FORECASTING MACROECONOMIC TIME SERIES WITH LOCALLY ADAPTIVE SIGNAL EXTRACTION

PAOLO GIORDANI AND MATTIAS VILLANI

Abstract. We introduce a non-Gaussian dynamic mixture model for macroeconomic forecasting. The Locally Adaptive Signal Extraction and Regression (*LASER*) model is designed to capture relatively persistent AR processes (signal) contaminated by high frequency noise. The distribution of the innovations in both noise and signal is robustly modeled using mixtures of normals. The mean of the process and the variances of the signal and noise are allowed to shift suddenly or gradually at unknown locations and number of times. The model is then capable of capturing movements in the mean and conditional variance of a series as well as in the signal-to-noise ratio. Four versions of the model are estimated by Bayesian methods and used to forecast a total of nine quarterly macroeconomic series from the US, Sweden and Australia. We observe that allowing for infrequent and large parameter shifts while imposing normal iid errors often leads to erratic forecasts, but that the model typically forecasts well if robustified by allowing for non-normal errors and time varying variances. Our main finding is that, for the nine series we analyze, specifications with infrequent and large shifts in error variances outperform both fixed parameter specifications as well as smooth, continuous shifts when it comes to interval coverage.

Keywords: Bayesian inference, Forecast evaluation, Regime switching, State-space modeling, Dynamic Mixture models.

1. Introduction

This paper is concerned with the forecasting performance for macroeconomic time series of a class of dynamic mixture models. The widespread instability of coefficients in standard ARMA models for these types of data series has been widely documented (see for example Stock and Watson, 1996). Multiple shifts in local means, error variances and autocorrelation structure in inflation, interest rates and other nominal time series are detected by various frequentist and Bayesian procedures on the last four decades of data (Levin and Piger 2003, Stock and Watson 2006, Koop and Potter 2008, Giordani and Kohn, 2008). According to Clements and Hendry (1999), such shifts are the main cause of forecasting failure of univariate and multivariate linear models.

A variety of models has been formulated to tackle various forms of non-Gaussian behavior. Some models are used successfully, but one cannot but be surprised by the overall difficulty

Giordani: *Research department, Sveriges Riksbank, SE-103 37 Stockholm, Sweden. E-mail: paolo.giordani@riksbank.se.* Villani: *Research department, Sveriges Riksbank, SE-103 37 Stockholm, Sweden* and *Department of Statistics, Stockholm University.* The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

in outperforming standard AR processes even when they exhibit large parameter instabilities (Stock and Watson, 1996). Some simple models have withstood the test of time and are largely adopted by practitioners and academics. Some of these simple models are cast in state space form, others are estimated by least squares, but all essentially involve some exponential discounting of past observations (discounted least squares, exponential smoothing) and/or over-differencing (ARIMA, local trends models).[1 2] Marcellino (2008) recently reports good forecasting performance for time varying parameter models (TVP) for macroeconomic series, while Stock and Watson (1996) report more ambiguous results. Markov switching models (Hamilton, 1989) and the closely associated multiple change-point models (e.g. Chib, 1998) are very popular for off-line analysis, but little has been published on the forecasting performance of the former and, as far as we are aware, nothing on the latter. The few available studies show disappointing results for Markov switching models (Clements and Krolzig (1998) and Bessec and Bouabdallah, 2005), at least for point forecasts.

We summarize the discussion above as follows. Even though in sample analysis indicates that parameter instability in AR(MA) models is widespread in macroeconomic time series, fixed parameter specifications are competitive with simple models that assume continuous and smooth time variation and superior to complex Markov switching models when it comes to forecasting.

The goal of this paper is to shed some light on these seemingly conflicting observations. Rather than forecasting a large number of series, we choose to provide a more detailed analysis of three macroeconomic series of particular interest (real GDP growth, CPI inflation and a short interest rate) using quarterly US, Swedish and Australian data (for a total of nine series). Our tool of analysis is a Bayesian dynamic mixture model recently developed for forecasting at the Swedish Central Bank. The model, denoted *LASER* (Locally Adaptive Signal Extraction and Regression), allows for a variety of generalizations of standard ARMA models that can account for Gaussian or non-Gaussian shifts in local mean, error variance and persistence, as well as for non-Gaussian innovations. By switching on and off various features of the model and monitoring the real-time forecasts we can therefore try to understand which features are responsible for good and poor performance.

From a technical perspective, our innovation is to expand on the work of Giordani and Kohn (2008) (henceforth GK) and introduce a more general extension of the ARMA class than currently available in the literature. The Markov Chain Monte Carlo (MCMC) technology of GK is used to achieve fast and efficient Bayesian inference, which allows us to perform the first (to the best of our knowledge) serious forecasting evaluation of change-point and mixture innovation models. Based on this forecasting evaluation, our main conclusions are as follows. First, it is much easier to outperform fixed coefficient models when considering interval coverage rather than point forecasts. Second, infrequent and large shifts in error

[1]We define over-differencing informally as the differencing of a series that cannot reasonably be considered unbounded, like the real interest rate or the consumption over income ratio.

[2]Harvey (1989) and West and Harrison (1997) provide detailed expositions of many such models from a frequentist and Bayesian perspective respectively.

variance provide better conditional interval coverage than continuous smooth shifts. Third, models that allow infrequent and large shifts in conditional mean but normal independent identically distributed (iid) innovations are very fragile to the presence of outliers and of shifts in error variance, whereas they perform well when the Gaussian iid assumption is removed.

Our interpretation of the first two results is that shifts in error variance are large and persistent in our series, and therefore easy to detect and model with a change-point approach. The intuition for the third results is that when normality is imposed on the errors, any outlier (or increased variance) will be interpreted as a parameter shift in real time, generating excessively volatile forecasts.

Section 2 presents *LASER* in rather general terms, and discusses some options to model shifts in local mean and variances. Section 3 specifies four models that are nested in the *LASER* framework and differ in the specification of the error structure and of time variation. Section 4 shows how two of these models imply time variation in the persistence of the process, not only in its mean and variance. Section 5 presents the forecasting experiment and discusses results, focusing on point forecasts first and then on interval coverage. Section 6 concludes.

2. Locally adaptive signal extraction and regression

LASER (Locally Adaptive Signal Extraction and Regression) is a state space model for filtering and forecasting recently developed at the Swedish Central Bank and employing the approach to shifts in conditional mean and variance proposed by Giordani and Kohn (2008). The univariate random variable y_t is decomposed into three processes[3], all of which can have mixture of normals (MN) innovations:

(1) A local mean μ_t, which can be any conditionally Gaussian process (i.e. Gaussian conditional on a vector of latent indicators and parameters).
(2) A latent, unobserved, process x_t, modeled as a finite-order, stationary, AR model with (i) MN innovations in the log variance process (ii) MN errors (iii) unknown lag length p.
(3) An independently distributed measurement error/additive outlier process with (i) MN innovations in the log variance process (ii) MN errors.

We can write the observation equation and the transition equation for x_t as

$$y_t = \mu_t + x_t + \epsilon_t \tag{2.1}$$

$$x_t = \rho_1 x_{t-1} + ... + \rho_p x_{t-p} + u_t \tag{2.2}$$

$$\epsilon_t \sim MoN(k_y, \boldsymbol{\pi}_y, \boldsymbol{\alpha}_y, \sigma^2_{y,t}\boldsymbol{\xi}^2_y) \tag{2.3}$$

$$u_t \sim MoN(k_x, \boldsymbol{\pi}_x, \boldsymbol{\alpha}_x, \sigma^2_{x,t}\boldsymbol{\xi}^2_x). \tag{2.4}$$

[3]In fact there is a fourth component, a regression effect (also time-varying) that we omit in our discussion since the paper focuses on univariate forecasting.

Equations (2.3) and (2.4) are to be read as follows: ϵ_t has a mixture of normals (MN) distribution with k_y components, and parameters specified by a k_y vector of probabilities $\boldsymbol{\pi}_y$, a k_y vector or means $\boldsymbol{\alpha}_y$,and a k_y vector of variances $\sigma^2_{y,t}\boldsymbol{\xi}^2_y$, where $\sigma^2_{y,t}$ is a (possibly time-varying) scalar. For identification purposes, we set $[\boldsymbol{\alpha}_y]_1 = [\boldsymbol{\alpha}_x]_1 = 0$ and $\left[\boldsymbol{\xi}_y\right]_1 = [\boldsymbol{\xi}_x]_1 = 1$, where $[\mathbf{a}]_i$ denotes the ith element of a vector $\mathbf{a}$. The autoregressive parameters $\rho_1, ..., \rho_p$ are constrained to lie in the stationary region. The lag length is unknown and we compute its posterior by adding an updating step in the MCMC algorithm, with the user specifying the maximum number of lags and the prior lag probabilities.[4] At this level the model is still very general and requires several choices to be made operational. We now discuss some options for the local mean process μ_t.

Modelling shifts in the local mean $\boldsymbol{\mu_t}$. We refer to μ_t as the "local mean" because we typically assume that μ_t changes infrequently. One possible specification is the random walk with a two component mixture distribution for the errors, and one component degenerate:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \sigma_{\mu,1} u_{\mu,t} \text{ with prob. } \pi_1 \\ &= \mu_{t-1} \text{ with prob. } 1 - \pi_1, \end{aligned} \tag{2.5}$$

where $u_{\mu,t}$ is $NID(0, 1)$ and π_1 is the probability of a shift. In this case y_t is globally non-stationary, though it may behave as a stationary series for long stretches. For $\pi_1 = 1$ the innovations are normal as in the well-known local level model. An attractive alternative when prior information on the long-run mean of the series $\overline{\mu}$ is available or when the sample is large is the globally stationary specification

$$\begin{aligned} \mu_t &= \overline{\mu} + \sigma_{\mu,2} u_{\mu,t} \text{ with prob. } \pi_1 \\ &= \mu_{t-1} \text{ with prob. } 1 - \pi_1. \end{aligned} \tag{2.6}$$

It is also possible to allow for both types of shifts:

$$\mu_t = \mu_{t-1} + \sigma_{\mu,1} u_{\mu,t} \text{ with prob. } \pi_1 \tag{2.7}$$

$$= \overline{\mu} + \sigma_{\mu,2} u_{\mu,t} \text{ with prob. } \pi_2 \tag{2.8}$$

$$= \mu_{t-1} \text{ with prob. } 1 - \pi_1 - \pi_2.$$

Finally, when the shifts are infrequent but possibly large, all these specifications are unappealing for most macroeconomic series in that they suggest an immediate jump of y_t to the new local mean. Since we believe that large shifts (e.g. from high to low inflation) typically take place over the course of several quarters, we generalize the specification in (2.7) as follows:

[4]Our prior assumes that if $\rho_i \neq 0$, then also $\rho_j \neq 0$ if $i > j$. This assumption could be easily relaxed.

(2.9)
$$\begin{aligned}\Delta\mu_t &= (1-\rho_\mu)(\widetilde{\mu}_t - \mu_{t-1})\\ \widetilde{\mu}_t &= \mu_{t-1} + \sigma_{\mu,1}u_{\mu,t} \text{ with prob. } \pi_1\\ \widetilde{\mu}_t &= \overline{\mu} + \sigma_{\mu,2}u_{\mu,t} \text{ with prob. } \pi_2\\ \widetilde{\mu}_t &= \widetilde{\mu}_{t-1} \text{ with prob. } 1-\pi_1-\pi_2,\end{aligned}$$

where ρ_μ determines how gradual the transition is and $\rho_\mu = 0$ retrieves (2.7). Here $\widetilde{\mu}_t$ jumps and μ_t moves gradually to $\widetilde{\mu}_t$.

Modelling shifts in variances. We model shifts in log variances as random walks with a two component mixture distribution for the errors, and one component degenerate:

(2.10)
$$\begin{aligned}\ln\sigma_{y,t} &= \ln\sigma_{y,t-1} + \delta_y e_{y,t} \text{ with prob. } \pi_y^\sigma\\ \ln\sigma_{y,t} &= \ln\sigma_{y,t-1} \text{ with prob. } 1-\pi_y^\sigma,\end{aligned}$$

and

(2.11)
$$\begin{aligned}\ln\sigma_{x,t} &= \ln\sigma_{x,t-1} + \delta_x e_{x,t} \text{ with prob. } \pi_x^\sigma\\ \ln\sigma_{x,t} &= \ln\sigma_{x,t-1} \text{ with prob. } 1-\pi_x^\sigma,\end{aligned}$$

with $e_{y,t}$ and $e_{x,t}$ both $NID(0,1)$. This formulation allows for infrequent, large shifts as well as for continuos, small shifts ($\pi_y^\sigma = \pi_x^\sigma = 1$).

3. Four forecasting models

3.1. **Models.** If $\mu_t, \sigma_{x,t}$ and $\sigma_{y,t}$ are constant and all innovations normal, *LASER* simplifies to the state space representation of an ARMA(p,p) process. We wish to understand which additional features of *LASER* can be expected to contribute to forecasting accuracy (both in terms of point forecasts and of interval coverage). For this purpose we will compare the forecasting performance for several versions of *LASER*[5]. These versions will effectively differ only in the prior and not in the way inference is performed (which is by MCMC with lag selection and stationarity imposed in all cases; see the Appendix for a description of the MCMC scheme). The four models can be broadly characterized as follows:

(1) ARMA. $\mu_t, \sigma_{x,t}$ and $\sigma_{y,t}$ constant and all innovations normal.
(2) Shifts. Infrequent shifts in μ_t, constant $\sigma_{x,t}$ and $\sigma_{y,t}$ and normal innovations.
(3) Robust TVP. Normal innovations in μ_t, $\ln(\sigma_{x,t})$ and $\ln(\sigma_{y,t})$ and MN innovations elsewhere.

[5]It would be interesting to also evaluate the forecasting performance of a Bayesian model average of these four models. Computing marginal likelihoods for dynamic mixture models is a very difficult and time consuming endeavour, see e.g. Fruewirth-Schnatter (2006). Moreover, we would have to compute marginal likelihoods in every time period of the evaluation sample. Since our main motivation is to compare the four models to better understand their differences in a forecasting environment, we will not compute a Bayesian model average of the models.

(4) Robust Shifts. Infrequent shifts in μ_t, $\ln(\sigma_{x,t})$ and $\ln(\sigma_{y,t})$ and MN innovations elsewhere.

The exact priors for each model are given in the next section. The ARMA model acts as a benchmark. The Shifts model has shifts in mean but normal iid errors. The Robust TVP specification is meant, by comparison with the Robust Shifts specification, to evaluate the relative merits of frequent, small shifts versus infrequent, larger shifts.

3.2. **Priors.** This section presents the priors used in this paper for blocks of parameters. Unless otherwise specified, the priors are common to all four models.

Priors for $\rho_1, ..., \rho_p, \boldsymbol{\alpha}_y, \boldsymbol{\xi}_y, \sigma_{y,0}, \boldsymbol{\alpha}_x, \boldsymbol{\xi}_x, \sigma_{x,0}$. We assume the following probabilities for lag lengths p from 1 to 4 (longer lags have zero probability): $prob(p = 1, p = 2, p = 3, p = 4) =$ (0.4, 0.3, 0.2, 0.1).

$$\rho_1, ..., \rho_p | p \sim N\left(\begin{bmatrix} \underline{\rho} \\ 0 \\ ... \\ 0 \end{bmatrix}, \underline{\sigma}_\rho^2 \begin{bmatrix} 1 & \underline{\rho} & \underline{\rho}^2 & \underline{\rho}^3 \\ & 1 & \underline{\rho} & \underline{\rho}^2 \\ & & 1 & \underline{\rho} \\ & & & 1 \end{bmatrix}\right) \text{ if } \rho_1, ..., \rho_p \text{ lie in the stationary region,}$$

$p(\rho_1, ..., \rho_p) = 0$ otherwise. The prior attempts to be coherent in the sense that the correlation structure of the prior covariance matrix is what would be observed in a large sample generated with parameters fixed at the prior mean. In our applications on quarterly data we let the prior be somewhat informative at $\underline{\rho} = 0.8$ and $\underline{\sigma}_\rho = 0.2$.

The priors on $\ln(\sigma_{y,0})$ and $\ln(\sigma_{x,0})$ are normal, centered on the residual OLS variance from an AR(4) but very disperse. In Robust TVP and Robust Shifts we add a second component to both distributions, with a somewhat informative prior centered on a scale mixture of normals (i.e. symmetric leptokurtic) specification:

$$\begin{aligned} [\boldsymbol{\alpha}_y]_2 &\sim N(0, 0.5\widehat{\sigma}^2) \\ [\boldsymbol{\xi}_y]_2 &\sim IG(20 \times 3^2, 20), \end{aligned}$$

where the inverse gamma prior on ξ_y^2 can be interpreted as twenty prior observation with variance 9, and $\widehat{\sigma} = 1.42 \cdot median|e|$, with e the OLS residuals from an AR(4) regression, is an outlier-robust measure of the standard deviation of the errors. The prior probability of the second component is beta, with prior probability 0.2 and prior sample size 10 (see Gelman et al. 2004). The prior for $\boldsymbol{\alpha}_x, \boldsymbol{\xi}_x^2, \boldsymbol{\pi}_x$ is identical. This prior formalizes the idea that we are quite confident that the distribution is leptokurtic (hence the informative prior on the variance of the second component), but less confident of its precise shape (hence the weakly informative prior on the mean and probability of the second component).

Priors for $\rho_\mu, \pi_1, \sigma_{\mu,1}, \mu_0$. The process for μ_t is described in equation (2.9). In Shifts and Robust Shifts, we fix ρ_μ to 0.8 as we don't expect to have enough mean shifts to estimate this parameter accurately. In Robust TVP we of course set $\rho_\mu = 0$. In this paper we fix $\pi_2 = 0$ since we do not want to run the risk of using hindsight in specifying a long-run mean for all variables. However, in actual applications we would use domain knowledge if available (e.g. an inflation target). In Robust TVP we set $\pi_1 = 1$. For Shifts and Robust Shifts we fix π_1 to 0.02, which corresponds to an average interval between shifts of around 12 years (the distribution of the number of shifts is multinomial with parameters π_1 and n, the sample size). The corresponding prior distribution for the number of shifts in each regime is binomial, with most mass in the 0-8 interval in the period 1959-2007 and probability of a shift equally distributed across all time periods (see Koop and Potter (2007) for a discussion).

The prior for μ_0 and $\widetilde{\mu}_0$ is bivariate normal, centered on the sample mean, with high variance but near-one correlation. This effectively assumes that $\mu_0 = \widetilde{\mu}_0$.

The prior for $\sigma^2_{\mu,1}$ is inverse gamma

$$\xi_y^2 \sim IG(10 \times \lambda^2 \widehat{\sigma}_y^2, 10),$$

where $\widehat{\sigma}_y^2$ is the sample variance of $y_t^s = 0.8y_{t-1}^s + 0.2y_t$. We smooth y_t to construct a data-driven prior to reduce the impact of outliers and high frequency noise. In actual applications we would use available domain knowledge of the series to construct a prior on the likely size of the shifts.

We notice that given a prior that a sample of a few decades will most likely contain at most a few shifts, it is important to use informative priors on their variance. As the prior sample size (here 10) goes to infinity, the standard deviation of a shift concentrates on $\lambda\widehat{\sigma}_y$. We set $\lambda = 0.25$ in Robust Shifts and Shifts and $\lambda = 0.25\sqrt{0.02}$ in Robust TVP, which gives approximately the same variance for $(\mu_t - \mu_{t-1})$. λ is fixed at zero for ARMA.

Priors for $\pi_y, \pi_x, \delta_y, \delta_x$. For δ_y and δ_x we use the same inverse gamma prior:

$$\delta_y^2 \sim IG(10 \times \lambda_y^2, 10).$$

For Robust Shifts, we set $\lambda_y = \lambda_x = 0.7$ and we fix $\pi_y = \pi_x = 0.02$. For interpretation, notice that $\delta = 0.7$ means that a plus (minus) one standard deviation shock to $\log(\sigma_{y,t})$ increases (decreases) $\sigma_{y,t}$ by 50% (25%). For Robust TVP $\pi_y = \pi_x = 1$ and $\lambda_y = \lambda_x = 0.7\sqrt{0.02}$.

4. Modeling time variation in persistence

Even though it would be technically possible to let the AR parameters $\rho_1, ..., \rho_p$ be time varying, *LASER* treats them as constant. In models with constant $\sigma_{y,t}$ and $\sigma_{x,t}$, such as Shifts and ARMA, this does imply that the local autocorrelation properties of the series y_t are also constant. However, shifts in the local mean y_t can give rise to a slowly decaying autocorrelation function (see Granger and Hyung, 1999), and MN innovations can also affect the persistence properties of y_t. If, on the other hand, we let $\sigma_{y,t}$ and $\sigma_{x,t}$ be time varying, the local autocorrelation structure can change substantially. For example, a drop in $\sigma_{x,t}$ effectively means that the persistent component accounts for a smaller share of the variance of y_t and hence all autocorrelations decrease. This approach to modeling time variation in local persistence is particularly parsimonious (compared to letting $\rho_1, ..., \rho_p$ all be time varying) in the case of a long lag length p. To make a simple example, consider the following version of *LASER* with fixed μ and normal iid errors:

$$\begin{aligned} y_t &= \mu + x_t + \sigma_{y,t}\epsilon_t \\ x_t &= \rho x_{t-1} + \sigma_{x,t}u_t, \end{aligned} \tag{4.1}$$

and shifts in $\sigma_{y,t}$ and $\sigma_{x,t}$ modeled as in Robust Shifts.

Let us define the local autocorrelation at lag i and time t as the correlation between y_t and y_{t+i} obtained assuming (as a heuristic approximation) that σ_y and σ_x are constant between time t and time $t+i$. This takes the simple form

$$corr_t(y_t, y_{t+i}) = \rho^i \frac{\sigma_{x,t}^2/(1-\rho^2)}{\sigma_{x,t}^2/(1-\rho^2) + \sigma_{y,t}^2},$$

from which it is clear that variations in $\sigma_{y,t}$ and $\sigma_{x,t}$ will affect local autocorrelations. As an illustration, Figure 1 reports the first-order local autocorrelation for quarterly US inflation over time (see Section 5 for a description of the data) from the Robust Shifts model with one lag. The parameters are set to their posterior means based on the full sample from 1959q2 to 2006q4. It is evident that the local persistence of the process has dropped dramatically.

5. Data and results

We use three widely monitored series of US data and the three corresponding Australian and Swedish series: (i) real GDP growth (ii) CPI inflation (iii) three month treasury bill. The data cover the period 1980q2-2006q4 for the US, and 1980q2-2007q4 for Sweden and Australia.[6] For all counties, we start forecasting with forty observations and then update model parameters and forecasts one observation at a time. All forecasts are out of sample in the sense that priors are formulated based on recursive estimation of all model parameters, but we do not consider publishing delays nor (for GDP growth) data revision. A new MCMC (see the Appendix for details) is run as each observation is added to the sample. Given the large number of MCMC run, we do not assess convergence and mixing for each run. Based on our experience, these are usually adequate, particularly when the object of interest is the forecast distribution rather than the model parameters and when informative priors are used (as in this paper).[7]

We evaluate one and four quarter ahead forecasts using the RMSFE loss function and correspondingly taking the mean of the forecast distribution as our point forecast. We note that

[6] US data are from the Fred database (http://research.stlouisfed.org/fred2/). GDP growth is the annualized log growth rate of real DGP (GDP251). Inflation is the log growth rate of CPI for all urban consumers, aggregated from monthly (CPIAUCSL). The three month interest rate is a quarterly average of daily values (FYGM3). Swedish data are obtained using the same transformations as for US data. The Ecowin database codes for the original series are swe01850, swe11899, swe14010. Australian data are from the Reserve Bank of Australia website. The codes for the original series are GGDPCVRGDI, GCPIAGQP, and FIRMMBAB90.

[7] An analysis of mixing for similar models is given in Giordani and Kohn (2008). They also notice that the use of informative priors is in some case important in ensuring both good forecasting performance and good mixing of the chain.

the forecast distribution averages over parameter uncertainty and, in more complex models, also over uncertainty over latent indicators for the mixture of normals used to model the error distributions as well as uncertainty over the number, timing and size of shifts in mean and log variances.

5.1. **Point forecasts.** The RMSFE statistics are summarized in Table 1. We highlight the following results:

(1) Robust Shifts performs better than Shifts on average.
(2) Robust Shifts performs substantially better than Robust TVP one quarter ahead and about as well four quarters ahead on average.
(3) Robust Shifts outperforms Robust TVP for interest rate forecasting in all three countries and forecast horizons, often by a large margin.
(4) On average Robust Shifts performs slightly better than ARMA on US data and substantially better on Swedish data.
(5) All models largely outperform a random walk on all series except for Swedish interest rates.

We interpret result (1) as a warning that allowing for large shifts in parameters while assuming normal iid errors may be dangerous, in particular for forecasting. Under a normal iid assumption, an outlier or an increase in error variance is likely to be interpreted as a mean shift and produce excessively volatile forecasts. This problem will be more severe the more volatile are the time series. As an illustration, Figure 2 shows an extreme case for US interest rates on a longer sample starting already in $1959q2$: during a period (1979-1982) of extremely volatile US interest rates, Shifts exhibits disastrously volatile forecasts while Robust Shifts does not. Note that Figure 2 displays the actual data at time t and the four-steps ahead forecast made at time t. Conversely, if the error variance decreases, a model with iid errors will take longer to recognize a mean shifts that a model with time varying variances. An example of this can be seen in Figure 3, which shows that Robust Shifts adjusted the local mean of the Swedish interest rate downward at a much faster pace than Shifts.

Our sample is too small for us to interpret result (2) as strong evidence of superior forecasting performance of infrequent and large shifts. However, we can put result (1), (2) and (4) together and state with some confidence that well specified change-point models that allow for non-normal errors and non-constant variances are a promising forecasting tool. Result (3) is probably a reflection of the smoothness of interest rates from quarter to quarter. This smoothness means that shifts in mean are very large compared to quarterly prediction errors, facilitating the detection of shifts.

Finally, we observe that inflation and interest rates exhibit a decade long downward shift (approximately 1980-1990), followed by a period of stability. The driftless random walk specification for the local mean outlined in equation (2.9) does not capture these features of the data particularly well.

5.2. **Interval forecasts.** Turning to the evaluation of forecast intervals, let us define the sequence of hit indicators of a 1-step-ahead forecast interval[8] with coverage probability α as: $I_t^\alpha = 1$ if the realized y_t falls inside the interval, and $I_t^\alpha = 0$ otherwise. Christoffersen (1998) develops asymptotic likelihood ratio tests of correct conditional coverage using that $I_t^\alpha \overset{iid}{\sim} Bernoulli(\alpha)$ for a correct forecast interval. Adolfson, Lindé and Villani (2007) propose a Bayesian alternative to these tests by computing posterior probabilities of the following three hypotheses

$$\begin{aligned} H_0 &: I_t^\alpha \overset{iid}{\sim} Bern(\alpha) \\ H_1 &: I_t^\alpha \overset{iid}{\sim} Bern(\pi),\ \pi \text{ unrestricted} \\ H_2 &: I_t^\alpha \sim Markov(\pi_{01}, \pi_{11}),\ \pi_{01} \text{ and } \pi_{11} \text{ unrestricted.} \end{aligned} \tag{5.1}$$

The notation $Markov(\pi_{01}, \pi_{11})$ is here used to denote a general two-state Markov chain with transition probabilities $\pi_{01} = \Pr(0 \to 1)$ and $\pi_{11} = \Pr(1 \to 1)$. If H_0 is supported, the forecast intervals are correct, both in terms of coverage and independence of interval hits. If data supports H_1, the hit indicators are independent, but do not generate the intended coverage α. A large posterior probability of H_2 suggests a violation of the independence property of the interval. Note that even if H_2 receives the largest posterior probability, the coverage of the interval may still be correct. Whether or not the interval has the correct coverage when the evidence is in favor of H_2 is indicated by the relative distribution of the remaining probability mass on H_0 and H_1. We will for simplicity use a uniform prior on π, π_{01} and π_{11}, with prior independence between π_{01} and π_{11}.[9]The posterior probabilities of all three hypoteses in (5.1) can be computed analytically by integration, see Adolfson et al. (2007).

Table 2 presents the results from an out-of-sample evaluation of the 70% predictive intervals of the four models. The columns labelled *Mean* present the posterior mean of the actual coverage, i.e. the posterior mean of π in Equation 5.1. Bold numbers in the *Mean* columns indicate that the target value of 0.7 lies inside the 95% posterior interval of π. The remaining columns of Table 2 display the posterior probabilities of the three hypotheses in equation 5.1. Focusing first on the results for the US sample, it is clear that first two models produce to wide forecast intervals for US GDP growth, whereas Robust TVP and Robust Shifts seem to have the correct coverage (the 95% probability intervals for π covers 0.7) and also independent interval hits (the posterior probabilities of H_0 are 0.714 and 0.774, respectively). Basically the same results apply to US inflation, here Robust Shifts has excellent coverage. Looking at the US interest rates, it is clear that only Robust Shifts accurately captures the distribution of the US interest rate. The other three models have much to wide forecast intervals and ARMA and Robust TVP seem to generate dependent intervals hits.

Turning to the analysis of the Swedish data in Table 2 we see that all four models have fairly accurate intervals for GDP growth. Robust TVP and Robust Shifts perform better than

[8]We shall here restrict attention to forecast intervals with equal tail probabilities.

[9]The working paper version of our paper (available at www.riksbank.com/research/villani) presents results for a whole range of Beta priors.

their competitors on the inflation series, and, similar to the US analysis, Robust Shifts give a much more accurate coverage of the interest rate than the other models. Finally, Table 2 shows that only Robust Shifts has fairly accurate forecast intervals of Australian GDP growth. Robust TVP and Robust Shifts do much better than ARMA and Shifts on Australian inflation. All four models generate to wide 70% intervals for the Australian interest rate, but Robust Shifts is again doing much better than the other models.

A less formal, but more detailed, way of investigating the quality of models' forecast densities is by plotting their normalized forecast error over time. The normalized forecast error at time t is defined as $\Phi^{-1}[F_t(y_{t+1})]$, where $F_t(y_{t+1})$ is the model's one-step-ahead predictive CDF evaluated at the realization y_{t+1}, and $\Phi^{-1}(\cdot)$ is the inverse Gaussion CDF. If the model is correct then $F_t(y_{t+1})$ should be independent and uniformly distributed over time, and hence $\Phi^{-1}[F_t(y_{t+1})]$ should be $NID(0,1)$.

Figure 4 displays the normalized forecast errors for the US data. The ARMA normalized forecast errors in the top row of Figure 4 show clear periods of low and high variance that last at least several years, a feature that neither of the two constant variance models ARMA and Shifts can capture. Robust TVP also struggles here because its smoothly varying variance needs to move rapidly and then stay approximately constant for long periods, the end result here is that Robust TVP behaves more like a fixed-parameter model. Similar to the evidence from the conditional coverage analysis in Table 2, the Robust Shifts model does well one all three variables. Its shifting mean and variance adapts quickly, and its heavy tailed errors serve the dual purpose of moderating the shifting processes and accounting for the largest outliers in the earlier part of the sample. The normalized forecast errors for the Swedish data in Figure 5 show similar behavior as in the US data. The four models give similarly well-behaved normalized forecast errors for Swedish GDP growth and inflation, but they all have a hard time capturing the Swedish interest rate, with Robust Shifts doing a much better job than its competitors. Finally, the normalized residuals in Figure 6 shows clear evidence of over-coverage for the first three models on Australian GDP growth. It is also striking from Figure 6 that Robust Shifts produces much more well-behaved normalized residual for the Australian interest rate than the other three models, who all fail miserably in this respect.

6. Conclusions

Any conclusion based on forecasts made on only nine series should be proposed and received with caution. This being said, the results offer several interesting interpretations. The first point to emerge is that models that allow for large shifts in the conditional mean parameters should be coupled with at least a fat tailed distribution for the errors, and, if possible, shifts in error variance. This observation is likely to extend to regime-switching models and to be particularly important when the focus is on forecasting. With these extensions, models that allow for occasional shifts or regime switches can be useful forecasting tools. Whether smooth, continuous shifts or infrequent and large ones are more appropriate will of course depend on the data, but in our case infrequent shifts perform somewhat better for point forecasts and considerably better for interval coverage. The good performance of the model with infrequent, large shifts in variance is probably due to the fact that in our series the error variance does seem to move in this fashion, with periods of high or low volatility that come about rather quickly and last several years. Mean shifts, in contrast, can take a decade or more and be immediately reversed. Finally, all inflation series show large shifts in the local persistence of the process, which can be fruitfully modeled.

References

[1] Adolfson, M., Lindé, J. and Villani, M., 2007, Forecasting performance of an open economy dynamic stochastic general equilibrium model, *Econometric Reviews*, 26, 289-328.

[2] Bessec, M., and Bouabdallah, O., 2005, What causes the forecasting failure of Markov-Switching models? A Monte Carlo study, working paper, available at http://ideas.repec.org/p/wpa/wuwpem/0503018.html.

[3] Carter, C.K., and Kohn, R., 1994, On Gibbs sampling for state-space models, *Biometrika*, 83, 589-601.

[4] Chib, S. 1998, Estimation and comparison of multiple change-point models, *Journal of Econometrics*, 86, 221-241.

[5] Clements, M.P., and Hendry, D.F., 1999, *Forecasting non-stationary economic time series*, Cambridge, Mass., MIT Press.

[6] Clements M.P., Krolzig H.-M., 1998, A Comparison of the Forecast Performance of Markov-Switching and Threshold Autoregressive Models of US GNP, *Econometrics Journa*l, 1, C47-C75.

[7] Fruhwirth-Schnatter, S., 1994, Data augmentation and dynamic linear models, *Journal of Time Series Analysis*, 15, 183-202.

[8] Fruhwirth-Schnatter, S., 2006, *Finite Mixture and Markov Switching Models*, New York, Springer.

[9] Geweke, J., 2005, Contemporary Bayesian econometrics and statistics, Wiley, New York.

[10] Gerlach, R., C. Carter and R. Kohn, 2000, Efficient Bayesian inference for dynamic mixture models, *Journal of the American Statistical Association*, 95, 819–828.

[11] Giordani, P., and Kohn, R., 2008, Efficient Bayesian inference for multiple change-point and mixture innovation models, *Journal of Business and Economic Statistics*, 26, 66-77.

[12] Giordani, P., and Kohn, R., 2007, Adaptive independent Metropolis-Hastings by fast estimation of mixtures of normals, working paper, University of New South Wales.

[13] Granger, C., and Hyung, Occasional Structural Breaks and Long Memory, *Journal of Empirical Finance*, Discussion Paper 99-14.

[14] Hamilton, J.D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57, 357-84.

[15] Hamilton, J.D., 1994, Time series analysis, Princeton University Press, Princeton.

[16] Harvey, A.C., 1989, Forecasting, structural time series models and the Kalman filter, Cambridge University Press.

[17] Koop, G., and Potter, S., 2007, Prior elicitation in multiple change-point models, forthcoming in the *International Economic Review.*

[18] Koop, G., and Potter, S.M., 2008, Forecasting and estimating multiple change-point models, *Review of Economic Studies*, 74, 763-789.

[19] Levin, A.T., and Piger, J., 2003, Is inflation persistence intrinsic in industrial economies? Federal Reserve Bank of St. Louis Working Paper 2002-023E.

[20] Marcellino M., 2008, A Benchmark for models of growth and inflation, forthcoming in *Journal of Forecasting.*

[21] Shephard, N.E, 1994, Partial non-Gaussian state-space models, *Biometrika*, 81, 115-131.

[22] Stock, J., and Watson, M., 1996, Evidence on structural instability in macroeconomic time series relations, *Journal of Business and Economic Statistics*, 14, 11-30.

[23] Stock, J.H., and Watson, M.W., 2006, Why has inflation become harder to forecast?, NBER Working Paper 12324.

[24] West, M., and Harrison, J., 1997, *Bayesian forecasting and dynamic models*, Springer-Verlag, New York.

Appendix A. Markov Chain Monte Carlo inference

Sampling of parameters and latents is done by Markov Chain Monte Carlo. *LASER* is not a conditionally Gaussian model. However, it falls into the class of multiplicative models (Shephard, 1994) which can be made conditionally Gaussian by sequential conditioning. Efficient inference for conditionally Gaussian sub-models can then be performed as described in Giordani and Kohn (2008). In their paper the time variation in log variances is limited to one equation (measurement or one transition) only. However, this already makes the model multiplicative, so they draw conditional mean parameter and latents given the evolution of the log variance and viceversa. *LASER* extends the time variation to transition equation variances. This results in a rather different model, but the inferential procedure is only trivially modified. We therefore sketch the steps of the Gibbs sampler and refer to Giordani and Kohn (2008) for further details. The MCMC steps are given for Robust Shifts, since this is the more general of the models considered in this paper.

For inferential purposes, it is convenient to rewrite the model in equations (2.1)-(2.4) as

$$y_t = \mu_t + x_t + \alpha_y(S_{y,t}) + \sigma_{y,t}\xi_{y,t}(S_{y,t})\epsilon_t \tag{A.1}$$

$$x_t = \rho_1 x_{t-1} + ... + \rho_p x_{t-p} + \alpha_x(S_{x,t}) + \sigma_{x,t}\xi_{x,t}(S_{x,t})u_t, \tag{A.2}$$

where $S_{y,t}$ and $S_{x,t}$ are discrete latent variables which index components of a mixture of normals.

The MCMC sampler is now sketched. First initialize parameters and latents. The complete update of all parameters and states (one iteration) involves nine steps:

1. Draw latent variables (interventions) in μ_t one at a time given $y_t, \sigma_{y,t}, \sigma_{x,t}, \alpha_y(S_{y,t})$, $\xi_y(S_{y,t}), a_x(S_{x,t}), \xi_x(S_{x,t}), \rho_1, ..., \rho_p$ for $t = 1, ..., n$, with $x^{1,n}$ and $\mu^{1,n}$ integrated out, as described in Giordani and Kohn (2007). ($x^{1,n}$ stands for $x_1, ..., x_n$, with n the sample size. When there is no risk of misunderstanding, we also write x for $x^{1,n}$.)
2. Draw the states $x^{1,n}$ and $\mu^{1,n}$ in one block conditional on the same variables as in step (1) using for example the algorithm of Carter and Kohn (1994).
3. Draw the lag lenght and $\rho_1, ..., \rho_p$ in one block imposing stationarity as described below.
4. Update $S_y^{1,n}$ in one block given $y^{1,n}, \mu^{1,n}, x^{1,n}, \sigma_{y,t}$ and parameters of the MN. This is accomplished by defining $y_t^* = (y_t - x_t - \mu_t)/\sigma_{y,t}$ and noticing that $y_t^* = \alpha_y(S_t)/\sigma_{y,t} + \xi_y(S_{y,t})\epsilon_t$. The problem is then a standard one of drawing the latent indicators of a

standard mixture of normal (except that the division by $\sigma_{y,t}$ must be accounted for) described for example in Geweke (2005).

(5) Update $S_x^{1,n}$ in one block given $x^{1,n}, \rho_1, ..., \rho_p, \sigma_{x,t}$ and the parameters of the MN as described in step (4).

(6) Update the parameters of the MN for the observation equation and the transition equation using a sequence of standard Gibbs steps described in Geweke (2005).

(7) Define $y_t^* = (y_t - x_t - \mu_t - \boldsymbol{\alpha}_y(S_{y,t}))/\xi_y(S_{y,t}) = \sigma_{\epsilon,t}\epsilon_t$. Update $\sigma_{\epsilon,t}$ as in Giordani and Kohn. This again involves updating latent indicators one at a time and then updating $\ln(\sigma_y^{1,n})$ in one block.

(8) Similarly define $x_t^* = (x_t - \rho_1 x_{t-1} - ... - \rho_p x_{t-p} - a_x(S_{y,t}))/\xi_x(S_{y,t}) = \sigma_{x,t}u_t$ and update $\sigma_{x,t}$.

(9) Update the variances of the shifts in $\mu_t, \ln(\sigma_{y,t})$ and $\ln(\sigma_{x,t})$ given $x^{1,n}, \mu^{1,n}$, $\ln(\sigma_y^{1,n})$ and $\ln(\sigma_x^{1,n})$ drawing from independent inverse gamma distributions.

Drawing the lag length while imposing stationarity. Updating the lag lenght p in step 3 while imposing stationarity is not particularly difficult, but since statements to the contrary and cumbersome approaches are sometimes seen in the literature, we outline this step in more detail.

Let P be the maximum number of lags (four in out paper). We draw p and $\rho_1, ..., \rho_p$ jointly conditional on $x_{-P+1}, ..., x_n$, on $\sigma_x^{1,n}, S_x^{1,n}$ and the parameters of the MN.[10] This is done as follows:

(1) Define $x_t^* = (x_t - a_x(S_{y,t}))/(\xi_x(S_{y,t}) \times \sigma_{x,t})$ and $x_{t-i}^* = x_{t-i}/(\xi_x(S_{y,t})\sigma_{x,t})$ for $t = 1, ..., n$ and $i = 1, ..., P$, where $x_{-P+1}, ..., x_0$ are generated by data augmentation at each draw. Then $x_t^* \sim N(\rho_1 x_{t-1}^* + ... + \rho_P x_{t-P}^*, 1)$.

(2) Propose a lag length p as a random draw from $1, ..., P$.

(3) Draw $\rho_1^{'}, ..., \rho_p^{'}$ given $x_{-P}^*, .., x_n^*$. Given a normal prior, the posterior is also normal with standard form.

(4) Reject the draw of $p^{'}$ and $\rho_1^{'}, ..., \rho_p^{'}$ if the parameters imply that x_t^* lies outside the stationary region (see Hamilton, 1994). Otherwise accept with probability

$$\min \left\{ 1, \frac{f(x^*|\rho^{'})f(\rho^{'}|p^{'})f(p^{'})}{f(x^*|\rho)f(\rho|p)f(p)} \frac{f(\rho|x^*, p)}{f(\rho^{'}|x^*, p^{'})} \right\},$$

where $\rho = (\rho_1, ..., \rho_P)$.

We notice that the acceptance rate is one for $p^{'} = p$ and ρ stationary, so this procedure is as efficient as the more common practice of integrating out ρ, but more practical to impose stationarity or any other truncated prior on ρ.

[10] An alternative would be to perform a MH step without conditioning on $x_{-p^*+1}, ..., x_n$. This more expensive approach will probably work substantially better when x_t accounts for a small share of the variance of y_t (Fruhwirth-Schnatter, (1994) and Giordani and Kohn, 2007).

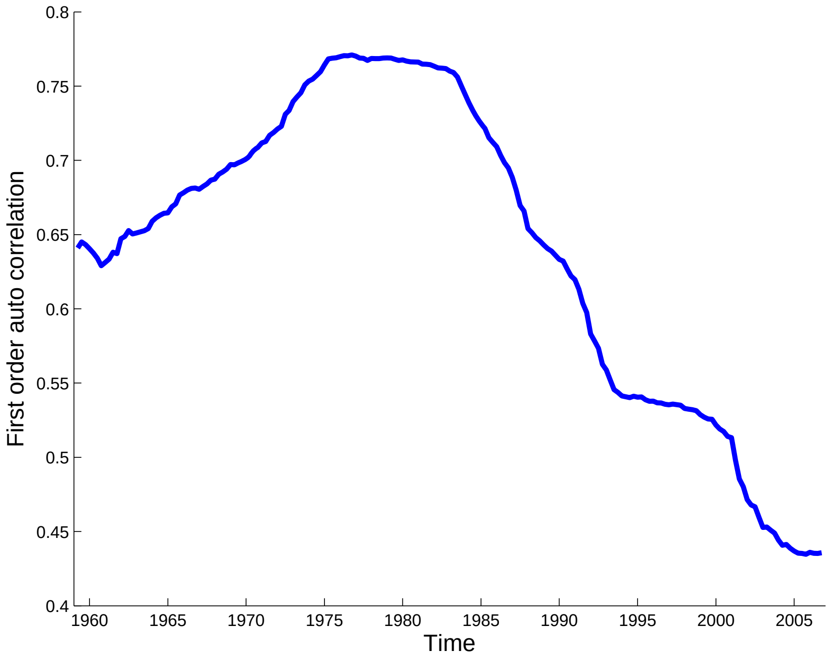

FIGURE 1. First-order local autocorrelation for the US inflation using the Robust Shifts model with one lag. The parameters of the model is estimated using the full sample, but the autocorrelation is computed at different points in time.

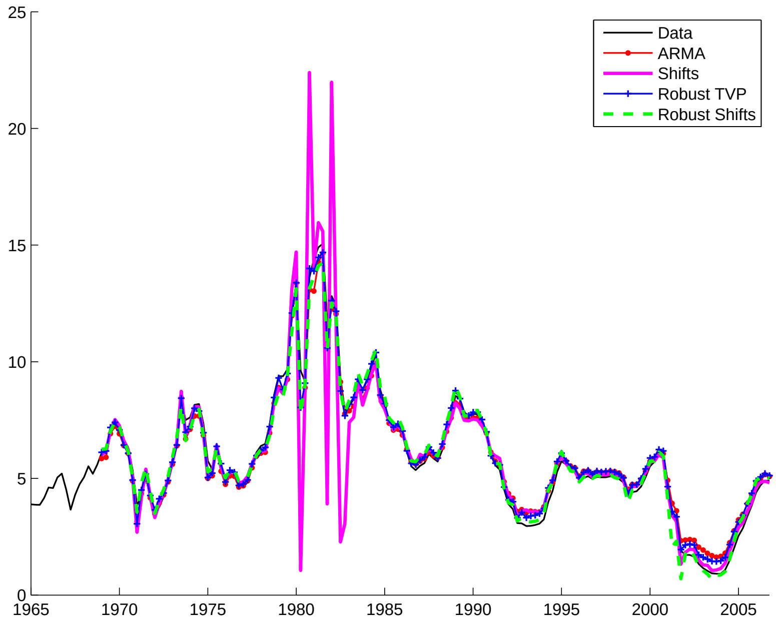

FIGURE 2. Sequential 4-steps-ahead forecasts of the U.S. interest rate. The figure displays the actual data at time t and the 4-steps-ahead forecasts produced at time t.

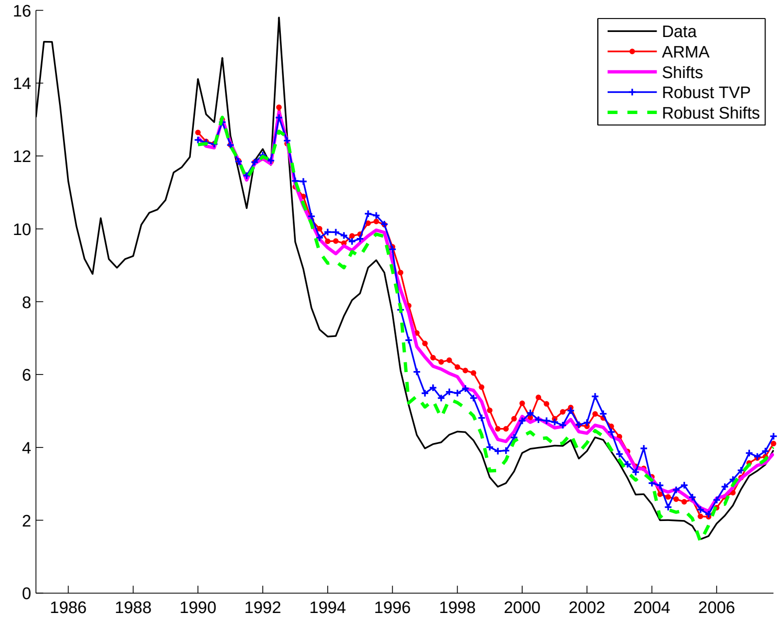

FIGURE 3. Sequential 4-steps-ahead forecasts of the Swedish interest rate. The figure displays the actual data at time t and the 4-steps-ahead forecasts produced at time t.

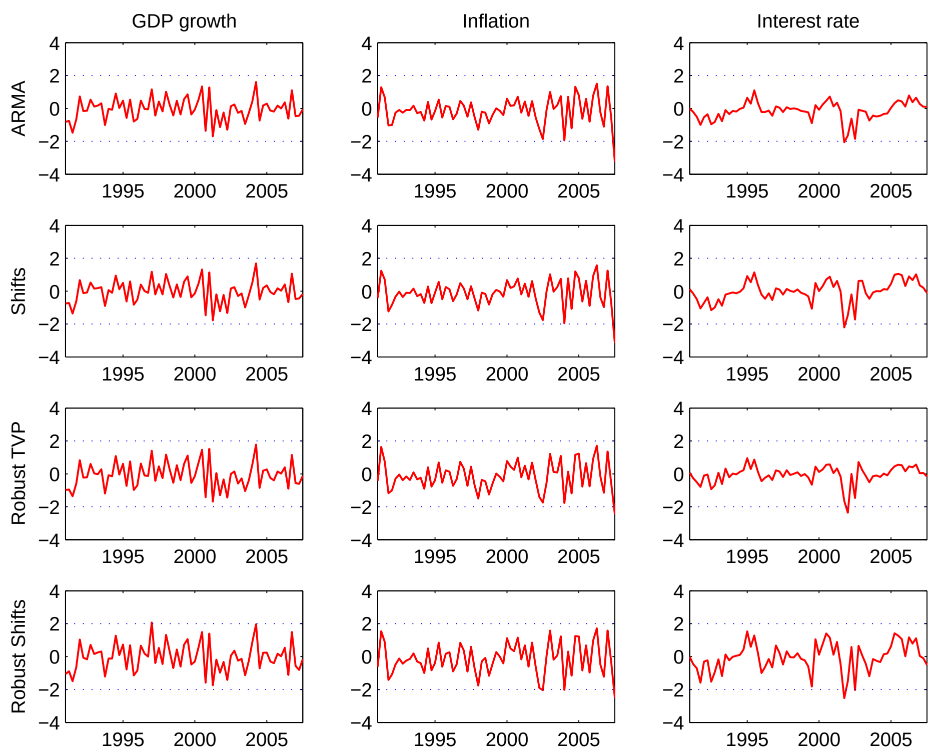

FIGURE 4. Normalized forecast errors plotted over time for the U.S. data.

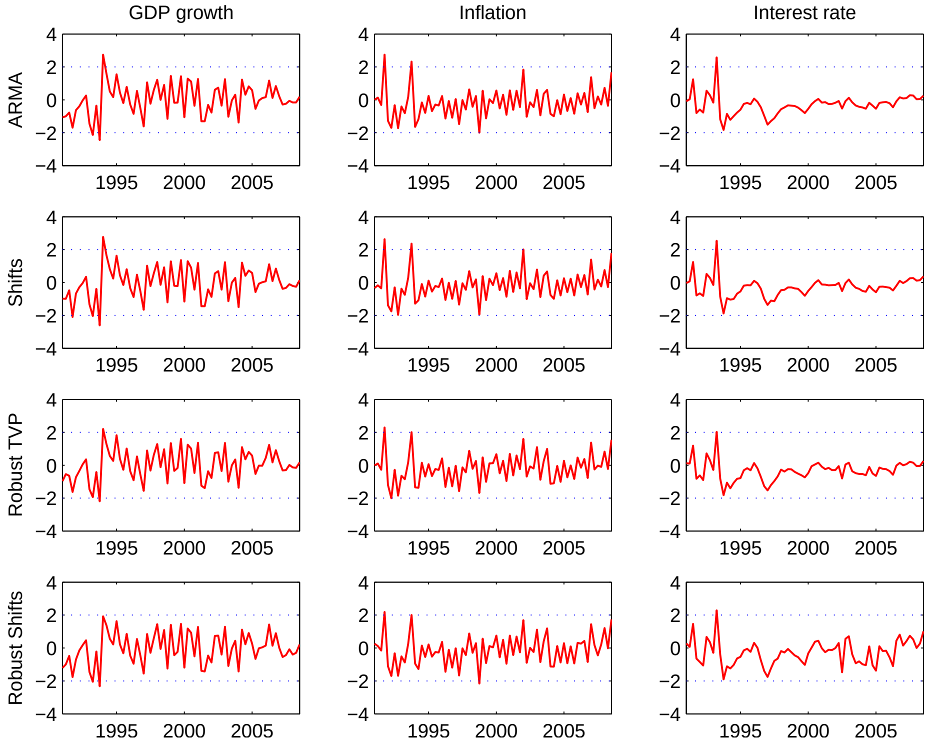

FIGURE 5. Normalized forecast errors plotted over time for the Swedish data.

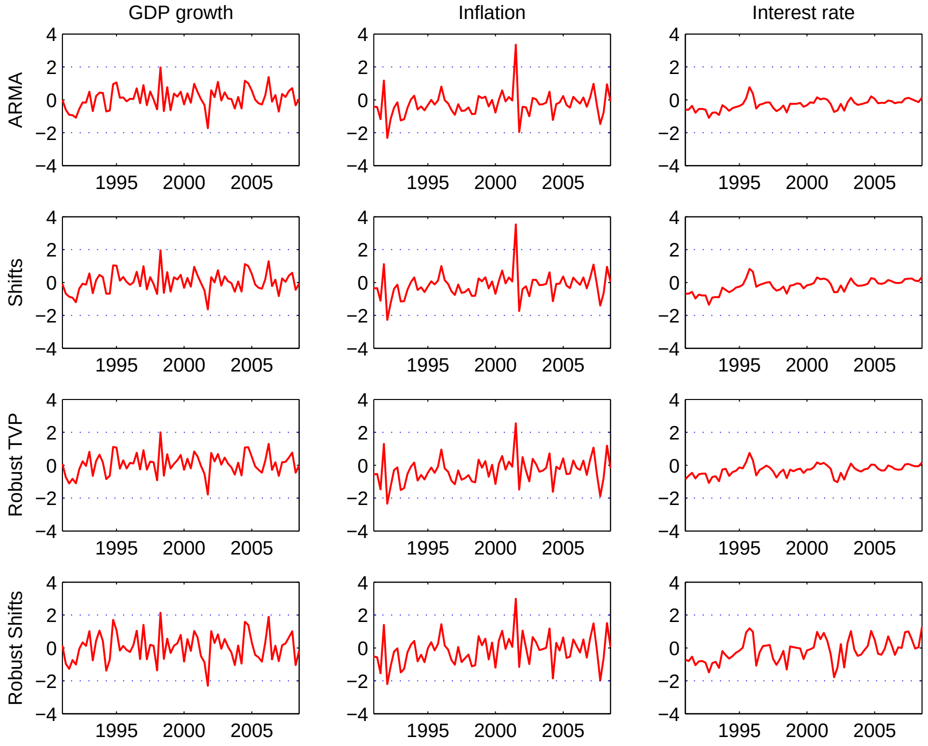

FIGURE 6. Normalized forecast errors plotted over time for the Australian data.

U.S. 1990Q2-2006Q4

	1Q			4Q		
	GDP	Infl.	Rate	GDP	Infl	Rate
ARMA	1.991	1.289	0.409	**1.950**	1.410	1.419
Shifts	0.985	**0.993**	1.011	1.033	1.005	0.997
Robust TVP	1.002	1.023	1.133	1.014	**0.972**	0.974
Robust Shifts	**0.974**	1.029	**0.866**	1.052	1.001	**0.942**
Random walk	2.086	1.345	1.900	2.402	1.710	1.201

Sweden 1990Q2-2007Q4

	1Q			4Q		
	GDP	Infl.	Rate	GDP	Infl.	Rate
ARMA	2.095	3.117	1.071	2.805	3.337	2.404
Shifts	1.010	0.989	0.962	0.996	0.966	0.935
Robust TVP	0.986	0.946	0.971	**0.983**	0.951	0.938
Robust Shifts	**0.985**	**0.909**	**0.873**	0.989	**0.935**	**0.837**
Random walk	1.115	1.261	1.161	1.184	1.051	0.918

Australia 1990Q2-2007Q4

	1Q			4Q		
	GDP	Infl.	Rate	GDP	Infl.	Rate
ARMA	2.909	2.715	0.703	2.906	3.160	2.407
Shifts	**0.999**	0.996	0.968	**0.999**	0.971	0.956
Robust TVP	1.001	1.011	1.134	1.000	1.042	1.148
Robust Shifts	1.014	**0.950**	**0.925**	1.003	**0.964**	**0.976**
Random walk	1.391	1.244	1.510	1.380	1.237	1.540

TABLE 1. Root Mean Squared Forecast Errors (RMSFE). The first row of each subtable displays the RMSFE of the ARMA model. The following rows give the ratio of the RMSFE of the ARMA model to the RMSFE of the indicated model. The best model for a given country, variable and horizon is indicated in bold.

	US 1990Q2 - 2006Q4											
	GDP				Inflation				Interest rate			
	Mean	H_0	H_1	H_2	Mean	H_0	H_1	H_2	Mean	H_0	H_1	H_2
ARMA	0.838	0.088	0.503	0.407	0.823	0.187	0.485	0.327	0.926	0.000	0.496	0.503
Shifts	0.823	0.200	0.521	0.277	0.823	0.223	0.580	0.195	0.867	0.019	0.753	0.227
RobustTVP	**0.764**	0.714	0.219	0.065	**0.750**	0.749	0.169	0.081	0.941	0.000	0.328	0.671
RobustShifts	**0.750**	0.774	0.174	0.050	**0.706**	0.839	0.118	0.041	**0.720**	0.834	0.127	0.038
	Sweden 1990Q2 - 2007Q4											
	GDP				Inflation				Interest rate			
	Mean	H_0	H_1	H_2	Mean	H_0	H_1	H_2	Mean	H_0	H_1	H_2
ARMA	**0.666**	0.748	0.119	0.132	**0.778**	0.613	0.276	0.110	0.875	0.000	0.032	0.967
Shifts	**0.680**	0.761	0.108	0.130	**0.792**	0.473	0.340	0.186	0.899	0.001	0.469	0.528
RobustTVP	**0.694**	0.793	0.107	0.098	**0.750**	0.761	0.170	0.068	0.861	0.000	0.008	0.991
RobustShifts	**0.666**	0.748	0.119	0.132	**0.736**	0.812	0.144	0.042	**0.792**	0.402	0.2901	0.307
	Australia 1990Q - 2007Q4											
	GDP				Inflation				Interest rate			
	Mean	H_0	H_1	H_2	Mean	H_0	H_1	H_2	Mean	H_0	H_1	H_2
ARMA	0.888	0.002	0.767	0.230	0.847	0.003	0.032	0.966	0.972	0.000	0.660	0.340
Shifts	0.903	0.001	0.785	0.214	0.833	0.017	0.087	0.895	0.972	0.000	0.660	0.340
RobustTVP	0.861	0.021	0.663	0.314	**0.777**	0.573	0.258	0.168	0.972	0.000	0.660	0.340
RobustShifts	**0.777**	0.633	0.285	0.086	**0.722**	0.750	0.113	0.137	0.847	0.061	0.727	0.212

TABLE 2. Out-of-sample evaluation of the models' 70% one-step-ahead predictive intervals. The column labelled 'Mean' present the posterior mean estimate of each model's actual coverage probability (π in equation 5.1). Numbers in bold indicate that the point 0.7 (the intended coverage) lies inside the 95% posterior intervals for the actual coverage, and hence that the intervals has the correct coverage. The columns labelled H_0, H_1 and H_2 display the posterior probability of the three hypotheses in equation 5.1 for the uniform prior on the π's.