

Moving Average Stochastic Volatility Models with Application to Inflation Forecast

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

May 2012

Abstract

Moving average and stochastic volatility are two important components for modeling and forecasting macroeconomic and financial time series. The former aims to capture short-run dynamics, whereas the latter allows for volatility clustering and time-varying volatility. We introduce a new class of models that includes both of these useful features. The new models allow the conditional mean process to have a state space form. As such, this general framework includes a wide variety of popular specifications, including the unobserved components and time-varying parameter models. Having a moving average process, however, means that the errors in the measurement equation are no longer serially independent, and estimation becomes more difficult. We develop a posterior simulator that builds upon recent advances in precision-based algorithms for estimating this new class of models. In an empirical application involving U.S. inflation we find that these moving average stochastic volatility models provide better in-sample fitness and out-of-sample forecast performance than the standard variants with only stochastic volatility.

Keywords: state space, unobserved components model, precision, sparse, density forecast

JEL classification codes: C11, C51, C53

1 Introduction

Since the pioneering works of Box and Jenkins, autoregressive moving average (ARMA) models have become standard tools for modeling and forecasting time series. The theoretical justification of these ARMA models, as is well-known, is the Wold decomposition theorem, which states that any zero mean covariance-stationary time series has an infinite moving average representation. One implication is that any such process can be approximated arbitrarily well by a sufficiently high order ARMA model. In practice, it is found that simple univariate ARMA models often outperform complex multivariate models in forecasting.

However, despite the theoretical justification and empirical success of this class of models, a voluminous literature has highlighted the importance of allowing for time-varying volatility in macroeconomic and financial data for estimation and forecasting. Standard ARMA models that assume constant variance are seemingly not flexible enough. One way to accommodate this time-variation in variance is via the GARCH model (Bollerslev, 1986). For example, Nakatsuma (2000) considers a linear regression model with ARMA-GARCH errors. Another popular way to allow for time-varying volatility is via the stochastic volatility (SV) model (e.g., Taylor, 1994; Kim, Shephard, and Chib, 1998). The popularity of this approach can be seen through the numerous extensions of the basic SV setup in recent years, such as the SV models with jump and Student's t error (Chib, Nardari, and Shephard, 2002), SV with leverage (Jacquier, Polson, and Rossi, 2004; Omori, Chib, Shephard, and Nakajima, 2007), SV with asymmetric, heavy-tailed error (Nakajima and Omori, 2012), semiparametric SV models via the Dirichlet process mixture (Jensen and Maheu, 2010), etc., to name but a few examples.

Several recent studies have attempted to bridge these two literatures on ARMA and SV models, and have considered various flexible autoregressive models with stochastic volatility (e.g., Cogley and Sargent, 2005; Primiceri, 2005; Cogley, Primiceri, and Sargent, 2010). But there are few papers that investigate moving average models with SV. The purpose of this article is to fill this gap: we introduce a class of models that include both the moving average and stochastic volatility components, where the conditional mean process has a flexible state space representation. As such, the setup includes a wide variety of popular specifications as special cases, including the unobserved components and time-varying parameter models. Of course, any invertible MA process can be approximated by a sufficiently high order AR model. In practice, however, forecasts based on these AR models — since they have many parameters to estimate — often compare poorly to parsimonious ARMA models (e.g., Stock and Watson, 2007; Athanasopoulos and Vahid, 2008). In our empirical work that involves quarterly inflation, we find that there is substantial support for the proposed models against their counterparts with only SV. The forecasting results suggest that addition of the MA component further improves the forecast performance of standard SV models, particularly at short forecast horizons.

A second contribution of this paper is to develop an efficient Markov chain Monte Carlo (MCMC) sampler for estimating this class of models. Since the conditional mean process

has a state space form, estimation might appear to be straightforward. However, under our models the errors in the measurement equation are no longer serially independent due to the presence of the MA component. As such, application of Kalman filter-based methods would first require a suitable transformation of the data to make the errors serially independent. Instead of using the Kalman filter, we take a different approach: we extend previous work on precision-based algorithms for state space models in Chan and Jeliazkov (2009), which is shown to be more efficient than Kalman filter-based methods; see also Rue, Martino, and Chopin (2009). By exploiting the sparse structure of the covariance matrix of the observations, we develop an easy and fast method for estimating these new models.

A third contribution involves a substantive empirical application on modeling and forecasting U.S. quarterly consumer price index (CPI) inflation. A vast literature on this topic has emerged over the last two decades; recent studies include Koop and Potter (2007), Stock and Watson (2007, 2010), Cogley and Sbordone (2008), Cogley et al. (2010), Clark and Doh (2011), Korobilis (2012), Koop and Korobilis (2012), among many others. One key finding in this literature is that both persistence and volatility in the inflation process have changed considerably over time. In particular, inflation volatility decreased gradually from the high level in the late 1970s to a trough at around 1990, until it increased again and peaked at the aftermath of the global financial crisis. Empirically, it is often found that models with stochastic volatility provide substantially better point and density forecasts than those obtained from constant error variance models (e.g., Clark and Doh, 2011; Chan, Koop, Leon-Gonzalez, and Strachan, 2012a).

Another key finding in this literature is that for forecasting inflation, both at short and long horizons, it is often difficult to improve upon univariate models using only information in observed inflation (e.g., Stock and Watson, 2007, 2010; Chan et al., 2012a). One reason for this lack of predictive power of a wide range of seemingly relevant variables — such as unemployment rate and GDP growth — might be because variables useful for forecasting change over time (e.g., oil price might be an important predictor for inflation in the 1970s but is less important in the 2000s) and/or over business cycle (e.g., some variables may predict well in expansions but not in recessions). In fact, Koop and Korobilis (2012) find evidence that the set of relevant predictors for inflation does change over time. Given these findings, we consider univariate time series models using only information in observed inflation. Additional explanatory variables, of course, can be incorporated if desired.

We focus on univariate MA-SV models in this paper, partly because for our empirical work these models are sufficient. We note that the univariate framework developed here can be used to construct multivariate models in a straightforward manner. For example, in the multivariate SV models of Chib, Nardari, and Shephard (2006), SV is induced by a number of latent factors, each of which follows an independent univariate SV process. In this setup, we can, for example, replace the SV process with the univariate MA-SV process introduced in this paper to construct a multivariate SV model with autocorrelated errors. We leave the multivariate case for future research.

The rest of this article is organized as follows. In Section 2 we introduce the general framework, and discuss how this state space form includes a variety of popular specifications as special cases. Section 3 develops an efficient posterior simulator to estimate this new class of models. Section 4 presents empirical results for modeling and forecasting U.S. CPI inflation. In the last section we conclude our findings and discuss future research direction.

2 Moving Average Stochastic Volatility Models

The general framework we consider is the following q -th-order moving average model with stochastic volatility:

$$y_t = \mu_t + \varepsilon_t^y, \quad (1)$$

$$\varepsilon_t^y = u_t + \psi_1 u_{t-1} + \cdots + \psi_q u_{t-q}, \quad u_t \sim \mathcal{N}(0, e^{h_t}), \quad (2)$$

$$h_t = h_{t-1} + \varepsilon_t^h, \quad \varepsilon_t^h \sim \mathcal{N}(0, \sigma_h^2). \quad (3)$$

The errors u_t and ε_t^h are independent of each other for all leads and lags. We further assume that $u_0 = u_{-1} = \cdots = u_{-q+1} = 0$. One can, of course, treat these initial error terms as parameters if desired, and the estimation procedures discussed in the next section can be easily extended to allow for this possibility. For typical situations where $T \gg q$, whether these errors are modeled explicitly or not makes little difference in practice.

By choosing a suitable conditional mean process μ_t in (1), this setup includes a variety of popular specifications, such as:

1. the autoregressive model:

$$\mu_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p};$$

2. the linear regression model:

$$\mu_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt},$$

where $\mathbf{x}_t = (x_{1t}, \dots, x_{kt})$ is a vector of explanatory variables;

3. the unobserved components model:

$$\begin{aligned} \mu_t &= \tau_t, \\ \tau_t &= \tau_{t-1} + \varepsilon_t^\tau, \quad \varepsilon_t^\tau \sim \mathcal{N}(0, \sigma_\tau^2); \end{aligned}$$

4. the time-varying parameter model:

$$\begin{aligned} \mu_t &= \beta_{0t} + \beta_{1t} x_{1t} + \cdots + \beta_{kt} x_{kt}, \\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\varepsilon}_t^\beta, \quad \boldsymbol{\varepsilon}_t^\beta \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\beta), \end{aligned}$$

where $\boldsymbol{\beta}_t = (\beta_{0t}, \beta_{1t}, \dots, \beta_{kt})'$.

Some other flexible time-varying models recently introduced also fall within this general framework. Examples include an autoregressive unobserved components model discussed in Clark and Doh (2011), as well as various bounded trend inflation models proposed in Chan, Koop, and Potter (2012b). The framework in (1)–(3) is a natural extension of the standard stochastic volatility setting. In particular, the latter is a special case of this general framework with $\psi_1 = \dots = \psi_q = 0$. For identification purposes, we assume the usual invertibility conditions, i.e., the roots of the characteristic polynomial associated with the MA coefficients are all outside the unit circle.

It is well known that moving average models have a state space representation, and the likelihood function can be evaluated using the Kalman filter (Harvey, 1985). This approach can be slow, however, especially when we need to make tens of thousands functional evaluations of the likelihood in the MCMC algorithm. We therefore introduce in the next section a direct way to evaluate the likelihood function. By utilizing fast sparse matrix routines, the new approach is simple and easy to program. Another complication in our setting is that the MA component induces serial dependence in observations. Consequently, in order to apply conventional Kalman filter-based methods to simulate the states, one would first need to transform the data so that the errors in the new measurement equation are serially independent (e.g., one such transformation is suggested in Chib and Greenberg, 1994). Instead of using Kalman filter, we introduce a direct approach that builds upon previous work on precision-based algorithms (Chan and Jeliazkov, 2009) for fitting this new class of models.

3 Estimation

We introduce a direct approach for estimating the class of models in (1)–(3) that exploits the special structure of the problem, particularly that the covariance matrix of the joint distribution for $\mathbf{y} = (y_1, \dots, y_T)'$ is sparse, i.e., it contains only a few non-zero elements. We first introduce a fast and simple way to evaluate the likelihood function that is useful for both maximum likelihood and Bayesian estimation. It is followed by a detailed discussion on a new posterior simulator for estimating the log-volatilities and the parameters in the conditional mean process.

3.1 Likelihood Evaluation

To obtain the likelihood function, we first derive the joint distribution of the observations $\mathbf{y} = (y_1, \dots, y_T)'$. To this end, we rewrite (1)–(2) in matrix notations:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{H}_\psi \mathbf{u}, \tag{4}$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_T)'$, $\mathbf{u} = (u_1, \dots, u_T)' \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_y)$, $\mathbf{S}_y = \text{diag}(e^{h_1}, \dots, e^{h_T})$, and \mathbf{H}_ψ is a $T \times T$ lower triangular matrix with ones on the main diagonal, ψ_1 on first lower

diagonal, ψ_2 on second lower diagonal, and so forth. For example, for $q = 2$, we have

$$\mathbf{H}_\psi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \psi_1 & 1 & 0 & 0 & \cdots & 0 \\ \psi_2 & \psi_1 & 1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \psi_1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_2 & \psi_1 & 1 \end{pmatrix}.$$

It is important to note that in general \mathbf{H}_ψ is a banded $T \times T$ matrix that contains only $(T - q/2)(q + 1) < T(q + 1)$ non-zero elements, which is substantially less than T^2 for typical applications where $T \gg q$. This special structure can be exploited to speed up computation. For instance, obtaining the Cholesky decomposition of a banded $T \times T$ matrix with fixed bandwidth involves only $\mathcal{O}(T)$ operations as opposed to $\mathcal{O}(T^3)$ for a full matrix of the same size. Similar computational savings can be generated in operations such as multiplication, forward and backward substitution by using block-banded or sparse matrix algorithms.

Now, by a simple change of variable, it follows from (4) that

$$(\mathbf{y} \mid \boldsymbol{\psi}, \boldsymbol{\mu}, \mathbf{h}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}_\mathbf{y}),$$

where $\mathbf{h} = (h_1, \dots, h_T)'$, $\boldsymbol{\Omega}_\mathbf{y} = \mathbf{H}_\psi \mathbf{S}_\mathbf{y} \mathbf{H}'_\psi$. Since $\mathbf{S}_\mathbf{y} = \text{diag}(e^{h_1}, \dots, e^{h_T})$ is a diagonal matrix, and \mathbf{H}_ψ is a lower triangular sparse matrix, the product $\boldsymbol{\Omega}_\mathbf{y}$ is sparse. In fact, it is a banded matrix with only a narrow band of non-zero elements around the main diagonal. Moreover, since $|\mathbf{H}_\psi| = 1$ for any $\boldsymbol{\psi} = (\psi_1, \dots, \psi_q)'$, we have $|\boldsymbol{\Omega}_\mathbf{y}| = |\mathbf{S}_\mathbf{y}| = \exp\left(\sum_{t=1}^T h_t\right)$. The log joint density of \mathbf{y} is therefore given by

$$\log p(\mathbf{y} \mid \boldsymbol{\psi}, \boldsymbol{\mu}, \mathbf{h}) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T h_t - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Omega}_\mathbf{y}^{-1} (\mathbf{y} - \boldsymbol{\mu}). \quad (5)$$

It is important to realize that one need not obtain the $T \times T$ inverse matrix $\boldsymbol{\Omega}_\mathbf{y}^{-1}$ in order to evaluate the log density in (5) — it would involve $\mathcal{O}(T^3)$ operations. Instead, it can be computed in three steps, each of which requires only $\mathcal{O}(T)$ operations. To this end, we introduce the following notations: given a lower (upper) triangular $T \times T$ non-singular matrix \mathbf{A} and a $T \times 1$ vector \mathbf{c} , let $\mathbf{A} \setminus \mathbf{c}$ denote the unique solution to the triangular system $\mathbf{A}\mathbf{x} = \mathbf{c}$ obtained by forward (backward) substitution, i.e., $\mathbf{A} \setminus \mathbf{c} = \mathbf{A}^{-1}\mathbf{c}$. Now, we first obtain the Cholesky decomposition $\mathbf{C}_\mathbf{y}$ of the banded matrix $\boldsymbol{\Omega}_\mathbf{y}$ such that $\mathbf{C}_\mathbf{y} \mathbf{C}'_\mathbf{y} = \boldsymbol{\Omega}_\mathbf{y}$, which involves only $\mathcal{O}(T)$ operations. Then compute

$$\mathbf{x}_1 = \mathbf{C}'_\mathbf{y} \setminus (\mathbf{C}_\mathbf{y} \setminus (\mathbf{y} - \boldsymbol{\mu}))$$

by forward followed by backward substitution, each of which requires $\mathcal{O}(T)$ operations since $\mathbf{C}_\mathbf{y}$ is also banded. By definition,

$$\mathbf{x}_1 = \mathbf{C}_\mathbf{y}^{-1'} (\mathbf{C}_\mathbf{y}^{-1} (\mathbf{y} - \boldsymbol{\mu})) = (\mathbf{C}_\mathbf{y} \mathbf{C}'_\mathbf{y})^{-1} (\mathbf{y} - \boldsymbol{\mu}) = \boldsymbol{\Omega}_\mathbf{y}^{-1} (\mathbf{y} - \boldsymbol{\mu}).$$

Finally, compute

$$\mathbf{x}_2 = -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})' \mathbf{x}_1 = -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Omega}_{\mathbf{y}}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

which gives the quadratic term in (5). Thus, given $\boldsymbol{\mu}$, $\boldsymbol{\psi}$ and \mathbf{h} , one can efficiently evaluate the likelihood function without the need of the Kalman filter.

3.2 Posterior Analysis

Now, we discuss an efficient posterior sampler for estimating the moving average stochastic volatility model in (1)–(3). To keep the discussion concrete, we consider in particular the unobserved components specification; estimation for other conditional mean processes follows similarly. Specifically, the measurement equation is given by (4) with $\boldsymbol{\mu} = \boldsymbol{\tau} = (\tau_1, \dots, \tau_T)'$, whereas the transition equations are

$$\begin{aligned} \tau_t &= \tau_{t-1} + \varepsilon_t^\tau, & \varepsilon_t^\tau &\sim \mathcal{N}(0, \sigma_\tau^2), \\ h_t &= h_{t-1} + \varepsilon_t^h, & \varepsilon_t^h &\sim \mathcal{N}(0, \sigma_h^2). \end{aligned}$$

The transition equation for τ_t is initialized with $\tau_1 \sim \mathcal{N}(\tau_0, \sigma_{0\tau}^2)$, where τ_0 and $\sigma_{0\tau}^2$ are some known constants. Similarly, the transition equation for h_t is initialized with $h_1 \sim \mathcal{N}(h_0, \sigma_{0h}^2)$, where h_0 and σ_{0h}^2 are fixed. In particular, we set $\tau_0 = h_0 = 0$.

We assume independent priors for $\boldsymbol{\psi}$, σ_τ^2 , and σ_h^2 , i.e., $p(\boldsymbol{\psi}, \sigma_\tau^2, \sigma_h^2) = p(\boldsymbol{\psi})p(\sigma_\tau^2)p(\sigma_h^2)$. For $\boldsymbol{\psi}$, we consider a multivariate normal prior with support in the region where the invertibility conditions on $\boldsymbol{\psi}$ hold. As for the variances, we assume the following independent inverse-gamma priors:

$$\sigma_\tau^2 \sim \mathcal{IG}(\nu_\tau, S_\tau), \quad \sigma_h^2 \sim \mathcal{IG}(\nu_h, S_h).$$

Then posterior draws can be obtained by sequentially sampling from:¹

1. $p(\boldsymbol{\tau} \mid \mathbf{y}, \mathbf{h}, \boldsymbol{\psi}, \sigma_h^2, \sigma_\tau^2)$;
2. $p(\mathbf{h} \mid \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\psi}, \sigma_h^2, \sigma_\tau^2)$;
3. $p(\boldsymbol{\psi}, \sigma_h^2, \sigma_\tau^2 \mid \mathbf{y}, \boldsymbol{\tau}, \mathbf{h}) = p(\boldsymbol{\psi} \mid \mathbf{y}, \boldsymbol{\tau}, \mathbf{h}) p(\sigma_h^2 \mid \mathbf{y}, \boldsymbol{\tau}, \mathbf{h}) p(\sigma_\tau^2 \mid \mathbf{y}, \boldsymbol{\tau}, \mathbf{h})$.

We first discuss an efficient way to sample from $p(\boldsymbol{\tau} \mid \mathbf{y}, \mathbf{h}, \boldsymbol{\psi}, \sigma_h^2, \sigma_\tau^2)$. First note that by pre-multiplying (4) by $\mathbf{H}_\boldsymbol{\psi}^{-1}$, we have

$$\tilde{\mathbf{y}} = \tilde{\boldsymbol{\tau}} + \mathbf{u},$$

¹MATLAB codes for estimating this unobserved components moving average stochastic volatility model are available at <http://people.anu.edu.au/joshua.chan/>.

where $\tilde{\mathbf{y}} = \mathbf{H}_\psi^{-1} \mathbf{y}$ and $\tilde{\boldsymbol{\tau}} = \mathbf{H}_\psi^{-1} \boldsymbol{\tau}$. In other words, the log density for $\tilde{\mathbf{y}}$ is

$$\log p(\tilde{\mathbf{y}} | \boldsymbol{\psi}, \tilde{\boldsymbol{\tau}}, \mathbf{h}) \propto -\frac{1}{2} \sum_{t=1}^T h_t - \frac{1}{2} (\tilde{\mathbf{y}} - \tilde{\boldsymbol{\tau}})' \mathbf{S}_y^{-1} (\tilde{\mathbf{y}} - \tilde{\boldsymbol{\tau}}), \quad (6)$$

Hence it is more convenient to work with $\tilde{\boldsymbol{\tau}}$ instead of the original parameterization $\boldsymbol{\tau}$. Once we have a draw for $\tilde{\boldsymbol{\tau}}$, we simply need to pre-multiply it by \mathbf{H}_ψ to get a draw for $\boldsymbol{\tau}$.

Next, we derive the prior density for $\tilde{\boldsymbol{\tau}}$. To this end, we first obtain the prior density for $\boldsymbol{\tau}$. Rewrite the transition equation for τ_t in matrix notations:

$$\mathbf{H}\boldsymbol{\tau} = \boldsymbol{\varepsilon}^\tau,$$

where $\boldsymbol{\varepsilon}^\tau = (\varepsilon_1^\tau, \dots, \varepsilon_T^\tau)' \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_\tau)$, $\mathbf{S}_\tau = \text{diag}(\sigma_{0\tau}^2, \sigma_\tau^2, \dots, \sigma_\tau^2)$, and \mathbf{H} is the first difference matrix

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

That is, $(\boldsymbol{\tau} | \sigma_\tau^2) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_\tau)$, where $\boldsymbol{\Omega}_\tau^{-1} = \mathbf{H}' \mathbf{S}_\tau^{-1} \mathbf{H}$. Recall that $\sigma_{0\tau}^2$ is the variance for the initial underlying trend τ_1 , and is assumed to be a fixed hyper-parameter (although it is straightforward to treat it as a parameter).

It is important to realize that in this case the precision matrix $\boldsymbol{\Omega}_\tau^{-1}$ is sparse, in contrast to the case for \mathbf{y} where the covariance matrix $\boldsymbol{\Omega}_y$ is sparse. Now, by a simple change of variable, we have $(\tilde{\boldsymbol{\tau}} | \sigma_\tau^2) \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_\psi^{-1} \boldsymbol{\Omega}_\tau \mathbf{H}_\psi^{-1})$. Noting that $|\mathbf{H}| = |\mathbf{H}_\psi| = 1$ and $|\boldsymbol{\Omega}_\tau| = \sigma_{0\tau}^2 (\sigma_\tau^2)^{T-1}$, the log prior density for $\tilde{\boldsymbol{\tau}}$ is therefore given by

$$\log p(\tilde{\boldsymbol{\tau}} | \sigma_\tau^2) \propto -\frac{T-1}{2} \log \sigma_\tau^2 - \frac{1}{2} \tilde{\boldsymbol{\tau}}' \mathbf{H}'_\psi \boldsymbol{\Omega}_\tau^{-1} \mathbf{H}_\psi \tilde{\boldsymbol{\tau}}, \quad (7)$$

Combining (6) and (7), and using standard results from linear regression (see, e.g., Koop, 2003), we obtain the log conditional density $\log p(\tilde{\boldsymbol{\tau}} | \tilde{\mathbf{y}}, \mathbf{h}, \boldsymbol{\psi}, \sigma_h^2, \sigma_\tau^2)$ as follows:

$$\begin{aligned} \log p(\tilde{\boldsymbol{\tau}} | \tilde{\mathbf{y}}, \mathbf{h}, \boldsymbol{\psi}, \sigma_h^2, \sigma_\tau^2) &\propto -\frac{1}{2} (\tilde{\mathbf{y}} - \tilde{\boldsymbol{\tau}})' \mathbf{S}_y^{-1} (\tilde{\mathbf{y}} - \tilde{\boldsymbol{\tau}}) - \frac{1}{2} \tilde{\boldsymbol{\tau}}' \mathbf{H}'_\psi \boldsymbol{\Omega}_\tau^{-1} \mathbf{H}_\psi \tilde{\boldsymbol{\tau}} \\ &\propto -\frac{1}{2} (\tilde{\boldsymbol{\tau}}' (\mathbf{S}_y^{-1} + \mathbf{H}'_\psi \boldsymbol{\Omega}_\tau^{-1} \mathbf{H}_\psi) \tilde{\boldsymbol{\tau}} - 2 \tilde{\boldsymbol{\tau}}' \mathbf{S}_y^{-1} \tilde{\mathbf{y}}) \\ &\propto -\frac{1}{2} (\tilde{\boldsymbol{\tau}} - \hat{\boldsymbol{\tau}})' \mathbf{D}_{\tilde{\boldsymbol{\tau}}}^{-1} (\tilde{\boldsymbol{\tau}} - \hat{\boldsymbol{\tau}}), \end{aligned}$$

where $\mathbf{D}_{\tilde{\boldsymbol{\tau}}} = (\mathbf{S}_y^{-1} + \mathbf{H}'_\psi \boldsymbol{\Omega}_\tau^{-1} \mathbf{H}_\psi)^{-1}$, and $\hat{\boldsymbol{\tau}} = \mathbf{D}_{\tilde{\boldsymbol{\tau}}} \mathbf{S}_y^{-1} \tilde{\mathbf{y}}$. That is,

$$(\tilde{\boldsymbol{\tau}} | \tilde{\mathbf{y}}, \mathbf{h}, \boldsymbol{\psi}, \sigma_h^2, \sigma_\tau^2) \sim \mathcal{N}(\hat{\boldsymbol{\tau}}, \mathbf{D}_{\tilde{\boldsymbol{\tau}}}).$$

Since $\mathcal{N}(\hat{\boldsymbol{\tau}}, \mathbf{D}_{\hat{\boldsymbol{\tau}}})$ is typically high-dimensional, sampling from it using a brute-force approach is time-consuming. Here we adopt the precision-based sampling method in Chan and Jeliazkov (2009) to obtain draws from $\mathcal{N}(\hat{\boldsymbol{\tau}}, \mathbf{D}_{\hat{\boldsymbol{\tau}}})$ efficiently. To proceed, first note that the precision matrix $\mathbf{D}_{\hat{\boldsymbol{\tau}}}^{-1} = (\mathbf{S}_{\mathbf{y}}^{-1} + \mathbf{H}'_{\boldsymbol{\psi}} \boldsymbol{\Omega}_{\boldsymbol{\tau}}^{-1} \mathbf{H}_{\boldsymbol{\psi}})$ is sparse. Thus $\hat{\boldsymbol{\tau}}$ can be computed quickly using the same approach for evaluating the likelihood function discussed earlier: obtain the Cholesky decomposition $\mathbf{C}_{\hat{\boldsymbol{\tau}}}$ of $\mathbf{D}_{\hat{\boldsymbol{\tau}}}^{-1}$, and compute

$$\hat{\boldsymbol{\tau}} = \mathbf{C}'_{\hat{\boldsymbol{\tau}}} \backslash (\mathbf{C}_{\hat{\boldsymbol{\tau}}} \backslash (\mathbf{S}_{\mathbf{y}}^{-1} \tilde{\mathbf{y}}))$$

by forward and backward substitution. A draw from $\mathcal{N}(\hat{\boldsymbol{\tau}}, \mathbf{D}_{\hat{\boldsymbol{\tau}}})$ can now be obtained as follows: sample T independent standard normal draws $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and return

$$\tilde{\boldsymbol{\tau}} = \hat{\boldsymbol{\tau}} + \mathbf{C}'_{\hat{\boldsymbol{\tau}}} \backslash \mathbf{z}.$$

Since $\tilde{\boldsymbol{\tau}}$ is an affine transformation of a normal random vector, it is also a normal random vector. It is easy to check that its expectation is $\hat{\boldsymbol{\tau}}$ and its covariance matrix is

$$\mathbf{C}_{\hat{\boldsymbol{\tau}}}^{-1'} \mathbf{C}_{\hat{\boldsymbol{\tau}}}^{-1} = (\mathbf{C}_{\hat{\boldsymbol{\tau}}} \mathbf{C}'_{\hat{\boldsymbol{\tau}}})^{-1} = \mathbf{D}_{\hat{\boldsymbol{\tau}}}$$

as desired. Finally, given the draw $\tilde{\boldsymbol{\tau}} \sim \mathcal{N}(\hat{\boldsymbol{\tau}}, \mathbf{D}_{\hat{\boldsymbol{\tau}}})$, return $\boldsymbol{\tau} = \mathbf{H}_{\boldsymbol{\psi}} \tilde{\boldsymbol{\tau}}$.

In Step 2 of the MCMC sampler, we sample from $p(\mathbf{h} | \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\psi}, \sigma_h^2, \sigma_{\boldsymbol{\tau}}^2)$. To proceed, first define $\mathbf{y}^* = \mathbf{H}_{\boldsymbol{\psi}}^{-1}(\mathbf{y} - \boldsymbol{\tau})$. It follows from (4) that $\mathbf{y}^* = \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_{\mathbf{y}})$, where $\mathbf{S}_{\mathbf{y}} = \text{diag}(e^{h_1}, \dots, e^{h_T})$. The transition equation for h_t implies

$$\mathbf{H}\mathbf{h} = \boldsymbol{\varepsilon}^{\mathbf{h}},$$

where $\boldsymbol{\varepsilon}^{\mathbf{h}} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_{\mathbf{h}})$ and $\mathbf{S}_{\mathbf{h}} = \text{diag}(\sigma_{0h}^2, \sigma_h^2, \dots, \sigma_h^2)$. That is, $(\mathbf{h} | \sigma_h^2) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{h}})$, where $\boldsymbol{\Omega}_{\mathbf{h}}^{-1} = \mathbf{H}' \mathbf{S}_{\mathbf{h}}^{-1} \mathbf{H}$. With these transformations, the auxiliary mixture sampling approach in Kim, Shepherd, and Chib (1998) can be applied to draw \mathbf{h} efficiently; see also Koop and Korobilis (2010), p. 308–310, for a textbook treatment.

For Step 3, note that $\boldsymbol{\psi}, \sigma_h^2$, and $\sigma_{\boldsymbol{\tau}}^2$ are conditionally independent given $\mathbf{y}, \boldsymbol{\tau}$, and \mathbf{h} . Hence, we can sample each sequentially. Given the prior $p(\boldsymbol{\psi})$, it follows from (5) that

$$\begin{aligned} \log p(\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h}) &\propto \log p(\boldsymbol{\psi}) + \log p(\mathbf{y} | \boldsymbol{\psi}, \boldsymbol{\tau}, \mathbf{h}) \\ &\propto \log p(\boldsymbol{\psi}) - \frac{1}{2} \mathbf{y}^{*'} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{y}^*, \end{aligned}$$

where we use the transformation $\mathbf{y}^* = \mathbf{H}_{\boldsymbol{\psi}}^{-1}(\mathbf{y} - \boldsymbol{\tau})$. Hence, $\log p(\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h})$ can be quickly evaluated for any $\boldsymbol{\psi}$ given $\mathbf{y}, \boldsymbol{\tau}$ and \mathbf{h} using the method discussed in the previous subsection. Since in typical applications $\boldsymbol{\psi}$ is low-dimensional — only a few moving average terms are needed — one can maximize $\log p(\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h})$ numerically and obtain the mode and the negative Hessian evaluated at the mode, denoted as $\hat{\boldsymbol{\psi}}$ and \mathbf{K} , respectively. Then, draws from $p(\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h})$ are obtained using an independence-chain Metropolis-Hasting step with proposal density $\mathcal{N}(\boldsymbol{\psi}, \mathbf{K}^{-1})$, for example. Finally, both $p(\sigma_h^2 | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h})$ and

$p(\sigma_\tau^2 | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h})$ are inverse-gamma densities, and can therefore be sampled using standard methods. In fact, we have

$$\begin{aligned} (\sigma_\tau^2 | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h}) &\sim \mathcal{IG} \left(\nu_\tau + (T - 1)/2, S_\tau + \sum_{t=2}^T (\tau_t - \tau_{t-1})^2/2 \right) \\ (\sigma_h^2 | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h}) &\sim \mathcal{IG} \left(\nu_h + (T - 1)/2, S_h + \sum_{t=2}^T (h_t - h_{t-1})^2/2 \right). \end{aligned}$$

We note that the estimation for other conditional mean processes can be implemented analogously; we provide the estimation details for the autoregressive moving average stochastic volatility model in the appendix.

4 Modeling and Forecasting U.S. Inflation Rate

Now we use the proposed moving average stochastic volatility models to analyze the behavior of U.S. quarterly CPI inflation, and contrast the results with those obtained from the standard variants with only stochastic volatility. In addition, we also compare the forecast performance of the two classes of models at various forecast horizons. Since allowing for stochastic volatility is found to be empirically important, all the models considered in the section have stochastic volatility in the measurement equation.

4.1 Competing Models

We consider three popular specifications for modeling inflation, and for each we have two versions: with and without the MA component. The primary goal of this exercise is not to find the best model *per se*. Rather, our objective is to investigate if the addition of the MA component improves model-fit and forecast performance, and how it affects the estimates of the states and other parameters of interest. The first specification is the unobserved components model, which we reproduce here for convenience:

$$\begin{aligned} y_t &= \tau_t + u_t, & u_t &\sim \mathcal{N}(0, e^{h_t}), \\ \tau_t &= \tau_{t-1} + \varepsilon_t^\tau, & \varepsilon_t^\tau &\sim \mathcal{N}(0, \sigma_\tau^2), \\ h_t &= h_{t-1} + \varepsilon_t^h, & \varepsilon_t^h &\sim \mathcal{N}(0, \sigma_h^2). \end{aligned}$$

We refer this version of unobserved components model as the **UC** model. Stock and Watson (2007) extend this specification to include stochastic volatility in the transition equation for the underlying trend τ_t . Specifically, the variance for ε_t^τ , instead of being fixed to be constant, is allowed to be time-varying: $\varepsilon_t^\tau \sim \mathcal{N}(0, e^{g_t})$. The log-volatility g_t , in turn, evolves as a random walk:

$$g_t = g_{t-1} + \varepsilon_t^g, \quad \varepsilon_t^g \sim \mathcal{N}(0, \sigma_g^2).$$

This version of unobserved components model, where both the measurement and transition equations have stochastic volatility, is referred to as the **UCSV** model. The third specification is the **AR** model, where the measurement equation is an $\text{AR}(p)$ process

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t,$$

where u_t has the same stochastic volatility specification as before. Throughout we set $p = 2$ for simplicity. In addition, we impose the conditions that the roots of the characteristic polynomial associated with the AR coefficients all lie outside the unit circle, so that the AR process is stationary.

Table 1: A list of competing models.

| Model | Description |
|----------------|---|
| UC | unobserved components model |
| UC-MA | same as UC but with an MA(1) component |
| UCSV | same as UC but the state equation for τ_t has stochastic volatility |
| UCSV-MA | same as UCSV but with an MA(1) component |
| AR | autoregressive model with 2 lags |
| AR-MA | same as AR but with an MA(1) component |

For each of the three models — **UC**, **UCSV** and **AR** — we include the moving average stochastic volatility variants as specified in (1)–(2), and we refer them as **UC-MA**, **UCSV-MA** and **AR-MA** respectively. For these models with the moving average components, we set q , the number of moving average terms, to be one. We find little benefits, at least for this empirical exercise, of having additional moving average terms. We summarize all six specifications in Table 1.

4.2 Data and Prior

The data consist of U.S. quarterly CPI inflation from 1947Q1 to 2011Q3. More specifically, given the quarterly CPI figures z_t , we use $y_t = 400 \log(z_t/z_{t-1})$ as the CPI inflation. For easy comparison, we choose broadly similar priors across models. In particular, we use exactly the same priors for the common model parameters in each pair of models with and without the MA component. For the MA coefficient ψ_1 , we assume the normal prior $\psi_1 \sim \mathcal{N}(\psi_0, V_\psi)$ with the restriction $|\psi_1| < 1$, i.e., the MA process is assumed to be invertible. We set $\psi_0 = 0$ and $V_\psi = 5$. The prior distribution thus centers around 0 and has support within the interval $(-1, 1)$. Given the large prior variance, it is also relatively non-informative.

As discussed in Section 3.2, we assume independent inverse-gamma priors for σ_τ^2 and σ_h^2 : $\sigma_\tau^2 \sim \mathcal{IG}(\nu_\tau, S_\tau)$ and $\sigma_h^2 \sim \mathcal{IG}(\nu_h, S_h)$. We choose relatively small — hence relatively non-informative — values for the degrees of freedom parameters: $\nu_\tau = \nu_h = 10$. For the scale parameters, we set $S_\tau = 0.18$ and $S_h = 0.45$. These values imply $\mathbb{E}\sigma_\tau^2 =$

0.141^2 and $\mathbb{E} \sigma_h^2 = 0.224^2$. 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 et al. (2012b) and Stock and Watson (2007). For the **UCSV** and **UCSV-MA** models, we also need to specify a prior for σ_g^2 . We assume $\sigma_g^2 \sim \mathcal{IG}(\nu_g, S_g)$, where $\nu_g = 10$ and $S_g = 0.45$. This prior is the same as that for σ_h^2 . Finally, we assume a normal prior for the AR coefficients $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2)'$ in the **AR** and **AR-MA** models: $\boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\phi}_0, \mathbf{V}_\phi)$, where $\boldsymbol{\phi}_0 = \mathbf{0}$ and $\mathbf{V}_\phi = 5 \times \mathbf{I}$, with the restrictions that the characteristic polynomial associated with $\boldsymbol{\phi}$ has all its roots outside the unit circle, so that the corresponding AR process is stationary.

4.3 Full Sample Estimation Results

We present in this subsection the empirical results for the six models listed in Table 1, obtained using the full sample from 1947Q1 to 2011Q3. All the posterior moments and quantiles are based on 50,000 draws from the MCMC algorithm introduced in Section 3.2 after a burnin period of 5,000.

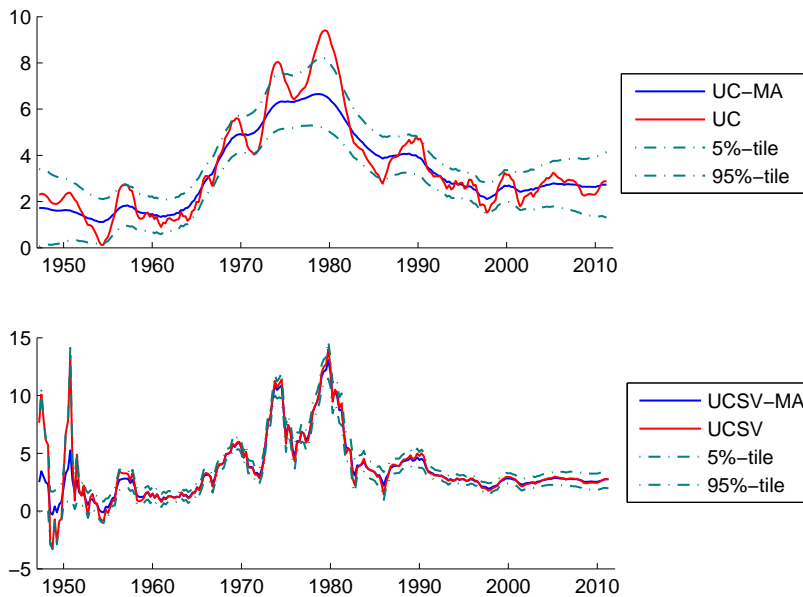


Figure 1: Posterior means and quantiles for τ_t under the **UC**, **UC-MA**, **UCSV** and **UCSV-MA** models. The posterior quantiles are obtained under the MA variants.

Figure 1 presents point estimates (posterior means) of the underlying inflation τ_t for four models, as well as credible intervals obtained under the MA variants. It is clear from the figure that there are large differences between estimates under the **UC** and **UC-MA** models. In particular, by allowing for an extra channel for persistence through the moving

average errors, the latter model produces much smoother estimates, which are more in line with the notion of a smooth, gradually changing underlying inflation. This finding is broadly consistent with those reported in earlier studies, such as Clark and Doh (2011) and Chan et al. (2012b), who also find that by explicitly modeling short-run dynamics one often obtains smoother, more reasonable, underlying inflation estimates.

On the other hand, the underlying inflation estimates for the **UCSV** and **UCSV-MA** models are very similar, except for the early sample where those for the former appear to be more erratic. By allowing the volatility in the transition equation for τ_t to be time-varying, the model attributes much of the variation in observed inflation to variation in τ_t , with the consequence that it traces closely the actual inflation. As such, the moving average errors in the measurement equation play a lesser role in channeling the short-run dynamics of the series. The erratic underlying inflation estimates, which change rapidly in short periods, cast doubt on the overall appropriateness of the **UCSV** model. Regardless, for our purpose it suffices to note that the MA component does help to obtain smoother estimates for the underlying inflation, though to a lesser extent than in the **UC** model.

In Figure 2 we report the estimates for the log-volatilities h_t . Under all specifications we find substantial variation in the inflation volatility estimates, highlighting the importance of allowing for stochastic volatility in the measurement equation. In addition, these results also suggest that the log-volatility estimates are somewhat sensitive to modeling assumptions, especially the treatment of the error process in the **UCSV** model.

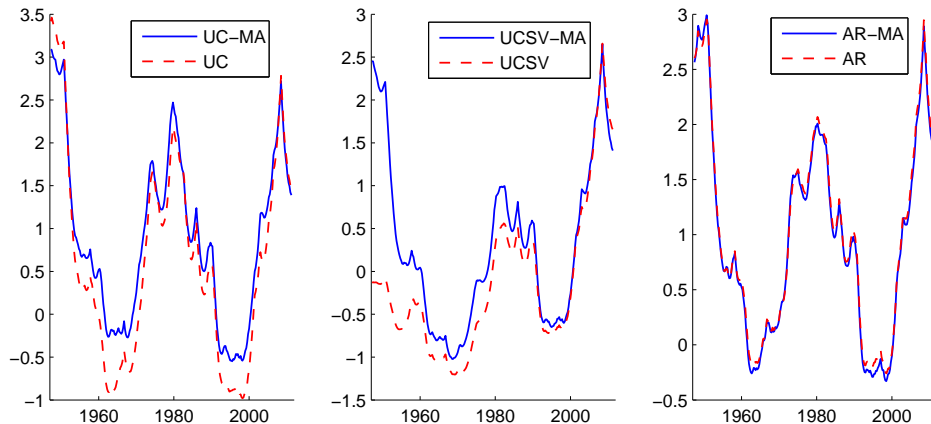


Figure 2: Posterior means for h_t under the six models.

A key parameter of interest is the MA coefficient ψ_1 — if the posterior density for ψ_1 is concentrated around 0, it would indicate that the MA component might not be necessary. Table 2 reports the posterior means and standard deviations for ψ_1 , as well as the probabilities that ψ_1 is positive, under the three moving average stochastic volatility models. It is clear from the table that the posterior means for ψ_1 are all quite large in absolute value. For instance, the MA coefficient is estimated to be about 0.463 in the **UC-MA** model, indicating that even after controlling for the underlying inflation and stochastic

volatility, there is still considerable positive autocorrelation in the observed inflation. In fact, the posterior probability that ψ_1 is positive is estimated to be 1, i.e., out of the 50,000 posterior draws for ψ_1 obtained, none are less than zero.

Table 2: Posterior means, standard deviations, and the probabilities of being positive for ψ_1 under the **UC-MA**, **UCSV-MA**, and **AR-MA** models.

| | UC-MA | UCSV-MA | AR-MA |
|--|--------------|----------------|--------------|
| $\mathbb{E}(\psi_1 \mathbf{y})$ | 0.463 | 0.267 | -0.382 |
| $\sqrt{\text{Var}(\psi_1 \mathbf{y})}$ | 0.070 | 0.122 | 0.133 |
| $\mathbb{P}(\psi_1 > 0 \mathbf{y})$ | 1.00 | 0.976 | 0.005 |

Another interesting point to note is that under both versions of unobserved components models, ψ_1 is estimated to be positive, whereas its estimate is negative under the **AR-MA** model. This difference is perhaps not surprising: in the former two models the autocorrelation of the observed inflation is modeled only through the random walk underlying trend, which is seemingly not sufficient to capture all the observed positive autocorrelation. In contrast, past observed inflation rates enter directly the conditional mean process in **AR-MA**. The result appears to suggest that this AR component (with two lags) “over-captures” the observed positive autocorrelation, which in turn induces a negative autocorrelation in the residual.

We report in Figure 3 the kernel density estimates for ψ_1 under the three models. The graph in the left panel shows that most of the mass for ψ_1 is concentrated between 0.2 and 0.7 in **UC-MA**. The posterior densities for the remaining two models tell a similar story — ψ_1 is unlikely to be a small value around zero. Taken together, these results suggest the relevance of extending the standard stochastic volatility models to include an MA component.

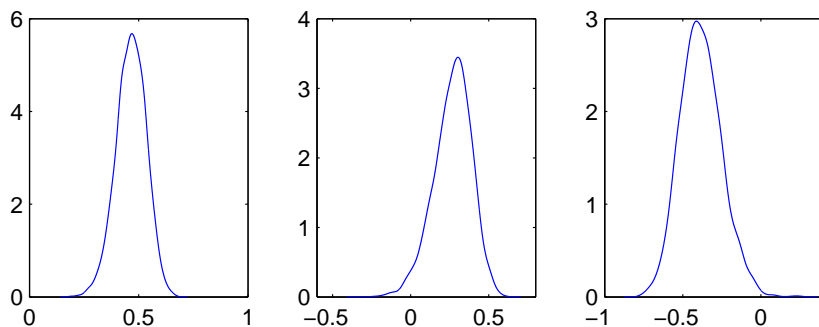


Figure 3: Kernel density estimates for ψ_1 under the **UC-MA** (left), **UCSV-MA** (middle), and **AR-MA** (right) models.

The results in Table 2 and Figure 3 may be viewed as suggestive evidence showing the relevance of the proposed MA component in the stochastic volatility models. In what

follows, we perform a formal Bayesian model comparison exercise to compare each pair of stochastic volatility models (i.e., with and without the MA component) using Bayes factors (see, e.g., Koop, 2003, p. 3–4). Since we are comparing nested models, the Bayes factor in favor of the model that has the MA component against the standard variant can be computed using the Savage-Dickey density ratio (Verdinelli and Wasserman, 1995):

$$\text{BF} = \frac{p(\psi_1 = 0)}{p(\psi_1 = 0 | \mathbf{y})}.$$

In other words, we simply need to evaluate the marginal prior and posterior densities for ψ_1 at 0. The ratio of the two values then gives the relevant Bayes factor. The numerator density is a univariate truncated normal, and can be easily evaluated. The denominator density, however, is not of standard form. In fact, neither is the full conditional density $p(\psi_1 | \mathbf{y}, \mathbf{h}, \boldsymbol{\mu})$. Hence, the usual trick of approximating $p(\psi_1 = 0 | \mathbf{y})$ using the Monte Carlo average

$$\frac{1}{R} \sum_{i=1}^R p(\psi_1 = 0 | \mathbf{y}, \mathbf{h}^{(i)}, \boldsymbol{\mu}^{(i)})$$

by summing over the posterior draws $\mathbf{h}^{(i)}$ and $\boldsymbol{\mu}^{(i)}$ cannot be applied. Instead, we approximate $p(\psi_1 = 0 | \mathbf{y})$ with a Gaussian kernel estimator obtained using the posterior draws for ψ_1 (see, e.g., Geweke, 2010, p 84–85). The results are reported in Table 3. For each of the three pairwise comparisons — **UC-MA** against **UC**, **UCSV-MA** against **UCSV**, and **AR-MA** against **AR** — there is strong evidence that the data prefer the variants with the MA component, and in the case for **UC-MA** vs **UC**, the evidence is overwhelming.

Table 3: Bayes factors in favor of the proposed models against the standard variants with only stochastic volatility.

| | UC-MA | UCSV-MA | UC-MA |
|----|------------------------|----------------|--------------|
| BF | 2.15×10^{127} | 4.56 | 6.20 |

Remember that each pair of the models only differs in an extra parameter ψ_1 , and the stochastic volatility models are standard in the literature. Given the context, these full sample estimation results present strong evidence in favor of the proposed models against their counterparts without the MA component. Not only do the former models fit the data better, they also give more sensible underlying inflation estimates. In the next subsection, we present forecasting results that show the proposed models also provide more accurate point and density forecasts.

4.4 Forecasting Results

We now perform a recursive out-of-sample forecast exercise to evaluate the performance of the models listed in Table 1 for forecasting U.S. quarterly CPI inflation at various

horizons. We use each of the models to produce both point and density k -step-ahead iterated forecasts with $k = 1, 4, 8, 12$ and 16 . Specifically, given the data up to time t , denoted as $\mathbf{y}_{1:t}$, we use the predictive mean $\mathbb{E}(y_{t+k} | \mathbf{y}_{1:t})$ as the point forecast and the predictive density $p(y_{t+k} | \mathbf{y}_{1:t})$ as the density forecast. These forecasts are then evaluated for $t = t_0, \dots, T - k$, where t_0 is 1975Q1.

In practice, neither the predictive mean nor the predictive density of y_{t+k} can be computed analytically. Instead, they are obtained using predictive simulation. More precisely, at every MCMC iteration, given the model parameters and states (up to time t), we simulate future states from time $t + 1$ to $t + k$ using the relevant transition equations. We also simulate future errors $u_s \sim \mathcal{N}(0, e^{h_s})$ for $s = t + 1, \dots, t + k - 1$. Given these draws, y_{t+k} is a normal random variable as specified in (1), and one can easily produce the point and density forecasts for y_{t+k} . Hence, we have a pair of forecasts (point and density) at every MCMC iteration. These forecasts are then averaged over all the posterior draws to produce estimates for $\mathbb{E}(y_{t+k} | \mathbf{y}_{1:t})$ and $p(y_{t+k} | \mathbf{y}_{1:t})$. The whole exercise is then repeated using data up to time $t + 1$ to produce $\mathbb{E}(y_{t+k+1} | \mathbf{y}_{1:t+1})$ and $p(y_{t+k+1} | \mathbf{y}_{1:t+1})$, and so forth.

Let y_{t+k}^o denote the observed value of y_{t+k} that is known at time $t + k$. The metric used to evaluate the point forecasts is the root mean squared forecast error (RMSFE) defined as

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

To evaluate the density forecast $p(y_{t+k} | \mathbf{y}_{1:t})$, one natural measure is the predictive likelihood $p(y_{t+k} = y_{t+k}^o | \mathbf{y}_{1:t})$, i.e., the predictive density of y_{t+k} evaluated at the observed value y_{t+k}^o . Clearly, if the actual outcome y_{t+k}^o is unlikely under the density forecast, the value of the predictive likelihood will be small, and vice versa; see, e.g., Geweke and Amisano (2011) for a more detailed discussion of the predictive likelihood and its connection to the marginal likelihood. We evaluate the density forecasts using the mean of log predictive likelihoods:

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

Table 4 presents the point forecast results for the six models. For easy comparison, we report the RMSFEs for **UC** at different horizons; in other rows the ratios of RMSFEs of a certain model to those of **UC** are reported. Hence, values smaller than unity indicate better forecast performance than **UC**.

A few broad observations can be drawn from these forecasting results. First, for all the specifications considered, there is clear evidence that the proposed models perform better than the standard variants for short-horizon forecasts. For instance, in comparing one-quarter-ahead forecasts, the RMSFE for **UC-MA** is 92.7% of the value for **UC**. Remember that the latter is among the best forecasting models in the literature. Clearly, these results provide strong evidence that the addition of the MA component gives substantial benefits. Second, for medium-horizon forecasts (e.g., $k \geq 8$), each of the two

pairs — **UCSV** and **UCSV-MA**; **AR** and **AR-MA** — give almost identical RMSFEs. For **UC-MA**, however, it consistently gives better point forecasts compared to **UC**, even at four-year forecast horizon. This might seem surprising at first glance, as the MA component only models short-run dynamics. However, the results in Section 4.3, especially Figure 1, suggest one explanation: by including the MA component, one obtains smoother and more reasonable estimates for underlying inflation, which in turn help produce better forecasts at longer horizons.

Table 4: RMSFEs for forecasting quarterly CPI inflation: level for **UC**, and ratio of RMSFE to that of **UC** for all other models

| | $k = 1$ | $k = 4$ | $k = 8$ | $k = 12$ | $k = 16$ |
|----------------|---------|---------|---------|----------|----------|
| UC | 2.332 | 2.703 | 3.112 | 3.354 | 3.412 |
| UC-MA | 0.927 | 0.984 | 0.937 | 0.916 | 0.928 |
| UCSV | 0.930 | 0.978 | 1.057 | 1.071 | 1.066 |
| UCSV-MA | 0.920 | 0.973 | 1.051 | 1.069 | 1.060 |
| AR | 0.961 | 0.958 | 0.960 | 0.954 | 0.942 |
| AR-MA | 0.942 | 0.951 | 0.965 | 0.956 | 0.947 |

It is of interest to note that **UCSV** performs better than **UC** at short-horizon forecasts but worse at longer horizons. This is consistent with the estimation results presented in Figure 1: by including stochastic volatility in the transition for τ_t , the estimates for underlying inflation under **UCSV** traces closely the actual inflation, which is good for short-horizon forecasts but not for longer horizons.

It is also worth mentioning that the simple **AR** model performs better than **UCSV** at all forecast horizons except $k = 1$. This seems to contradict the results in, e.g., Stock and Watson (2007), who find the **UCSV** model outperforms simple ARMA models. However, our **AR** model differs from those ARMA models in two key aspects. First, in our specification the stationary conditions on the AR coefficients are imposed. This is important for medium- and long-horizon forecasts — without the stationary conditions, point forecasts tend to be bad due to the presence of explosive draws in the AR process, as documented in Chan et al. (2012b). Second, the error variance has a stochastic volatility specification in **AR** as opposed to being fixed as a constant in those ARMA models. Since there is substantial time-variation in inflation volatility, models that do not allow for time-varying volatility tend to produce inferior forecasts. Our results suggest that by correctly specifying the error variance and imposing reasonable stationary conditions, simple ARMA models can outperform more complex models such as the unobserved components models.

We report the density forecast results in Table 4, which paint a very similar picture. In particular, the addition of the MA component provides substantial benefits for short-horizon forecast and in some cases for longer horizons also.

Table 5: Average log predictive likelihoods for forecasting quarterly CPI inflation

| | $k = 1$ | $k = 4$ | $k = 8$ | $k = 12$ | $k = 16$ |
|---------|---------|---------|---------|----------|----------|
| UC | -2.088 | -2.332 | -2.490 | -2.548 | -2.592 |
| UC-MA | -2.046 | -2.287 | -2.381 | -2.385 | -2.432 |
| UCSV | -2.052 | -2.323 | -2.494 | -2.562 | -2.624 |
| UCSV-MA | -2.047 | -2.313 | -2.483 | -2.549 | -2.622 |
| AR | -2.068 | -2.289 | -2.423 | -2.470 | -2.510 |
| AR-MA | -2.048 | -2.343 | -2.511 | -2.566 | -2.603 |

5 Concluding Remarks and Future Research

With the aim to expand the toolkit for analyzing time series data, we have introduced a new class of models that generalizes the standard SV specification to include a moving average component. The addition of the MA component leads to models that are more robust to misspecification, which often translates into better forecast performance in practice. A drawback of the new models is that the estimation is more difficult, as the MA component induces serial dependence in the observations. In view of this, we introduce a novel algorithm that exploits the special structure of the models — that the covariance matrix for the observations is sparse, and therefore fast routines for sparse matrices can be used.

We illustrate the relevance of the new class of models with an empirical application involving U.S. CPI inflation. Our empirical results show that the data strongly favor the new models over the standard variants with only SV. In a recursive forecast exercise, there is clear evidence that the new models deliver improvements in out-of-sample forecasts. A second finding is that by allowing for SV in the error variance and imposing stationary conditions on the conditional mean process, parsimonious ARMA models can outperform more complex models such as the unobserved components models. In this paper we have only considered univariate models. For future research it would be interesting to formulate multivariate versions of the proposed MA-SV models, using, e.g., factors with SV, and extend the new estimation methods to those settings.

Appendix

In this appendix we outline a posterior simulator for estimating the **AR-MA** model:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + u_t + \psi_1 u_{t-1} + \cdots + \psi_q u_{t-q},$$

$$h_t = h_{t-1} + \varepsilon_t^h,$$

where $u_t \sim \mathcal{N}(0, e^{h_t})$ and $\varepsilon_t^h \sim \mathcal{N}(0, \sigma_h^2)$. Let $\phi = (\phi_0, \phi_1, \dots, \phi_p)'$, and we assume independent priors for ϕ , ψ , and σ_h^2 . The priors for ψ and σ_h^2 are given in Section 3.2.

As for ϕ , we take the following truncated multivariate normal prior

$$\phi \sim \mathcal{N}(\phi_0, \mathbf{V}_\phi) \mathbf{1}(\phi \in \mathbf{A}_\phi),$$

where $\phi_0 = \mathbf{0}$, $\mathbf{V}_\phi = 5 \times \mathbf{I}$ and \mathbf{A}_ϕ is the set where the roots of the autoregressive polynomial associated with ϕ are all outside the unit circle. The posterior simulator sequentially draws from: (1) $p(\phi | \mathbf{y}, \mathbf{h}, \boldsymbol{\psi}, \sigma_h^2)$; (2) $p(\mathbf{h} | \mathbf{y}, \phi, \boldsymbol{\psi}, \sigma_h^2)$; and (3) $p(\boldsymbol{\psi}, \sigma_h^2 | \mathbf{y}, \phi, \mathbf{h}) = p(\boldsymbol{\psi} | \mathbf{y}, \phi, \mathbf{h}) p(\sigma_h^2 | \mathbf{y}, \phi, \mathbf{h})$. Steps 2 and 3 can be implemented as in the sampler given in Section 3.2, and here we focus on Step 1. To this end, we first write the observation equation in matrix form:

$$\mathbf{y} = \mathbf{X}\phi + \mathbf{H}_\psi \mathbf{u},$$

where \mathbf{X} is a $T \times p$ matrix of lagged observations appropriately defined. Now let $\tilde{\mathbf{y}} = \mathbf{H}_\psi^{-1} \mathbf{y}$ and $\tilde{\mathbf{X}} = \mathbf{H}_\psi^{-1} \mathbf{X}$, both of which can be computed quickly as \mathbf{H}_ψ is banded. Then,

$$(\tilde{\mathbf{y}} | \boldsymbol{\psi}, \phi, \mathbf{h}) \sim \mathcal{N}(\tilde{\mathbf{X}}\phi, \mathbf{S}_y),$$

where $\mathbf{S}_y = \text{diag}(e^{h_1}, \dots, e^{h_T})$. It follows that

$$(\phi | \mathbf{y}, \mathbf{h}, \boldsymbol{\psi}, \sigma_h^2) \sim \mathcal{N}(\hat{\phi}, \mathbf{D}_\phi) \mathbf{1}(\phi \in \mathbf{A}_\phi),$$

where $\mathbf{D}_\phi = (\tilde{\mathbf{X}}' \mathbf{S}_y^{-1} \tilde{\mathbf{X}} + \mathbf{V}_\phi^{-1})^{-1}$ and $\hat{\phi} = \mathbf{D}_\phi \tilde{\mathbf{X}}' \mathbf{S}_y^{-1} \tilde{\mathbf{y}}$. A draw from the truncated density can then be obtained via the acceptance-rejection sampling (e.g., Kroese, Taimre, and Botev, 2011, chapter 3) with proposal density $\mathcal{N}(\hat{\phi}, \mathbf{D}_\phi)$. This step is efficient when ϕ is low-dimensional, as is often the case.

References

- G. Athanasopoulos and F. Vahid. VARMA versus VAR for macroeconomic forecasting. *Journal of Business and Economic Statistics*, 26(2):237–252, 2008.
- T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.
- 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, G. Koop, R. Leon-Gonzalez, and R. Strachan. Time varying dimension models. *Journal of Business and Economic Statistics*, 2012a. Forthcoming.
- J. C. C. Chan, G. Koop, and S. M. Potter. A new model of trend inflation. Working paper, Centre for Applied Macroeconomic Analysis, Australian National University, 2012b.
- S. Chib and E. Greenberg. Bayes inference in regression models with ARMA(p, q) errors. *Journal of Econometrics*, 64(12):183 – 206, 1994.

- S. Chib, F. Nardari, and N. Shephard. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316, 2002.
- S. Chib, F. Nardari, and N. Shephard. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341371, 2006.
- T. E. Clark and T. Doh. A Bayesian evaluation of alternative models of trend inflation. Working paper, Federal Reserve Bank of Cleveland, 2011.
- 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 and A. Sbordone. Trend inflation, indexation, and inflation persistence in the new keynesian phillips curve. *American Economic Review*, 98:2101–2126, 2008.
- T. Cogley, G. Primiceri, and T. Sargent. Inflation-gap persistence in the U.S. *American Economic Journal: Macroeconomics*, 2:43–69, 2010.
- J. Geweke. *Complete and Incomplete Econometric Models*. Princeton University Press, Princeton, 2010.
- J. Geweke and G. Amisano. Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26:1–29, 2011.
- A. C. Harvey. Trends and cycles in macroeconomic time series. *Journal of Business and Economic Statistics*, 3(3):216–227, 1985.
- E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185212, 2004.
- M. J. Jensen and J. M. Maheu. Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics*, 157(2):306–316, 2010.
- 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. *Bayesian Econometrics*. Wiley & Sons, New York, 2003.
- G. Koop and D. Korobilis. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4):267–358, 2010.
- G. Koop and D. Korobilis. Forecasting inflation using dynamic model averaging. *International Economic Review*, 2012. Forthcoming.
- G. Koop and S. M. Potter. Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, 74:763–789, 2007.
- D. Korobilis. VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, 2012. Forthcoming.

- D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. John Wiley & Sons, New York, 2011.
- J. Nakajima and Y. Omori. Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t -distribution. *Computational Statistics & Data Analysis*, 2012. Forthcoming.
- T. Nakatsuma. Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach. *Journal of Econometrics*, 95(1):57–69, 2000.
- Y. Omori, S. Chib, N. Shephard, and J. Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449, 2007.
- G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852, 2005.
- 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.
- S. J. Taylor. Modelling stochastic volatility. *Mathematical Finance*, 4:183–204, 1994.
- I. Verdinelli and L. Wasserman. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, 90(430):614–618, 1995.