Stochastic volatility models with skewness selection

Igor Martins[*]and Hedibert Freitas Lopes [†]

November 2023

Abstract

This paper expands traditional stochastic volatility models by allowing for time-varying skewness without imposing it. While dynamic asymmetry may capture the likely direction of future asset returns, it comes at the risk of leading to overparameterization. Our proposed approach mitigates this concern by leveraging sparsity-inducing priors to automatically selects the skewness parameter as being dynamic, static or zero in a data-driven framework. We consider two empirical applications. First, in a bond yield application, dynamic skewness captures interest rate cycles of monetary easing and tightening being partially explained by central banks' mandates. In an currency modeling framework, our model indicates no skewness in the carry factor after accounting for stochastic volatility which supports the idea of carry crashes being the result of volatility surges instead of dynamic skewness.

Key words: Stochastic Volatility; Sparsity; Skewness.

1 Introduction

Accurate representation of asset returns is one of the key topics in finance. Based on the theoretical results from Markowitz [1952] and Hansen and Richard [1987], standard approaches to asset pricing have largely focused on the first and second moments. Stochastic volatility (SV) models discussed, for example, on Jacquier et al. [2002], are one of the cornerstone models in modern financial econometrics. In its simplest form SV models asset returns via normal distribution with persistent volatility and a mean that is either constant or a linear function of explanatory variables. Such models capture the first two moments of asset returns in a simple and elegant matter being supported empirically and theoretically as discussed in Shephard [2005].

While we acknowledge the importance of the first two moments in asset pricing, we also notice the potential benefits of including skewness when modeling returns. Due to its ability of capturing the likely direction of returns,

---

[*]Insper. igorfbm@al.insper.edu.br
[†]Insper. hedibertfl@insper.edu.br

models with time-varying skewness may be more suitable to forecast periods with a higher concentration of same signal returns leading to better detection of periods of both overperformance and underperformance. Bianchi et al. [2022] is one key example of the empirical benefits of adding such feature for modeling cross-sectional stock momentum. While the momentum factor is known for delivering good mean-variance compensation, it is also subject to long period of negative performance. By capturing such prolonged periods of likely negative returns via dynamic skewness, Bianchi et al. [2022] improves the performance of the stock momentum factor upon traditional approaches which neglect skewness.

However, including dynamic skewness in traditional financial econometric models is not a free-lunch. While allowing for asymmetry may lead to a better representation of some financial time series, it may not be a vital feature and its inclusion risks overparameterization. Therefore, we wish to include dynamic skewness only when required by the data and remove such feature if is not necessary.

This paper expand stochastic volatility models by allowing dynamic skewness without having to impose it. We replace the traditional hypothesis of gaussian errors in favor of a skew-normal distribution. Such change preserves the usual features for the first two moments of SV models but allows for dynamic skewness. Since the inclusion of time-varying asymmetry may not always be necessary, we consider a sparsity - inducing scheme for its parameters. In particular, we consider a random-walk evolution for the asymmetry. When the standard deviation of the dynamic process is shrunk to zero, our model results in a SV with constant skewness. If, additionally, the level of the skewness is shrunk to zero, we recover a traditional SV model. By combining prior information with the likelihood of the model, our proposed approach automatically chooses between dynamic, static or no skewness.

We consider two empirical applications. We obtain three main results on a bond yield application for Brazil and the US. First, our proposed model indicates that bond yield changes for both countries are better represented by including time-varying skewness with out-of-sample improvements for forecasting the direction of future yields. Second, the recovered skewness is associated with interest rate cycles of monetary easing and tightening. Third, inflation and unemployment partially explain the recovered skewness linking it to central banks' mandates. We also model the carry factor for currency returns. Our model indicates not only the lack of dynamics but also no skewness at all for it after accounting for volatility. Therefore, similar to the crash mitigation via volatility scaling proposed by Barroso and Santa-Clara [2015], the carry factor also have its crashes mitigated once dynamic volatility is taken into account without requiring the inclusion skewness.

This paper intersects and contributes to multiple areas. First, it contributes to the stochastic volatility literature by expanding the static skewness model of Nakajima and Omori [2012] to the dynamic case while also expanding the sparsity - inducing scheme of Nakajima [2020] by allowing dynamic skewness to be shrunk towards the static case. Second, it contributes to the toolbox of methods for recovering dynamic skewness. Trolle and Schwartz [2014] rely on option data while Rafferty [2012] uses rolling windows. Both approaches have limitations. While theoretically sound, option based approaches require tradeable options which large collection of strikes with high

liquidity and continuous expiration dates. Such requirements are hard to meet in several applications, specially when for emerging markets such as in our Brazilian bond applications. Rolling-window based approaches artificially input dynamics into skewness by changing the sample period by period. Such change comes at the cost of outliers playing a large role in the estimation in addition to a trade-off based on window-size, which will affect the precision of the estimate, and the speed in which the dynamics change. Third, it contributes to the interest rate literature by showing that inflation and unemployment partially explain skewness and linking it to central banks' mandates expanding the traditional mean and variance analysis of papers such as Litterman and Scheinkman [1991] and Joslin [2018]. Fourthly, it contributes to the debate of whether the carry factor presents dynamic skewness after accounting for heteroskedasticity, discussed Burnside et al. [2011] and Jurek [2014], by claiming that skewness is unlikely to be dynamic and, in fact, is more likely to be zero after accounting for SV effects.

The paper is organized as follows. It starts by describing traditional SV models and moves on to our proposal with time-varying skewness on Section 2. Section3 discusses the sparsity-inducing framework. Section 4 shows our HMC approach to simulate from the joint posterior. Section 5 presents the bond yield forecasting application while Section 6 shows the carry factor application. Section 7 concludes.

2   SV model with time-varying skewness

A vanilla stochastic volatility model is given by Equations (1) - (5). Equation (1) represent asset returns with mean 0 and dynamic volatility $exp(h_t/2)$. Equation (2) indicates persistent log-volatility with level μ and persistence ϕ with initial values given by the stationary distribution shown in Equation (3). In its simplest version both the measure and state equations have Gaussian errors as represented in Equations (4) and (5).

$$y_t = exp\left(\frac{h_t}{2}\right)\varepsilon_t \tag{1}$$

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \sigma_h\eta_t \tag{2}$$

$$h_0 \sim N\left(\mu_h, \frac{\sigma_h^2}{1 - \phi_h^2}\right) \tag{3}$$

$$\varepsilon_t \sim N(0,1) \tag{4}$$

$$\eta_t \sim N(0,1) \tag{5}$$

In order to introduce asymmetry into the model, we replace the normal variable ε_t in Equation (4) by a skew-normal random variable z_t. We consider the skew-normal representation of Azzalini [1985] and Azzalini [2013] shown in Equation (6) where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution function, respectively. λ controls the degree of asymmetry in the distribution as show in Figure (1). In particular, for $\lambda = 0$, the skew-normal reduces to a standard normal distribution. Also, we may add location ξ and scale ω parameters to the

skew-normal denoting $X = \xi + \omega z$ by $X \sim SN(\xi, \omega, \lambda)$ which has its distribution presented in Equation (7). Appendix A describes how ξ, λ and λ relate to mean, variance and skewness. Therefore, our model can be viewed as an expansion of the stochastic volatility models with static skewness proposed by Nakajima and Omori [2012], Nakajima [2020] and Li and Scharth [2020].

$$\text{If } Z \sim SN(\lambda), \text{ then } p(z; \lambda) = 2\phi(z)\Phi(\lambda z) \tag{6}$$

$$\text{If } X \sim SN(\xi, \omega, \lambda), \text{ then } p(x; \xi, \omega, \lambda) = 2\frac{1}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\lambda\frac{x-\xi}{\omega}\right) \tag{7}$$

Figure 1: Skew-normal densities with $\xi = 0$ and $\omega = 1$ and λ equal to -2, 0, and 1 for the red, black and blue lines, respectively. If $\lambda < 0$, the distribution is skewed to the left. If $\lambda = 0$, SN(0) reduces to the standard Gaussian distribution. Finally, if $\lambda > 0$ the distribution is skewed to the right.

As shown in Bianchi et al. [2022] and Carr and Wu [2007], its plausible that asset returns have time-varying skewness. Thus, we allow for this possibility by replacing the static λ for λ_t which evolves according to a random-walk starting at λ_0 as represented by Equations (8) to (9). Therefore, we change our observation equation by allowing both volatility and skewness to be time-varying changing the observation equation to Equation (11).

$$z_t \sim SN(\lambda_t) \tag{8}$$

$$\lambda_t = \lambda_{t-1} + \sigma_\lambda \eta_{\lambda,t} \text{ with } \eta_{\lambda,t} \sim N(0,1) \tag{9}$$

$$\lambda_0 = \alpha_0 \tag{10}$$

$$y_t = exp\left(\frac{h_t}{2}\right)z_t \tag{11}$$

3 Sparsity-inducing approach

The inclusion of asymmetry may lead to a better representation of time series specially by capturing periods with a higher mass of returns with the same signal, as shown, for example, in Bianchi et al. [2022], Azzalini [2013], and Rachev et al. [2005]. However, it may not always be necessary since it requires the estimation of additional parameters. To mitigate this overparametrization risk, we aim to include time-varying skewness only when required by the data. While one can estimate multiple models varying the skewness specification and then performing selection via Bayes factor, we propose a sparse-inducing method which conveniently performs model selection without requiring estimating multiples models while also avoiding Bayes factor completely.

Equation (9) drives the dynamics of the asymmetry parameter. If σ_λ is close enough to zero, then λ_t is static and assume the value of its initial point λ_0. Additionally, if the initial point is also zero, then the model reduces to the vanilla SV model. Therefore, if we induce sparsity for σ_λ and α_0 we can achieve all 3 cases of interest.

In the non-Bayesian literature, shrinkage is based on maximizing the likelihood of a model subject to a penalty function with the LASSO of Tibshirani [1996] being the most commonly employed approach. From a Bayesian point of view, shrinkage problems can be represented as a penalization to the log-likelihood via log-prior. In fact, the posterior mode of a linear model with Double Exponential prior with location 0 and scale $2/\psi$ is equal to the point estimate of the LASSO with penalty ψ as shown in Park and Casella [2008].

Thus, in order to shrink σ_λ and λ_0 towards zero, we consider both having Double Exponential priors shown in Equation (12) and (13) similarly to the approach of Belmonte et al. [2014] in the context of dynamic regression models. Priors such as the ones discussed in Lopes et al. [2022] and Bitto and Frühwirth-Schnatter [2019] also are reasonable approaches to induce sparsity on time-varying parameter models.

$$\sigma_\lambda \sim \text{DoubleExponential}(0, 1/\kappa_{\sigma_\lambda}) \text{ with } \kappa_{\sigma_\lambda} \sim Gamma(a, b) \tag{12}$$

$$\alpha_\lambda \sim \text{DoubleExponential}(0, 1/\kappa_\alpha) \text{ with } \kappa_\alpha \sim Gamma(c, d) \tag{13}$$

While, to the best of our knowledge, we are the first to induce sparsity on the dynamic skewness framework, Nakajima [2020] have used the spike and slab prior of George and McCulloch [1993] to produce a data-driven framework which obtains the posterior probability of the presence of static skewness.

4 Remaining priors and posterior inference

Our modeling approach leads to the following unknown quantities $\Theta = \{\mu_h, \phi_h, \sigma_h, \alpha_0, \sigma_\lambda, \{h_t\}, \{\lambda_t\}\}$. We aim to recover the joint posterior distribution $p(\Theta|y)$ which by Bayes' rule can be computed as follows

$$p(\Theta|y) = \frac{p(y|\Theta)p(\Theta)}{\int p(y|\Theta)p(\Theta)d\Theta}$$

$p(y|\Theta)$ is the likelihood component being characterized by our proposed sampling model. We consider independent components for each member of $p(\Theta)$. $p(\mu_h)$ and $p(\phi_h)$ follow Gaussian distributions, $p(\sigma_h^2)$ has inverse gamma distribution while $p(\alpha_0)$ and $p(\sigma_\lambda)$ have double exponential distribution as discussed previously. The exact values for the prior parameters are described in Appendix B and Appendix C.

While we can provide a full description of $p(y|\Theta)$ and $p(\Theta)$, the problem of recovering $p(\Theta|y)$ is not analytically tractable. Thus, we resort to simulation methods to sample from $p(\Theta|y)$ and then compute summary posterior statistics to characterize our results. This paper uses a particular Markov Chain Monte Carlo (MCMC) method known as Hamiltonian Monte Carlo (HMC) instead of the more traditional Random - Walk Metropolis-Hastings (RWMH) algorithm to sample from $p(\Theta|y)$.

As discussed in Gamerman and Lopes [2006], the RWMH algorithm is a type o MCMC algorithm which generates a sequence of values Θ which approximates $p(\Theta|y)$. Each iteration i generates $\Theta^{(i)}$ which is defined in part by $q(\Theta^{prop}|\Theta^{i-1})$ where Θ^{prop} is a proposal for the next value in the chain and Θ^{i-1} represents the value of Θ on the previous iteration. The name RWMH is due to the proposal being generate as a random - walk from the previously sampled Θ. Each proposed value for Θ^i may be accepted or rejected based on Equation (14) where acc represents the acceptance ratio. While there are little restrictions for the use of RWMH algorithms, its limitations come from the computational side since the acceptance ratio is usually low requiring a large amount of iterations.

$$acc = min\left(1, \frac{f(\Theta^{prop})q(\Theta^{t-1}|\Theta^{prop})}{f(\Theta^{t-1})q(\Theta^{prop}|\Theta^{t-1})}\right) \tag{14}$$

One of the main advantages of HMC comes exactly due to its higher acceptance rate than RWMH. HMC improves upon RWMH by employing guided proposals based on the gradient of the log posterior to direct the Markov chain towards regions of higher posterior density while also sampling the tail areas properly as discussed in Betancourt [2017]. Thomas and Tu [2021] and Betancourt [2017] are good introductions to HMC.

Over time, HMC travels on trajectories that are governed by the Hamiltonian equations:

$$\frac{dp}{dt} = -\frac{\partial H(\Theta, p)}{\partial \Theta} = \nabla_\Theta log f(\Theta|y)$$

$$\frac{d\Theta}{dt} = \frac{\partial H(\Theta, p)}{\partial p} = \frac{\partial K(p)}{\partial p} = M^{-1}p$$

where $H(\Theta, p)$ represents the Hamilton function. In the context of this paper $H(\Theta, p) = -log f(\Theta|y) + \frac{1}{2}p^T M^{-1} p$. Additionally, $\nabla_\Theta log f(\Theta|y)$ is the gradient of the log posterior density which is the main responsible for the improvement of HMC over RWMH in acceptance ratio.

As discussed in Betancourt [2017], proposals generated from the exact solution of the Hamilton equations would be accepted with probability 1. However, they usually are not analytically tractable and solutions comes from numerical methods, typically, via the leapfrog method. Since the leapfrog is an approximation, an acceptance step is added to ensure that proposals does not deviate far from the specified Hamiltonian $H(\theta, p)$. Therefore, on practice, the acceptance rate of HMC is less than 100% but higher than RWMH. Thus, we employ HMC to sample from $p(\Theta|y)$.

An additional benefit of HMC apporaches are their easy implementation via Stan introduce by Carpenter et al. [2017]. In this paper, we combine Stan with R via rStan introduced by Guo et al. [2020]. In every application, we run our HMC scheme for 30000 iterations with the first 15000 being used as burn-in draws.

5   Empirical application: Bond Yields

The bond market is one of the largest in the world being key for investors and policy makers. Most papers focus on the first two moments e.g. Litterman and Scheinkman [1991], Collin-Dufresne and Goldstein [2002], Cochrane and Piazzesi [2005] and Joslin [2018]. Our paper focus on the much less explored third moment. As discussed previously, skewness captures the likely direction of returns. Therefore, it allows an interest rate investor to improve their forecast about sign of future yields allowing larger investments when positive returns are likely and buying protection or going short otherwise.

In our first application, we model monthly yield changes in fixed 1-year maturity for both American and Brazilian bonds. The sample of US bonds comes from the updated dataset of Gürkaynak et al. [2007] made available by the Federal Reserve Board starting at 07/1981 and going until 08/2023. The sample of Brazilian bonds is based on the DI interest rate contracts available on the Brazilian stock and futures exchange B3. The DI contracts are interpolated to form 1-year fixed maturity bonds using the same Nelson-Siegel procedure described in Gürkaynak et al. [2007] resulting in a sample starting in 02/2004 and ending in 08/2023. Figure (2) plots both time series of monthly yield changes. Notably, both series fluctuate around zero with volatility clusters. Both series hint at the possibility of skewness. For example, the US series presents a persistent and negative yield change period in the early 90 and early 2000. In the Brazilian series, the negative and persistent yield changes are even clearer with the period from 08/2016 to 03/2018 being one example.

We apply our proposed model to both US and Brazilian bonds with the priors presented in Appendix B. To account for the Brazilian time series being shorter than the American, the estimation sample comes closer to the present than the American. Precisely, we consider an estimation period which goes up to 12/2018 for the US
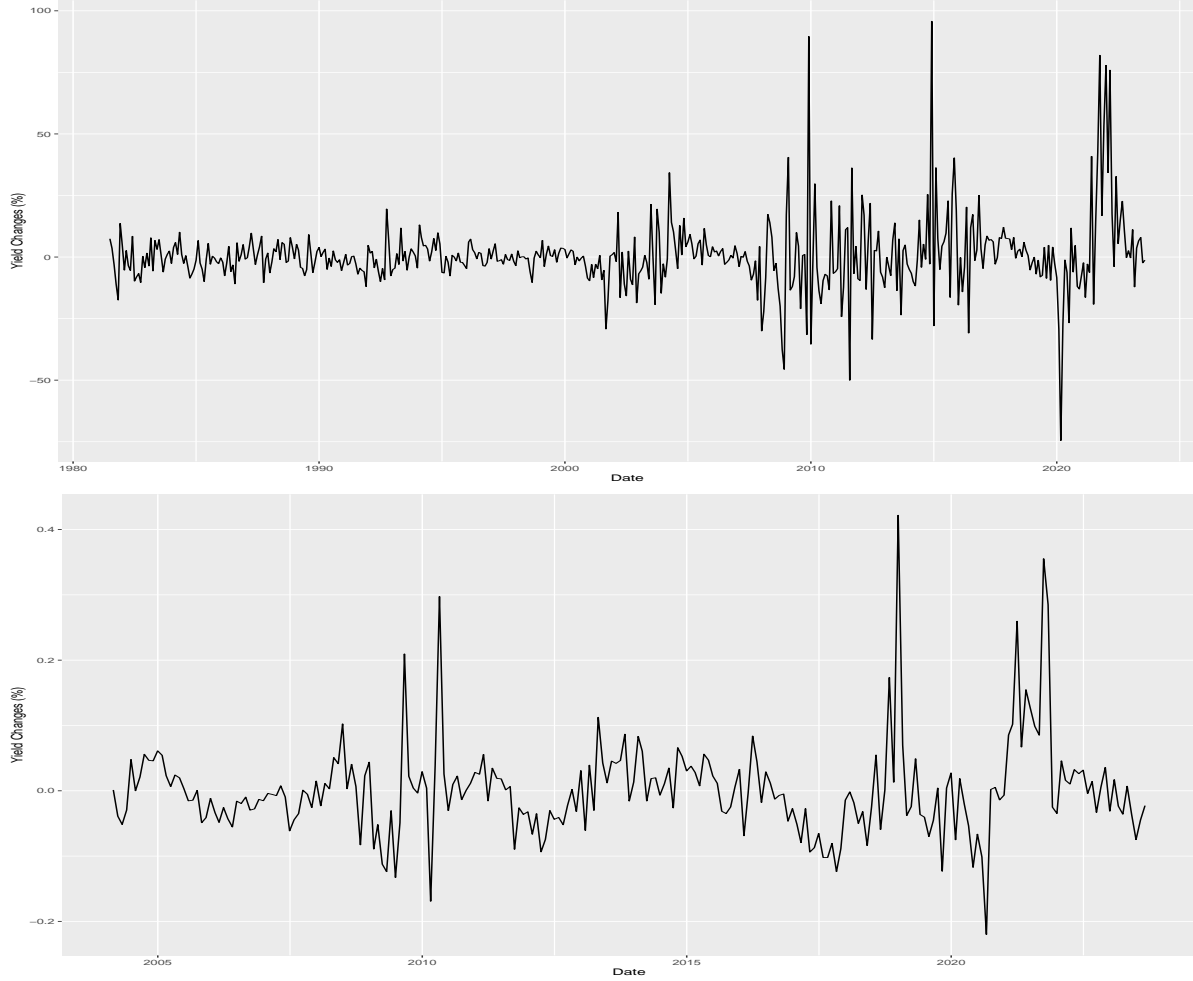
Figure 2: Change in yields from fixed 1-year maturity bonds from the US and Brazil. The top panel presents monthly changes in yields from 1-year bonds from the US government starting in 07/1981 and going up to 08/2023. Similarly, the bottom panel shows monthly changes in 1-year Brazilian bonds from 02/2004 up to 08/2023.

and up to 12/2019 to the Brazilian case. Table (1) presents the posterior summaries for the parameters of both time-series. The values of ϕ_h indicate high persistence in the log-scale which is reflected in Figure (3) which plots posterior summaries of $\{h_t\}$. In both cases, our approach captures the surges in volatility indicated in Figure (2). For example, the US series captures spikes of volatility in the early 2000's and on the financial crisis of 2008. The Brazilian time series also presents a spike in volatility in 2008 which is also captured by our model.

Additionally, both series are likely to have a dynamic skewness as show by the posterior summaries of σ_λ with the Brazilian bonds having a bigger range of variation for λ_t than bonds from the US. Such, evidence is supported by the posterior summaries of $\{\lambda_t\}$ in Figure (4) as well.

Figure (5) plots the posterior mean of $\{\lambda_t\}$ alongside with monetary easing-tightening cycles. Green shaded areas represent monetary easing periods characterized by interest rates cuts by the countries' central bank. Conversely, red shaded areas represent monetary tightening periods characterized by interest rates hikes. Our approach identifies negative skewness in the yield changes during easing and large positive skewness when a central bank is hiking

	μ_h	ϕ_h	σ_h	σ_λ	α_0
US q05	2.94	0.92	0.32	0.09	-0.62
US Mean	3.88	0.96	0.42	0.17	-0.06
US q95	4.84	0.99	0.55	0.27	0.42
BR q05	2.59	0.59	0.38	0.63	-2.67
BR Mean	3.08	0.79	0.63	1.09	-0.46
BR q95	3.59	0.94	0.91	1.75	0.79

Table 1: Posterior summaries of the parameters from our proposed model. μ_h, ϕ_h and σ_h represent the log-scale level, persistence and standard deviation, respectively.σ_λ captures the dynamic of the skewness component and α_0 is the initial level of the skewness. Both series are likely to have a dynamic skewness as show by the posterior summaries of σ_λ with the Brazilian bonds having a bigger range of skewness variation than bonds from the US
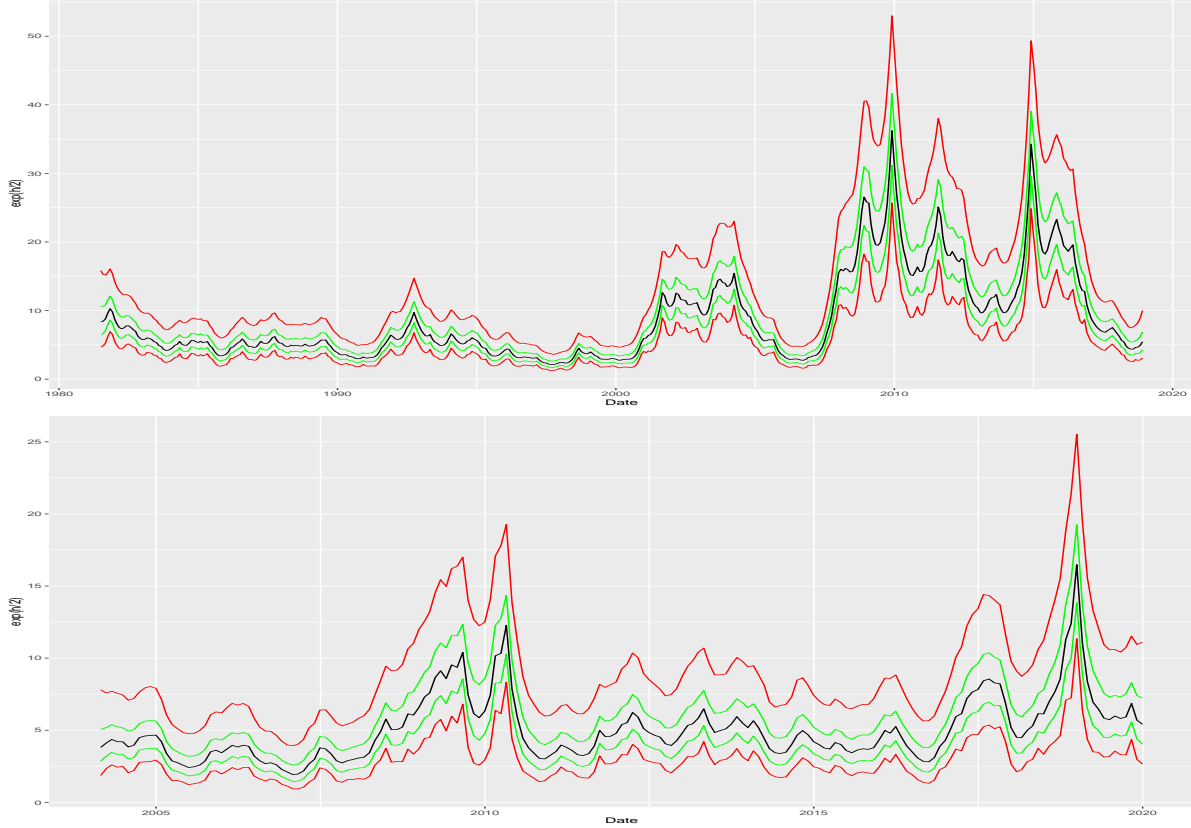

Figure 3: Scale recovered for the US and Brazilian bonds during the estimation period which goes up to 12/2018 for the US and up to 12/2019 for the Brazilian case. Green lines represent quantiles .25 and .75 while solid red lines represent quantiles .05 and .95. For both series, time-varying volatility is plausible. Top panel indicates rises in yield change volatility during the early 00's and around the global financial crisis for the US. While the bottom panel indicate peaks of volatility for Brazilian yield changes between 2009 to 2010 with a second spike in late 2018 and early 2019.

interest rates. US monetary police cycles are based on the FED effective fund rate while for the Brazilian case they are recovered from the target rate policy rate decision of each meeting. [1]

If the asymmetry of yield changes is connected to interest rate cycles, then drivers of the monetary policy should help explain skewness. We test this hypothesis by regressing the posterior mean of $\{\lambda_t\}$, denoted $\hat\lambda$, into inflation and unemployment as represented by Equation (15). Both variables are reasonable ex-ante since central banks have

---

[1] Brazilian target policy rate is available at https://www.bcb.gov.br/controleinflacao/historicotaxasjuros
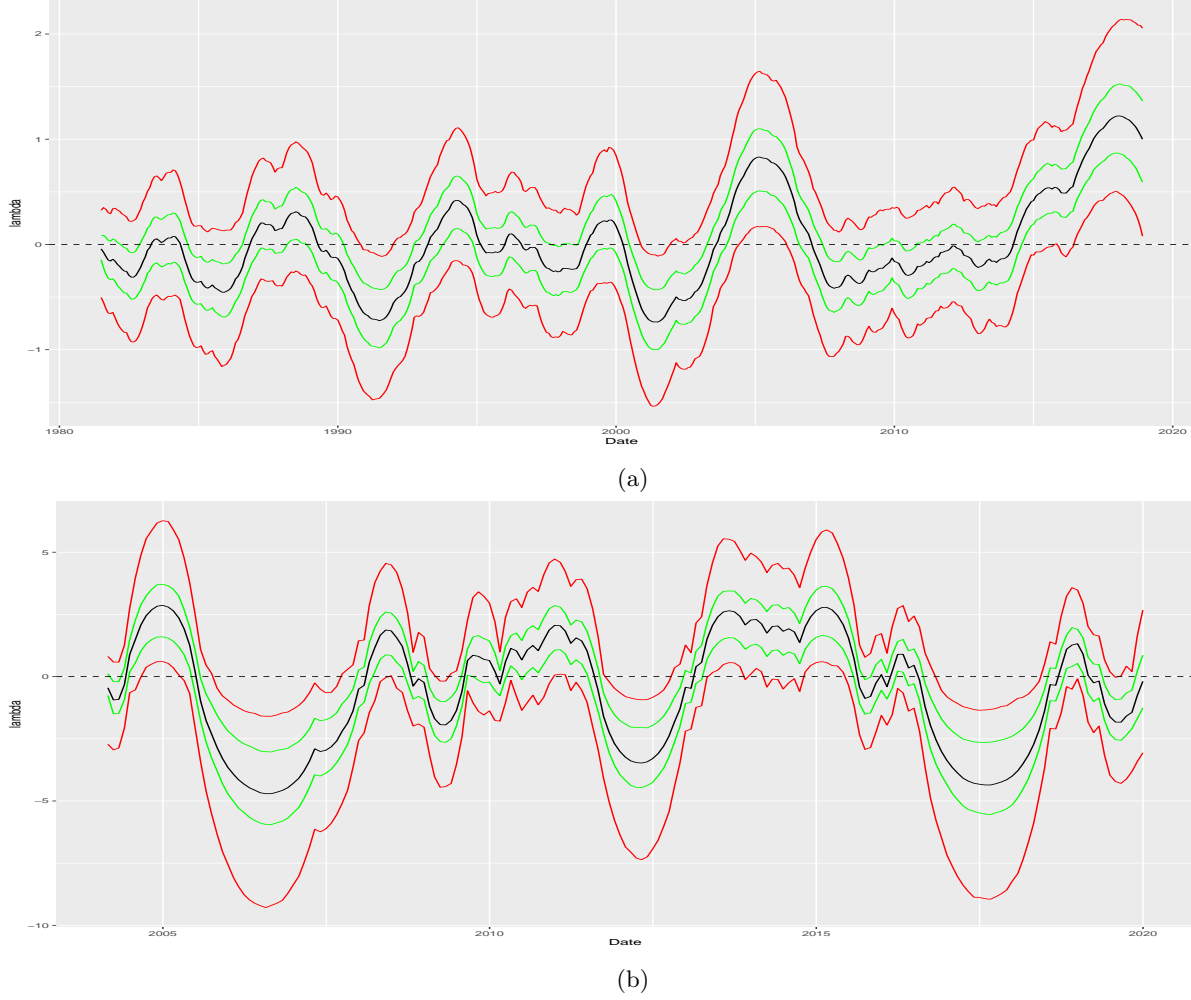
(a)

(b)

Figure 4: $\{\lambda_t\}$ recovered for the US and Brazilian bonds during the estimation period which goes up to 12/2018 for the US and up to 12/2019 for the Brazilian case. While λ is not the skewness itself, it can be transformed by using the formula in Appendix A. Green lines represent quantiles .25 and .75 while solid red lines represent quantiles .05 and .95. Time-varying skewness is likely in both cases. We obtain stronger evidence o dynamic skewness for Brazil in the bottom panel than for the US in the top one.

mandates of price stability and full employment.

$$\hat{\lambda}_t \sim N(X\beta, \sigma^2) \text{ with } X = [1, Inflation_t, Unemployment_t] \tag{15}$$

Table (2) indicates that unemployment levels partially explains the asymmetry in yield changes for both countries. The negative sign of $\beta_{unemployment}$ is reasonable since central banks usually fights high unemployment levels with interest rate cuts to promote consumption and, as show in Figure (5), easing cycles are associated with negative skewness on yield changes. Additionally, for the Brazilian case, inflation is also likely to play a role. The sign of $\beta_{inflation}$ is also reasonable since central banks combat surges in inflation with interest rake hikes which, as show in Figure (5), is associated with high values of skewness. Finally, while inflation was a problem in the US in the late 70's and early 80's, for the majority of our estimation sample the American economy didn't face inflationary
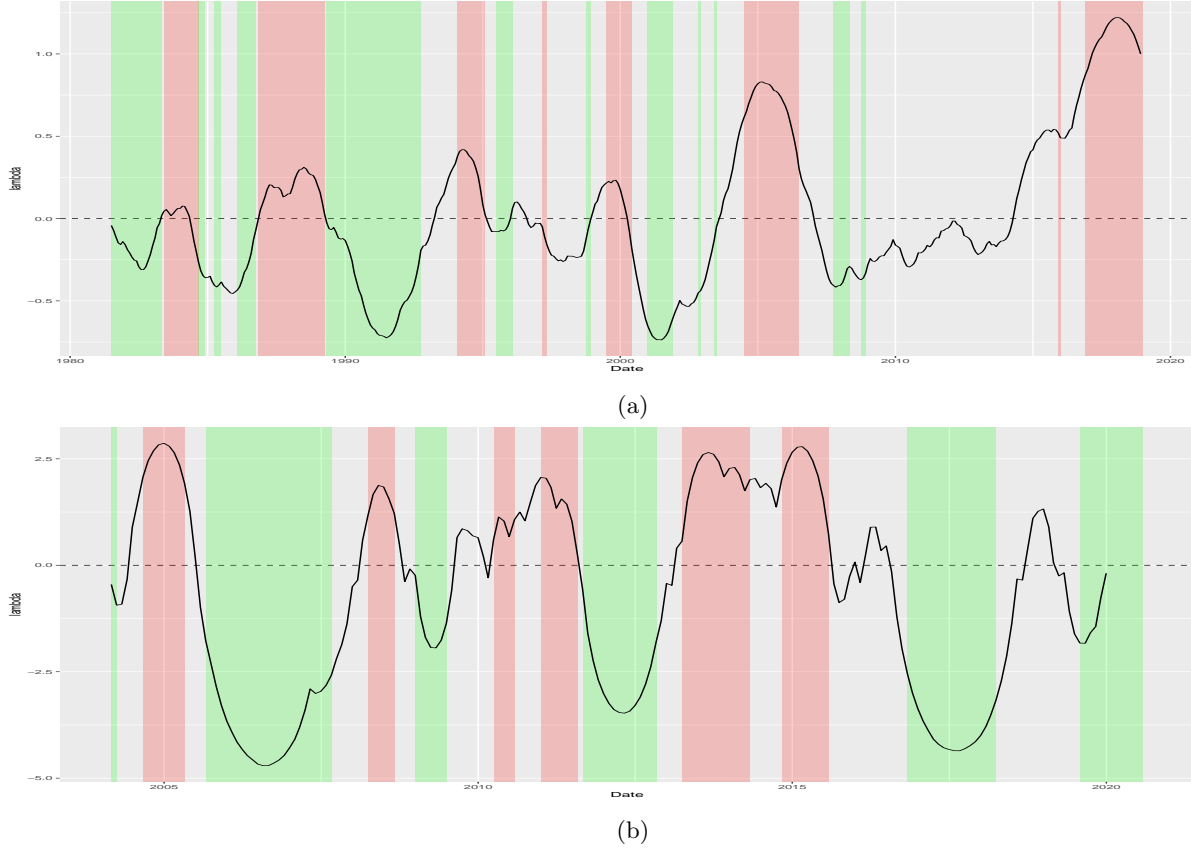
10

(a)

(b)

Figure 5: $\{\lambda_t\}$ recovered for the US, top panel, and Brazilian bonds, bottom panel, during the estimation period which goes up to 12/2018 for the US and up to 12/2019 for the Brazilian case. Green shaded area indicate monetary easing periods while red ones indicate monetary tightening. In both cases, λ_t seems able to capture the monetary cycles and reflect future direction of yield change. US monetary police cycles are based on the FED effective fund rate while for the Brazilian case they are recovered from the target rate policy rate decision of each meeting.

pressures. Since such pressures didn't exist, the yields won't react to them. Therefore, the near 0 effect of β_λ is justifiable.

	Intercept	$\beta_{Inflation}$	$\beta_{Unemployment}$
US q05	0.39	-0.01	-0.11
US mean	0.58	0.00	-0.09
US q95	0.73	0.01	-0.06
BR q05	-0.72	1.16	-0.35
BR mean	0.39	2.47	-0.23
BR q95	1.57	3.52	-0.09

Table 2: Linear regression of the posterior mean of $\{\lambda_t\}$ into inflation and unemployment during the estimation sample. For both countries unemployment partially explains the skewness in yield changes while the asymmetry in the Brazilian case is also partially explained by inflation.

We also evaluate the out of sample performance of our proposed model. If skewness is informative about the likely direction of changes in bond yields, then $sign(E_t\hat{\lambda}_{t+1})$ should agree with $sign(y_{t+1})$. We verify this claim within a increasing window framework where the first window corresponds to the estimation sample. For each window, we run our HMC procedure and obtain $sign(E_t\hat{\lambda}_{t+1})$. As shown in Table (3), our procedure indicates For US

11

bonds, $sign(y_{t+1})$ is equal to $sign(E_t\lambda_{t+1})$ 66.1% of cases with average value of change being 20.6% when correctly forecasted and only 8.7% when wrong. Similarly, for the Brazilian case, $sign(y_{t+1})$ is equal to $sign(E_t[\lambda_{t+1}])$ in 72.7% of cases. Additionally, the average magnitude of correctly forecasted yield changes is 8.2% and only 2.4% when wrongly predicted. Thus, the out of sample analysis support our claim of dynamic skewness in bond yield changes being a forecasting variable of future yield changes.

	US (%)	Brazil (%)
Hit Ratio	66.1	72.7
Avg when right	20.6	8.2
Avg when wrong	8.7	2.4
$y_{t+1} > 0$	44.6	54.5

Table 3: If bond yield skewness captures the likely direction of yield changes, then we should expected yield increases when $E_t[\lambda_{t+1}] > 0$ and, otherwise, decreases. Our model provides evidence supporting such claims. In both cases, our model correctly predicts the correct yield change direction in at least 66.1% of times.

6 Empirical application: Carry factor

In addition to the Bond yield application described in Section 5, we also consider an application for the currency market. Lustig et al. [2011] indicates that two factors, dollar factor and carry, explain the cross-section of currency returns. This application focus on the latter which captures interest rate differentials between countries by going long on countries with high interest rate differentials with respect to the US and, conversely, goes short countries with the smallest interest rate differentials with respect to the US. While we focus on FX markets, Koijen et al. [2018] argues in favor of carry-based factors being a suitable to explain a the cross-section of a large number of asset classes such as commodities and equities. We consider and updated sample of the carry factor describe in Lustig et al. [2011][2], presented in Figure (6), which starts on 11/1983 and goes up to 05/2021.

We are not the first to study the skewness of carry returns. For example, Burnside et al. [2011] and Rafferty [2012] identify a time-varying crash-risk on the carry factor. This risk would materialize in some occasions leading to large negative returns to the carry factor and skewing its distribution to the left. Conversely, Jurek [2014] uses out-of-the-money currency options hedging against large crashes and show that carry remains profitable indicating a small role for tail risk on currency returns. Additionally, Jurek [2014] presents evidence in favor of time varying volatility for carry returns and that the largest negative return of its sample, 10/2008, occurs in a period of high volatility. Our model is well suited to evaluate such claims. Jurek [2014] is not an isolated case. Barroso and Santa-Clara [2015] provides another example of tail risk mitigation in tradeable factors after accounting for volatility without skewness playing a major role.

We consider the prior specification shown in Appendix C and sample from the posterior using the HMC scheme describe in Section 4. Figure (7) plots the posterior mean (black), interquartile range (green) and q05-q95 interval

---

[2]You can check Lustig's carry factor on gsb-faculty.stanford.edu/hanno-lustig/files/2022/05/CurrencyPortfolios.xls
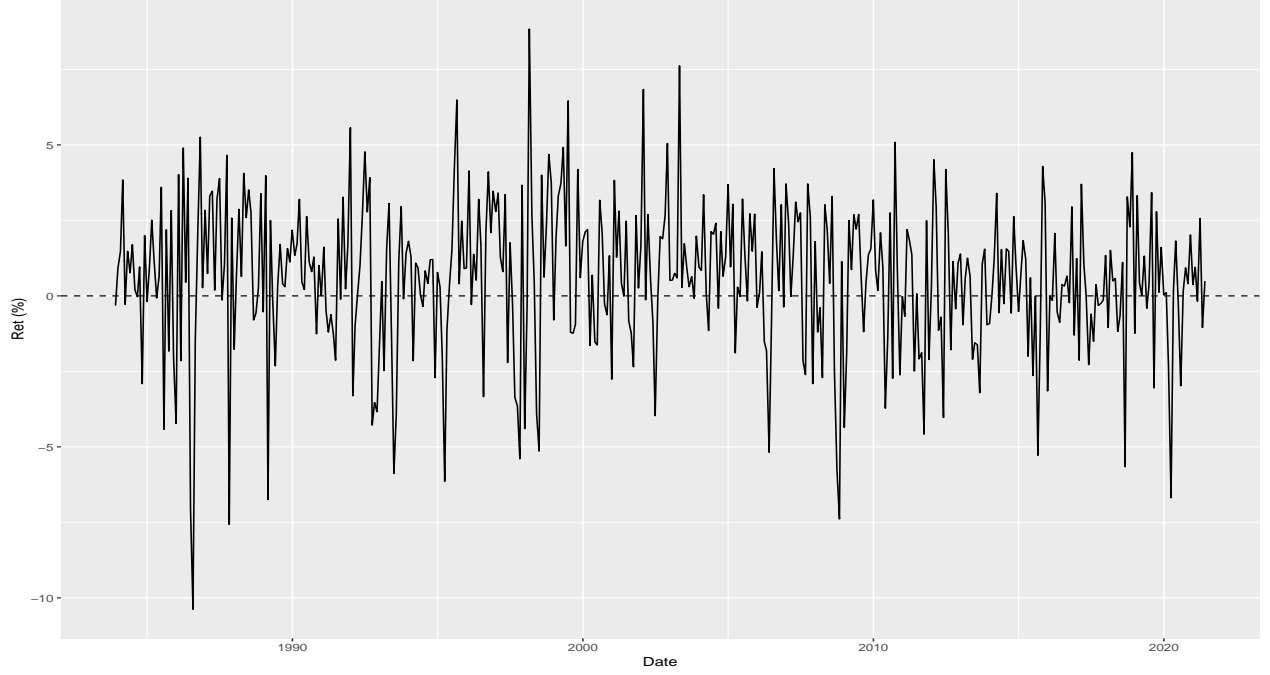
Figure 6: Updated version of the time series for the Carry factor introduced by Lustig et al. [2011]. Our sample goes from 11/1983 to 05/2021.

(red) for both $\{exp\left(\frac{h_t}{2}\right)\}$ and $\{\lambda_t\}$ which are shown at the top and bottom panel, respectively. The top panel corroborates with the evidence of time-varying volatility with a surge in volatility around 10/2008 similarly to the description of Jurek [2014]. The bottom panel indicate that is likely that there is no skewness at all after accounting for stochastic volatility.

Additionally, we verify how the carry factor performs in period with high volatility when compared to low volatility. We say that volatility is high if is above the average volatility recovered for the full sample and we say its low otherwise. Table (4) reports the results of our analysis. While the average return for the carry factor in both environments is the same, crash indicators such as the return on the 5th percentile and minimum return are severely improved. Therefore, in combination with the results shown in Figure (7b), our results indicate that after accounting for volatility, its unlikely that skewness play a large role in affecting the returns of the carry factor.

	High Vol	Low Vol
Mean	0.58	0.57
Sd	3.34	1.41
Q05	-5.38	-2.07
Min	-10.38	-3.21

Table 4: Summary statistics of the carry factor in high and low volatility environments for the sample starting in 11/1983 and going up to 12/2021. While the average return for the carry factor in both environments is the same, crash indicators such as the return on the 5-th percentile and minimum return are severely improved.
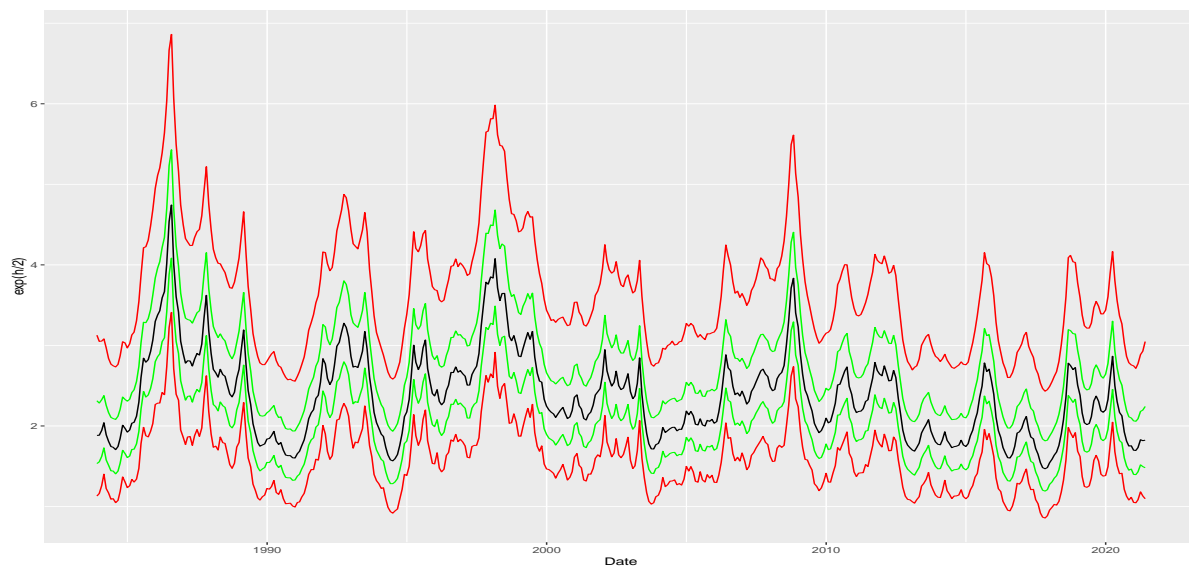
(a)

(b)

Figure 7: Skewness recovered from the carry factor. Green lines represent quantiles .25 and .75 while solid red lines represent quantiles .05 and .95. After controlling for volatility, we find no evidence of skewness in the carry factor.

7 Conclusion

This paper expands stochastic volatility models by allowing for time-varying skewness without having to imposing it. By considering a LASSO-type regularization for both the standard deviation and starting level of the skewness dynamics, our model chooses between dynamic, static and no skewness in a data-driven approach. On our bond yield application, we highlight the benefits of dynamic skewness by showing its connection to monetary easing/tightening cycles and with central banks' mandates. Additionally, we show that asymmetry is informative about the likely direction of future bond yield changes. On the second application, we shed light into the debate of carry average returns reflecting time-varying skewness versus no skewness but time-varying volatility. Our model indicates no skewness after accounting for stochastic volatility.

References

A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178, 1985.

A. Azzalini. *The skew-normal and related families*, volume 3. Cambridge University Press, 2013.

P. Barroso and P. Santa-Clara. Momentum has its moments. *Journal of Financial Economics*, 116(1):111–120, 2015.

C. L. Bayes and M. D. Branco. Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics*, pages 141–163, 2007.

M. A. Belmonte, G. Koop, and D. Korobilis. Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94, 2014.

M. Betancourt. A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*, 2017.

D. Bianchi, A. De Polis, and I. Petrella. Taming momentum crashes. *Available at SSRN 4182040*, 2022.

A. Bitto and S. Frühwirth-Schnatter. Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210(1):75–97, 2019.

C. Burnside, M. Eichenbaum, I. Kleshchelski, and S. Rebelo. Do peso problems explain the returns to the carry trade? *The Review of Financial Studies*, 24(3):853–891, 2011.

B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76, 2017.

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

J. H. Cochrane and M. Piazzesi. Bond risk premia. *American economic review*, 95(1):138–160, 2005.

P. Collin-Dufresne and R. S. Goldstein. Do bonds span the fixed income markets? theory and evidence for unspanned stochastic volatility. *The Journal of Finance*, 57(4):1685–1730, 2002.

D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC press, 2006.

E. I. George and R. E. McCulloch. Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.

J. Guo, J. Gabry, B. Goodrich, and S. Weber. Package 'rstan'. *URL https://cran.r—project.org/web/packages/rstan/*, 2020.

R. S. Gürkaynak, B. Sack, and J. H. Wright. The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, 54(8):2291–2304, 2007.

L. P. Hansen and S. F. Richard. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica: Journal of the Econometric Society*, pages 587–613, 1987.

E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 20(1):69–87, 2002.

S. Joslin. Can unspanned stochastic volatility models explain the cross section of bond volatilities? *Management Science*, 64(4):1707–1726, 2018.

J. W. Jurek. Crash-neutral currency carry trades. *Journal of Financial Economics*, 113(3):325–347, 2014.

R. S. Koijen, T. J. Moskowitz, L. H. Pedersen, and E. B. Vrugt. Carry. *Journal of Financial Economics*, 127(2): 197–225, 2018.

M. Li and M. Scharth. Leverage, asymmetry and heavy tails in the high-dimensional factor stochastic volatility model. *Journal of Business & Economic Statistics*, (just-accepted):1–38, 2020.

R. B. Litterman and J. Scheinkman. Common factors affecting bond returns. *The journal of fixed income*, 1(1): 54–61, 1991.

H. F. Lopes, R. E. McCulloch, and R. S. Tsay. Parsimony inducing priors for large scale state–space models. *Journal of Econometrics*, 230(1):39–61, 2022.

H. Lustig, N. Roussanov, and A. Verdelhan. Common risk factors in currency markets. *The Review of Financial Studies*, 24(11):3731–3777, 2011.

H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

J. Nakajima. Skew selection for factor stochastic volatility models. *Journal of Applied Statistics*, 47(4):582–601, 2020.

J. Nakajima and Y. Omori. Stochastic volatility model with leverage and asymmetrically heavy-tailed error using gh skew student's t-distribution. *Computational Statistics & Data Analysis*, 56(11):3690–3704, 2012.

T. Park and G. Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

S. T. Rachev, C. Menn, and F. J. Fabozzi. *Fat-tailed and skewed asset return distributions: implications for risk management, portfolio selection, and option pricing.* John Wiley & Sons, 2005.

B. Rafferty. Currency returns, skewness and crash risk. *Skewness and Crash Risk (March 15, 2012)*, 2012.

N. Shephard. *Stochastic volatility: selected readings*. OUP Oxford, 2005.

S. Thomas and W. Tu. Learning hamiltonian monte carlo in r. *The American Statistician*, 75(4):403–413, 2021.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

A. B. Trolle and E. S. Schwartz. The swaption cube. *The Review of Financial Studies*, 27(8):2307–2353, 2014.

Appendix A: Moments of a skew-normal distribution

Following Azzalini [1985] and Bayes and Branco [2007], ξ, ω and λ relate to mean, variance and skewness by the following expressions:

If $Z \sim SN(\xi, \omega, \lambda)$, then

$$E[Z] = \xi + \sqrt{\omega}\delta\sqrt{\frac{2}{\pi}}$$

$$Var[Z] = \omega\left(1 - \frac{2}{\pi}\delta^2\right)$$

$$\gamma = \sqrt{\frac{2}{\pi}}\left(\frac{4}{\pi} - 1\right)\left(\frac{\lambda}{\sqrt{1 + \lambda^2}}\right)^3\left(1 - \frac{2}{\pi}\frac{\lambda^2}{1 + \lambda^2}\right)^{3/2}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ and γ is the skewness index. Note that γ depends only λ.

Appendix B: Priors and additional results for the Bond applications

We consider the following structure for the bond yield applications

$$\mu_h \sim N(4, 10^2)$$

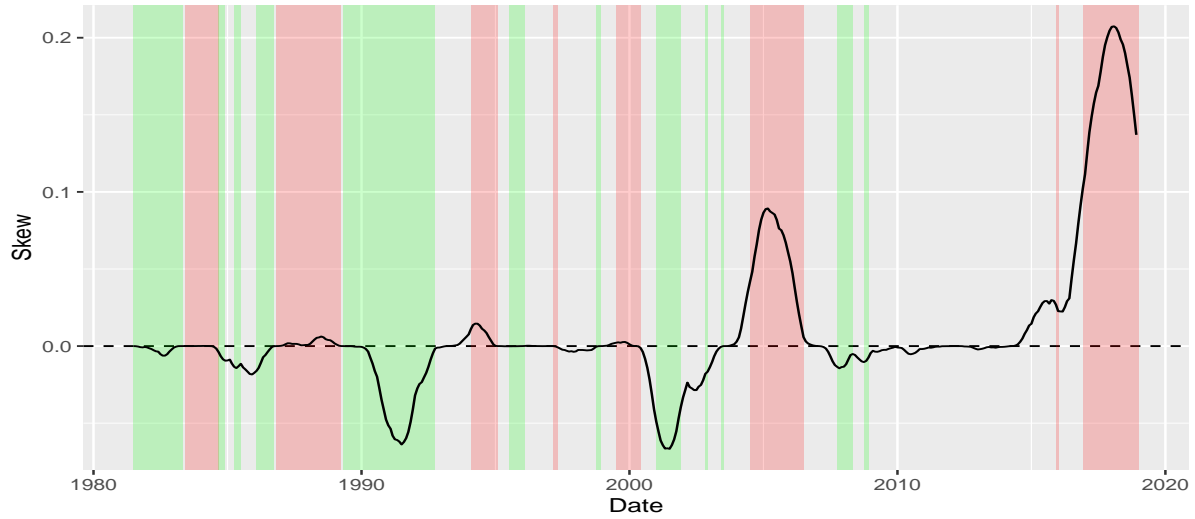$$\phi_h \sim N(0.95, 0.5^2)$$

$$\sigma \sim IG(0.1, 0.1)$$

$$\alpha_0 \sim DoubleExp(0, 1/\kappa_\alpha)$$

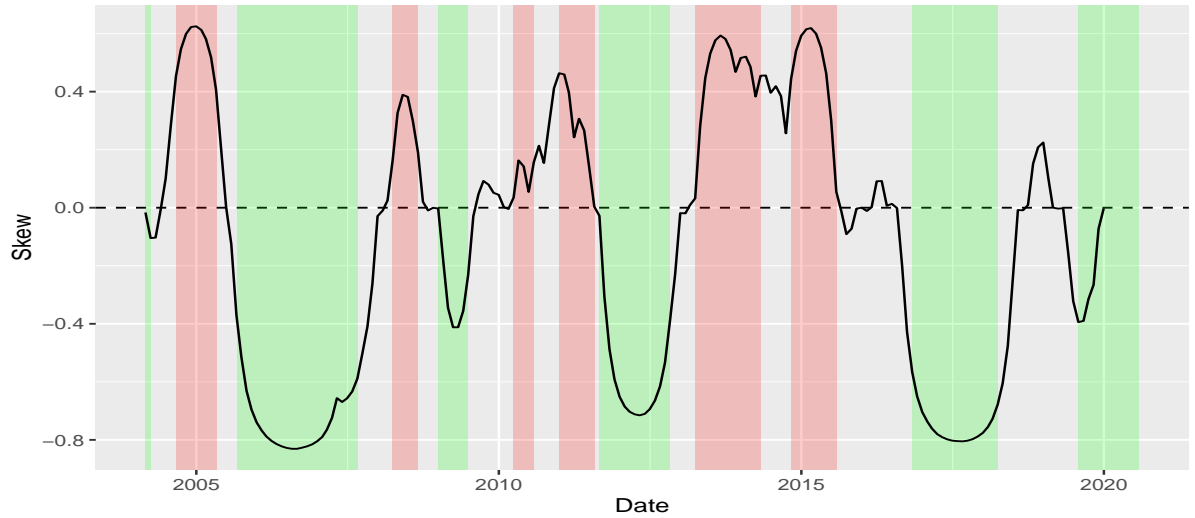$$\sigma_\lambda \sim DoubleExp(0, 1/\kappa_{\sigma_\lambda})$$

$$\kappa_\alpha \sim Gamma(0.1, 0.1)$$

$$\kappa_{\sigma_\lambda} \sim Gamma(0.1, 0.1)$$

We also plot the skewness itself using green and red shades to indicate easing and tightening periods in addition to the posterior for $\{\lambda_t\}$. While the conclusions are the same, the version with skewness may help the reader to make a better assessment of the magnitude of skewness in each period.

Figure 8: Skewness recovered based on the posterior median of $\{\lambda_t\}$ for the US, top panel, and Brazilian bonds, bottom panel, during the estimation period which goes up to 12/2018 for the US and up to 12/2019 for the Brazilian case. Green shaded area indicate monetary easing periods while red ones indicate monetary tightening. US monetary police cycles are based on the FED funds effective rate changes while for the Brazilian case they are recovered from the target rate policy rate decision of each meeting. In both countries, positive skewness is associated with tight monetary policy periods while negative skewness is connected to easing cycles.

Appendix C: Priors for the FX application

We consider the following structure for the bond yield applications

$$\mu_h \sim N(4, 10^2)$$

$$\phi_h \sim N(0.95, 0.5^2)$$

$$\sigma_h \sim IG(2, 2)$$

$$\alpha_0 \sim DoubleExp(0, 1/\kappa_\alpha)$$

$$\sigma_\lambda \sim DoubleExp(0, 1/\kappa_{\sigma_\lambda})$$

$$\kappa_\alpha \sim Gamma(0.1, 0.1)$$

$$\kappa_{\sigma_\lambda} \sim Gamma(0.1, 0.1)$$