Semiparametric Bayesian inference for time-varying parameter regression models with stochastic volatility

Dimitrakopoulos Stefanos[1]

Warwick University, UK

Abstract

We develop a Bayesian semiparametric method to estimate a time-varying parameter regression model with stochastic volatility, where both the error distributions of the observations and parameter-driven dynamics are unspecified. We illustrate our methodology with an application to inflation.

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

JEL CODE: C11, C14, C15, C22

---

[1]Correspondence to: Department of Economics, Warwick University, Coventry, CV4 7ES, UK. E-mail address: s.dimitrakopoulos@warwick.ac.uk.

1 Introduction

A vast literature has demonstrated the gains from allowing for time-varying parameters in stochastic volatility models (TVP-SV models), when analyzing (macro)financial data (Primiceri, 2005; Cogley and Sargent, 2005; Stock and Watson, 2007; D'Agostino et al., 2013; Clark and Ravazzolo, 2015). Due to the presence of the stochastic volatility component the likelihood function for this class of models is intractable. As a result, researchers have developed Markov chain Monte Carlo (MCMC) algorithms for estimating the model parameters (see, for example, Nakajima (2011)).

In this paper, we consider two semiparametric extensions of the TVP-SV model, utilising a popular Bayesian prior for modelling unknown distributions, the Dirichlet process (DP) prior (Ferguson, 1973). We first use this prior to model in a flexible way the distribution of the dependent variable's innovation and second, to consider wider class of the distribution of the time-varying parameter's innovation. The resulting semiparametric TVP-SV model is referred to as the S-TVP-SV model. To estimate the model parameters and the unknown distributions, we propose an efficient MCMC algorithm.

The first semiparametric extension has already been applied in the context of standard stochastic volatility models (Jensen and Maheu, 2010; Delatola and Griffin, 2011). The second semiparametric extension is novel and constitutes our main contribution to the Bayesian semiparametric literature on TVP-SV models.

The motivation behind the S-TVP-SV model stems from the empirical literature on inflation modelling. Recently, evidence has been found of non-normality in modelling inflation persistence, leading to increased interest in non-Gaussian (fat-tailed) distributions for modelling inflation dynamics (Lanne and Saikkonen, 2011; Lanne et al., 2012; Chiu et al., 2014; Lanne, 2015). Our point of departure is an autoregressive version of the unobserved components with stochastic volatility (UC-SV) model, proposed by Stock and Watson (2007). Stock and Watson (2007) considered a UC-SV model that decomposed inflation into a trend and a transitory component and assumed fat-tailed error distributions for the observation and state equations to control for outliers.

In this paper, we generalize the approach of Stock and Watson (2007) to account for shocks that may not be symmetrically distributed, as economic systems may react differently in recessions and expansionary periods. Furthermore, if there are different regimes operating within the sample period, a fat-tailed distribution may be inadequate to capture this data characteristic. In our proposed model, each of the unconditional error distributions for the observations and the parameter-driven dynamics is allowed to follow an infinite mixture of normals.

2 Econometric set up

2.1 The TVP-SV model

Consider the following TVP-SV model

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

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

$$h_{t+1} = \mu_h + \phi h_t + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0, \sigma_\eta^2). \tag{3}$$

Equation (1) contains two types of coefficients: the constant coefficient vector, $\boldsymbol{\beta}$**,** of dimension $k \times 1$ and time-varying coefficients, $\boldsymbol{\alpha}_t$, of dimension $p \times 1$. $\mathbf{x}_t$ and $\mathbf{z}_t$ are the design matrices which do not include an intercept and h_t is the log-volatility at time t.

Equation (2) is a random walk process which is initialized with $\boldsymbol{\alpha}_0 = \mathbf{0}$ and $\mathbf{u}_0 \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0)$, where $N(\cdot, \cdot)$ denotes the normal distribution with the initial state error variance $\boldsymbol{\Sigma}_0$ being known.

The error terms ε_t and η_t are assumed to be independent[1] for all t. The error term

[1]In the context of stochastic volatility models, Jensen and Maheu (2014) assumed that the errors ε_t and η_t are correlated and modelled them nonparametrically, using DP priors.

ε_t follows a normal distribution with mean μ and time-varying variance $\sigma_t^2 = \exp(h_t)$. The dynamics of the log-volatility $h_t = \log(\sigma_t^2)$ are described by equation (3) which is a stationary ($|\phi| < 1$) first-order autoregressive process. This process is initialized with $h_1 \sim N(\mu_h/(1-\phi), \sigma_\eta^2/(1-\phi^2)$. The parameter ϕ is the persistence volatility that measures the degree of autocorrelation in h_t, and σ_η is the standard deviation of the shock to log-volatility.

We assume the following priors over the set of parameters $(\boldsymbol{\beta}, \sigma_\eta^2, \boldsymbol{\Sigma}, \mu_h, \mu)$,

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

$$\mu_h \sim N(\bar{\mu_h}, \bar{\sigma_h}^2), \quad \mu \sim N(\bar{\mu}, \bar{\sigma}^2),$$

where IW and $\mathcal{IG}$ denote the Inverse-Wishart distribution and the inverse gamma distribution, respectively. To guarantee that the persistence parameter ϕ satisfies the stationarity restriction, we assume $(\phi + 1)/2 \sim Beta(\phi_a, \phi_\beta)$.

2.2 Two semiparametric extensions

The advantage of Dirichet process modelling results from its theoretical properties, one of which is the clustering property. A detailed exposition of the statistical properties of the DP prior is given, among others, by Ghosal (2010).

The error term ε_t, is assumed to have an unspecified functional form based on the following Dirichlet process mixture (DPM) model

$$\begin{aligned}
&\varepsilon_t|\vartheta_t, h_t \sim N(\mu_t, \lambda_t^2 \exp(h_t)),\ \vartheta_t = (\mu_t, \lambda_t^2), t = 1..., T, \\
&\vartheta_t \overset{i.i.d}{\sim} G, \\
&G|a, G_0 \sim DP(a, G_0), \\
&G_0 = N(\mu_t; \mu_0, \tau_0\lambda_t^2)\mathcal{IG}(\lambda_t^2; \tfrac{e_0}{2}, \tfrac{f_0}{2}), \\
&a \sim \mathcal{G}(\underline{c}, \underline{d}),
\end{aligned} \tag{4}$$

where μ_h in the stochastic volatility equation is set to zero for identification reasons. The unspecified functional form of the distribution of ε_t, given in (4), was first proposed by Jensen and Maheu (2010).

According to specification (4), the conditional distribution of ε_t given h_t and ϑ_t is Gaussian with mean μ_t and variance $\lambda_t^2 \exp(h_t)$. The $\vartheta_t = (\mu_t, \lambda_t^2)$ is generated from an unknown distribution G. For the prior base distribution G_0 we assume a conjugate normal-inverse gamma, $N(\mu_t; \mu_0, \tau_0 \lambda_t^2) \mathcal{IG}(\lambda_t^2; \frac{e_0}{2}, \frac{f_0}{2})$. A gamma prior distribution $\mathcal{G}(\underline{c}, \underline{d})$ is placed upon a, which is the precision parameter (positive scalar). As a tends to infinity G converges pointwise to G_0.

One can show that the unconditional distribution of ε_t follows an infinite mixture model with time-varying means and variances. So our DPM model is able to capture asymmetries and multiple modes that may characterize the data.

Furthermore, to capture the uncertainty about the distribution of $\mathbf{u}_t$, we impose on it the following novel flexible structure,

$$\begin{aligned}
&\mathbf{u}_t | \omega_t, \mathbf{\Sigma} \sim N(0, \omega_t^{-1} \mathbf{\Sigma}),\ t = 1, ..., T-1, \\
&\omega_t \overset{i.i.d}{\sim} G_\omega, \\
&G_\omega | a_\omega, G_{0\omega} \sim DP(a_\omega, G_{0\omega} = \mathcal{G}(\tfrac{e_\omega}{2}, \tfrac{e_\omega}{2})), \\
&a_\omega \sim \mathcal{G}(\underline{c}_\omega, \underline{d}_\omega).
\end{aligned} \tag{5}$$

The positive scale parameter ω_t in (5) comes from an unknown discrete distribution G_ω. The Dirichlet process prior in (5) is defined by the parameter a_ω and the base gamma distribution $G_{0\omega}$. As the precision parameter a_ω tends to infinity, G_ω converges pointwise to $G_{0\omega}$. In this case, the unconditional distribution of $\mathbf{u}_t$ is a multivariate Student-t distribution with e_ω degrees of freedom and as e_ω increases the error distribution mimics the Normal distribution. For small values of a_ω the unconditional distribution of $\mathbf{u}_t$ is a finite mixture of multivariate normals, each of which has the same mean. Therefore, our semiparametric approach for the distribution of $\mathbf{u}_t$ can capture the potential clustering in the mixing scalar parameter of the innovation's covariance matrix.

The TVP-SV model combined with the DPM models of (4) and (5) produces the semiparametric TVP-SV model (S-TVP-SV model).

3 Posterior analysis

3.1 The MCMC algorithm for the S-TVP-SV model

Define

$$\boldsymbol{y} = (y_1, ..., y_T),\ \boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_T),\ \mathbf{h} = (h_1, ..., h_T),$$
$$\boldsymbol{\theta} = (\vartheta_1, ..., \vartheta_T),\ \vartheta_t = (\mu_t, \lambda_t^2),\ \boldsymbol{\omega} = (\omega_1, ..., \omega_{T-1}).$$

Our MCMC scheme for the semiparametric model consists of two parts. In part I, we update the parameters ($\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, σ_η^2, $\boldsymbol{\alpha}$, $\mathbf{h}$, ϕ) and recover the error terms $\{\varepsilon_t\}_{t=1}^T$ and $\{\mathbf{u}_t\}_{t=1}^{T-1}$. We sample $\boldsymbol{\alpha}$ using the simulation smoothing algorithm of De Jong and Shephard (1995). To update the volatility vector $\mathbf{h}$ we apply the approach of Chan (2015), which is not based on Kalman-filter methods but on the precision sampler of Chan and Jeliazkov (2009).

Having calculated the error terms $\{\varepsilon_t\}_{t=1}^T$ and $\{\mathbf{u}_t\}_{t=1}^{T-1}$, we update, in part II, the DP parameters $(\boldsymbol{\theta}, \boldsymbol{\omega}, a, a_\omega)$ using marginal methods, since the DP is integrated out; see, for example, Escobar and West (1995) and MacEachern and Müller (1998).

Details of the MCMC algorithm for the semiparametric model along with a simulation study are given in the Online Appendix.

3.2 Posterior predictive density of the error term ε_{T+1}

A key quantity of interest in density estimation and an important feature of Bayesian inference is the posterior predictive density. With respect to the S-TVP-SV we obtain from the sampler the out-of-sample posterior predictive density for the (one-step ahead) error term ε_{T+1} conditional on the data $\boldsymbol{\Omega}_T = (\boldsymbol{y}, \mathbf{X}_T, \mathbf{Z}_T)$, where $\mathbf{X}_T =$

$(\mathbf{x}_1, ..., \mathbf{x}_T)$ and $\mathbf{Z}_T = (\mathbf{z}_1, ..., \mathbf{z}_T)$ which is given by

$$f(\varepsilon_{T+1}|\boldsymbol{\Omega}_T) = \int f(\varepsilon_{T+1}|\boldsymbol{\theta}^*, h_{T+1}, a)\pi(\boldsymbol{\theta}^*, h_{T+1}, a|\boldsymbol{\Omega}_T)d\boldsymbol{\theta}^* dh_{T+1} da. \quad (6)$$

$$\approx \tfrac{1}{L}\textstyle\sum_{l=1}^{L} f(\varepsilon_{T+1}|\boldsymbol{\theta}^{*(l)}, h_{T+1}^{(l)}, a^{(l)}),$$

where $\boldsymbol{\theta}^* = (\vartheta_1^*, ..., \vartheta_M^*)'$, $M \leq T$ is the set of unique values from $\boldsymbol{\theta}$, with ϑ_m^*=(μ_m^*, λ_m^{*2}), $m = 1, ..., M$ and M is the number of clusters in $\boldsymbol{\theta}$ (see also Online Appendix for further details). $\boldsymbol{\theta}^{*(l)}$ and $a^{(l)}$ are simulated samples of $\boldsymbol{\theta}^*$ and a respectively and $h_{T+1}^{(l)}$ is a posterior draw generated from $N(\phi^{(l)} h_T^{(l)}, \sigma_\eta^{2(l)})$. L is the number of iterations after the burn-in period. The predictive density of ε_{T+1} conditional on $\boldsymbol{\theta}^{*(l)}, h_{T+1}^{(l)}$ and $a^{(l)}$ is a mixture of a Student-t density and Normal densities, namely,

$$f(\varepsilon_{T+1}|\boldsymbol{\theta}^{*(l)}, h_{T+1}^{(l)}, a^{(l)}) = \frac{a^{(l)}}{a^{(l)} + T} q_t(\varepsilon_{T+1}|\mu_0, (\exp(h_{T+1}^{(l)}) + \tau_0) f_0/e_0, e_0)$$

$$+\tfrac{1}{a^{(l)}+T} \textstyle\sum_{m=1}^{M^{(l)}} n_m^{(l)} N(\varepsilon_{T+1}|\mu_m^{*(l)}, \exp(h_{T+1}^{(l)})\lambda_m^{*2(l)}), \quad (7)$$

where $q_t(.|m, v, u)$ is the Student-t distribution with mean m, degrees of freedom u and scale factor v. The quantity n_m is explained in the Online Appendix.

3.3 Model comparison

We conduct Bayesian model comparison, using the Deviance information criterion (DIC) (Spiegelhalter et al., 2002) and cross-validation predictive densities. Further details on how to implement these methods are provided in the Online Appendix.

4 Empirical application

We use data on US quarterly consumer price index (CPI) inflation from 1948Q1 to 2013Q2. For modelling inflation persistence, we consider the following autoregressive TVP-SV (AR-TVP-SV) model,

$$y_t = \alpha_{1,t} + \alpha_{2,t} y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \exp(h_t)),\ t = 1, ..., T,$$

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

$$h_{t+1} = \mu_h + \phi h_t + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0, \sigma_\eta^2),$$

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.

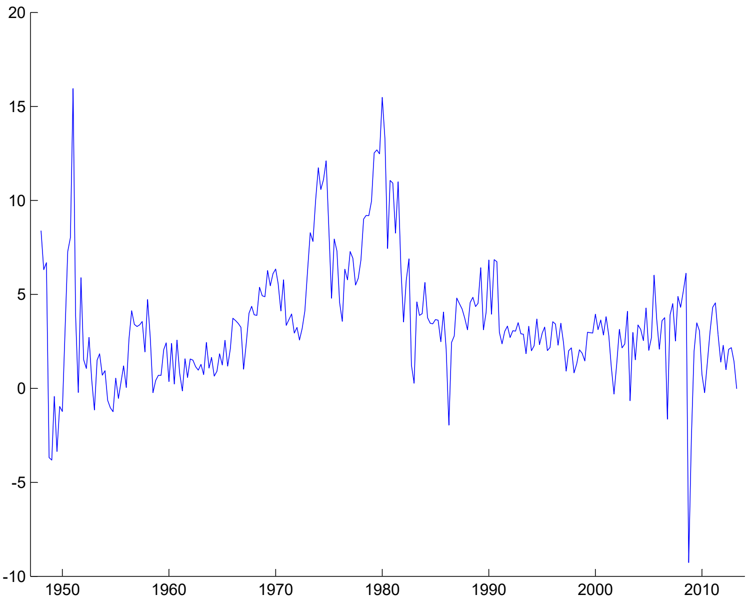

Figure 1: The inflation path from 1948Q1 to 2013Q2.

In the semiparametric version of the AR-TVP-SV model, denoted as the AR-S-TVP-SV model, the error terms ε_t and $\mathbf{u}_t$ follow the DPM models of (4) and (5), respectively[2]. For comparison purposes, we also estimated the AR-TVP-SV model,

[2]A limitation of the semiparametric model is that mixing over the time-varying parameters

with the errors ε_t and $\mathbf{u}_t$ being Student-t distributed. We refer to this model as AR-St-TVP-SV. The St-TVP-SV model is presented in the Online Appendix.

After discarding the first 80000 draws, we run the sampler for 150000 iterations. To monitor convergence we use the CD statistics of Geweke (1992) and the inefficiency factor (IF); see, for example, Chib (2001). For the AR-S-TVP-SV model, we chose the same hyperparameters for the priors as in the simulation study (see Online Appendix).

The estimation results are presented in Table 1. Across all models of Table 1, all the parameters but μ_h are significant. Based on the DIC and CV values (Table 1), the AR-S-TVP-SV model has the best fit to the data. The AR-TVP-SV model is the least preferred model.

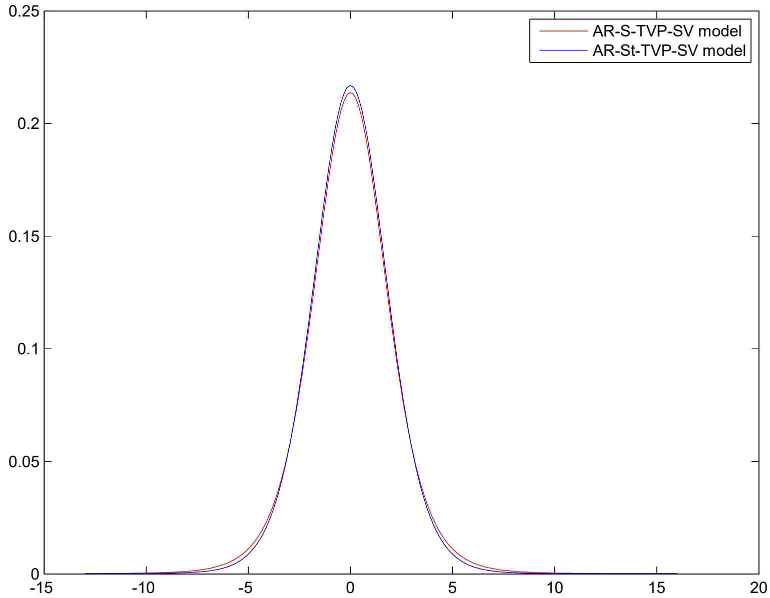

Figure 2: Posterior predictive densities for ε.

The posterior predictive density of the error term ε_t for the AR-S-TVP-SV model, which is plotted against that of the AR-St-TVP-SV model (Figure 2), indicates that the distribution of the dependent's variable innovation is nonnormal (with kurtosis 5.6254 and skewness 1.9735). This empirical finding is supported by the fact that

scaled covariance matrix fails to capture the regime switching behavior of the Sims and Zha (2006) model. A change from one regime's parameter values to another is only possible if the mixture representation of the parameter innovations is mixed over the mean vector of the normal kernel.

the semiparametric model requires $M = 4.3764$ clusters to fit the data (Table 1). The parametric models inflate the volatility parameter σ_η to compensate for the excess kurtosis found in the data; the estimated degrees of freedom $v1$ for the AR-St-TVP-SV is 9.3842.

The path of the posterior estimates of $\exp(h_t)$ obtained from the semiparametric model shows high inflation volatility during the Great Moderation and the Great Recession (Figure 3).

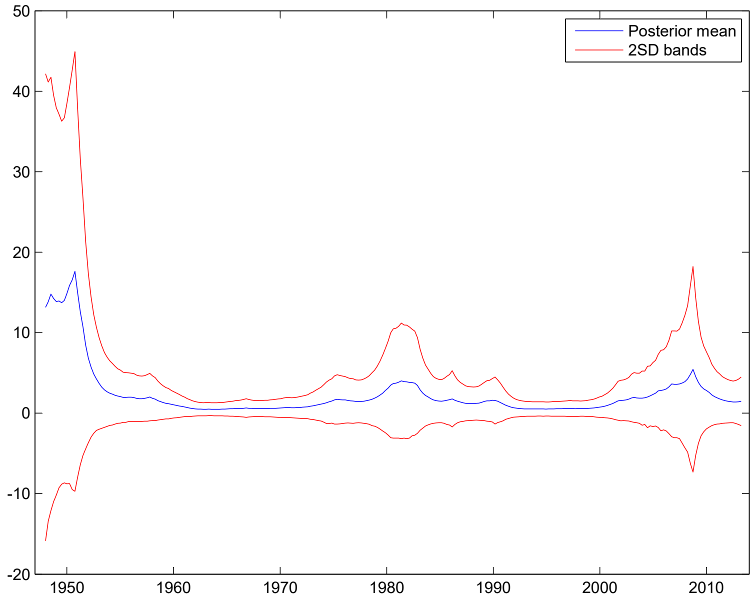

Figure 3: Evolution of $\exp(h_t)$ obtained from the AR-S-TVP-SV model; posterior mean (blue), two standard deviation bands (red).

Also, the AR-S-TVP-SV model highlights some degree of clustering in the mixing scalar parameter of the innovation's covariance; the number of clusters in $\boldsymbol{\omega}$ was found to be M_ω =4.0796 (Table 1)-M_ω is explained in the Online Appendix.

Figure 4 presents the estimates of $\boldsymbol{\alpha}_t$ for the AR-S-TVP-SV model. As can be seen, there is apparent time-variation in these estimates, highlighting the importance of allowing for time-varying parameters. Similar results were produced by the rest of the models (see Online Appendix).

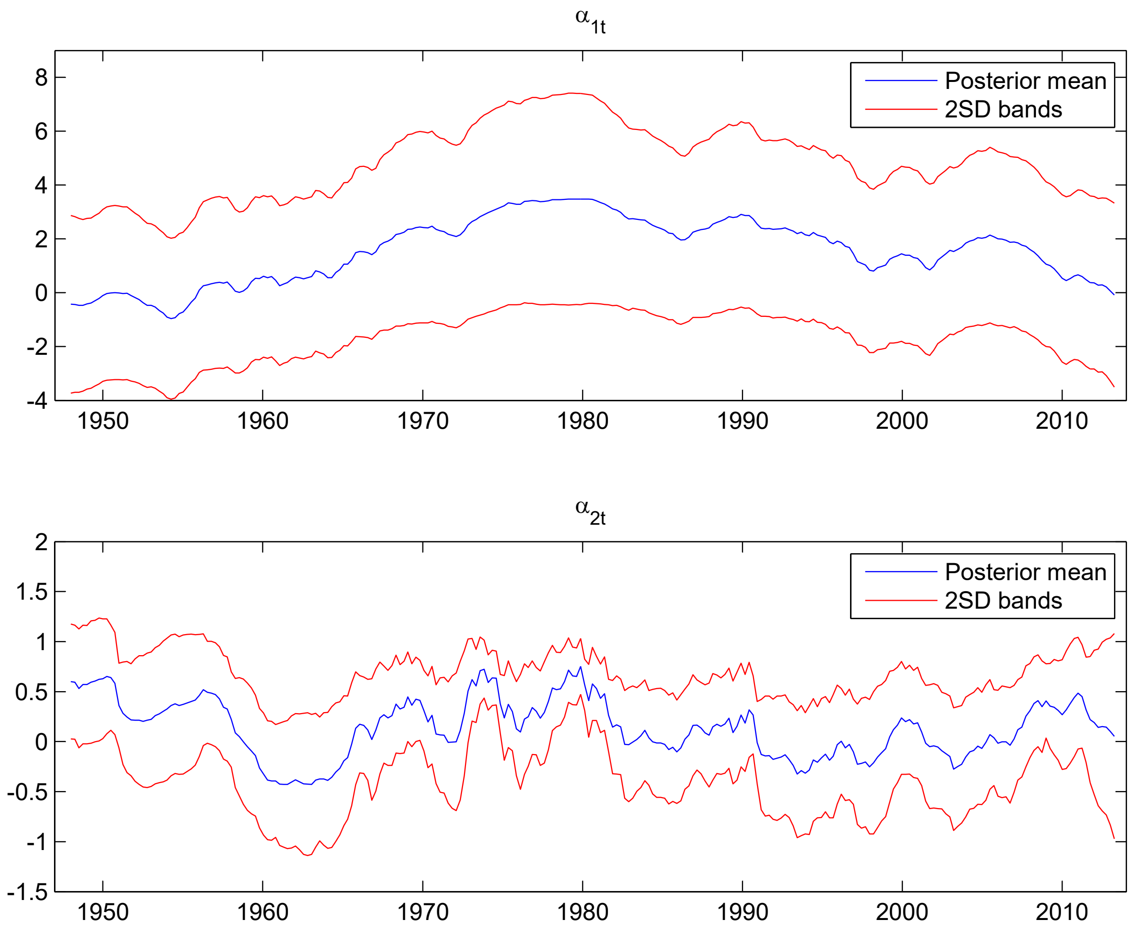

Figure 4: Evolution of $\boldsymbol{\alpha}_t$ obtained from the AR-S-TVP-SV model; posterior mean (blue), two standard deviation bands (red).

5 Conclusions

We proposed a novel Bayesian semiparametric time-varying parameter regression model with stochastic volatility (TVP-SV), where both the error distributions of the observations and parameter-driven dynamics were left unspecified. The Dirichlet process was used as a prior to these unknown distributions. We devised an efficient Markov chain Monte Carlo algorithm to estimate the model parameters and unknown distributions. An autoregressive version of the proposed model was applied to inflation persistence. The empirical results showed that the proposed model had better fit to the data than competing parametric models. For future research it would be interesting to enrich the proposed semiparametric model with a nonparametric leverage effect, as macro shocks that have the greatest effect on the economy are often asymmetrically distributed. Extending the TVP-SV model of this paper in this direction could prove fruitful relative to existing TVP-SV findings.

Table 1: Empirical results

Model	AR-TVP-SV			AR-S-TVP-SV			AR-St-TVP-SV		
	Mean	CD	IF	Mean	CD	IF	Mean	CD	IF
$\mathbf{\Sigma}_{11}$	0.2157*	0.642	63.84	0.2808*	0.771	59.20	0.1544*	0.846	37.68
	(0.1076)			(0.1590)			(0.0573)		
$\mathbf{\Sigma}_{22}$	0.0483*	0.271	17.92	0.0679*	0.017	26.14	0.0358*	0.347	6.09
	(0.0114)			(0.0216)			(0.0070)		
ϕ	0.9586*	0.079	53.77	0.9655*	0.004	47.31	0.9674*	0.000	13.03
	(0.0276)			(0.0232)			(0.0233)		
μ_h	0.3664	0.630	7.48				0.2110	0.933	9.22
	(0.9169)						(0.8628)		
σ_η	0.3964*	0.207	115.85	0.2577*	0.009	113.23	0.2964*	0.010	70.76
	(0.1022)			(0.0830)			(0.0633)		
M				4.3764*	0.333	75.07			
				(2.4396)					
M_ω				4.0796*	0.477	22.67			
				(2.1991)					
$v1$							9.3842*	0.242	142.56
							(14.0543)		
$v2$							65.8133*	0.088	50.84
							(21.5601)		
CV	0.4909			0.5234			0.5122		
DIC	2501.9			1948.2			2018.8		

*Significant based on the 95% highest posterior density interval. Standard errors in parentheses.

References

Chan JCC. 2015. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics, forthcoming.*

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.

Chiu CWJ, Haroon M, Pinter G. 2014. Fat-tails in VAR models. *Working paper.*

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

Cogley T, Sargent TJ. 2005. Monetary policy and learning the conquest of US inflation: Learning and robustness to model uncertainty. *Review of Economic Dynamics* **8**(2): 528 – 563.

D'Agostino A, Gambetti L, Giannone D. 2013. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* **28**(1): 82–101.

De Jong P, Shephard N. 1995. The simulation smoother for time series models. *Biometrika* **82**(2): 339–350.

Delatola EI, Griffin JE. 2011. Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Analysis* **6**(4): 1–26.

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

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

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).

Ghosal S. 2010 . *The Dirichlet process, related priors and posterior asymptotics.* in Hjort N, Holmes C, Mller P, Walker SG. (Eds.). Bayesian nonparametrics: principles and practice. Cambridge University Press.

Jensen M, Maheu JM. 2010. Bayesian semiparametric stochastic volatility modelling. *Journal of Econometrics* **157**(2): 306–316.

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

Lanne M. 2015. Noncausality and inflation persistence. *Studies in Nonlinear Dynamics and Econometrics* **19**(4): 469481.

Lanne M, Luoto J, Saikkonen P. 2012. Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting* **28**(3): 623 – 631.

Lanne M, Saikkonen P. 2011. Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics.*

MacEachern S, Müller P. 1998. Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7**(2): 223–238.

Nakajima J. 2011. Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications. *Monetary and Economic Studies* **29**: 107–142.

Primiceri GE. 2005. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies* **72**(3): 821–852.

Sims CA, Zha T. 2006. Were there regime switches in U.S. monetary policy? *The American Economic Review* **96**(1): 54–81.

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

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: Semiparametric Bayesian inference for time-varying parameter regression models with stochastic volatility

Dimitrakopoulos Stefanos[1]

[1]Department of Economics, Warwick University, Coventry, CV4 7ES, UK

1 MCMC algorithm for the S-TVP-SV model

Part 1

Posterior sampling of $\boldsymbol{\beta}$

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

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

where

$$D_0 = \left(\mathbf{B}^{-1} + \sum_{t=1}^{T} \frac{\mathbf{x}_t \mathbf{x}_t'}{\exp(h_t)\lambda_t^2}\right)^{-1}, \; d_0 = \mathbf{B}^{-1}\boldsymbol{\beta}_0 + \sum_{t=1}^{T} \frac{\mathbf{x}_t(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t - \mu_t)}{\exp(h_t)\lambda_t^2}.$$

Posterior sampling of $\boldsymbol{\Sigma}$

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

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

Posterior sampling of σ_η^2

Update σ_η^2 by sampling from

$$\sigma_\eta^2|v_a, v_\beta, \phi, \mathbf{h} \sim \mathcal{IG}\left(\frac{v_a + T}{2}, \frac{v_\beta + {h_1}^2(1-\phi^2) + \sum_{t=1}^{T-1} (h_{t+1} - \phi h_t)^2}{2}\right).$$

Posterior sampling of $\boldsymbol{\alpha}$

Update the time-varying parameters by applying the simulation smoother of De Jong and Shephard (1995) to the following model

$$\tilde{y}_t = \mathbf{z}_t'\boldsymbol{\alpha}_t + \exp(h_t/2)\lambda_t \epsilon_t, \; \epsilon_t \sim N(0,1), \; t = 1, ..., T,$$

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

where $\tilde{y}_t = y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mu_t$.

Posterior sampling of h

Apply the sampler of Chan (2015) to the following model

$$y_t^\star = \exp(h_t/2)\epsilon_t,\ \epsilon_t \sim N(0,1),\ t = 1,...,T, \tag{A.1}$$
$$h_{t+1} = \phi h_t + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0,\sigma_\eta^2), \tag{A.2}$$

with $cov(\epsilon_t, \eta_t) = 0$ and $y_t^\star = \frac{y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t - \mu_t}{\lambda_t}$.

To be more specific, the posterior distribution of the volatility vector $\mathbf{h}$ is given by

$$p(\mathbf{h}|\phi, \sigma_\eta^2, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{y}^\star) \propto p(\boldsymbol{y}^\star|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{h})p(\mathbf{h}|\phi, \sigma_\eta^2), \tag{A.3}$$

where $\boldsymbol{y}_t^\star = (y_1^\star, ..., y_T^\star)$. In order to sample from the posterior distribution $p(\mathbf{h}|\phi, \sigma_\eta^2, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{y}^\star)$, we approximate it by a Gaussian distribution, which is then used as a proposal density within the Acceptance-Rejection Metropolis-Hastings (ARMH) algorithm (see, for example, Tierney (1994) and Chib and Greenberg (1995)). Candidate draws from the Gaussian approximation are generated using the precision sampler of Chan and Jeliazkov (2009), instead of Kalman-filter based methods.

In particular, it can be shown that the density $p(\mathbf{h}|\phi, \sigma_\eta^2)$ in expression (A.3) is Gaussian, that is, $\mathbf{h}|\phi, \sigma_\eta^2 \sim N(\boldsymbol{H}^{-1}\hat{\mathbf{h}}, (\boldsymbol{H}'\Sigma^{-1}\boldsymbol{H})^{-1})$, where $\hat{\mathbf{h}} = (0, ..., 0)'$, $\Sigma = diag(\sigma_\eta^2/(1-\phi^2), \sigma_\eta^2, ..., \sigma_\eta^2)$ and $\boldsymbol{H}$ is a lower triangular sparse matrix (with determinant 1-hence, it is invertible)

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

The logarithm of $p(\mathbf{h}|\phi, \sigma_\eta^2)$ can be written as

$$\log p(\mathbf{h}|\phi, \sigma_\eta^2) = const - \tfrac{1}{2}(\mathbf{h}'\boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}\mathbf{h} - 2\mathbf{h}'\boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}\boldsymbol{H}^{-1}\hat{\mathbf{h}}). \tag{A.4}$$

The density $p(\boldsymbol{y}^\star|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{h})$ in expression (A.3) can also be approximated by a normal density. By taking the second order Taylor expansion of the logarithm of $p(\boldsymbol{y}^\star|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{h})$ around $\tilde{\mathbf{h}}$, which is the mode of the posterior $\log p(\mathbf{h}|\phi, \sigma_\eta^2, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{y}^\star)$ (see below), we have,

$$\begin{aligned}\log p(\boldsymbol{y}^\star|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{h}) &\approx \log p(\boldsymbol{y}^\star|\boldsymbol{\beta}, \boldsymbol{\theta}, \tilde{\mathbf{h}}) + (\mathbf{h} - \tilde{\mathbf{h}})'\mathbf{f} - \tfrac{1}{2}(\mathbf{h} - \tilde{\mathbf{h}})'\mathbf{G}(\mathbf{h} - \tilde{\mathbf{h}}), \\ &= const - \tfrac{1}{2}(\mathbf{h}'\mathbf{G}\mathbf{h} - 2\mathbf{h}'(\mathbf{f} + \mathbf{G}\tilde{\mathbf{h}})),\end{aligned} \tag{A.5}$$

where $\mathbf{f} = (f_1, ..., f_T)'$ is the gradient vector with $f_t = \frac{d \log p(y_t^\star|\boldsymbol{\beta}, \vartheta_t, h_t)}{dh_t} = -\frac{1}{2} + \frac{1}{2}y_t^{2\star}\exp(-h_t)$

evaluated at $\tilde{h}_t$, $t = 1, ..., T$ and $\mathbf{G} = diag(G_1, ..., G_T)$ is the diagonal negative Hessian of the $\log p(\boldsymbol{y}^\star|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{h})$, with $G_t = -\frac{d^2 \log p(y_t^\star|\boldsymbol{\beta}, \vartheta_t, h_t)}{dh_t^2} = \frac{1}{2} y_t^{2\star} \exp(-h_t)$ evaluated at $\tilde{h}_t$, $t = 1, ..., T$.

From (A.4) and (A.5) the logarithm of the posterior distribution of the volatility vector becomes

$$\log p(\mathbf{h}|\phi, \sigma_\eta^2, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{y}^\star) \approx const - \tfrac{1}{2}(\mathbf{h}'\boldsymbol{K_h}\mathbf{h} - 2\mathbf{h}'\boldsymbol{k_h}) = \log\ g(\mathbf{h}), \tag{A.6}$$

where $\boldsymbol{K_h} = \boldsymbol{H}'\Sigma^{-1}\boldsymbol{H} + \mathbf{G}$, $\boldsymbol{k_h} = \mathbf{f} + \mathbf{G}\tilde{\mathbf{h}} + \boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}\boldsymbol{H}^{-1}\hat{\mathbf{h}}$ and $g(\mathbf{h}) \propto N(\hat{\boldsymbol{m}}, \boldsymbol{K_h}^{-1})$, with $\hat{\boldsymbol{m}} = \boldsymbol{K_h}^{-1}\boldsymbol{k_h}$. In other words, the posterior $p(\mathbf{h}|\phi, \sigma_\eta^2, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{y}^\star)$ can be approximated by a normal density with mean $\hat{\boldsymbol{m}}$ and precision matrix $\boldsymbol{K_h}$. This Gaussian approximation is then used as a proposal density in the ARMH step, where candidate values are obtained using the precision sampler of Chan and Jeliazkov (2009), instead of Kalman-filter based methods.

Typically, $N(\hat{\boldsymbol{m}}, \boldsymbol{K_h}^{-1})$ is a high-dimensional density and sampling from it can be time-consuming. Here, we use the precision-based sampler of Chan and Jeliazkov (2009), which exploits the fact that the precision matrix $\boldsymbol{K_h}$ is a band matrix since $\boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}$ and $\mathbf{G}$ are also band matrices. In particular, $\boldsymbol{K_h}$ is a tridiagonal matrix as its non-zero elements appear only on the main diagonal and the diagonals above and below the main one. Consequently, we can compute fast and efficiently the mean $\hat{\boldsymbol{m}}$ without calculating the inverse $\boldsymbol{K_h}^{-1}$, by solving the linear system $\boldsymbol{K_h}\hat{\boldsymbol{m}} = \boldsymbol{k_h}$. Furthermore, a draw $\tilde{\boldsymbol{m}}$ from $N(\hat{\boldsymbol{m}}, \boldsymbol{K_h}^{-1})$ can be obtained, using the precision sampler of Chan and Jeliazkov (2009): calculate the Cholesky factor $\boldsymbol{C}$ of $\boldsymbol{K_h}$ such that $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{K_h}$, sample T independent standard normal draws, $\boldsymbol{z} \sim N(0, \boldsymbol{I})$, solve $\boldsymbol{C}'\boldsymbol{x} = \boldsymbol{z}$ for $\boldsymbol{x}$ by backward substitution and return $\tilde{\boldsymbol{m}} = \hat{\boldsymbol{m}} + \boldsymbol{x}$.

The point $\tilde{\mathbf{h}}$ around which the second order Taylor expansion is taken in expression (A.5) is desirable to be the mode of the posterior $\log p(\mathbf{h}|\phi, \sigma_\eta^2, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{y}^\star)$ for an efficient sampling. The negative Hessian of this posterior distribution evaluated at $\mathbf{h} = \tilde{\mathbf{h}}$ is $\boldsymbol{K_h}$ and the gradient evaluated at $\mathbf{h} = \tilde{\mathbf{h}}$ is -$\boldsymbol{K_h}\tilde{\mathbf{h}}$ +$\boldsymbol{k_h}$. To find the mode, we apply the Newton-Raphson method as follows: 1) Initialize $\mathbf{h} = \tilde{\mathbf{h}}^{(1)}$ for some constant vector $\tilde{\mathbf{h}}^{(1)}$. 2) Set $\tilde{\mathbf{h}} = \tilde{\mathbf{h}}^{(l)}$ for $l = 1, 2, ...$, and compute $\boldsymbol{K_h}$, $\boldsymbol{k_h}$ and $\mathbf{h}^{(l+1)} = \mathbf{h}^{(l)} + \boldsymbol{K_h}^{-1}(-\boldsymbol{K_h}\mathbf{h}^{(l)} + \boldsymbol{k_h}) = \boldsymbol{K_h}^{-1}\boldsymbol{k_h}$. This process is repeated until convergence is achieved.

Posterior sampling of ϕ

Update ϕ using an independence Metropolis-Hastings algorithm, as in Kim et al. (1998).

Posterior sampling of ε_t

Calculate ε_t from $\varepsilon_t = y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t$, $t = 1, ..., T$.

Posterior sampling of $\mathbf{u}_t$

Calculate $\mathbf{u}_t$ from $\mathbf{u}_t = \boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t$, $t = 1, ..., T-1$.

Part 2

Posterior sampling of $\{\psi_t\}$ **and** $\{\vartheta_m^*\}$

Since $\vartheta_t = (\mu_t, \lambda_t^2) \overset{iid}{\sim} G$, with G being a random (discrete) distribution generated from a DP prior (see subsection 2.2 of the main paper), the vector $\boldsymbol{\theta}$ will contain ties. Let $\boldsymbol{\theta}^* = (\vartheta_1^*, ..., \vartheta_M^*)'$, $M \leq T$ be the set of unique values from $\boldsymbol{\theta}$. Instead of simulating $\boldsymbol{\theta}$, we sample the vector of unique values $\boldsymbol{\theta}^*$ and the vector of the latent indicator variables $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_T)'$, where $\psi_t = m$ when $\vartheta_t = \vartheta_m^*$, $m = 1, ..., M$. This reparametrisation improves mixing (MacEachern, 1994).

Let the vector $\boldsymbol{\theta}^{(t)}$ denote all the elements of $\boldsymbol{\theta}$ with ϑ_t removed. The vector $\boldsymbol{\theta}^{(t)}$ will contain $M^{(t)}$ clusters, that is, $\boldsymbol{\theta}^{*(t)} = (\vartheta_1^{*(t)}, ..., \vartheta_{M^{(t)}}^{*(t)})'$, where $M^{(t)}$ is the number of unique values in $\boldsymbol{\theta}^{(t)}$. The number of elements in $\boldsymbol{\theta}^{(t)}$ that take the distinct value $\vartheta_m^{*(t)}$ will be $n_m^{(t)} = \sum_j \mathbf{1}(\psi_j = m,\, j \neq t)$, $m = 1, ..., M^{(t)}$.

The sampler for updating $\{\psi_t\}$ and $\{\vartheta_m^*\}$ consists of two steps.

<u>**Step 1**</u>: Sample each ψ_t according to the probabilities

$$P(\psi_t = m|\boldsymbol{\theta}^{*(t)}, \psi^{(t)}, n_m^{(t)}, h_t) \propto \begin{cases} q_{tm} & \text{if} \quad m = 1, ..., M^{(t)} \\ q_{t0} & \text{if} \quad m = M^{(t)}+1, \end{cases} \tag{A.7}$$

where $\psi^{(t)} = \boldsymbol{\psi} \backslash \{\psi_t\}$ and the weights q_{t0} and q_{tm} in $(A.7)$ are defined respectively as

$$q_{t0} \propto a \int f(\varepsilon_t|h_t, \vartheta_t) dG_0(\vartheta_t), \; q_{tm} \propto n_m^{(t)} f(\varepsilon_t|h_t, \vartheta_m^{*(t)}).$$

From $(A.7)$, ψ_t can take the value $m = 1, ..., M^{(t)}$ with probability proportional to q_{tm}. In this case ϑ_t, $t = 1, ..., T$, is assigned to an existing cluster $\vartheta_m^{*(t)}$, $m = 1, ..., M^{(t)}$. The term q_{tm} is proportional to $n_m^{(t)}$ times the normal distribution of ε_t evaluated at $\vartheta_m^{*(t)}$, that is, $q_{tm} \propto n_m^{(t)} \exp(-\frac{1}{2}\left(\varepsilon_t - \mu_m^{*(t)}\right)^2 / \exp(h_t)\lambda_m^{*2,(t)})$.

Also from $(A.7)$, ψ_t can take a new value $m = M^{(t)} + 1$ with probability proportional to q_{t0}. In this case, we set $\vartheta_t = \vartheta_{M^{(t)}+1}^*$ and sample $\vartheta_{M^{(t)}+1}^*$ from the posterior baseline distribution

$$\vartheta_t = (\mu_t, \lambda_t^2)|\varepsilon_t, h_t, \mu_0, \tau_0, e_0, f_0 \sim N(\mu_t|\overline{\mu_0}, \overline{\tau_0}\lambda_t^2)\mathcal{IG}(\lambda_t^2|\tfrac{\overline{e_0}}{2}, \tfrac{\overline{f_0}}{2}),$$

where
$$\overline{\mu_0} = \frac{\mu_0 + \exp(-h_t)\tau_0\varepsilon_t}{1+\exp(-h_t)\tau_0}, \qquad \overline{\tau_0} = \frac{\tau_0}{1+\exp(-h_t)\tau_0},$$

$$\overline{e_0} = e_0 + 1, \qquad \overline{f_0} = f_0 + \frac{(\varepsilon_t - \mu_0)^2}{\tau_0 + \exp(h_t)}.$$

The term q_{t0} is proportional to the concentration parameter a times the marginal density of ε_t. This density is equal to the Student-t distribution $q_t(\varepsilon_t|\mu_0, (\exp(h_t) + \tau_0)f_0/e_0, e_0)$, where μ_0 is the mean, e_0 is the degrees of freedom and the remaining term $(\exp(h_t)+\tau_0)f_0/e_0$ is the scale factor.

<u>**Step 2**</u>:
Sample ϑ_m^*, $m = 1, ..., M$ from the following baseline posterior

$$\vartheta_m^* = (\mu_m^*, \lambda_m^{*2})|\{\varepsilon_t\}_{t\in F_m}, \{h_t\}_{t\in F_m}, \mu_0, \tau_0, e_0, f_0 \sim N(\mu_m^*|\overline{\mu_m}, \overline{\tau_m}\lambda_m^{*2})\mathcal{IG}(\lambda_m^{*2}|\tfrac{\overline{e_m}}{2}, \tfrac{\overline{f_m}}{2}),$$

where
$$\overline{\mu_m} = \frac{\mu_0+\tau_0 \sum\limits_{t\in F_m} \varepsilon_t \exp(-h_t)}{1+\tau_0 \sum\limits_{t\in F_m} \exp(-h_t)}, \qquad \overline{\tau_m} = \frac{\tau_0}{1+\tau_0 \sum\limits_{t\in F_m} \exp(-h_t)},$$

$$\overline{e_m} = e_0 + n_m, \quad \overline{f_m} = f_0 + \frac{(\tilde{\varepsilon}_t-\mu_0)^2}{\tau_0+\sum\limits_{t\in F_m} \exp(h_t)} + \sum_{t\in F_m} [\exp(-h_t/2)(\varepsilon_t - \tilde{\varepsilon}_t)]^2,$$

$$\tilde{\varepsilon}_t = \frac{\sum\limits_{t\in F_m} \varepsilon_t \exp(-h_t)}{\sum\limits_{t\in F_m} \exp(-h_t)},$$

and $F_m = \{t : \vartheta_t = \vartheta_m^*\}$ is the set of ϑs sharing the parameter ϑ_m^*.

Posterior sampling of a

We sample a following Escobar and West (1995). In particular, we first sample ξ, a latent random variable, 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 $\{\omega_t\}_{t=1}^{T-1}$

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

$$p\left(\omega_t|\boldsymbol{\omega}^{(t)}, \mathbf{u}_t, e_\omega, \boldsymbol{\Sigma}, G_{0\omega}\right) \propto \tilde{q}_{t0} p(\omega_t|\mathbf{u}_t, e_\omega, \boldsymbol{\Sigma}) + \sum_{m_\omega=1}^{M_\omega^{(t)}} \tilde{q}_{tm_\omega} \delta_{\omega_{m_\omega}^{*(t)}}(\omega_t),$$
$$t = 1, ..., T-1, \qquad (A.8)$$

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.8) but instead update the latent indicators in an analogous way to that for ϑ_t's and resample the clusters $\omega_{m_\omega}^*$, $m_\omega = 1, ..., M_\omega$ from the posterior gamma distribution

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

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

Updating a_ω is similar to the updating of the precision parameter a.

2 Model comparison

The DIC (Spiegelhalter et al., 2002) adds together the fit of the model and its complexity and is defined as DIC $= \overline{D(\mathbf{\Theta})} + p_D$. Model fit is measured by the deviance $D(\mathbf{\Theta}) = -2\log f(\boldsymbol{y}|\mathbf{\Theta})$ where $logf(\boldsymbol{y}|\mathbf{\Theta})$ is the log-likelihood function and $\mathbf{\Theta}$ denotes the vector of all parameters in the model, that is, $\mathbf{\Theta}$=($\boldsymbol{\beta}$, $\mathbf{\Sigma}$, σ^2_η, $\boldsymbol{\alpha}$, $\mathbf{h}$, $\phi, \boldsymbol{\theta}, \boldsymbol{\omega}, a, a_\omega)'$. $\overline{D(\mathbf{\Theta})} = -2\mathbf{E}_{\mathbf{\Theta}}[\log f(\boldsymbol{y}|\mathbf{\Theta})|\boldsymbol{y}]$ is the posterior mean deviance. Model complexity is measured by the effective number of model parameters and is defined as $p_D = \overline{D(\mathbf{\Theta})} - D(\overline{\mathbf{\Theta}})$ where $D(\overline{\mathbf{\Theta}}) = -2\log f(\boldsymbol{y}|\overline{\mathbf{\Theta}})$ and $\log f(\boldsymbol{y}|\overline{\mathbf{\Theta}})$ is the log-likelihood evaluated at $\overline{\mathbf{\Theta}}$, the posterior mean of $\mathbf{\Theta}$. Therefore, DIC $= 2\overline{D(\mathbf{\Theta})} - D(\overline{\mathbf{\Theta}})$. The smaller the DIC, the better the model fit.

The deviance for the S-TVP-SV model is

$$D(\mathbf{\Theta}) = -2\sum_{t=1}^T \log\left(\frac{1}{\sqrt{2\pi\lambda_t^2\exp(h_t)}} exp\left(-\frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t - \mu_t)^2}{2\lambda_t^2\exp(h_t)}\right)\right).$$

We also apply the leave-one-out cross validation (CV) method that requires the calculation of the conditional predictive ordinate (CPO),

$$CPO_t = f(y_t|y_{-t}) = \int f(y_t|\mathbf{\Theta})f(\mathbf{\Theta}|y_{-t}) = \mathbf{E}_{\mathbf{\Theta}|y_{-t}}[f(y_t|\mathbf{\Theta})],\ t = 1, ..., T,$$

where $y_{-t} = \boldsymbol{y} \setminus \{y_t\}$. Gelfand and Dey (1994) and Gelfand (1996) proposed a Monte carlo integration of CPO. More specifically,

$$C\hat{P}O_t = \hat{f}(y_t|y_{-t}) = \left(\frac{1}{M}\sum_{m=1}^M \left(f(y_t|y_{-t}, \mathbf{\Theta}^{(m)})\right)^{-1}\right)^{-1},$$

where

$$f(y_t|y_{-t}, \mathbf{\Theta}^{(m)}) = \frac{1}{\sqrt{2\pi\lambda_t^{2(m)}\exp(h_t^{(m)})}} exp\left(-\frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta}^{(m)} - \mathbf{z}_t'\boldsymbol{\alpha}_t^{(m)} - \mu_t^{(m)})^2}{2\lambda_t^{2(m)}\exp(h_t^{(m)})}\right),$$

and M is the number of iterations after the burn-in period. Then for each model we calculate the average of the estimated CPO values, $\frac{1}{T}\sum_{t=1}^T \hat{f}(y_t|y_{-t})$. Higher values of this average are associated with better "goodness of fit" of a model.

3 The St-TVP-SV model

Consider the following TVP-SV model,

$$y_t = \mu + \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \exp(h_t/2)\sqrt{\rho_{1t}}z_t,\ z_t \sim N(0,1)$$

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

$$h_{t+1} = \mu_h + \phi h_t + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0, \sigma_\eta^2),$$

where ρ_{1t} follows an inverse gamma prior, $\rho_{1t} \sim \mathcal{IG}(v1/2, v1/2)$ and $v1$ follows a uniform prior distribution on the set [3, 120]. The marginal distribution of $e_t = \sqrt{\rho_{1t}}z_t$ is a Student-t distribution with $v1$ degrees of freedom[1].

Furthermore, the positive scale variable ρ_{2t} follows a Gamma distribution, $\rho_{2t} \sim \mathcal{G}(v2/2, v2/2)$ and $v2$ is uniformly distributed on the set [3, 120]. The unconditional distribution of $\mathbf{u}_t$ is a multivariate Student-t distribution.

To update $v1$ and $v2$ we use independence Metropolis-Hastings steps, where the candidate draws are generated from a truncated normal proposal density in the interval $[3, 120]$.

To update the volatilities we apply the approach of Chan (2015), where now $\hat{\mathbf{h}} = (\mu_h/(1-\phi), \mu_h, ..., \mu_h)'$.

4 Monte Carlo experiments

In this section we evaluate the efficiency of the proposed MCMC algorithm for the S-TVP-SV model.

We generated T=500 data points from the proposed model setting $k = 2$ (number of fixed coefficients), $p = 2$ (number of time-varying coefficients) and assuming also the following set of true parameter values

$$\boldsymbol{\beta} = (3, 0.8)', \boldsymbol{\Sigma} = diag(2, 2), \phi = 0.8, \boldsymbol{\alpha}_1 = (-2, 7)',\ \sigma_\eta = 0.1,$$

where $diag(\cdot)$ is a diagonal matrix. The elements of $\mathbf{x}_t = (x_{1t}, x_{2t})'$ and $\mathbf{z}_t = (z_{1t}, z_{2t})'$ for $t = 1, ..., T$ are generated respectively as $x_{jt} \sim U(0,1)$ and $z_{it} \sim 0.1U(0,1)$ for $j, i = 1, 2$, where $U(a, b)$ is the uniform distribution defined on the domain (a, b).

Regarding the generation of the true innovations for the regressions of y_t and of $\boldsymbol{\alpha}_t$, we examine the following case (case 1):

- ε_t =exp$(h_t/2)\epsilon_t$, $\epsilon_t \sim 0.5N(0.2, 1) + 0.5N(-1, 6)$, with $\mu = Exp(\epsilon_t) = -0.4$
- $\mathbf{u}_t \sim MVt(\mathbf{0}, \boldsymbol{\Sigma}, 4)$, where MVt is the multivariate-t distribution with mean $\mathbf{0}$, covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom 4.

[1] From a Bayesian perspective, stochastic volatility models with fat tails have been considered by Jacquier et al. (2004), among others.

We also considered the following case (case 2): T=1500 and the following set of true parameter values

$$\boldsymbol{\beta} = (-1, 3)', \boldsymbol{\Sigma} = diag(5, 5), \phi = 0.8, \boldsymbol{\alpha}_1 = (2, -2)',\ \sigma_\eta = 0.5,$$

with $x_{jt} \sim 2U(0, 1)$, $z_{it} \sim N(2, 0.01)$ for $j, i = 1, 2$ and
- ε_t =exp$(h_t/2)\epsilon_t$, $\epsilon_t \sim 0.5N(-1, 1) + 0.5N(3, 3)$, with $\mu = Exp(\epsilon_t) = 1$
- $\mathbf{u}_t \sim MVt(\mathbf{0}, \boldsymbol{\Sigma}, 6)$.

In either case, we assume the following proper but sufficiently diffuse prior distributions

$$\boldsymbol{\beta} \sim N(0, 20 \times I_{2\times2}),\ \boldsymbol{\alpha}_1 \sim N(0, 20 \times I_{2\times2}),\ \sigma^2_\eta \sim \mathcal{IG}(50/2, 0.5/2),$$

$$(\phi + 1)/2 \sim Beta(80, 14),\ \boldsymbol{\Sigma} \sim IW(1, 20 \times I_{2\times2}),$$

$$\mu_t \sim N(0, 4 \times \lambda^2_t),\ \lambda^2_t \sim I\mathcal{G}(5/2, 5/2), \quad \omega_t \sim \mathcal{G}(50/2, 50/2),$$

where $I_{2\times2}$ is a 2×2 identity matrix.

After discarding the first 30000 draws, we run the sampler 60000. To monitor convergence we use the CD statistics of Geweke (1992) and the inefficiency factor (IF); see, for example, Chib (2001).

Tables 1 (case 1) and 2 (case 2) report the posterior estimates of the mean and standard deviation for all the parameters of the S-TVP-SV model. In both cases, the S-TVP-SV model produces satisfactory estimation accuracy.

Furthermore, in Figures 1 (case 1) and 2 (case 2), the posterior means of α_{1t} and α_{2t}, follow closely the path of their corresponding true values. Also, almost all the true states fall inside the two standard deviation bands.

In Figures 3 (case 1) and 4 (case 2) we plotted the true and the estimated out-of-sample posterior predictive distribution of the error term ε obtained from the semiparametric model. As can be seen, the S-TVP-SV model is able to mimic well the true nonnormal distribution.

The sampler of the S-TVP-SV model performs well overall.

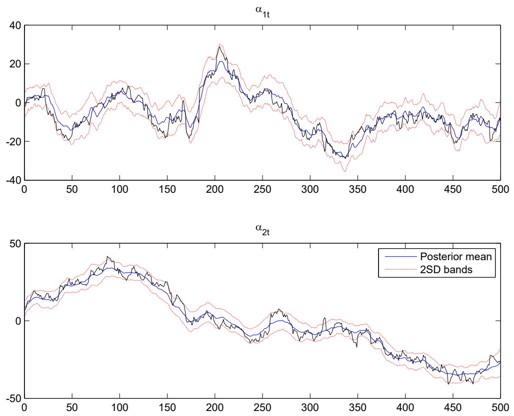

Figure 1: Simulated data: Path of the estimated α_{1t} and α_{2t} for the S-TVP-SV model; T=500. True path (black), posterior mean (blue), two standard deviation bands (red).

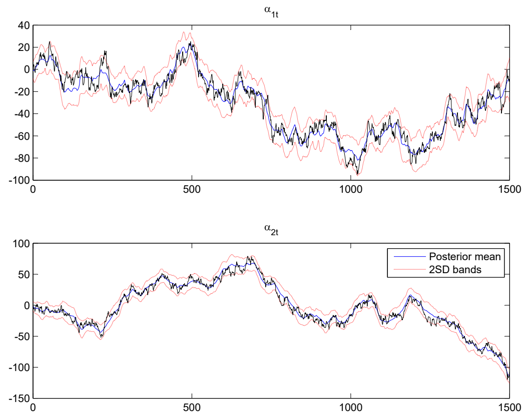

Figure 2: Simulated data: Path of the estimated α_{1t} and α_{2t} for the S-TVP-SV model; T=1500. True path (black), posterior mean (blue), two standard deviation bands (red).

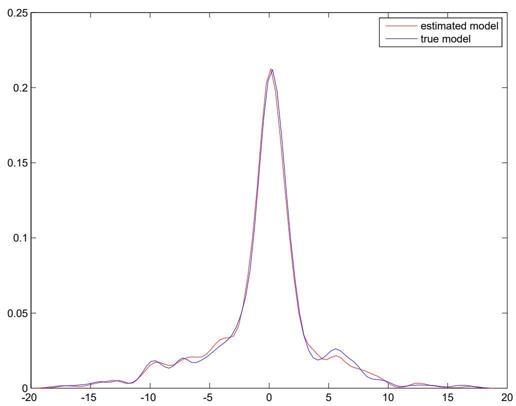

Figure 3: Simulated data. True and estimated out-of-sample posterior predictive distribution of the error term ε obtained from the S-TVP-SV model; T=500.

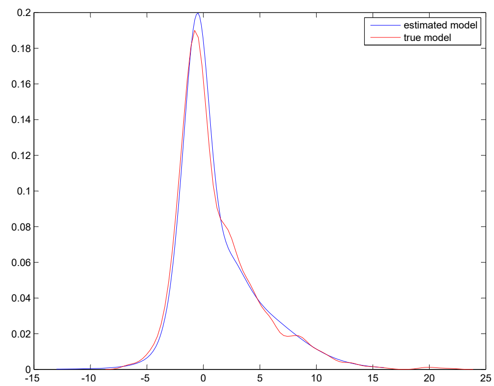

Figure 4: Simulated data. True and estimated out-of-sample posterior predictive distribution of the error term ε obtained from the S-TVP-SV model; T=1500.

Table 1: Simulated data: T=500

Model	S-TVP-SV			
True values	Mean	Stdev	CD	IF
$\beta_1 = 3$	2.9775	0.3637	0.549	20.05
$\beta_2 = 0.8$	1.0464	0.2714	0.941	23.81
$\mathbf{\Sigma}_{11} = 2$	2.9195	3.0824	0.073	52.52
$\mathbf{\Sigma}_{22} = 2$	1.5804	1.6299	0.962	58.35
$\phi = 0.8$	0.7976	0.0730	0.148	21.56
$\sigma_\eta = 0.1$	0.1041	0.0110	0.972	28.49

Table 2: Simulated data: T=1500

Model	S-TVP-SV			
True values	Mean	Stdev	CD	IF
$\beta_1 = -1$	-0.9750	0.1140	0.001	29.19
$\beta_2 = 3$	2.9770	0.0924	0.007	40.44
$\mathbf{\Sigma}_{11} = 5$	6.3537	1.4614	0.647	61.05
$\mathbf{\Sigma}_{22} = 5$	6.5565	2.6642	0.961	61.63
$\phi = 0.8$	0.8582	0.0677	0.001	128.52
$\sigma_\eta = 0.5$	0.3496	0.0156	0.024	128.86

5 Empirical results

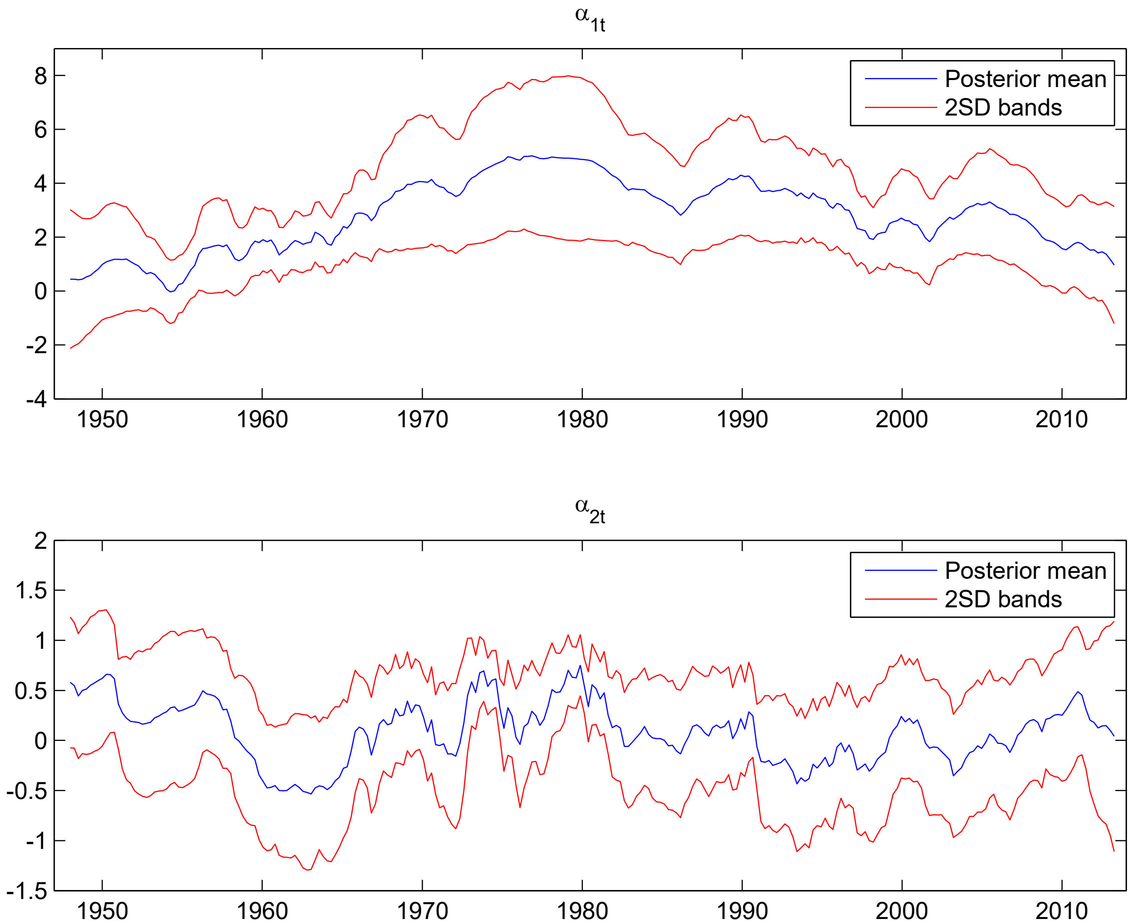

Figure 5: Evolution of $\boldsymbol{\alpha}_t$ obtained from the AR-TVP-SV model; posterior mean (blue), two standard deviation bands (red).

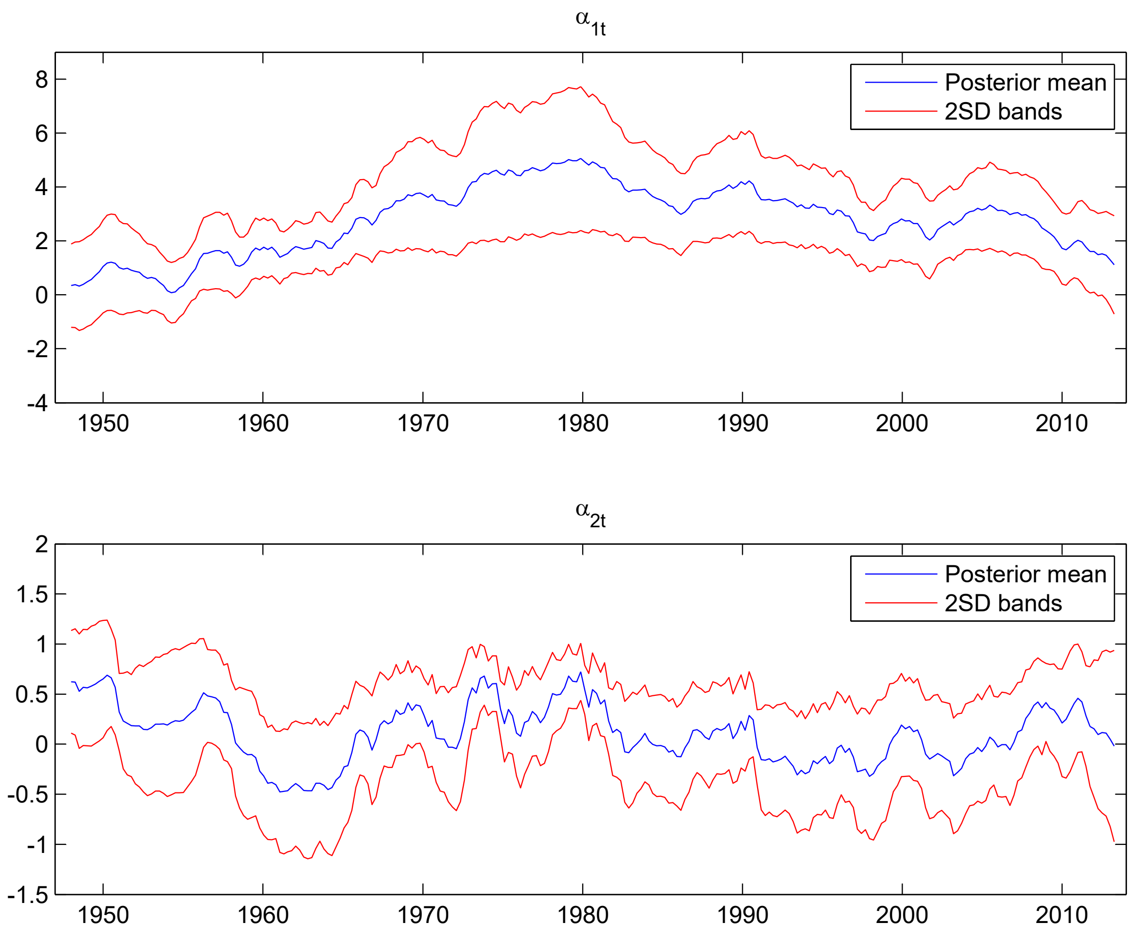

Figure 6: Evolution of $\boldsymbol{\alpha}_t$ obtained from the AR-St-TVP-SV model; posterior mean (blue), two standard deviation bands (red).

References

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

Chan JCC. 2015. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics, forthcoming.*

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.

De Jong P, Shephard N. 1995. The simulation smoother for time series models. *Biometrika* **82**(2): 339–350.

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

Gelfand AE. 1996. Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice,* ed. W. R. Gilks, S. Richardson, D. J. Spiegelhalter. London, UK: Chapman & Hall chapter 9, pp. 145–161.

Gelfand AE, Dey DK. 1994. Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B* **56**(3): 501–514.

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).

Jacquier E, Polson NG, Rossi PE. 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* **122**(1): 185–212.

Kim S, Shephard N, Chib S. 1998. Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies* **65**(3): 361–393.

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

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

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