The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling

Joshua C.C. Chan*
Research School of Economics,
Australian National University

May 2015

Abstract

This paper generalizes the popular stochastic volatility in mean model of Koopman and Hol Uspensky (2002) to allow for time-varying parameters in the conditional mean. The estimation of this extension is nontrival since the volatility appears in both the conditional mean and the conditional variance, and its coefficient in the former is time-varying. We develop an efficient Markov chain Monte Carlo algorithm based on band and sparse matrix algorithms instead of the Kalman filter to estimate this more general variant. We illustrate the methodology with an application that involves US, UK and Germany inflation. The estimation results show substantial time-variation in the coefficient associated with the volatility, highlighting the empirical relevance of the proposed extension. Moreover, in a pseudo out-of-sample forecasting exercise, the proposed variant also forecasts better than various standard benchmarks.

Keywords: nonlinear, state space, inflation forecasting, inflation uncertainty

JEL classification: C11, C15, C53, C58, E31

*We thank seminar participants at University of Melbourne, University of Bucharest, the 8th Rimini Bayesian Econometrics Workshop, Workshop on Empirical Methods in Macroeconomic Policy Analysis 2014 and the 8th International Conference on Computational and Financial Econometrics for helpful comments and suggestions. In particular, this paper has benefited from the discussion with Eric Eisenstat and Ole Maneesoonthorn, among others. Financial support from the Australian Research Council via a Discovery Early Career Researcher Award (DE150100795) is gratefully acknowledged.

1 Introduction

A voluminous literature has demonstrated that structural instability is present in a wide variety of macroeconomic time series (see, e.g., Canova, 1993; Cogley and Sargent, 2001; Koop and Potter, 2007, among many others). One popular approach to model this structural instability is via a time-varying parameter (TVP) model in which the parameters in the conditional mean can evolve gradually over time. A related literature has highlighted the importance of allowing for time-varying volatility in macroeconomic time series, where the heteroscedastic errors are typically modeled using a stochastic volatility specification (see, e.g., Cogley and Sargent, 2005; Primiceri, 2005). For macroeconomic forecasting, D'Agostino, Gambetti, and Giannone (2013) find that both features are crucial in producing accurate forecasts.

We contribute to this line of research by developing a model where the stochastic volatility has a direct and time-varying impact on the variable of interest. More specifically, we build upon the stochastic volatility in mean (SVM) model of Koopman and Hol Uspensky (2002)—originally developed for financial time series as an alternative of the ARCH-M model of Engle, Lilien, and Robins (1987)—in which the volatility enters the conditional mean as a covariate. More recently, the SVM model has been used to fit macroeconomic data, as in Berument, Yalcin, and Yildirim (2009) for investigating the inflation-inflation uncertainty relationship and Mumtaz and Zanetti (2013) for examining the impact of monetary shocks volatility. However, so far all studies are limited to models with constant coefficients. We extend the standard SVM model to allow for time-varying coefficients in the conditional mean as this feature is generally recognized as important for typical macroeconomic time series. In fact, in our application involving US inflation, the TVP variant fits the inflation data better even after penalizing the additional model complexity. Further, it outperforms the standard version with constant coefficients in a recursive forecasting exercise.

A second contribution of this paper is to develop an efficient Markov chain Monte Carlo (MCMC) sampler for estimating this new model. In the original setting with constant coefficients, Koopman and Hol Uspensky (2002) propose a simulated maximum likelihood estimator based on the Kalman filter to fit the model. However, this approach cannot be easily generalized to the time-varying parameter setting as likelihood evaluation would involve "integrating out" both the time-varying coefficients and the volatility by Monte Carlo methods. Instead, we adopt a Bayesian approach and propose an MCMC algorithm to simulate from the joint posterior distribution. A key novel feature of our approach is that it is based on band and sparse matrix algorithms instead of the Kalman filter. This achieves efficiency gains by exploiting the special structure of the problem: the Hessian of the log-conditional density of the log-volatility is a band matrix—i.e., it contains only a few nonzero elements confined along a narrow diagonal band. This feature is essential for reducing the computational costs.

A third contribution involves an empirical application on modeling and forecasting US quarterly consumer price index (CPI) inflation and real-time GDP deflator. Our point

of departure is a version of the unobserved components model with stochastic volatility in Stock and Watson (2007). By adding the SVM component in the conditional mean, the proposed SVM model allows us to study the impact of the volatility feedback—how inflation volatility affects the level of inflation. There is a large literature that examines the relationship between inflation and inflation uncertainty. Empirical work generally supports the claim that higher inflation generates higher inflation uncertainty, as first postulated in the Nobel lecture of Friedman (1977) and formalized in Ball (1992), whereas there is mixed support in the reverse direction. On the one hand, the game-theoretic model of Cukierman and Meltzer (1986) shows that the central bank has an incentive to create inflation surprises in times of high inflation uncertainty to exploit the trade-off between inflation and unemployment—hence, the model predicts higher inflation uncertainty leads to higher inflation. On the other hand, Holland (1995) suggests that if central bankers view inflation uncertainty as costly, they would be induced to reduce inflation uncertainty when inflation is high, thus creating a negative relationship between inflation and inflation uncertainty.

To shed some light on this issue, we use the proposed model to investigate if inflation uncertainty affects the level of inflation using US inflation. In contrast to previous studies that consider only models with constant coefficients in the conditional mean, our time-varying parameter model allows us to assess if the inflation-inflation uncertainty relationship has changed over time. In fact, the coefficient associated with volatility is estimated to be positive in the early sample and it drops substantially or even to negative values since the early 1980s. In addition, we examine if this time-variation can be exploited to improve inflation forecasts. In a pseudo out-of-sample forecasting exercise, we find that the TVP variant outperforms standard benchmarks for both point and density forecasts.

The rest of this article is organized as follows. In Section 2 we first introduce the SVM model with time-varying parameters. Then, an efficient posterior simulator is developed to estimate this new model. Section 3 presents empirical results for modeling and forecasting US inflation. Additional estimation results for UK and Germany data are also provided. In the last section we conclude our findings and discuss some future research directions.

2 SVM with Time-Varying Parameters

In this section we first introduce the time-varying parameter stochastic volatility in mean (TVP-SVM) model that generalizes the original specification in Koopman and Hol Uspensky (2002). In particular, we allow the coefficients in the conditional mean to be time-varying. We then propose an efficient MCMC sampler—based on fast band and sparse operations rather than on the conventional Kalman filter—to estimate this model.

2.1 The Model

We consider a time-varying parameter model with stochastic volatility where the stochastic volatility also enters the conditional mean equation. Specifically, let y_t denote the time series of interest. Then, consider

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_t + \alpha_t \mathrm{e}^{h_t} + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \mathrm{e}^{h_t}), \tag{1}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \varepsilon_t^h, \qquad \varepsilon_t^h \sim \mathcal{N}(0, \sigma^2), \tag{2}$$

where $\mathbf{x}_t$ is a $k \times 1$ vector of covariates, $\boldsymbol{\beta}_t$ is the associated $k \times 1$ vector of time-varying parameters, and the disturbances ε_t^y and ε_t^h are mutually and serially uncorrelated. The log-volatility h_t follows a stationary AR(1) process with $|\phi| < 1$, and it is initialized with $h_1 \sim \mathcal{N}(\mu, \sigma^2/(1-\phi^2))$. The model (1)–(2) generalizes the original setup in Koopman and Hol Uspensky (2002) by allowing the conditional mean of y_t to have time-varying parameters, i.e., both α_t and $\boldsymbol{\beta}_t$ are time-varying. This feature is empirically important for typical macroeconomic applications. The vector of coefficients $\boldsymbol{\gamma}_t = (\alpha_t, \boldsymbol{\beta}_t')'$ in turn evolves according to a random walk process:

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + \boldsymbol{\varepsilon}_t^{\gamma}, \qquad \boldsymbol{\varepsilon}_t^{\gamma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \tag{3}$$

where $\boldsymbol{\Omega}$ is a $(k+1) \times (k+1)$ covariance matrix. Following the literature on time-varying parameters vector autoregressions (e.g., Cogley and Sargent, 2005; Cogley, Primiceri, and Sargent, 2010), we allow for a generic correlation structure among the innovations to the random-walk coefficients. Finally, the random walk process in (3) is initialized with $\boldsymbol{\gamma}_1 \sim \mathcal{N}(\boldsymbol{\gamma}_0, \boldsymbol{\Omega}_0)$ for constant matrices $\boldsymbol{\gamma}_0$ and $\boldsymbol{\Omega}_0$.

Of course, when $\alpha_t = 0$ for all $t = 1, \ldots, T$, the model in (1)–(3) reduces to a standard TVP regression with stochastic volatility. By allowing α_t to be nonzero, the model permits an additional channel of persistence—since the log-volatility follows an AR(1) process, a shock to h_{t-1} would affect h_t, which in turns has a direct impact on the conditional mean of y_t. This channel is shown to be empirically important in both the model comparison exercise and the forecasting exercise below. Also note that (1)–(3) define a Gaussian state space model with two types of states, namely, $\boldsymbol{\gamma}_t$ and h_t. The model is linear in $\boldsymbol{\gamma}_t$ but nonlinear in h_t, and the this nonlinearity makes estimation more difficult.

2.2 Bayesian Estimation

We now introduce an MCMC sampler to simulate from the posterior distribution of the model in (1)–(3). In the original setting with constant coefficients, Koopman and Hol Uspensky (2002) develop a simulated maximum likelihood estimator based on the Kalman filter to fit the SVM model. Specifically, the likelihood—or more precisely the *observed-data* or *integrated* likelihood, obtained by integrating out h_t—can be evaluated using importance sampling, where the importance density is constructed by approximating the *conditional* likelihood—the conditional density of y_t given h_t—using a Gaussian

density. Independent draws from this high-dimensional Gaussian density can be sampled using Kalman filter-based algorithms such as those in Carter and Kohn (1994), Früwirth-Schnatter (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002). Since there are only a few parameters in a constant coefficient SVM model, one can maximize the likelihood numerically to obtain the maximum likelihood estimates.

However, this approach cannot be easily generalized to our time-varying parameter setting as likelihood evaluation in this case would involve "integrating out" both types of states, namely, $\boldsymbol{\gamma}_t$ and h_t, by Monte Carlo methods. Since the states are much higher dimensional in our case, this likelihood evaluation step is in general nontrival. To get around this difficulty, we adopt a Bayesian approach and develop an MCMC algorithm to simulate from the joint posterior distribution. Due to the modular nature of MCMC algorithms, we can simulate each type of states one at a time, which reduces the dimension of the problem and makes estimation much easier.

Another novel feature of our proposed approach is that it builds upon recent advances in band and sparse matrix algorithms rather than using the conventional Kalman filter. Recent papers using the former approach include Rue (2001) for linear Gaussian Markov random fields; Chan and Jeliazkov (2009) and McCausland, Miller, and Pelletier (2011) for linear Gaussian state space models; Rue, Martino, and Chopin (2009) for nonlinear Markov random fields; McCausland (2012), Chan, Koop, and Potter (2013), Chan and Strachan (2014) and Djegnéné and McCausland (2014) for nonlinear state space models. Our new approach achieves efficiency gains by exploiting the special structure of the problem, particularly that the Hessian of the log-conditional density of the log-volatilities (given the data and other parameters) is a band matrix, i.e., it contains only a few nonzero elements arranged along a diagonal band. This feature turns out to be important in developing efficient sampling algorithms.

To elaborate, there are two related notions of efficiency. The first is the computational speed—how much time it takes to obtain a certain number of posterior draws. The second is related to the autocorreleations of the MCMC draws—by construction MCMC draws are autocorrelated; the lower the autocorrelation, the closer they are to the ideal case of *independent* draws. A sampler is efficient if it produces draws with low autocorrelations. The proposed algorithm is efficient in both senses. The first criterion is easy to evaluate; we document the speed of the algorithm at the end of the section. To achieve efficiency in the second sense, the proposed approach samples all the log-volatiliies $h_1, \dots, h_T$ jointly, in contrast to the single-move sampler of Mumtaz and Zanetti (2013) that samples each h_t at a time. (Note that the auxiliary mixture sampler of Kim, Shepherd, and Chib (1998) cannot be applied to the SVM model as h_t also enters the conditional mean equation).

To complete the model specification, we assume independent priors for σ^2, μ, ϕ and $\boldsymbol{\Omega}$:

$$\mu \sim \mathcal{N}(\mu_0, V_\mu), \quad \phi \sim \mathcal{N}(\phi_0, V_\phi)\mathbb{1}(|\phi| < 1), \quad \sigma^2 \sim \mathcal{IG}(\nu_{\sigma^2}, S_{\sigma^2}), \quad \boldsymbol{\Omega} \sim \mathcal{IW}(\nu_{\boldsymbol{\Omega}}, \mathbf{S}_{\boldsymbol{\Omega}}), \tag{4}$$

where $\mathcal{IG}(\cdot, \cdot)$ denotes the inverse-gamma distribution and $\mathcal{IW}(\cdot, \cdot)$ represents the inverse-Wishart distribution. Note that we impose the stationarity condition $|\phi| < 1$ on the

prior for ϕ. For notational convenience, let $\mathbf{x}$ denote the covariates, $\mathbf{y} = (y_1, \ldots, y_T)'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_T')'$ and $\mathbf{h} = (h_1, \ldots, h_T)'$. Then posterior draws can be obtained by sequentially sampling from:

1. $p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2, \boldsymbol{\Omega}) = p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$;
2. $p(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mathbf{x}, \mathbf{h}, \mu, \phi, \sigma^2, \boldsymbol{\Omega}) = p(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mathbf{x}, \mathbf{h}, \boldsymbol{\Omega})$;
3. $p(\boldsymbol{\Omega}, \sigma^2 \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mathbf{h}, \mu, \phi) = p(\boldsymbol{\Omega} \,|\, \boldsymbol{\gamma}) p(\sigma^2 \,|\, \mathbf{h}, \mu, \phi)$;
4. $p(\mu, \phi \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mathbf{h}, \sigma^2, \boldsymbol{\Omega}) = p(\mu, \phi \,|\, \mathbf{h}, \sigma^2)$.

In Step 1 the joint conditional density $p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$ is high-dimensional and non-standard. To simulate from this density, we follow the approach in Chan and Strachan (2014) by exploiting the fact that the Hessian of $\log p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$ is a band matrix. As such, a Gaussian approximation can be quickly obtained, which is then used as a proposal density in an acceptance-rejection Metropolis-Hastings algorithm. Moreover, since the Hessian of this Gaussian proposal density is also a band matrix, we use the precision sampler in Chan and Jeliazkov (2009)—so called as it exploits the band structure of the precision matrix (inverse covariance matrix) of the states—to obtain candidate draws instead of Kalman filter-based algorithms.

We first discuss an efficient way to obtain a Gaussian approximation of $p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$. By Bayes' Theorem, we have

$$p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2) \propto p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\gamma}, \mathbf{h}) p(\mathbf{h} \,|\, \mu, \phi, \sigma^2).$$

In what follows, we derive explicit expressions for the two densities on the right-hand side. It turns out that the prior density $p(\mathbf{h} \,|\, \mu, \phi, \sigma^2)$ is Gaussian (see below). If we approximate the likelihood $p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\gamma}, \mathbf{h})$ by a Gaussian density in $\mathbf{h}$, we immediately obtain a Gaussian approximation of $p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$. To that end, note that given a point $\widetilde{\mathbf{h}} = (\widetilde{h}_1, \ldots, \widetilde{h}_T)' \in \mathbb{R}^T$, we can approximate the log conditional likelihood $\log p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\gamma}, \mathbf{h}) = \sum_{t=1}^T \log p(y_t \,|\, \mathbf{x}_t, \boldsymbol{\gamma}_t, h_t)$ using a second-order Taylor expansion around $\widetilde{\mathbf{h}}$ to obtain

$$\begin{aligned} \log p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\gamma}, \mathbf{h}) &\approx \log p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\gamma}, \widetilde{\mathbf{h}}) + (\mathbf{h} - \widetilde{\mathbf{h}})' \mathbf{f} - \frac{1}{2} (\mathbf{h} - \widetilde{\mathbf{h}})' \mathbf{G} (\mathbf{h} - \widetilde{\mathbf{h}}) \\ &= -\frac{1}{2} (\mathbf{h}' \mathbf{G} \mathbf{h} - 2 \mathbf{h}' (\mathbf{f} + \mathbf{G} \widetilde{\mathbf{h}})) + c_1, \end{aligned} \tag{5}$$

where c_1 is a constant independent of $\mathbf{h}$, $\mathbf{f} = (f_1, \ldots, f_T)'$ and $\mathbf{G} = \text{diag}(G_1, \ldots, G_T)$ with

$$f_t = \frac{\partial}{\partial h_t} \log p(y_t \,|\, \mathbf{x}_t, \boldsymbol{\gamma}_t, h_t)|_{h_t = \widetilde{h}_t}, \qquad G_t = -\frac{\partial^2}{\partial h_t^2} \log p(y_t \,|\, \mathbf{x}_t, \boldsymbol{\gamma}_t, h_t)|_{h_t = \widetilde{h}_t}.$$

In other words, $\mathbf{G}$ is the negative Hessian of the log conditional likelihood evaluated at $\widetilde{\mathbf{h}}$. Also note that $\mathbf{G}$ is diagonal (hence a band matrix). Since the log-conditional density

of y_t given the latent variables h_t and $\boldsymbol{\gamma}_t = (\alpha_t, \boldsymbol{\beta}_t')'$ is given by

$$\begin{aligned}\log p(y_t \,|\, \mathbf{x}_t, \boldsymbol{\gamma}_t, h_t) = &-\frac{1}{2}\log(2\pi) - \frac{1}{2}h_t - \frac{1}{2}\mathrm{e}^{-h_t}(y_t - \mathbf{x}_t'\boldsymbol{\beta}_t - \alpha_t \mathrm{e}^{h_t})^2 \\ &-\frac{1}{2}\log(2\pi) - \frac{1}{2}h_t - \frac{1}{2}\left(\alpha_t^2 \mathrm{e}^{h_t} + \mathrm{e}^{-h_t}(y_t - \mathbf{x}_t'\boldsymbol{\beta}_t)^2 - 2\alpha_t(y_t - \mathbf{x}_t'\boldsymbol{\beta}_t)\right),\end{aligned}$$

it is easy to check that

$$\begin{aligned}\frac{\partial}{\partial h_t}\log p(y_t \,|\, \mathbf{x}_t, \boldsymbol{\gamma}_t, h_t) &= -\frac{1}{2} - \frac{1}{2}\alpha_t^2 \mathrm{e}^{h_t} + \frac{1}{2}\mathrm{e}^{-h_t}(y_t - \mathbf{x}_t'\boldsymbol{\beta}_t)^2, \\ \frac{\partial^2}{\partial h_t^2}\log p(y_t \,|\, \mathbf{x}_t, \boldsymbol{\gamma}_t, h_t) &= -\frac{1}{2}\alpha_t^2 \mathrm{e}^{h_t} - \frac{1}{2}\mathrm{e}^{-h_t}(y_t - \mathbf{x}_t'\boldsymbol{\beta}_t)^2.\end{aligned}$$

Next, we derive the prior density $p(\mathbf{h} \,|\, \mu, \phi, \sigma^2)$. To that end, let

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

Note that the determinant of $\mathbf{H}_\phi$ is 1—hence, it is invertible—regardless of the value of ϕ. Also observe that $\mathbf{H}_\phi$ is a *lower bidiagonal* matrix, i.e., its nonzero elements are confined along the main diagonal and the diagonal below. Now, write the state equation of h_t in (2) as:

$$\mathbf{H}_\phi \mathbf{h} = \widetilde{\boldsymbol{\delta}}_{\mathbf{h}} + \boldsymbol{\varepsilon}^h, \qquad \boldsymbol{\varepsilon}^h \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_h)$$

where $\widetilde{\boldsymbol{\delta}}_h = (\mu, (1-\phi)\mu, \ldots, (1-\phi)\mu)'$, $\boldsymbol{\varepsilon}^h = (\varepsilon_1^h, \ldots, \varepsilon_T^h)'$ and $\mathbf{S}_h = \mathrm{diag}(\sigma^2/(1-\phi^2), \sigma^2, \ldots, \sigma^2)$. That is, $(\mathbf{h} \,|\, \mu, \phi, \sigma^2) \sim \mathcal{N}(\boldsymbol{\delta}_h, (\mathbf{H}_\phi' \mathbf{S}_h^{-1} \mathbf{H}_\phi)^{-1})$ with log-density

$$\log p(\mathbf{h} \,|\, \mu, \phi, \sigma^2) = -\frac{1}{2}(\mathbf{h}'\mathbf{H}_\phi'\mathbf{S}_h^{-1}\mathbf{H}_\phi\mathbf{h} - 2\mathbf{h}'\mathbf{H}_\phi'\mathbf{S}_h^{-1}\mathbf{H}_\phi\boldsymbol{\delta}_h) + c_2, \tag{6}$$

where $\boldsymbol{\delta}_h = \mathbf{H}_\phi^{-1}\widetilde{\boldsymbol{\delta}}_h$ and c_2 is a constant independent of $\mathbf{h}$. Finally, combining (5) and (6), we have

$$\begin{aligned}\log p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2) &= \log p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\gamma}, \mathbf{h}) + \log p(\mathbf{h} \,|\, \mu, \phi, \sigma^2) + c_3, \\ &\approx -\frac{1}{2}(\mathbf{h}'\mathbf{K}_{\mathbf{h}}\mathbf{h} - 2\mathbf{h}'\mathbf{k}_{\mathbf{h}}) + c_4,\end{aligned} \tag{7}$$

where c_3 and c_4 are constants independent of $\mathbf{h}$, $\mathbf{K}_{\mathbf{h}} = \mathbf{H}_\phi'\mathbf{S}_h^{-1}\mathbf{H}_\phi + \mathbf{G}$ and $\mathbf{k}_{\mathbf{h}} = \mathbf{f} + \mathbf{G}\widetilde{\mathbf{h}} + \mathbf{H}_\phi'\mathbf{S}_h^{-1}\mathbf{H}_\phi\boldsymbol{\delta}_{\mathbf{h}}$. It can be shown that the expression in (7) is the log-kernel of the $\mathcal{N}(\widehat{\mathbf{h}}, \mathbf{K}_{\mathbf{h}}^{-1})$ density, where $\widehat{\mathbf{h}} = \mathbf{K}_{\mathbf{h}}^{-1}\mathbf{k}_{\mathbf{h}}$ (see, e.g., Kroese and Chan, 2014, p. 238). In other words, $p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$ can be approximated by the Gaussian density with mean vector $\widehat{\mathbf{h}}$ and precision matrix $\mathbf{K}_{\mathbf{h}}$. For our purpose it is important to realize that the precision matrix $\mathbf{K}_{\mathbf{h}}$ is *tridiagonal*—i.e., its nonzero elements appear only on the

main diagonal and the diagonals above and below the main one. One consequence is that $\widehat{\mathbf{h}}$ can be computed efficiently by solving the linear system $\mathbf{K_h x} = \mathbf{k_h}$ for $\mathbf{x}$ without calculating the inverse $\mathbf{K_h}^{-1}$. In addition, draws from $\mathcal{N}(\widehat{\mathbf{h}}, \mathbf{K_h}^{-1})$ can be quickly obtained using the precision sampler in Chan and Jeliazkov (2009). This Gaussian approximation is then used as the proposal density in the acceptance-rejection Metropolis-Hastings step. It remains to select the point $\widetilde{\mathbf{h}}$ for the Taylor expansion in (5). We choose $\widetilde{\mathbf{h}}$ to be the mode of $p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$, which can be quickly obtained by the Newton-Raphson method (see Appendix A for details).

To implement Step 2, first note that we can rewrite (1) as

$$y_t = \mathbf{z}_t' \boldsymbol{\gamma}_t + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \mathrm{e}^{h_t}), \tag{8}$$

where $\mathbf{z}_t = (\exp(h_t), \mathbf{x}_t')'$ and $\boldsymbol{\gamma}_t = (\alpha_t, \boldsymbol{\beta}_t')'$. Then (3) and (8) define a linear Gaussian state space model in $\boldsymbol{\gamma}_t$. Hence, a draw from $p(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mathbf{x}, \mathbf{h}, \boldsymbol{\Omega})$ can be obtained either by conventional Kalman filter-based algorithms (e.g., Carter and Kohn, 1994; Durbin and Koopman, 2002) or the more efficient precision sampler (Chan and Jeliazkov, 2009; McCausland, Miller, and Pelletier, 2011); we adopt the latter approach and the details are given in Appendix A. Lastly, Step 3 to Step 4 are standard and we refer the readers to Appendix A for details.

To get a sense of the speed of this MCMC sampler, we fit the model using US quarterly CPI inflation with a total of $T = 262$ observations. In particular, we consider the unobserved components model—i.e., $\mathbf{x}_t$ includes only an intercept (see Section 3 for details). The algorithm is implemented using MATLAB on a desktop with an Intel Core i7-870 @2.93 GHz process. It takes about 31 seconds to obtain 10000 posterior draws.

3 Modeling and Forecasting Inflation

There is a large literature on understanding the relationship between inflation and inflation uncertainty. Friedman (1977) in his Nobel lecture postulates that higher inflation might potentially generate higher inflation uncertainty in the future. This insight is formalized in a model of asymmetric information in Ball (1992), which predicts that the public is more unsure about future inflation—due to the uncertainty around the direction of government policy—when the current inflation is high. In the reverse direction of causation, Cukierman and Meltzer (1986) consider a game-theoretic model, where the central bank has a stronger incentive to create inflation surprises in time of high inflation uncertainty. This model thus predicts that high inflation uncertainty leads to higher inflation.

Early empirical work using survey data as a measure of inflation uncertainty generally finds evidence of Granger causality in both directions (e.g., Holland, 1995). However, in contrast to what Cukierman and Meltzer (1986) predict—high inflation uncertainty leads to higher inflation—the relationship is often found to be negative. Holland (1995)

suggests one possible reason: central bankers view inflation uncertainty as costly. When the current inflation is high, they have an incentive to reduce it in the future so as to reduce the welfare cost of inflation uncertainty. Survey based measures of inflation uncertainty are later criticized as inappropriate, as they only reflect dispersion of inflation forecasts across professional forecasters, but not the individual's uncertainty about her own forecast. In view of this, Grier and Perry (1998) construct instead an inflation uncertainty proxy using the conditional variance estimates from GARCH models. Using this as a measure of inflation uncertainty, they find evidence that inflation uncertainty Granger-causes inflation in the G7 countries. However, in some countries higher inflation uncertainty lowers inflation, whereas in others the direction is reversed. Treating conditional variance *estimates* from GARCH models as observed might be problematic, especially when these estimates cannot be pinned down precisely. More recently, Berument, Yalcin, and Yildirim (2009) consider a direct approach to study the inflation-inflation uncertainty relationship. Specifically, they consider a constant coefficient SVM model in which the stochastic volatility enters the conditional mean equation. Using US data they find evidence suggesting that higher inflation uncertainty increases inflation, supporting the prediction of Cukierman and Meltzer (1986).

We use the proposed TVP-SVM model to analyze the behavior of US inflation and investigate if inflation uncertainty affects the level of inflation. In contrast to previous studies that consider only models with constant coefficients in the conditional mean, our time-varying parameter model allows us to assess if the inflation-inflation uncertainty relationship has changed over time. In fact, estimation results strongly suggest that it does. Additional results for UK and Germany show a similar pattern. Next, we examine if this information can be used to improve inflation forecasts. In a pseudo out-of-sample forecasting exercise using US vintage CPI and real-time GDP deflator, we find that the proposed model outperforms standard benchmarks for both point and density forecasts.

3.1 Full Sample Estimation Results

A popular specification for modeling inflation is the unobserved components model popularized by Stock and Watson (2007). In particular, they decompose the inflation into a trend and a transitory component, and both components have time-varying volatilities. Our point of departure is a version where the transitory component has stochastic volatility whereas the variance of the trend is constant. This variant is used in various applications such as those in Chan (2013) and Chan, Koop, and Potter (2013). Here we allow for the possibility of volatility feedback—the inflation volatility may impact the level of inflation. In addition, we also allow past inflation to affect the current inflation volatility. Specifically, we consider the following TVP-SVM model:

$$y_t = \tau_t + \alpha_t \mathrm{e}^{h_t} + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \mathrm{e}^{h_t}), \tag{9}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \beta y_{t-1} + \varepsilon_t^h, \qquad \varepsilon_t^h \sim \mathcal{N}(0, \sigma^2), \tag{10}$$

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + \boldsymbol{\varepsilon}_t^{\boldsymbol{\gamma}}, \qquad \boldsymbol{\varepsilon}_t^{\boldsymbol{\gamma}} \sim \mathcal{N}(0, \boldsymbol{\Omega}), \tag{11}$$

where $\boldsymbol{\gamma}_t = (\alpha_t, \tau_t)'$ and $\boldsymbol{\Omega}$ is a 2×2 covariance matrix. Since $\exp(h_t)$ is the variance of the transitory component, its associated coefficient in the conditional mean equation, α_t, may be interpreted as the impact of the transitory volatility on the level of inflation. We assume the independent priors as described in (4) and the values of the hyper-parameters are detailed in Appendix C. The basic setup can be extended to more complex stochastic volatility processes, such as the one with structural breaks in Eisenstat and Strachan (2015).

The model in (9)–(11) adds a feature to the general model introduced in Section 2.1 by letting past inflation y_{t-1} affect volatility h_t. The sampler in Section 2.2 can be easily modified to handle this new feature. In particular, we add an extra block to sample β from its full conditional distribution: $(\beta \,|\, \mathbf{y}, \mathbf{h}, \mu, \phi, \sigma^2) \sim \mathcal{N}(\widehat{\beta}, D_\beta)$, where $D_\beta^{-1} = V_\beta^{-1} + \mathbf{X}_\beta'\mathbf{X}_\beta/\sigma^2$ and $\widehat{\beta} = D_\beta(V_\beta^{-1}\beta_0 + \mathbf{X}_\beta'\mathbf{z}_\beta/\sigma^2)$ with $\mathbf{X}_\beta = (y_1, \ldots, y_{T-1})'$ and $\mathbf{z}_\beta = (h_2 - \phi h_1 - \mu(1-\phi), \ldots, h_T - \phi h_{T-1} - \mu(1-\phi))'$. Other blocks can be adjusted accordingly. For instance, in sampling $\mathbf{h}$, we only need to change $\widetilde{\boldsymbol{\delta}}_h$ to $\widetilde{\boldsymbol{\delta}}_h = (\mu, (1-\phi)\mu + \beta y_1, \ldots, (1-\phi)\mu + \beta y_{T-1}\beta)'$.

We first fit the model using US quarterly CPI inflation from 1948Q1 to 2013Q2. More specifically, given the quarterly CPI figures z_t, we use $y_t = 400\log(z_t/z_{t-1})$ as the CPI inflation. A plot of the data is given in Figure 1.

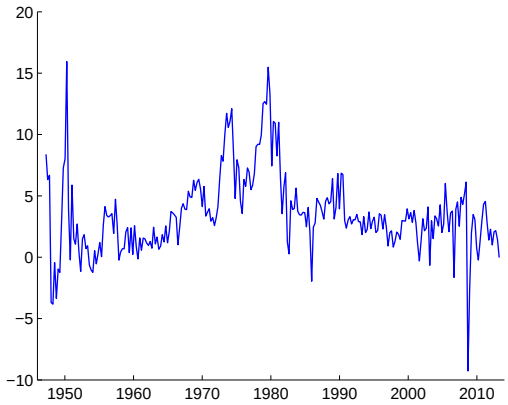

Figure 1: US quarterly CPI inflation from 1948Q1 to 2013Q2.

The results reported below are based on 50000 posterior draws—after a burn-in period of 50000—obtained using the sampler described in Section 2.2. The acceptance rate of the acceptance-rejection Metropolis-Hastings step to sample $\mathbf{h}$ is 98%, indicating that the Gaussian proposal well approximates the conditional posterior density of $\mathbf{h}$.

To further assess the efficiency of the sampler, we report the inefficiency factors of the posterior draws, defined as

$$1 + 2\sum_{l=1}^{L} \rho_l,$$

where ρ_l is the sample autocorrelation at lag length l and L is chosen to be large enough so that the autocorrelation tapers off. In the ideal case where the posterior draws are independent, the corresponding inefficiency factor is 1. More generally, the inefficiency factor measures how many extra draws are needed to give results equivalent to this ideal case. For example, an inefficiency factor of 100 indicates that roughly 10000 posterior draws are required to give the same information as 100 independent draws.

In Figure 2 we report the inefficiency factors corresponding to the posterior draws of $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_T)'$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)'$ and $\mathbf{h} = (h_1, \ldots, h_T)'$. Note that each vector is of length T, so we have a total of $3T$ inefficiency factors. To present the information visually, boxplots are reported, where the middle line of the box denotes the median, while the lower and upper lines represent respectively the 25- and the 75-percentiles. The whiskers extend to the maximum and minimum. For example, the boxplot associated with $\mathbf{h}$ indicates that about 75% of the log-volatilities have inefficiency factors less than 15, and the maximum is about 40. These values are comparable to those of the standard stochastic volatility model—i.e., when $\alpha_t = 0$—estimated by the auxiliary mixture sampler of Kim, Shepherd, and Chib (1998). The inefficiency factors of the model parameters are reported in Table 2. All in all, the results suggest that the proposed sampler is quite efficient in terms of producing posterior draws that are not highly autocorrelated.

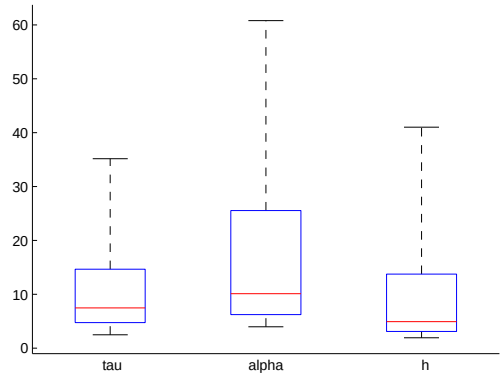

Figure 2: Boxplots of the inefficiency factors corresponding to the posterior draws of $\boldsymbol{\tau}$ (tau), $\boldsymbol{\alpha}$ (alpha) and $\mathbf{h}$ (h).

Next, we compare the unobserved components SVM model with time-varying parameters in (9)–(11)—which we denote as the **UC-SVM**—to two variants to see which model fits the inflation data best while controlling for model complexity. The first variant is the standard unobserved components model with stochastic volatility, which we call the **UC**, where we simply set $\alpha_t = 0$ for $t = 1, \ldots, T$. The second version is a constant coefficient SVM where $\alpha_1 = \cdots = \alpha_T = \alpha$, which we denote as the **UC-SVM-const**.

We use the Bayes factor as our model comparison criterion. Specifically, to compare two

models, say, M_1 and M_2, the *Bayes factor* in favor of M_1 against M_2 is defined as

$$\mathrm{BF}_{12} = \frac{p(\mathbf{y} \mid M_1)}{p(\mathbf{y} \mid M_2)},$$

where $p(\mathbf{y} \mid M_i)$ is the *marginal likelihood* for M_i, $i = 1, 2$, which is simply the marginal data density under model M_i evaluated at the observed data $\mathbf{y}$. If the observed data are unlikely under the model, the associated marginal likelihood would be "small" and vice versa. Moreover, the Bayes factor is related to the posterior odds ratio PO_{12} as follows:

$$\mathrm{PO}_{12} = \frac{\mathbb{P}(M_1 \mid \mathbf{y})}{\mathbb{P}(M_2 \mid \mathbf{y})} = \frac{\mathbb{P}(M_1)}{\mathbb{P}(M_2)} \times \mathrm{BF}_{12},$$

where $\mathbb{P}(M_1)/\mathbb{P}(M_2)$ is the prior odds ratio. Thus, under equal prior model probabilities, i.e., $\mathbb{P}(M_1) = \mathbb{P}(M_2)$, the posterior odds ratio in favor of M_1 reduces to the Bayes factor BF_{12}. In that case, if, for example, $\mathrm{BF}_{12} = 10$, then model M_1 is 10 times more likely than model M_2 given the data.

We compute the marginal likelihoods for the three models using the method of Chib and Jeliazkov (2001) and the results are reported in Table 1. The numerical standard errors are computed based on 20 independent chains, each of which of length 50000. Using a total of one million draws for each model, the marginal likelihoods are estimated accurately. Out of the three models, the SVM model with time-varying parameters is the most preferred. In fact, its Bayes factor against the next best model is about 1.2 million ($\exp(-536.4 + 550.4)$), indicating overwhelming support for the proposed model. It is also interesting to note that the data favor the **UC** and the **UC-SVM-const** about equally. Apparently, adding the SVM term but restricting the associated coefficient to be constant does not help fit the data substantially better compared to the simplest unobserved components model.

Table 1: The estimated log marginal likelihoods for the unobserved components model (**UC**), the UC-SVM model with constant coefficients (**UC-SVM-const**), and the UC-SVM model with time-varying coefficients (**UC-SVM**).

	log marginal likelihood	numerical standard error
UC-SVM	−536.4	0.08
UC-SVM-const	−550.5	0.08
UC	−550.4	0.03

We report the posterior moments and quantiles of the model parameters, which are summarized in Table 2. The estimates associated with the parameters in the AR(1) process of h_t are similar to those obtained in previous studies. For instance, the transition of h_t is highly persistent, with the posterior mean of ϕ estimated to be 0.963 with a 90% credible interval (0.928, 0.995). It is also interesting to note that β, the coefficient associated with the lagged inflation, is estimated to be quite small. In fact, its 90%

credible interval includes 0, indicating that past inflation has a limited impact on current log-volatility.

Table 2: Estimated posterior moments and quantiles of model parameters as well as the corresponding inefficiency factors.

parameter	posterior mean	posterior standard deviation	5%-tile	95%-tile	inefficiency factor
μ	0.121	0.947	-1.445	1.217	73.0
β	0.003	0.005	-0.005	0.011	4.7
ϕ	0.963	0.021	0.928	0.995	16.4
σ^2	0.072	0.020	0.045	0.110	25.9
ω_α^2	0.044	0.039	0.014	0.103	154.1
$\omega_{\alpha\tau}$	0.011	0.033	-0.033	0.055	120.3
ω_τ^2	0.117	0.057	0.053	0.224	63.7

In Figure 3 we plot the evolution of h_t and α_t and the associated 90% credible intervals. The estimates of h_t are similar to those reported in the literature—e.g., the inflation volatility increases substantially during the Great Inflation period and subsides during the Great Moderation, until it peaks again following the aftermath of the Global Financial Crisis. The right panel of the figure reports the estimates of α_t. There is apparently substantial time-variation in the estimates, highlighting the relevance of the time-varying parameter extension. For example, before 1980 the estimates are mostly between 0.5 and 2, whereas they become much smaller and even negative after the early 1980s. Using a time-invariant model that averages these two episodes is likely to give an estimate that is close to zero.

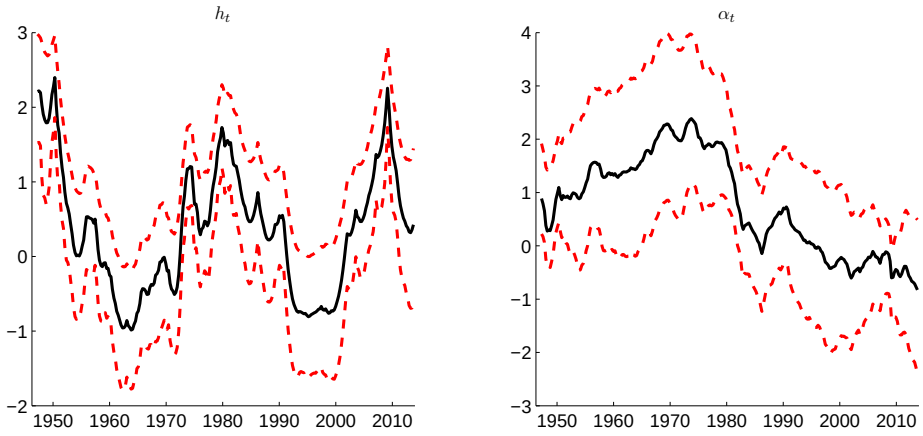

Figure 3: The evolution of h_t (left panel) and α_t (right panel). The solid lines are the estimated posterior means and the dotted lines are the 5- and 95-percentiles.

As mentioned above, using the conditional variance estimates from GARCH models as

a measure of inflation uncertainty, Grier and Perry (1998) find mixed evidence on the impact of inflation uncertainty on inflation—for some G7 countries (including US) higher inflation uncertainty in fact *lowers* inflation. In contrast, our results show that inflation uncertainty—measured by the stochastic volatility—has a *positive* impact on inflation from the beginning of the sample till early 1980s, whereas in the latter part of the sample the impact is close to zero or even negative. The differences in results may come from two sources. First, Grier and Perry (1998) use a constant coefficient model whereas we allow for time-variation in the impact of the inflation uncertainty. Second, the measure of inflation uncertainty is different—under a GARCH model the conditional variance is a deterministic function of the data and the parameters, whereas under a stochastic volatility model the log-volatility is a random variable.

Earlier we compared the proposed **UC-SVM** to a restricted version where $\alpha_t = 0$ for all t, and found that the data favor the former model. In other words, the restriction $\alpha_t = 0$ is unlikely to hold over the *whole* sample period. It can be seen in Figure 3, however, that the 90% credible intervals for α_t typically exclude zero before 1980, but they tend to include zero after the early 1990s. This can be taken as casual evidence against the restriction $\alpha_t = 0$ in the early sample and in favor of it afterward. To formally assess the evidence for this restriction over time, we adopt the approach in Koop, Leon-Gonzalez, and Strachan (2010) to compute the dynamic posterior probabilities $\mathbb{P}(\alpha_t \neq 0 \,|\, \mathbf{y})$ for $t = 1, \ldots, T$.

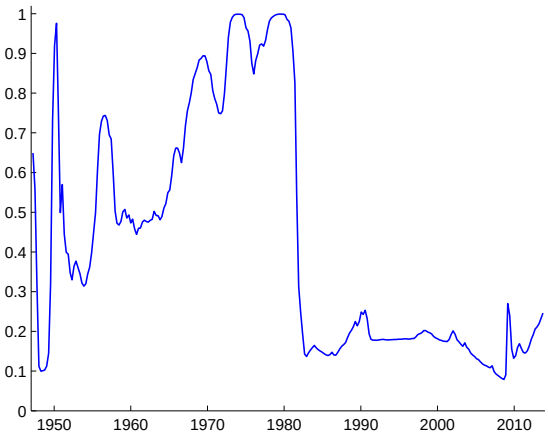

Figure 4: The dynamic posterior probabilities that $\alpha_t \neq 0$.

More specifically, we assume that *a priori* it is equally likely that the restriction $\alpha_t = 0$ holds or not—i.e., $\mathbb{P}(\alpha_t = 0) = \mathbb{P}(\alpha_t \neq 0) = 0.5$. Then, the posterior odds ratio in favor of the restriction $\alpha_t = 0$ can be computed via the Savage-Dickey density ratio

$$\mathrm{PO}_t = \frac{p(\alpha_t = 0 \,|\, \mathbf{y})}{p(\alpha_t = 0)},$$

where the numerator quantity is the value of the marginal posterior density of α_t evaluated

at zero, whereas the denominator term is the marginal prior density evaluated at zero. Koop, Leon-Gonzalez, and Strachan (2010) describe how one can compute both quantities using Monte Carlo methods based on the Kalman filter. We follow a similar approach, but using a direct method based on band matrix routines instead of the Kalman filter; the computational details are given in Appendix B. Once we have obtained the posterior odd ratios, we can calculate the posterior probabilities as $\mathbb{P}(\alpha_t \neq 0 \,|\, \mathbf{y}) = 1/(1 + \mathrm{PO}_t)$. These dynamic probabilities are reported in Figure 4.

In agreement with the estimation results of α_t, Figure 4 indicates that there is substantial time-variation in the support for the restriction $\alpha_t = 0$. For example, for much of the 1960s up till 1980, it is highly likely that $\alpha_t \neq 0$. However, there is a sharp drop in the posterior probability that $\alpha_t \neq 0$ in around 1980, coinciding with the Volcker's October 1979 revolution. Since then it becomes more likely that the restriction $\alpha_t = 0$ holds. Once again these results highlight the relevance of extending the constant coefficient SVM model to one with time-varying parameters. In the next section we further show that the proposed **UC-SVM** forecasts better than the constant coefficient variant and other standard benchmarks. Additional estimation results for UK and Germany data are provided in Section 3.3.

3.2 Forecasting Results

There is a large and growing literature on forecasting inflation using time-varying models. Recent papers include Koop and Potter (2007), Canova (2007), Stock and Watson (2007, 2010), Clark and Doh (2011), Chan, Koop, Leon-Gonzalez, and Strachan (2012), Koop and Korobilis (2012, 2013), Chan (2013), D'Agostino, Gambetti, and Giannone (2013) and Clark and Ravazzolo (2014). Using both vintage and real-time data, we investigate the out-of-sample forecast performance of the unobserved components SVM model in (9)–(11) (recall that it is denoted as the **UC-SVM**). We consider a few standard benchmarks for comparison. In particular, we include **UC-SVM-const**, **UC** and the following random walk model, denoted as the **RW**:

$$y_t = y_{t-1} + \varepsilon_t^y \qquad \varepsilon_t^y \sim \mathcal{N}(0, \sigma_y^2).$$

Moreover, we also consider the unobserved components model in Stock and Watson (2007), denoted as the **UCSV**, where both the trend and transitory components follow stochastic volatility processes.

For easy comparison, we choose broadly similar priors across models. In particular, we use exactly the same priors for the common model parameters in the four unobserved components models. The priors for the **UC-SVM** are exactly those described in Section 3.1. For the **UC-SVM-const**, we consider a Gaussian prior for the time-invariant α: $\alpha \sim \mathcal{N}(0, V_\alpha)$ with $V_\alpha = 5$. Finally, for the **RW** without stochastic volatility, the error variance of the measurement equation σ_y^2 is assumed to have an inverse-gamma prior $\sigma_y^2 \sim \mathcal{IG}(\nu_y, S_y)$ with $\nu_y = 10$ and $S_y = 9$. This implies $\mathbb{E}\,\sigma_y^2 = 1$, which is

comparable to the four specifications with stochastic volatility, where the prior mean for μ, the unconditional mean of the log-volatility, is $\mu = 0$ (and hence $e^{\mu} = 1$).

In addition to point forecasts, we also compute density forecasts—it is generally acknowledged that when assessing the quality of competing models, both the ability to accurately predict the central tendency and the uncertainty around the forecasts should be taken into account (see, e.g., Amisano and Giacomini, 2007; Clark, 2011). Specifically, we use each of the five models to produce both point and density m-step-ahead iterated forecasts with $m = 1$ and $m = 4$, i.e., one-quarter-ahead and one-year-ahead forecasts, respectively. We run the MCMC sampler described in Section 2.2 multiple times with an expanding sample. More precisely, we first obtain posterior draws given the data up to time t, which we denote as $\mathbf{y}_{1:t}$. Then, we compute the predictive mean $\mathbb{E}(y_{t+m} \,|\, \mathbf{y}_{1:t})$ as the point forecast and the predictive density $p(y_{t+m} \,|\, \mathbf{y}_{1:t})$ as the density forecast. Next, we move one period ahead and repeat the whole exercise with data $\mathbf{y}_{1:t+1}$, and so forth. These forecasts are then evaluated for $t = t_0, \ldots, T - m$, where t_0 is 1975Q1. Note that the predictive mean and the predictive density of y_{t+m} can be computed using predictive simulation. We refer the readers to the discussion in Chan (2013) for details.

The metric used to evaluate the point forecasts is the root mean squared forecast error (RMSFE) defined as

$$\text{RMSFE} = \sqrt{\frac{\sum_{t=t_0}^{T-m} (y_{t+m}^{\text{o}} - \mathbb{E}(y_{t+m} \,|\, \mathbf{y}_{1:t}))^2}{T - m - t_0 + 1}},$$

where y_{t+m}^{o} is the observed value of y_{t+m} that is known at time $t+m$. For the evaluation of the density forecast $p(y_{t+m} \,|\, \mathbf{y}_{1:t})$, we use a metric that is based on the predictive likelihood $p(y_{t+m} = y_{t+m}^{\text{o}} \,|\, \mathbf{y}_{1:t})$, i.e., the predictive density of y_{t+m} evaluated at the observed value y_{t+m}^{o}. If the actual observation y_{t+m}^{o} is likely under the density forecast, the value of the predictive likelihood will be large, and vise versa; we refer the readers to Geweke and Amisano (2011) for a more detailed discussion of the predictive likelihood. Finally, we evaluate the density forecasts using the average log predictive likelihoods:

$$\frac{1}{T - m - t_0 + 1} \sum_{t=t_0}^{T-m} \log p(y_{t+m} = y_{t+m}^{\text{o}} \,|\, \mathbf{y}_{1:t}).$$

For this metric, a larger value indicates better forecast performance.

Table 3 presents the point and density forecast results of the five models using US vintage CPI data. For easy comparison, all results are computed relative to the **RW** model. In particular, we present ratios of RMSFEs relative to those of the **RW**, where values smaller than unity indicate better forecast performance than the benchmark. For density forecasts, we report differences of average log predictive likelihoods of a given model from those of the **RW**—hence positive values indicate better forecast performance compared to the **RW**.

A few broad observations can be drawn from these forecasting results. First, the four unobserved components model all forecast better than the benchmark **RW** in both point

and density forecasts. For instance, in comparing one-quarter-ahead point forecasts, the RMSFE for the **UCSV** is 9% lower than the value for the **RW**. Second, the proposed **UC-SVM** performs the best in both point and density forecasts and for both forecast horizons. In particular, it outperforms the popular **UCSV**, showing the value of incorporating the inflation-inflation uncertainty relationship into the forecasting model. Furthermore, the proposed model also forecasts better than the **UC-SVM-const**—the SVM model with constant coefficients—highlighting the empirical importance of allowing for time-varying parameters.

Table 3: Point and density forecast performance of competing models relative to the random walk model (**RW**); vintage US CPI inflation.

	relative RMSFE		relative average predictive likelihood	
	1-quarter-ahead	1-year-ahead	1-quarter-ahead	1-year-ahead
RW	1.00	1.00	0.00	0.00
UC	0.94	0.88	0.27*	0.46*
UCSV	0.91*	0.88*	0.26*	0.45*
UC-SVM-const	0.92*	0.90	0.23*	0.45*
UC-SVM	0.89*	0.88*	0.31*	0.56*#

Note: $*$ indicates superior forecast performance relative to the **RW** at significance level 0.05 using an asymptotic test in Diebold and Mariano (1995); # indicates superior performance relative to the **UCSV**.

Table 3 reports the average forecast performance of the five models over the whole evaluation period. Since a growing body of work has shown that forecast performance is often not stable over time, we investigate the forecast performance of the proposed model over time by plotting the cumulative sums of log predictive likelihoods in Figure 5.

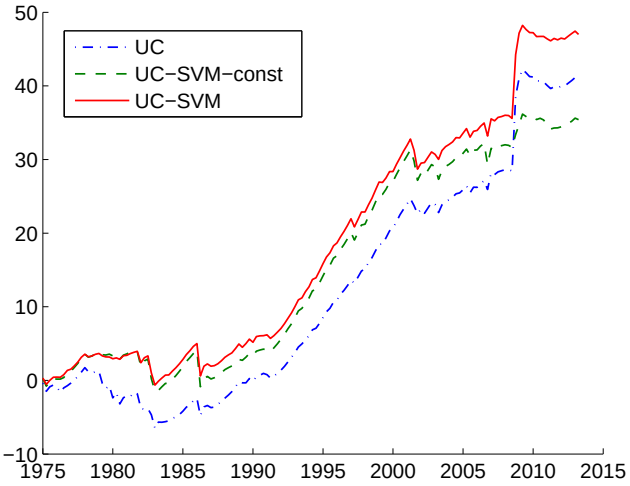

Figure 5: Cumulative sums of log predictive likelihoods for one-quarter-ahead forecasts relative to the random walk model (**RW**); vintage US CPI inflation.

Overall, it can be seen that the **UC-SVM** consistently forecasts better than the alternatives. Moreover, a few particular patterns are worth mentioning. First, the two SVM models consistently forecast better than the random walk benchmark, whereas the **UC** outperforms the benchmark only after the early 1990s. Second, the **UC-SVM-const** forecasts better than the **UC** for most of the sample, but it is overtaken by the latter at around the Global Financial Crisis, whereas the time-varying **UC-SVM** remains the best. This shows that models with time-varying parameters are better at capturing the abrupt shock of the Global Financial Crisis than models with constant coefficients.

Next, we evaluate the performance of the proposed model in a real-time forecasting exercise. Specifically, real-time data on GDP price index (vintages from 1965Q4 to 2013Q4) are obtained from the Real-Time Data Set for Macroeconomists compiled by the Philadelphia Federal Reserve Bank. As before, the evaluation period starts from 1975Q1 till the end of the sample. To produce a one-quarter-ahead forecast for the inflation at time t, we use the vintage dated time t, which contains data up to only time $t-1$. To evaluate this forecast, we use the relevant observation in the vintage dated time $t+1$. Similarly, to produce a four-quarter-ahead forecast for the inflation at time t, we use the vintage dated time $t-3$, which contains data up to time $t-4$. Again, the vintage dated time $t+1$ is used for forecast evaluation.

Table 4: Point and density forecast performance of competing models relative to the random walk model (**RW**); real-time US GDP deflator.

	relative RMSFE		relative average predictive likelihood	
	1-quarter-ahead	1-year-ahead	1-quarter-ahead	1-year-ahead
RW	1.00	1.00	0.00	0.00
UC	0.90	0.93	0.18*	0.37*
UCSV	0.89*	0.95*	0.15*	0.28*
UC-SVM-const	0.87*	0.95	0.22*#	0.40*#
UC-SVM	0.87*	0.95	0.23*#	0.40*#

Note: $*$ indicates superior forecast performance relative to the **RW** at significance level 0.05 using an asymptotic test in Diebold and Mariano (1995); # indicates superior performance relative to the **UCSV**.

As before, both point and density forecasts are produced for the five models and the results are reported in Table 4. The main conclusions remain the same: while all unobserved components models do better than the **RW**, the proposed **UC-SVM** generally outperforms other UC variants.

3.3 Additional Results for UK and Germany Inflation

In this section we report additional estimation results using data from the UK and Germany. We select these two countries primarily because of their very different inflation experience after World War II. Figure 6 plots the quarterly CPI inflation of both countries from 1955Q1 to 2013Q4 obtained from the OECD economic database.

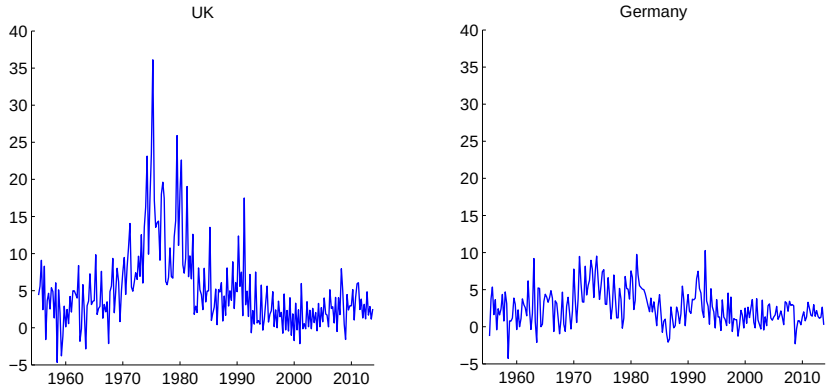

Figure 6: UK and Germany quarterly CPI inflation from 1955Q1 to 2013Q4.

It can be seen from the left panel that UK inflation had been remarkably volatile before the Bank of England adopted inflation targeting and gained independence in the 1990s. Since then UK inflation has become less volatile and remained low. In contrast, Germany essentially escaped the Great Inflation and always had very low and stable inflation. Hence, comparing these two countries might show interesting results.

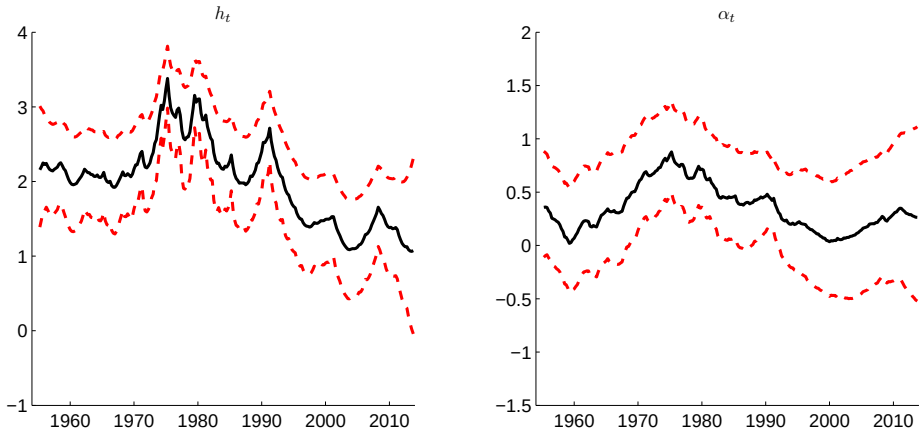

Figure 7: The evolution of h_t (left panel) and α_t (right panel) for UK data.

Figures 7 and 8 plot the evolution of the log-volatility h_t and its time-varying impact α_t for UK and Germany data, respectively, as well as the the associated 90% credible intervals (dashed lines). It is also not surprising that the log-volatility estimates of the UK are generally larger than those of Germany, indicating that UK inflation is typically more volatile. The estimates of h_t for the UK are qualitatively similar to US estimates—e.g., inflation volatility increases substantially during the Great Inflation, followed by an overall decline starting from the early 1980s (except for a spike during the Exchange Rate Mechanism Crisis in 1992). In contrast, the estimates of h_t for Germany are remarkably flat during the Great Inflation. They show a similar general decline throughout the 1980s,

which is only interrupted briefly following the unification of West and East Germany in 1990.

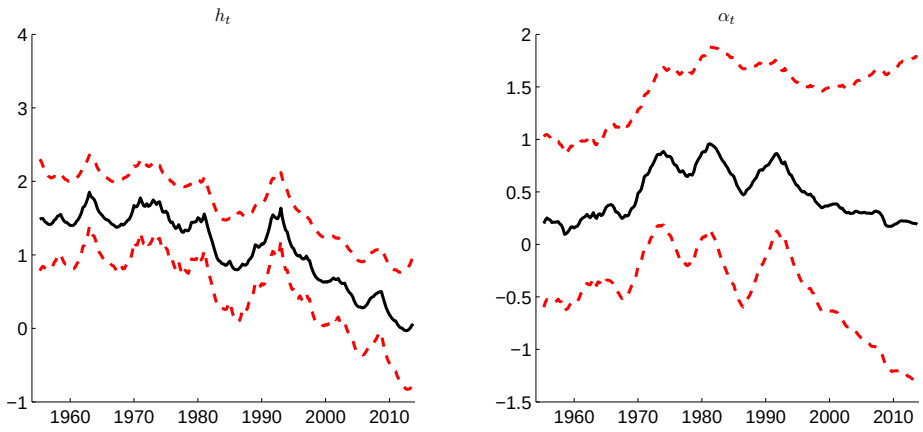

Figure 8: The evolution of h_t (left panel) and α_t (right panel) for Germany data.

In contrast to the results in Grier and Perry (1998) who find mixed evidence on the impact of inflation uncertainty on inflation—both positive and negative estimates are obtained for G7 countries—our estimates of α_t for UK and Germany are always positive. Similar to the results of US data, UK estimates of α_t show substantial time-variation—the estimates increase from about 0 in 1960 to around 1 in mid-1970s, which is followed by a steady decline in the 1980s and 1990s. The corresponding 90% credible intervals follow a similar pattern: they tend to exclude 0 from the early 1970s onwards, but they become wider and include 0 from the early 1990s, where the timing seems to coincide with the adoption of inflation targeting by the Bank of England in 1992. Although the estimates for Germany are always positive and follow a qualitatively similar pattern, the estimates are substantially more imprecise—the 90% credible intervals of α_t are much wider and typically include 0 except for a few episodes. For Germany inflation volatility feedback seems to be a less important channel compared to the UK and the US.

4 Concluding Remarks and Future Research

We have extended the popular stochastic volatility in mean model of Koopman and Hol Uspensky (2002) to allow for time-varying parameters in the conditional mean. An efficient sampler building upon recent advances in state space simulation techniques is developed to estimate this more general model. We demonstrated the methodology with an empirical application involving inflation modeling and forecasting. The estimation results showed substantial time-variation in the coefficient associated with the stochastic volatility, highlighting the empirical relevance of the proposed extension. Furthermore, in a recursive forecasting exercise, the time-varying parameter variant also performed better than various standard benchmarks in both point and density forecasts.

For future research, it would be interesting to investigate if the proposed time-varying parameter variant also fits other macroeconomic or financial time series better than the standard SVM model. Furthermore, it would also be of interest to generalize the proposed model to a multivariate setting, e.g., in the spirit of Mumtaz and Zanetti (2013).

Appendix A: Estimation of the TVP-SVM Model

In this appendix we provide details of the Metropolis-within-Gibbs sampler for estimating the TVP-SVM model in (1)–(3).

We first discuss how one can quickly locate the maximum of $p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$, or equivalently that of $\log p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$, using the Newton-Raphson method (see, e.g., Kroese, Taimre, and Botev, 2011, pp. 688-689). It follows from (7) that the negative Hessian of $\log p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{x}, \boldsymbol{\gamma}, \mu, \phi, \sigma^2)$ evaluated at $\mathbf{h} = \widetilde{\mathbf{h}}$ is $\mathbf{K}_\mathbf{h}$ and the gradient at $\mathbf{h} = \widetilde{\mathbf{h}}$ is $-\mathbf{K}_\mathbf{h}\widetilde{\mathbf{h}} + \mathbf{k}_\mathbf{h}$. Hence, we can implement the Newton-Raphson method as follows: initialize with $\mathbf{h} = \mathbf{h}^{(1)}$ for some constant vector $\mathbf{h}^{(1)}$. For $t = 1, 2, \ldots$, use $\widetilde{\mathbf{h}} = \mathbf{h}^{(t)}$ in the evaluation of $\mathbf{K}_\mathbf{h}$ and $\mathbf{k}_\mathbf{h}$, and compute

$$\mathbf{h}^{(t+1)} = \mathbf{h}^{(t)} + \mathbf{K}_\mathbf{h}^{-1}(-\mathbf{K}_\mathbf{h}\mathbf{h}^{(t)} + \mathbf{k}_\mathbf{h}) = \mathbf{K}_\mathbf{h}^{-1}\mathbf{k}_\mathbf{h}.$$

Repeat this procedure until some convergence criterion is reached, e.g., when $\|\mathbf{h}^{(t+1)} - \mathbf{h}^{(t)}\| < \varepsilon$ for some prefixed tolerance level ε.

Next, we discuss Step 2 of the sampler: to efficiently sample from $p(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mathbf{x}, \mathbf{h}, \boldsymbol{\Omega})$. Stacking (8) over $t = 1, \ldots, T$, we have

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^y, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_\mathbf{y}), \tag{12}$$

where $\boldsymbol{\varepsilon}^y = (\varepsilon_1^y, \ldots, \varepsilon_T^y)'$,

$$\mathbf{S}_\mathbf{y} = \begin{pmatrix} \mathrm{e}^{h_1} & 0 & \cdots & 0 \\ 0 & \mathrm{e}^{h_2} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathrm{e}^{h_T} \end{pmatrix}, \qquad \mathbf{Z} = \begin{pmatrix} \mathbf{z}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_2' & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{z}_T' \end{pmatrix}.$$

To derive the prior density of $\boldsymbol{\gamma}$ implied by (3), rewrite the state equation of $\boldsymbol{\gamma}_t$ in matrix notations:

$$\mathbf{H}\boldsymbol{\gamma} = \widetilde{\boldsymbol{\delta}}_{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon}^\gamma, \qquad \boldsymbol{\varepsilon}^\gamma \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_\gamma)$$

where $\widetilde{\boldsymbol{\delta}}_{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}_0', \mathbf{0}, \ldots, \mathbf{0})'$, $\boldsymbol{\varepsilon}^\gamma = (\boldsymbol{\varepsilon}_1^{\gamma\,\prime}, \ldots, \boldsymbol{\varepsilon}_T^{\gamma\,\prime})'$, $\mathbf{S}_\gamma = \mathrm{diag}(\boldsymbol{\Omega}_0, \boldsymbol{\Omega}, \ldots, \boldsymbol{\Omega})$ and $\mathbf{H}$ is the first difference matrix

$$\mathbf{H} = \begin{pmatrix} \mathbf{I}_{k+1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_{k+1} & \mathbf{I}_{k+1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_{k+1} & \mathbf{I}_{k+1} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{I}_{k+1} & \mathbf{I}_{k+1} \end{pmatrix}.$$

That is, $(\boldsymbol{\gamma} \,|\, \boldsymbol{\Omega}) \sim \mathcal{N}(\boldsymbol{\delta}_{\boldsymbol{\gamma}}, (\mathbf{H}'\mathbf{S}_{\boldsymbol{\gamma}}^{-1}\mathbf{H})^{-1})$, where $\boldsymbol{\delta}_{\boldsymbol{\gamma}} = \mathbf{H}^{-1}\widetilde{\boldsymbol{\delta}}_{\boldsymbol{\gamma}}$. Now, using standard results from linear regression (see, e.g., Kroese and Chan, 2014, Corollary 8.1), we have,

$$(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mathbf{h}, \boldsymbol{\Omega}) \sim \mathcal{N}(\widehat{\boldsymbol{\gamma}}, \mathbf{D}_{\boldsymbol{\gamma}}), \tag{13}$$

where $\mathbf{D}_{\boldsymbol{\gamma}}^{-1} = \mathbf{H}'\mathbf{S}_{\boldsymbol{\gamma}}^{-1}\mathbf{H} + \mathbf{Z}'\mathbf{S}_{\mathbf{y}}^{-1}\mathbf{Z}$ and $\widehat{\boldsymbol{\gamma}} = \mathbf{D}_{\boldsymbol{\gamma}}(\mathbf{H}'\mathbf{S}_{\boldsymbol{\gamma}}^{-1}\mathbf{H}\boldsymbol{\delta}_{\boldsymbol{\gamma}} + \mathbf{Z}'\mathbf{S}_{\mathbf{y}}^{-1}\mathbf{y})$. Since $\mathbf{D}_{\boldsymbol{\gamma}}$ is typically a high-dimensional full covariance matrix, the conventional sampling approach that requires the Cholesky factor of $\mathbf{D}_{\boldsymbol{\gamma}}$ is time-consuming. However, since the precision $\mathbf{D}_{\boldsymbol{\gamma}}^{-1}$ is a band matrix, one can efficiently sample from $\mathcal{N}(\widehat{\boldsymbol{\gamma}}, \mathbf{D}_{\boldsymbol{\gamma}})$ using band and sparse matrix routines; we refer the readers to Chan and Jeliazkov (2009) for details.

To implement Step 3, first note that $\boldsymbol{\Omega}$ and σ^2 are conditionally independent given the latent states $\boldsymbol{\gamma}$ and $\mathbf{h}$. In fact, both conditional distributions are standard distributions:

$$\begin{aligned}
(\boldsymbol{\Omega} \,|\, \boldsymbol{\gamma}) &\sim \mathcal{IW}\left(\nu_{\boldsymbol{\Omega}} + T - 1, \mathbf{S}_{\boldsymbol{\Omega}} + \sum_{t=2}^{T} (\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_{t-1})(\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_{t-1})'\right), \\
(\sigma^2 \,|\, \mathbf{h}, \mu, \phi) &\sim \mathcal{IG}\left(\nu_{\sigma^2} + \frac{T}{2}, \widetilde{S}_{\sigma^2}\right),
\end{aligned}$$

where $\widetilde{S}_{\sigma^2} = S_{\sigma^2} + ((1-\phi^2)(h_1 - \mu)^2 + \sum_{t=2}^{T}(h_t - \mu - \phi(h_{t-1} - \mu))^2)/2$.

Lastly, we sample μ and ϕ jointly to improve efficiency. Specifically, we implement an independence-chain Metropolis-Hastings step with a bivariate Student-t proposal with degree of freedom parameter set to 5. Recall that ϕ is bounded between -1 and 1. For computational convenience, we consider the transformation $\eta = \tanh(\phi)$ so that η takes any real value. Then the mode of the joint density of (μ, η) can be obtained using the Newton-Raphson method.

Appendix B: Computation of Dynamic Probabilities

In this appendix we provide details of computing the time-varying probabilities that $\alpha_t \neq 0$ as reported in Figure 4. Recall that the posterior odds ratio in favor of the restriction $\alpha_t = 0$ can be obtained via the Savage-Dickey density ratio

$$\mathrm{PO}_t = \frac{p(\alpha_t = 0 \,|\, \mathbf{y})}{p(\alpha_t = 0)}.$$

Hence, it suffices to calculate the two quantities $p(\alpha_t = 0 \,|\, \mathbf{y})$ and $p(\alpha_t = 0)$. First, note that even though the *marginal* posterior density $p(\alpha_t \,|\, \mathbf{y})$ is unknown, the *conditional* posterior density $p(\alpha_t \,|\, \mathbf{y}, \mathbf{h}, \boldsymbol{\Omega})$ is Gaussian. To see this, recall from (13) that the joint density $p(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mathbf{h}, \boldsymbol{\Omega})$ is Gaussian, where $\boldsymbol{\gamma} = (\alpha_1, \tau_1, \ldots, \alpha_T, \tau_T)'$. Since all marginals of a jointly Gaussian density is Gaussian, so is $p(\alpha_t \,|\, \mathbf{y}, \mathbf{h}, \boldsymbol{\Omega})$. Therefore, we can evaluate $p(\alpha_t = 0 \,|\, \mathbf{y}, \mathbf{h}, \boldsymbol{\Omega})$ exactly. It follows that $p(\alpha_t = 0 \,|\, \mathbf{y})$ can be estimated using the Monte Carlo average:

$$\widehat{p(\alpha_t = 0 \,|\, \mathbf{y})} = \frac{1}{R}\sum_{i=1}^{R} p(\alpha_t = 0 \,|\, \mathbf{y}, \mathbf{h}^{(i)}, \boldsymbol{\Omega}^{(i)}),$$

where $(\mathbf{h}^{(1)}, \mathbf{\Omega}^{(1)}), \ldots, (\mathbf{h}^{(R)}, \mathbf{\Omega}^{(R)})$ are posterior draws. Similarly, we can estimate $p(\alpha_t = 0)$ using Monte Carlo methods since the prior density $p(\boldsymbol{\gamma} \,|\, \mathbf{\Omega})$ is also Gaussian; see Appendix A for the exact expression. Finally, we can calculate the posterior probabilities using $\mathbb{P}(\alpha_t \neq 0 \,|\, \mathbf{y}) = 1/(1 + \mathrm{PO}_t)$.

Appendix C: Hyper-parameters for the Application

In this appendix we provide the details of the hyper-parameters used in the inflation application. The prior distributions are described in (4). In particular, μ follows a Gaussian distribution with mean $\mu_0 = 0$ and variance $V_\mu = 10$; ϕ has a truncated Gaussian distribution with $\phi_0 = 0.97$ and $V_\phi = 0.1^2$. In addition, β has a normal prior $\mathcal{N}(\beta_0, V_\beta)$ with $\beta_0 = 0$ and $V_\beta = 10$.

The variance parameters are independently distributed as $\sigma^2 \sim \mathcal{IG}(\nu_{\sigma^2}, S_{\sigma^2})$ and $\mathbf{\Omega} \sim \mathcal{IW}(\nu_{\mathbf{\Omega}}, \mathbf{S}_{\mathbf{\Omega}})$. We choose relatively small—hence relatively noninformative—values for the degrees of freedom parameters: $\nu_{\sigma^2} = \nu_{\mathbf{\Omega}} = 10$. For the scale parameters, we set $S = 0.36$ and $\mathbf{S}_{\mathbf{\Omega}} = \mathrm{diag}(0.13, 0.8125)$. These values imply $\mathbb{E}\,\sigma^2 = 0.2^2$, $\mathbb{E}\,\mathbf{\Omega} = \mathrm{diag}(0.1^2, 0.25^2)$. The chosen prior means reflect the desired smoothness of the corresponding state transition, and are comparable to those used in previous studies, such as Chan (2013) and Stock and Watson (2007).

References

G. Amisano and R. Giacomini. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, 25(2):177–190, 2007.

L. Ball. Why does high inflation raise inflation uncertainty? *Journal of Monetary Economics*, 29(3):371–388, 1992.

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

F. Canova. Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model. *Journal of Economic Dynamics and Control*, 17:233–261, 1993.

F. Canova. G-7 inflation forecasts: random walk, Phillips curve or what else? *Macroeconomic Dynamics*, 11(1):1–30, 2007.

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

J. C. C. Chan. Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176(2):162–172, 2013.

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

J. C. C. Chan and R. Strachan. The zero lower bound: Implications for modelling the interest rate. Working paper, Research School of Economics, Australian National University, 2014.

J. C. C. Chan, G. Koop, R. Leon-Gonzalez, and R. Strachan. Time varying dimension models. *Journal of Business and Economic Statistics*, 30(3):358–367, 2012.

J. C. C. Chan, G. Koop, and S. M. Potter. A new model of trend inflation. *Journal of Business and Economic Statistics*, 31(1):94–106, 2013.

S. Chib and I. Jeliazkov. Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, 96:270–281, 2001.

T. E. Clark. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29(3), 2011.

T. E. Clark and T. Doh. A Bayesian evaluation of alternative models of trend inflation. Working paper, Federal Reserve Bank of Cleveland, 2011.

T. E. Clark and F. Ravazzolo. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 2014. Forthcoming.

T. Cogley and T. J. Sargent. Evolving post-world war II US inflation dynamics. *NBER Macroeconomics Annual*, 16:331–388, 2001.

T. Cogley and T. J. Sargent. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262 – 302, 2005.

T. Cogley, G. Primiceri, and T. Sargent. Inflation-gap persistence in the U.S. *American Economic Journal: Macroeconomics*, 2:43–69, 2010.

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

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

P. de Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, 82:339–350, 1995.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):134–144, 1995.

B. Djegnéné and W. J. McCausland. The HESSIAN method for models with leverage-like effects. *Journal of Financial Econometrics*, 2014. Forthcoming.

J. Durbin and S. J. Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89:603–615, 2002.

E. Eisenstat and R. W. Strachan. Modelling inflation volatility. *Journal of Applied Econometrics*, 2015. Forthcoming.

R. F. Engle, D. M. Lilien, and R. P. Robins. Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica*, 55(2):391–407, 1987.

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

S. Früwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994.

J. Geweke and G. Amisano. Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26:1–29, 2011.

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

A. S. Holland. Inflation and uncertainty: tests for temporal ordering. *Journal of Money, Credit and Banking*, 27(3):827–837, 1995.

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

G. Koop and D. Korobilis. Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886, 2012.

G. Koop and D. Korobilis. Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198, 2013.

G. Koop and S. M. Potter. Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, 74:763–789, 2007.

G. Koop, R. Leon-Gonzalez, and R. W. Strachan. Dynamic probabilities of restrictions in state space models: an application to the Phillips curve. *Journal of Business & Economic Statistics*, 28(3):370–379, 2010.

S. J. Koopman and E. Hol Uspensky. The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of Applied Econometrics*, 17(6): 667–689, 2002.

D. P. Kroese and J. C. C. Chan. *Statistical Modeling and Computation*. Springer, New York, 2014.

D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. John Wiley & Sons, New York, 2011.

W. J. McCausland. The HESSIAN method: Highly efficient simulation smoothing, in a nutshell. *Journal of Econometrics*, 168(2):189–206, 2012.

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

H. Mumtaz and F. Zanetti. The impact of the volatility of monetary policy shocks. *Journal of Money, Credit and Banking*, 45(4):535–558, 2013.

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

H. Rue. Fast sampling of Gaussian Markov random fields with applications. *Journal of the Royal Statistical Society Series B*, 63:325–338, 2001.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace. *Journal of the Royal Statistical Society Series B*, 71:319–392, 2009.

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

J. H. Stock and M. W. Watson. Modeling inflation after the crisis. Working Paper 16488, National Bureau of Economic Research, October 2010.