

# CAMA

Centre for Applied Macroeconomic Analysis

---

## Stochastic Volatility Models with ARMA Innovations: An Application to G7 Inflation Forecasts

---

CAMA Working Paper 32/2018  
June 2018

**Bo Zhang**

Research School of Economics, ANU

**Joshua C.C. Chan**

University of Technology Sydney

Centre for Applied Macroeconomic Analysis, ANU

**Jamie L. Cross**

Research School of Economics, ANU

### Abstract

We introduce a new class of stochastic volatility models with autoregressive moving average (ARMA) innovations. The conditional mean process has a flexible form that can accommodate both a state space representation and a conventional dynamic regression. The ARMA component introduces serial dependence which renders standard Kalman filter techniques not directly applicable. To overcome this hurdle we develop an efficient posterior simulator that builds on recently developed precision-based algorithms. We assess the usefulness of these new models in an inflation forecasting exercise across all G7 economies. We find that the new models generally provide competitive point and density forecasts compared to standard benchmarks, and are especially useful for Canada, France, Italy and the US.

## Keywords

autoregressive moving average errors, stochastic volatility, inflation forecast, state space models, unobserved components model

## JEL Classification

C11, C52, C53, E37

## Address for correspondence:

(E) [cama.admin@anu.edu.au](mailto:cama.admin@anu.edu.au)

ISSN 2206-0332

[The Centre for Applied Macroeconomic Analysis](#) in the Crawford School of Public Policy has been established to build strong links between professional macroeconomists. It provides a forum for quality macroeconomic research and discussion of policy issues between academia, government and the private sector.

**The Crawford School of Public Policy** is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

# Stochastic Volatility Models with ARMA Innovations: An Application to G7 Inflation Forecasts

Bo Zhang

Research School of Economics,  
Australian National University

Joshua C. C. Chan

University of Technology Sydney  
and CAMA

Jamie L. Cross

Research School of Economics,  
Australian National University

June 2018

## **Abstract**

We introduce a new class of stochastic volatility models with autoregressive moving average (ARMA) innovations. The conditional mean process has a flexible form that can accommodate both a state space representation and a conventional dynamic regression. The ARMA component introduces serial dependence which renders standard Kalman filter techniques not directly applicable. To overcome this hurdle we develop an efficient posterior simulator that builds on recently developed precision-based algorithms. We assess the usefulness of these new models in an inflation forecasting exercise across all G7 economies. We find that the new models generally provide competitive point and density forecasts compared to standard benchmarks, and are especially useful for Canada, France, Italy and the US.

**Keywords:** autoregressive moving average errors, stochastic volatility, inflation forecast, state space models, unobserved components model.

**JEL classification codes:** C11, C52, C53, E37

# 1 Introduction

Following the seminal work of Box and Jenkins (1970), autoregressive moving average (ARMA) models have become the standard tool for modeling and forecasting univariate time series. More recently, coefficient instability in macroeconomic time series models has been widely acknowledged (see, e.g., Stock and Watson, 1996; Ludbergh et al., 2003; Marcellino, 2004; Stock and Watson, 2007; Cross and Poon, 2016). For example, Stock and Watson (2007) show that US CPI inflation is best modeled by an unobserved components model in which both the transitory and trend equations allow for time-varying volatility. It is therefore necessary to extend the ARMA framework to allow for the possibility of heteroscedasticity.

There are two popular approaches to achieve this objective: autoregressive conditional heteroscedasticity models of Engle (1982)—or their generalized counterparts introduced in Bollerslev (1986) called GARCH models—and stochastic volatility (SV) models (Taylor, 1994). In a recent paper, Clark and Ravazzolo (2015) put these two classes of models head-to-head in a forecasting exercise involving a few key US macroeconomic time series. They find that stochastic volatility models generally provide superior point and density forecasts across all variables. Thus, SV seems to be a more appropriate specification, at least for macroeconomic forecasting. In addition, given the historical success of ARMA models and their more recent time-varying extensions, one might expect a more flexible class of dynamic models with ARMA and SV errors could further improve the forecast performance. This idea is partially investigated by Chan (2013), who shows that MA-SV errors are useful in forecasting US inflation. Nonetheless, it remains to be seen whether more general ARMA-SV errors can further enhance forecast accuracy.<sup>1</sup>

With this idea in mind, our objective in this paper is to investigate the forecast performance of a new class of ARMA-SV models. By allowing the conditional mean process to have a flexible state space representation, our general framework is able to accommodate numerous popular specifications, such as unobserved components and time varying parameter models, as special cases. Thus, from a methodological perspective, our models can be viewed as an extension of the unobserved components model in Stock and Watson (2007) to allow for ARMA errors. In addition, we also extend the recent work of Chan (2013) to allow for a more flexible error structure.

---

<sup>1</sup>Nakatsuma (2000) develops a posterior simulator for estimating ARMA-GARCH models. Thus, our paper can also be viewed as filling an important gap in completing the econometricians toolbox of possible ARMA error models.

A second contribution of our paper is to develop an efficient posterior simulator to estimate this new class of models. A major computational hurdle is that the ARMA component introduces serial dependence into the measurement equation, which makes standard Kalman filter techniques not directly applicable. To overcome this issue, one may seek a suitable transformation of the data to make the errors serially independent (Chib and Greenberg, 1994). Here we take a different route and build upon recent advances in precision-based algorithms for state space models, which have been shown to be computationally more efficient than Kalman filter based methods (Chan and Jeliazkov, 2009; McCausland et al., 2011). The key to our efficiency gain is that despite having a full covariance matrix implied by the ARMA structure, we can work with only sparse matrices, which substantially speed up the computations. In this way, we are able to overcome the computational challenge of having a full covariance matrix and maintain the same type of advantages obtained by Chan (2013).

A third contribution of our paper is that we provide a substantive forecasting exercise involving two commonly used inflation measures: CPI and the GDP Deflator, across the G7 economies. Since inflation plays a key role in modern monetary policy, any forecast improvement over the standard set of benchmark models will have substantive practical significance. Our primary result is that the proposed ARMA-SV models generally provide superior out-of-sample point and density forecasts, and are especially useful for Canada, France, Italy and the US. From an empirical perspective, given that our study includes state-of-the-art models such as UC-SV and UC-MA-SV, our analysis can be viewed as an extension of the results presented in Stock and Watson (2007) and Chan (2013).

The rest of the paper is structured as follows. In Section 2 we present the general ARMA-SV framework, discuss how this embeds a variety of popular model specifications, and develop an efficient posterior simulator to estimate this new class of models. In Section 3, we discuss our application of forecasting CPI and GDP Deflator inflation measures in each of the G7 countries. We then present the results in Section 4 and conclude in Section 5.

## 2 Stochastic Volatility Models with ARMA Errors

Let  $y_t$  denote a variable of interest at date  $t$ , where  $t = 1, \dots, T$ . Then, the general state space representation of the ARMA( $p, q$ )-SV framework is given by:

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

$$\varepsilon_t^y = \phi_1 \varepsilon_{t-1}^y + \dots + \phi_p \varepsilon_{t-p}^y + u_t + \psi_1 u_{t-1} + \dots + \psi_q u_{t-q}, \quad u_t \sim \mathcal{N}(0, e^{h_t}), \quad (2)$$

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

where the error terms  $\varepsilon_t^\tau$  and  $u_t$  are independent across all leads and lags. Note that setting  $\phi_i = \psi_j = 0$  for all  $i = 1, \dots, p$  and  $j = 1, \dots, q$  returns the standard stochastic volatility framework. Similarly, setting  $h_1 = \dots = h_T = c$ , where  $c$  is some constant, returns the standard ARMA framework.

Equation (1) is referred to as either the measurement or observation equation. It is composed of a time-varying conditional mean  $\mu_t$  and an error terms  $\varepsilon^y$ . By choosing a suitable equation for the conditional mean, it is easily seen that this framework embeds a variety of popular model specifications. Two such examples are:

1. The autoregressive model:

$$\mu_t = \rho_0 + \sum_{i=1}^m \rho_i y_{t-i} \quad (4)$$

2. The unobserved components model:

$$\tau_t = \tau_{t-1} + \varepsilon_t^\tau, \quad \varepsilon_t^\tau \sim \mathcal{N}(0, \sigma_\tau^2), \quad (5)$$

Of course, many other frameworks could be considered, such as the standard linear regression model with constant or time-varying coefficients.

Equations (2) and (3) then denote the state equations for the ARMA( $p, q$ ) error structure in which the variance follows a SV process. Notice that, in contrast to GARCH models, the SV parameter  $h_t$  enters this specification as an instantaneous volatility of the model. Moreover, (2) can be rewritten in terms of a polynomial:

$$\phi(L)\varepsilon_t = \psi(L)u_t,$$

where  $L$  is the lag operator,  $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ , and  $\psi(L) = 1 + \psi_1 L + \dots + \psi_q L^q$ . As is standard in the ARMA literature, herein we assume that all roots of  $\phi(L)$  stay outside the unit circle for stationarity, and that all roots of  $\psi(L)$  fall outside the unit circle for invertibility (Chib and Greenberg, 1994).

## 2.1 Estimation

We estimate the model with Bayesian methods and simulate the joint distributions of interest through an efficient Metropolis-within-Gibbs sampler that builds upon recent developments in precision-based algorithms. The main challenge is that due to the ARMA error structure, the covariance matrix of the joint distribution for  $\mathbf{y} = (y_1, \dots, y_T)'$  is a full matrix. Thus, in order to use the conventional Kalman filter, the original data need to be transformed so that the transformed errors are independent. Here, however, we build on results in Chan (2013) and employ a direct approach using precision-based algorithms. The efficiency of our algorithm stems from the fact that despite the covariance matrix being full, we can work with only banded matrices—i.e., matrices that are sparse and the few non-zero elements are arranged along the main diagonal. We will discuss this in more detail in the following subsections.

### 2.1.1 Likelihood Evaluation

We first investigate the likelihood function of our model and present a fast and simple way to evaluate it for both Bayesian and maximum likelihood estimation. Since the likelihood function is the joint distribution of the data, we seek so stack the system of equations in (1)–(3) over  $t = 1, \dots, T$ . To this end, note that (2) can be written as:

$$\mathbf{H}_\phi \boldsymbol{\varepsilon}^y = \mathbf{H}_\psi \mathbf{u}, \quad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_\mathbf{u}), \quad (6)$$

where  $\boldsymbol{\varepsilon}^y = (\varepsilon_1^y, \dots, \varepsilon_T^y)'$ ,  $\mathbf{u} = (u_1, \dots, u_T)'$ ,  $\boldsymbol{\Omega}_\mathbf{u} = \text{diag}(e^{h_1}, \dots, e^{h_T})$ ,  $\mathbf{H}_\phi$  is a  $T \times T$  difference matrix and  $\mathbf{H}_\psi$  is a  $T \times T$  lower triangular matrix with ones along the main diagonal and  $\psi_j$  on the  $j$ -th lower diagonal,  $j = 1, \dots, q$ . For example, if we have an

ARMA(2,2) error structure then  $\mathbf{H}_\phi$  and  $\mathbf{H}_\psi$  are defined to be:

$$\mathbf{H}_\phi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -\phi_1 & 1 & 0 & 0 & \cdots & 0 \\ -\phi_2 & -\phi_1 & 1 & 0 & \cdots & 0 \\ 0 & -\phi_2 & -\phi_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\phi_2 & -\phi_1 & 1 \end{pmatrix}, \quad \mathbf{H}_\psi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \psi_1 & 1 & 0 & 0 & \cdots & 0 \\ \psi_2 & \psi_1 & 1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \psi_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_2 & \psi_1 & 1 \end{pmatrix}.$$

Since  $\mathbf{H}_\phi$  is a lower triangular matrix with ones along the main diagonal,  $|\mathbf{H}_\phi| = 1$  for any  $\phi = (\phi_1, \dots, \phi_p)'$ . Thus,  $\mathbf{H}_\phi$  is invertible and (6) can be written as:

$$\boldsymbol{\varepsilon}^y = \mathbf{H}_\phi^{-1} \mathbf{H}_\psi \mathbf{u}. \quad (7)$$

Finally, stacking (1) over all dates and substituting (7) gives:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{H}_\phi^{-1} \mathbf{H}_\psi \mathbf{u}, \quad (8)$$

where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_T)'$ . By a change of variable, it follows that:

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

where  $\boldsymbol{\psi} = (\psi_1, \dots, \psi_p)'$ ,  $\mathbf{h} = (h_1, \dots, h_T)'$  and  $\boldsymbol{\Omega}_y = \mathbf{H}_\phi^{-1} \mathbf{H}_\psi \boldsymbol{\Omega}_u (\mathbf{H}_\phi^{-1} \mathbf{H}_\psi)'$ . Thus, the log-likelihood function is given by:

$$\log p(\mathbf{y} \mid \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\mu}, \mathbf{h}) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T h_t - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Omega}_y^{-1} (\mathbf{y} - \boldsymbol{\mu}). \quad (9)$$

Evaluation of the log-likelihood function requires the computation of the  $T \times T$  inverse matrix  $\boldsymbol{\Omega}_y$ . In general, this is an intensive procedure, requiring  $\mathcal{O}(T^3)$  operations. The key to overcoming this computational hurdle is to notice that  $\mathbf{H}_\phi^{-1} \mathbf{H}_\psi = \mathbf{H}_\psi \mathbf{H}_\phi^{-1}$ , a claim that is proven in Appendix. Given this fact, (8) can be written as:

$$\tilde{\mathbf{y}} = \tilde{\boldsymbol{\mu}} + \mathbf{H}_\psi^{-1} \mathbf{u}, \quad (10)$$

where  $\tilde{\mathbf{y}} = \mathbf{H}_\psi^{-1} \mathbf{y}$  and  $\tilde{\boldsymbol{\mu}} = \mathbf{H}_\psi^{-1} \boldsymbol{\mu}$ . Thus, by a change of variable:

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



where  $\mathbf{S}_{\tilde{\mathbf{y}}}^{-1} = \mathbf{H}'_{\phi} \boldsymbol{\Omega}_u^{-1} \mathbf{H}_{\phi}$  and the log-likelihood for  $\tilde{\mathbf{y}}$  is:

$$\log p(\tilde{\mathbf{y}} | \tilde{\boldsymbol{\mu}}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\psi}) \propto -\frac{1}{2} \sum_{t=1}^T h_t - \frac{1}{2} (\tilde{\mathbf{y}} - \tilde{\boldsymbol{\mu}})' \mathbf{S}_{\tilde{\mathbf{y}}}^{-1} (\tilde{\mathbf{y}} - \tilde{\boldsymbol{\mu}}). \quad (11)$$

The fact that  $\mathbf{S}_{\tilde{\mathbf{y}}}$  is invertible stems from noting that  $|\mathbf{H}_{\phi}| = 1$  for any vector  $\boldsymbol{\phi}$  and that  $|\boldsymbol{\Omega}_u| = e^{\sum_{t=1}^T h_t} > 0$  for any  $h_1, \dots, h_T$ . Moreover, since  $\boldsymbol{\Omega}_u$  is a diagonal matrix, its inverse is simply given by taking the reciprocal of the diagonal elements, i.e.,  $\boldsymbol{\Omega}_u^{-1} = \text{diag}(e^{-h_1}, \dots, e^{-h_T})$ . Thus, compared with (9), (11) can be calculated much faster. More precisely, the Cholesky decomposition of a  $T \times T$  band matrix takes  $\mathcal{O}(T)$  operations, which is substantially less than the  $\mathcal{O}(T^3)$  operations required for the same operation on a full matrix of the same size.<sup>2</sup>

Given these facts, we can employ a simple 3-step procedure to evaluate (11) (see also the discussion in Chan and Jeliazkov, 2009). First, we obtain the Cholesky decomposition  $\mathbf{C}_{\tilde{\mathbf{y}}}$  of the band matrix  $\mathbf{S}_{\tilde{\mathbf{y}}}$ , which involves  $\mathcal{O}(T)$  operations. Second, we implement forward substitution and backward substitution to get:

$$\mathbf{A} = \mathbf{C}'_y \setminus (\mathbf{C}_y \setminus (\mathbf{y} - \boldsymbol{\mu})),$$

which, by definition, is equivalent to  $\mathbf{A} = \mathbf{C}_y^{-1'} (\mathbf{C}_y^{-1} (\mathbf{y} - \boldsymbol{\mu})) = \mathbf{S}_{\tilde{\mathbf{y}}}^{-1} (\mathbf{y} - \boldsymbol{\mu})$ . Finally, we compute:

$$\mathbf{B} = -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \mathbf{A} = -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \mathbf{S}_{\tilde{\mathbf{y}}}^{-1} (\mathbf{y} - \boldsymbol{\mu})$$

Thus, conditional on  $\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\psi}$  and  $\mathbf{h}$ , the log-likelihood function (11) can be efficiently evaluated without implementing the Kalman filter.

### 2.1.2 Posterior Analysis and Simulation

After discussing an efficient way to evaluate the likelihood function, we now outline an efficient posterior sampler for estimating the ARMA-SV model presented in (1)–(3). In our empirical application we will consider the forecast performance of various nested specifications of both the UC-ARMA-SV and AR-ARMA-SV models. In this section we will therefore restrict our focus to the UC-ARMA-SV model, noting that the estimation for all alternative models considered in this paper can be obtained in a similar manner.

The UC-ARMA-SV model is given by using (5) in place of (1). The transition equa-

<sup>2</sup>For a textbook treatment see Chapter 4 in Golub and Van Loan (2013).

tions for the time-varying trend and log-volatilities are respectively initialized with  $\tau_1 \sim \mathcal{N}(\tau_0, \sigma_{0\tau}^2)$  and  $h_1 \sim \mathcal{N}(h_0, \sigma_{0h}^2)$  where  $\tau_0, \sigma_{0\tau}^2, h_0,$  and  $\sigma_{0h}^2$  are known constants.

Let  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_T)$ . The priors for  $\boldsymbol{\tau}, \boldsymbol{\phi}, \boldsymbol{\psi}, \sigma_\tau^2$  and  $\sigma_h^2$  are assumed to be independent. In particular, we set:

$$\begin{aligned} \sigma_\tau^2 &\sim \mathcal{IG}(\nu_\tau, S_\tau), & \sigma_h^2 &\sim \mathcal{IG}(\nu_h, S_h), \\ \boldsymbol{\phi} &\sim \mathcal{N}(\boldsymbol{\phi}_0, \mathbf{V}_\phi) \mathbf{1}(\boldsymbol{\phi} \in \mathbf{A}_\phi), & \boldsymbol{\psi} &\sim \mathcal{N}(\boldsymbol{\psi}_0, \mathbf{V}_\psi) \mathbf{1}(\boldsymbol{\psi} \in \mathbf{A}_\psi), \end{aligned}$$

where  $\mathcal{IG}$  denotes the inverse-gamma distribution, while  $\mathbf{1}(\boldsymbol{\phi} \in \mathbf{A}_\phi)$  and  $\mathbf{1}(\boldsymbol{\psi} \in \mathbf{A}_\psi)$  are indicator functions in which  $\mathbf{A}_\phi$  and  $\mathbf{A}_\psi$  are sets that satisfy the stationarity restriction to the roots of the respective autoregressive polynomials.

Posterior draws are then obtained by sequentially sampling from:

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

In what follows we discuss an efficient way to sample from each of these conditional distributions.

### Step 1: Sampling $\boldsymbol{\tau}$

To efficiently sample from  $p(\boldsymbol{\tau} \mid \mathbf{y}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\psi}, \sigma_\tau^2)$ , first note that the log posterior density for  $\tilde{\boldsymbol{\tau}}$  is:

$$\log p(\tilde{\boldsymbol{\tau}} \mid \tilde{\mathbf{y}}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\psi}, \sigma_\tau^2) \propto \log p(\tilde{\boldsymbol{\tau}} \mid \sigma_\tau^2) + \log p(\tilde{\mathbf{y}} \mid \tilde{\boldsymbol{\tau}}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\psi}), \quad (12)$$

where  $p(\tilde{\boldsymbol{\tau}} \mid \sigma_\tau^2)$  is the prior for  $\tilde{\boldsymbol{\tau}}$  and  $p(\tilde{\mathbf{y}} \mid \tilde{\boldsymbol{\tau}}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\psi})$  is the likelihood for  $\tilde{\mathbf{y}}$  which is obtained by setting  $\boldsymbol{\tau} = \boldsymbol{\mu}$  in (10).

The prior density is obtained by noting that (5) can be stacked over all dates  $t = 1, \dots, T$  to give:

$$\mathbf{H}\boldsymbol{\tau} = \boldsymbol{\varepsilon}_\tau, \quad \boldsymbol{\varepsilon}_\tau \sim \mathcal{N}(0, \boldsymbol{\Omega}_{\boldsymbol{\varepsilon}_\tau}), \quad (13)$$

where  $\boldsymbol{\Omega}_{\boldsymbol{\varepsilon}_\tau} = \text{diag}(\sigma_{0\tau}^2, \sigma_\tau^2, \dots, \sigma_\tau^2)$  and  $\mathbf{H}$  is a first-difference matrix of size  $T$ . Since  $|\mathbf{H}| = 1$ , (13) can be written as:

$$\boldsymbol{\tau} = \mathbf{H}^{-1}\boldsymbol{\varepsilon}_\tau. \quad (14)$$

Thus  $\boldsymbol{\tau} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_\tau)$ , where  $\boldsymbol{\Omega}_\tau^{-1} = \mathbf{H}'\boldsymbol{\Omega}_{\varepsilon_\tau}^{-1}\mathbf{H}$ . Pre-multiplying both sides of (14) by  $\mathbf{H}_\psi^{-1}$  gives

$$\tilde{\boldsymbol{\tau}} = \mathbf{H}_\psi^{-1}\mathbf{H}^{-1}\boldsymbol{\varepsilon}_\tau. \quad (15)$$

so that  $\tilde{\boldsymbol{\tau}} \sim \mathcal{N}(0, \boldsymbol{\Omega}_{\tilde{\boldsymbol{\tau}}})$  and  $\boldsymbol{\Omega}_{\tilde{\boldsymbol{\tau}}}^{-1} = \mathbf{H}'_\psi\boldsymbol{\Omega}_\tau^{-1}\mathbf{H}_\psi$ . The log prior density for  $\tilde{\boldsymbol{\tau}}$  is therefore given by:

$$\log p(\tilde{\boldsymbol{\tau}} | \sigma_\tau^2) \propto -\frac{T-1}{2} \log \sigma_\tau^2 - \frac{1}{2} \tilde{\boldsymbol{\tau}}' \mathbf{H}'_\psi \boldsymbol{\Omega}_\tau^{-1} \mathbf{H}_\psi \tilde{\boldsymbol{\tau}}, \quad (16)$$

Finally, substituting (11) and (16) into (12), gives:

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

where  $\mathbf{D}_{\tilde{\boldsymbol{\tau}}} = (\mathbf{H}'_\psi \boldsymbol{\Omega}_\tau^{-1} \mathbf{H}_\psi + \mathbf{H}'_\phi \boldsymbol{\Omega}_u^{-1} \mathbf{H}_\phi)^{-1}$  and  $\hat{\boldsymbol{\tau}} = \mathbf{D}_{\tilde{\boldsymbol{\tau}}} \mathbf{H}'_\phi \boldsymbol{\Omega}_u^{-1} \mathbf{H}_\phi \tilde{\mathbf{y}}$ . The conditional posterior distribution is therefore:

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

Sampling from this distribution can be efficiently conducted via the precision-based algorithm discussed in the previous section. In particular, since  $\mathbf{D}_{\tilde{\boldsymbol{\tau}}}$  is a band matrix, we first use the Cholesky decomposition  $\mathbf{C}_{\tilde{\boldsymbol{\tau}}}$  of the precision matrix  $\mathbf{D}_{\tilde{\boldsymbol{\tau}}}^{-1}$  and then compute  $\hat{\boldsymbol{\tau}}$ . Next, forward and backward substitution give:

$$\hat{\boldsymbol{\tau}} = \mathbf{C}'_{\tilde{\boldsymbol{\tau}}} \setminus (\mathbf{C}_{\tilde{\boldsymbol{\tau}}} \setminus (\mathbf{H}'_\phi \boldsymbol{\Omega}_u^{-1} \mathbf{H}_\phi \tilde{\mathbf{y}})),$$

so that the draws of  $\tilde{\boldsymbol{\tau}}$  can be obtained by:

$$\tilde{\boldsymbol{\tau}} = \hat{\boldsymbol{\tau}} + \mathbf{C}'_{\tilde{\boldsymbol{\tau}}} \setminus \mathbf{R}, \quad \mathbf{R} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Finally, we obtain the draw of  $\boldsymbol{\tau}$  by the transformation  $\boldsymbol{\tau} = \mathbf{H}_\psi \tilde{\boldsymbol{\tau}}$ .

## Step 2: Sampling $\mathbf{h}$

To sample from  $p(\mathbf{h} | \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \boldsymbol{\psi}, \sigma_\tau^2, \sigma_h^2)$ , note that (8) can be written as:

$$\mathbf{y}^* = \mathbf{u}, \quad (17)$$

where  $\mathbf{y}^* = \mathbf{H}_\psi^{-1} \mathbf{H}_\phi (\mathbf{y} - \boldsymbol{\tau})$ . Thus,  $\mathbf{y}^* \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_u)$ . With this transformation the aux-

iliary mixture sampler proposed by Kim et al. (1998) can be directly applied. The only difference here is that we replace their forward-backward smoothing algorithm with the precision-based sampler discussed in the previous section.

### Step 3: Sampling $\sigma_h^2$ and $\sigma_\tau^2$

Since the inverse-gamma distribution is a conjugate prior for the multivariate normal, draws from the conditional distribution for  $\sigma_h^2$  and  $\sigma_\tau^2$  can be obtained by directly sampling from the known distribution. For instance, the key to sampling  $\sigma_\tau^2$  is to note that, by a change of variable in (5),  $\tau_t \sim \mathcal{N}(\tau_{t-1}, \sigma_\tau^2)$ . Thus, given the inverse-gamma prior for  $\sigma_\tau^2$ , we have:

$$\begin{aligned} p(\sigma_\tau^2 | \boldsymbol{\tau}) &\propto p(\boldsymbol{\tau} | \sigma_\tau^2) + p(\sigma_\tau^2), \\ &= (\sigma_\tau^2)^{-\frac{T}{2}} \exp\left(-\frac{1}{2\sigma_\tau^2} \sum_{t=2}^T (\tau_t - \tau_{t-1})^2\right) \cdot (\sigma_\tau^2)^{-(\nu_0-1)} \exp\left(-\frac{S_\tau}{\sigma_\tau^2}\right), \\ &\propto (\sigma_\tau^2)^{-((\frac{T}{2}+\nu_0)-1)} \exp\left(-\frac{1}{\sigma_\tau^2} \left(\sum_{t=2}^T (\tau_t - \tau_{t-1})^2/2 + S_\tau\right)\right). \end{aligned}$$

Hence:

$$(\sigma_\tau^2 | \boldsymbol{\tau}) \sim \mathcal{IG}\left(T/2 + \nu_\tau, \sum_{t=2}^T (\boldsymbol{\tau}_t - \boldsymbol{\tau}_{t-1})^2/2 + S_\tau\right).$$

Similarly, given that (3) implies that  $h_t \sim \mathcal{N}(h_{t-1}, \sigma_h^2)$ , the posterior density of  $\sigma_h^2$  is given by:

$$(\sigma_h^2 | \mathbf{h}) \sim \mathcal{IG}\left(T/2 + \nu_h, \sum_{t=2}^T (\mathbf{h}_t - \mathbf{h}_{t-1})^2/2 + S_h\right).$$

### Step 4: Sampling $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$

Unlike  $\mathbf{y}$ ,  $\boldsymbol{\tau}$ ,  $\sigma_h^2$  and  $\sigma_\tau^2$ , which all possess conditional posterior distributions from the exponential family, the distributions of  $\boldsymbol{\psi}$  and  $\boldsymbol{\phi}$  are unknown. We therefore seek suitable candidate densities to implement and independence-chain Metropolis-Hastings algorithm Kroese et al. (2011). Since these densities turn out to be different, we now consider each in turn.

First, to sample the moving average term  $\boldsymbol{\psi}$ , we note that stacking (1) and (2) over all dates gives:

$$\mathbf{y}^{**} = \mathbf{H}_\psi \mathbf{u}, \tag{18}$$

where  $\mathbf{y}^{**} = \mathbf{H}_\phi(\mathbf{y} - \boldsymbol{\tau})$ . By a change of variable,  $(\mathbf{y}^{**} | \boldsymbol{\psi}, \boldsymbol{\tau}, \mathbf{h}) \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_\psi \boldsymbol{\Omega}_u \mathbf{H}_\psi')$ . Thus, given the normal prior, the log-likelihood of the conditional posterior is:

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

Since the dimension of  $\boldsymbol{\psi}$  is typically low, it's feasible to use numerical optimization to obtain the mode and negative Hessian of  $\log p(\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h})$  evaluated at the mode, denoted  $\widehat{\boldsymbol{\psi}}$  and  $\mathbf{H}_\psi$  respectively. These values can then be used as the mean and variance in the multivariate normal density  $q(\boldsymbol{\psi})$ . Given this density, we obtain candidate draws  $\boldsymbol{\psi}^c$  and update in accordance with the acceptance probability:

$$\min\left\{1, \frac{p(\boldsymbol{\psi}^c | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h})}{p(\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h})} \cdot \frac{q(\boldsymbol{\psi})}{q(\boldsymbol{\psi}^c)}\right\}.$$

Second, to sample  $\phi$ , first note that (18) can be written as:

$$\mathbf{H}_\phi \mathbf{z} = \mathbf{H}_\psi \mathbf{u}, \quad (19)$$

where  $\mathbf{z} = \mathbf{y} - \boldsymbol{\tau}$ . Since  $\mathbf{H}_\psi$  is invertible, this equation can be expressed as:

$$\mathbf{z} = \mathbf{X}_z \boldsymbol{\phi} + \mathbf{H}_\psi \mathbf{u}, \quad (20)$$

where  $\mathbf{X}_z = (\mathbf{z}_1, \dots, \mathbf{z}_{T-1})'$ . So  $\boldsymbol{\phi}$  is equivalent to the coefficient of a standard linear regression model. Thus, given the truncated normal prior, the posterior density of  $\boldsymbol{\phi}$  is:

$$(\boldsymbol{\phi} | \mathbf{y}, \boldsymbol{\tau}, \mathbf{h}) \sim \mathcal{N}(\widehat{\boldsymbol{\phi}}, \mathbf{D}_\phi) \mathbb{1}(\boldsymbol{\phi} \in \mathbf{A}_\phi),$$

where  $\mathbf{D}_\phi^{-1} = \mathbf{V}_\phi^{-1} + \mathbf{X}_z' (\mathbf{H}_\psi \boldsymbol{\Omega}_u \mathbf{H}_\psi')^{-1} \mathbf{X}_z$  and  $\widehat{\boldsymbol{\phi}} = \mathbf{D}_\phi \cdot (\mathbf{V}_\phi^{-1} \boldsymbol{\phi}_0 + \mathbf{X}_z (\mathbf{H}_\psi \boldsymbol{\Omega}_u \mathbf{H}_\psi')^{-1} \mathbf{z})$ . Sampling from this distribution can then be conducted with a standard acceptance-rejection algorithm (see, e.g. Koop (2003) or Kroese et al. (2011)).

### 3 Application to Inflation Forecasting

In this section we assess the proposed ARMA-SV frameworks ability to forecast two commonly used inflation measures: CPI and the GDP Deflator, in each of the G7 countries.

More specifically, we present both out-of-sample point and density forecast performance to various benchmarks. Given the quarterly index, denoted  $z_t$ , we use  $y_t = 100 \log(z_t/z_{t-1})$  as the associated inflation measure. The resulting series are plotted in Figures 1 and 2.

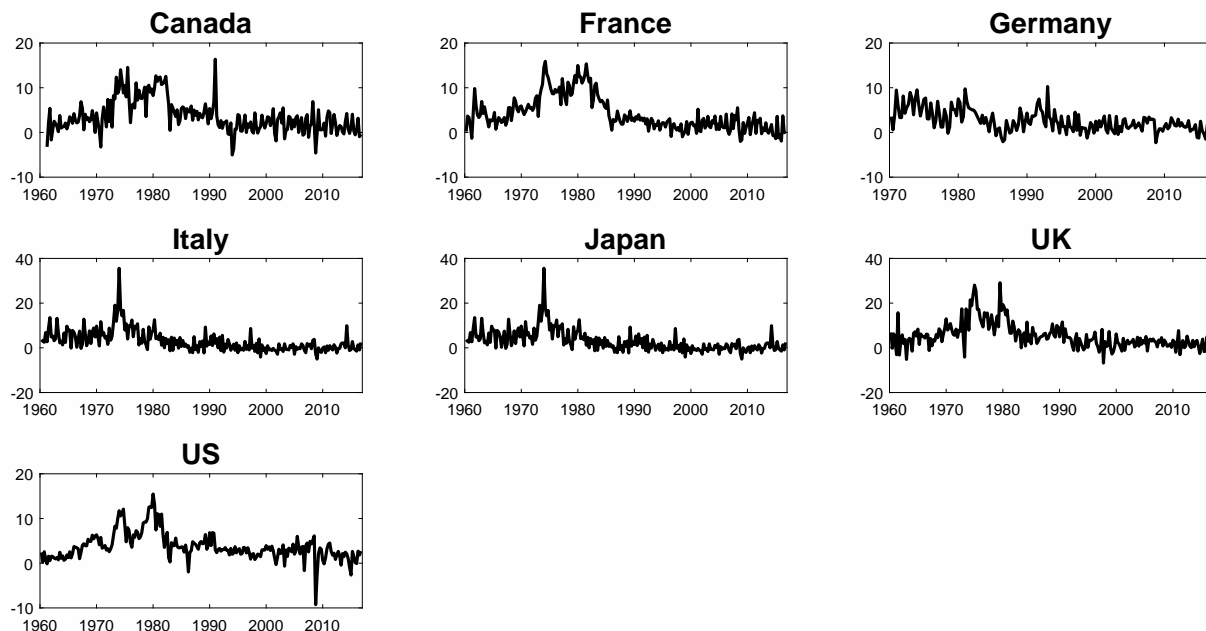


Figure 1: CPI inflation across G7 countries.

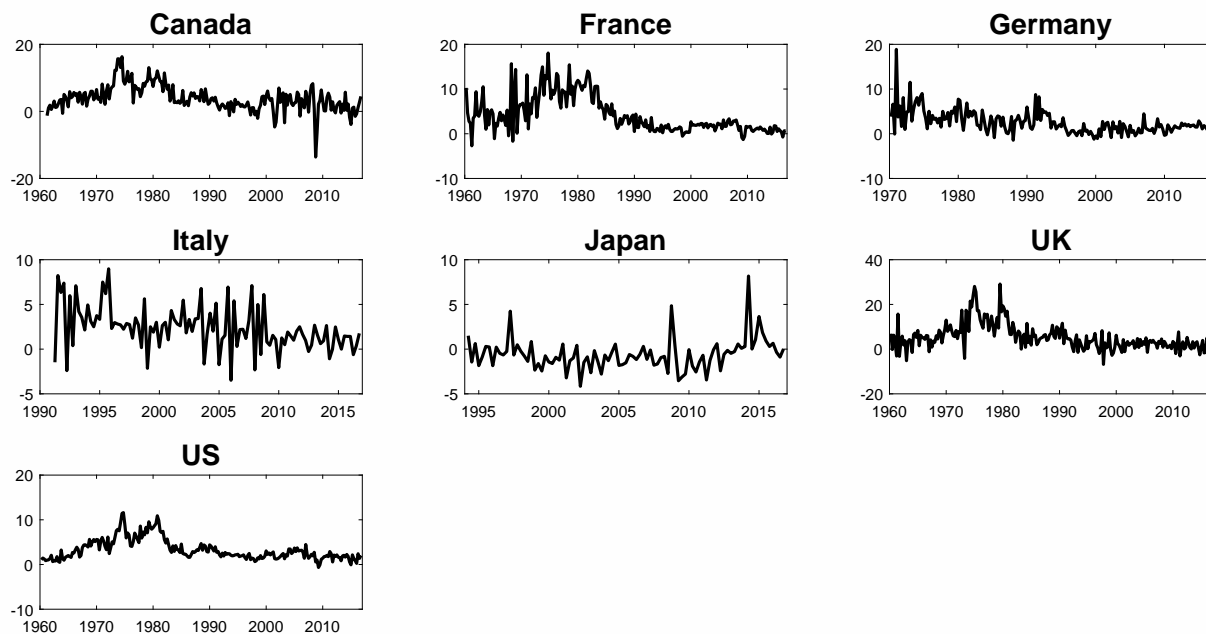


Figure 2: GDP Deflator inflation across G7 countries.

While each series ends in 2016Q4, due to different starting dates, there are some differ-

ences in sample size. For instance, while most CPI series start in 1960Q2, Canada’s is only available from 1961Q1 and Germany’s from 1970Q2.

### 3.1 Competing Models

The primary benchmark model is taken to be a stationary  $AR(m)$  model—as in (4)—with homoscedastic variance. The reason for using this benchmark is two-fold. First, it is still a competitive model among both univariate and multivariate models. Second, given the parsimonious structure of the model, any finding that it provides competitive forecasts has practical significance.

In addition to the homoscedastic  $AR(m)$  model, we also consider a more general class of heteroscedastic autoregressive (AR) and unobserved components (UC) models. These include specifications with and without SV, MA-SV and ARMA-SV errors. In each country, the lag length is selected via the Bayesian information criteria (BIC). The results from these tests are summarized in Table 1.

Table 1: Lag length selection for  $AR(m)$  models using BIC.

Country	CPI Inflation	GDP Deflator Inflation
Canada	4	5
France	5	6
Germany	4	4
Italy	4	1
Japan	8	1
UK	5	3
US	3	2

In each case we limit the ARMA-SV component to the observation equation. While the extension to ARMA-SV errors in the state equation is conceptually straight forward to implement, this specification allows for a more direct comparison with the broader literature (e.g., Stock and Watson (2007), Chan (2013) and Chan et al. (2013)). In total, this constitutes 10 models, each of which is summarized in Table 2.

Table 2: A list of competing models.

Model	Description
AR( $m$ )	Autoregressive model with homoscedastic errors
AR( $m$ )-SV	Autoregressive model with SV errors
AR( $m$ )-MA-SV	Autoregressive model with MA-SV errors
AR( $m$ )-ARMA-SV	Autoregressive model with ARMA-SV errors
AR( $m$ )-ARMA	Autoregressive model with ARMA errors
UC	Unobserved components model with homoscedastic errors
UC-SV	Unobserved components model with SV errors
UC-MA-SV	Unobserved components model with MA-SV errors
UC-ARMA-SV	Unobserved components model with ARMA-SV errors
UC-ARMA	Unobserved components model with ARMA errors

### 3.2 Priors and Initial Conditions

In each of the UC models we set the initial value of  $\tau_t$  as  $\tau_1 \sim \mathcal{N}(\tau_0, \sigma_{0\tau}^2)$ , where  $\tau_0, h_0, \sigma_{0\tau}^2$ . In particular, we set  $\tau_0 = 0$  and  $\sigma_{0\tau}^2 = 5$ . Similarly, we initialize UC models with SV with  $h_1 \sim \mathcal{N}(h_0, \sigma_{0h}^2)$ , where  $h_0 = 0$  and  $\sigma_{0h}^2 = 5$ .

For the AR models, we set an independent truncated prior for the conditional mean coefficients. In particular, the prior mean is the zero vector and the variance is  $\mathbf{V}_\rho = 5 \times \mathbf{I}_n$ , where  $\mathbf{I}_n$  denotes the identity matrix of size  $n$ . The posterior distribution of this model is then obtained by following the procedure in Section 2.1.2, where Step 1 is replaced with draws from  $p(\boldsymbol{\rho} | \mathbf{y}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\psi}, \sigma_\tau^2, \sigma_h^2)$ , with  $\boldsymbol{\rho} = (\rho_1, \dots, \rho_m)'$ . To this end, note that by following similar steps in Section 2.1.2, the conditional posterior distribution is:

$$(\boldsymbol{\rho} | \mathbf{y}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\psi}, \sigma_\tau^2, \sigma_h^2) \sim \mathcal{N}(\hat{\boldsymbol{\rho}}, \mathbf{D}_\rho) \mathbf{1}(\boldsymbol{\rho} \in \mathbf{A}_\rho), \quad (21)$$

where  $\mathbf{D}_\rho = (\mathbf{V}_\rho + \mathbf{H}'_\phi \boldsymbol{\Omega}_u^{-1} \mathbf{H}_\phi)^{-1}$ ,  $\hat{\boldsymbol{\rho}} = \mathbf{D}_\rho \mathbf{H}'_\phi \boldsymbol{\Omega}_u^{-1} \mathbf{H}_\phi \tilde{\mathbf{y}}$ , and the truncated region ensures all roots lay outside of the unit circle. Samples from this truncated density are then obtained using the acceptance-rejection method discussed in Section 2.1.2 using the proposal density  $\mathcal{N}(\hat{\boldsymbol{\rho}}, \mathbf{D}_\rho)$ .

Finally, following Chan (2013) we set the moving average order in the MA-SV model variants to be one. For consistency, we also set each of the specifications with ARMA( $p, q$ ) errors to be ARMA(1,1).



### 3.3 Forecasting

We conduct a pseudo out-of-sample forecasting exercise in which we consider both point and density forecasts. In each exercise, we divide the data into three sub-samples. The first part contains the first  $m$  observations used to initialize the AR( $m$ ) models. This guarantees that all AR and UC model variants have the same initial observation. The second part is the *estimation period*, which consists of the next 40 observations. The third part is the *hold-out period*, which contains the remaining observations that are used to assess the forecast performance of the model.

To see how the forecasts are conducted, let  $\mathbf{y}_{1:t}$  denote the data from the *estimation period* and  $\hat{\mathbf{y}}_{t+k}$  represent the vector of  $k$ -steps-ahead forecasts with  $k=1,4,8,12$  and 16. Density forecasts are obtained by the predictive density:  $f(\mathbf{y}_{t+k}|\mathbf{y}_{1:t})$ , and point forecasts are taken to be the mean of this density:  $\hat{\mathbf{y}}_{t+k} = \mathbb{E}[\mathbf{y}_{t+k}|\mathbf{y}_{1:t}]$ . These forecasts are conducted with predictive simulation. For concreteness, suppose we want to produce a 4-step ahead forecast of US CPI inflation from 1975Q1 to 1976Q1. Then, given the MCMC draws up to 1975Q1 along with the relevant transition equations, we simulate the future states up to time 1975Q4. The conditional expectation of this equation is then taken to be the point forecast and the observed data is used to evaluate implied density to produce a density forecast. The exercise is then repeated using data from 1975Q2 up to the end of the hold-out period, i.e. 2012Q4.

In each period, the parameter estimates are based on 45,000 draws from the posterior simulator discussed in 2.1, after discarding the first 5,000 draws as a burn-in period.

#### 3.3.1 Forecast Metrics

The accuracy of the point and density forecasts are respectively assessed with the mean square forecast error (MSFE) and log-predictive-likelihood (LPL).

To compute the MSFE, we first evaluate the forecast  $\hat{y}_{T_0+t+k-1}$  by averaging all the posterior means  $\mathbb{E}(y_{T_0+t+k-1} | \mathbf{y}_{1:T_0+t})$  when it is time  $T_0 + t$ . Given that the forecasting error is  $\mathbf{e}_{T_0+t+k-1}^2 = \mathbf{y}_{T_0+t+k-1}^0 - \mathbb{E}(y_{T_0+t+k-1} | \mathbf{y}_{1:T_0+t})$ , where  $k$  denotes a  $k$ -step-ahead forecast, the MSFE is defined as:

$$\text{MSFE} = \frac{1}{T - T_0 - k + 1} \sum_{t=1}^{T-T_0-k+1} \mathbf{e}_{T_0+t+k-1}^2$$

Since a smaller forecast error corresponds to a smaller MSFE, a relatively smaller MSFE indicates better forecast performance. As mentioned in the previous section, we use a stationary AR( $m$ ) model as the benchmark. To facilitate our discussion when presenting the results we therefore standardize the MSFE of each model to the MSFE of the AR( $m$ ) model. Hence the relative MSFE (RMSFE) for the AR( $m$ ) model is 1.00. Moreover, if a given model produces a RMSFE less than one then this indicates superior forecast performance to the AR( $m$ ), and vice-versa.

To compute the predictive likelihood we simply sum the evaluated likelihood function implied by the forecasts. That is:

$$\text{LPL} = \sum_{t=1}^{T-T_0-k+1} \log p(\hat{y}_{T_0+t+k-1} = y_{T_0+t+k-1} | \mathbf{y}_{1:T_0+t})$$

Since a larger likelihood implies better fit, for interpretation purposes, a larger predictive likelihood value implies a better density forecast performance. To facilitate our discussion we also standardize the LPL to the AR( $m$ ) benchmark. Since it is in logs, the relative LPL (RLPL) for the AR( $m$ ) model is 0.00. Moreover, if a given model produces a positive LPL then this indicates superior forecast performance to the AR( $m$ ), and vice-versa.

## 4 Forecast Results

To facilitate the discussion, we separately present the point and density forecast results and defer the tables containing the results to Appendix B.

### 4.1 Point Forecast Results

The point forecast results for CPI are presented in Tables 3–9. While there is no single best model across all countries, there are some general results. For instance, in line with previous studies (e.g., Stock and Watson, 2007; Chan, 2013; Clark and Ravazzolo, 2015), the AR-SV and UC-SV models tend to dominate their constant volatility counterparts in CPI forecasts. Interestingly, this results does not extend to the GDP Deflator. In that case, the simple AR model often provides better forecasts than its SV counterpart. One exception is the US, in which the SV variants dominate their homoscedastic counterparts.

Table 3: MSFEs relative to AR benchmark: Canada CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	0.99	0.97	0.95	0.93	0.90
AR-ARMA	0.99	0.96	0.85	0.77	0.70
AR-MA-SV	0.99	0.97	0.95	0.93	0.89
AR-ARMA-SV	<b>0.97</b>	<b>0.95</b>	<b>0.82</b>	<b>0.73</b>	<b>0.67</b>
UC	1.01	0.99	0.95	0.91	0.88
UC-SV	1.05	1.03	0.96	0.89	0.88
UC-ARMA	1.02	0.96	0.88	0.80	0.74
UC-MA-SV	1.01	1.04	0.96	0.88	0.87
UC-ARMA-SV	1.02	1.03	0.90	0.80	0.74

Table 4: MSFEs relative to AR benchmark: France CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	0.95	0.91	0.86	0.83	0.83
AR-ARMA	0.82	0.87	0.89	0.93	0.95
AR-MA-SV	0.83	0.89	0.85	0.82	0.83
AR-ARMA-SV	<b>0.81</b>	0.88	0.90	0.94	0.95
UC	0.84	0.82	0.79	0.81	0.82
UC-SV	0.91	0.93	0.88	0.87	0.85
UC-ARMA	0.85	0.77	0.66	0.63	0.59
UC-MA-SV	0.93	1.02	0.91	0.87	0.84
UC-ARMA-SV	0.84	<b>0.73</b>	<b>0.57</b>	<b>0.52</b>	<b>0.46</b>

Table 5: MSFEs relative to AR benchmark: Germany CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	<b>1.00</b>	1.00	1.00	1.00	1.00
AR-SV	1.01	<b>0.98</b>	<b>0.96</b>	0.92	0.88
AR-ARMA	1.01	1.01	1.00	0.97	0.88
AR-MA-SV	1.01	0.99	<b>0.96</b>	0.92	0.88
AR-ARMA-SV	1.01	1.01	0.98	0.95	0.89
UC	1.08	1.10	1.01	0.95	0.89
UC-SV	1.13	1.14	1.00	0.93	0.85
UC-ARMA	1.14	1.16	1.00	0.92	0.83
UC-MA-SV	1.14	1.15	0.99	<b>0.91</b>	<b>0.83</b>
UC-ARMA-SV	1.14	1.15	0.99	<b>0.91</b>	<b>0.83</b>

Table 6: MSFEs relative to AR benchmark: Italy CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	<b>0.89</b>	0.67	0.63	0.60	0.57
AR-ARMA	0.93	0.65	0.59	0.55	0.49
AR-MA-SV	<b>0.89</b>	0.67	0.63	0.61	0.58
AR-ARMA-SV	<b>0.89</b>	0.63	0.56	0.50	0.43
UC	1.00	0.77	0.79	0.83	0.86
UC-SV	1.11	0.89	0.87	0.86	0.84
UC-ARMA	1.04	0.67	0.70	0.75	0.81
UC-MA-SV	1.08	0.95	0.86	0.81	0.75
UC-ARMA-SV	0.99	<b>0.60</b>	<b>0.51</b>	<b>0.47</b>	<b>0.41</b>

Table 7: MSFEs relative to AR benchmark: Japan CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	1.04	0.83	<b>0.72</b>	0.76	<b>0.73</b>
AR-ARMA	<b>0.70</b>	0.79	0.94	0.95	0.99
AR-MA-SV	1.06	0.82	<b>0.72</b>	0.75	<b>0.73</b>
AR-ARMA-SV	<b>0.70</b>	0.79	0.94	0.95	0.99
UC	0.78	<b>0.76</b>	0.78	<b>0.74</b>	0.75
UC-SV	0.80	0.81	0.83	0.77	0.76
UC-ARMA	0.83	0.79	0.87	0.83	0.84
UC-MA-SV	0.80	0.81	0.82	0.77	0.75
UC-ARMA-SV	0.73	0.80	0.84	0.76	0.74

Table 8: MSFEs relative to AR benchmark: UK CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	0.97	0.95	0.77	0.77	0.76
AR-ARMA	<b>0.75</b>	0.65	0.85	0.89	0.94
AR-MA-SV	0.86	0.91	0.78	0.77	0.76
AR-ARMA-SV	<b>0.75</b>	0.65	0.85	0.89	0.94
UC	0.96	0.74	0.80	0.79	0.81
UC-SV	0.99	0.81	0.81	0.75	0.73
UC-ARMA	0.99	0.68	0.70	<b>0.64</b>	<b>0.63</b>
UC-MA-SV	1.02	0.79	0.78	0.72	0.69
UC-ARMA-SV	0.98	<b>0.63</b>	<b>0.68</b>	<b>0.64</b>	0.64

Table 9: MSFEs relative to AR benchmark: US CPI inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	0.99	0.97	0.86	0.84	0.82
AR-ARMA	1.03	1.01	0.88	0.83	0.75
AR-MA-SV	0.99	0.97	0.86	0.84	0.80
AR-ARMA-SV	1.00	0.97	<b>0.82</b>	<b>0.73</b>	<b>0.63</b>
UC	0.99	0.97	0.88	0.91	0.91
UC-SV	0.96	<b>0.93</b>	0.92	0.97	0.88
UC-ARMA	1.00	0.97	0.83	0.79	0.70
UC-MA-SV	<b>0.94</b>	0.94	0.90	0.92	0.81
UC-ARMA-SV	1.00	1.07	0.86	0.79	0.70

Table 10: MSFEs relative to AR benchmark: Canada GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	1.01	0.97	0.93	0.92	0.91
AR-ARMA	0.95	0.96	0.93	0.83	0.78
AR-MA-SV	0.96	0.97	0.93	0.92	0.91
AR-ARMA-SV	0.96	0.95	<b>0.90</b>	<b>0.78</b>	<b>0.73</b>
UC	0.94	0.98	0.96	0.91	0.89
UC-SV	0.95	<b>0.94</b>	0.93	0.89	0.86
UC-ARMA	0.92	0.97	0.91	0.82	0.78
UC-MA-SV	<b>0.89</b>	0.95	0.93	0.89	0.86
UC-ARMA-SV	0.92	1.15	0.98	0.84	0.81

Table 11: MSFEs relative to AR benchmark: France GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	1.00	0.91	0.89	0.88	0.87
AR-ARMA	0.71	0.94	0.96	0.98	0.99
AR-MA-SV	1.18	0.91	0.88	0.88	0.87
AR-ARMA-SV	<b>0.66</b>	0.82	0.80	0.81	0.83
UC	1.18	0.98	0.92	0.89	0.87
UC-SV	1.18	1.02	0.95	0.90	0.88
UC-ARMA	0.70	0.78	0.71	0.67	0.64
UC-MA-SV	0.70	1.04	0.96	0.91	0.89
UC-ARMA-SV	0.67	<b>0.77</b>	<b>0.66</b>	<b>0.61</b>	<b>0.55</b>

Table 12: MSFEs relative to AR benchmark: Germany GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	1.00	<b>0.98</b>	<b>0.95</b>	<b>0.95</b>	0.94
AR-ARMA	0.98	1.06	1.18	1.30	1.42
AR-MA-SV	0.98	0.99	0.96	<b>0.95</b>	0.95
AR-ARMA-SV	<b>0.96</b>	1.01	1.06	1.13	1.15
UC	<b>0.96</b>	1.02	0.97	0.96	0.95
UC-SV	<b>0.96</b>	1.02	0.97	0.96	0.95
UC-ARMA	1.01	1.04	0.99	0.98	0.96
UC-MA-SV	1.01	1.01	<b>0.95</b>	<b>0.95</b>	0.94
UC-ARMA-SV	1.00	1.01	<b>0.95</b>	<b>0.95</b>	<b>0.93</b>

Table 13: MSFEs relative to AR benchmark: Italy GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	1.00	1.01	1.01	1.01	1.01
AR-ARMA	1.13	0.92	0.97	1.00	1.00
AR-MA-SV	1.13	1.01	1.01	1.01	1.01
AR-ARMA-SV	1.23	0.98	1.00	1.01	1.01
UC	1.23	0.75	0.76	<b>0.75</b>	<b>0.74</b>
UC-SV	1.23	0.78	0.79	0.81	0.81
UC-ARMA	<b>0.96</b>	0.75	0.76	0.76	0.75
UC-MA-SV	1.23	0.76	0.77	0.77	0.78
UC-ARMA-SV	0.97	<b>0.73</b>	<b>0.75</b>	0.76	0.77

Table 14: MSFEs relative to AR benchmark: Japan GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	<b>1.00</b>
AR-SV	1.00	1.01	1.00	1.00	1.00
AR-ARMA	<b>0.93</b>	<b>0.92</b>	<b>0.93</b>	<b>0.93</b>	1.01
AR-MA-SV	<b>0.93</b>	1.01	1.00	1.00	1.00
AR-ARMA-SV	1.04	1.00	1.00	1.00	1.00
UC	1.00	0.98	1.03	1.06	1.06
UC-SV	1.00	0.98	1.02	1.04	1.04
UC-ARMA	1.00	0.99	1.04	1.08	1.08
UC-MA-SV	1.00	0.98	1.02	1.05	1.04
UC-ARMA-SV	1.00	0.98	1.02	1.05	1.04

Table 15: MSFEs relative to AR benchmark: UK GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	<b>1.00</b>	1.00	1.00	1.00	1.00
AR-SV	1.00	0.96	0.91	0.87	0.83
AR-ARMA	1.04	1.03	1.00	0.97	0.93
AR-MA-SV	1.04	0.95	0.90	0.86	0.82
AR-ARMA-SV	1.02	1.03	1.01	0.98	0.94
UC	1.00	0.92	0.89	0.91	0.87
UC-SV	1.00	0.98	0.89	0.84	0.78
UC-ARMA	1.02	<b>0.90</b>	<b>0.81</b>	<b>0.75</b>	<b>0.68</b>
UC-MA-SV	1.02	0.98	0.88	0.83	0.77
UC-ARMA-SV	1.00	1.09	0.95	0.83	0.72

Table 16: MSFEs relative to AR benchmark: US GDP Deflator inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	1.00	1.00	1.00	1.00	1.00
AR-SV	0.97	0.87	0.81	0.80	0.78
AR-ARMA	0.97	0.86	0.73	0.64	0.60
AR-MA-SV	0.96	0.87	0.82	0.82	0.82
AR-ARMA-SV	0.95	<b>0.84</b>	<b>0.67</b>	<b>0.54</b>	<b>0.49</b>
UC	<b>0.93</b>	0.89	0.87	0.92	0.93
UC-SV	0.99	0.98	0.98	0.98	0.96
UC-ARMA	0.96	0.88	0.83	0.80	0.77
UC-MA-SV	1.05	1.07	1.02	0.97	0.94
UC-ARMA-SV	0.96	0.88	0.82	0.77	0.75

Since our major innovation is the estimation of ARMA-SV error models, it is useful to compare the relative forecast performance of these models to those made by their nested variants: ARMA, SV, and MA-SV models. In the case of CPI forecasts, models with ARMA-SV errors provide the best point forecasts across all horizons in both Canada, France and Italy. They also do quite well in the UK, and at longer horizons in the US (i.e.  $k = 8, 12, 16$ ). Finally, we highlight the fact that Japan's CPI inflation is quite difficult to forecast.

A slightly weaker result exists when forecasting GDP Deflator inflation. In that case, ARMA-SV models still provide the best forecasts in France, and are only bested by the simple UC model at the one-step-ahead horizon in the US. They also provide good medium term forecasts in Italy and longer term forecasts in both Canada and Germany. Finally, despite having no clearly dominant model in the CPI forecasts, the AR-ARMA

model does quite well in forecasting Japan’s GDP Deflator inflation.

Taken together, our main conclusion from this exercise is that ARMA-SV error models provide good point forecasts of CPI and GDP deflator inflation measures. In particular, with few exceptions they are able to improve on the point forecast performance gained by the simpler SV model in Stock and Watson (2007) and the MA-SV model in Chan (2013).

## 4.2 Density Forecast Results

In this subsection we report the results corresponding to the density forecasts. In particular, the forecast performance of the models are presented in Tables 17–30.

Table 17: Sum of log predictive likelihoods relative to AR benchmark: Canada CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	5.5	9.6	5.7	9.6	3.4
AR-ARMA	<b>8.6</b>	<b>17.1</b>	<b>30.7</b>	<b>46.0</b>	<b>49.3</b>
AR-MA-SV	5.4	9.5	5.9	9.6	3.6
AR-ARMA-SV	4.4	7.2	18.2	29.7	30.2
UC	-2.5	1.7	5.9	8.6	9.9
UC-SV	-1.5	7.4	9.4	15.5	8.4
UC-ARMA	-3.3	4.3	13.8	21.7	26.4
UC-MA-SV	0.6	5.1	9.6	17.0	10.2
UC-ARMA-SV	-0.8	-11.9	-4.4	6.5	4.9

Table 18: Sum of log predictive likelihoods relative to AR benchmark: France CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	5.0	15.9	25.0	27.5	31.2
AR-ARMA	15.2	24.1	9.5	-13.8	-33.6
AR-MA-SV	22.9	15.5	23.4	25.9	29.6
AR-ARMA-SV	<b>24.0</b>	17.8	7.5	-6.1	-13.5
UC	18.4	22.0	30.4	28.3	26.5
UC-SV	17.3	27.9	38.6	37.6	40.1
UC-ARMA	18.6	21.3	34.7	40.9	47.7
UC-MA-SV	14.1	21.6	36.0	39.8	43.7
UC-ARMA-SV	21.1	<b>30.1</b>	<b>55.2</b>	<b>65.4</b>	<b>77.6</b>



Table 19: Sum of log predictive likelihoods relative to AR benchmark: Germany CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	2.0	-2.4	-1.4	4.8	11.4
AR-ARMA	<b>3.8</b>	<b>0.3</b>	-1.2	-0.8	0.5
AR-MA-SV	1.6	-3.1	-2.8	3.3	9.7
AR-ARMA-SV	-0.1	-11.4	-12.1	-8.9	-1.7
UC	-12.0	-13.2	-2.4	2.3	6.6
UC-SV	-11.7	-14.0	<b>0.4</b>	<b>12.3</b>	<b>24.1</b>
UC-ARMA	-17.1	-18.4	-3.1	5.0	12.9
UC-MA-SV	-12.9	-17.3	-1.9	9.2	22.3
UC-ARMA-SV	-12.8	-16.8	-2.1	9.2	22.0

Table 20: Sum of log predictive likelihoods relative to AR benchmark: Italy CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	63.9	84.7	90.1	88.0	88.5
AR-ARMA	25.0	57.2	69.9	79.0	87.3
AR-MA-SV	<b>65.2</b>	84.7	89.9	86.8	86.3
AR-ARMA-SV	59.9	80.1	88.8	92.8	100.3
UC	0.5	27.0	29.7	25.6	20.0
UC-SV	38.2	66.9	63.6	56.1	49.5
UC-ARMA	0.4	4.4	0.4	-4.5	-9.9
UC-MA-SV	48.7	70.1	77.8	78.4	78.6
UC-ARMA-SV	46.9	<b>87.5</b>	<b>102.0</b>	<b>105.9</b>	<b>110.9</b>

Table 21: Sum of log predictive likelihoods relative to AR benchmark: Japan CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	-26.4	50.1	<b>68.7</b>	60.6	63.2
AR-ARMA	47.1	53.1	39.1	33.3	24.2
AR-MA-SV	-27.5	51.4	67.1	60.6	62.0
AR-ARMA-SV	<b>47.8</b>	50.1	32.3	19.0	5.1
UC	11.5	29.8	31.5	35.9	35.6
UC-SV	33.4	55.8	57.7	66.2	<b>68.9</b>
UC-ARMA	11.3	25.4	22.3	25.9	26.1
UC-MA-SV	32.9	<b>56.5</b>	57.0	<b>64.5</b>	66.8
UC-ARMA-SV	46.8	52.2	41.2	40.7	37.4

Table 22: Sum of log predictive likelihoods relative to AR benchmark: UK CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	21.1	17.3	60.8	51.7	51.9
AR-ARMA	45.3	69.7	48.3	37.9	26.2
AR-MA-SV	41.8	21.1	59.2	50.9	51.3
AR-ARMA-SV	<b>57.7</b>	<b>72.3</b>	34.4	15.0	-0.9
UC	-7.6	29.3	34.2	34.8	31.3
UC-SV	20.9	63.0	<b>68.7</b>	<b>71.2</b>	<b>70.7</b>
UC-ARMA	-4.7	31.6	36.0	39.8	40.4
UC-MA-SV	17.4	62.6	68.3	70.3	69.3
UC-ARMA-SV	31.2	59.7	53.5	50.5	46.3

Table 23: Sum of log predictive likelihoods relative to AR benchmark: US CPI Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	30.8	18.5	19.4	22.5	35.4
AR-ARMA	11.1	5.1	27.3	39.0	55.0
AR-MA-SV	30.6	17.0	17.8	20.0	34.4
AR-ARMA-SV	28.6	22.6	30.6	<b>40.9</b>	<b>61.2</b>
UC	-0.6	-0.3	11.5	2.9	9.3
UC-SV	<b>32.3</b>	25.9	26.2	25.2	36.9
UC-ARMA	-0.2	2.8	17.5	18.7	37.8
UC-MA-SV	30.5	<b>27.4</b>	<b>31.7</b>	34.6	50.8
UC-ARMA-SV	26.3	13.6	26.8	37.4	55.7

Table 24: Sum of log predictive likelihoods relative to AR benchmark: Canada GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	28.5	6.8	1.3	3.7	4.0
AR-ARMA	11.9	3.4	-2.6	25.8	33.0
AR-MA-SV	33.7	6.1	0.8	2.9	3.1
AR-ARMA-SV	35.5	<b>15.1</b>	<b>13.9</b>	<b>28.7</b>	<b>35.4</b>
UC	-67.2	-68.4	-60.3	-57.1	-46.1
UC-SV	33.0	14.4	6.9	9.8	11.1
UC-ARMA	14.0	-0.3	13.8	26.6	31.7
UC-MA-SV	35.5	12.1	6.5	11.4	11.2
UC-ARMA-SV	<b>37.2</b>	0.1	2.3	10.7	15.1

Table 25: Sum of log predictive likelihoods relative to AR benchmark: France GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	66.7	60.0	65.1	57.7	58.2
AR-ARMA	19.7	-2.8	2.3	8.2	14.3
AR-MA-SV	68.3	59.8	64.6	57.4	58.4
AR-ARMA-SV	74.0	57.6	<b>66.2</b>	<b>69.4</b>	<b>72.3</b>
UC	-3.8	4.6	17.4	23.1	25.5
UC-SV	69.7	<b>59.9</b>	64.3	59.9	56.5
UC-ARMA	3.1	9.6	23.5	31.9	37.8
UC-MA-SV	<b>70.6</b>	58.0	62.4	58.8	56.1
UC-ARMA-SV	3.1	9.6	23.5	31.9	37.8

Table 26: Sum of log predictive likelihoods relative to AR benchmark: Germany GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	<b>15.5</b>	<b>15.0</b>	19.5	17.3	16.8
AR-ARMA	4.3	-12.1	-20.9	-29.4	-36.0
AR-MA-SV	15.4	14.2	18.0	15.6	15.3
AR-ARMA-SV	11.2	2.3	1.8	-2.8	-3.7
UC	-5.2	-3.0	4.3	4.2	4.5
UC-SV	13.4	13.7	<b>21.6</b>	<b>20.3</b>	<b>21.4</b>
UC-ARMA	-3.1	-3.2	4.5	6.1	7.9
UC-MA-SV	13.4	14.0	20.9	19.3	20.3
UC-ARMA-SV	13.9	13.9	19.8	18.0	19.1

Table 27: Sum of log predictive likelihoods relative to AR benchmark: Italy GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	-1.8	-1.6	-2.0	-3.6	-3.8
AR-ARMA	6.6	1.4	0.3	-0.8	-1.1
AR-MA-SV	-0.5	-1.6	-2.1	-3.8	-3.8
AR-ARMA-SV	3.5	-0.4	-1.6	-3.4	-3.6
UC	7.6	4.1	3.8	3.7	3.4
UC-SV	11.8	7.5	5.8	3.5	2.3
UC-ARMA	10.2	6.9	6.5	<b>6.5</b>	<b>6.3</b>
UC-MA-SV	13.0	9.5	7.9	5.2	4.1
UC-ARMA-SV	<b>11.9</b>	<b>10.3</b>	<b>8.7</b>	6.1	4.9

Table 28: Sum of log predictive likelihoods relative to AR benchmark: Japan GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	7.9	5.7	7.5	10.6	8.3
AR-ARMA	-9.1	-1.7	-6.0	-2.7	-18.4
AR-MA-SV	8.0	5.9	7.5	10.8	8.4
AR-ARMA-SV	8.4	7.0	8.8	<b>12.3</b>	<b>9.0</b>
UC	8.1	<b>9.1</b>	7.4	6.3	3.6
UC-SV	7.7	7.3	9.3	11.1	9.1
UC-ARMA	-4.9	-3.9	-6.0	-9.5	-5.9
UC-MA-SV	<b>9.0</b>	7.3	9.1	10.8	8.7
UC-ARMA-SV	8.6	7.4	<b>9.5</b>	11.0	8.8

Table 29: Sum of log predictive likelihoods relative to AR benchmark: UK GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	21.1	17.3	<b>60.8</b>	<b>51.7</b>	<b>51.9</b>
AR-ARMA	45.3	69.7	48.3	37.9	26.2
AR-MA-SV	41.8	21.1	59.2	50.9	51.3
AR-ARMA-SV	<b>57.7</b>	<b>72.3</b>	34.4	15.0	-0.9
UC	-10.2	23.8	27.5	25.9	23.9
UC-SV	8.0	42.6	48.8	52.3	55.3
UC-ARMA	-7.5	23.8	27.0	29.3	29.9
UC-MA-SV	6.8	42.0	48.1	52.6	55.0
UC-ARMA-SV	8.1	10.1	9.7	12.2	16.3

Table 30: Sum of log predictive likelihoods relative to AR benchmark: US GDP Deflator Inflation

	$k = 1$	$k = 4$	$k = 8$	$k = 12$	$k = 16$
AR	0.0	0.0	0.0	0.0	0.0
AR-SV	15.5	34.6	38.3	38.9	42.0
AR-ARMA	13.2	35.4	48.6	60.9	68.8
AR-MA-SV	17.2	28.1	29.9	29.6	31.6
AR-ARMA-SV	<b>18.8</b>	<b>40.7</b>	<b>59.7</b>	<b>71.6</b>	<b>80.8</b>
UC	-4.8	13.8	17.3	10.5	6.3
UC-SV	17.5	35.3	35.0	28.3	25.2
UC-ARMA	4.9	13.5	17.3	19.2	20.6
UC-MA-SV	14.1	31.6	33.3	31.5	32.2
UC-ARMA-SV	18.1	40.7	49.6	54.6	58.6

In line with the point forecast results, the AR-SV and UC-SV models tend to dominate their constant volatility counterparts in CPI forecasts. However, the result here is stronger in the sense that it also extends to the GDP deflator. More generally, models with ARMA-SV errors tend to provide better forecasts than those with SV, ARMA or homoscedastic errors. This is especially true in forecasting the GDP deflator in the US, or CPI in France. One notable exception is that the AR-ARMA model provides the best forecast of CPI in Canada across all horizons. Another is the one-step-ahead forecast of US CPI, in which the simpler UC-SV model provides the best density forecast. This result is important because it is in contrast with those in Chan (2013). In that paper, the UC-MA-SV model is found to provide better forecasts of the same variable on an earlier sample. More precisely, Chan (2013) uses CPI inflation from 1947Q1 to 2011Q3, while we extend this up to 2017Q4. Thus, our result suggests that, in the case of US inflation forecasts, the relative performance of the UC-MA-SV model has declined since the end of 2011. That being said, the UC-MA-SV model still provides good forecasts of US CPI inflation at the one- and two-year-ahead forecast horizons.

Thus, our main conclusion from this exercise is that ARMA-SV error models again provide competitive forecasts across the sample of countries considered in this study. This is especially true in France, Canada and the US, where they provide the best forecasts of CPI for France and the GDP Deflator for Canada and the US across all forecast horizons.

## 5 Concluding Remarks

We have introduced a new class of dynamic models with ARMA-SV errors, provided details on how to estimate them, and shown that they can be useful in forecasting inflation. The main difficulty in estimating such models is that the ARMA component induces serial dependence in the measurement errors, making the standard Kalman filter not directly applicable. We showed that this could be overcome by carefully designing the order of matrix operations. Moreover, by exploiting the model structure, we were able to develop an efficient algorithm that avoids the forward and backward recursions in the Kalman filter. To illustrate the usefulness of the models, we assessed their forecast performance of two commonly used inflation measures: CPI and the GDP Deflator, in each of the G7 countries. More specifically, we presented both out-of-sample point and density forecast performance to various nested AR and UC models.

While there was no clearly dominant model across each of the countries, the AR-ARMA-

SV model provided highly competitive forecasts of both inflation measures. In particular, they provided the best one-step-ahead point forecasts of CPI in all countries except Germany and the US. The model also dominated the CPI point forecasts at all other horizons in Canada, while the UC-ARMA-SV variant dominated in both France and Italy. This latter result extended to the density forecasts. In the former case, however, the simpler AR-ARMA model produced the best density forecasts across all forecast horizons. Finally, ARMA-SV models dominated the GDP Deflator density forecasts in Canada the US, provided good short-term forecasts in Italy and the UK, and good long-term forecasts in France.

## Appendix: Proof of $\mathbf{H}_\psi^{-1}\mathbf{H}_\phi = \mathbf{H}_\phi\mathbf{H}_\psi^{-1}$

**Proposition:** Suppose  $\mathbf{H}_\phi$  and  $\mathbf{H}_\psi$  are the following matrices of size  $T$ :

$$\mathbf{H}_\phi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -\phi_1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ -\phi_p & \cdots & -\phi_1 & 1 & \cdots & 0 \\ \vdots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & -\phi_p & \cdots & -\phi_1 & 1 \end{pmatrix}, \quad \mathbf{H}_\psi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \psi_1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \psi_q & \cdots & \psi_1 & 1 & \cdots & 0 \\ \vdots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \psi_q & \cdots & \psi_1 & 1 \end{pmatrix}.$$

Then:  $\mathbf{H}_\phi^{-1}$  and  $\mathbf{H}_\psi$  commute, i.e.,  $\mathbf{H}_\psi^{-1}\mathbf{H}_\phi = \mathbf{H}_\phi\mathbf{H}_\psi^{-1}$ .

**Proof:** Let  $\mathbf{L}_i$  be a  $T \times T$  matrix which only has the nonzero elements 1 on the  $i$ -th lower diagonal for  $i = 0, \dots, T-1$ , i.e.,

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

In particular,  $\mathbf{L}_0 = \mathbf{I}_T$  (identity matrix). It is easy to check that  $\mathbf{L}_i\mathbf{L}_j = \mathbf{L}_{i+j} = \mathbf{L}_j\mathbf{L}_i$ , when  $i, j \geq 0$  and  $i+j \leq T-1$ . Then we can write  $\mathbf{H}_\phi$  and  $\mathbf{H}_\psi$  as:

$$\mathbf{H}_\phi = \mathbf{I}_T - \sum_{i=1}^p \phi_i \mathbf{L}_i, \quad \mathbf{H}_\psi = \mathbf{I}_T + \sum_{j=1}^q \psi_j \mathbf{L}_j.$$

So that:

$$\begin{aligned}
\mathbf{H}_\phi \mathbf{H}_\psi &= \left( \mathbf{I}_T - \sum_{i=1}^p \phi_i \mathbf{L}_i \right) \left( \mathbf{I}_T + \sum_{j=1}^q \psi_j \mathbf{L}_j \right) \\
&= \mathbf{I}_T - \sum_{i=1}^p \phi_i \mathbf{L}_i + \sum_{j=1}^q \psi_j \mathbf{L}_j - \sum_{i=1}^p \sum_{j=1}^q \phi_i \psi_j \mathbf{L}_i \mathbf{L}_j \\
&= \mathbf{I}_T + \sum_{j=1}^q \psi_j \mathbf{L}_j - \sum_{i=1}^p \phi_i \mathbf{L}_i - \sum_{j=1}^q \sum_{i=1}^p \psi_j \phi_i \mathbf{L}_j \mathbf{L}_i \\
&= \left( \mathbf{I}_T + \sum_{j=1}^q \psi_j \mathbf{L}_j \right) \left( \mathbf{I}_T - \sum_{i=1}^p \phi_i \mathbf{L}_i \right) \\
&= \mathbf{H}_\psi \mathbf{H}_\phi.
\end{aligned}$$

Hence:

$$\begin{aligned}
\mathbf{H}_\psi^{-1} (\mathbf{H}_\phi \mathbf{H}_\psi) \mathbf{H}_\psi^{-1} &= \mathbf{H}_\psi^{-1} (\mathbf{H}_\psi \mathbf{H}_\phi) \mathbf{H}_\psi^{-1} \\
\mathbf{H}_\psi^{-1} \mathbf{H}_\phi &= \mathbf{H}_\phi \mathbf{H}_\psi^{-1}.
\end{aligned}$$



## References

- Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.
- George EP Box and Gwilym M Jenkins. *Time series analysis: Forecasting and control*. Holden-Day, 1970.
- Joshua C. C. Chan. Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176(2):162–172, 2013.
- Joshua C. C. Chan and Ivan Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(2):101–120, 2009.
- Joshua C. C. Chan, Gary Koop, and Simon M. Potter. A new model of trend inflation. *Journal of Business and Economic Statistics*, 31(1):94–106, 2013.
- Siddhartha Chib and Edward Greenberg. Bayes inference in regression models with ARMA( $p$ ,  $q$ ) errors. *Journal of Econometrics*, 64:183–206, 1994.
- Todd E. Clark and Francesco Ravazzolo. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575, 2015.
- Jamie Cross and Aubrey Poon. Forecasting structural change and fat-tailed events in Australian macroeconomic variables. *Economic Modelling*, 58:34–51, 2016.
- Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.
- Gene H. Golub and Charles F. Van Loan. *Matrix computations*, volume 3. The Johns Hopkins University Press, 2013.
- Sangjoon Kim, Neil Shepherd, and Siddhartha Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65: 361–393, 1998.
- Gary Koop. *Bayesian Econometrics*. John Wiley and Sons, New York, 2003.

- Dirk P. Kroese, Thomas Taimre, and Zdravko I. Botev. *Handbook of Monte Carlo methods*. John Wiley and Sons, New York, 2011.
- Stefan Ludbergh, Timo Teräsvirta, and Dick van Dijk. Time-varying smooth transition autoregressive models . *Journal of Business and Economic Statistics*, 21(1):104–121, 2003.
- Massimiliano Marcellino. Forecasting EMU macroeconomic variables . *International Journal of Forecasting*, 20:359–372, 2004.
- William J. McCausland, Shirley Miller, and Denis Pelletier. Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics and Data Analysis*, 55(1):199–212, 2011.
- Teruo Nakatsuma. Bayesian analysis of ARMA–GARCH models: A Markov chain sampling approach. *Journal of Econometrics*, 95(1):57–69, 2000.
- James H. Stock and Mark W. Watson. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14:11–30, 1996.
- James H. Stock and Mark W. Watson. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33, 2007.
- Stephen J. Taylor. Modeling stochastic volatility: A review and comparative study. *Mathematical finance*, 4(2):183–204, 1994.