Bayesian Inference and Forecasting in the Stationary Bilinear Model

ROBERTO LEON-GONZALEZ † and FUYU YANG ‡

†*National Graduate Institute for Policy Studies,7-22-1 Roppongi, Minato-ku, Tokyo, Japan, 106-8677*
(e-mail: email: rlg@grips.ac.jp)
‡*School of Economics, University of East Anglia, Norwich, NR4 7TJ, UK*
(e-mail: fuyu.yang@uea.ac.uk)

Abstract

A stationary bilinear (SB) model can be used to describe processes with a time-varying degree of persistence that depends on past shocks. This study develops methods for Bayesian inference, model comparison, and forecasting in the SB model. Using monthly U.K. inflation data, we find that the SB model outperforms the random walk, first order autoregressive AR(1), and autoregressive moving average ARMA(1,1) models in terms of root mean squared forecast errors . In addition, the SB model is superior to these three models in terms of predictive likelihood for the majority of forecast observations.

JEL classification numbers: C5, C11, C32

Keywords: Stationary bilinear model, Markov chain Monte Carlo, model comparison

1   Introduction

The class of bilinear processes was first proposed by Granger and Andersen (1978a) and was found to be able to, as stated in Raeburn et al. (1995), 'approximate any nonlinear model to an arbitrary degree of accuracy over a finite time interval'. Brunner and Hess (1995) note that the bilinear model's capacity to approximate any well-behaved nonlinear relationship is analogous to the ability of an autoregressive moving average (ARMA) model to approximate well-behaved linear relationships. Bilinear models have been successfully applied to analyse macroeconomic and financial series to capture data non-linearity; see, for example, Byers and Peel (1995), Charemza et al. (2005), and Hristova (2005).

The SB model is specified as the following:

$$y_t = (a + b\varepsilon_{t-1}) y_{t-1} + \varepsilon_t, \tag{1}$$

where b is the bilinear term and $\varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2)$. Granger and Andersen (1978a) note that a second order stationarity condition for the SB model is

$$a^2 + b^2 \sigma_\varepsilon^2 < 1, \tag{2}$$

where a, $b \neq 0$. Along with equation (2), Sesay and Subba Rao (1988) and Kim et al. (1990) established the following necessary restrictions (equations 3–5) to meet the assumption that the first four moments of $\{y_t\}_{t=1}^T$ are finite.

$$|a| < 1, \tag{3}$$

$$\left| a^3 + 3ab^2 \sigma_\varepsilon^2 \right| < 1, \tag{4}$$

and

$$a^4 + 6a^2 b^2 \sigma_\varepsilon^2 + 3b^4 \sigma_\varepsilon^4 < 1. \tag{5}$$

The dynamics in an SB process are driven by the idiosyncratic shocks ε_t, whereas the persistence parameter, $a + b\varepsilon_{t-1}$, is driven by the past shock ε_{t-1}. Because the persistence parameter is defined as 'the sum of the coefficients on the lagged dependent variable' (see O'Reilly and Whelan, 2005), therefore, a non-zero bilinear term b together with the error term ε_{t-1} would induce a time-varying persistence that changes corresponding to the lagged shock ε_{t-1}.

According to equation (1), the one-step-ahead forecast of y_{t+1} using the SB model is as follows:

$$E_t \left(y_{t+1} \right) = \left(a + b \varepsilon_t \right) y_t,$$

which indicates that a large shock in the system may affect both the expectations of future values and induce changes in the series' persistence. Moreover, if a series is modelled with an SB process, all of the past shocks $\{ \varepsilon_i \}_{i=1}^{t}$ are inevitably propagated via equation (1) to affect the one-step-ahead forecast $E_t \left(y_{t+1} \right)$. If $E_t \left(y_{t+1} \right)$ is constructed in such a way that the bilinear term b and ε_t are neglected, the forecast is distorted; see Charemza et al. (2005) for empirical illustrations. For instance, if b is misspecified as 0, the SB model is equivalent to a linear AR(1) model without an intercept:

$$y_t = a y_{t-1} + \varepsilon_t.$$

Hence, the forecast is distorted by the amount, $b \varepsilon_t y_t$, if the forecasting model is misspecified as the above lower order autoregressive AR model.

The SB process is very useful for empirical applications. However, as noted in the work of Brunner and Hess (1995), the estimation of an SB model could be problematic, especially when at least one of the four moment conditions in equations (2)–(5) is close to being violated.

Brunner and Hess (1995) simulate 10,000 series of data using a first order bilinear process (specified as in equation (1)) with a few sets of 'true' values, where the first moment condition (which is also one of the stationarity conditions) is close to being violated. They demonstrate that the expected negative log-likelihood function will be characterized with a long narrow spike under the 'true' values. Therefore, because of this characteristic in the expected log-likelihood function, the estimates of the bilinear parameters from a standard optimization routine, e.g., maximum likelihood method, are more likely to be biased away from the 'true' values. However, if the moment conditions are easily satisfied, the well-behaved expected log-likelihood function has a global optimum located over the 'true' parameter values, and the global optimum can be found easily.

In this study, we focus on making Bayesian inference in an SB model, which has a single bilinear term and a, b, and σ_{ε}^2 jointly meet the moment restrictions in equations (2)–(5). The Bayesian estimation method for the SB model

and the model comparison method for non-nested models that we propose may overcome some of the difficulties highlighted in the literature. Despite the simple specification of the SB model, the dynamics in the changing persistence of inflation can be nicely captured by the SB. The SB model specification satisfies the stationarity conditions, which makes it distinct from many other bilinear model specifications that have been investigated in the literature, such as in Chen (1992a), Charemza et al. (2005),Bibi and Lessak (2009), and Feng et al. (2013). In this study, we aim to develop an efficient sampling algorithm that enables us to simulate posterior distributions for all parameters of interest jointly satisfying the stationarity restrictions.

Chen (1992a) has also proposed to estimate the bilinear models using Bayesian techniques. However, in Chen (1992a), the stationarity conditions are not given special considerations to elicit the priors. Because a posterior simulator is normally constructed by combining the prior and the likelihood function in the Bayesian framework, in particular, the elicited priors should reflect a 'priori' distributional belief in the parameters of interest to a certain extent, if not fully. Especially considering the characteristics of the expected log-likelihood function of the bilinear models in Brunner and Hess (1995), the prior elicitation is crucial for making Bayesian inference in the SB model. The Gibbs sampling algorithm proposed by Chen (1992a) is efficient using the untruncated elicited priors because the data simulated for applications in Chen (1992a) can easily satisfy the stationarity conditions. However, if a bilinear data generating process does not easily satisfy the first moment condition in equation (2), the Gibbs sampling algorithm proposed in Chen (1992a), and forecasting method in Chen (1992b), will not be applicable.

In particular, the sampling algorithm proposed by Chen (1992a) is not suitable for bilinear model specifications with the first moment condition close to being violated. However, for macroeconomic modellers, such bilinear model specifications are more appealing and intuitive because this type of underlying process allows the data to be highly persistent, and it also allows the persistence to be correlated with shocks in previous periods.

In this study, we refer the lower order autoregressive AR model with an intercept as AR(1) model from here onwards. With an application to monthly U.K. inflation, we compare the in-sample fit of four non-nested models, the SB model,

the RW model, an AR(1) model, and an ARMA(1,1) using the likelihood and the marginal likelihood. We focus on illustrating the flexible model comparison method that can be used for both nested and non-nested models, rather than exploring all possible competing models and compare them with the SB model.

To compare the out-of-sample forecasting capacity of competing models, besides evaluating the root mean squared forecast errors (RMSFE) and applying the Diebold-Mariano test for forecast accuracy comparison, the comparison is also carried out using the predictive likelihood; (see Geweke and Amisano, 2010). We find that the SB model consistently provides the highest predictive likelihood for forecast observations. This paper provides strong empirical evidence that the variation in the estimated inflation persistence appears to be lower after the 1990s compared with those in the 1970s and 1980s. The results agree with previous findings in the literature that the structure of inflation persistence may have changed and that inflation was much more anchored during the recent financial crisis than during the 1980s (see Watson, 2014).

The remainder of the paper is organized as follows. Section 2 presents methods in Bayesian inference in the SB model specified in equation (1). Section 3 uses monthly U.K. inflation data to estimate the SB model and to compare it to other models in terms of forecasting accuracy at different horizons. Section 4 concludes.

2 Bayesian Inferences in the SB Model

We start by introducing some notation: a time series with a sample size of N is denoted as $\mathbf{y} = (y_1, \cdots, y_n)'$, and we assume the first observation y_1 is the initial observation. The error disturbances ε_t for $t = 1, \cdots, n$ are $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)'$. Following the recommendations in Charemza et al. (2005), the initial disturbance value is set as $\varepsilon_1 = 0$. We denote the error precision $h_\varepsilon = \sigma_\varepsilon^{-2}$ and collect all parameters of interest in θ, where $\theta = (a, b, h_\varepsilon)'$. Let $N(\mu, V)$ denote the normal distribution with mean μ and variance V. Therefore, the normal density is denoted by $f_N(\mu, V)$. Suppose a random variable x follows a gamma distribution, $x \sim \Gamma(\alpha, \beta)$, where α is the shape parameter and β is the scale parameter, then the gamma density is $f_\Gamma(\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)$. [1]

---

[1]Later on in this study, the gamma prior is truncated to a region, where a second order stationarity condition is met.

Using equation (1) recursively, we can write

$$\varepsilon_{t-1} = y_{t-1} - (a + b\varepsilon_{t-2})\, y_{t-2},$$

for $t = 2, \cdots, n$, so that y_t can be written as

$$y_t = a f_1(t, b) + f_2(t, b) + \varepsilon_t, \tag{6}$$

where

$$f_1(t, b) = \sum_{i=1}^{t-1} \left[(-b)^{i-1} \prod_{j=1}^{i} y_{t-j} \right],$$

and

$$f_2(t, b) = \sum_{i=1}^{t-2} \left[(-1)^{i+1} b^i y_{t-i} \prod_{j=1}^{i} y_{t-j} \right].$$

Note that $f_2(t, b) = 0$ when $t = 2$. The corresponding derivations can be found in Appendix A. Denote $F_{t-1} = (y_1, \cdots, y_{t-1})'$. According to equation (6), the likelihood function of $p(y_t \mid \theta, F_{t-1})$ for $t \geq 2$ is as follows:

$$p(y_t \mid \theta, F_{t-1}) = \frac{h_\varepsilon^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \exp \left\{ -\frac{h_\varepsilon}{2} [y_t - a f_1(t, b) - f_2(t, b)]^2 \right\},$$

which leads to a joint likelihood function of $p(\mathbf{y} \mid \theta)$ expressed as follows:

$$p(\mathbf{y} \mid \theta) = \frac{h_\varepsilon^{\frac{1}{2}(N-1)}}{(2\pi)^{\frac{1}{2}(N-1)}} \exp \left\{ -\frac{h_\varepsilon}{2} \sum_{t=2}^{N} [y_t - a f_1(t, b) - f_2(t, b)]^2 \right\}. \tag{7}$$

The density function of the prior, which incorporates the moment restrictions, is

$$p(a, b, h_\varepsilon) = p(b \mid h_\varepsilon, a)\, p(h_\varepsilon \mid a)\, p(a)\, I_{\left[a^2 + b^2 h_\varepsilon^{-1} < 1 \right]}(a, b, h_\varepsilon)\, I_{[M]}(a, b, h_\varepsilon), \tag{8}$$

where M is the region in which the inequalities (4) and (5) are satisfied. From equation (7), we are not able to elicit a prior for b such that the conditional posterior follows a standard distribution. Therefore, for simplicity, we assume b to be uniformly distributed within a region such that the moment condition in equation (2) is satisfied. With the derived likelihood in equation (7) and the joint prior in equation (8), the joint posterior is given by Bayes theorem as

$$
\begin{aligned}
p(a, b, h_\varepsilon \mid y) \;\; &\propto \;\; p(y \mid a, b, h_\varepsilon)\, p(a, b, h_\varepsilon) \\
&\propto \;\; p(y \mid a, b, h_\varepsilon)\, p(b \mid h_\varepsilon, a)\, p(h_\varepsilon \mid a)\, p(a) \\
&\quad\;\; I_{\left[a^2 + b^2 h_\varepsilon^{-1} < 1 \right]}(a, b, h_\varepsilon)\, I_{[M]}(a, b, h_\varepsilon).
\end{aligned}
$$

The conditional posterior of h_ε is a truncated gamma and that of a is a truncated normal.

A truncated normal prior for a is elicited, $p(a) = f_N\left(\overline{\mu}_a, \overline{V}_a\right) I_{[|a|<1]}(a)$, where $I_{[A]}(\cdot)$ is the indicator function indicating that the moment restriction in equation (3) is satisfied. By choosing $\overline{\mu}_a = 0$ and $\overline{V}_a = 10^4$, the truncated normal prior is roughly equivalent to a uniform prior in the interval that verifies the restriction. We also try to depart from this value as well in a sensitivity analysis. We elicit a gamma prior for the error precision h_ε as $p(h_\varepsilon \mid a) = f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$. Note that the typical improper non-informative prior $(p(h_\varepsilon \mid a) \propto |h_\varepsilon|^{-1})$ corresponds to $p(h_\varepsilon \mid a) = f_{\Gamma\varepsilon}\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$ with $\underline{\alpha}_\varepsilon = 0$ and $\underline{\beta}_\varepsilon = \infty$. To make the prior for h_ε as non-informative as possible, yet ensure it is a proper prior, we choose $\underline{\alpha}_\varepsilon = 1$. By choosing $\underline{\alpha}_\varepsilon = 1$, the distribution of $\sigma_\varepsilon^2 = h_\varepsilon^{-1}$ is an inverted gamma with one degree of freedom. Therefore, it has no prior moments with infinite mean and variance, which implies great uncertainty. We may try a range of values for $\underline{\beta}_\varepsilon$ for prior robustness analysis. The smaller $\underline{\beta}_\varepsilon$ is, the greater prior weight is given to large values of σ_ε^2, which results in a greater chance that y_t is large in absolute value. To analyse macroeconomic series, such as inflation, we have elicited $\underline{\beta}_\varepsilon = 0.01$ as a reasonable value. For model comparison purpose, $\underline{\alpha}_\varepsilon = 1$ and $\underline{\beta}_\varepsilon = 0.01$ are also elicited to estimate the error variance in an AR(1) and an ARMA(1,1) model.

The biggest challenge in developing the sampling algorithm is to take random draws from the truncated posteriors. Because the truncation area could be far in the tail of the posterior distributions, to draw from these truncated distributions, we use the mixed rejection algorithm proposed by Geweke (1991). The following Gibbs sampling scheme uses the conditional posterior densities outlined in Appendix A to get draws from the posterior.

Gibbs sampling for a, b, and h_ε

1. Give initial values to a and b.

2. Repeat (a)–(c) S times and discard the first S_0 draws.

   (a) Sample h_ε from $h_\varepsilon \mid y, a, b \sim \Gamma\varepsilon\left(\overline{\alpha}_\varepsilon, \overline{\beta}_\varepsilon\right) \cdot I_{[h_\varepsilon > b^2/(1-a^2)]}(a, b, h_\varepsilon) I_{[M]}(a, b, h_\varepsilon)$ using mixed rejection sampling.

(b) Sample a from $a \mid y, b, h_\varepsilon \sim N\left(\overline{\mu}_a, \overline{V}_a\right) \cdot I_{\left[|a| < \sqrt{1 - b^2 h_\varepsilon^{-1}}\right]}\left(a, b, h_\varepsilon\right) I_{[M]}\left(a, b, h_\varepsilon\right)$ using mixed rejection sampling.

(c) Sample b from a region, where $p\left(b | y, a, h_\varepsilon\right) \propto \exp\left\{-\frac{h_\varepsilon}{2} \sum_{t=2}^{N} \left[y_t - af_1\left(t, b\right) - f_2(t, b)\right]^2\right\}$ $I_{\left[|b| < \sqrt{(1 - a^2) h_\varepsilon}\right]}\left(a, b, h_\varepsilon\right) I_{[M]}\left(a, b, h_\varepsilon\right)$ using the Griddy–Gibbs sampling algorithm; for example, see Ritter and Tanner (1992).

The averages of the draws from the Gibbs sampling are estimates of the posterior means. Convergence diagnostics can be conducted to examine the efficiency of the proposed Markov chain Monte Carlo (MCMC) algorithm.

3   Application to U.K. Inflation Rates

Because the SB specification accommodates nonlinearity and allows for time varying persistence, the SB model can be a natural candidate to model inflation. Analysing inflation persistence in order to improve on the inflation forecasting and understanding inflation in response to idiosyncratic shocks have been the primary interests of macroeconomic modellers. In particular, the analysis of the dynamics of inflation persistence has received increasing attention because it is widely believed that inflation persistence is closely related to the monetary regime and the effectiveness of monetary policies. As noted in the work of Meller and Nautz (2012), if inflation persistence is high, then shocks to inflation have long-lived effects. However, if inflation persistence is low, steering inflation expectations is more effective and inflation can be brought in line with the target quickly.

It is suggested in O'Reilly and Whelan (2005) and Osborn and Sensier (2009), that the persistence of inflation may have changed in the last three decades in the U.K. and Euro area due to substantial changes in the monetary regime over time. Meller and Nautz (2012) use the introduction of European Monetary Union (EMU) as a natural experiment, and they find that the persistence of inflation has significantly decreased in the Euro area, probably, as a result of effective monetary policy of the ECB.

In this study, we apply the U.K. inflation to the SB model. The data are 12-month inflation values. For instance, the inflation to July 2014 indicates the

changes in prices between July 2014 and July 2013. The data series is collected from the Office for National Statistics (ONS), which has 703 observations covering the period from January 1956 to July 2014. Two important events occurred in the sample period. In 1992, a new framework for monetary policy commenced following the U.K.'s exit from the European Community's Exchange Rate Mechanism. In 1997, the central bank was granted operational independence over monetary policy, and inflation was targeted.

With an application to the SB model, we provide strong empirical evidence that both the level and variation of U.K. inflation persistence have decreased after the 1990s. Moreover, we illustrate how the persistence and inflation dynamics change in response to the shocks. However, evaluating whether the monetary policies carried out by the Bank of England are effective is beyond the scope of this study.

Model Estimation

We fit the data with the first order SB model in equation (1) using the MCMC algorithm proposed in Section 2. Table 1 provides the estimation results using the Gibbs sampler from 15000 iterations. The first 5000 draws are discarded. For prior sensitivity analysis, we choose two different priors of a: a tight prior with $\underline{V}_a = 0.01$, and a flat prior with $\underline{V}_a = 10^4$. A Griddy–Gibbs sampler algorithm is applied to sample b. Because the sampling algorithm is efficient and the computational cost of having numerous grids is small, we set the grid number as 1500. From table 1, all absolute convergence diagnostic (CD) values are smaller than 1.96. Therefore, we do not find evidence that the chain does not converge. Because there is not much difference in the estimated results when a flat prior or a tight prior is used, the algorithm is not sensitive to elicited priors.

Figure 1 plots the posterior draws and histogram plots. The results from the Bayesian MCMC approach indicate that a is close to unity and that the bilinearity term b is non-negligible. A positive value of b indicates that a positive shock in period $t-1$ will lead to higher inflation persistence in period t. In other words, inflation is unlikely to be anchored if a sequence of positive idiosyncratic shocks occur.

Figure 2 and figure 3 plot the filtered shocks, and estimated persistence with

Table 1: *Estimation results from the SB model applied to monthly U.K. inflation data. The elicited prior a is $a \sim N\left(\overline{\mu}_a, \overline{V}_a\right) I_{[|a|<1]}(a)$, where $\overline{\mu}_a = 0.98$. For prior sensitivity analysis, $\overline{V}_a = 10^4$ and $\overline{V}_a = 0.01$ are elicited for comparison. The prior for h_ε is $h_\varepsilon \sim \Gamma\varepsilon\left(\underline{\alpha}_\varepsilon, \underline{\beta}_\varepsilon\right)$, where $\underline{\alpha}_\varepsilon = 1$ and $\underline{\beta}_\varepsilon = 0.01$. The table reports posterior means, modes, and medians together with standard deviations. CD represents convergence diagnostic values.*

	$a \sim N\left(\underline{\mu}_a, \underline{V}_a\right) I_{[|a|<1]}$ with $\underline{\mu}_a = 0.98$, $\underline{V}_a = 10^4$				
$\underline{\alpha}_\varepsilon = 1, \underline{\beta}_\varepsilon = 0.01$	Mean	Mode	Median	St.Dev	CD
a	0.9943	0.9951	0.9946	0.0032	-0.4163
b	0.0196	0.0202	0.0196	0.0021	0.6814
σ_ε^2	0.3219	0.3227	0.3214	0.0172	1.6185

|	$a \sim N\left(\underline{\mu}_a, \underline{V}_a\right) I_{[|a|<1]}$ with $\underline{\mu}_a = 0.98$, $\underline{V}_a = 0.01$				
$\underline{\alpha}_\varepsilon = 1, \underline{\beta}_\varepsilon = 0.01$	Mean	Mode	Median	St.Dev	CD
a	0.9943	0.9944	0.9945	0.0031	-0.4739
b	0.0196	0.0201	0.0197	0.0022	-1.3336
σ_ε^2	0.3217	0.3191	0.3213	0.0170	-0.8262

corresponding inflation, respectively. Not surprisingly, both large positive and negative shocks occurred in the 1970s and early 1980s. As a result, there is a large variation in the inflation persistence between 1975 and 1985. The changes in inflation persistence after 1991 are smaller relative to those before the early 1980s, which correspond to the smaller size of shocks. These results indicate that inflation has been more anchored in the last two decades. Figure 4 plots the filtered shocks, inflation persistence, and data in two sub-sample periods. This figure illustrates that the level of inflation persistence fell during the 2000s, and there was less variation compared with the 1980s. Watson (2014) has obtained similar findings using U.S. data.

Model Comparison

In a Bayesian framework, a non-nested model comparison can be achieved by quantifying the uncertainty associated with each competing model. An initial analysis suggests a unit root in the U.K. inflation series after an application to the Phillips - Perron (PP) and augmented Dickey - Fuller (ADF) tests. However, we wish to establish whether the underlying process can be better modelled with an SB process.

In this study, three methods are applied to compare the RW, AR(1), ARMA(1,1)

Figure 1: Plots of the MCMC draws. The top panel plots the posterior draws (from left to right) for a, b, and σ_ε^2. The bottom panel plots the simulated posterior distributions (from left to right) of a, b, and σ_ε^2.
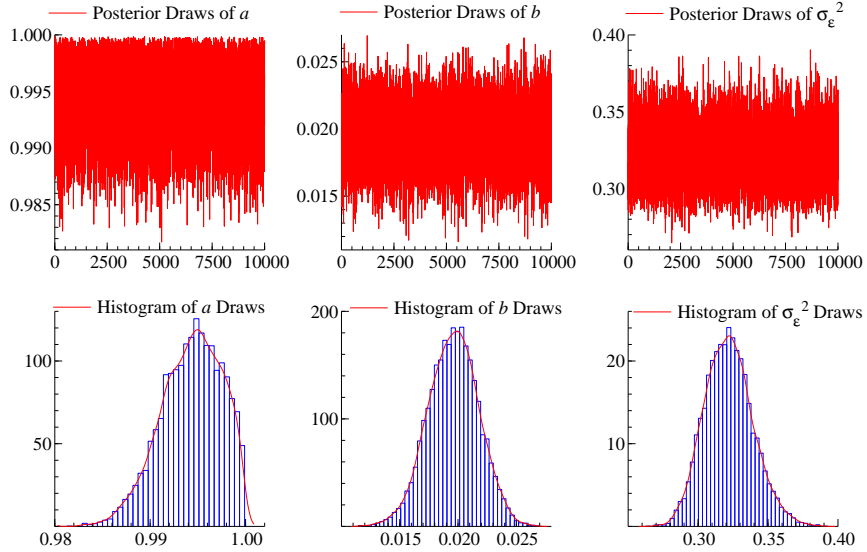

Figure 2: Inflation data with corresponding filtered shocks $\widehat{\varepsilon}_t$; The entire sample period is from January 1956 to July 2014; $t = 2, \cdots, 703$. Inflation is on left scale and filtered shocks are on right scale.
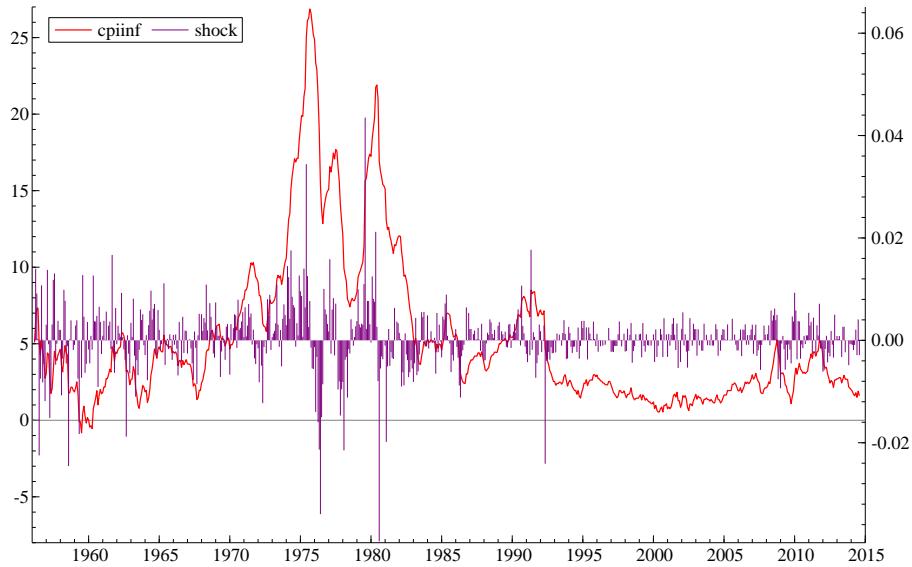

and SB models. The first method is to calculate the log-likelihood using the fitted parameter estimates. The second method is to calculate the marginal likelihood

Figure 3: Inflation data and time-varying persistence parameter $\left(\widehat{a} + \widehat{b}\widehat{\varepsilon}_{t-1}\right)$. The entire sample period is from January 1956 to July 2014; $t = 2, \cdots, 703$. Inflation is on left scale and persistence is on right scale.
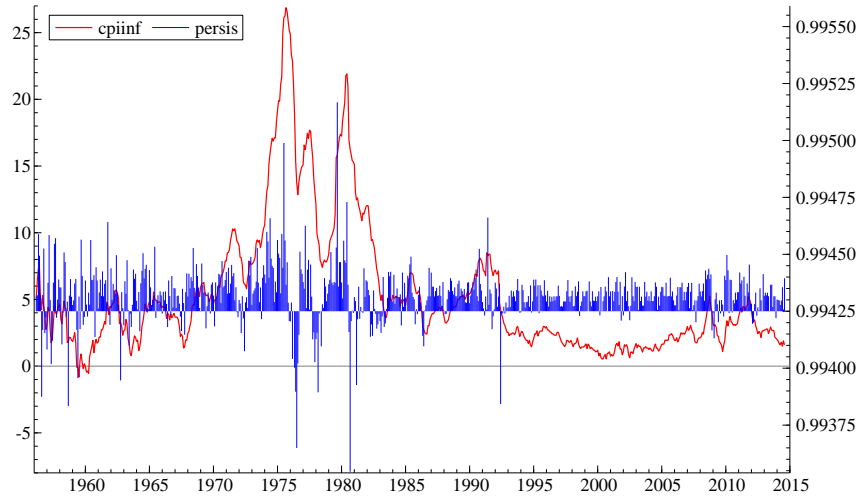
Figure 4: From top to bottom: Filtered shocks $\widehat{\varepsilon}_t$; Time-varying persistence parameter $\left(\widehat{a} + \widehat{b}\widehat{\varepsilon}_{t-1}\right)$; Inflation data. The left panels plot the sample period January 1975 – December 1984. The right panels plot the sample period January 2004 – December 2013.
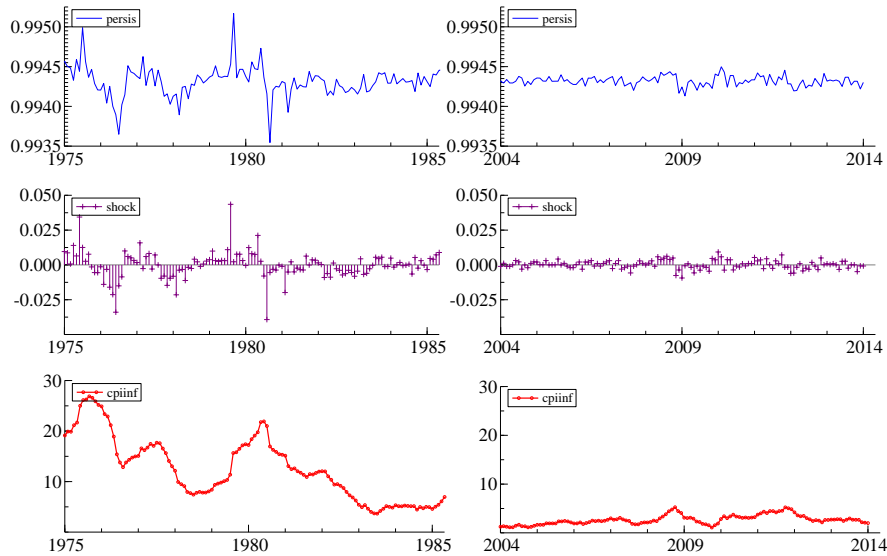
using a Gelfand–Dey method; see Gelfand and Dey (1994). The third method is to calculate the predictive likelihood proposed by Geweke (2001), which is less sensitive to the choice of priors relative to the second method using the marginal likelihood.

In Geweke (2001), the predictive likelihood is calculated by integrating the likelihood function over the posterior distribution of the unobservable at the time the prediction is made. In line with the idea that a model is as good as its predictions, the predictive likelihood provides an intuitive tool to compare across models within the Bayesian approach.

Denote as M_{SB} the SB model, denote as θ_{SB} the parameters in M_{SB} and denote as $\mathbf{y}_{t-1}^o$ all observations available up to $t-1$. The one-step-ahead predictive likelihood, evaluated at time t, is defined as

$$PL_{SB}(t) = p\left(y_t^o \mid \mathbf{y}_{t-1}^o, M_{SB}\right) = \int_{\ominus_A} p\left(y_t^o \mid \mathbf{y}_{t-1}^o, \theta_{SB}\right) p\left(\theta_{SB} \mid \mathbf{y}_{t-1}^o\right) d\theta_{SB},$$

which can be approximated by

$$S^{-1} \sum_{s=1}^{S} p\left(y_t^o \mid \mathbf{y}_{t-1}^o, \theta_{SB}^{(s)}\right)$$

where $\theta_{SB}^{(s)}$ are the MCMC draws from the posterior of θ_{SB} given data up to $t-1$. The predictive likelihood for all observations from $t = 2$ up to T is equal to the marginal likelihood and is defined as

$$p\left(\mathbf{y}_t^o \mid M_{SB}\right) = \prod_{t=2}^{T} p\left(y_t^o \mid \mathbf{y}_{t-1}^o, M_{SB}\right),$$

which implies the additive decomposition:

$$\log p\left(\mathbf{y}_t^o \mid M_{SB}\right) = \sum_{t=2}^{T} \log PL_{SB}(t).$$

Because one of the competing models is the RW model, denoted by M_{RW}, the log Bayes factor can then be decomposed as

$$\log\left[\frac{p\left(\mathbf{y}_t^o \mid M_{SB}\right)}{p\left(\mathbf{y}_t^o \mid M_{RW}\right)}\right] = \sum_{t=2}^{T} \log\left[\frac{PL_{SB}(t)}{PL_{RW}(t)}\right].$$

As stated in Geweke and Amisano (2010), the decomposition shows how individual observations contribute to the evidence in favour of the nonlinear SB model versus the linear RW model, the ARMA(1,1) model or the AR(1) model.

Figure 5: 243 observations from May 1994 to July 2014 are reserved to calculate the predictive likelihood. Top left figure plots the predictive likelihoods of the RW model, SB model, AR(1) model and ARMA(1,1) over this reserved sample period. The other three figures separately plot the predictive likelihoods of the SB model against the AR(1) model, ARMA(1,1), and RW over the period from April 2007 to July 2014.
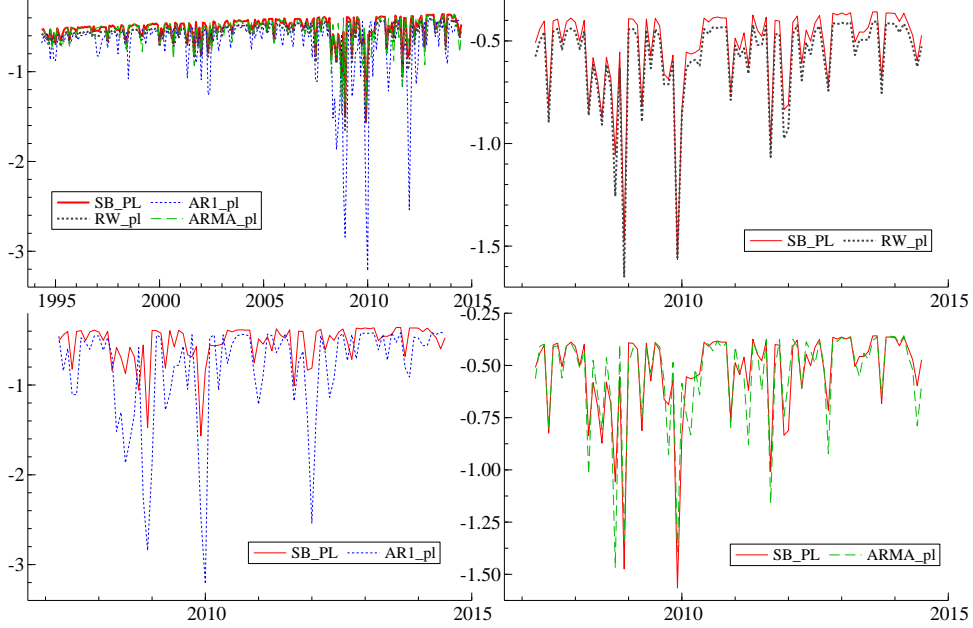
If we calculate $\log PL_{SB}(t)$ from $t = k + 1$ rather than from $t = 2$, the initial k observations are then used as a training sample. The predictive likelihood for the last $T - k$ observations is

$$\log p\left(\mathbf{y}_t^o \mid \mathbf{y}_k^o, M_{SB}\right) = \log p\left(\mathbf{y}_{(k+1):T}^o \mid \mathbf{y}_k^o, M_{SB}\right) = \sum_{t=k+1}^{T} \log PL_{SB}(t).$$

In this study, the initial $k = 460$ observations are chosen as the training sample, and the log predictive likelihood for the last 243 observations (May 1994-July 2007) is used for model comparison. Furthermore, we plot the contribution of each data point in the log predictive likelihood. In this way, we can identify the observations that contribute to the evidence in favour of the SB model versus the RW model, the ARMA(1,1) or an AR(1) model. In table 2, the log likelihoods are calculated by evaluating the log likelihood function at the posterior means of the parameters. We can see that the SB model has the highest likelihood value. Furthermore, the log marginal likelihood of the SB model, calculated using the Gelfand-Dey method, is the highest among the competing models.

14

Table 2: *Model Comparison Results: Log likelihood and log marginal likelihood are calculated using the whole sample of 703 observations. Log predictive likelihood is calculated using the last 243 observations. The competing models are SB, RW, ARMA(1,1), and AR(1). In order to calculate the log marginal likelihood using the Gelfand-Dey method (Gelfand and Dey, 1994), we have to calculate the densities of a truncated normal distribution at posterior draws. An arbitrarily selected value of p determines the size of the truncated tails. The bigger p is, the more posterior draws are dropped and the less number of draws can be used to calculate the log marginal likelihood. We selected $p = 0.01$ in this case. This value of p means the tails of normal distribution are truncated, which has 0.01 probability in them.*

	log likelihood	*log marginal likelihood* $p = 0.01$	*log predictive likelihood* *460 training sample*
SB flat prior	-597.1195	-315.0464	-126.1173
RW	-631.4283	-635.2478	-138.4045
AR(1) with intercept	-632.7988	-637.6467	-167.4863
ARMA(1,1) with intercept	-598.5505	-516.7798	-130.1119

The predictive likelihood of the last 243 observations also indicates that the SB model outperforms the RW model, the ARMA(1,1) model and the AR(1) model. At the top left of figure 5, how each observation contributes to the predictive likelihood for each of the four competing models is plotted. Comparing the predictive likelihood over multiple horizons for observed data using the four competing models, we can see that the SB model consistently obtains the highest predictive likelihood , specifically, in 208 out of 243 observations compared with AR(1), 242 out of 243 compared with RW, and in 156 out of 243 compared with ARMA(1,1).

From the top plot in figure 5, the observations from May 1994 to April 2007 contribute almost equally well to the predictive likelihood using either RW or SB model. The AR(1) and ARMA(1,1) models also receive high predictive likelihoods with observations from this sample period. However, the RW model and SB model often receive higher predictive likelihoods than the AR(1) or ARMA(1,1) models with observations from April 2007 to December 2011. This result suggests that an SB model is better in predictions compared with an AR(1) model, especially over the crisis period and immediately after.

Forecasting Exercise

Regarding inflation forecasting, a considerable section of the literature focuses on investigating the forecasting performance of linear and nonlinear econometric models in both a univariate and multivariate time series framework, e.g., Stock and Watson (1999, 2014), Barnett et al. (2014), Maheu and Song (2014). In Canova (2007), the bivariate and trivariate models suggested by the economic theory add marginal predictive accuracy compared with the univariate model after a horse race between multivariate models and univariate models. Based on the forecasting comparison results, it turns out that the random walk (RW) model or a lower order AR model often provides the smallest RMSFE, and therefore, are hard to beat, see Atkeson and Ohanian (2001).

In this section, with an application to monthly U.K. inflation, we simulate forecast densities by taking advantage of the Bayesian framework. We then compare the forecasting capacity of the SB model with other leading forecasting models, such as the AR(1) model, ARMA(1,1) and the RW model.

Granger and Andersen (1978a, p. 74) discussed the invertibility conditions of the bilinear models. Additionally, in Subba Rao (1981), the invertibility conditions for a generalized bilinear time series model are provided. Using the definition of invertibility in Granger and Andersen (1978b), the SB process is invertible if $|b\sigma_\varepsilon| < 0.606$. Limiting ourselves to a simple case, based on the estimation results from table 1, this nonlinear SB model is not subject to a non-invertibility problem. Therefore, the SB forecasting model can be used to associate the present events with past events in a unique manner.

Denoting $A_{t-1} = a + b\varepsilon_{t-1}$, the SB data generating process can be rewritten as

$$y_t = (a + b\varepsilon_{t-1}) y_{t-1} + \varepsilon_t = A_{t-1} y_{t-1} + \varepsilon_t.$$

Therefore, we can write $y_{t+1} = A_t y_t + \varepsilon_{t+1}$, where $A_t = a + b\varepsilon_t$. The one-step-ahead point forecast of y_{t+1} conditional on the current observation y_t, denoted $\widetilde{y}^{SB}_{t+1|t}$ is

$$
\begin{aligned}
\widetilde{y}^{SB}_{t+1|t} &= E_t\left[y_{t+1} \mid y_t\right] = E_t\left[A_t y_t + \varepsilon_{t+1} \mid y_t\right] \\
&= \widehat{A}_t y_t + E\left(\varepsilon_{t+1} \mid y_t\right) = \widehat{A}_t y_t,
\end{aligned}
$$

where $\widehat{A}_t = \widehat{a} + \widehat{b}\widehat{\varepsilon}_t$. The estimates of all past shocks $\widehat{\varepsilon}_i$ with $i = 2, ..., t$ can be retrieved using $\widehat{\varepsilon}_t = y_t - \left(\widehat{a} + \widehat{b}\widehat{\varepsilon}_{t-1}\right) y_{t-1}$, where the initial shock ε_1 is assumed to be 0.

If we denote the two-step-ahead point forecast of y_{t+2} conditional on the current observation y_t as $\widetilde{y}^{SB}_{t+2|t}$, given $\varepsilon_{t+1} \sim N\left(0, \sigma_\varepsilon^2\right)$, then the two-step-ahead forecast $\widetilde{y}^{SB}_{t+2|t}$ is

$$
\begin{aligned}
\widetilde{y}^{SB}_{t+2|t} &= E_t\left[y_{t+2} \mid y_t\right] = E_t E_{t+1}\left[y_{t+2} \mid y_t\right] = E_t E_{t+1}\left[A_{t+1} y_{t+1} + \varepsilon_{t+2} \mid y_t\right] \\
&= E_t E_{t+1}\left[A_{t+1}\left(A_t y_t + \varepsilon_{t+1}\right) \mid y_t\right] = \widehat{a}\left(\widehat{a} + \widehat{b}\widehat{\varepsilon}_t\right) y_t + \widehat{b}\widehat{\sigma}_\varepsilon^2.
\end{aligned}
$$

According to the law of iterated expectations, an h-step-ahead point forecast can be summarized as

$$
\begin{aligned}
\widetilde{y}^{SB}_{t+h|t} &= E_t\left[y_{t+h} \mid y_t\right] = E_t E_{t+1} \ldots E_{t+h-1}\left[y_{t+h} \mid y_t\right] \\
&= \widehat{a}^{h-1}\left(\widehat{a} + \widehat{b}\widehat{\varepsilon}_t\right) y_t + \frac{\widehat{a}^{h-1} - 1}{\widehat{a} - 1}\widehat{b}\widehat{\sigma}_\varepsilon^2.
\end{aligned}
\tag{9}
$$

In the RW forecasting model, the one-step-ahead or multi-step-ahead point forecasts will all be the same, conditional on current information:

$$
\begin{aligned}
\widetilde{y}^{RW}_{t+1|t} &= E_t\left(y_{t+1} \mid y_t\right) = y_t, \\
\widetilde{y}^{RW}_{t+h|t} &= E_t E_{t+1} \ldots E_{t+h-1}\left[y_{t+h} \mid y_t\right] = y_t.
\end{aligned}
$$

To evaluate forecasting capacity, we used the conventional RMSFE for both the one-step-ahead and multi-step-ahead forecasting exercises. Both the fixed rolling window forecast and the expanding rolling window forecast are applied. This ad-hoc rolling window approach then allows the estimated coefficients using the AR model and the ARMA(1,1) model to vary over time.

To conduct the fixed rolling window forecast, we choose a window size containing T ($T = 460$) observations for parameter estimations, where T is smaller than the total number of observations N ($N = 703$). Iterated estimation is carried out for periods $N - T$ with a fixed number of observations T that is, a fixed window size. Forecasting using an expanding rolling window allows the calibration sample to grow for each calibration period. The initial T observations are
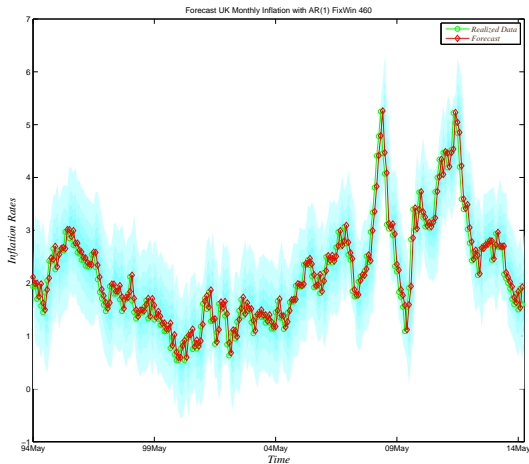
Figure 6: SB: Forecast density fan chart with a fixed rolling window; the calibration sample size is 460.
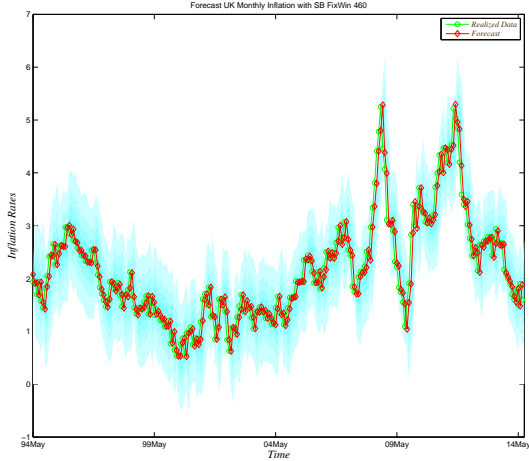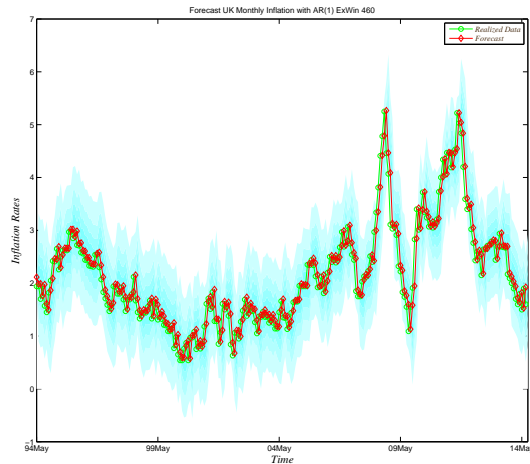
Figure 7: SB: Forecast density fan chart with an expanding rolling window; the calibration sample size is 460.
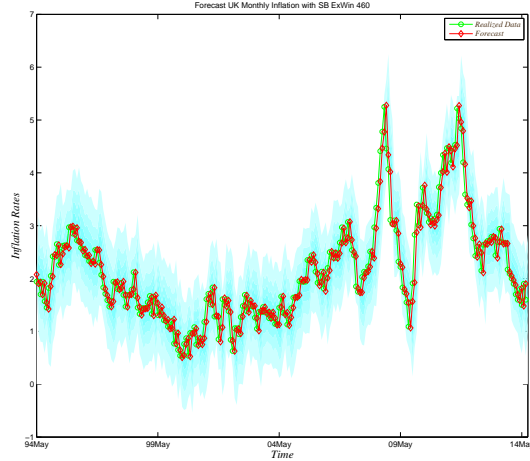
Figure 8: AR(1): Forecast density fan chart with a fixed rolling window; the calibration sample size is 460.

Figure 9: AR(1): Forecast density fan chart with an expanding rolling window; the calibration sample size is 460.

used as the first calibration sample. Then, in each iteration with an expanding rolling window, the window expands on adding one more observation.

Table 3 summarizes all RMSFE using the SB, RW, ARMA(1,1) and AR(1) forecasting models, where the fixed rolling window size is chosen as 460, and the first expanding rolling window also starts with the initial 460 observations. The SB forecasting model outperforms the RW, ARMA(1,1) and AR(1) forecasting

Table 3: *Root mean squared forecast errors (RMSFE) calculated using the SB, RW, AR(1)and ARMA(1,1) forecasting models. The forecast horizon extends from one-step-ahead forecasts to 24-step-ahead forecasts. The fixed rolling window and expanding rolling window approaches are applied. The forecasting period covers from May 1994 to July 2014.*

Fix460	1-step	2-step	3-step	4-step	5-step	6-step	7-step	8-step	9-step	10-step	11-step	12-step
SB	**0.2722**	**0.4107**	**0.5330**	**0.6310**	**0.7110**	**0.7769**	**0.8262**	**0.8716**	**0.9080**	**0.9424**	**0.9720**	**0.9984**
RW	0.2749	0.4157	0.5390	0.6385	0.7204	0.7883	0.8396	0.8870	0.9258	0.9625	0.9941	1.0216
AR(1)	0.2770	0.4160	0.5383	0.6370	0.7184	0.7858	0.8367	0.8839	0.9226	0.9592	0.9908	1.0183
ARMA(1,1)	0.2957	0.4157	0.5389	0.6384	0.7203	0.7882	0.8395	0.8870	0.9258	0.9625	0.9941	1.0216
	13-step	14-step	15-step	16-step	17-step	18-step	19-step	20-step	21-step	22-step	23-step	24-step
SB	**0.9951**	**0.9867**	**0.9674**	**0.9487**	**0.9356**	**0.9216**	**0.9270**	**0.9357**	**0.9529**	**0.9787**	**1.0024**	**1.0296**
RW	1.0193	1.0115	0.9934	0.9759	0.9642	0.9526	0.9615	0.9740	0.9954	1.0261	1.0544	1.0862
AR(1)	1.0161	1.0084	0.9905	0.9734	0.9619	0.9506	0.9597	0.9723	0.9939	1.0245	1.0527	1.0844
ARMA(1,1)	1.0193	1.0115	0.9934	0.9759	0.9642	0.9526	0.9615	0.9740	0.9954	1.0261	1.0544	1.0862
Ex460	1-step	2-step	3-step	4-step	5-step	6-step	7-step	8-step	9-step	10-step	11-step	12-step
SB	**0.2722**	**0.4108**	**0.5332**	**0.6313**	**0.7112**	**0.7772**	**0.8264**	**0.8717**	**0.9081**	**0.9422**	**0.9715**	**0.9977**
RW	0.2749	0.4157	0.5390	0.6385	0.7204	0.7883	0.8396	0.8870	0.9258	0.9625	0.9941	1.0216
AR(1)	0.2763	0.4157	0.5380	0.6367	0.7182	0.7855	0.8363	0.8835	0.9221	0.9586	0.9901	1.0175
ARMA(1,1)	0.2940	0.4157	0.5389	0.6384	0.7203	0.7882	0.8395	0.8870	0.9258	0.9625	0.9941	1.0216
	13-step	14-step	15-step	16-step	17-step	18-step	19-step	20-step	21-step	22-step	23-step	24-step
SB	**0.9943**	**0.9858**	**0.9668**	**0.9482**	**0.9353**	**0.9217**	**0.9279**	**0.9369**	**0.9544**	**0.9808**	**1.0047**	**1.0324**
RW	1.0193	1.0115	0.9934	0.9759	0.9642	0.9526	0.9615	0.9740	0.9954	1.0261	1.0544	1.0862
AR(1)	1.0152	1.0074	0.9895	0.9723	0.9608	0.9495	0.9586	0.9713	0.9928	1.0236	1.0519	1.0838
ARMA(1,1)	1.0193	1.0115	0.9934	0.9759	0.9642	0.9526	0.9615	0.9740	0.9954	1.0261	1.0544	1.0862

models on all occasions providing the smallest RMSFE. Comparing the fixed rolling window method with the expanding rolling window method, both methods provide very similar RMSFE.

Note that RMSFE in table 3 are obtained with a very large rolling window. Therefore a linear model might have been potentially disadvantaged. It is intuitive that when the calibration period is big, rich dynamics in the data become inevitable. Therefore, a nonlinear model, which allows for more dynamics, may perform better than a simple linear model in forecasting.

Table 4 provides RMSFE of the competing forecasting models by using a rolling window with a size of 48, i.e. four years of inflation data. The forecasting period covers from October 2005 to July 2014.

Also table 4 shows that linear models perform no worse than the advocated SB model. The linear AR(1) model becomes attractive by providing the smallest RMSFE most of the time. The SB model only provides the smallest RMSFE for two-step-ahead forecasts. However, the SB model provides the second smallest RMSFE for all other multi-step-ahead forecasts.

Nonetheless, we have to address that it is not striking that AR(1) only forecasts marginally better than the SB model. Intuitively, a small sample with less dynamics might be better modelled with a simple linear model.

In order to compare the forecasting accuracy of all linear forecasting models relative to the SB forecasting model, table 5 presents the Diebold-Mariano test statistics. The forecasting errors are obtained from fixed rolling window forecasting with sizes of 460 and 48, separately.

Table 5 shows that when a window size is big, i.e. 460, the SB model performs better than the linear AR(1), RW(1) and ARMA(1,1) models with negative Diebold-Mariano test statistics. However, the SB model's forecasting accuracy is only significantly better, at the 10 percent level, than the linear models with one-step-ahead forecasting. When a fixed rolling window with a size of 48 is applied, the linear AR(1) model in general performs marginally better than the SB model. Moreover, in no cases does the AR(1) predict significantly better than the SB model.

Using the MCMC draws, it is straightforward to simulate the one-step-ahead and multi-step-ahead predictive distributions. Because the calibrating sample changes as the window rolls, the simulated forecast distributions would, therefore,

Table 4: *Root mean squared forecast errors (RMSFE) calculated using the SB, RW, AR(1) and ARMA(1,1) forecasting models. The forecast horizon extends from one-step-ahead forecasts to 24-step-ahead forecasts. We applied a fixed rolling window approach with a window size of 48. The forecasting period covers from October 2005 to July 2014.*

Fix48	1-step	2-step	3-step	4-step	5-step	6-step	7-step	8-step	9-step	10-step	11-step	12-step
SB	0.3499	**0.5630**	0.7513	0.9026	1.0304	1.1412	1.2273	1.2956	1.3473	1.3873	1.4167	1.4476
RW	0.3495	0.5640	0.7522	0.9091	1.0405	1.1575	1.2500	1.3223	1.3832	1.4298	1.4684	1.4912
AR(1)	0.3851	0.5761	0.7466	0.8884	1.006	1.1107	1.1952	1.2608	1.3164	1.3573	1.3910	1.4107
ARMA(1,1)	0.3598	0.5639	0.7522	0.9090	1.0405	1.1575	1.2500	1.3223	1.3832	1.4298	1.4684	1.4912

	13-step	14-step	15-step	16-step	17-step	18-step	19-step	20-step	21-step	22-step	23-step	24-step
SB	1.4489	1.4548	1.4307	1.3961	1.3816	1.3876	1.3995	1.4226	1.4568	1.5076	1.5513	1.5850
RW	1.4792	1.4570	1.4177	1.3789	1.3484	1.3266	1.3277	1.3389	1.3644	1.4105	1.4517	1.5036
AR(1)	1.3999	1.3801	1.3462	1.3133	1.2906	1.2760	1.2805	1.2954	1.3221	1.3669	1.4073	1.4572
ARMA(1,1)	1.4792	1.4570	1.4177	1.3789	1.3484	1.3266	1.3277	1.3389	1.3644	1.4105	1.4517	1.5036

change over time. We can use fan charts to illustrate the time-varying forecasting distributions, as well as the time-varying forecasting uncertainty. Figures 6 – 9 plot the fan charts using the SB and AR(1) forecasting models. In figure 6, the forecast uncertainties simulated using a fixed rolling window are very similar to those in figure 7. This is probably because the fixed rolling window size covers 460 observations, which is large. Therefore, the estimated $\widehat{\sigma}_\varepsilon^2$ from a fixed rolling window method is not significantly different from the $\widehat{\sigma}_\varepsilon^2$ achieved from an expanding rolling window method. As a result, the forecasting distributions achieved from these two methods are similar. In particular, the training sample covers from January 1956 to April 1994, which contains a period of high inflation. Empirical results show that the level of inflation persistence fell during the 2000s, and there is less variation compared with the 1980s. These results suggest that inflation has been more anchored in the last two decades.

4 Conclusion

This study investigated the first order bilinear model with a single bilinear term using a Bayesian approach. Upon application to monthly U.K. inflation data from January 1956 to July 2014, we found that the SB model fits the underlying process of U.K. inflation well.

The high persistence in the underlying process varies over time, and the estimated time-varying autoregressive coefficients are below one most of the time. However, the coefficients are close to one (or exceed one) if the idiosyncratic shock in the previous period is large and positive. Ignoring the bilinear term b may lead to incorrect inferences, thereby inducing severely distorted forecasts. Motivated by Brunner and Hess's (1995) views regarding the potential problems in estimating the bilinear model with the MLE, we developed an efficient MCMC sampling algorithm to estimate the SB model where the moment conditions are met.

Overall, the SB model outperforms the RW, ARMA(1,1) and AR(1) models for both one-step-ahead and multi-step-ahead out-of-sample forecast when the training data contains rich dynamics, or covers a long period of time. In terms of predictive likelihood, the SB model is superior to the RW, ARMA(1,1) and AR(1) models for the majority of forecast observations (May 1994 – July 2014).

Table 5: *We present Diebold—Mariano forecast accuracy comparison tests of our SB model forecasts against the RW, AR(1) and ARMA(1,1). The null hypothesis is that the two forecasts have the same mean squared error. Negative values indicate superiority of our SB model forecasts, and bold denote significance relative to the asymptotic null distribution at the 10 percent level.*

Fix48	1-step	2-step	3-step	4-step	5-step	6-step	7-step	8-step	9-step	10-step	11-step	12-step
RW	0.0632	-0.0481	-0.0270	-0.1388	-0.1672	-0.2155	-0.2469	-0.2475	-0.2955	-0.3201	-0.3533	-0.2779
AR(1)	**-1.9161**	-0.4448	0.1382	0.4011	0.5876	0.5775	0.5054	0.4612	0.3436	0.2952	0.2206	0.2902
ARMA(1,1)	-0.8591	-0.0481	-0.0270	-0.1388	-0.1672	-0.2155	-0.2469	-0.2475	-0.2955	-0.3201	-0.3533	-0.2779
	13-step	14-step	15-step	16-step	17-step	18-step	19-step	20-step	21-step	22-step	23-step	24-step
RW	-0.1801	-0.0128	0.0754	0.0971	0.1915	0.3452	0.3896	0.4294	0.4555	0.4351	0.3966	0.3217
AR(1)	0.3491	0.5107	0.5519	0.5097	0.5687	0.6831	0.6929	0.7079	0.7276	0.6908	0.6371	0.5709
ARMA(1,1)	-0.1801	-0.0128	0.0754	0.0971	0.1915	0.3452	0.3896	0.4294	0.4555	0.4351	0.3966	0.3217

Fix460	1-step	2-step	3-step	4-step	5-step	6-step	7-step	8-step	9-step	10-step	11-step	12-step
RW	**-1.6925**	-1.2165	-1.0534	-0.9466	-0.8794	-0.8115	-0.7840	-0.7765	-0.8000	-0.8250	-0.8409	-0.8420
AR(1)	-1.4618	-0.8113	-0.6144	-0.5423	-0.5330	-0.5151	-0.5153	-0.5318	-0.5706	-0.6133	-0.6435	-0.6549
ARMA(1,1)	**-2.9093**	-1.2165	-1.0534	-0.9466	-0.8794	-0.8115	-0.7840	-0.7765	-0.8000	-0.8250	-0.8409	-0.8420
	13-step	14-step	15-step	16-step	17-step	18-step	19-step	20-step	21-step	22-step	23-step	24-step
RW	-0.8271	-0.8018	-0.7829	-0.7653	-0.7530	-0.7597	-0.8029	-0.8470	-0.8972	-0.9403	-0.9758	-0.9966
AR(1)	-0.6523	-0.9836	-0.9788	-0.9744	-0.9758	-1.0040	-1.0785	-1.1505	-1.2282	-1.2937	-1.3465	-1.3795
ARMA(1,1)	-0.8271	-0.8018	-0.7829	-0.7653	-0.7530	-0.7597	-0.8029	-0.8470	-0.8972	-0.9403	-0.9758	-0.9966

Despite the simple specification of the SB model, the SB model that allows for changing persistence in a series turns out to be an appealing model for inflation. Compared with the AR(1) and ARMA(1,1) models, the SB model predicts better during the period of and after the recent financial crisis, especially when a long period of historical data is considered for analysis.

Appendix

A   Derivation of the Full Posterior Conditionals

The likelihood function of the SB model can be derived recursively. The simplest SB model is specified as

$$y_t = (a + b\varepsilon_{t-1}) y_{t-1} + \varepsilon_t, \quad t = 2, \cdots, n,$$

where y_1 is the initial observation and $\varepsilon_1 = 0$. Denote $\mathbf{y} = (y_2, \cdots, y_n)'$ and $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)'$. With $n \geq 2$,

$$
\begin{aligned}
y_n &= a y_{n-1} + b y_{n-1} \varepsilon_{n-1} + \varepsilon_n, \\
y_{n-1} &= a y_{n-2} + b y_{n-2} \varepsilon_{n-2} + \varepsilon_{n-1}.
\end{aligned}
$$

Therefore,

$$\varepsilon_{n-1} = y_{n-1} - a y_{n-2} - b y_{n-2} \varepsilon_{n-2},$$

and also

$$\varepsilon_{n-2} = y_{n-2} - a y_{n-3} - b y_{n-3} \varepsilon_{n-3}.$$

If we take the substitute equations recursively,

$$
\begin{aligned}
y_n &= a y_{n-1} + b y_{n-1} \left(y_{n-1} - a y_{n-2} - b y_{n-2} \varepsilon_{n-2} \right) + \varepsilon_n \\
&= a y_{n-1} + b y_{n-1} y_{n-1} - a b y_{n-1} y_{n-2} - b^2 y_{n-1} y_{n-2} y_{n-2} + a b^2 y_{n-1} y_{n-2} y_{n-3} \\
&\quad + b^3 y_{n-1} y_{n-2} y_{n-3} \varepsilon_{n-3} + \varepsilon_n.
\end{aligned}
\tag{10}
$$

The component with ε_{n-3} can be further substituted. The last component with error term ε_2 is then

$$
\begin{aligned}
(-1)^{n-1} \left[b^{n-2} y_{n-1} \cdots y_2 \varepsilon_2 \right] &= (-1)^{n-1} \left[b^{n-2} y_{n-1} \cdots y_2 \left(y_2 - a y_1 - b \varepsilon_1 y_1 \right) \right] \\
&= (-1)^{n-1} b^{n-2} y_{n-1} \cdots y_2 y_2 + a \left(-b \right)^{n-2} y_{n-1} \cdots y_1 + (-1)^n b^{n-1} y_{n-1} \cdots y_1 \varepsilon_1 .
\end{aligned}
$$

Because we assume $\varepsilon_1 = 0$, equation (10) can be generalized for $t = 3, \cdots, n$ as

$$
y_t = a \sum_{i=1}^{t-1} \left[(-b)^{i-1} \prod_{j=1}^{i} y_{t-j} \right] + \sum_{i=1}^{t-2} \left[(-1)^{i+1} b^i y_{t-i} \prod_{j=1}^{i} y_{t-j} \right] + \varepsilon_t .
$$

If we denote

$$
f_1 (t, b) = \sum_{i=1}^{t-1} \left[(-b)^{i-1} \prod_{j=1}^{i} y_{t-j} \right] ,
$$

$$
f_2 (t, b) = \sum_{i=1}^{t-2} \left[(-1)^{i+1} b^i y_{t-i} \prod_{j=1}^{i} y_{t-j} \right] , \quad t \geq 3 ,
$$

equation (10) can be summarized as

$$
y_t = a f_1 (t, b) + f_2 (t, b) + \varepsilon_t ,
$$

where $f_2 (t, b) = 0$ with $t = 2$.

Recall equation (8), the joint prior is

$$
p (a, b, h_\varepsilon) = p (b \mid h_\varepsilon, a) \, p (h_\varepsilon \mid a) \, p (a) \, I_{\left[a^2 + b^2 h_\varepsilon^{-1} < 1 \right]} (a, b, h_\varepsilon) \, I_{[M]} (a, b, h_\varepsilon) ,
$$

where M is the region in which the inequalities (4) and (5) are satisfied.

To meet the stationarity condition in equation (3), a is restricted as $|a| < 1$ and the density of elicited prior for a is

$$
p(a) \propto \frac{1}{(2\pi \underline{V}_a)^{\frac{1}{2}}} \exp \left\{ -\frac{\left(a - \underline{\mu}_a \right)^2}{2 \underline{V}_a} \right\} I_{[|a|<1]} (a) , \tag{11}
$$

where $\underline{\mu}_a$ is the prior mean and $\underline{V}_a$ is the prior variance. Because equation (11) does not integrate up to 1, it is preferable to have a prior that allows $p(a)$ to integrate to 1 within the truncated region for model comparison reasons. Hence, a

truncated prior of a can be obtained by dividing the unnormalized prior density by a normalizing constant $\Pr\left(1 \,||a| < 1\right)$. Therefore, the prior of a is the following:

$$p(a) = \frac{1}{\Pr\left(1 \,||a| < 1\right)\left(2\pi \underline{V}_a\right)^{\frac{1}{2}}} \exp\left\{-\frac{\left(a - \underline{\mu}_a\right)^2}{2\underline{V}_a}\right\} I_{[|a| < 1]}(a), \tag{12}$$

where the normalizing constant $\Pr\left(1 \,||a| < 1\right)$ can be calculated as

$$\Pr\left(1 \,||a| < 1\right) = \Phi_N\left(\frac{1 - \underline{\mu}_a}{\sqrt{\underline{V}_a}}\right) - \Phi_N\left(\frac{-1 - \underline{\mu}_a}{\sqrt{\underline{V}_a}}\right).$$

Φ_N is used to indicate the c.d.f of a normal distribution.

The joint prior of h_ε and b conditional on a can be expressed as

$$p\left(h_\varepsilon, b \mid a\right) \propto \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon} \Gamma\left(\underline{\alpha}_\varepsilon\right)} h_\varepsilon^{\underline{\alpha}_\varepsilon - 1} \exp\left(-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}\right) I_{\left[a^2 + b^2 h_\varepsilon^{-1} < 1\right]}\left(a, b, h_\varepsilon\right). \tag{13}$$

Note that h_ε and b are both random variables, and therefore the normalizing constant in equation (13), that we omit for simplicity, does not depend on either h_ε or b. The conditional prior for $p\left(h_\varepsilon \mid a\right)$, then, can be obtained by marginalizing over b in equation (13) as follows:

$$p\left(h_\varepsilon \mid a\right) \propto \int p\left(h_\varepsilon, b \mid a\right) db.$$

To meet the moment conditions, b is restricted as $|b| < \sqrt{\left(1 - a^2\right) h_\varepsilon}$ in equation(2). The conditional prior of h_ε, $p\left(h_\varepsilon \mid a\right)$, can then be obtained as the following:

$$\begin{aligned}
p\left(h_\varepsilon \mid a\right) &\propto \int p\left(h_\varepsilon, b \mid a\right) I_{\left[|b| < \sqrt{\left(1 - a^2\right) h_\varepsilon}\right]}\left(a, b, h_\varepsilon\right) db \\
&\propto p\left(h_\varepsilon, b \mid a\right) \int_{-\sqrt{\left(1 - a^2\right) h_\varepsilon}}^{\sqrt{\left(1 - a^2\right) h_\varepsilon}} db \\
&\propto \frac{1}{\underline{\beta}_\varepsilon^{\underline{\alpha}_\varepsilon} \Gamma\left(\underline{\alpha}_\varepsilon\right)} h_\varepsilon^{\underline{\alpha}_\varepsilon - 1} \exp\left(-\frac{h_\varepsilon}{\underline{\beta}_\varepsilon}\right) \cdot 2\sqrt{\left(1 - a^2\right) h_\varepsilon}.
\end{aligned} \tag{14}$$

Then, a conditional prior $p\left(b \mid h_\varepsilon, a\right)$ for b is

$$p\left(b \mid h_\varepsilon, a\right) \propto \frac{1}{2\sqrt{\left(1 - a^2\right) h_\varepsilon}} I_{\left[|b| < \sqrt{\left(1 - a^2\right) h_\varepsilon}\right]}\left(a, b, h_\varepsilon\right) I_{[M]}\left(a, b, h_\varepsilon\right). \tag{15}$$

The full posterior conditionals can be derived by combining the elicited prior in equation (12, 14, and 15) together with the derived likelihood in equation (7). Therefore, the posterior conditional of a is $a \sim N\left(\overline{\mu}_a, \overline{V}_a\right) I_{\left[|a| < \sqrt{1 - b^2 h_\varepsilon^{-1}}\right]}(a, b, h_\varepsilon)$, a truncated normal distribution with variance $\overline{V}_a$:

$$\overline{V}_a = \left[\frac{1}{\sigma_\varepsilon^2} \sum_{t=2}^N \left[f_1\left(t, b\right)\right]^2 + \frac{1}{\underline{V}_a}\right]^{-1},$$

and mean $\overline{\mu}_a$:

$$\overline{\mu}_a = \overline{V}_a \cdot \left[h_\varepsilon \sum_{t=2}^N f_1\left(t, b\right)\left[y_t - f_2\left(t, b\right)\right] + \frac{\underline{\mu}_a}{\underline{V}_a}\right]. \tag{16}$$

The posterior conditional for h_ε follows a truncated gamma distribution $h_\varepsilon \sim \Gamma\varepsilon\left(\overline{\alpha}_\varepsilon, \overline{\beta}_\varepsilon\right) I_{[h_\varepsilon > b^2/(1-a^2)]}(a, b, h_\varepsilon)$, where

$$\overline{\alpha}_\varepsilon = \underline{\alpha}_\varepsilon + \frac{1}{2}\left(N - 1\right),$$

and

$$\overline{\beta}_\varepsilon = \left[\frac{1}{\underline{\beta}_\varepsilon} + \frac{1}{2}\sum_{t=2}^N \left[y_t - a f_1\left(t, b\right) - f_2(t, b)\right]^2\right]^{-1}.$$

If the value of $b^2/\left(1 - a^2\right)$ is extremely large, h_ε has to be sampled from the upper tail of a gamma distribution. For simplicity, we can approximate the tail with an exponential distribution,

$$f_{\exp}\left(x\right) = \lambda \exp\left(-x\lambda\right),$$

where $\lambda = b^2/\left(1 - a^2\right)$. The inverse $c.d.f$ algorithm together with an exponential rejection algorithm can be applied to generate random draws of h_ε.

The value of b depends on the values of a and h_ε. The posterior conditional of b is

$$p\left(b|y, a, h_\varepsilon\right) \propto \exp\left\{-\frac{h_\varepsilon}{2}\sum_{t=2}^N \left[y_t - a f_1\left(t, b\right) - f_2(t, b)\right]^2\right\}$$
$$I_{\left[|b| < \sqrt{(1-a^2)h_\varepsilon}\right]}(a, b, h_\varepsilon) I_{[M]}(a, b, h_\varepsilon).$$

Because b does not follow a standard distribution from which we can generate random samples, the Griddy–Gibbs sampling method proposed in Ritter and Tanner (1992) is used to get random draws of b. Based on the above posterior conditionals of a, b and h_ε, a Gibbs sampling incorporated with a Griddy–Gibbs sampling algorithm can be carried out in a straightforward way to simulate the posterior distributions of a, b and h_ε.

References

Atkeson, A., Ohanian, L. E., 2001. Are phillips curves useful for forecasting inflation. Federal Reserve Bank of Minneapolis Quarterly Review 25 (1), 2 – 11.

Barnett, A., Mumtaz, H., Theodoridis, K., 2014. Forecasting uk gdp growth and inflation under structural change. a comparison of models with time varying parameters. International Journal of Forecasting 30 (1), 129 – 143.

Bibi, A., Lessak, R., 2009. On stationarity and -mixing of periodic bilinear processes. Statistics & Probability Letters 79 (1), 79 – 87.

Brunner, A. D., Hess, G. D., 1995. Potential problems in estimating bilinear time-series models. Journal of Economic Dynamics and Control 19 (4), 663 – 681.

Byers, J. D., Peel, D. A., 1995. Bilinear quadratic arch and volatility spillovers in inter-war exchange rates. Applied Economics Letters 2 (7), 215 – 219.

Canova, F., 2 2007. G-7 inflation forecasts: Random walk, phillips curve or what else? Macroeconomic Dynamics 11, 1 – 30.

Charemza, W. W., Lifshits, M., Makarova, S., 2005. Conditional testing for unit-root bilinearity in financial time series: some theoretical and empirical results. Journal of Economic Dynamics and Control 29 (1 - 2), 63 – 96.

Chen, C. W., 1992a. Bayesian analysis of bilinear time series models : a gibbs sampling approach. Communications in Statistics - Theory and Methods 21 (12), 3407 – 3425.

Chen, C. W. S., 1992b. Bayesian inferences and forecasting in bilinear time series models. Communications in Statistics - Theory and Methods 21 (6), 1725 – 1743.

Feng, H., Peng, L., Zhu, F., 2013. Interval estimation for a simple bilinear model. Statistics & Probability Letters 83 (10), 2152 – 2159.

Gelfand, A. E., Dey, D. K., 1994. Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society. Series B (Methodological), 501 – 514.

Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface, Alexandria, VA: American Statistical Association, 571 – 578.

Geweke, J., 2001. Bayesian econometrics and forecasting. Journal of Econometrics 100 (1), 11 – 15.

Geweke, J., Amisano, G., 2010. Comparing and evaluating bayesian predictive distributions of asset returns. International Journal of Forecasting 26 (2), 216 – 230.

Granger, C., Andersen, A., 1978a. An introduction to bilinear time series models. Banderhoeck and Ruprecht, Gottingen.

Granger, C., Andersen, A., 1978b. On the invertibility of time series models. Stochastic Processes and their Applications 8 (1), 87 – 92.

Hristova, D., 2005. Maximum likelihood estimation of a unit root bilinear model with an application to prices. Studies in Nonlinear Dynamics and Econometrics 9, 1199.

Kim, W. K., Billard, L., Basawa, I. V., 1990. Estimation for the first-order diagonal bilinear time series model. Journal of Time Series Analysis 11 (3), 215 – 229.

Maheu, J. M., Song, Y., 2014. A new structural break model, with an application to canadian inflation forecasting. International Journal of Forecasting 30 (1), 144 – 160.

Meller, B., Nautz, D., 2012. Inflation persistence in the euro area before and after the european monetary union. Economic Modelling 29 (4), 1170 – 1176.

O'Reilly, G., Whelan, K., Nov. 2005. Has euro-area inflation persistence changed over time? Review of Economics and Statistics 87 (4), 709 – 720.

Osborn, D. R., Sensier, M., 2009. Uk inflation: Persistence, seasonality and monetary policy. Scottish Journal of Political Economy 56 (1), 24 – 44.

Raeburn, E., Lane, J., Peel, D., 1995. Some empirical evidence on the time series properties of four uk asset prices. Economica 63 (251), 405 – 426.

Ritter, C., Tanner, M. A., 1992. Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler. Journal of the American Statistical Association 87 (419), 861 – 868.

Sesay, S. A. O., Subba Rao, T., 1988. Yule-walker type difference equations for higher-order moments and cumulants for bilinear time series models. Journal of Time Series Analysis 9 (4), 385 – 401.

Stock, J. H., Watson, M. W., 1999. Forecasting inflation. Journal of Monetary Economics 44 (2), 293 – 335.

Subba Rao, T., Jan. 1981. On the theory of bilinear time series models. Journal of the Royal Statistical Society. Series B (Methodological) 43 (2), 244 – 255.

Watson, M. W., 2014. Inflation persistence, the nairu, and the great recession. American Economic Review 104 (5), 31 – 36.