace{\widetilde{y}_t}_{w_t} = \underbrace{\left[\lambda_t^{1/2} - c_{11}(\nu)\right]}_{Z_t}\underbrace{\delta_t}_{\kappa_t} + \underbrace{\epsilon_t}_{e_t}, \tag{B-16}$$

$$\underbrace{\delta_{t+1}}_{\kappa_{t+1}} = \underbrace{\left[\mu_\delta(1-\phi_\delta)\right]}_{d_t} + \underbrace{\left[\phi_\delta\right]}_{T_t}\underbrace{\delta_t}_{\kappa_t} + \underbrace{\left[1\right]}_{R_t}\underbrace{\eta_t^\delta}_{\eta_t}, \tag{B-17}$$

where $\widetilde{y}_t = (y_t - X_t\beta)e^{-h_t/2}$ and with $H_t = \lambda_t$ and $Q_t = \sigma_\delta^2$. Note that the observation Equation (B-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 $\delta = (\delta_1, ..., \delta_T)$. If δ_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 (B-17).

Block 6: Sample the constant trend volatility μ_h from $p(\mu_h|h, \phi_h, \sigma_h^2)$ and the constant trend asymmetry μ_δ 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 μ_τ, where $\tau = (h, \delta)$, is

$$\mu_\tau|\tau, \phi_\tau, \sigma_\tau^2 \sim \mathcal{N}(\hat{\mu}_\tau, D_{\mu_\tau}), \tag{B-18}$$

with

$$D_{\mu_\tau} = (V_{\mu_\tau}^{-1} + X_{\mu_\tau}'\Sigma_\tau^{-1}X_{\mu_\tau})^{-1}, \tag{B-19}$$

$$\hat{\mu}_\tau = D_{\mu_\tau}(V_{\mu_\tau}^{-1}\mu_{\tau 0} + X_{\mu_\tau}'\Sigma_\tau^{-1}z_{\mu_\tau}), \tag{B-20}$$

where $X_{\mu_\tau} = (1, 1-\phi_\tau, ..., 1-\phi_\tau)'$, $z_{\mu_\tau} = (\tau_1, \tau_2 - \phi_\tau\tau_1, ..., \tau_T - \phi_\tau\tau_{T-1})'$ and $\Sigma_\tau = \text{diag}(\sigma_\tau^2/(1-\phi_\tau^2), \sigma_\tau^2, ..., \sigma_\tau^2)$. If δ_t is specified as a (driftless) random walk as in Section 4, sampling μ_δ is simply omitted.

Block 7: Sample the volatility AR(1) coefficient ϕ_h from $p(\phi_h|h, \mu_h, \sigma_h^2)$ and the asymmetry AR(1) coefficient ϕ_δ from $p(\phi_\delta|\delta, \mu_\delta, \sigma_\delta^2)$

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

$$p(\phi_\tau|\tau, \mu_\tau, \sigma_\tau^2) \propto p(\phi_\tau)g(\phi_\tau)\exp\left(-\frac{1}{2\sigma_\tau^2}\sum_{t=2}^{T}(\tau_t - \mu_\tau - \phi_\tau(\tau_{t-1} - \mu_\tau))^2\right), \tag{B-21}$$

with

$$g(\phi_\tau) = (1-\phi_\tau^2)^{1/2}\exp\left(-\frac{1}{2\sigma_\tau^2}(1-\phi_\tau^2)(\tau_1 - \mu_\tau)^2\right), \tag{B-22}$$

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 $\mathcal{N}(\hat{\phi}_\tau, D_{\phi_\tau})\mathbb{I}(|\phi_\tau| < 1)$ with

$$D_{\phi_\tau} = (V_{\phi_\tau}^{-1} + X_{\phi_\tau}' X_{\phi_\tau}/\sigma_\tau^2)^{-1}, \tag{B-23}$$

$$\hat{\phi}_\tau = D_{\phi_\tau}(V_{\phi_\tau}^{-1}\phi_{\tau 0} + X_{\phi_\tau}' z_{\phi_\tau}/\sigma_\tau^2), \tag{B-24}$$

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 ϕ_τ, a proposal ϕ_τ^* is accepted with probability $\min(1, g(\phi_\tau^*)/g(\phi_\tau))$. In case of rejection, the Markov chain remains at the current state ϕ_τ. If δ_t is specified as a (driftless) random walk as in Section 4, sampling ϕ_δ is simply omitted.

Block 8: Sample the shock variances σ_h^2 from $p(\sigma_h^2|h, \mu_h, \phi_h)$ and σ_δ^2 from $p(\sigma_\delta^2|\delta, \mu_\delta, \phi_\delta)$

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

$$\sigma_\tau^2|\tau, \mu_\tau, \phi_\tau \sim \mathcal{IG}(c_{\tau 0} + T/2, C_\tau), \tag{B-25}$$

where notation follows Chan and Hsiao (2014) and

$$C_\tau = C_{\tau 0} + \left[(1-\phi_\tau^2)(\tau_1 - \mu_\tau)^2 + \sum_{t=2}^{T}(\tau_t - \mu_\tau - \phi_\tau(\tau_{t-1} - \mu_\tau))^2\right]/2. \tag{B-26}$$

If δ_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 β 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 β

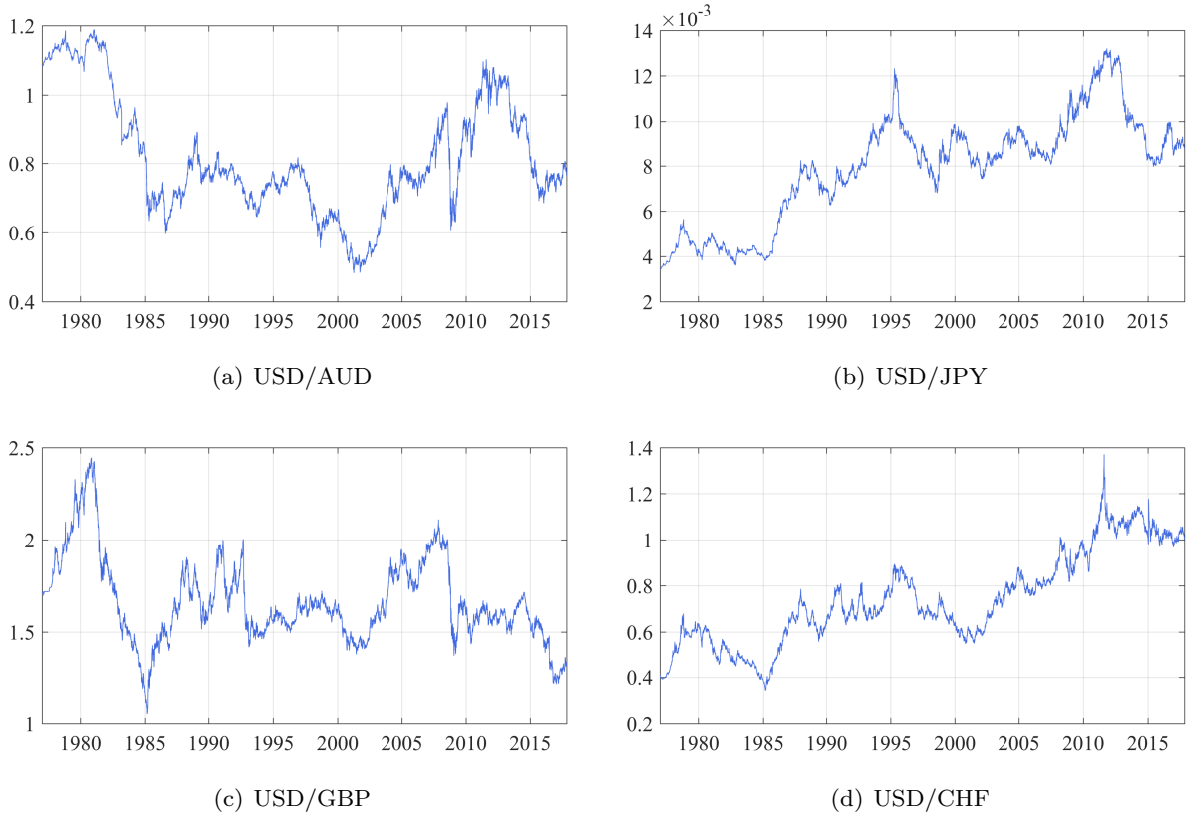
can be sampled as in Tsionas (2002). The conditional posterior distribution is

$$\beta|y, X, h, \delta, \lambda, \nu \sim \mathcal{N}\left([X'\Lambda^{-1}X]^{-1}X'\Lambda^{-1}(\widetilde{y} - \delta \odot e^{h/2} \odot \lambda^{1/2}), e^{h}[X'\Lambda^{-1}X]^{-1}\right), \tag{B-27}$$

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

Appendix C Data

Figure C-1: Nominal exchange rates



(a) USD/AUD

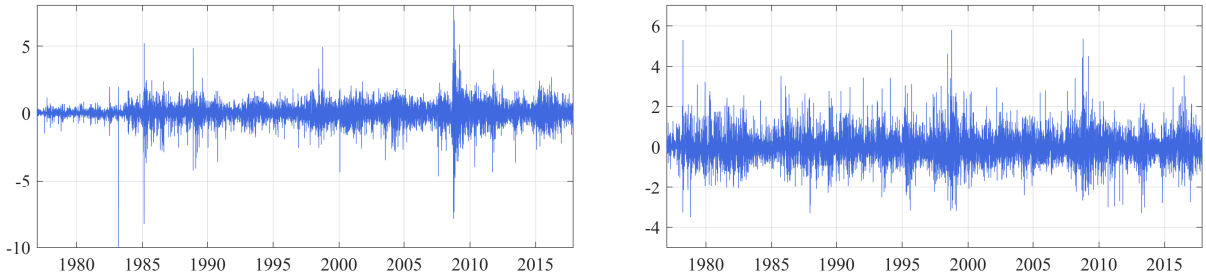
(b) USD/JPY

(c) USD/GBP

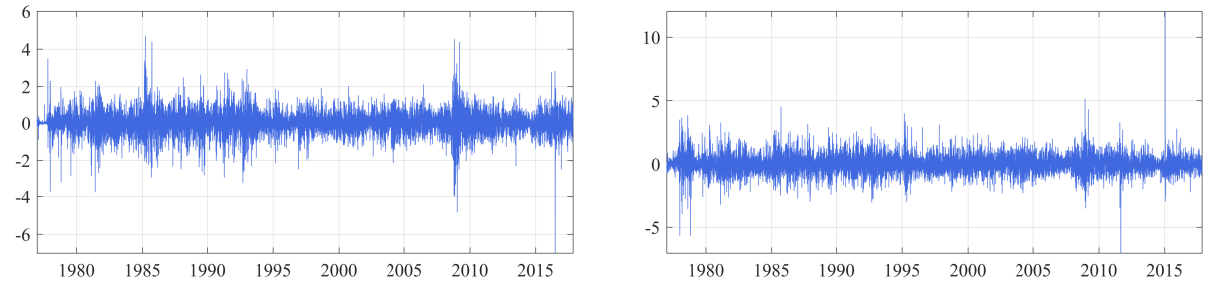
(d) USD/CHF

Note: The figure displays the four daily nominal exchange rates measured as USD per foreign currency unit over the period 01/01/1977–31/10/2017 ($T = 10,255$). Source: Board of Governors of the Federal Reserve System (US), retrieved from FRED, Federal Reserve Bank of St. Louis.

Figure C-2: Nominal exchange rate returns



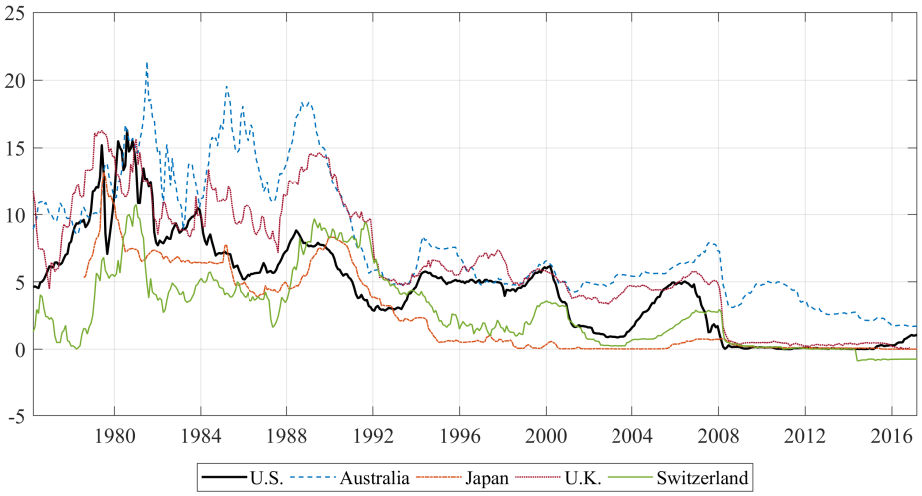
(a) USD/AUD (b) USD/JPY



(c) USD/GBP (d) USD/CHF

Note: The figure displays the daily nominal exchange rate returns used in the estimations. These are calculated as $y_t = (S_t - S_{t-1})/S_{t-1} \times 100$ where S_t is the nominal exchange rate.

Figure C-3: Short-term interest rates



Note: The figure displays the short-term (3-months) interest rates for the sample countries and the U.S. as the reference country. Source: Board of Governors of the Federal Reserve System (US), retrieved from FRED, Federal Reserve Bank of St. Louis.

Appendix D Additional results

Monte Carlo simulation with alternative prior configurations

The goal of the Monte Carlo experiments in Section 3 is to show that the developed MCMC algorithm and its building blocks, such as the extended mixture approximation described in Section 2.3, succeed in recovering the true volatility and skewness dynamics. While the priors used in these simulations are relatively uninformative, for most parameters they are centered around the true values. This applies to the intercepts and the AR(1) coefficients and, to a lesser extent, to the shock variances. Priors centered around the true values have been used to avoid biasing the results through the prior configuration.

Table D-1: Monte Carlo simulation results: Alternative prior specifications

Sample size	Parameter	Mean	SE	2.5%	97.5%	Bias	CD	IF
T = 1,000	μ_h	−0.06	0.30	−0.63	0.54	−0.06	0.46	10.44
	ϕ_h	0.97	0.04	0.89	0.99	−0.02	0.48	60.06
	σ_h^2	0.02	0.00	0.01	0.02	0.01	0.49	76.96
	μ_δ	−0.12	0.50	−1.23	0.64	−0.12	0.39	352.04
	ϕ_δ	0.43	0.23	0.16	0.99	−0.56	0.38	687.27
	σ_δ^2	0.02	0.00	0.02	0.02	0.01	0.49	109.48
	ν	6.29	2.71	4.01	13.13	1.29	0.44	100.29
T = 5,000	μ_h	0.00	0.14	−0.28	0.25	0.00	0.42	7.66
	ϕ_h	0.99	0.00	0.98	0.99	0.00	0.47	63.98
	σ_h^2	0.01	0.00	0.01	0.02	0.00	0.47	152.72
	μ_δ	−0.02	0.16	−0.32	0.29	−0.02	0.46	77.39
	ϕ_δ	0.89	0.21	0.23	0.99	−0.10	0.37	511.61
	σ_δ^2	0.02	0.00	0.01	0.02	0.01	0.43	399.74
	ν	5.25	0.49	4.46	6.32	0.25	0.45	57.36
T = 10,000	μ_h	0.01	0.10	−0.19	0.19	0.01	0.46	3.72
	ϕ_h	0.99	0.00	0.98	0.99	0.00	0.46	79.22
	σ_h^2	0.01	0.00	0.01	0.01	0.00	0.46	184.81
	μ_δ	0.00	0.11	−0.22	0.22	0.00	0.48	8.56
	ϕ_δ	0.98	0.05	0.97	0.99	−0.01	0.41	247.85
	σ_δ^2	0.02	0.00	0.01	0.02	0.01	0.41	552.50
	ν	5.16	0.31	4.60	5.78	0.16	0.48	40.17

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

In addition, it is interesting to see how the model and the estimation algorithm behave if the prior distributions are not centered around the true values and even more flat. To this end, the simulation reported in Table 3 has been re-done using the same DGP as before but the following prior configurations: $\mu_h \sim \mathcal{N}(-5, 10)$, $\mu_\delta \sim \mathcal{N}(-5, 10)$, $\phi_h \sim \mathcal{N}(0.95, 1)$, $\phi_\delta \sim \mathcal{N}(0.95, 1)$, $\sigma_h^2 \sim$

$\mathcal{IG}(2.5, 0.08)$ and $\sigma_\delta^2 \sim \mathcal{IG}(2.5, 0.08)$. The prior distributions of the shock variances imply a prior expectation of 0.05 and a prior standard deviation of 0.08 (Chan and Grant, 2016a). The prior specification of the degrees of freedom parameter remains unchanged at $\nu \sim \mathcal{U}(0, 50)$.

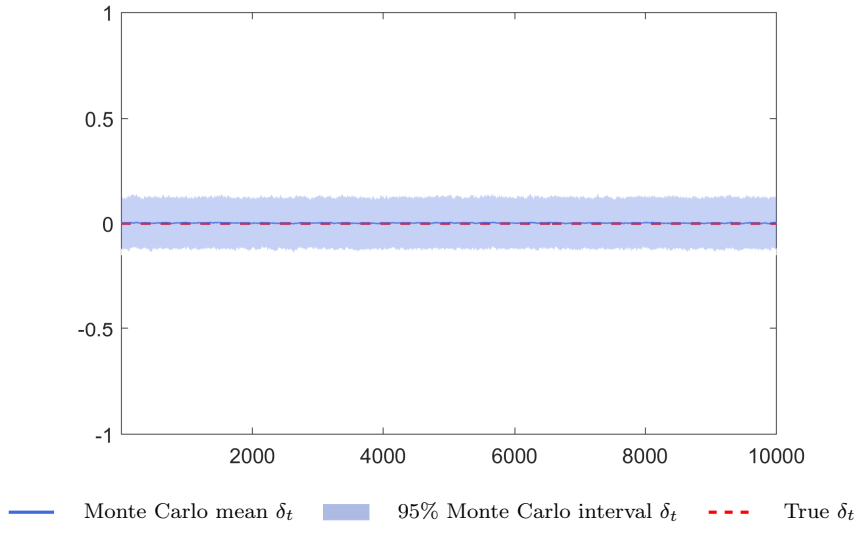
Table D-1 presents the results of this prior sensitivity analysis. Overall, the results remain relatively similar compared to the baseline simulation. However, when using a completely flat prior for the AR(1) parameter of the unobserved asymmetry process, ϕ_δ, it cannot be identified anymore in a relatively small sample of $T = 1,000$. The precision of the estimation improves quickly though with increasing sample size and for $T = 10,000$, the results are comparable to the baseline. Interestingly, these alternative prior configurations result, on average, in significantly lower inefficiency factors. In summary, this sensitivity analysis shows that the estimation algorithm performs well even when the prior distributions are not centered around the true values and, in addition, are even more flat.

Monte Carlo simulation with symmetric DGP

This section takes a closer look at the performance of the proposed SVSS model in a situation where the underlying DGP is symmetric, i.e. skewness does not exist. To this end, the small illustrative example presented at the end of Section 3 is extended to a large scale Monte Carlo simulation. Specifically, 1,000 datasets are simulated ($T = 10,000$) from the restricted (symmetric) SVSS model, i.e. the true values in the DGP are: $\mu_h = 0$, $\phi_h = 0.99$, $\sigma_h^2 = 0.01$, $\mu_\delta = 0$, $\phi_\delta = 0.99$, $\sigma_\delta^2 = 0$, $\nu = 5$. Afterwards, the unrestricted SVSS model is estimated using each of these datasets and the posterior mean series of the asymmetry parameter δ are retained.

Figure D-1 presents the average estimated asymmetry parameter δ along with the 95% density interval. The results clearly point towards a good performance of the model even in a case where the SVSS model is overparametrized since the underlying DGP is symmetric. The Monte Carlo mean of δ_t is very close to 0 $\forall\, t = 1, ..., T$ and the surrounding 95% interval is relatively narrow.

Figure D-1: Monte Carlo results for δ_t and symmetric DGP



GARCH with conditional skewness vs. SVSS

This section compares the results obtained from the stochastic volatility - stochastic skewness model with a competitor model from the GARCH family that features deterministic time variation in the higher moments. As discussed in Section 1, higher moment dynamics have been extensively modelled in the literature on generalized autoregressive conditional heteroscedasticity (GARCH) models. Hansen (1994) has developed the so-called autoregressive conditional density (ACD) class of models, where not only the latent variance process but also the unobserved scale and shape parameters of the standardized shock distribution are assumed to follow GARCH dynamics.

For the purpose of this section, we specify the following ACD model, where the conditional mean is again modelled as an AR(4) process, the conditional variance has GARCH(1,1) dynamics and the standardized shocks follow a skewed-t distribution (Fernández and Steel, 1998) where the unconstrained asymmetry parameter exhibits quadratic dynamics:

$$y_t = \beta_0 + \sum_{l=1}^{4} \beta_l y_{t-l} + \varepsilon_t, \tag{D-1}$$

$$\varepsilon_t = \sigma_t z_t, \qquad z_t \sim \mathcal{SKT}(0, 1, \delta_t, \nu), \tag{D-2}$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \sigma_{t-1}^2, \tag{D-3}$$

$$\delta_t = L + \frac{U - L}{1 + \exp(-\bar{\delta}_t)}, \tag{D-4}$$

$$\bar{\delta}_t = c_0 + c_1 z_{t-1} + c_2 z_{t-1}^2 + c_3 \bar{\delta}_{t-1}. \tag{D-5}$$

The model specified in Equations (D-1)-(D-5) can be considered a simplified version of the original model in Hansen (1994) where the degrees of freedom parameter ν is assumed constant to allow for a better comparison with the results from the SVSS model. The unconstrained asymmetry parameter $\bar{\delta}_t$ is logistically transformed to ensure that its constrained counterpart δ_t lies within a sensible interval $\delta_t \in [L, U]$. To estimate the outlined ACD model, we use the R package *racd* of Ghalanos (2014).[20] When comparing this model to the SVSS model, it should be noted that an exact comparison remains difficult since not only the dynamics of skewness and volatility differ but also the underlying distributional assumption about the error term. While the SVSS model employs the noncentral t-distribution, the default choice for the ACD model is the skewed t-distribution of Fernández and Steel (1998). Keeping this in mind, the results presented in the following still provide many interesting insights into the performance of both models.

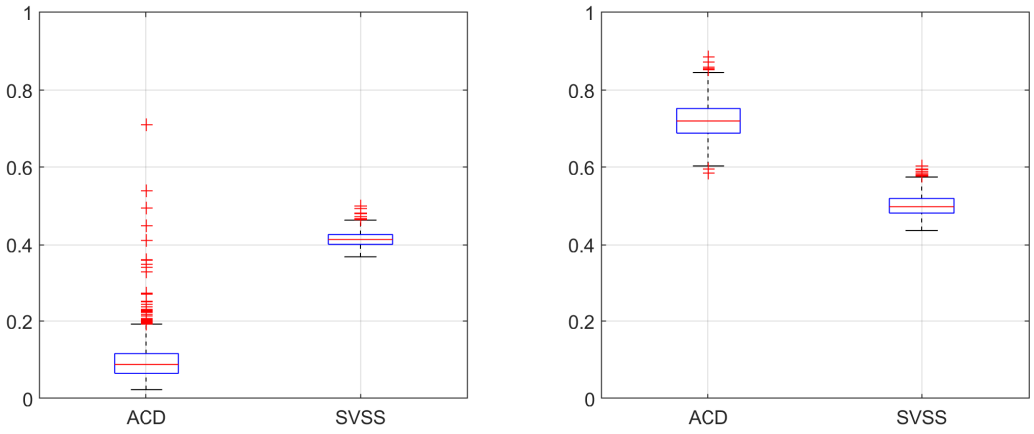
We start by comparing both models in a simulation exercise where samples are generated from the models and both the ACD and the SVSS model are estimated on these samples resulting in four possible combinations of true and estimated model.[21] To measure model performance, we compute for each Monte Carlo sample the average absolute deviation (over time) of the estimated

[20] For the non-linear optimization problem the *ucminf* solver is used with tolerance levels $grtol = 10^{-4}$ and $xtol = 10^{-6}$ and the maximum number of function evaluations is set to $maxeval = 500$.

[21] The simulation exercise follows the set-up of Section 3. The true parameter values for the SVSS model are as in Table 3. Moreover, samples from the ACD model in Equations (D-1)-(D-5) are generated using the following parameter values: $\omega = 0.005, \alpha_1 = 0.04, \alpha_2 = 0.95, c_0 = 0, c_1 = 0.2, c_2 = -0.005, c_3 = 0.9, \nu = 5$. The conditional mean specification is dropped for the simulation exercise, i.e. $\beta_l = 0 \; \forall \; l = 0, ..., 4$. We set $L = 0.5$ and $U = 2$, which, given $\nu = 5$, ensures that conditional skewness lies approximately in the interval $[-2, 2]$.

skewness series from the true skewness series. Figure D-2 presents boxplots of these statistics for both DGPs (ACD and SVSS) and both estimated models.

Figure D-2: Monte Carlo simulation: ACD vs. SVSS



(a) DGP: ACD model

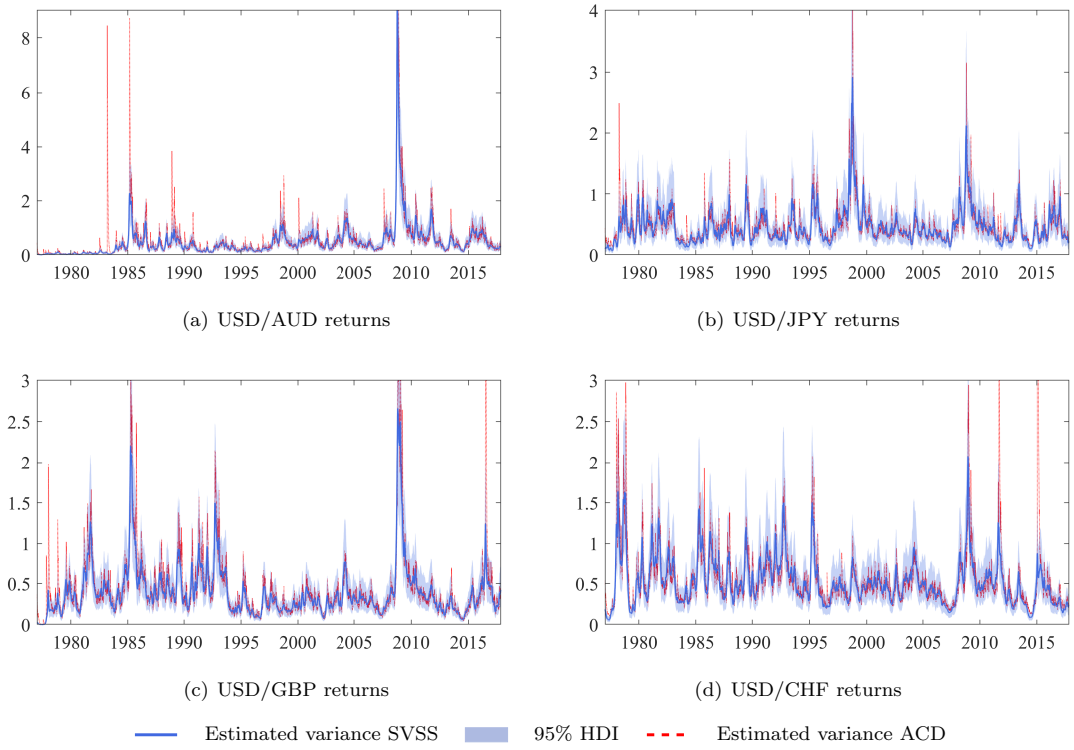
(b) DGP: SVSS model

Note: The figures display boxplots of the average absolute deviations of the estimated skewness series from the true skewness series for the ACD and the SVSS model (for the 1,000 Monte Carlo samples).

When inspecting the plots, a few conclusions can be drawn. First, as expected the model that corresponds to the DGP estimates skewness on average more precisely. Second, the most precise results are obtained in the case where the ACD model is estimated on samples generated from the same model. However, the existence of a few large outliers suggests that the ACD model can occasionally perform poorly. Third, while the SVSS model clearly outperforms the ACD model for samples from the SVSS model, overall stochastic skewness seems to be more difficult to capture than deterministic skewness as indicated by the higher levels of the absolute deviation measures. Finally, in general both models have difficulties capturing skewness dynamics generated by the other model. While this cannot be seen directly from the presented boxplots, the estimated skewness series of both models when using data generated from the other model, are mostly relatively flat and thus reflect the average level of skewness in the data while not capturing the dynamics over time appropriately.

We now turn to comparing both models when estimating them using the four exchange rate return series. In order to get an idea of how the estimated latent variables relate, Figures D-3 and D-4 contrast the estimated variance and skewness series obtained from both models. Even though the estimated GARCH variances show sometimes larger volatility peaks, overall the evolution of the estimated time-varying variances across both models is relatively similar. In contrast, the estimated skewness series differ greatly. Naturally, the different parametric assumptions underlying both models lead to very different time series patterns of the estimated skewness series. While the SVSS model produces a rather smooth and slowly evolving skewness series, the ACD model yields a highly erratic estimated asymmetry series.

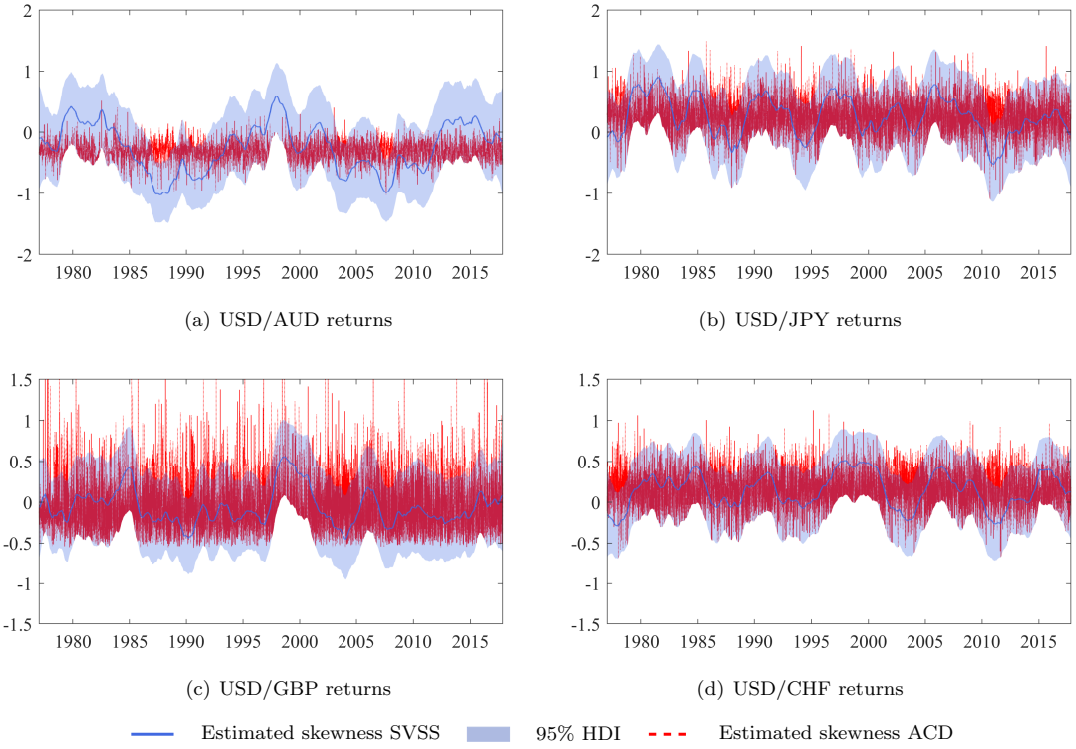
Figure D-3: Estimated variance series: ACD vs. SVSS



(a) USD/AUD returns (b) USD/JPY returns

(c) USD/GBP returns (d) USD/CHF returns

Figure D-4: Estimated skewness series: ACD vs. SVSS



(a) USD/AUD returns (b) USD/JPY returns

(c) USD/GBP returns (d) USD/CHF returns

To assess its forecasting performance, the exercise presented in Section 4.4 has been done also for the ACD model.[22] The results of the Value-at-Risk and expected shortfall forecasts are presented in Tables D-2 and D-3. The ACD model does very well when used for VaR forecasting and its performance is overall comparable with the SVSS model. When used to forecast expected shortfall, the ACD model again performs very well and, on average, performs even better then the SVSS model. However, the SVSS model has advantages when forecasting expected shortfall in the left tail of all four currency return distributions thus maintaining a distinct gain when forecasting the expected size of potential crashes.

Table D-2: ACD model: LR tests for unconditional coverage, independence and conditional coverage of the VaR violations

α	\$/AUD	\$/JPY	\$/GBP	\$/CHF	\$/AUD	\$/JPY	\$/GBP	\$/CHF
	(a) Number of VaR violations				(b) p-values for LR-UC test			
0.5%	11	11	13	9	0.46	0.46	0.87	0.18
1%	19	26	23	17	0.09	0.81	0.40	**0.03**
5%	153	135	165	122	0.14	0.92	**0.01**	0.21
95%	143	130	125	122	0.55	0.59	0.32	0.21
99%	26	20	18	25	0.81	0.14	0.06	0.66
99.5%	10	12	7	14	0.30	0.66	**0.05**	0.92
	(c) p-values for LR-IND test				(d) p-values for LR-CC test			
0.5%	0.77	0.77	0.72	0.81	0.73	0.73	0.93	0.40
1%	0.61	0.48	0.53	0.64	0.22	0.76	0.58	0.10
5%	0.07	0.77	0.15	0.23	0.06	0.95	**0.02**	0.22
95%	0.15	0.32	0.43	0.83	0.29	0.53	0.45	0.44
99%	0.49	0.59	0.11	0.23	0.76	0.30	**0.05**	0.44
99.5%	0.79	0.74	0.85	0.06	0.57	0.86	0.14	0.17

Note: This table contains the number of VaR violations based on the out-of-sample forecasting exercise for the ACD model and each exchange rate (a) as well as the p-values corresponding to the likelihood ratio (LR) test for unconditonal coverage following Kupiec (1995) (b) and the LR tests for independence of the violations (c) and conditional coverage (d) following Christoffersen (1998). Bold numbers indicate p-values < 0.05.

In summary, this short section provides interesting insights into the relative performance of the ACD and the SVSS model. Nevertheless, a broader and more detailed comparison of stochastic skewness models and deterministic skewness models would certainly be fruitful but is beyond the scope of this paper.

[22] We focus here on comparing the out-of-sample performance of both models as an in-sample comparison is not straightforward. This is because standard measures for model comparison are not available for the SVSS model while the conditional Deviance Information Criterion (DIC) used to compare the SV models in Section 4.2 is not readily available for the ACD model.

Table D-3: ACD model: precision of expected shortfall forecasts

α	USD/AUD	USD/JPY	USD/GBP	USD/CHF
0.5%	**0.065**	**0.219**	0.256	**0.400**
1%	**0.094**	**0.051**	**0.164**	**0.205**
5%	0.056	0.015	**0.039**	0.013
95%	0.050	0.004	0.030	0.070
99%	0.184	**0.084**	0.059	0.486
99.5%	0.288	0.048	0.073	0.892

Note: This table contains the expected shortfall accuracy measure of Embrechts et al. (2005). Bold numbers indicate a better performance of the SVSS model compared to the ACD model.

Subsample analysis

Since the sample of exchange rate returns used for the baseline estimations spans more than forty years, the aspect of parameter stability over this long period needs to be discussed. While the SVSS model allows the conditional variance and skewness to vary over time, the parameters governing the unobserved states h and δ as well as the degrees of freedom parameter ν, are assumed to be constant over the sample period.

Table D-4: Parameter estimates of the SVSS model over subsamples

		USD/AUD		USD/JPY		USD/GBP		USD/CHF	
	Coef.	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1977–1989 (T = 3,257)	μ_h	−2.586	0.460	−1.568	0.121	−1.775	0.675	−0.987	0.175
	ϕ_h	0.992	0.003	0.959	0.015	0.995	0.002	0.983	0.005
	σ_h^2	0.027	0.007	0.055	0.024	0.018	0.004	0.021	0.005
	ν	4.890	0.480	6.769	3.247	5.690	0.574	15.478	6.977
	σ_δ^2	0.004	0.002	0.005	0.002	0.004	0.002	0.007	0.004
1990–2003 (T = 3,520)	μ_h	−1.592	0.200	−1.293	0.142	−1.565	0.144	−0.951	0.103
	ϕ_h	0.992	0.003	0.987	0.005	0.983	0.006	0.983	0.006
	σ_h^2	0.006	0.002	0.008	0.003	0.016	0.005	0.007	0.002
	ν	7.367	1.010	7.049	0.907	15.664	6.932	24.341	9.472
	σ_δ^2	0.005	0.002	0.004	0.001	0.007	0.004	0.008	0.004
2004–2017 (T = 3,478)	μ_h	−0.902	0.267	−1.396	0.152	−1.328	0.321	−1.269	0.286
	ϕ_h	0.993	0.003	0.987	0.004	0.995	0.002	0.994	0.002
	σ_h^2	0.008	0.002	0.010	0.003	0.005	0.001	0.005	0.001
	ν	30.915	9.183	7.904	1.063	16.297	4.253	8.379	1.067
	σ_δ^2	0.008	0.004	0.004	0.001	0.007	0.004	0.005	0.002

Note: All estimations include a conditional mean specification as in Equation (14).

In case this assumption does not hold, the estimates from the full sample analysis either

reflect a time average or the unobserved states could absorb what are actually dynamics in the underlying parameters. In order to shed light on these issues, this appendix presents the results of a small subsample analysis. Table D-4 contains the posterior means and standard deviations of the parameters when estimating the SVSS model over three non-overlapping subsamples of similar size. The results can be summarized as follows. First, with a few exceptions the parameters governing the unobserved (log-)volatility process h and the asymmetry process δ seem to be relatively stable across the three subsamples considered. This certainly holds for the persistence parameter, ϕ_h, while the variance parameter of volatility shocks, σ_h^2, is lower in the second and third subsample. Most notably, the degrees of freedom parameter, ν, shows some fluctuation, especially in the cases of the USD/AUD and the USD/CHF returns. The forecasting exercise presented in Section 4.4 partly accounts for potential instabilities in the underlying parameters by re-estimating the candidate models in each period of the out-of-sample period. While one could in principle also allow for time-variation in these (currently fixed) parameters, the model complexity would significantly increase and the effect on, for example, forecasting performance, would be highly uncertain ('bias-variance trade-off').