

Forecasting Real Time Inflation with Time Varying Vector Autoregressions

Antonello D'Agostino

CBFSAI*

Luca Gambetti

UAB and RECent†

Domenico Giannone

European Central Bank‡

First Draft: September 2008,
Very Preliminary

Abstract

We perform an out-of-sample real time forecasting exercise using a model with the unemployment rate, inflation and the short term interest rate. We compare the time-varying autoregressions forecast with those produced by alternative fixed coefficient models. Our findings show that TV-VAR are very powerful in forecasting. In particular the time-varying autoregressions model is the only model producing forecasts that are accurate for all the three variables. Moreover the time-varying autoregressions model produces by far the most accurate forecast for inflation.

*Contact: Central Bank and Financial Services Authority of Ireland - Economic Analysis and Research Department, PO Box 559 - Dame Street, Dublin 2, Ireland. E-mail: antonello.dagostino@centralbank.ie.

†Contact: Office B3.174, Departament d'Economia i Historia Econòmica, Edifici B, Universitat Autònoma de Barcelona, Bellaterra 08193, Barcelona, Spain. Tel (+34) 935811289; e-mail: luca.gambetti@uab.cat

‡Contact: European Central Bank - Monetary Policy Research, Postfach 16 03 19, 600 66, Frankfurt am Main, Germany; ; e-mail: domenico.giannone@ecb.int. The views expressed in this paper are those of the authors, and do not necessarily reflect those of the Central Bank and Financial Services Authority of Ireland or the European Central Bank.

1 Introduction

The US economy has undergone many structural changes in the post war period. On the real side, we have witnessed a strong moderation of business cycle fluctuations. On the nominal side we had the rise of inflation in the 70s and the subsequent fall in the course of the 80s. The conduct of policy has undergone substantial changes, in particular monetary policy that has become more transparent and more aggressive against inflation (Clarida, Gali and Gertler, 2001). The relation among the real and the nominal side of the economy has changed drastically, the most well know case is the recent break down of the Phillips curve which as very strong and stable before the 80s (Stock and Watson, 2007).

To study these evolution the literature has developed dynamic models that allow for time variation in the dynamic interrelations among economic variables. Since the seminal paper by Cogley and Sargent (2003), a growing amount of works have conducted the analysis using time-varying vector autoregressions (e.g. Primiceri, 2006, Canova and Gambetti, 2008). These reduced form models are essentially forecasting models in which the dynamic interrelation and the size of the forecast errors are allowed to change over time. These model have been proved to be very powerful for studying, explain and tracking the structural changes that have affected affecting the economy. Taking into account structural changes might be helpful for forecasting. There are at least two reasons why these could be helpful. First, the models could be able to detect structural changes in real time and hence help forecasting them, e.g. the declined level of inflation in the 80. Structural changes could hide short term relations among macroeconomic variables which could be could be reestablished once the structural break/trend component is removed (Cogley, Primiceri and Sargent, 2008 using a TV VAR find that the Phillips curve is reestablished once the trend component of inflation is removed; Cogley and Sbordone, find that the Phillips curve relation could be distorted by the presence of structural break).

The existing literature is based however on in-sample ex-post estimation and it is not known whether these model are also useful in real-time. This paper is a first attempt to fill this gap. We perform an out-of-sample forecasting exercixe using a model with Unemployment rate, inflation and the policy interest rate. The exercise if real-time since the forecast is estimated recursively using the data that were available at the time the forecast is made (we use real-time data).

We compare the TV-VAR forecast with those produced by alternative fixed co-

efficient models. We estimate them using all historical data available at each point the forecast is made. We also consider estimate these models using a rolling windows which is an heuristic way of taking structural changes into account.

Our findings show that TV-VAR are very powerful in forecasting. In particular the TV-VAR model is the only model producing forecasts that are accurate for all the three variables. Precisely, the TV-VAR produces by far the most accurate forecast for inflation. For the unemployment rate the best forecasts are produced by the TV VAR and a fixed coefficient models estimated over rolling windows. For the interest rate the best forecast are produced by the TV-VAR and the TV-AR.

These implies that TV-VAR models are faster than rolling VARs in recognizing structural changes in inflation. Moreover, the TV multivariate models work better than TV univariate models and the Random walk model indicating that interrelations among macroeconomic variables carry out important information once we take structural change explicitly into account. These results holds true in different sub-samples. In particular, they are also confirmed in the great moderation sample when it is known that inflation has become very hard to forecast since most of the model are outperformed by the RW (Atkenson and Ohanian) and simple univariate TV models.

2 The Dataset

The dataset used in the paper consists of three variables: GDP deflator (GDPD), Unemployment Rate (UR) and the three month treasury bills (denoted as IR). We use real time databases for GDPD (quarterly vintages on quarterly frequency) and UR (quarterly vintages on monthly frequency).¹ For the three month interest rate we use the actual series.² The unemployment and the interest rate series are sampled monthly, following CS2001, CS2005 and CPS, we convert them into quarterly series by taking the middle month and the first month values in each quarter, respectively for UR and IR. The time span is therefore on quarterly basis from 1948:Q1 to 2007:Q4. It is covered, in the case of GDPD and UR, by the quarterly vintages 1966:Q1 to 2007:Q4.

¹The data are available on the Federal Reserve Bank of Philadelphia website at: <http://www.phil.frb.org/econ/forecast/reaindex.html>.

²The series is available on the FRED dataset of the Federal Reserve Bank of St. Louis (mnemonics TB3MS), at: <http://research.stlouisfed.org/fred2/series/TB3MS>

3 The Forecasting Exercise

Our objective is to predict the h -period ahead unemployment rate UR_{t+h} , the interest rate IR_{t+h} and the annualized price inflation $\pi_{t+h}^h = \frac{400}{h} \log(\frac{P_{t+h}}{P_t})$, where P_{t+h} is the GDP deflator at time $t+h$ and $\frac{400}{h}$ is the normalization term.

3.1 The forecasting model

Let y_t be a vector including the inflation rate (π_t), the unemployment rate (UR_t), and the short term interest rate (IR_t). We assume that y_t admits the following time varying coefficients VAR (TV-VAR) representation:

$$y_t = A_{0,t} + A_{1,t}y_{t-1} + \dots + A_{p,t}y_{t-p} + \varepsilon_t \quad (1)$$

where $A_{0,t}$ contains time-varying intercepts, $A_{i,t}$ are matrices of time-varying coefficients, $i = 1, \dots, p$ and ε_t is a Gaussian white noise with zero mean and time-varying covariance matrix Σ_t . Let $A_t = [A_{0,t}, A_{1,t}, \dots, A_{p,t}]$, and $\theta_t = \text{vec}(A_t)$, where $\text{vec}(\cdot)$ is the column stacking operator. We assume that all the roots of the VAR polynomial lie outside the unit circle at every t - this is sufficient to make y_t locally stationary. Given this condition, we postulate the following law of motion for θ_t :

$$\theta_t = \theta_{t-1} + \omega_t \quad (2)$$

where ω_t is a Gaussian white noise with zero mean and covariance Ω . We let $\Sigma_t = F_t D_t F_t'$, where F_t is lower triangular, with ones on the main diagonal, and D_t a diagonal matrix. Let σ_t be the vector of the diagonal elements of $D_t^{1/2}$ and $\phi_{i,t}$, $i = 1, \dots, n-1$ the column vector formed by the non-zero and non-one elements of the $(i+1)$ -th row of F_t^{-1} . We assume:

$$\log \sigma_t = \log \sigma_{t-1} + \xi_t \quad (3)$$

$$\phi_{i,t} = \phi_{i,t-1} + \psi_{i,t} \quad (4)$$

where ξ_t and $\psi_{i,t}$ are Gaussian white noises with zero mean and covariance matrix Ξ and Ψ_i , respectively. Let $\phi_t = [\phi'_{1,t}, \dots, \phi'_{n-1,t}]$, $\psi_t = [\psi'_{1,t}, \dots, \psi'_{n-1,t}]$, and Ψ be the covariance matrix of ψ_t . We assume that $\psi_{i,t}$ is independent of $\psi_{j,t}$, for $j \neq i$, and that ξ_t , ψ_t , ω_t , ε_t are mutually uncorrelated at all leads and lags. In principle, one could make ε_t and ω_t correlated. However, it is well known that such a model can be equivalently represented with a setup where shocks are mutually uncorrelated but

ε_t is serially correlated. Since our measurement equation is a VAR, such a flexibility is unneeded here. Note that the specification in (3) and (4) is similar to the one employed by Primi. Relative to CS2005, it allows $\psi_t \neq 0$ at each t . Details of estimation appear in Appendix.

Equation (1) has the following companion form

$$\mathbf{y}_t = \mu_t + \mathbf{A}_t \mathbf{y}_{t-1} + \epsilon_t$$

where $\mathbf{y}_t = [y'_t \dots y'_{t-p+1}]'$, $\epsilon_t = [\varepsilon'_t 0 \dots 0]'$ and $\mu_t = [A'_{0,t} 0 \dots 0]'$ are $np \times 1$ vectors and

$$\mathbf{A}_t = \begin{pmatrix} & A_t \\ I_{n(p-1)} & 0_{n(p-1),n} \end{pmatrix}$$

where $A_t = [A_{1,t} \dots A_{p,t}]$ is an $n \times np$ matrix, $I_{n(p-1)}$ is an $n(p-1) \times n(p-1)$ identity matrix and $0_{n(p-1),n}$ is a $n(p-1) \times n$ matrix of zeros. Let $\hat{\mu}_t$ and $\hat{\mathbf{A}}_t$ denote the median of the joint posterior distribution of μ_t , \mathbf{A}_t (see Appendix for the details). The forecast of \mathbf{y}_{t+1} 1-step ahead is:

$$\hat{\mathbf{y}}_{t+1|t} = \hat{\mu}_t + \hat{\mathbf{A}}_t \mathbf{y}_t \quad (5)$$

Forecasts at time $t+h$ are computed iteratively from the previous forecasts:

$$\hat{\mathbf{y}}_{t+h|t} = \hat{\mu}_t + \hat{\mathbf{A}}_t \hat{\mathbf{y}}_{t+h-1} = \sum_{j=1}^h \hat{\mathbf{A}}_t^{j-1} \hat{\mu}_t + \hat{\mathbf{A}}_t^h \mathbf{y}_t \quad (6)$$

Forecasts of π_{t+h}^h are computed by cumulating the first h forecasts of the first entries (which correspond to π_t) of the forecasted vector $\hat{\mathbf{y}}_{t+h|t}$:

$$\hat{\pi}_{t+h|t}^h = \frac{1}{h} \sum_{i=1}^h \hat{\mathbf{y}}_{1,t+i|t}$$

where $\hat{\mathbf{y}}_{1,t+h-i|t} = \pi_{t+h-i}$ if $i \geq h$. Forecast of UR_{t+h} and IR_{t+h} correspond to the second and third entries of the of the forecasted vector $\hat{\mathbf{y}}_{t+h|t}$.

3.2 Other forecasting models

The forecasting ability of the time varying VAR is compared with that of other time series models. Particular emphasis will be posed on the comparison between the time varying VAR and the non-time varying (coefficient) counterpart. Such comparison allows us disentangle and evaluate the eventual contribution to the forecast accuracy due to the richer model specification.

- **Naïve Forecasting Models (Benchmarks):**

The first approach used to forecast the variables is a simple, naïve forecasting model. In the case of inflation we choose the Atkeson and Ohanian (2001) (AO) benchmark. AO demonstrate that, since 1984, structural models of US inflation have been outperformed by a naïve forecasts based on the average rate of inflation over the current and previous three quarters. This is essentially a "no change" forecast for inflation:

$$\hat{\pi}_{t+h|t}^{h,ao} = \pi_t^4 = \frac{1}{4}(\pi_t + \pi_{t-1} + \pi_{t-2} + \pi_{t-3}) \quad (7)$$

In the case of unemployment rate and interest rate the naïve benchmark is slightly modified. In these cases, the last available data point is used as an estimate of the future values at every horizon.

- **Time Varying Autoregression (TV-AR)**

Inflation (π_t), unemployment rate (UR_t) and interest rate (IR_t) are modelled as an autoregressive model with drifting coefficients and drifting stochastic volatility in the error term:³

$$x_t = \alpha_{0,t} + \alpha_{1,t}x_{t-1} + \alpha_{2,t}x_{t-2} + \dots + \alpha_{p,t}x_{t-p} + \varepsilon_t \quad (8)$$

The 1-step ahead ahead forecast is computed as:

$$\hat{x}_{t+1|t}^{tv-ar} = \hat{\alpha}_{0,t} + \hat{\alpha}_{1,t}x_t + \hat{\alpha}_{2,t}x_{t-1} + \dots + \hat{\alpha}_{p,t}x_{t-p+1}$$

where the *hat* denotes coefficients estimates. Details of the estimation are given in appendix 1. In general, at time $t+h$, the forecast can be computed recursively from the previous forecasts:

$$\hat{x}_{t+h|t}^{ar-tv} = \hat{\alpha}_{0,t} + \sum_{i=1}^p \hat{\alpha}_i \hat{x}_{t+h-i|t}^{ar-tv}$$

where $\hat{x}_{t+h-i|t}^{ar-tv} = x_{t+h-i}$ if $i \geq h$. In the case of inflation, $x_{1,t} = \pi_t$, predictions at time $t+h$, $\hat{\pi}_{t+h|t}^{h,ar-tv}$, are computed by cumulating the first h forecasts of the (log) price changes $\frac{1}{h}(\sum_{i=1}^h \hat{\pi}_{t+h|t}^{ar-tv})$

³AR coefficients and the residual volatility are assumed to evolve according to the equations discussed in the previous subsection.

- **Recursive Autoregression (AR-REC)**

Each of the three variables π_t , UR_t and IR_t is modelled as an autoregressive process:

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \epsilon_t \quad (9)$$

Forecasts are computed exactly as in the previous after the parameters are estimated by Ordinary Least Squares (OLS).

- **Rolling Autoregression (AR-ROL)**

This model is equivalent to the previous one AR-REC, with the difference that estimates of the parameters are computed on a ten years rolling window, instead of using the whole sample length.

- **Recursive VAR (VAR-REC)** This is a VAR model in the three variables, π_t , UR_t and IR_t , collected in a (3×1) vector of variables y_t :

$$y_t = A_0 + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + e_t \quad (10)$$

Forecasts of the variable π_t are computed using formulas (5) and (6), but replacing A_t with the estimated matrix of parameters computed by OLS.

- **Rolling VAR (VAR-ROL)** Equivalent to VAR-REC, with estimates computed by OLS on a ten years rolling window.

3.3 The Simulation Exercise

To analyze the predictive power of the previous model we perform a standard out-of-sample forecast exercise. The procedure is as follows; the exercise begins by estimating all the parameters of the models on a sample span called the estimation window (1948:Q2 to 1969:Q4). The estimated parameters are then used to forecast the variables h -steps ahead outside the estimation window.⁴ The estimation window is updated sequentially with one observation and the parameters are re-estimated based on the new sub-sample.⁵ The h -steps ahead forecasts are again computed outside the new sample. This procedure is then iterated until the end of the sample. Forecasts of the vector of variables $x_{t+h,v}^h$ labeled as $\hat{x}_{t+h|t}^{h,i}$, with $i \in$

⁴Following the related literature, we use a two lags specifications for all the analyzed models.

⁵In the case of VAR-ROL the estimation window is kept fixed to ten years. The first observation is then dropped every time a new observation is added in the iteration.,

$\{NAIVE, AR - REC, AR - ROL, AR - TV, VAR - REC, VAR - ROL, VAR - TV\}$ are stored and used to compute the Mean Square Forecast Error ($MSFE_h$) at horizon h .⁶ For horizon h the $MSFE_h$ is defined as the following:

$$MSFE_h^i = \frac{1}{t_1 - (t_0 + h)} \sum_{s=t_0+h}^{t_1} (x_{s,s+2}^h - \hat{x}_{s|t}^{h,i})^2$$

with $i \in \{NAIVE, AR - REC, AR - ROL, AR - TV, VAR - REC, VAR - ROL, VAR - TV\}$. Since data are continuously revised at each quarter, several measures of inflation available. Following Romer and Romer (2000), we consider the figures published after the next two subsequent quarters as the true realized values.

The MSFE is a measure of the average forecast accuracy over the out-of-sample window. In the empirical exercise $t_0 = 1970 : Q1 + h$ and $t_1 = 2007 : Q4 - 12 + h$ with $h = 1, 2, \dots, 12$.⁷

To facilitate comparisons between the various models, the results are reported in terms of the relative $MSFE$ statistics, where the $MSFE$ of the $NAIVE$ model, used as benchmark, is at the denominator:

$$\frac{MSFE_h^i}{MSFE_h^{NAIVE}}$$

When the relative MSFE is less than one, the model i improves the forecast of the benchmark model. For example, a value of 0.8 would indicate that model i improves the forecast performance of the benchmark model by 20%.

4 Results

Forecasting results for the three variables are reported in table 1 to 3 in appendix appendix 2 and are discussed below individually for each variable.

4.1 Inflation

Results for inflation are reported in table 1. The table is divided in three sub-tables describing the results for the all sample, the first sub-sample (1970:Q1 to 1984:Q4)

⁶The ex-post realized inflation $x_{t+h,v}^h$ can be computed on different vintages. The second subscript v denotes the vintage used for such computation.

⁷In the simulation exercise forecasts for horizon $h = 1$ correspond to nowcast, given that in the real time dataset figures are available only up to the previous quarter.

and the second sub-sample (1985:Q1 to 2007:Q4) respectively. In each sub-table, the first row refers to the forecasting horizon, ranging from 1 to 12. In the second row there is the MSFE for the naïve model, used as benchmark. The following rows show the ratio between the MSFE of obtained with a particular model to the MSFE obtained with the benchmark.

Results for the full sample point to a superior forecasting ability of the VAR-TV model. Such model dominates the benchmark specifications over all the horizons with an average improvement of about 35%. A relative good performance is observed also for the univariate time varying specification (AR-TV) with improvements of about 10% over the horizons 3 to 7. The other specifications, univariate and multivariate fail to improve the forecasts accuracy of the benchmark.

The analysis split for the two sub-samples shows that in the first sub-sample (1970:Q1 to 1984:Q4), the VAR-TV model again displays a superior performance, with sizeable improvements on the benchmark over all the horizons. The AR-TV model also improves the forecast accuracy of the naïve model, with improvements greater than 10% in the middle horizons; none of the other models does better than the benchmark.

On the second sub-sample (1985:Q1 to 2007:Q4), results are somehow similar. The VAR-TV model is once again the best performing model; its forecast accuracy increases with the horizon. We observe an improvement of about 10% over the benchmark model, for horizon 4, which steadily increases to an outstanding 50% for the three years predictions (horizon 12). On this sub-sample, the model displays a performance similar to the benchmark for the shorter horizon. The other models fail to improve upon the simple univariate specification, the only exception is the AR-TV in the long run.

A visual inspection of the predictions can highlights the forecasting properties of the different model. Figure 1, in appendix 3, plots, for the two years horizon, the forecasts of three models, VAR-TV (green line), VAR-REC (dotted, red line) and AR-TV (dashed, grey line) against the ex-post realized, true values (blue line). The time varying VAR tracks extraordinarily well the persistent movements of the two years changes in prices (blue line), both before and after the great moderation period. Also the AR-TV model exhibits a good ability in tracking the underlying, persistent movements in the trend, even if it is less accurate than the three variables multivariate specifications. The VAR-REC model, on the other side, fails in tracking such movements as quickly as the other models and the underlying trend is tracked

only after a significant time delay.

4.2 Unemployment

Table 2 reports the results for unemployment. Over all the sample, all the models have a good forecasting performance, especially in the long run, the only exception is the AR-ROL model, which never beats the benchmark specification. Particularly good is the performance of the VAR-REC and VAR-TV models. In the long run (horizon 12) their improvements on the naïve model are about 53% and 40% respectively.

On the first sub-sample (1970:Q1 to 1984:Q4), there is, in general, a good performance of the multivariate models, while all the univariate specifications fails to improve upon the benchmark model. On this sub-sample, unlike the inflation case, a richer information set is more important than the richer dynamics, which is captured by the drifting coefficients.

On the second sub-sample (1985:Q1 to 2007:Q4) results are quite similar. In this case also the AR-TV and AR-ROL (in the long run) have a good forecasting performance. Particularly, good are once again the performances of VAR-REC (especially in the long-run) and VAR-TV (both in the short and long-run) models, with improvements, at horizon 12, of about 65% and 50% respectively.

Figure 2 displays the forecasts, at horizon 8, for the three models, AR-TV, VAR-REC and AR-TV against the realized ex-post values of the unemployment rate. The two VAR models have quite collinear forecasts. At the beginning of the 80s there is a spike in the forecast path of the three models, which is more severe for the time varying specification. The univariate specification does not track very well the unemployment rate series in the pre-85 sample, while in the great moderation period its predictions are comparable with those obtained with the multivariate specifications.

4.3 Interest Rate

Results for interest rate are shown in table 3. Over all the sample, the time varying specifications improve the accuracy of the benchmark model, while the other models perform worse. The improvements at three years horizon are of about 9% and 16% respectively for AR-TV and VAR-TV.

During the first period (1970:Q1 to 1984:Q4), none of the models provide forecasts more accurate than the naïve specification. The VAR-TV displays, on average, the same forecasting performance.

The second sub-sample (1985:Q1 to 2007:Q4) is characterized by an accurate forecasting performance of VAR-TV, VAR-REC and AR-TV. The other models do not improve on the accuracy of the benchmark specification (AR-ROL shows some improvements only on the longer horizons). Those results confirm the findings of D’Agostino Giannone and Surico (2006). They find that predictability of interest rates have increased during the great moderation period. The interpretation of this result is consistent with the new course of the monetary policy management undertaken by the Federal Reserve during those years. The improved transparency and the better communication strategy policies have probably contributed to enlarge the information set of the agents, who more easily might have predicted the future path of interest rates.

Figure 3 below displays the two years ahead forecasts for interest rates. The three models (VAR-TV, AR-TV and VAR-REC) deliver quite collinear forecasts. At the beginning of the 80s however the VAR-REC does not track very well the movements of interest rates and it takes some time before the forecasts of this model realign to those of the time varying specifications to capture the persistent movements in the interest rate’s trend.

4.4 Discussion

The forecasting results described in the previous section highlights some important features which is worth stressing. Time varying specifications, especially the VAR one, offer valid alternatives to traditional models the literature and the practitioners have used so far for forecasting purposes. In fact, the VAR-TV model, calibrated with priors as in CPS⁸, provides outstanding forecasts for inflation over all the sample. Results are very good especially over the post-85 sample, where the improvements on the AO benchmark, which is hard to beat **[see] AO, SW_JMBC2007 are quite remarkable.*

Good predictions are also provided for the interest rate series, which displays dynamic properties, such as the persistence, similar to inflation. In the pre-85 sample the forecast accuracy is comparable, on average, to that of the benchmark model and on the post-85 sample it is much better.

Finally, results for unemployment are very good as well. This series however exhibits movements which are less persistent than those of the other two series; therefore

⁸Results are robust to alternative priors specifications and are available upon request from the authors

also the other time series models, which are designed to describe more stationary dynamics, have a good forecasting performance.

5 Conclusions

References

- ATKESON, A., AND L. E. OHANIAN (2001): “Are Phillips Curves Useful for Forecasting Inflation?,” *Quarterly Review*, pp. 2–11.
- CARTER, C. K., AND R. KOHN (1994): “On Gibbs Sampling for State Space Models,” *Biometrika*, 81, 541–553.
- COGLEY, T., G. E. PRIMICERI, AND T. J. SARGENT (2008): “Inflation-Gap Persistence in the U.S.,” NBER Working Paper 13749, NBER.
- COGLEY, T., AND T. J. SARGENT (2001): “Evolving Post WWII U.S. Inflation Dynamics,” in *Macroeconomics Annual*, pp. 331–373. MIT Press.
- (2005): “Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US,” *Review of Economic Dynamics*, 8, 262–302.
- PRIMICERI, G. E. (2005): “Time Varying Structural Vector Autoregressions and Monetary Policy,” *Review of Economic Studies*, 72, 821–852.
- STOCK, J. H., AND M. W. WATSON (2007): “Why Has U.S. Inflation Become Harder to Forecast?,” *Journal of Money, Credit and Banking*, 39, No.1.

Appendix 1

Estimation is done using Bayesian methods. We refer the reader to the Appendix for the detail, while here we describe the general line of the procedure. We make the following assumptions for priors densities. \hat{x} denotes the OLS estimate of parameter x obtained in the initial sample.

$$\begin{aligned}
 P(\theta_0) &= N(\hat{\theta}, \hat{V}_\theta) \\
 P(\phi_{i0}) &= N(\hat{\phi}_i, \hat{V}_{\phi_i}) \\
 P(\log \sigma_0) &= N(\log \hat{\sigma}, I_n) \\
 P(\Omega) &= IW(\Omega_0^{-1}, \rho_1) \\
 P(\Xi) &= IW(\Xi_0^{-1}, \rho_2) \\
 P(\Psi_i) &= IW(\Psi_0^{-1}, \rho_{3i})
 \end{aligned}$$

where $\bar{\Omega}_0^{-1} = (\lambda_1 \rho_1 \hat{V}_\theta)^{-1}$, $\Xi_0^{-1} = \lambda_2 \rho_2 I_n$ and $\Psi_0^{-1} = \lambda_{3i} \rho_{3i} \hat{\phi}_i, \hat{V}_{\phi_i}$.

Discuss parameters $\lambda \dots$

To draw from the joint posterior distribution of model parameters we use a Gibbs sampling algorithm along the lines described in Primiceri (2005). The basic idea of the algorithm is to draw sets of coefficients from known conditional posterior distributions. The algorithm is initialized⁹ at some values and, under some regularity conditions, the draws converge to a draw from the joint posterior after a burn in period. Let z be $(q \times 1)$ vector, we denote z^T the sequence $[z'_1, \dots, z'_T]'$. Each repetition is composed of the following steps:

1. $p(\sigma^T | x^T, \theta^T, \phi^T, \Omega, \Xi, \Psi, s^T)$
2. $p(s^T | x^T, \theta^T, \sigma^T, \phi^T, \Omega, \Xi, \Psi)^{10}$
3. $p(\phi^T | x^T, \theta^T, \sigma^T, \Omega, \Xi, \Psi, s^T)$
4. $p(\theta^T | x^T, \sigma^T, \phi^T, \Omega, \Xi, \Psi, s^T)$
5. $p(\Omega | x^T, \theta^T, \sigma^T, \phi^T, \Xi, \Psi, s^T)$
6. $p(\Xi | x^T, \theta^T, \sigma^T, \phi^T, \Omega, \Psi, s^T)$
7. $p(\Psi | x^T, \theta^T, \sigma^T, \phi^T, \Omega, \Xi, s^T)$

9

¹⁰See below the definition of s^T .

Gibbs sampling algorithm

- Step 1: sample from $p(\sigma^T | y^T, \theta^T, \phi^T, \Omega, \Xi, \Psi, s^T)$

To draw σ^T we use the algorithm of Kim, Shephard and Chibb (KSC) (1998). Consider the system of equations $y_t^* \equiv F_t^{-1}(y_t - X_t' \theta_t) = D_t^{1/2} u_t$, where $u_t \sim N(0, I)$, $X_t = (I_n \otimes x_t')$, and $x_t = [1_n, y_{t-1} \dots y_{t-p}]$. Conditional on y^T, θ^T , and ϕ^T , y_t^* is observable. Squaring and taking the logarithm, we obtain

$$y_t^{**} = 2r_t + v_t \quad (11)$$

$$r_t = r_{t-1} + \xi_t \quad (12)$$

where $y_{i,t}^{**} = \log((y_{i,t}^*)^2 + 0.001)$ - the constant (0.001) is added to make estimation more robust - $v_{i,t} = \log(u_{i,t}^2)$ and $r_t = \log \sigma_{i,t}$. Since, the innovation in (11) is distributed as $\log \chi^2(1)$, we use, following KSC, a mixture of 7 normal densities with component probabilities q_j , means $m_j - 1.2704$, and variances v_j^2 ($j=1, \dots, 7$) to transform the system in a Gaussian one, where $\{q_j, m_j, v_j^2\}$ are chosen to match the moments of the $\log \chi^2(1)$ distribution. The values are:

j	q_j	m_j	v_j^2
1.0000	0.0073	-10.1300	5.7960
2.0000	0.1056	-3.9728	2.6137
3.0000	0.0000	-8.5669	5.1795
4.0000	0.0440	2.7779	0.1674
5.0000	0.3400	0.6194	0.6401
6.0000	0.2457	1.7952	0.3402
7.0000	0.2575	-1.0882	1.2626

Let $s^T = [s_1, \dots, s_T]'$ be a matrix of indicator variables selecting at each point in time the member of the mixture to be used for each element of v_t . Conditional on s^T , $(v_{i,t} | s_{i,t} = j) \sim N(m_j - 1.2704, v_j^2)$. Therefore we can use the algorithm of Carter and R.Kohn (1994) to draw r_t ($t=1, \dots, T$) from $N(r_t | r_{t+1}, R_t | r_{t+1})$, where $r_t | r_{t+1} = E(r_t | r_{t+1}, y^t, \theta^T, \phi^T, \Omega, \Xi, \Psi, s^T)$ and $R_t | r_{t+1} = Var(r_t | r_{t+1}, y^t, \theta^T, \phi^T, \Omega, \Xi, \Psi, s^T)$.

- Step 2: sample from $p(s^T | y^T, \theta^T, \sigma^T, \phi^T, \Omega, \Xi, \Psi)$

Conditional on $y_{i,t}^{**}$ and r^T , we independently sample each $s_{i,t}$ from the discrete density defined by $Pr(s_{i,t} = j | y_{i,t}^{**}, r_{i,t}) \propto f_N(y_{i,t}^{**} | 2r_{i,t} + m_j - 1.2704, v_j^2)$, where $f_N(y | \mu, \sigma^2)$ denotes a normal density with mean μ and variance σ^2 .

- Step 3: sample from $p(\phi^T | y^T, \theta^T, \sigma^T, \Omega, \Xi, \Psi, s^T)$

Consider again the system of equations $F_t^{-1}(y_t - X_t'\theta_t) = F_t^{-1}\hat{y}_t = D_t^{1/2}u_t$. Conditional on θ^T , \hat{y}_t is observable. Since F_t^{-1} is lower triangular with ones in the main diagonal, each equation in the above system can be written as

$$\hat{y}_{1,t} = \sigma_{1,t}u_{1,t} \quad (13)$$

$$\hat{y}_{i,t} = -\hat{y}_{[1,i-1],t}\phi_{i,t} + \sigma_{i,t}u_{i,t} \quad i = 2, \dots, n \quad (14)$$

where $\sigma_{i,t}$ and $u_{i,t}$ are the i th elements of σ_t and u_t respectively, $\hat{y}_{[1,i-1],t} = [\hat{y}_{1,t}, \dots, \hat{y}_{i-1,t}]$. Under the block diagonality of Ψ , the algorithm of Carter and R.Kohn (1994) can be applied equation by equation, obtaining draws for $\phi_{i,t}$ from a $N(\phi_{i,t}|\phi_{i,t+1}, \Phi_{i,t}|\phi_{i,t+1})$, where $\phi_{i,t}|\phi_{i,t+1} = E(\phi_{i,t}|\phi_{i,t+1}, y^t, \theta^T, \sigma^T, \Omega, \Xi, \Psi)$ and $\Phi_{i,t}|\phi_{i,t+1} = Var(\phi_{i,t}|\phi_{i,t+1}, y^t, \theta^T, \sigma^T, \Omega, \Xi, \Psi)$.

- Step 4: sample from $p(\theta^T|y^T, \sigma^T, \phi^T, \Omega, \Xi, \Psi, s^T)$

Conditional on all other parameters and the observables we have

$$y_t = X_t'\theta_t + \varepsilon_t \quad (15)$$

$$\theta_t = \theta_{t-1} + \omega_t \quad (16)$$

Draws for θ_t can be obtained from a $N(\theta_t|\theta_{t+1}, P_{t|t+1})$, where $\theta_t|\theta_{t+1} = E(\theta_t|\theta_{t+1}, y^T, \sigma^T, \phi^T, \Omega, \Xi, \Psi)$ and $P_{t|t+1} = Var(\theta_t|\theta_{t+1}, y^T, \sigma^T, \phi^T, \Omega, \Xi, \Psi)$ are obtained with the algorithm of Carter and R.Kohn (1994).

- Step 5: sample from $p(\Omega|y^T, \theta^T, \sigma^T, \phi^T, \Xi, \Psi, s^T)$

Conditional on the other coefficients and the data, Ω has an Inverse-Wishart posterior density with scale matrix $\Omega_1^{-1} = (\Omega_0 + \sum_{t=1}^T \Delta\theta_t(\Delta\theta_t)')^{-1}$ and degrees of freedom $df_{\Omega_1} = df_{\Omega_0} + T$, where Ω_0^{-1} is the prior scale matrix, df_{Ω_0} are the prior degrees of freedom and T is length of the sample use for estimation. To draw a realization for Ω make df_{Ω_1} independent draws z_i ($i=1, \dots, df_{\Omega_1}$) from $N(0, \Omega_1^{-1})$ and compute $\Omega = (\sum_{i=1}^{df_{\Omega_1}} z_i z_i')^{-1}$ (see Gelman et. al., 1995).

- Step 6: sample from $p(\Xi_{i,i}|y^T, \theta^T, \sigma^T, \phi^T, \Omega, \Psi, s^T)$

Conditional the other coefficients and the data, Ξ has an Inverse-Wishart posterior density with scale matrix $\Xi_1^{-1} = (\Xi_0 + \sum_{t=1}^T \Delta \log \sigma_t (\Delta \log \sigma_t)')^{-1}$ and degrees of freedom $df_{\Xi_1} = df_{\Xi_0} + T$ where Ξ_0^{-1} is the prior scale matrix and df_{Ξ_0} the prior degrees of freedom. Draws are obtained as in step 5.

- Step 7: sample from $p(\Psi|y^T, \theta^T, \sigma^T, \phi^T, \Omega, \Xi, s^T)$.

Conditional on the other coefficients and the data, Ψ_i has an Inverse-Wishart posterior density with scale matrix $\Psi_{i,1}^{-1} = (\Psi_{i,0} + \sum_{t=1}^T \Delta\phi_{i,t}(\Delta\phi_{i,t})')^{-1}$ and degrees of freedom $df_{\Psi_{i,1}} = df_{\Psi_{i,0}} + T$ where $\Psi_{i,0}^{-1}$ is the prior scale matrix and $df_{\Psi_{i,0}}$ the prior degrees of freedom. Draws are obtained as in step 5 for all i .

6 Appendix 2: Tables

Table 1: *Relative Forecasting Performance for Inflation*

All Sample	Horizon											
	1	2	3	4	5	6	7	8	9	10	11	12
Naïve	2.15	1.95	2.03	2.24	2.52	2.77	2.94	3.06	3.17	3.25	3.30	3.31
AR-REC	1.13	1.19	1.17	1.17	1.18	1.17	1.18	1.19	1.21	1.22	1.25	1.28
AR-ROL	1.08	1.09	1.05	1.03	1.06	1.09	1.11	1.13	1.16	1.18	1.21	1.24
AR-TV	1.03	0.99	0.91	0.88	0.89	0.90	0.91	0.93	0.94	0.96	0.98	1.00
VAR-REC	1.15	1.26	1.31	1.37	1.42	1.46	1.53	1.60	1.68	1.76	1.84	1.93
VAR-ROL	1.01	1.15	1.22	1.22	1.25	1.28	1.33	1.38	1.44	1.48	1.54	1.60
VAR-TV	0.86	0.74	0.64	0.62	0.63	0.64	0.65	0.66	0.68	0.69	0.70	0.72
First Sample	Horizon											
1	2	3	4	5	6	7	8	9	10	11	12	
Naïve	4.02	4.03	4.43	4.98	5.67	6.21	6.62	6.89	7.10	7.20	7.15	6.98
AR-REC	1.11	1.15	1.15	1.16	1.17	1.17	1.18	1.19	1.21	1.23	1.28	1.35
AR-ROL	1.05	1.02	0.97	0.96	0.99	1.00	1.01	1.01	1.01	1.01	1.04	1.07
AR-TV	0.97	0.90	0.85	0.84	0.86	0.87	0.89	0.91	0.92	0.94	0.98	1.03
VAR-REC	1.09	1.16	1.20	1.25	1.29	1.32	1.37	1.42	1.49	1.53	1.56	1.62
VAR-ROL	0.89	0.99	1.05	1.02	1.03	1.06	1.12	1.16	1.21	1.26	1.33	1.42
VAR-TV	0.82	0.66	0.58	0.58	0.60	0.61	0.63	0.65	0.67	0.68	0.72	0.76
Second Sample	Horizon											
1	2	3	4	5	6	7	8	9	10	11	12	
Naïve	0.93	0.59	0.47	0.45	0.47	0.52	0.54	0.57	0.60	0.67	0.78	0.92
AR-REC	2.61	3.90	5.06	5.76	6.25	6.23	6.39	6.39	6.37	5.94	5.25	4.61
AR-ROL	1.19	1.40	1.51	1.54	1.63	1.72	1.89	2.09	2.26	2.34	2.24	2.10
AR-TV	1.21	1.37	1.30	1.16	1.13	1.13	1.11	1.08	1.09	1.07	0.96	0.86
VAR-REC	1.29	1.70	2.01	2.22	2.47	2.61	2.80	3.03	3.18	3.42	3.50	3.47
VAR-ROL	1.35	1.84	2.31	2.64	2.91	2.93	3.04	3.11	3.21	3.05	2.74	2.51
VAR-TV	0.98	1.06	1.01	0.91	0.88	0.86	0.82	0.77	0.76	0.71	0.60	0.52

Table 2: *Relative Forecasting Performance for Unemployment Rate*

All Sample	Horizon											
	1	2	3	4	5	6	7	8	9	10	11	12
Naïve	0.15	0.40	0.72	1.07	1.43	1.75	2.07	2.39	2.65	2.86	3.06	3.22
AR-REC	1.00	0.99	1.01	1.03	1.04	1.02	0.99	0.95	0.92	0.90	0.87	0.85
AR-ROL	1.08	1.12	1.17	1.24	1.24	1.19	1.14	1.14	1.14	1.12	1.12	1.12
AR-TV	1.00	0.97	0.99	1.01	1.02	1.01	0.98	0.95	0.93	0.90	0.88	0.86
VAR-REC	0.99	0.85	0.74	0.67	0.61	0.52	0.47	0.45	0.44	0.43	0.45	0.47
VAR-ROL	1.18	1.13	1.04	0.91	0.80	0.69	0.64	0.63	0.67	0.72	0.79	0.85
VAR-TV	1.02	0.92	0.84	0.78	0.76	0.68	0.64	0.62	0.60	0.57	0.57	0.59
First Sample	Horizon											
1	2	3	4	5	6	7	8	9	10	11	12	
Naïve	0.30	0.81	1.46	2.15	2.81	3.28	3.68	4.03	4.22	4.39	4.56	4.72
AR-REC	0.98	1.00	1.05	1.09	1.11	1.14	1.16	1.17	1.19	1.20	1.19	1.18
AR-ROL	1.06	1.11	1.18	1.26	1.28	1.24	1.23	1.28	1.33	1.34	1.37	1.41
AR-TV	0.98	0.99	1.03	1.06	1.09	1.11	1.13	1.14	1.17	1.18	1.18	1.18
VAR-REC	0.97	0.80	0.67	0.59	0.54	0.47	0.45	0.46	0.49	0.50	0.53	0.56
VAR-ROL	1.18	1.10	0.98	0.82	0.71	0.60	0.57	0.59	0.67	0.76	0.86	0.93
VAR-TV	1.02	0.91	0.82	0.76	0.75	0.67	0.64	0.64	0.64	0.61	0.62	0.65
Second Sample	Horizon											
1	2	3	4	5	6	7	8	9	10	11	12	
Naïve	0.05	0.13	0.23	0.37	0.53	0.75	1.02	1.33	1.63	1.87	2.08	2.25
AR-REC	2.80	3.06	3.13	3.00	2.79	2.38	2.01	1.72	1.50	1.38	1.28	1.22
AR-ROL	1.16	1.15	1.13	1.15	1.12	1.03	0.94	0.86	0.82	0.78	0.75	0.72
AR-TV	1.07	0.93	0.84	0.82	0.80	0.70	0.62	0.56	0.52	0.48	0.45	0.43
VAR-REC	1.09	1.07	1.03	0.97	0.85	0.68	0.53	0.42	0.35	0.33	0.33	0.35
VAR-ROL	1.17	1.29	1.27	1.23	1.08	0.93	0.81	0.72	0.66	0.67	0.69	0.73
VAR-TV	1.02	0.96	0.90	0.88	0.82	0.73	0.63	0.57	0.52	0.50	0.51	0.50

Table 3: *Relative Forecasting Performance for Interest Rate*

All Sample	Horizon											
	1	2	3	4	5	6	7	8	9	10	11	12
Naïve	0.87	1.72	2.48	3.46	4.49	5.72	6.80	7.54	8.18	9.04	9.81	10.28
AR-REC	1.12	1.08	1.07	1.05	1.04	1.05	1.06	1.05	1.05	1.06	1.06	1.08
AR-ROL	1.23	1.18	1.22	1.20	1.17	1.17	1.18	1.18	1.18	1.18	1.16	1.15
AR-TV	1.04	1.02	0.98	0.95	0.93	0.93	0.93	0.92	0.91	0.91	0.91	0.91
VAR-REC	0.99	0.97	0.98	0.96	0.98	0.99	1.00	0.99	1.00	1.01	1.02	1.03
VAR-ROL	1.09	1.14	1.37	1.39	1.45	1.43	1.44	1.44	1.42	1.36	1.33	1.32
VAR-TV	0.97	0.96	0.95	0.92	0.91	0.90	0.89	0.88	0.86	0.85	0.85	0.84
First Sample	Horizon											
	1	2	3	4	5	6	7	8	9	10	11	12
Naïve	1.80	3.22	4.18	5.55	7.06	8.96	10.49	11.18	11.81	12.97	13.90	14.26
AR-REC	1.15	1.13	1.15	1.15	1.15	1.18	1.21	1.22	1.25	1.28	1.31	1.35
AR-ROL	1.26	1.19	1.25	1.22	1.18	1.20	1.24	1.28	1.32	1.35	1.34	1.36
AR-TV	1.09	1.10	1.06	1.04	1.02	1.03	1.04	1.05	1.05	1.06	1.10	1.14
VAR-REC	1.02	1.02	1.08	1.07	1.12	1.16	1.19	1.22	1.26	1.28	1.29	1.31
VAR-ROL	1.11	1.21	1.51	1.50	1.58	1.56	1.58	1.62	1.56	1.47	1.46	1.47
VAR-TV	1.01	1.01	1.02	0.99	0.98	0.98	0.97	0.97	0.97	0.98	1.00	1.04
Second Sample	Horizon											
	1	2	3	4	5	6	7	8	9	10	11	12
Naïve	0.27	0.74	1.37	2.09	2.82	3.61	4.39	5.16	5.82	6.47	7.15	7.69
AR-REC	3.64	2.52	1.94	1.74	1.66	1.67	1.63	1.53	1.47	1.47	1.46	1.44
AR-ROL	1.08	1.13	1.16	1.17	1.16	1.13	1.09	1.05	1.00	0.96	0.92	0.89
AR-TV	0.83	0.81	0.81	0.81	0.79	0.77	0.76	0.74	0.73	0.70	0.67	0.63
VAR-REC	0.87	0.81	0.79	0.78	0.75	0.72	0.70	0.67	0.65	0.66	0.67	0.70
VAR-ROL	1.02	0.96	1.11	1.20	1.23	1.22	1.21	1.20	1.24	1.21	1.15	1.13
VAR-TV	0.82	0.80	0.81	0.81	0.80	0.78	0.76	0.74	0.72	0.69	0.64	0.61

7 Figures

Figure 1: *Inflation Predictions: Two Years Ahead*

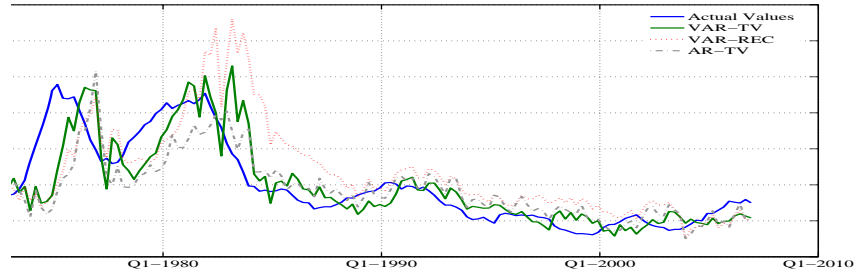


Figure 2: *Unemployment Predictions: Two Years Ahead*

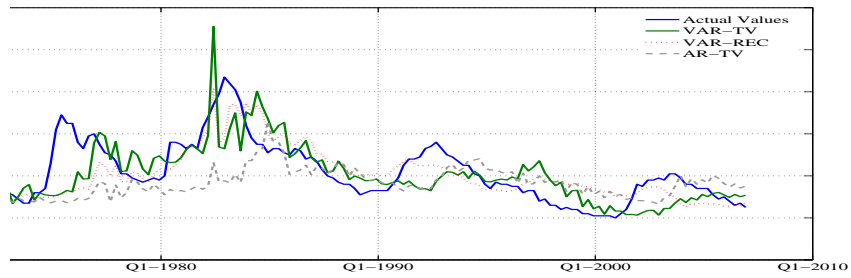


Figure 3: *Interest Rate Predictions: Two Years Ahead*

