The semiparametric asymmetric stochastic volatility model with time-varying parameters: The case of US inflation

Stefanos Dimitrakopoulos[1]

Department of Accounting, Finance and Economics, Oxford Brookes University, Oxford, OX33 1HX, UK

Abstract

We propose a semiparametric extension of the time-varying parameter regression model with asymmetric stochastic volatility. For parameter estimation we use Bayesian methods. We illustrate our methods with an application to US inflation.

Keywords: asymmetric stochastic volatility, Dirichlet process, Markov chain Monte Carlo, time-varying parameters, inflation

JEL CODE: C11, C14, C15, C22

[1]Correspondence to: Stefanos Dimitrakopoulos, E-mail: sdimitrakopoulos@brookes.ac.uk.

1 Introduction

Time varying-parameter regression models with stochastic volatility (TVP-SV models) have been successfully applied to inflation modeling (Stock and Watson, 2007; Clark and Ravazzolo, 2015; Chan, 2017).

In this paper, we focus on the relationship between inflation and volatility that has been examined by many researchers. (Friedman, 1977) points out the potential positive association between inflation and volatility. There are also many empirical evidences, including (Baillie et al., 1996), (Grier and Perry, 1998), and (Fountas, 2001). (Chan, 2017) developed a stochastic volatility in mean model with time-varying parameters and applied it to estimate inflation. (Chan, 2017) found positive relationship between inflation and volatility before early 1980s, and zero or even negative after early 1980s.

The contribution of this paper is threefold. First, we capture the correlation between inflation and volatility by modeling jointly the distribution of inflation and log-volatility within a TVP-SV model. Furthermore, the joint distribution of inflation and volatilities is modelled semiparametrically. The intuition behind this semiparametric extension is that macroeconomic shocks that have the greatest effect on the economy are often not symmetric, suggesting that innovations have a distribution that is skewed to the left or to the right.

(Dimitrakopoulos, 2017) extended semiparametrically the TVP-SV model by using mixtures of Dirichlet processes (Ferguson, 1973) for the observations' errors and the errors of the parameter-driven dynamics. (Dimitrakopoulos, 2017)'s mixture approach over both the mixture's means and variances of the observation distribution can capture this skewness. An alternative flexible approach to capturing skewness is to jointly model nonparametrically the bivariate distribution of the observations and the log-volatilities. This approach was proposed by (Jensen and Maheu, 2014) who used a bivariate Dirichlet process mixture model for the innovations of a SV model with leverage to examine the behaviour of daily returns.

Following (Jensen and Maheu, 2014), we extend the model of (Dimitrakopoulos, 2017) by accounting for a semiparametric asymmetric stochastic volatility that captures in a flexible way the joint distribution of the empirical skewness of inflation. The resulting model specification is novel and constitutes our second contribution.

We use Bayesian methods and develop an efficient Markov chain Monte Carlo algorithm for estimating the parameters of the model. This is our third contribution.

2 Econometric set up

2.1 The TVP-SV model with correlated errors

Consider the following time-varying parameter regression model with asymmetric stochastic volatility

$$y_t = \mu + \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \exp(h_t/2)\varepsilon_t,\ t = 1, ..., T, \tag{1}$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t,\ \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}),\ t = 0, 1, ..., T-1, \tag{2}$$

$$h_{t+1} = \mu_h + \phi(h_t - \mu_h) + \eta_t,\ |\phi| < 1, \tag{3}$$

where the errors ε_t and η_t are independently and identically distributed following the bivariate normal distribution,

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_h \\ \rho\sigma_h & \sigma_h^2 \end{pmatrix}\right]. \tag{4}$$

In equation (1), μ is the intercept, $\boldsymbol{\beta}$ is the constant coefficient vector of dimension $k \times 1$ and $\boldsymbol{\alpha}_t$ are the time-varying coefficients of dimension $p \times 1$. No constant is included in the design matrices $\mathbf{x}_t$ and $\mathbf{z}_t$.

The parameter-driven dynamics in equation (2) follow a random walk process which is initialized with $\boldsymbol{\alpha}_0 = \mathbf{0}$ and $\mathbf{u}_0 \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0)$, for known initial covariance matrix $\boldsymbol{\Sigma}_0$.

In equation (3), the term h_t is the log-volatility at time t and ϕ is a persistence parameter that satisfies the stationarity restriction ($|\phi| < 1$). The AR(1) stochastic volatility process is initialized with $h_1 \sim N(\mu_h, \sigma_h^2/(1-\phi^2))$.

The model given by expressions (1)-(4) is the TVP-SV model with correlated errors[1] (TVP-SVC model). Furthermore, when the correlation parameter ρ equals zero, the TVP-SVC model reduces to the standard TVP-SV model.

We also assume the following priors

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \mathbf{B}),\ \sigma_h^2 \sim \mathcal{IG}(v_a/2, v_\beta/2),\ \boldsymbol{\Sigma} \sim IW(\delta, \Delta^{-1}),$$

$$\mu_h \sim N(\bar{\mu_h}, \bar{\sigma_h}^2),\quad \mu \sim N(\bar{\mu}, \bar{\sigma}^2),\ \rho \sim N(\rho_0, \sigma_\rho^2) I_{|\rho|<1},\ \phi \sim N(\phi_0, \sigma_\phi^2) I_{|\phi|<1},$$

where IW and $\mathcal{IG}$ denote the Inverse-Wishart distribution and the inverse gamma distribution, respectively. $I_{|\rho|<1}$ is an indicator function that equals one for the stationary region and zero otherwise and $N(\rho_0, \sigma_\rho^2) I_{|\rho|<1}$ is a normal density truncated in the stationary region. Similar analysis holds for the prior of ϕ.

2.2 The semiparametric TVP-SV model with correlated errors

We relax the parametric assumption for the joint distribution of ε_t and η_t by letting this distribution be unspecified. To this end, we use the Dirichlet process prior which is a powerful tool for modelling unknown distributions. For a detailed description of this prior see (Navarro et al., 2006).

[1] In finance, the negative correlation between ε_t and η_t is called leverage effect: as asset prices decline, companies become mechanically more leveraged since the relative value of their debt rises relative to that of their equity. As a result, it is natural to expect that their stock becomes riskier, hence more volatile. It is difficult to imagine that a similar economic argument exists for inflation. For this reason, we avoid using the term "leverage" throughout the paper.

The unspecified functional form of $(\varepsilon_t, \eta_t)'$ is given by the following Dirichlet process mixture (DPM) model

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} |\Lambda_t \sim N(\mathbf{0}, \Lambda_t),$$

$$\begin{aligned} &\Lambda_t \overset{i.i.d}{\sim} G, \quad G|a, G_0 \sim DP(a, G_0), \\ &G_0 = IW(s_0, S_0), \quad a \sim \mathcal{G}(\underline{c}, \underline{d}), \end{aligned} \tag{5}$$

where $\Lambda_t = \begin{pmatrix} \sigma^2_{y,t} & \sigma_{yh,t} \\ \sigma_{yh,t} & \sigma^2_{h,t} \end{pmatrix}$. μ_h in expresson (3) is set to zero for identification reasons.

Model (5) was first proposed by (Jensen and Maheu, 2014). According to this model, the conditional distribution of the error vector $(\varepsilon_t, \eta_t)'$ given Λ_t is a bivariate Gaussian with mean zero and random variance-covariance matrix Λ_t. Λ_t is generated from an unknown distribution G on which a Dirichlet process prior is imposed. For the prior base distribution G_0 we assume an Inverse-Wishart distribution and for the positive scalar (precision parameter) a we use a gamma prior distribution. Depending on the value of a, the DPM model in expression (5) can mimic a variety of distributions (bivariate Student-t, bivariate normal, finite mixture of bivariate normals).

Furthermore, for the distribution of $\mathbf{u}_t$ we assume the following DPM structure,

$$\begin{aligned} &\mathbf{u}_t|\omega_t, \mathbf{\Sigma} \sim N(0, \omega_t^{-1}\mathbf{\Sigma}),\ t = 1, ..., T-1, \\ &\omega_t \overset{i.i.d}{\sim} G_\omega, \quad G_\omega|a_\omega, G_{0\omega} \sim DP(a_\omega, G_{0\omega}), \\ &G_{0\omega} = \mathcal{G}(\tfrac{e_\omega}{2}, \tfrac{e_\omega}{2}),\ a_\omega \sim \mathcal{G}(\underline{c}_\omega, \underline{d}_\omega). \end{aligned} \tag{6}$$

Model (6) was considered by (Dimitrakopoulos, 2017). The positive-valued scale

mixing variable ω_t is generated from a random distribution G_ω. $G_{0\omega}$ is the prior baseline gamma distribution and a_ω is the dispersion parameter. The unconditional distribution of the error $\mathbf{u}_t$ follows an infinite mixture of multivariate Gaussians, where this mixture arises from the convolution of G_ω (the mixing distribution) with a multivariate Gaussian kernel. Due to the clustering property of the Dirichlet process, this infinite mixture model reduces to a finite mixture of multivariate Gaussians with a random number of components. The resulting mixture model includes as special cases the multivariate Student-t and the multivariate Normal.

The TVP-SVC model combined with the DPM models of (5) and (6) produces the semiparametric TVP-SVC model (S-TVP-SVC model).

3 Posterior analysis

3.1 MCMC algorithm for the S-TVP-SVC model

Our MCMC algorithm updates the parameters ($\boldsymbol{\beta}$, $\mathbf{h}$, ϕ, μ, $\boldsymbol{\alpha}$, $\boldsymbol{\Sigma}$), where $\mathbf{h} = (h_1, ..., h_{T+1})'$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_T)$ as well as the DPM parameters. In the Online Appendix we provide details of the MCMC algorithm for the S-TVP-SVC model.

3.2 Density forecasts

We evaluate the performance of the proposed semiparametric model against that of competing models by conducting a recursive out-of-sample forecasting exercise. In particular, the comparison of the models is done using density forecasts. Further details are given in the Online Appendix.

4 Empirical application

4.1 Modeling strategies

Our dataset consists of US quarterly consumer price index (CPI) inflation, covering the period 1948Q1-2013Q2. The empirical model for capturing inflation dynamics is an autoregressive TVP-SVC (AR-TVP-SVC) model,

$$y_t = \alpha_{1,t} + \alpha_{2,t} y_{t-1} + \exp(h_t/2)\varepsilon_t,\ t = 1, ..., T,$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t,\ \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}),$$

$$h_{t+1} = \mu_h + \phi(h_t - \mu_h) + \eta_t,\ |\phi| < 1,$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_h \\ \rho\sigma_h & \sigma_h^2 \end{pmatrix} \right],$$

where $y_t = 400 * log(l_t/l_{t-1})$ denotes the CPI inflation and l_t is the quarterly CPI figure. We plot y_t in Figure 1.

For comparison purposes, we considered three alternative model specifications: The first model is the the semiparametric version of the AR-TVP-SVC (AR-S-TVP-SVC) model, where the disturbances $(\varepsilon_t, \eta_t)'$ and $\mathbf{u}_t$ follow the DPM structures of (5) and (6), respectively. The second model is the the AR-S-TVP-SV model, proposed by (Dimitrakopoulos, 2017) and the third one is the AR-TVP-SVC model, where the errors $(\varepsilon_t, \eta_t)'$ and $\mathbf{u}_t$ are Student-t distributed. This model is referred to as the AR-St-TVP-SVC model[2].

We threw away the first 100000 draws and kept the next 150000 MCMC draws. The hyperparameters for the priors of the AR-S-TVP-SVC model are the same as those used in the simulation experiment of the Online Appendix.

[2]The St-TVP-SVC model is described in the Online Appendix.

4.2 Empirical results

Table 1 presents the estimation results. The posterior estimate of ρ is positive and significant in the parametric models (AR-TVP-SVC and AR-St-TVP-SVC), with the correlation being stronger in the AR-TVP-SVC model than in the AR-St-TVP-SVC model. The time plot of the expected values of $p(\rho_t|y_1, ..., y_T)$, where $\rho_t = \sigma_{yh,t}/(\sigma_{y,t}\sigma_{h,t})$, $t = 1, ..., 261$ obtained from the AR-S-TVP-SVC also shows a positive correlation between inflation and volatility (Figure 2). This correlation attains its largest values during the period of Great Moderation and the Great Recession. The same holds for the dynamic evolution of the expected values of $p(\sigma^2_{y,t}|y_1, ..., y_T)$ and $p(\sigma^2_{h,t}|y_1, ..., y_T)$, $t = 1, ..., 261$ (Figures 3 and 4, respectively).

In Table 1, the estimated ρ was found to be large, positive and statistically significant; ρ=0.5434 in the AR-St-TVP-SVC model and ρ=0.5530 in the AR-TVP-SVC model. Also, from Figure 2 (AR-S-TVP-SVC model) the expected values of ρ_t are also positive and around 0.4. The positive relationship between inflation and inflation uncertainty has also been supported by previous studies; see for example (Cukierman and Meltzer, 1986) and (Berument et al., 2009).

In addition, Figure 2 suggests that for the US economy the values of ρ_t do not fluctuate substantially around 0.4, and therefore the relationship between inflation and inflation volatility is not time-varying. This empirical finding holds throughout the period in question and it is in contrast to the empirical findings of (Chan, 2017) for the US economy. (Chan, 2017) proposed a stochastic volatility in mean model with time-varying parameters and found that there is a positive and time-varying relationship between inflation and inflation uncertainty before 1980s, but no relationship afterward.

Table 1 (last row) reports the density forecast results of the four models. For the forecasting exercise the evaluation period is from 2013Q3 to 2014Q2. The AR-S-TVP-SVC model provides better density forecasts than the rest of the models as

this model has the highest log-predictive score[3] (LPS). The forecast performance of the AR-TVP-SVC model is lower than that of the AR-St-TVP-SVC model. The AR-S-TVP-SV model produces the worst density forecasts. The forecasting results clearly show that by modeling jointly the distribution of inflation and log-volatility we substantially improve the forecast performance of the AR-TVP-SV models. Furthermore, these results verify the forecast gains from modeling nonparametrically the error distributions of the AR-TVP-SVC model.

In Figure 5 the posterior estimates of the coefficients $\boldsymbol{\alpha}_t$ for the AR-S-TVP-SVC model exhibit time-variation the path of which is similar to that obtained from the rest of the models of Table 3 (Figures 6-8).

Additional empirical results are presented in the Online Appendix.

5 Conclusions

Using MCMC methods, we estimated a semiparametric time-varying parameter regression model with asymmetric stochastic volatility. The proposed model had better fit to US inflation data than competing models. We also found positive correlation between inflation and log-volatility, volatility fluctuation and time-variation in coefficients.

[3]The log-predictive score is explained in the Online Appendix.

Table 1: Empirical results

Model	AR-S-TVP-SV			AR-S-TVP-SVC			AR-TVP-SVC			AR-St-TVP-SVC		
	Mean	CD	IF	Mean	CD	IF	Mean	CD	IF	Mean	CD	IF
$\mathbf{\Sigma}_{11}$	0.2845*	0.617	63.15	0.0923*	-0.4571	94.25	0.0637*	-1.5574	93.178	0.0656*	0.7755	77.47
	(0.1551)			(0.0683)			(0.0415)			(0.0423)		
$\mathbf{\Sigma}_{22}$	0.0665*	0.286	53.95	0.0126*	3.559	27.052	0.0109*	0.9139	27.058	0.0110*	-1.245	25.751
	(0.0229)			(0.0047)			(0.0036)			(0.0035)		
ϕ	0.9681*	0.070	52.73	0.9751*	0.4035	98.86	0.9525*	-0.0949	67.643	0.9567*	1.5542	51.552
	(0.0231)			(0.0266)			(0.0290)			(0.0390)		
μ_h							0.6101	-0.2322	2.3447	0.6098	1.3335	14.976
							(0.7760)			(1.0238)		
σ_h^2	0.046*	0.8413	83.131				0.1278*	-0.0095	83.43	0.1100*	-1.1943	84.01
	(0.0349)						(0.0569)			(0.0566)		
ρ							0.5530*	-0.0101	119.26	0.5434*	0.1185	105.45
							(0.1916)			(0.1895)		
M	4.9697*	0.323	75.11	2.3632*	0.7541	95.22						
	(2.4398)			(1.196)								
M_ω	4.0795*	0.475	21.94	2.3704*	0.8025	117.17						
	(2.1991)			(1.396)								
$v1$										43.455*	-0.5186	106.03
										(39.139)		
$v2$										72.909*	0.5064	136.69
										(34.097)		
LPS	0.0798			0.6501			0.60283			0.6167		

*Significant based on the 95% highest posterior density interval. LPS stands for log-predictive score; see Online Appendix.

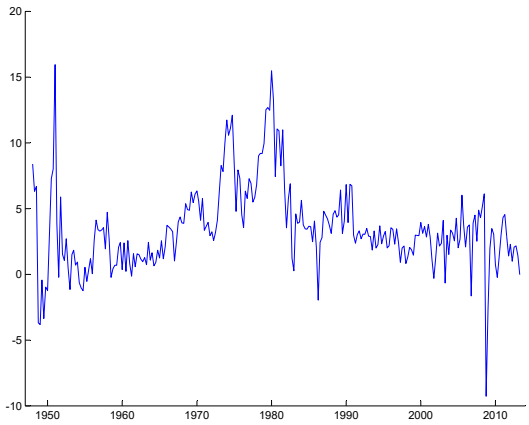

Figure 1: The inflation path from 1948Q1 to 2013Q2

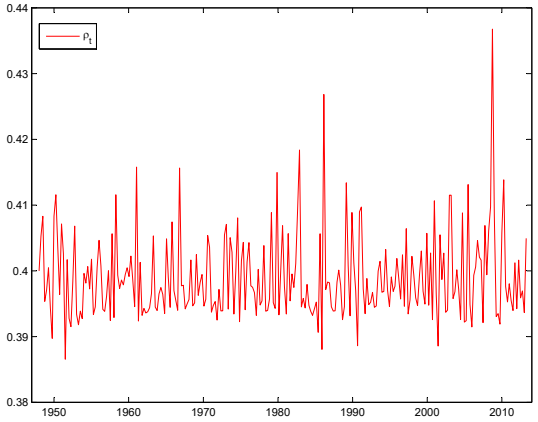

Figure 2: Time plot of the expected value of $p(\rho_t|y_1, ..., y_T)$, $t = 1, ..., 261$, obtained from the AR-S-TVP-SVC model.

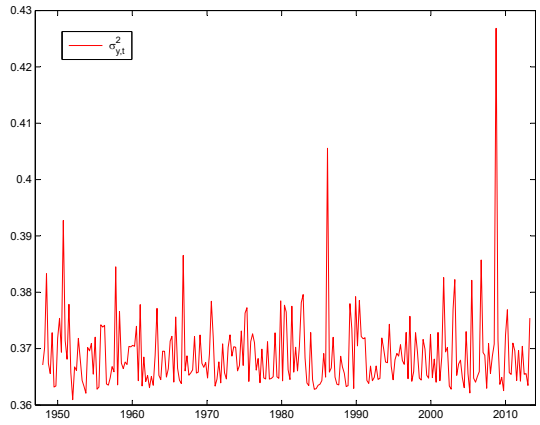

Figure 3: Time plot of the expected value of $p(\sigma^2_{y,t}|y_1, ..., y_T)$, $t = 1, ..., 261$, obtained from the AR-S-TVP-SVC model.

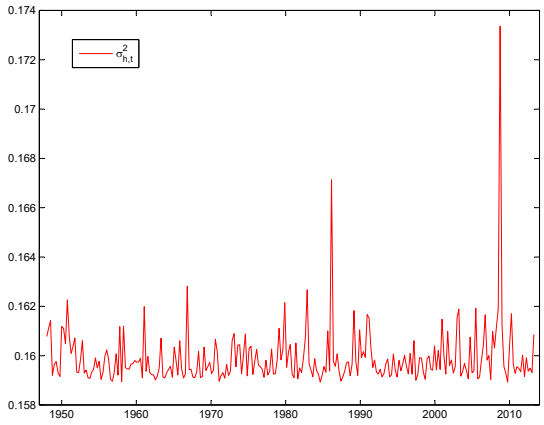

Figure 4: Time plot of the expected value of $p(\sigma^2_{h,t}|y_1, ..., y_T)$, $t = 1, ..., 261$, obtained from the AR-S-TVP-SVC model.

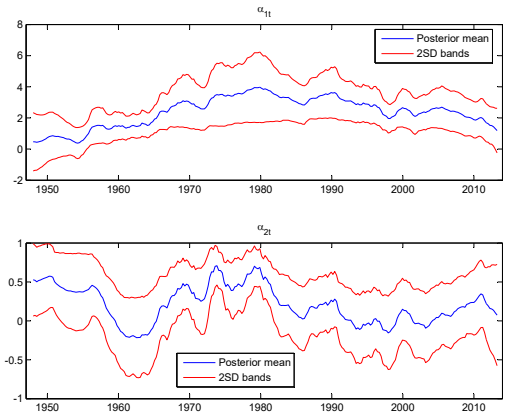

Figure 5: Evolution of $\boldsymbol{\alpha}_t$, $t = 1, ..., 261$, obtained from the AR-S-TVP-SVC model; posterior mean (blue), two standard deviation bands (red).

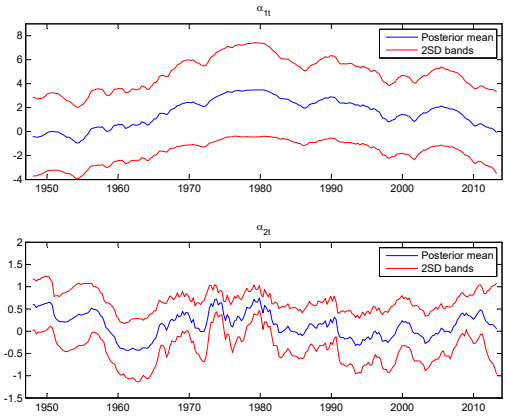

Figure 6: Evolution of $\boldsymbol{\alpha}_t$, $t = 1, ..., 261$, obtained from the AR-S-TVP-SV model; posterior mean (blue), two standard deviation bands (red).

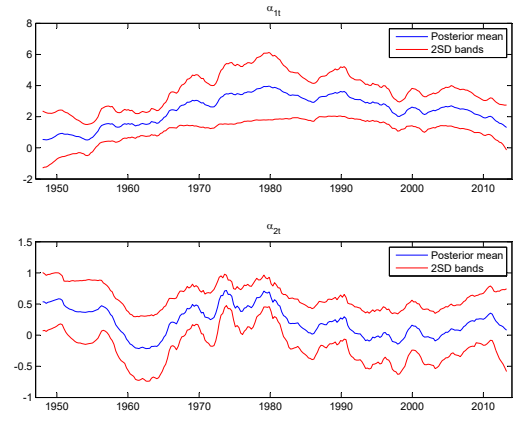

Figure 7: Evolution of $\boldsymbol{\alpha}_t$, $t = 1, ..., 261$, obtained from the AR-St-TVP-SVC model; posterior mean (blue), two standard deviation bands (red).

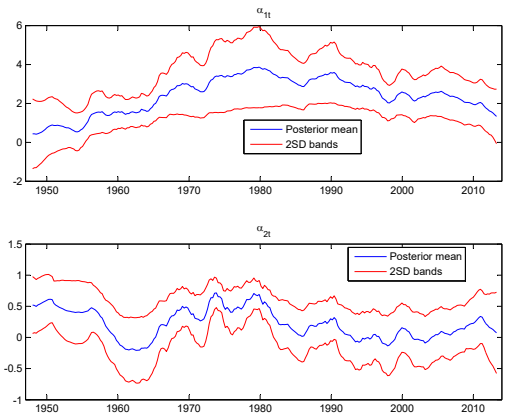

Figure 8: Evolution of $\boldsymbol{\alpha}_t$, $t = 1, ..., 261$, obtained from the AR-TVP-SVC model; posterior mean (blue), two standard deviation bands (red).

References

Baillie RT, Chung CF, Tieslau MA. 1996. Analysing inflation by the fractionally integrated ARFIMA–GARCH model. *Journal of Applied Econometrics* **11**(1): 23–40.

Berument H, Yalcin Y, Yildirim J. 2009. The effect of inflation uncertainty on inflation: Stochastic volatility in mean model within a dynamic framework. *Economic Modelling* **26**(6): 1201–1207.

Chan JCC. 2017. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics* **35**(1): 17–28.

Clark TE, Ravazzolo F. 2015. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics* **30**(4): 551–575.

Cukierman A, Meltzer AH. 1986. A theory of ambiguity, credibility, and inflation under discretion and asymmetric information. *Econometrica* **54**(5): 1099–1128.

Dimitrakopoulos S. 2017. Semiparametric Bayesian inference for time-varying parameter regression models with stochastic volatility. *Economics Letters* **150**: 10–14.

Ferguson TS. 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**(2): 209–230.

Fountas S. 2001. The relationship between inflation and inflation uncertainty in the UK: 1885–1998. *Economics Letters* **74**(1): 77 – 83.

Friedman M. 1977. Nobel lecture: Inflation and unemployment. *Journal of Political Economy* **85**(3): 451–472.

Grier KB, Perry MJ. 1998. On inflation and inflation uncertainty in the G7 countries. *Journal of International Money and Finance* **17**(4): 671 – 689.

Jensen M, Maheu JM. 2014. Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture. *Journal of Econometrics* **178**: 523–538.

Navarro DJ, Griffiths TL, Steyvers M, Lee MD. 2006. Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology* **50**(2): 101 – 122.

Stock JH, Watson MW. 2007. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit, and Banking* **39**: 3–33.

Online Appendix for: The semiparametric asymmetric stochastic volatility model with time-varying parameters: The case of US inflation

Stefanos Dimitrakopoulos*[1]

[1]Department of Accounting, Finance and Economics, Oxford Brookes University, UK

1 MCMC algorithm for the S-TVP-SVC model

The estimation of the S-TVP-SVC model is nontrivial, due to the intractability of the likelihood function under the presence of time-varying parameters, time-varying conditional variances and nonparametric error structures. Our posterior algorithm is as follows:

Posterior sampling of $\{\Lambda_t\}$

Since Λ_t, $t = 1, ..., T$, follows a random discrete probability distribution on which a DP prior is imposed, the set $\boldsymbol{\Lambda} = \{\Lambda_t\}$ will contain $m = 1, ..., M$, $M \leq T$ unique covariance matrices L_m, where

$$L_m = \begin{pmatrix} \sigma_{y,m}^2 & \sigma_{yh,m} \\ \sigma_{yh,m} & \sigma_{h,m}^2 \end{pmatrix},$$

and M is the number of unique matrices in $\boldsymbol{\Lambda}$. Define $\boldsymbol{L} = \{L_m\}$ and $\rho_m = \sigma_{yh,m}/(\sigma_{y,m}\sigma_{h,m})$. In addition, let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_T)'$ be the vector of the latent indicator variables, where $\psi_t = m$ when $\Lambda_t = L_m$, $m = 1, ..., M$.

By introducing ψ_t in the S-TVP-SVC model, we can orthogonalise the correlated errors ε_t and η_t. Following (Jensen and Maheu, 2014), one can show that equations (1) and (3) in the main paper can be rewritten in terms of the orthogonal errors w_t and u_t as follows

$$\begin{aligned} y_t &= \mu + \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \rho_{\psi_t}\sigma_{y,\psi_t}\exp(h_t/2)(h_{t+1} - \phi h_t)/\sigma_{h,\psi_t} \\ &+ \exp(h_t/2)\sqrt{1-\rho_{\psi_t}^2}\sigma_{y,\psi_t}w_t,\ w_t \sim N(0,1),\ t = 1, ..., T, \end{aligned} \tag{A.1}$$

$$h_{t+1} = \phi h_t + \sigma_{h,\psi_t}u_t,\ |\phi| < 1,\ u_t \sim N(0,1), \tag{A.2}$$

where $(w_t, u_t) \sim N(\mathbf{0}, \mathbf{I}_2)$ and $\mathbf{I}_2$ is a 2×2 identity matrix. The posterior sampling of μ, $\boldsymbol{\beta}$, $\mathbf{h}$, $\boldsymbol{\alpha}$ and ϕ, which is presented below, is based on the equations (A.1) and (A.2).

*Correspondence to: Stefanos Dimitrakopoulos, Oxford Brookes University, Department of Accounting, Finance and Economics, Oxford, OX33 1HX, UK, Tel: +44(0) 1865 485478, E-mail: sdimitrakopoulos@brookes.ac.uk.

Let the set $\mathbf{\Lambda}^{(t)}$ denote $\mathbf{\Lambda}$ with Λ_t removed. $\mathbf{\Lambda}^{(t)}$ will contain $M^{(t)}$ clusters, that is, $\boldsymbol{L}^{(t)} = (L_1^{(t)}, ..., L_{M^{(t)}}^{(t)})'$. The number of matrices in $\mathbf{\Lambda}^{(t)}$ that correspond to the distinct covariance matrix $L_m^{(t)}$ will be $n_m^{(t)} = \sum_j \mathbf{1}(\psi_j = m,\, j \neq t)$, $m = 1, ..., M^{(t)}$.

Instead of simulating $\mathbf{\Lambda}$, we sample $\boldsymbol{L}$ and $\boldsymbol{\psi}$ to improve mixing (MacEachern, 1994). The sampler for updating $\{\psi_t\}$ and $\{L_m\}$ consists of two steps.

Step 1: Sample each ψ_t according to the probabilities

$$P(\psi_t = m|\boldsymbol{L}^{(t)}, \psi^{(t)}, n_m^{(t)}) \propto \begin{cases} q_{tm} & \text{if} \quad m = 1, ..., M^{(t)} \\ q_{t0} & \text{if} \quad m = M^{(t)}+1, \end{cases} \tag{A.3}$$

where $\psi^{(t)} = \boldsymbol{\psi} \backslash \{\psi_t\}$ and the weights q_{t0} and q_{tm} in $(A.3)$ are defined respectively as

$$q_{tm} \propto n_m^{(t)} f_N(\boldsymbol{\epsilon}_t|\mathbf{0}, L_m^{(t)}),\; q_{t0} \propto a \int\; f(\boldsymbol{\epsilon}_t|\Lambda_t) dG_0(\Lambda_t),$$

where $\boldsymbol{\epsilon}_t = (\varepsilon_t, \eta_t)' = (y_t^* / \exp(h_t/2), h_{t+1} - \phi h_t)'$ and $y_t^* = y_t - \mu - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t$.

According to $(A.3)$, ψ_t can take the existing value $m = 1, ..., M^{(t)}$ with probability proportional to q_{tm}. In this case, Λ_t, $t = 1, ..., T$, is assigned to an existing (unique) covariance matrix $L_m^{(t)}$, $m = 1, ..., M^{(t)}$. $f_N(\boldsymbol{\epsilon}_t|\mathbf{0}, L_m^{(t)})$ is the bivariate normal distribution of $\boldsymbol{\epsilon}_t$ evaluated at $L_m^{(t)}$.

Also, according to $(A.3)$, ψ_t can take a new value $m = M^{(t)} + 1$ with probability proportional to q_{t0}. In this case, we set Λ_t=$L_{M^{(t)}+1}$ and sample $L_{M^{(t)}+1}$ from the posterior baseline distribution

$$\Lambda_t|\boldsymbol{\epsilon}_t, S_0, s_0 \sim IW(s_0 + 1, S_0 + \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t').$$

The term q_{t0} is proportional to the dispersion parameter a times an integral which is the marginal density of $\boldsymbol{\epsilon}_t$. This density is equal to the bivariate Student-t distribution $q_{MSt}(\boldsymbol{\epsilon}_t|\mathbf{0}, S_0/(s_0 - 1), s_0 - 1)$, with mean $\mathbf{0}$, covariance $S_0/(s_0 - 1)$ and degrees of freedom $s_0 - 1$.

Step 2:
Sample L_m, $m = 1, ..., M$ from the following baseline posterior

$$L_m|\{\boldsymbol{\epsilon}_t\}_{t \in F_m}, S_0, s_0 \sim IW(s_0 + n_m, S_0 + \sum_{t \in F_m} \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'),$$

where $F_m = \{t : \Lambda_t = L_m\}$ is the set of Λs equaling L_m.

Posterior sampling of a

We sample a as in (Escobar and West, 1995). So we first sample the latent random variable ξ from $p(\xi|a, M) \sim\; Beta(a + 1, T)$ and then we sample a from a mixture of two gammas, $p(a|\xi, M) \sim\; \pi_\xi \mathcal{G}(\underline{c} + M, \underline{d} - \log(\xi)) + (1 - \pi_\xi)\mathcal{G}(\underline{c} + M - 1, \underline{d} - \log(\xi))$, where $\pi_\xi/(1 - \pi_\xi) = (\underline{c} + M - 1)/T(\underline{d} - \log(\xi))$.

Posterior sampling of h

Apply the sampler of (Chan, 2017) to the following model

$$y_t = \mu + \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \rho_{\psi_t}\sigma_{y,\psi_t}\exp(h_t/2)(h_{t+1} - \phi h_t)/\sigma_{h,\psi_t} + \exp(h_t/2)\sqrt{1-\rho_{\psi_t}^2}\sigma_{y,\psi_t}w_t,\ w_t \sim N(0,1),\ t = 1, ..., T,$$

$$h_{t+1} = \phi h_t + \sigma_{h,\psi_t}u_t,\ |\phi| < 1,\ u_t \sim N(0,1).$$

In particular, one can show that the logarithm of the posterior distribution of the volatility vector $\mathbf{h} = (h_1, ..., h_{T+1})'$ is given by

$$\log p(\mathbf{h}|\boldsymbol{y}, \boldsymbol{\beta}, \mu, \boldsymbol{\alpha}, \{\Lambda_t\}, \phi) \approx const - \tfrac{1}{2}(\mathbf{h}'\boldsymbol{K_h}\mathbf{h} - 2\mathbf{h}'\boldsymbol{k_h}) = \log\ g(\mathbf{h}), \tag{A.4}$$

where $\boldsymbol{y} = (y_1, ..., y_T)$, $\boldsymbol{K_h} = \boldsymbol{H}'\Sigma^{-1}\boldsymbol{H} + \mathbf{G}$ and $\boldsymbol{k_h} = \mathbf{f} + \mathbf{G}\tilde{\mathbf{h}} + \boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}\boldsymbol{H}^{-1}\hat{\mathbf{h}}$. Also, $\hat{\mathbf{h}} = (0, ..., 0)'$ and $\Sigma = diag(\sigma_{h,0}^2/(1-\phi^2), \sigma_{h,\psi_1}^2, ..., \sigma_{h,\psi_T}^2)$, where $\sigma_{h,0}^2 = E[G_0]_{(2,2)} = [\frac{S_0}{s_0-p-1}]_{(2,2)}$. The point $\tilde{\mathbf{h}}$ is the mode of the posterior $\log p(\mathbf{h}|\bullet)$ in (A.4). $\boldsymbol{H}$ is a lower triangular sparse matrix

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi & 1 & 0 & \cdots & 0 \\ 0 & -\phi & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\phi & 1 \end{pmatrix}.$$

The gradient vector $\mathbf{f} = (f_1, ..., f_{T+1})'$ and the negative Hessian matrix

$$\mathbf{G} = \begin{pmatrix} G_{11} & G_{12} & 0 & \cdots & 0 \\ G_{21} & G_{22} & G_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & G_{T,T-1} & G_{TT} & G_{T,T+1} \\ 0 & \cdots & 0 & G_{T+1,T} & G_{T+1,T+1} \end{pmatrix},$$

are calculated as follows:

The logarithm of the conditional distribution $p(y_t|h_t, h_{t+1}, \boldsymbol{\beta}, \mu, \boldsymbol{\alpha}_t, \Lambda_t, \phi)$ is given by

$$\log p(y_t|h_t, h_{t+1}, \boldsymbol{\beta}, \mu, \boldsymbol{\alpha}_t, \Lambda_t, \phi) = -\tfrac{1}{2}\log\left(2\pi(1-\rho_{\psi_t}^2)\sigma_{y,\psi_t}^2\right) - \tfrac{h_t}{2}$$

$$-\frac{exp(-h_t)}{2(1-\rho_{\psi_t}^2)\sigma_{y,\psi_t}^2}\left(y_t^* - \rho_{\psi_t}\sigma_{y,\psi_t}\exp(h_t/2)\left(h_{t+1} - \phi h_t\right)/\sigma_{h,\psi_t}\right)^2.$$

Setting $p_t = p(y_t|h_t, h_{t+1}, \boldsymbol{\beta}, \mu, \boldsymbol{\alpha}_t, \Lambda_t, \phi)$ for notational convenience, we have for $t = 2, ..., T+1$,

$$f_1 = \tfrac{d\log p_t}{dh_t}, \quad f_t = \tfrac{d(\log p_t + \log p_{t-1})}{dh_t},$$

$$G_{11} = -\frac{d^2 \log p_t}{dh_t^2}, \quad G_{tt} = -\frac{d^2(\log p_t + \log p_{t-1})}{dh_t^2}, \quad G_{t-1,t} = -\frac{d^2 \log p_t}{dh_t dh_{t+1}},$$

evaluated at $\mathbf{h} = \tilde{\mathbf{h}}$, where

$$\begin{aligned}
\frac{d \log p_t}{dh_t} &= -\frac{1}{2} - \frac{1}{2(1-\rho_{\psi_t}^2)} \left(-\frac{y_t^{2*}}{\sigma_{y,\psi_t}^2 \exp(h_t)} - 2\rho_{\psi_t}^2 \phi (h_{t+1} - \phi h_t)/\sigma_{h,\psi_t}^2 \right. \\
&\qquad \left. + \frac{y_t^* \rho_{\psi_t}}{\sigma_{y,\psi_t} \exp(h_t/2) \sigma_{h,\psi_t}} (h_{t+1} - \phi h_t + 2\phi) \right), \\
\frac{d^2 \log p_t}{dh_t^2} &= -\frac{1}{2(1-\rho_{\psi_t}^2)} \left(\frac{y_t^{2*}}{\sigma_{y,\psi_t}^2 \exp(h_t)} + 2\rho_{\psi_t}^2 \phi^2/\sigma_{h,\psi_t}^2 \right. \\
&\qquad \left. - \frac{y_t^* \rho_{\psi_t}}{2\sigma_{y,\psi_t} \exp(h_t/2) \sigma_{h,\psi_t}} (h_{t+1} - \phi h_t + 4\phi) \right), \\
\frac{d \log p_t}{dh_{t+1}} &= -\frac{1}{(1-\rho_{\psi_t}^2)} \left(\rho_{\psi_t}^2 (h_{t+1} - \phi h_t)/\sigma_{h,\psi_t}^2 - \frac{y_t^* \rho_{\psi_t}}{\sigma_{y,\psi_t} \exp(h_t/2) \sigma_{h,\psi_t}} \right), \\
\frac{d^2 \log p_t}{dh_{t+1}^2} &= -\frac{\rho_{\psi_t}^2}{(1-\rho_{\psi_t}^2)\sigma_{h,\psi_t}^2}, \\
\frac{d^2 \log p_t}{dh_t h_{t+1}} &= -\frac{1}{(1-\rho_{\psi_t}^2)} \left(\rho_{\psi_t}^2 \phi/\sigma_{h,\psi_t}^2 - \frac{y_t^* \rho_{\psi_t}}{2\sigma_{y,\psi_t} \exp(h_t/2) \sigma_{h,\psi_t}} \right).
\end{aligned}$$

According to expression (A.4), the posterior $p(\mathbf{h}|\boldsymbol{y}, \boldsymbol{\beta}, \mu, \boldsymbol{\alpha}, \{\Lambda_t\}, \phi)$ is approximated by a Gaussian $g(\mathbf{h}) \propto N(\hat{\boldsymbol{m}}, {\boldsymbol{K_h}}^{-1})$, with mean $\hat{\boldsymbol{m}} = {\boldsymbol{K_h}}^{-1}\boldsymbol{k_h}$ and precision matrix $\boldsymbol{K_h}$. This Gaussian approximation is used as a proposal density in the Acceptance-Rejection Metropolis-Hastings algorithm (see, for example, (Tierney, 1994) and (Chib and Greenberg, 1995)), where candidate draws are obtained using the precision sampler of (Chan and Jeliazkov, 2009), instead of Kalman-filter based methods.

The precision sampler of (Chan and Jeliazkov, 2009) works as follows. First of all, note that $\boldsymbol{K_h}$ is a sparse matrix and therefore we can compute fast and efficiently $\hat{\boldsymbol{m}}$ without the need to obtain the inverse ${\boldsymbol{K_h}}^{-1}$, which involves a time-consuming matrix operation due to its large size. Instead, we solve the linear system $\boldsymbol{K_h}\hat{\boldsymbol{m}} = \boldsymbol{k_h}$. Next, we obtain the Cholesky decomposition $\boldsymbol{K_h} = \boldsymbol{C}\boldsymbol{C}'$. Let $\boldsymbol{x} = \boldsymbol{C}^{-1}\boldsymbol{z}$, where $\boldsymbol{z} \sim N(0, \boldsymbol{I})$. Then, $\boldsymbol{x} \sim N(0, {\boldsymbol{K_h}}^{-1})$. Finally return $\tilde{\boldsymbol{m}} = \hat{\boldsymbol{m}} + \boldsymbol{x}$.

Posterior sampling of $\boldsymbol{\beta}$

Update $\boldsymbol{\beta}$ by sampling from

$$\boldsymbol{\beta}|\mathbf{B}, \boldsymbol{\beta}_0, \mu, \boldsymbol{\alpha}, \mathbf{h}, \phi, \boldsymbol{y}, \{\Lambda_t\} \sim N(D_0 d_0, D_0),$$

where

$$D_0 = \left(\mathbf{B}^{-1} + \sum_{t=1}^{T} \frac{\mathbf{x}_t \mathbf{x}_t'}{\exp(h_t)\sigma_{y,\psi_t}^2 (1-\rho_{\psi_t}^2)} \right)^{-1},$$

$$d_0 = \mathbf{B}^{-1}\boldsymbol{\beta}_0 + \sum_{t=1}^{T} \frac{\mathbf{x}_t (y_t - \mu - \mathbf{z}_t'\boldsymbol{\alpha}_t - \rho_{\psi_t}\sigma_{y,\psi_t} \exp(h_t/2)(h_{t+1} - \phi h_t)/\sigma_{h,\psi_t})}{\exp(h_t)\sigma_{y,\psi_t}^2 (1-\rho_{\psi_t}^2)}.$$

Posterior sampling of μ

Update μ by sampling from

$$\mu|\bar{\mu},\bar{\sigma}^2,\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},\phi,\boldsymbol{y},\{\Lambda_t\} \sim N(D_1 d_1, D_1),$$

where

$$D_1 = \left(\frac{1}{\bar{\sigma}^2} + \sum_{t=1}^{T} \frac{1}{\exp(h_t)\sigma^2_{y,\psi_t}(1-\rho^2_{\psi_t})}\right)^{-1},$$

$$d_1 = \frac{\bar{\mu}}{\bar{\sigma}^2} + \sum_{t=1}^{T} \frac{y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t - \rho_{\psi_t}\sigma_{y,\psi_t}\exp(h_t/2)(h_{t+1}-\phi h_t)/\sigma_{h,\psi_t}}{\exp(h_t)\sigma^2_{y,\psi_t}(1-\rho^2_{\psi_t})}.$$

Posterior sampling of ϕ

We update ϕ using an independence Metropolis-Hastings algorithm. In particular, at the $l-th$ iteration we draw a candidate value $\phi^{(p)}$ from the truncated normal distribution $N(\sigma^2_{\hat{\phi}}\hat{\phi}, \sigma^2_{\hat{\phi}})I_{|\phi|<1}$, where

$$\sigma^2_{\hat{\phi}} = \left(\frac{1}{\sigma^2_\phi} + \sum_{t=1}^{T}\frac{h_t^2}{\sigma^2_{h,\psi_t}} + \sum_{t=1}^{T}\frac{\rho^2_{\psi_t}h_t^2}{\sigma^2_{h,\psi_t}(1-\rho^2_{\psi_t})}\right)^{-1},$$

$$\hat{\phi} = \left(\frac{\phi_0}{\sigma^2_\phi} + \sum_{t=1}^{T}\frac{h_t h_{t+1}}{\sigma^2_{h,\psi_t}} + \sum_{t=1}^{T}\frac{\rho_{\psi_t}h_t(y_t^* - \rho_{\psi_t}h_{t+1}\sigma_{y,\psi_t}\exp(h_t/2)/\sigma_{h,\psi_t})}{\sigma_{h,\psi_t}\sigma_{y,\psi_t}\exp(h_t/2)(1-\rho^2_{\psi_t})}\right)^{-1}.$$

Given $\phi^{(p)}$ and the value from the previous iteration $\phi^{(l-1)}$, $\phi^{(p)}$ is accepted as a valid current draw ($\phi^{(l)}$=$\phi^{(p)}$) with probability

$$a_p(\phi^{(l-1)}, \phi^{(p)}) = \min\left(\frac{f(h_1|\sigma^2_{h,0},\phi^{(p)})}{f(h_1|\sigma^2_{h,0},\phi^{(l-1)})}, 1\right),$$

where $f(h_1|\sigma^2_{h,0},\phi) = N(0, \sigma^2_{h,0}/(1-\phi^2))$.

Posterior sampling of $\boldsymbol{\alpha}$

Apply the precision sampler of Chan and Jeliazkov (2009) to the following model

$$\tilde{y}_t = \mathbf{z}_t'\boldsymbol{\alpha}_t + \exp(h_t/2)\sqrt{1-\rho^2_{\psi_t}}\sigma_{y,\psi_t}w_t,\ w_t \sim N(0,1),\ t = 1,...,T,$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t,\ \mathbf{u}_t \sim N(\mathbf{0}, \omega_t^{-1}\boldsymbol{\Sigma}),\ t = 0,1,...,T-1,$$

where $\tilde{y}_t = y_t - \mu - \mathbf{x}_t'\boldsymbol{\beta} - \rho_{\psi_t}\sigma_{y,\psi_t}\exp(h_t/2)(h_{t+1}-\phi h_t)/\sigma_{h,\psi_t}$.
In particular, stacking the equation for $\tilde{y}_t$ over $t = 1,...,T$, we have

$$\tilde{\boldsymbol{y}} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\xi}, \boldsymbol{\xi}_t \sim N(\mathbf{0}, \boldsymbol{S}_y) \Leftrightarrow$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}_1,...,\boldsymbol{\xi}_T)'$,

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1' & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_2' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{z}_3' & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{z}_T' \end{pmatrix}, \boldsymbol{S}_{\tilde{\boldsymbol{y}}} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_T \end{pmatrix},$$

and $s_t = \exp(h_t)(1-\rho_{\psi_t}^2)\sigma_{y,\psi_t}^2$, $t = 1, ..., T$.

The state equation $\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t$, can be written in a matrix notation as follows,

$$\boldsymbol{\Gamma}\boldsymbol{\alpha} = \tilde{\boldsymbol{\delta}}_{\boldsymbol{\alpha}} + \mathbf{u}, \mathbf{u} \sim N(\mathbf{0}, \boldsymbol{S}_u),$$

where $\tilde{\boldsymbol{\delta}}_{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}_0, \mathbf{0}, ..., \mathbf{0})'$, $\mathbf{u} = (\mathbf{u}_0, \mathbf{u}_1, ..., \mathbf{u}_{T-1})'$, $\boldsymbol{S}_u = diag(\boldsymbol{\Sigma}_0, \omega_1^{-1}\boldsymbol{\Sigma}, ..., \omega_{T-1}^{-1}\boldsymbol{\Sigma})$ and $\boldsymbol{\Gamma}$ is the first difference matrix

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{I}_p & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ -\boldsymbol{I}_p & \boldsymbol{I}_p & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{I}_p & \boldsymbol{I}_p & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\boldsymbol{I}_p & \boldsymbol{I}_p \end{pmatrix}.$$

Hence, the prior distribution of $\boldsymbol{\alpha}$ is a normal distribution, that is, $\boldsymbol{\alpha}|\{\omega_t\}_{t=1}^{T-1}, \boldsymbol{\Sigma} \sim N(\boldsymbol{\delta}_{\boldsymbol{\alpha}}, (\boldsymbol{\Gamma}'\boldsymbol{S}_u^{-1}\boldsymbol{\Gamma})^{-1})$, where $\boldsymbol{\delta}_{\boldsymbol{\alpha}} = \boldsymbol{\Gamma}^{-1}\tilde{\boldsymbol{\delta}}_{\boldsymbol{\alpha}}$. The posterior distribution of $\boldsymbol{\alpha}$ is also normal

$$\boldsymbol{\alpha}|\{\omega_t\}_{t=1}^{T-1}, \boldsymbol{\Sigma}, \tilde{\boldsymbol{y}} \sim N(\hat{\boldsymbol{\alpha}}, \boldsymbol{D}_{\boldsymbol{\alpha}}^{-1}),$$

where

$$\boldsymbol{D}_{\boldsymbol{\alpha}} = \boldsymbol{\Gamma}'\boldsymbol{S}_u^{-1}\boldsymbol{\Gamma} + \mathbf{Z}'\boldsymbol{S}_{\tilde{\boldsymbol{y}}}^{-1}\mathbf{Z},$$

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{D}_{\boldsymbol{\alpha}}^{-1}(\boldsymbol{\Gamma}'\boldsymbol{S}_u^{-1}\boldsymbol{\Gamma}\boldsymbol{\delta}_{\boldsymbol{\alpha}} + \mathbf{Z}'\boldsymbol{S}_{\tilde{\boldsymbol{y}}}^{-1}\tilde{\boldsymbol{y}}).$$

Note that $\boldsymbol{D}_{\boldsymbol{\alpha}}$ is a high-dimensional covariance matrix and therefore sampling from this posterior can be time-consuming. However, $\boldsymbol{D}_{\boldsymbol{\alpha}}$ is a band matrix and we can sample from $N(\hat{\boldsymbol{\alpha}}, \boldsymbol{D}_{\boldsymbol{\alpha}}^{-1})$ efficiently and fast, using the precision sampler of Chan and Jeliazkov (2009) which is based on block-banded and sparse matrix algorithms and not on Kalman-filter related methods.

Posterior sampling of $\boldsymbol{\Sigma}$

Update $\boldsymbol{\Sigma}$ by sampling from

$$\boldsymbol{\Sigma}|\delta, \Delta, \boldsymbol{\alpha}, \boldsymbol{\omega} \sim IW\left(\delta + T - 1, \Delta^{-1} + \sum_{t=1}^{T-1} \omega_t(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)'\right),$$

where $\boldsymbol{\omega} = (\omega_1, ..., \omega_{T-1})$.

Posterior sampling of $\boldsymbol{\omega}$

Having calculated $\mathbf{u}_t$ from $\mathbf{u}_t = \boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t$, $t = 1, ..., T-1$, we update $\boldsymbol{\omega}$ as in (Dimitrakopoulos, 2017). Since ω_t, $t = 1, ..., T-1$ follows the Dirichlet process prior G_ω, realizations of ω_t from G_ω will lie in a set of $M_\omega \leq T-1$ distinct values or clusters $\omega^* = (\omega_1^*, ..., \omega_{M_\omega}^*)$, where $\omega_{m_\omega}^*$, $m_\omega = 1, ..., M_\omega$ is a random draw from $G_{0\omega}$.

Let $\boldsymbol{\omega}^{(t)}$ denote all the elements in $\{\omega_t\}_{t=1}^{T-1}$ excluding the component ω_t. The vector $\boldsymbol{\omega}^{(t)}$ will contain ties. Suppose that $\boldsymbol{\omega}^{(t)}$ contains $M_\omega^{(t)}$ unique values, $(\omega_1^{*(t)}, ..., \omega_{M_\omega^{(t)}}^{*(t)})$ and assume also that each of these values appears in $\boldsymbol{\omega}^{(t)}$, $n_{m_\omega}^{(t)}$ times, where $n_{m_\omega}^{(t)} = \sum_j \mathbf{1}(\psi_j^\omega = m_\omega, j \neq t)$, $m_\omega = 1, ..., M_\omega^{(t)}$. The term ψ_t^ω, $t = 1, ..., T-1$ is a latent indicator variable such that $\psi_t^\omega = m_\omega$ when $\omega_t = \omega_{m_\omega}^*$, $m_\omega = 1, ..., M_\omega$.

From the Pólya-urn process (Blackwell and MacQueen, 1973), one can easily show that $\{\omega_t\}_{t=1}^{T-1}$ can be updated from the conditional posterior (continuous-discrete) distribution

$$p\left(\omega_t|\boldsymbol{\omega}^{(t)}, \mathbf{u}_t, e_\omega, \boldsymbol{\Sigma}, G_{0\omega}\right) \propto \tilde{q}_{t0} p(\omega_t|\mathbf{u}_t, e_\omega, \boldsymbol{\Sigma}) + \sum_{m_\omega=1}^{M_\omega^{(t)}} \tilde{q}_{tm_\omega} \delta_{\omega_{m_\omega}^{*(t)}}(\omega_t),$$

$$t = 1, ..., T-1, \tag{A.5}$$

where the posterior density of ω_t under the prior $G_{0\omega}$ is a gamma density, namely

$$p(\omega_t|\mathbf{u}_t, e_\omega, \boldsymbol{\Sigma}) \propto p(\mathbf{u}_t|\boldsymbol{\Sigma}, \omega_t) G_{0\omega}(\omega_t) \propto \omega_t^{\frac{e_\omega + p}{2} - 1} e^{-\frac{\omega_t (e_\omega + \mathbf{u}_t' \boldsymbol{\Sigma}^{-1} \mathbf{u}_t)}{2}},$$

and the weights $\tilde{q}_{t0}$ and $\tilde{q}_{tm_\omega}$ are given respectively by $\tilde{q}_{t0} \propto a \int f(\mathbf{u}_t|\boldsymbol{\Sigma}) dG_{0\omega}(\omega_t) \propto a q_t(\mathbf{u}_t|\mathbf{0}, \boldsymbol{\Sigma}, e_\omega)$, where q_t denotes the multivariate t-density function, and $\tilde{q}_{tm_\omega} \propto n_{m_\omega}^{(t)} f_N(\mathbf{u}_t|\mathbf{0}, \frac{1}{\omega_{m_\omega}^{*(t)}} \boldsymbol{\Sigma})$, where f_N denotes the multivariate normal distribution.

We do not sample directly from expression (A.5) but instead update the latent indicators in an analogous way to that for Λs and resample the clusters $\omega_{m_\omega}^*$, $m_\omega = 1, ..., M_\omega$ from the posterior gamma distribution

$$p(\omega_{m_\omega}^*|\{\mathbf{u}_t\}_{t \in F_{m_\omega}}, e_\omega, \boldsymbol{\Sigma}) \propto \omega_{m_\omega}^{*\, \frac{e_\omega + p \times n_{m_\omega}}{2} - 1} e^{-\frac{\omega_{m_\omega}^* (e_\omega + \sum_{t \in F_{m_\omega}} \mathbf{u}_t' \boldsymbol{\Sigma}^{-1} \mathbf{u}_t)}{2}},$$

where $F_{m_\omega} = \{t : \omega_t = \omega_{m_\omega}^*\}$ is the set of ωs sharing the parameter $\omega_{m_\omega}^*$.

Posterior sampling of a_ω

We update a_ω in the same way we update a.

2 Density forecasts

Forecast evaluation is conducted in terms of density forecasts. Define $\boldsymbol{\Omega}_T = (\boldsymbol{y}, \mathbf{X}_T, \mathbf{Z}_T)$, where $\boldsymbol{y} = (y_1, ..., y_T)$, $\mathbf{X}_T = (\mathbf{x}_1, ..., \mathbf{x}_T)$ and $\mathbf{Z}_T = (\mathbf{z}_1, ..., \mathbf{z}_T)$.

Given $\boldsymbol{\Omega}_T$ and G (the prior baseline distribution), we compute the one-step-ahead out-of-sample predictive density of y_{T+1}, $f(y_{T+1}|G, \boldsymbol{\Omega}_T)$, which is used as the density forecast for y_{T+1}. As a natural metric for the evaluation of the density forecast we compute the logarithm

of the predictive likelihood, which is the logarithm of the predictive density evaluated at the observed y^o_{T+1}, namely, $f(y^o_{T+1}|G, \mathbf{\Omega}_T)$. Next, we move one period forward and repeat the same forecasting exercise using $\mathbf{\Omega}_{T+1}$ data. For the evaluation period $t = T+1, ..., T+k$, the sum of the log predictive likelihoods $\sum_{t=T}^{T+k-1} \log f(y^o_{t+1}|G, \mathbf{\Omega}_T)$ is known as the log predictive score of the model. Higher values entail better (out-of-sample) forecasting ability of the model.

The predictive density $f(y_{T+1}|G, \mathbf{\Omega}_T)$ is obtained as follows. Let $\mathbf{\Theta}$ denote the parameter vector of the model, that is, $\mathbf{\Theta}$=($\boldsymbol{\beta}$, $\mathbf{\Sigma}$, $\boldsymbol{\alpha}$, $\mathbf{h}$, ϕ, μ $\mathbf{\Lambda}, \boldsymbol{\omega}, a, a_\omega)'$, where $\mathbf{\Lambda} = \{\Lambda_t\}$ and $\boldsymbol{\omega} = (\omega_1, ..., \omega_{T-1})$. For the S-TVP-SVC model the (one-step ahead) joint predictive density of $f((y_{T+1}, h_{T+2})')$ conditional on the prior baseline distribution G and on the data $\mathbf{\Omega}_T$ is given by

$$f\left(\begin{array}{c} y_{T+1} \\ h_{T+2} \end{array} |G, \mathbf{\Omega}_T \right) = \int f\left(\begin{array}{c} y_{T+1} \\ h_{T+2} \end{array} |\mathbf{\Theta} \right) \pi(\mathbf{\Theta}|\mathbf{\Omega}_T) d\mathbf{\Theta}, \qquad (A.6)$$

with G having been integrated out of the distribution of the error vector $(\varepsilon_t, \eta_t)'$. Expression (A.6) is approximated via Monte Carlo simulation by

$$f\left(\begin{array}{c} y_{T+1} \\ h_{T+2} \end{array} |G, \mathbf{\Omega}_T \right) \approx \frac{1}{L} \sum_{l=1}^{L} \hat{f}\left(\begin{array}{c} y_{T+1} \\ h_{T+2} \end{array} |\mathbf{\Theta}^{(l)} \right), \qquad (A.7)$$

where the functional form of the density in the right hand side of expression (A.7) is given by

$$f\left(\begin{array}{c} y_{T+1} \\ h_{T+2} \end{array} |\mathbf{\Theta} \right) = \frac{a}{a+T} f_{MSt}\left(\left(\begin{array}{c} y_{T+1} \\ h_{T+2} \end{array} \right); \left(\begin{array}{c} \mu + \mathbf{x}'_{T+1}\boldsymbol{\beta} + \mathbf{z}'_{T+1}\boldsymbol{\alpha}_{T+1} \\ \phi h_{T+1} \end{array} \right), \frac{E_{T+1} S_0 E_{T+1}}{s_0 - 1}, s_0 - 1 \right)$$

$$+ \frac{1}{a+T} \sum_{m=1}^{M} n_m N\left(\left(\begin{array}{c} y_{T+1} \\ h_{T+2} \end{array} \right); \left(\begin{array}{c} \mu + \mathbf{x}'_{T+1}\boldsymbol{\beta} + \mathbf{z}'_{T+1}\boldsymbol{\alpha}_{T+1} \\ \phi h_{T+1} \end{array} \right), E_{T+1} L_m E_{T+1} \right),$$

where $f_{MSt}(r_1, r_2, r_3)$ is the bivariate Student-t distribution with mean r_1, covariance r_2 and degrees of freedom r_3, $\boldsymbol{\alpha}_{T+1}$ is obtained from equation (2) of the manuscript and $E_{T+1} = \begin{pmatrix} \exp(h_{(T+1)/2}) & 0 \\ 0 & 1 \end{pmatrix}$. L is the number of iterations (after the burn-in period).

From expression (A.6) we can obtain the marginal posterior predictive density of y_{T+1}, which is defined as

$$f(y_{T+1}|G, \mathbf{\Omega}_T) = \int f(y_{T+1}|\mathbf{\Theta}) f(\mathbf{\Theta}|\mathbf{\Omega}_T) d\mathbf{\Theta} \approx \frac{1}{L} \sum_{l=1}^{L} \hat{f}(y_{T+1}|\mathbf{\Theta}^{(l)}),$$

where $\hat{f}(y_{T+1}|\mathbf{\Theta}^{(l)})$ when evaluated at an observed value y_{T+1} is the predictive likelihood of the S-TVP-SVC model and is defined as

$$f(y_{T+1}|\mathbf{\Theta}) = \frac{a}{a+T} f_{St}(y_{T+1}; \mu + \mathbf{x}'_{T+1}\boldsymbol{\beta} + \mathbf{z}'_{T+1}\boldsymbol{\alpha}_{T+1}, \frac{S_{0_{(1,1)}} \exp(h_{(T+1)})}{s_0 - 1}, s_0 - 1)$$

$$+\frac{1}{a+T}\sum_{m=1}^{M} n_m N(y_{T+1}; \mu + \mathbf{x}'_{T+1}\boldsymbol{\beta} + \mathbf{z}'_{T+1}\boldsymbol{\alpha}_{T+1}, \exp(h_{(T+1)})\sigma^2_{y,m}),$$

where f_{St} is the univariate Student-t distribution with mean $\mu+\mathbf{x}'_{T+1}\boldsymbol{\beta}+\mathbf{z}'_{T+1}\boldsymbol{\alpha}_{T+1}$, variance $\frac{S_{0_{(1,1)}} \exp(h_{(T+1)})}{s_0-1}$ and degrees of freedom $s_0 - 1$. $S_{0_{(1,1)}}$ is the (1,1) element of S_0.

3 The St-TVP-SVC model

Consider the following TVP-SV model,

$$y_t = \mu + \mathbf{x}'_t\boldsymbol{\beta} + \mathbf{z}'_t\boldsymbol{\alpha}_t + \exp(h_t/2)\sqrt{\lambda_{1t}}\varepsilon_t,\ t = 1, ..., T,$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t,\ \mathbf{u}_t \sim N(\mathbf{0}, \lambda_{2t}^{-1}\mathbf{\Sigma}),$$

$$h_{t+1} = \mu_h + \phi(h_t - \mu_h) + \eta_t,\ |\phi| < 1,$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_h \\ \rho\sigma_h & \sigma_h^2 \end{pmatrix}\right],$$

where $\lambda_{1t} \sim \mathcal{IG}(v1/2, v1/2)$, $\lambda_{2t} \sim \mathcal{G}(v2/2, v2/2)$ and $v1$ and $v2$ follow a uniform prior on the domain [3, 120].

To update $v1$ and $v2$ we use Metropolis-Hastings steps.

4 Monte Carlo experiments

To evaluate the efficiency of the proposed MCMC algorithm for the semiparametric TVP-SVC model we conducted Monte Carlo experiments.

The simulated data set was generated from the following model

$$y_t = 0.2 + \mathbf{x}'_t\boldsymbol{\beta} + \mathbf{z}'_t\boldsymbol{\alpha}_t + \exp(h_t/2)\sqrt{\lambda_t}\varepsilon_t,\ \lambda_t \sim \mathcal{IG}(8/2, 8/2),\ t = 1, ..., 260,$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t,\ \mathbf{u}_t \sim MVt(\mathbf{0}, \mathbf{\Sigma}, 5),$$

$$h_{t+1} = 0.8h_t + \eta_t,$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.5 \times \sqrt{0.1} \\ -0.5 \times \sqrt{0.1} & 0.1 \end{pmatrix}\right],$$

where $\boldsymbol{\beta} = (-1, 3)'$, $\boldsymbol{\alpha}_t = (\alpha_{1t}, \alpha_{2t})'$ and $\rho = -0.5$. MVt is the multivariate-t distribution with mean $\mathbf{0}$, covariance matrix $\boldsymbol{\Sigma} = diag(2, 2)$ and degrees of freedom 5, where $diag(\cdot)$ denotes a diagonal matrix. Also, $\boldsymbol{\alpha}_1 = (-10, 20)'$. $T = 260$ is almost equal to the size of the empirical data set.

The elements of $\mathbf{x}_t = (x_{1t}, x_{2t})'$ and $\mathbf{z}_t = (z_{1t}, z_{2t})'$ for $t = 1, ..., T$ are drawn from a uniform distribution, that is, $x_{jt} \sim U(-0.5, 0.5)$ and $z_{it} \sim U(-0.5, 0.5)$ for $j, i = 1, 2$.

We assume the following prior distributions

$$\boldsymbol{\beta} \sim N(0, 20 \times I_{2\times 2}),\ \boldsymbol{\alpha}_1 \sim N(0, 10 \times I_{2\times 2}),\ \mu \sim N(0, 10),$$

$$\phi \sim N(0.97, 0.1^2) I_{|\phi|<1},\ \boldsymbol{\Sigma} \sim IW(1, 10 \times I_{2\times 2}),\ G_0 = IW(10, I_{2\times 2}),$$

$$G_{0\omega} = \mathcal{G}(\tfrac{5}{2}, \tfrac{5}{2}),\ a \sim \mathcal{G}(3, 9),\ a_\omega \sim \mathcal{G}(3, 9),$$

where $I_{2\times 2}$ is a 2×2 identity matrix.

After discarding the first 50000 draws, we run the sampler 150000. The code was written in Matlab and run on a desktop with an Intel Core i7-4710HQ @2.50 GHz 2.50 GHz. For $T = 260$, it takes about 1091 seconds for 10000 iterations.

In Table 1 we present the posterior means and standard deviations of the model parameters. We also report the CD statistics of (Geweke, 1992) and the inefficiency factor (IF); see, for example, (Chib, 2001). Given the small sample size ($T = 260$), the sampler of the S-TVP-SVC model leads to satisfactory estimation accuracy. This accuracy improves as the sample size increases; in Table 2 we present the estimation results, using the same simulated data set but for $T = 800$[1].

Furthermore, Figures 1 ($T = 260$) and 2 ($T = 800$) show the evolution of the estimated time-varying parameters α_{1t} and α_{2t}, along with their two standard deviation bands. The semiparametric model is able to capture the time variation of the coefficients, with their estimated posterior means tracing well the true path of the states.

Table 1: Simulated data: T=260

Model	S-TVP-SVC			
True values	Mean	Stdev	IF	CD
$\mu = 0.2$	0.1257	0.0896	5.5058	2.3879
$\beta_1 = -1$	-1.3472	0.3115	4.6442	-0.9717
$\beta_2 = 3$	3.3511	0.2919	4.5504	0.7794
$\boldsymbol{\Sigma}_{11} = 2$	2.957	1.729	97.171	0.9650
$\boldsymbol{\Sigma}_{22} = 2$	1.131	1.4036	92.666	0.9764
$\phi = 0.8$	0.8291	0.0775	86.991	2.1522

[1]For $T = 800$, it takes about 3319 seconds for 10000 iterations.

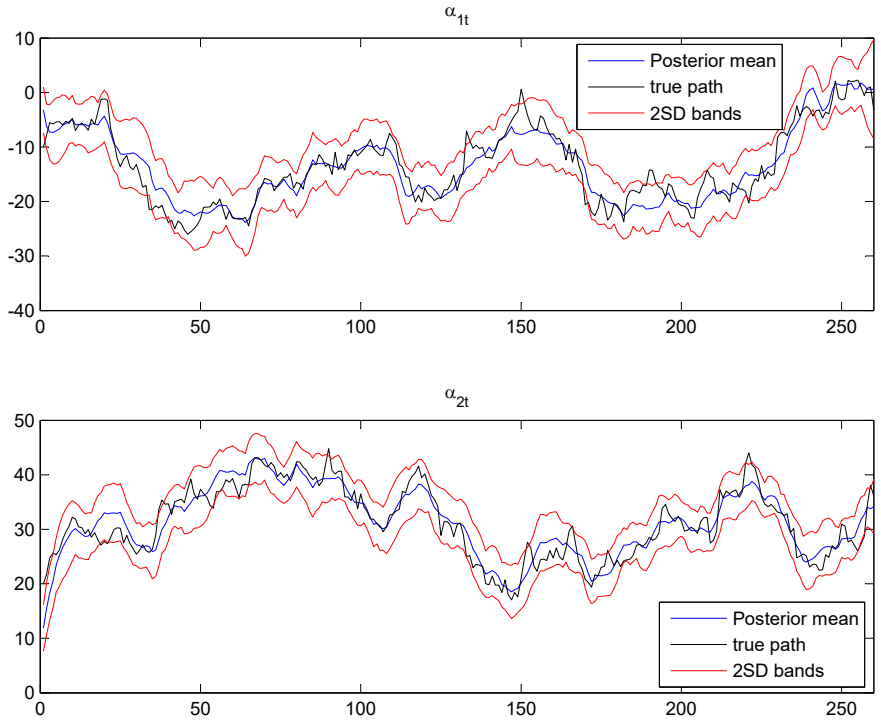

Figure 1: Simulated data: Path of the estimated α_{1t} and α_{2t} for the S-TVP-SVC model; T=260. True path (black), posterior mean (blue), two standard deviation bands (red).

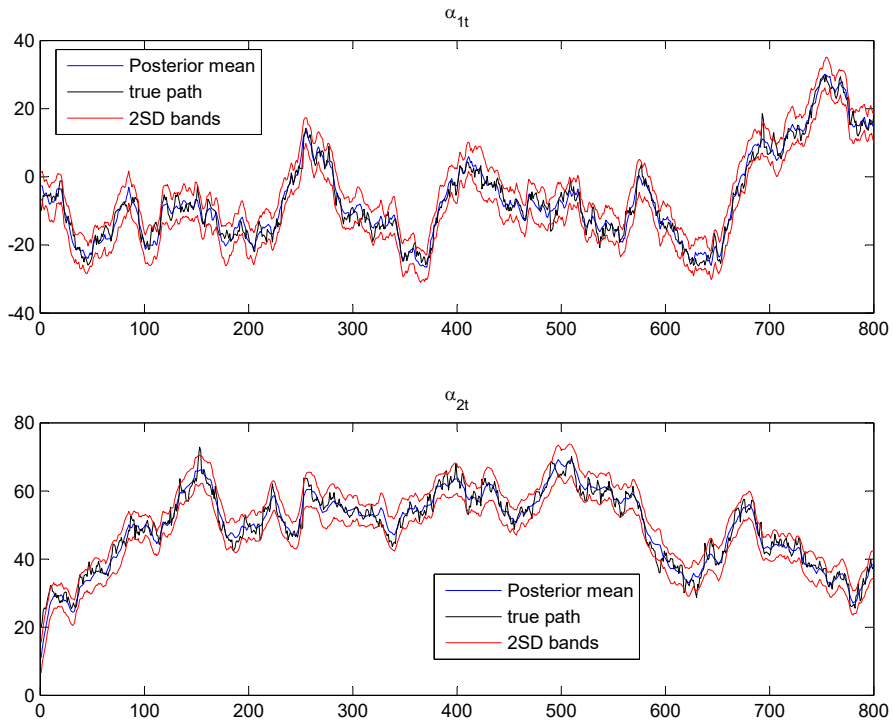

Figure 2: Simulated data: Path of the estimated α_{1t} and α_{2t} for the S-TVP-SVC model; T=800. True path (black), posterior mean (blue), two standard deviation bands (red).

Table 2: Simulated data: T=800

Model	S-TVP-SVC			
True values	Mean	Stdev	IF	CD
$\mu = 0.2$	0.2037	0.0467	3.8194	-0.6750
$\beta_1 = -1$	-0.9838	0.1644	4.2254	-0.6478
$\beta_2 = 3$	2.9987	0.1655	3.9814	-0.9239
$\mathbf{\Sigma}_{11} = 2$	2.2188	1.2914	72.032	0.5743
$\mathbf{\Sigma}_{22} = 2$	2.4661	0.6672	69.74	0.6740
$\phi = 0.8$	0.8127	0.0540	56.305	1.382

5 Additional empirical results

In Table 1 of the main paper, the DMP structure of expression (5) for the AR-S-TVP-SVC model produced $M = 2.3632$ clusters (M is explained in this Online Appendix). In other words, the proposed semiparametric model requires 2.3632 bivariate Gaussians to fit the data. For the AR-S-TVP-SV model, the estimated value of M was larger ($M = 4.3697$) to that of the AR-S-TVP-SVC model. Similarly, the AR-S-TVP-SV and AR-S-TVP-SVC models gave different degree of clustering, M_ω, in $\boldsymbol{\omega} = (\omega_1, ..., \omega_{T-1})$. M_ω is also explained in this Online Appendix. The nonnormality of the errors $(\varepsilon_t, \eta_t)'$ and $\mathbf{u}_t$ is also supported by the reported values of the degrees of freedom for the AR-St-TVP-SVC model (see Table 1 of the main paper).

In Figures 3-6 of this Appendix, we present the posterior mean of the time-varying volatility for the four empirical models of the main paper. This posterior mean is smoother in Figure 3 than in Figures 4 and 5. In Figure 6, the posterior estimates of volatilities for the AR-S-TVP-SVC model are much larger than those for the rest of the models. This is justified by the large expected values of $p(\sigma^2_{h,t}|y_1, ..., y_T)$, $t = 1, ..., 261$; see Figure 4 of the main paper.

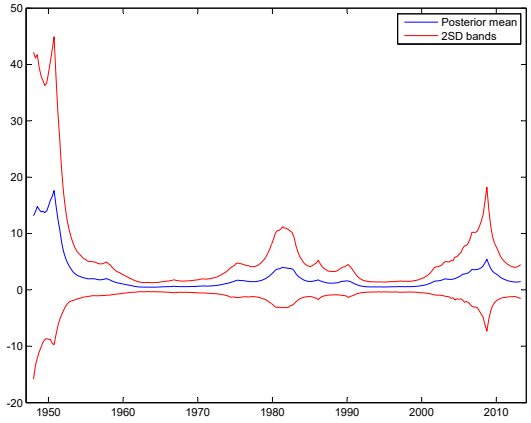

Figure 3: Evolution of $\exp(h_t)$, $t = 1, ..., 261$, obtained from the AR-S-TVP-SV model; posterior mean (blue), two standard deviation bands (red).

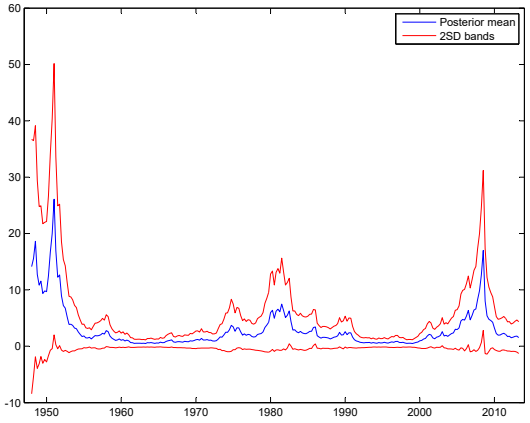

Figure 4: Evolution of $\exp(h_t)$, $t = 1, ..., 262$, obtained from the AR-TVP-SVC model; posterior mean (blue), two standard deviation bands (red).

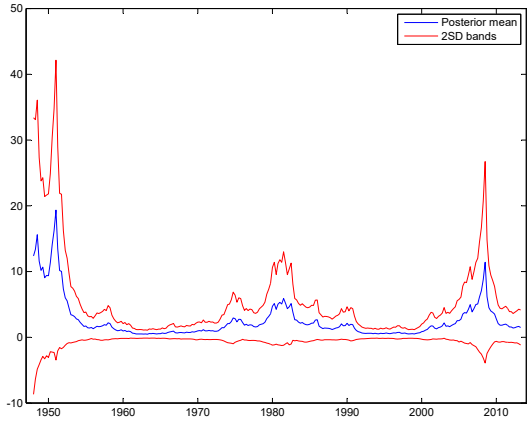

Figure 5: Evolution of $\exp(h_t)$, $t = 1, ..., 262$, obtained from the AR-St-TVP-SVC model; posterior mean (blue), two standard deviation bands (red).

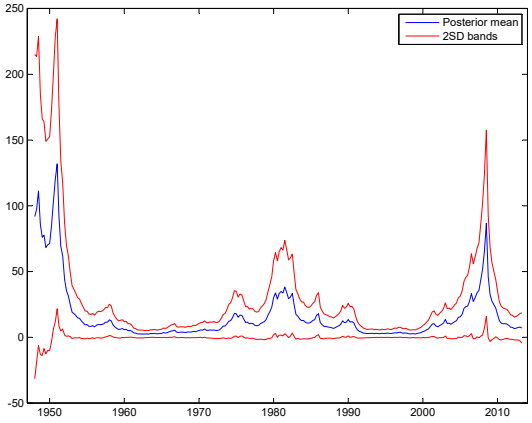

Figure 6: Evolution of $\exp(h_t)$, $t = 1, ..., 262$, obtained from the AR-S-TVP-SVC model; posterior mean (blue), two standard deviation bands (red).

References

Blackwell D, MacQueen J. 1973. Ferguson distributions via pólya urn schemes. *The Annals of Statistics* **1**(2): 353–355.

Chan JCC. 2017. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics* **35**(1): 17–28.

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

Chib S. 2001. *Markov Chain Monte Carlo methods: Computation and Inference.* In Heckman JJ. and Leamer E, (Eds.), Handbook of Econometrics, Volume 5, (pp. 3569-3649). Elsevier.

Chib S, Greenberg E. 1995. Understanding the Metropolis–Hastings algorithm. *The American Statistician* **49**(4): 327–335.

Dimitrakopoulos S. 2017. Semiparametric Bayesian inference for time-varying parameter regression models with stochastic volatility. *Economics Letters* **150**: 10–14.

Escobar M, West M. 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**(430): 577–588.

Geweke J. 1992. *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments.* In Bernardo J, Berger J, Dawid A and Smith A, (Eds.), Bayesian Statistics 4, Oxford: Clarendon Press, (pp. 641-649).

Jensen M, Maheu JM. 2014. Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture. *Journal of Econometrics* **178**: 523–538.

MacEachern SN. 1994. Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics. Simulation and Computation* **23**(3): 727–741.

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