International Journal of Forecasting 30 (2014) 129–143

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting UK GDP growth and inflation under structural change. A comparison of models with time-varying parameters

Alina Barnett [a], Haroon Mumtaz [b,*], Konstantinos Theodoridis [c]

[a] MPC unit, Bank of England, United Kingdom
[b] Center for Central Banking Studies, Bank of England, United Kingdom
[c] Monetary Assessment and Strategy Division, Bank of England, United Kingdom

ARTICLE INFO

Keywords:
Time-varying parameters
Regime switching
Vector autoregressions
Forecast comparison

ABSTRACT

Evidence from a large and growing body of empirical literature strongly suggests that there have been changes in the inflation and output dynamics in the United Kingdom. The majority of these papers base their results on a class of econometric models that allows for time-variation in the coefficients and volatilities of shocks. While these models have been used extensively for studying evolving dynamics and for structural analysis, there has been little evidence that they are useful for forecasting UK output growth and inflation. This paper attempts to fill this gap by comparing the performances of a wide range of time-varying parameter models in forecasting output growth and inflation. We find that allowing for time-varying parameters can lead to large and statistically significant gains in forecast accuracy.

1. Introduction

A large and growing body of literature has proposed and applied a number of empirical models that incorporate the possibility of structural shifts in the model parameters. The series of papers by Tom Sargent and his co-authors on the evolving dynamics of US inflation forms an often-cited example of this literature. In particular, Cogley, Primiceri, and Sargent (2008) and Cogley and Sargent (2002, 2005) use time-varying parameter VARs (TVP-VAR) to explore the possibility of shifts in inflation dynamics, with Benati (2007) applying this methodology to the modelling of the temporal shifts in UK macroeconomic dynamics. In contrast, Sims and Zha (2006) model changes in US macroeconomic dynamics using a regime-switching VAR (see Groen & Mumtaz, 2008, for an application to the United Kingdom). Balke (2000) highlights potential non-linearities in the output and inflation dynamics, and uses threshold VAR (TVAR) models to explore non-linear dynamics in output and inflation. Recent papers have estimated time-varying factor augmented VAR (TVP-FAVAR) models in order to incorporate more information into the empirical model. For example, Baumeister, Liu, and Mumtaz (2013) argue that incorporating a large information set can be important when modelling changes in the monetary transmission mechanism, and use a TVP-FAVAR to estimate the evolving response to US monetary policy shocks.

Most of this literature has focused on describing the evolution of macroeconomic dynamics. In contrast, studies on the forecasting ability of these models have been more limited in both number and scope. D'Agostino, Gambetti, and Giannone (2013) focus on TVP-VARs only, and show that they provide more accurate forecasts of US inflation and unemployment than fixed-coefficient VARs. In a recent contribution, Eickmeier, Lemke, and Marcellino (2011) present a comparison of the forecasting performances of

* Corresponding author.
E-mail addresses: Alina.Barnett@bankofengland.co.uk (A. Barnett), haroon.mumtaz@bankofengland.co.uk, h.mumtaz@qmul.ac.uk (H. Mumtaz), Konstantinos.Theodoridis@bankofengland.co.uk (K. Theodoridis).

0169-2070/$ – see front matter
http://dx.doi.org/10.1016/j.ijforecast.2013.06.002

the TVP-FAVAR, its fixed-coefficient counterpart, and AR models with time-varying parameters for US data over the period 1995–2007. The authors show that there are some gains (in terms of forecasting performances) from allowing time-variation in the model parameters and from exploiting a large information set.

The aim of this paper is to extend the forecast comparison exercises of D'Agostino et al. (2013) and Eickmeier et al. (2011) along two dimensions. First, our paper compares the forecast performances of a much wider range of multi-variate models with time-varying parameters. In particular, we compare the forecasting performances of (a range of) regime-switching models, TVP-VARs, TVP-FAVARs, TVARs, smooth transition VARs (ST-VARs), the unobserved component model with stochastic volatility proposed by Stock and Watson (2007), rolling VARs and recursive VARs. This extends the analysis of D'Agostino et al. (2013), where the focus is on the performances of TVP-VAR and TVP-AR models relative to those of fixed coefficient VARs;[1] our work instead provides a direct comparison of TVP-VARs with empirical models that provide an alternative specification for time-variation. Similarly, the focus of our application is broader than that of Eickmeier et al. (2011). Eickmeier et al. (2011) only consider an AR model with time-varying parameters as an alternative TVP model, whereas we compare the TVP-FAVAR model with a range of *multi-variate* forecasting models with time-varying and regime switching parameters. In addition, the forecast comparison is carried out recursively over the period 1976 Q1–2007 Q4, and thus covers a longer period than that of Eickmeier et al. (2011). Second, while previous papers have largely focused on the United States, we work with UK data and try to determine whether these time-varying parameter models are useful for forecasting UK inflation, GDP growth and the short-term interest rate. This question is highly relevant for policy, as the United Kingdom has experienced large changes in the dynamics of key macro variables over the last three decades. In addition, the recent financial crisis has been associated with large movements in inflation and output growth, again highlighting the possibility of structural change. Note also that the focus of our analysis differs from that of the analyses of Eklund, Kapetanios, and Price (2010) and Clark and McCracken (2009a). While these papers have focused largely on forecasting performances under structural change in a Monte Carlo setting, our exercise is a direct application to UK data using the time-varying parameter models that are currently popular in empirical work.[2]

The forecast comparison exercise brings out the following main results:

- On average, a VAR model estimated over rolling samples delivers the most accurate forecasts for GDP growth at the one-year forecast horizon, with a root mean squared error (RMSE) which is 6% lower than that of an AR(p) model. The ST-VAR and TVP-VAR models deliver similar performances.
- Models with time-varying parameters provide substantially better inflation forecasts. At the one-year horizon, the TVP-FAVAR model has an average RMSE which is 14% lower than that of an AR(p) model, and is the best performing model over the full forecast sample, which indicates the roles played by time-varying parameters and a large information set. The TVP-VAR model and Stock and Watson's unobserved component model also perform well, with the latter delivering the most accurate one-year-ahead forecasts over the post-1992 period.

The paper is organised as follows. Section 2 provides details of the data used in the study and describes the real time out-of-sample forecasting exercise. Section 3 describes the main forecasting models used in this study. Section 4 describes the main results in detail. Finally, Section 5 concludes.

2. Data and forecasting methodology

2.1. Data

Our main UK dataset consists of data on quarterly annualised real GDP growth, quarterly annualised inflation and the three month treasury bill rate. Quarterly data on these variables are available from 1955Q1 to 2010Q4.

The GDP growth series is constructed using real time GDP data obtained from the Office of National Statistics. Vintages of GDP data covering our sample period are available from 1976Q1 onwards, and these are used in our forecasting exercise as described below. GDP growth is defined as 400 times the log difference of GDP.

The inflation series is based on the seasonally adjusted harmonised index of consumer prices spliced with the retail price index excluding mortgage payments. These data are obtained from the Bank of England database. Inflation is calculated as 400 times the log difference of this price index. The three month treasury bill rate is obtained from Global Financial Data.

2.2. Forecasts and evaluation

The forecasting models (described in Section 3) are estimated recursively. The estimation starts from the initial sample from 1955Q1 to 1975Q4 and proceeds by adding one quarter of data at a time and re-estimating the forecasting model. Note that, in the case of GDP growth, we add a new vintage of GDP data (i.e., the vintage available in that quarter) at each iteration of this recursive estimation. The forecasting models are estimated recursively $R = 129$ times until 2007Q4.

At each iteration, we forecast GDP growth, inflation and the three-month treasury bill rate up to 12 quarters ahead. For models with time-varying parameters, we assume that the parameters are fixed over the forecast horizon when calculating forecasts. That is, following D'Agostino et al. (2013), the forecasts are calculated recursively using the last estimated value of the parameters, and do not account for parameter variation over the forecast period. An economic justification for this assumption is provided by Cogley and Sargent (2008).

[1] D'Agostino et al. (2013) compare the forecast performances of TVP-VARs and TVP-AR models with those of fixed coefficient VARs that allow for stochastic volatility and fixed coefficient VARs that are estimated using rolling or recursive windows.

[2] Faust and Wright (2011) compare the performances of a large number of models for forecasting US inflation. However, they do not focus exclusively on models with time-varying parameters. Ferrara, Marcellino, and Mogliani (2012) also compare the forecasting performances of a range of forecasting models, but limit their attention to single equation models.

The forecasts are evaluated via the following univariate and multivariate criteria.

Root mean squared error

In particular, we use the root mean squared error (RMSE), calculated as

$$RMSE = \sqrt{\sum_{t=T+1}^{T+h} \frac{\left(\hat{Z}_t - Z_t\right)^2}{h}}, \quad (1)$$

where $T+1, T+2, \ldots, T+h$ denotes the forecast horizon, $\hat{Z}_t$ denotes the forecast, and Z_t denotes the actual data. For GDP growth, the forecast error $\hat{Z}_t - Z_t$ is calculated using the latest available vintage.[3] We estimate the RMSE for $h = 1, 4, 8$ and 12 quarters.

In order to compare the performances of the different forecasting models, we use the RMSE of each model relative to a benchmark model: an AR(p) model estimated recursively over each subsample. We allow for the possibility that $p = 1 \ldots 4$, and choose the lag length that maximises the marginal likelihood at each date in the sample period.

Diebold–Mariano statistic

As a formal test of whether the predictive accuracy delivered by the non-linear models considered in this study is superior to that obtained using the AR(p) regression estimated recursively over each subsample via OLS, we use the statistic developed by Diebold and Mariano (1995).[4] The accuracy of each forecast is measured by using the squared error loss function: $L\left(\hat{Z}_t^i, Z_t\right) = \left(\hat{Z}_t^i - Z_t\right)^2$, where $t = T + 1, \ldots, T + R$ and R is the length of the forecast evaluation sample. Under the null hypothesis, the expected forecast loss from using one model instead of the other is the same:

$$H_o : E\left[L\left(\hat{Z}_t^i, Z_t\right)\right] = E\left[L\left(\hat{Z}_t^{AR}, Z_t\right)\right], \quad (2)$$

which can be rejected in favor of the alternative (the non-linear model delivers forecasts which are superior to those from the AR(p) model) if:

$$DM = \sqrt{R}\frac{\frac{1}{R}\sum_{t=T+1}^{T+R} d_t}{\hat{\sigma}_d} < c_\alpha, \quad (3)$$

where $d_t = \left(\hat{Z}_t^i - Z_t\right)^2 - \left(\hat{Z}_t^{i,AR} - Z_t\right)^2$, c_α is the critical value of the statistic at the significance level α, and $\hat{\sigma}_d^2$ is the heteroskedasticity and autocorrelation consistent variance estimator developed by Newey and West (1987). The information that the h-step-ahead forecast error follows a moving average process of order $h - 1$ is used to decide about the bandwidth of the kernel.

The estimation and evaluation process in this paper is carried out using real-time data. As was explained by Clark and McCracken (2009b), the distribution of the DM statistic is not standard normal under these circumstances. Since the null hypothesis can be imposed easily in our case by subtracting the sample mean $\left(\bar{d} = \frac{1}{R}\sum_{t=1}^{R} d_t\right)$, we derive the critical values of the DM statistic using moving block resampling techniques (see Goncalves & White, 2002, 2004). To be precise, the series

$$\tilde{d}_t = d_t - \bar{d} \quad (4)$$

is split into N blocks of length equal to $l = h - 1$, where $R = Nl$. These block are resampled with replacement, and the new pseudo data, $\tilde{d}_t^*$, are used to calculate the test statistic DM^*. The whole process is repeated $B = 999$ times, to deliver the bootstrapped distribution of the statistic under the null hypothesis of equal predictive accuracy ($\{DM_j^*\}_{j=1}^B$), which is used to calculate the critical value for the 10% significance level. Similar bootstrapping techniques which are employed to derive the critical values for equal predictive accuracy test statistics are given by Clark and McCracken (2011, chap. 14).[5]

Recursive Diebold–Mariano statistic

Motivated by the studies of Giacomini and Rossi (2010) and Rossi and Sekhposyan (2010), we also considered the recursive version of the DM statistic discussed in the previous section:

$$DM_t^{REC} = \sqrt{r}\frac{\frac{1}{r}\sum_{t=T+1}^{T+r} d_t}{\hat{\sigma}_d} < c_{\alpha,t}, \quad r = 1, \ldots, R. \quad (5)$$

This statistic is used to identify any possible changes in the forecasting performances of the competing models at different points over the forecasting sample, 1976–2007. Since the dynamic properties of the macroeconomic data vary over time, it seems fair to suspect that there will be variations in the models' abilities to forecast the future evolution of the macroeconomic series. For instance, there may be periods (perhaps during severe recessions and recoveries) where the forecasts derived by a nonlinear model are more accurate than those obtained by an $AR(p)$ model. The studies by Giacomini and Rossi (2010) and Rossi and Sekhposyan (2010) provide empirical evidence to support this view.

The critical values $c_{\alpha,t}$ (which now vary over time) are again constructed using the bootstrapping procedure considered in the previous section.

Trace statistic

In addition, we also calculate the trace of the forecast error covariance matrix (Ω) in order to assess the multivariate performances of the competing models. Consider the singular value decomposition of $\Omega = V\Lambda V'$, where V is the matrix of eigenvectors and Λ is the diagonal matrix with eigenvalues in descending order. The eigenvalues are the variances of the principal components, and the trace of Ω

[3] Qualitatively similar results are obtained if the comparison is made with the vintage published three years after the forecast is made. A table with these estimates is included in Appendix B.

[4] Note that the Diebold–Mariano (DM) statistic is calculated for the entire sample, not for each point in time like the RMSE. Furthermore, the DM statistic will coincide with the RMSE only if the forecasting horizon equals one. Finally, there could be cases when the DM test may be unable to distinguish between models even when there are quite large reductions in RMSE.

[5] The bootstrap procedure discussed here ignores parameter estimation uncertainty. Unfortunately, taking the estimation uncertainty into account is not a computationally feasible option here, given the computational cost of estimating the models considered in this study.

equals the sum of their eigenvalues. Based on these observations, Adolfson, Linde, and Villani (2007) argue that the trace will be determined to a large extent by the forecasting performances of the least predictable dimensions (largest eigenvalues). It should be mentioned that this statistic has its limitations. For instance, Clements and Hendry (1995) point out that the model ranking based on this statistic is affected by linear transformations of the forecasting variables. However, this is not the case in our exercise, since all of the variables are expressed in percentage terms.

3. Forecasting models

In this section we describe the forecasting models used in this study.

3.1. Regime switching VAR

In order to account for the possibility of structural shifts, we model inflation, output and the interest rate dynamics using a regime switching VAR of the following form:

$$Z_t = c_{S_t} + \sum_{j=1}^{K} B_{S_t} Z_{t-j} + \Omega_{H_t}^{1/2} \varepsilon_t \tag{6}$$

where Z_t is a $T \times 3$ data matrix that contains GDP growth, inflation and the interest rate, and B_S and Ω_h are regime dependent autoregressive coefficients and reduced-form variance covariance matrices. The VAR model allows for M breaks at unknown dates; as per Chib (1998), these are modelled via the latent state variable S_t for the VAR coefficients, and H_t for the error covariance matrix. In our most general regime switching model, the state variables S and H are assumed to evolve independently, with their transition governed by a first order Markov chain with $M + 1$ regimes, with restricted transition probabilities $p_{ij} = p\left(S_t = j | S_{t-1} = i\right)$ and $q_{ij} = p\left(H_t = j | H_{t-1} = i\right)$. The transition probability matrices are defined as

$$\begin{aligned} &p_{ij}, q_{ij} > 0 \quad \text{if } i = j \\ &p_{ij,} q_{ij} > 0 \quad \text{if } j = i + 1 \\ &p_{MM}, q_{MM} = 1 \\ &p_{ij}, q_{ij} = 0 \quad \text{otherwise.} \end{aligned} \tag{7}$$

For example, if $M = 3$, the transition matrices are defined as

$$\tilde{P} = \begin{pmatrix} p_{11} & 0 & 0 & 0 \\ 1 - p_{11} & p_{22} & 0 & 0 \\ 0 & 1 - p_{22} & p_{33} & 0 \\ 0 & 0 & 1 - p_{33} & 1 \end{pmatrix} \quad \text{and}$$

$$\tilde{Q} = \begin{pmatrix} q_{11} & 0 & 0 & 0 \\ 1 - q_{11} & q_{22} & 0 & 0 \\ 0 & 1 - q_{22} & q_{33} & 0 \\ 0 & 0 & 1 - q_{33} & 1 \end{pmatrix}.$$

Eqs. (6) and (7) define a Markov switching VAR with non-recurrent states, where transitions are allowed in a sequential manner. For example, to move from regime 1 to regime 3, the process has to visit regime 2. Similarly, transitions to past regimes are not allowed. However, this structure is not necessarily more restrictive than a standard Markov switching model; it simply implies that any new regimes are given new labels, rather than being linked to past states explicitly (as with a standard Markov switching model). This formulation implies that the regimes are identified by assumption, and no "label switching" problem exists when implementing the Gibbs sampler. This feature offers a clear computational advantage (relative to a regime switching VAR with unrestricted transition probabilities) by removing the need for regime normalisation, which can be computationally challenging as the number of regimes grows.[6]

We estimate three versions of this regime switching model: (i) the general switching model as set out in Eq. (6), which allows for independent breaks in the VAR coefficients and error covariance; (ii) a version of the regime switching VAR where the breaks in the VAR coefficients and the covariance matrix are restricted to occur jointly; and (iii) a version of the regime switching VAR where only the VAR coefficients allow breaks. Specification (ii) is estimated in order to determine whether allowing for different timings in the variance and coefficient breaks offers any advantage in terms of forecasting performance. Specification (iii), which does not include volatility breaks, is included in order to shed light on the role played by heteroscedasticity. In each case, we allow for up to 3 breaks or four regimes, and a lag length K of up to 4. The optimal numbers of regimes and lags for each model are chosen at each date in the sample by maximising the marginal likelihood. The computation of the marginal likelihood via the Chib (1995) method is described below, with details given in Appendix B.

Various different versions of this regime switching model have been used in a number of recent studies for describing the changing dynamics of key macroeconomic time series. For example, Sims and Zha (2006) argue that a model which incorporates regime switching dynamics provides a good description of the evolution of monetary policy and inflation dynamics in the US. Groen and Mumtaz (2008) provide a similar analysis for the UK, and show that a regime switching VAR is useful for describing the change in inflation persistence. It may also be argued that allowing for discrete shifts in the coefficients and error variances is especially appropriate, given the current crisis and its associated impact on macroeconomic variables.

The models are estimated using a Gibbs sampling algorithm. The prior distributions and conditional posteriors are described in Appendix B. Note that we employ a normal inverse Wishart prior on the VAR parameters in each regime. However, as is described in Appendix B, the tightness parameters are set to large values, rendering the prior distributions non-informative.

3.2. Time-varying VAR

In a recent paper, D'Agostino et al. (2013) showed that a VAR with time-varying parameters and stochastic volatility performs well in forecasting US macroeconomic data.

[6] This model is similar in spirit to the model considered by Pesaran, Pettenuzzo, and Timmermann (2006).

In addition, a voluminous body of literature has used the time-varying VAR model to investigate the possibility of a temporal shift in UK and US inflation dynamics. Prominent examples of papers that employ this model for the US include those of Cogley et al. (2008) and Cogley and Sargent (2002, 2005). Benati (2007) and Mumtaz and Sunder-Plassmann (2013) use the time-varying VAR model to capture the time-varying dynamics of UK macroeconomic and financial time series. Relative to the regime switching model, the time-varying VAR incorporates a more flexible specification for the time-varying parameters. In particular, it allows the time-variation in each VAR equation to be independent.

We use a general version of this model as a forecasting model for UK GDP growth, inflation and interest rates. In particular, we employ the following specification:

$$Z_t = c_t + \sum_{j=1}^{K} B_t Z_{t-j} + \Omega_t^{1/2} \varepsilon_t, \tag{8}$$

where the VAR coefficients $\Phi_t = vec(\{c_t, B_t\})$ evolve as random walks

$$\Phi_t = \Phi_{t-1} + \eta_t.$$

As per Cogley and Sargent (2005), the covariance matrix of the innovations v_t is factored as

$$VAR(v_t) \equiv \Omega_t = A_t^{-1} H_t (A_t^{-1})'. \tag{9}$$

The time-varying matrices H_t and A_t are defined as:

$$H_t \equiv \begin{bmatrix} h_{1,t} & 0 & 0 \\ 0 & h_{2,t} & 0 \\ 0 & 0 & h_{3,t} \end{bmatrix} \quad \text{and} \quad A_t \equiv \begin{bmatrix} 1 & 0 & 0 \\ \alpha_{21,t} & 1 & 0 \\ \alpha_{31,t} & \alpha_{32,t} & 1 \end{bmatrix}, \tag{10}$$

with the $h_{i,t}$ evolving as geometric random walks,

$$\ln h_{i,t} = \ln h_{i,t-1} + \tilde{\nu}_t.$$

Following Primiceri (2005), we postulate that the non-zero and non-one elements of the matrix A_t will evolve as driftless random walks,

$$\alpha_t = \alpha_{t-1} + \tau_t. \tag{11}$$

In our first specification, we consider a version of this TVP-VAR which was estimated recently by Baumeister and Benati (2010) and Cogley et al. (2008). These authors generalise the specification adopted by Cogley and Sargent (2005) to allow for stochastic volatility in η_t. In particular, we model $var(\eta_t) = Q_t$ as

$$Q_t = \tilde{A}^{-1} \tilde{H}_t \left(\tilde{A}^{-1}\right)', \tag{12}$$

where $\tilde{A}$ is a lower triangular matrix, while $\tilde{H}_t = diag\left(\tilde{h}_{1t} \dots \tilde{h}_{Mt}\right)$ with $M = N \times (N \times L + 1)$ and $\tilde{h}_{jt}$ evolving as

$$\ln \tilde{h}_{jt} = \ln \tilde{h}_{jt-1} + \tilde{u}_t. \tag{13}$$

As was discussed by Baumeister and Benati (2010), one advantage of this extended specification is that it allows for the possibility that the degree of time-variation may vary over the sample. For example, it accounts for the possibility that the VAR coefficients may change faster during crisis periods, while the degree of parameter drift will be smaller over tranquil periods. We consider two additional restricted specifications. First, following D'Agostino et al. (2013), we estimate a time-varying VAR with a constant degree of parameter drift, i.e., $Q_t = Q$. Second, in order to explore the role played by heteroscedasticity, we estimate a TVP-VAR with $\Omega_t = \Omega$ as per Cogley and Sargent (2002).

The models are estimated using a Gibbs sampling algorithm. The distributions of the priors and the conditional posteriors are described in Appendix B. We point out three aspects. First, the prior for Q (and for $\tilde{H}_0$ in the specification which allows for a time-varying Q) is set using a pre-sample of $T_0 = 40$ quarters. In particular, let Q_{OLS} denote the OLS estimate of the coefficient covariance matrix using the training sample. When Q is time-invariant, its prior distribution is assumed to be Inverse Wishart with a scale matrix given by $\bar{Q} = Q_{OLS} \times T_0 \times k$, where the scalar $k = 3.5e-04$, following Cogley and Sargent (2005). The prior degrees of freedom are set equal to $T_0 = 40$, the length of the training sample. When Q is time-varying, a prior distribution is required for the initial values of $\tilde{H}_t$. The mean of this log normal prior is set to be the log of $diag(A_{OLS} Q_{OLS} A'_{OLS}) \times T_0 \times k$, where A_{OLS} is the inverse of the Choleski decomposition of Q_{OLS}. The variance of the prior distribution is set to 1. On a log scale, this represents an agnostic prior about the initial value of $\tilde{H}_t$. Finally, note that we allow the lag length K to take a maximum value of 2, and select the model (at each date in the sample) that produces the largest value of the marginal likelihood. Larger values of K are not considered in our application, as they increase the computational burden considerably. In particular (as is standard in the literature), we employ rejection sampling in order to ensure that the VAR coefficients are stable at each point in the sample. This becomes increasingly time consuming to implement as K becomes larger. Given that the Gibbs sampling algorithm is implemented more than 100 times for each value of K, computational speed is a key concern in our application.

3.3. Time-varying factor augmented VAR

Recent empirical work on the evolving monetary transmission mechanism has employed time-varying factor augmented VAR (TVP-FAVAR) models as a way of incorporating additional information into the empirical specification (see for example Baumeister et al., 2013; Eickmeier et al., 2011, and the references therein). As was shown by Eickmeier et al. (2011), these models therefore offer a convenient way of combining a large information set and time-varying dynamics. As was shown by Eickmeier et al. (2011) for the US, the time-varying FAVAR delivers a forecasting performance which is superior to those of both its fixed coefficient counterpart and time-varying models that do not include information from a large data set. We estimate the following TVP-FAVAR model:

$$X_t = \beta F_t + e_{it} \tag{14}$$

$$F_t = c_t + \sum_{j=1}^{K} B_t F_{t-j} + \Omega_t^{1/2} \varepsilon_t,$$

where $X_t = [x_{it}, z_t]$, x_{it} is a $T \times M$ matrix of macroeconomic and financial variables, and z_t is the variable we are interested in forecasting, here GDP growth, inflation or the 3-month treasury bill rate. The matrix F_t contains the d latent factors that summarise the information in the panel x_{it} and z_t; that is, $z_t = [f_{1t}, \ldots f_{dt}, z_t]$. β denotes the factor loading matrix, while e_{it} represents the idiosyncratic component. We allow for first order serial correlation in e_{it}, with $e_{it} = \rho_i e_{it-1} + v_{it}$. More details on the observation equation are provided by Baumeister et al. (2013). The dynamics of F_t are described by a time-varying VAR model with stochastic volatility. The coefficients of this transition equation evolve as random walks (the shock on the random walk has a fixed variance). The variance of the shocks is specified as in Eq. (9).

The model is estimated using a Gibbs sampling algorithm. This algorithm is an extended version of the sampler used for the time-varying VAR, and is described in Appendix B. Note that the priors for the hyper-parameters of the transition equation are set as was described in the previous section. Appendix A describes the data set x_{it} used for the forecasting exercise. To summarise, x_{it} contains 43 variables that represent data on real activity, inflation, money supply, interest rates and exchange rates. This dataset is chosen because it is available consistently over the sample period used in our forecasting exercise.

We also consider two restricted versions of the model in Eq. (14). First, we fix $\Omega_t = \Omega$ and only allow time-variation in the coefficients of the transition equation. Second, we estimate a fixed coefficient FAVAR. These restricted models are used to gauge the role of time-varying parameters (in addition to the impact of the larger information set) in driving any change in forecasting performance.

For the time-varying FAVAR models, we allow for $K = 1, 2$ and $d = 2, 3$, and choose the model that maximises the marginal likelihood at each iteration of the recursive estimation procedure. For the fixed coefficient FAVAR model (where the computational burden is less severe), the specification search involves models estimated for $K = 1, 2, 3, 4$ and $d = 2, 3, 4$.

3.4. Unobserved component model with stochastic volatility

In a recent contribution, Stock and Watson (2007) show that a univariate unobserved component (UC) model with stochastic volatility performs well in forecasting US inflation. Following Stock and Watson (2007), we consider this model as a possible alternative specification for forecasting UK data. The UC model with stochastic volatility is given by:

$$\begin{aligned}
\tilde{Z}_t &= \beta_t + \sqrt{\sigma_t}\varepsilon_t \\
\beta_t &= \beta_{t-1} + \sqrt{\varpi_t} v_t \\
\ln \sigma_t &= \ln \sigma_{t-1} + e_{1t}, \quad var\,(e_{1t}) = g_1 \\
\ln \varpi_t &= \ln \varpi_{t-1} + e_{2t}, \quad var\,(e_{2t}) = g_2,
\end{aligned}$$

where $\tilde{Z}_t$ contains data on GDP growth, inflation or the 3-month treasury bill yield. The model is estimated using a MCMC algorithm. Note that, while Stock and Watson (2007) calibrate g_1 and g_2 for US data, we estimate these parameters (see Appendix B).

3.5. Threshold and smooth transition VAR models

Threshold and smooth transition VAR models allow for different VAR parameters in different regimes. In this case, the switching mechanism is intuitive and simple, making these models very attractive and, consequently, popular. In addition (unlike in regime switching and time-varying parameter models), the time-variation in the parameters is linked explicitly to a threshold variable. In other words, parameters are allowed to be different in expansions and recessions, periods of high and low inflation, and periods of high and low interest rates. While regime switching and time-varying models can allow for this possibility, the parameter change in these models is governed by a more general process.

These models can be expressed as follows:

$$Z_t = \sum_{m=1}^{M} \xi_{m,t} \left\{ c_m + \sum_{j=1}^{K} B_{j,m} + \Omega_m^{1/2} \varepsilon_t \right\}. \tag{15}$$

The state variable $\xi_{m,t}$ in the threshold case is a discrete variable that takes values of 1 or 0 according to the following rule:

$$\xi_{m,t} = \begin{cases} 1 & \text{if } Z_{i,t-d} \leq c_1 \\ 1 & \text{if } c_1 < Z_{i,t-d} \leq c_2 \\ \vdots & \\ 1 & \text{if } c_{M-2} < Z_{i,t-d} \leq c_{M-1} \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

where $Z_{i,t}$ is variable i of the Z vector, M is the number of regimes, and $c = (c_1, \ldots, c_{M-1})'$ is the vector of threshold values. In the smooth transition case, $\xi_{m,t}$ is a continuous variable given by

$$\xi_{m,t} = \begin{cases} 1 - \dfrac{1}{1 + \exp\left(-\gamma\left(Z_{i,t-d} - c_1\right)\right)} \\ \dfrac{1}{\exp\left(-\gamma\left(Z_{i,t-d} - c_1\right)\right)} \\ \quad - \dfrac{1}{\exp\left(-\gamma\left(Z_{i,t-d} - c_2\right)\right)} \\ \vdots \\ \dfrac{1}{\exp\left(-\gamma\left(Z_{i,t-d} - c_{M-2}\right)\right)} \\ \quad - \dfrac{1}{\exp\left(-\gamma\left(Z_{i,t-d} - c_{M-1}\right)\right)} \\ \dfrac{1}{1 + \exp\left(-\gamma\left(Z_{i,t-d} - c_{M-1}\right)\right)}. \end{cases} \tag{17}$$

In our exercise, i, M, d and K are all selected using data driven techniques. To be precise, we create a four dimensional grid ($i = 1, 2, 3, M = 2, \ldots, 4, d = 1, \ldots, 4$ and $K = 1, \ldots, 4$) and calculate the marginal likelihood of the threshold and smooth transition VAR model for each of these models. We select the threshold variable, the number of regimes (M^*), the delay parameter (d^*) and the number of lags (K^*) that maximise the marginal likelihood over this grid. This is done at each iteration of the recursive forecasting exercise.

$$\left(M^*, i^*, d^*, K^*\right)' = \underset{M=2,\dots,4, i=1,\dots 3, d=1,\dots,4, K=1,\dots,4}{\arg\max} \mathcal{M}_{TVAR}(Y). \quad (18)$$

If we assume conjugate priors for the VAR parameters, then, conditional on γ and c, the posterior distribution of the VAR coefficient vector is the conditional Normal-Wishart distribution. Unfortunately, the posterior distributions of γ and c conditional on the VAR parameter vector are unknown, meaning that we have to employ both the Gibbs and Metropolis-Hasting samplers in order to derive the full posterior distribution of the entire estimated parameter vector (Chen, 1998; Chen & Lee, 1995; Lopes & Salazar, 2006).

3.6. Rolling and recursive VAR models

Our final two forecasting models are based on the following VAR:

$$Z_t = c + \sum_{j=1}^{K} BZ_{t-j} + \Omega^{1/2}\varepsilon_t, \quad (19)$$

where Z_t is a $T \times 3$ data matrix that contains GDP growth, inflation and the interest rate. The recursive VAR is estimated recursively from 1976Q1 until the end of the sample period. The rolling VAR model uses a 10-year rolling window to estimate the model parameters. From an applied point of view, the main virtue of these models is the fact that they are simple to estimate. Thus, a finding that these models forecast well relative to the more sophisticated alternatives would be of practical importance. Note that we allow for the possibility that K varies between 1 and 4, and select the lag length by maximising the marginal likelihood at each date in the estimation sample.

3.7. Model averaging

We also consider the question of whether the average forecast from our forecasting models can improve upon the individual forecasts presented above. In particular, we combine the forecasts using Bayesian Model Averaging (BMA):

$$\hat{Z}_{t,BMA} = \sum_{m=1}^{14} \hat{Z}_{t,m} P(Z_t \setminus m), \quad (20)$$

where $\hat{Z}_{t,BMA}$ denotes the BMA forecast at time t, $\hat{Z}_{t,m}$ denotes the forecast from model m and $P(Z_t \setminus m)$ is the marginal likelihood.

The calculation of the marginal likelihood in Eq. (20) is the key task when estimating $\hat{Z}_{t,BMA}$. Following Chib (1995), we consider the following representation for the log marginal likelihood:

$$\ln P(Z_t \setminus m) = \ln F(Z_t \setminus \hat{\Xi}, m) + \ln p\left(\hat{\Xi}\right) - \ln G\left(\hat{\Xi} \setminus Z_t\right), \quad (21)$$

where $\ln F(Z_t \setminus \Xi, m)$ is the log likelihood, $\ln p\left(\hat{\Xi}\right)$ is the log prior density, and $\ln G\left(\hat{\Xi} \setminus Z_t\right)$ is the log posterior density, with all three terms evaluated at the posterior mean for the model parameters $\hat{\Xi}$. The prior density $\ln p\left(\hat{\Xi}\right)$ is easy to evaluate. Similarly, the log likelihood of the models we consider can be evaluated either directly or via non-linear filters. The final term $\ln G\left(\hat{\Xi} \setminus Z_t\right)$ requires more work. Following Chib (1995) and Chib and Jeliazkov (2001), we proceed by factorising $\ln G\left(\hat{\Xi} \setminus Z_t\right)$ into conditional and marginal densities of various parameter blocks and using additional Gibbs and Metropolis runs to approximate these densities. Details are provided in Appendix B.

In addition to BMA, we also consider forecast combination, where the weights are based on the past forecasting performances of the forecasting models. Following Stock and Watson (2004), we define these alternative weights computed at each date t as

$$w_{it} = \frac{M_{mt}^{-1}}{\sum_{m=1}^{14} M_{mt}^{-1}}, \quad (22)$$

with

$$M_{mt} = \sum_{s=t_0}^{t-h} \delta^{t-h-s} \left(\hat{Z}_{s+h} - Z_{s+h}\right)^2, \quad (23)$$

where t_0 denotes the start of the forecasting sample, the forecast horizon $h = 1$, and $\delta = 0.95$ is a discount factor that ensures that more recent forecast performances are assigned higher weights.[7]

For comparison, we also consider a simpler approach to forecast combination and report the performance of the average forecast $\hat{Z}_{t,average} = \frac{1}{14}\sum_{m=1}^{14} \hat{Z}_{t,m}$. As has been demonstrated in previous studies (see for example Stock & Watson, 2003), the simple average can provide a forecasting performance which is on par with those of more sophisticated approaches.

4. Results

4.1. Overall forecast performance

4.1.1. GDP growth

Consider the relative RMSE for GDP growth shown in Table 1. Over the full sample, 1976–2007, the rolling VAR model outperforms the other forecasting models at the one- and four-quarter horizons, with forecasting gains of 13% and 6% respectively. Note that this difference in forecasting performance is estimated to be significant at the 10% level at the one-quarter-ahead horizon. The rolling VAR is closely followed by the ST-VAR and the TVP-VAR in terms of forecasting performances over these forecasting horizons, while the unobserved component model performs well at the one-quarter horizon. The regime switching VAR with joint shifts in the coefficients and covariance and the recursive VAR also display gains relative to the AR(p) model at this horizon.[8] The TVP-VAR

[7] The forecasting results are similar for $h = 4$ and $\delta = 0.9$. These are available on request.

[8] Note that we use multiple equation regime switching/threshold models. Previous papers that have used single equation versions of these models suggest that their forecast performances are poor (see for instance Dacco & Satchell, 1999).

Table 1
GDP.

Models	1976–2007				1992–2007			
	1Q	4Q	8Q	12Q	1Q	4Q	8Q	12Q
RSVAR	1.198	1.076	1.079	1.077	1.238	1.084	1.080	1.064
RSVAR+	0.929	0.965	1.011	1.026	1.232	1.064	1.074	1.074
RSVAR++	1.055	1.089	1.098	1.098	1.093	1.024	1.048	1.049
TVP-VAR (general)	1.684	1.411	1.316	1.263	1.107	1.042	1.074	1.087
TVP-VAR	0.905	0.952	0.985	0.994	0.958	0.948	0.990	0.996
TVP-VAR (homoscedastic)	0.915	0.965	0.988	0.994	1.083	1.011	1.025	1.025
TVP-FAVAR	1.044	1.055	1.038	1.034	1.134	1.135	1.107	1.100
TVP-FAVAR (homoscedastic)	1.034	1.069	1.061	1.064	1.270	1.198	1.157	1.147
FAVAR	2.090	1.631	1.328	1.154	1.599	1.146	0.869	0.716
UC	0.922*	1.010	1.027	1.027	1.084	1.076	1.063	1.043
TVAR	0.989	1.049	1.082	1.099	1.054	0.993	1.047	1.065
ST-VAR	0.905	0.952	0.989	0.996	1.008	0.950	0.992	1.011
VAR (rolling)	0.865	0.939	0.997	1.013	0.921	0.977	1.051	1.052
VAR (recursive)	0.967	0.967	0.993	1.003	1.046	0.986	1.005	1.009
BMA	1.374	1.242	1.160	1.126	1.124	1.138	1.104	1.101
Simple average	0.874	0.938	0.974	0.984	0.904	0.914	0.959	0.971
MSFE average	0.863*	0.926	0.966	0.978	0.902	0.912	0.958	0.969

Notes:
RSVAR: Regime-switching VARs with independent switching in the coefficients and the covariance.
RSVAR+: Regime-switching VARs with joint switching in the coefficients and the covariance.
RSVAR++: Regime-switching VARs with only the coefficients switching.
TVP-VAR (general): Time-varying VAR with a time-varying degree of parameter drift.
TVP-VAR (standard): Time-varying VAR with a constant degree of parameter drift.
TVP-VAR (homoscedastic): Time-varying VAR with a constant variance-covariance matrix of the VAR residuals.
TVP-FAVAR: Time-varying factor augmented VAR model.
TVP-FAVAR (homoscedastic): Time-varying factor augmented VAR model with constant residual variance.
FAVAR: Factor augmented VAR model.
UC: Unobserved component model with stochastic volatility.
TVAR: Threshold VAR model.
ST-VAR: Smooth transition VAR model.
VAR (rolling): VAR estimated with a rolling constant window.
VAR (recursive): VAR estimated recursively.
BMA: The Bayesian weighted forecasting model.
Simple average: A simple average of the forecasts from all 14 models.
MSFE average: Combination forecast using weights based on past MSFEs.
* indicates a significant improvement over the benchmark at the 10% level, according to the DM statistic.

produces the most accurate forecast at longer horizons, but the gains relative to the AR(p) model are small.

The rolling VAR model continues to display the best one-quarter-ahead forecasting performance (in terms of relative RMSEs) over the great moderation period. The TVP-VAR model delivers the most accurate one-year-ahead GDP growth forecast over this period, while the FAVAR model performs well at longer horizons.

It is interesting to note that the simple average of the fourteen forecasts produces a lower relative RMSE than the BMA forecast over all forecasting horizons. This relatively poor performance of the BMA forecast supports the analysis by Eklund and Karlsson (2007), who highlight the danger of in-sample overfitting when using the marginal likelihood to compute the combination weights.[9] The performance of the MSFE weighted forecast is similar to that of the simple average, with the former delivering a marginally smaller relative RMSE. The MSFE weighted forecast also outperforms the BMA forecast, pointing to the superior performance of weights based on past forecast accuracies over those based on the marginal likelihood.

[9] The factors behind the good forecasting performance of the simple average (observed in several studies) have been investigated by Smith and Wallis (2009).

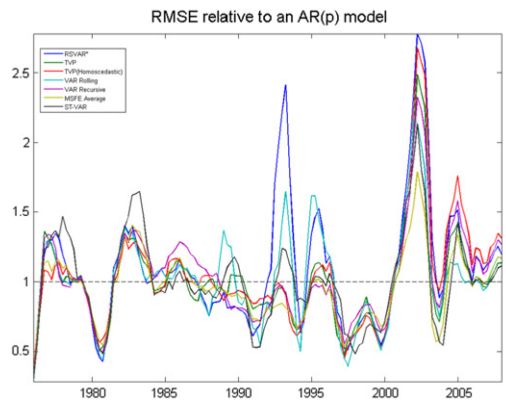

Fig. 1. Four-quarter moving average of the one-year-ahead RMSE relative to an AR(p) model for GDP growth over time.

Over the post-1993 period, the MSFE weighted forecast leads to the lowest relative RMSE over the one- and four-quarter horizons.

Fig. 1 explores the evolution of the (smoothed) one-year-ahead relative RMSEs of some of the best performing models over the forecasting period. The rolling VAR model does especially well over the pre-1990 period, though its performance then deteriorates over the early- to

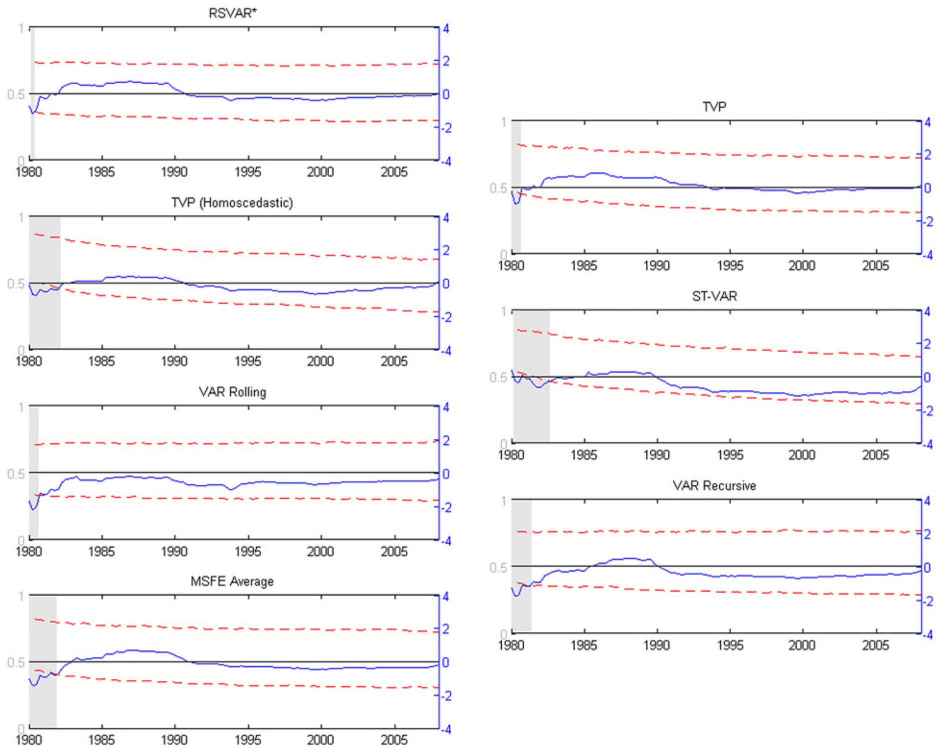

Fig. 2. Recursive DM statistics for one-year-ahead GDP growth forecasts from selected models.

mid-1990s. In contrast, the TVP-VAR and the ST-VAR perform well over this period. The performances of all of the models deteriorate in the early 2000s, with the deterioration being smallest for the rolling VAR and ST-VAR models. In the spirit of Rossi and Sekhposyan (2010), Fig. 2 plots the recursively calculated Diebold and Mariano (1995) statistic (blue line) and the corresponding 10% bootstrap critical values (red dotted line) for each of these models. A negative value of the statistic that is beyond the critical values indicates a significant improvement over the benchmark model—the shaded area in each figure highlights periods in which this is the case.[10] The estimates in Fig. 2 seem to confirm the general finding that the empirical models do not offer statistically significant gains over the AR(p) model *on average* at the four-quarter forecasting horizon. However, it is interesting to note that the early 1980s appears to be a period during which these models perform significantly better. This period was characterised by the deepest recession in our forecasting sample (covering the period from 1976 to 2007), perhaps suggesting that time-variation and regime switching can be useful during these volatile periods.

4.1.2. Inflation

Relative RMSEs for inflation are reported in Table 2. At the one-quarter forecast horizon, the homoscedastic TVP-FAVAR, the TVP-FAVAR and the TVP-VAR deliver the most accurate forecasts for inflation, on average. The improvement over the AR(p) benchmark is starker at the four-quarter horizon. At this horizon, the homoscedastic TVP-FAVAR delivers the most accurate forecast, with a RMSE 15% lower than the benchmark. The p-value for the Diebold and Mariano (1995) test suggests that this difference is statistically significant. The TVP-FAVAR model with stochastic volatility delivers a similar performance at this horizon, with gains over the benchmark of 10%. Note, however, that the inflation forecasts from the fixed coefficient FAVAR model are substantially less accurate. In addition, while the small scale TVP-VAR outperforms the AR(p) model, the gains are much more modest. These results suggest that the extra information included in the factor model *and* the presence of time-varying coefficients together lead to an improvement in the accuracy of inflation forecasts. The TVP-FAVAR models also deliver the most accurate forecasts over longer horizons.

Over the period of the great moderation, the rolling VAR model and Stock and Watson's unobserved component model deliver the best forecasting performances (in terms of the RMSE) at the one- and four-quarter horizons, with the relative RMSEs of the TVP-VAR and the homoscedastic TVP-FAVAR models being estimated to be similar. At the four-quarter horizon, a number of other models also display lower RMSEs than the AR(p) model. For example, forecasts from the regime switching VAR (with switching coefficients only), the general TVP-VAR, and ST-VAR have RMSEs which are 10%–20% lower than those of the benchmark. The fact that a wider range of models perform better over the great moderation period results in an improvement in the relative RMSE of the combination

10 The figure shows the recursive statistic and the critical values from 1980Q1 onwards, as the estimates over the first few years are unreliable, given the small numbers of observations.

Table 2
Inflation.

Models	1976–2007				1992–2007			
	1Q	4Q	8Q	12Q	1Q	4Q	8Q	12Q
RSVAR	1.232	1.180	1.199	1.146	1.108	0.992	0.924	0.887
RSVAR+	1.244	1.340	1.375	1.339	1.005	0.917	0.868*	0.828*
RSVAR++	1.264	1.245	1.232	1.159	0.902	0.887*	0.836*	0.799*
TVP-VAR (general)	1.177	1.083	1.040	0.973	0.904	0.782*	0.701*	0.656*
TVP-VAR	0.955	0.975	0.941	0.887	0.861*	0.737*	0.626*	0.569*
TVP-VAR (homoscedastic)	1.087	1.031	0.988	0.937	0.911*	0.826*	0.747*	0.705*
TVP-FAVAR	0.959	0.892	0.823*	0.764*	1.056	0.814*	0.663*	0.593*
TVP-FAVAR (homoscedastic)	0.924	0.855*	0.780*	0.726*	0.993	0.709*	0.562*	0.496*
FAVAR	1.063	1.020	1.006	0.984	1.106	0.910*	0.863*	0.856*
UC	1.028	0.981	0.943	0.903	0.849	0.701*	0.583*	0.520*
TVAR	1.057	1.124	1.142	1.071	0.994	0.930*	0.805*	0.706*
ST-VAR	1.036	1.115	1.097	1.026	0.939	0.836	0.738	0.680*
VAR (rolling)	1.194	1.185	1.171	1.101	0.840*	0.715*	0.615*	0.554*
VAR (recursive)	1.154	1.312	1.348	1.304	0.970	1.015	0.990	0.960
BMA	1.044	0.981	0.919	0.841*	1.047	0.803*	0.665*	0.599*
Simple average	0.974	0.991	0.982	0.934	0.859*	0.764*	0.680*	0.633*
MSFE average	0.989	0.992	0.976	0.927	0.862*	0.760*	0.670*	0.622*

Notes as for Table 1.

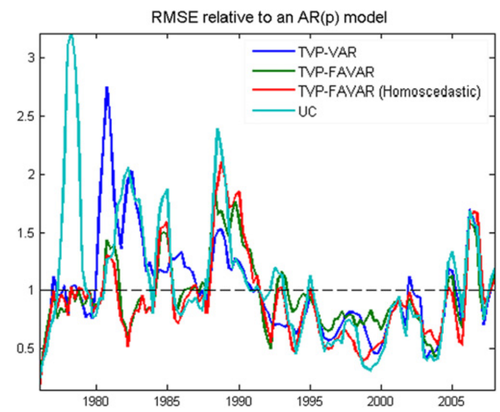

Fig. 3. Four-quarter moving average of the one-year-ahead RMSE relative to an AR(p) model for inflation over time.

forecasts. The Diebold and Mariano (1995) test provides stronger evidence that these forecasting models provide significantly more accurate forecasts than the benchmark over this sub-sample.

In Fig. 3, we examine the evolution of the smoothed relative RMSEs (at the four-quarter horizon) of the best performing inflation forecasting models. It is interesting to note that there is a distinct change in relative RMSEs after the early 1990s, with the inflation targeting period being characterised by improved performances by the four time-varying parameter models relative to the benchmark. Note that several studies have documented a change in the dynamics of UK inflation at this juncture (see for example Benati, 2007). Our results suggest that models with evolving parameters were able to adapt to this change better than recursively estimated fixed coefficient models. During the pre-1990 period, the TVP-FAVAR models have the lowest relative RMSEs, especially during the late 1970s and the early 1980s, when the other forecasting models in Fig. 3 appeared to be inaccurate relative to the benchmark.

In Fig. 4, we plot the recursive Diebold and Mariano (1995) statistic for these models. The homoscedastic TVP-FAVAR model consistently delivers forecasts that offer an improvement over the AR(p) model. The TVP-FAVAR model performs significantly better than the benchmark over the period of the great moderation. In contrast, the TVP-VAR and the UC model do not show any significant improvement over the AR(p) over the recursive sample.

Overall, our results for inflation point to the usefulness of time-varying parameters and a large information set in delivering accurate inflation forecasts.

4.1.3. Short term interest rates

The forecasting models perform poorly on average over the full forecasting period (see Table 3). At the four-quarter horizon, only the recursive VAR model improves upon the benchmark, albeit marginally. Over the one- and four-quarter horizons, the simple average of all forecasting models delivers the largest improvement relative to the benchmark.

In contrast, when considering the great moderation period, the differences between the performances of some of the forecasting models and that of the AR(p) benchmark are larger. At the one- and four-quarter horizons, the homoscedastic TVP-VAR model produces the most accurate interest rate forecasts, leading to a greater than 20% reduction in RMSE relative to the AR(p) model (with a significant Diebold–Mariano test statistic). Over this period, the TVAR model is the best performing model at longer horizons.

4.2. Model-specific results

In this subsection, we consider forecasting performances across different specifications of the estimated models. We focus on the estimated trace of the forecast error covariance matrix reported in Table 4 as an overall measure of forecasting performance. The table presents the estimated trace relative to that obtained using the AR(p) model. Consider, first, the regime switching models. If the focus is on the one-quarter horizon, then the trace of the forecast error covariance matrix suggests that allowing for independent regime shifts in the VAR coefficients and the error covariance matrix offers little benefit. At this horizon, the best forecasting performance (i.e., lowest trace) is

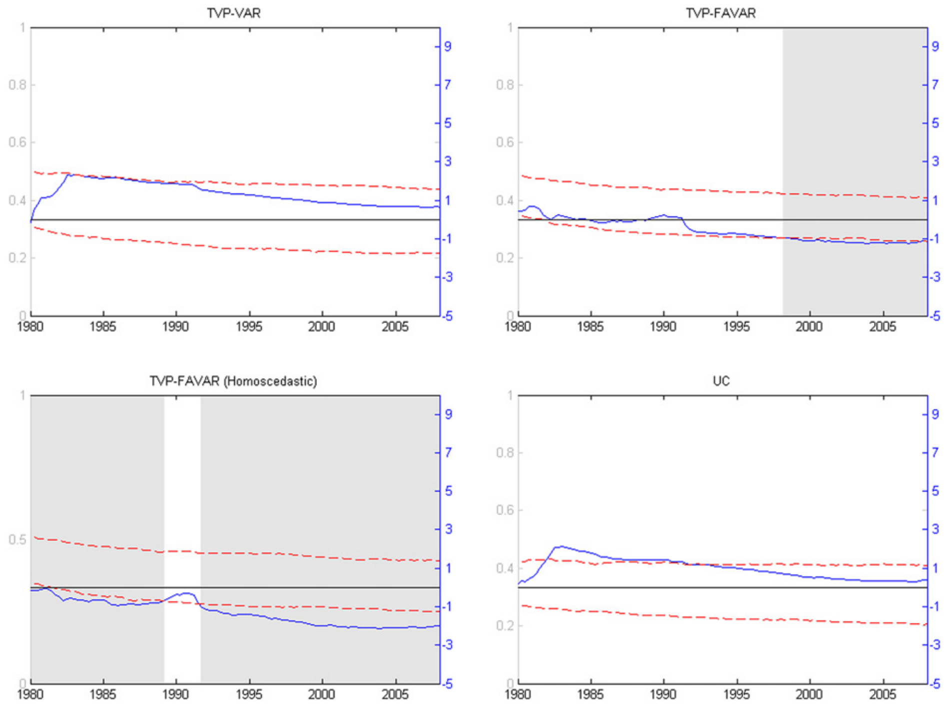

Fig. 4. Recursive DM statistics for one-year-ahead inflation forecasts from selected models.

Table 3
Policy rates.

Models	1976–2007				1992–2007			
	1Q	4Q	8Q	12Q	1Q	4Q	8Q	12Q
RSVAR	1.197	1.086	1.044	1.006	0.872	0.908	0.889*	0.838*
RSVAR+	1.104	1.008	0.976	0.933	1.054	1.083	1.059	0.993
RSVAR++	1.180	1.112	1.076	1.041	0.981	1.003	1.029	1.022
TVP-VAR (general)	1.173	1.263	1.276	1.234	0.999	1.107	1.111	1.058
TVP-VAR	1.015	1.020	1.005	0.981	0.756*	0.827*	0.811*	0.736*
TVP-VAR (homoscedastic)	1.076	1.067	1.056	1.033	0.737*	0.796*	0.813*	0.789*
TVP-FAVAR	1.094	1.095	1.104	1.083	0.875	0.914	0.860*	0.776*
TVP-FAVAR (homoscedastic)	1.018	1.015	1.020	1.003	0.899	0.898	0.886*	0.831*
FAVAR	2.490	1.972	1.741	1.605	3.631	2.559	2.109	1.884
UC	1.168	1.062	1.038	0.989	0.962	0.950	0.895*	0.803*
TVAR	1.036	1.093	1.057	1.024	0.937*	0.901*	0.806*	0.735*
ST-VAR	1.008	1.064	1.016	0.974	0.839*	0.876	0.875	0.836
VAR (rolling)	1.073	1.021	0.972	0.921*	0.902	0.971	0.968	0.873*
VAR (recursive)	1.003	0.991	0.985	0.982	0.968	0.967*	0.950*	0.939*
BMA	1.099	1.150	1.161	1.118	0.921	0.955	0.891*	0.809*
Simple average	0.997	0.973	0.956	0.925*	0.802*	0.861*	0.846*	0.797*
MSFE average	1.004	0.975	0.955*	0.923*	0.793*	0.856*	0.840*	0.791*

Notes as for Table 1.

delivered by the specification that imposes a common regime variable on the VAR coefficients and the covariance matrix. At longer horizons, however, the model with independent regime shifts delivers the lowest trace amongst the regime switching models.

From a forecast accuracy point of view, the general TVP-VAR proposed by Baumeister and Benati (2010) and Cogley et al. (2008) (which allows for a time-varying Q matrix) displays the least favorable performance of all of the TVP-VARs considered in this study. The best performance, on the basis of the trace of the forecast error covariance matrix and over all forecast horizons, is delivered by the standard TVP-VAR that allows for stochastic volatility. Note that the homoscedastic TVP-VAR has a somewhat larger trace than its heteroscedastic counterpart at all forecast horizons, suggesting that allowing for stochastic volatility helps to improve the forecasting performance within these models. One possible explanation for this finding is implied by the analysis of Stock (2001), who argues that ignoring the heteroscedasticity in TVP-VARs may bias the degree of time-variation. One can conjecture that this may have a potential negative impact on the forecasting performance.

According to the trace criteria, the TVP-FAVAR with homoscedastic shocks delivers the best forecast performance

Table 4
Trace statistic.

Models	1976–2007				1992–2007			
	1Q	4Q	8Q	12Q	1Q	4Q	8Q	12Q
RSVAR	1.577	1.245	1.244	1.079	1.531	0.934	0.877	0.741
RSVAR+	0.900	1.709	1.826	1.834	1.222	0.894	0.773	0.685
RSVAR++	1.233	1.643	1.438	1.313	0.961	0.881	0.843	0.779
TVP-VAR (general)	3.003	1.756	1.480	1.298	0.943	0.819	0.720	0.649
TVP-VAR	0.782	1.047	1.010	0.904	0.691	0.608	0.513	0.420
TVP-VAR (homoscedastic)	0.950	1.158	1.085	0.968	0.914	0.751	0.695	0.606
TVP-FAVAR	0.822	1.064	1.006	0.928	1.124	0.785	0.516	0.415
TVP-FAVAR (homoscedastic)	0.932	0.942	0.841	0.765	1.144	0.738	0.512	0.416
FAVAR	4.584	3.812	2.615	2.358	8.404	3.306	2.087	1.783
UC	1.015	1.126	0.986	0.875	0.917	0.702	0.461	0.355
TVAR	0.949	1.213	1.275	1.143	1.044	0.854	0.692	0.558
ST-VAR	0.931	1.127	1.236	1.199	0.894	0.939	1.010	0.988
VAR (rolling)	0.840	1.513	1.455	1.346	0.720	0.733	0.563	0.413
VAR (recursive)	0.941	1.642	1.664	1.560	0.924	1.078	1.023	0.950
BMA	1.728	1.258	1.080	0.952	1.095	0.778	0.518	0.421
Simple average	0.726	1.048	1.007	0.903	0.661	0.659	0.592	0.512
MSFE average	0.702	1.020	0.959	0.891	0.659	0.646	0.579	0.498

Notes as for Table 1.

(within the FAVAR models) at the four-, eight- and twelve-quarter horizons. Note also that this model performs better than all of the other competing models at these forecasting horizons (and over the full forecast sample). This again brings out the influence of both time-varying parameters *and* a large information set on forecast performance.

Over most forecast horizons (and on average over the full forecast period), the ST-VAR model has a lower trace than its threshold counterpart. This provides some evidence that allowing for smooth transitions rather than abrupt regime shifts improves the forecasting performance for this dataset.

The rolling VAR model appears to perform better than the recursively estimated VAR model. The rolling VAR has a lower trace (on average over the period 1976–2007) at all forecast horizons.

Finally, the simple and MSFE average combination forecasts outperform the BMA forecast. This difference in the performances of the simple average and the BMA is especially stark at forecast horizons of up to one year.

4.3. *Forecast performances and the recent financial crisis*

In this subsection, we consider how the forecasts from our models perform when considering the period 2008Q1–2010Q4, a period over which the financial crisis intensified. In particular, we consider how the forecasting models perform given the information set at 2007Q4. Fig. 5 plots the one- and four-step-ahead recursive forecasts for these variables from the fourteen forecasting models (and the two combination forecasts plus forecasts from the benchmark model) alongside the actual realised values over these quarters.

Consider the one-step-ahead GDP forecast in the top left panel. One striking feature of this panel is that most models predicted a zero or positive growth over the second half of 2008 and 2009, when the actual growth was strongly negative. In 2008Q4, the actual annualised GDP growth was −3.6%. The homoscedastic TVP-FAVAR was the only model to predict negative growth in this quarter, albeit with a forecast which was much higher than the realised value. In 2009Q1, negative growth was predicted by the TVP-FAVAR, FAVAR and UC models. However, as before, the prediction of GDP growth by these models was substantially more optimistic than the actual growth of −8%. It took until the third quarter of 2009 for the average one-step-ahead forecast to approach the actual realised growth rate. The top right panel of the figure shows that a similar interpretation can be placed on the four-step-ahead GDP forecasts. The benchmark AR(p) model performed relatively poorly over this period, indicating positive growth throughout the crisis.

The middle panel of Fig. 5 shows the one- and four-step-ahead forecasts for inflation. In 2008Q4, the annualised quarterly inflation was 5.5%. It then dropped to 0.8% in the next quarter, before reaching a trough at 0.1%. Inflation then rose to around 4% by mid-2010. Most of the forecasting models failed to predict (at the one-quarter horizon), the large drop in inflation between 2008Q4 and 2009Q1. Of all of the forecasting models, the largest drop in inflation over this quarter is predicted by Stock and Watson's unobserved component model; however, its projection of 3.5% inflation is still substantially higher than the actual. The rolling VAR and the homoscedastic TVP-FAVAR model predict values of inflation which are closest to the actual in 2009Q2. Both the TVAR and ST-VAR models are able to predict the 2009Q3 bounce-back in inflation, with the TVP-FAVAR models 'catching-up' in Q4. At the four-quarter forecast horizon, most forecasting models predicted higher than actual inflation over 2008 and the first half of 2009, then underpredicted inflation over the second half of 2009 and 2010Q1.

As with the one-step-ahead inflation forecast, the forecasting models have a hard time in matching the sharp fall in the short term interest rate in the last quarter of 2008 (bottom left panel). However, after 2009Q2, the models predict interest rates of close to zero. The bottom right panel shows that the four-quarter forecasts are quite far away from the true realised values, possibly reflecting the atheoretical nature of the forecasting models.

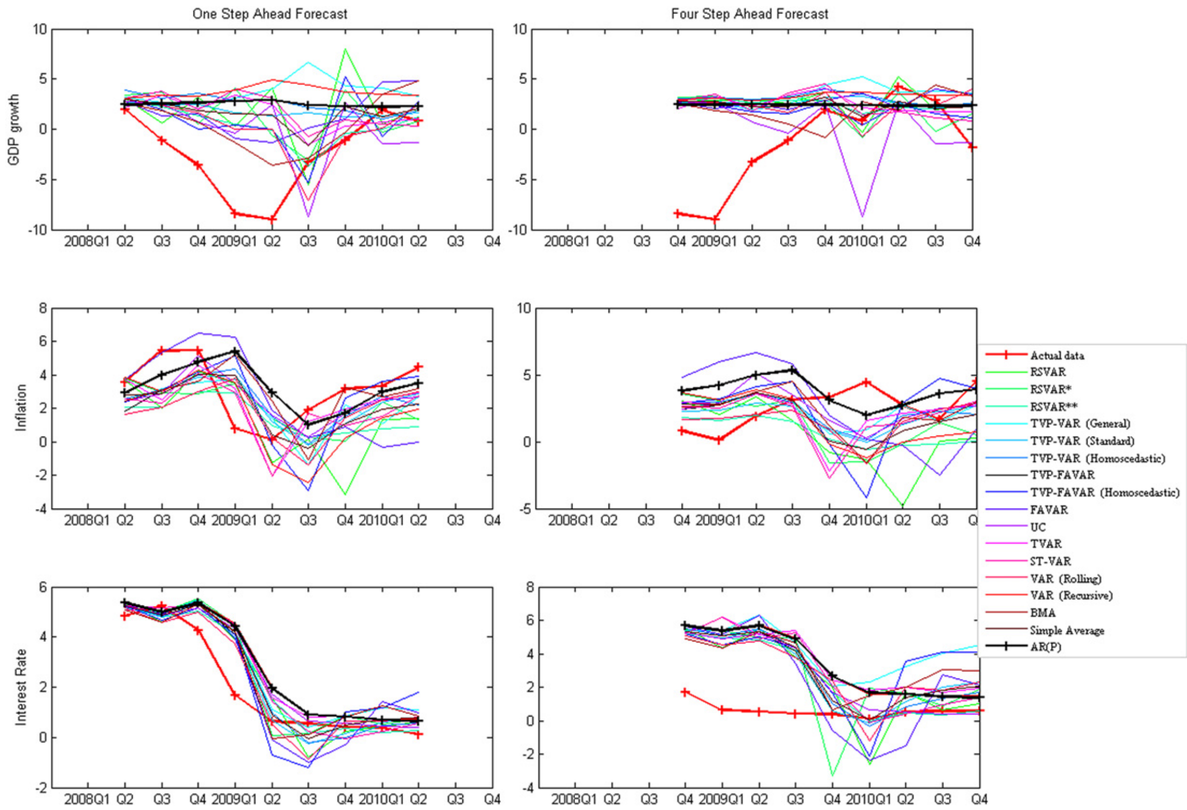

Fig. 5. Forecasts over the crisis period.

5. Conclusions

This paper investigates the performances of a variety of models with time-varying parameters in forecasting UK GDP growth, inflation and the short-term interest rate. Our results suggest that the rolling VAR, TVP-VAR and ST-VAR models all provide forecasts of GDP growth with lower average RMSEs than an AR(p) model.

Models with time-varying parameters lead to large improvements in inflation forecasting performances over the AR(p) benchmark. In particular, the TVP-FAVAR model, the TVP-VAR and the UC model all have RMSEs that are substantially lower than that of the benchmark model.

The TVP-VAR and the TVAR stand out when considering the interest rate forecasts, but, in general, it appears that the models considered in this study are less successful at forecasting interest rates than for either GDP growth or inflation.

Across the three variables, the TVP-FAVAR model stands out as delivering the most accurate forecasts on average over the sample at the four-quarter horizon. This highlights the role played by a large information set and time-varying parameters in delivering accurate forecasts.

In future research, it would be of interest to investigate whether models with time-varying parameters also perform well using US and Euro-area data. In addition, it would also be useful to consider the performances of DSGE models that incorporate switching parameters.

Acknowledgments

The authors thank Charlotte Dendy, Laurent Ferrara and the participants at the 7th ECB workshop on forecasting techniques for useful comments. The paper also benefited from comments by Michael Clements and two anonymous referees.

Appendix A. Data for the FAVAR models

The data set used to estimate the FAVAR models is listed in Table A.1. Note that, when estimating the model for forecasting inflation, we include GDP growth and the short-term interest rate in x_{it} (see Eq. (14)) along with the variables in Table A.1. When estimating the model for forecasting GDP growth, we include inflation and the interest rate in x_{it} along with the variables in Table A.1. Similarly, GDP growth and inflation are added to the panel when estimating the model for forecasting interest rates.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2013.06.002.

Table A.1
Data used to estimate the FAVAR model.

Variable no.	Variable name	Source	Transformation
1	General government: Final consumption expenditure	ONS	Log difference
2	ESA95 output index: F: Construction	ONS	Log difference
3	Total exports	ONS	Log difference
4	Total imports	ONS	Log difference
5	Gross fixed capital formation	ONS	Log difference
6	IOP: Manufacturing	ONS	Log difference
7	SA95 output index: Transport, storage & communication	ONS	Log difference
8	SA95 output index: Total	ONS	Log difference
9	ESA95 output index: Distribution, hotels & catering; repairs	ONS	Log difference
10	IOP: All production industries	ONS	Log difference
11	IOP: Electricity, gas and water supply	ONS	Log difference
12	IOP: Manuf. of food, drink & tobacco	ONS	Log difference
13	IOP: Manuf. of coke/petroleum prod/nuclear fuels	ONS	Log difference
14	IOP: Manuf. of chemicals & man-made fibres	ONS	Log difference
15	Consumption	ONS	Log difference
16	Trade balance	ONS	None
17	RPI total food	ONS	Log difference
18	RPI total non-food	ONS	Log difference
19	RPI total all items other than seasonal food	ONS	Log difference
20	GDP deflator	ONS	Log difference
21	Wages	ONS	Log difference
22	Import prices	IFS	Log difference
23	Export prices	IFS	Log difference
24	M4 deposits	BOE	Log difference
25	M4 lending	BOE	Log difference
26	Real nationwide house prices	Nationwide Building Society	Log difference
27	Dividend yield	GFD	None
28	PE ratio	GFD	None
29	FTSE All-Share index	GFD	Log difference
30	Pounds US dollar rate	GFD	Log difference
31	Pounds euro rate	GFD	Log difference
32	Pounds yen rate	GFD	Log difference
33	NEER	GFD	Log difference
34	Pounds Canadian dollar rate	GFD	Log difference
35	Pounds Australian dollar rate	GFD	Log difference
36	Corporate bond yield	GFD	None
37	Unemployment rate	GFD	None
38	5-year govt bond yield	GFD	None
39	10-year govt bond yield	GFD	None
40	20-year govt bond yield	GFD	None
41	Commodity price index	GFD	Log difference
42	Brent oil price	GFD	Log difference
43	Industrial production index	GFD	Log difference
44	United Kingdom composite leading indicators	GFD	Log difference

Notes:
ONS: Office for National Statistics.
IFS: International Financial Statistics.
GFD: Global Financial Data.

References

Adolfson, M., Linde, J., & Villani, M. (2007). Forecasting performance of an open economy DSGE model. *Econometric Reviews*, *26*(2–4), 289–328.

Balke, N. S. (2000). Credit and economic activity: credit regimes and nonlinear propagation of shocks. *Review of Economics and Statistics*, *82*(2), 344–349.

Baumeister, C., & Benati, L. (2010). *Unconventional monetary policy and the great recession — estimating the impact of a compression in the yield spread at the zero lower bound*. Working paper series: vol. 1258. European Central Bank.

Baumeister, C., Liu, P., & Mumtaz, H. (2013). Changes in the effects of monetary policy on disaggregate price dynamics. *Journal of Economic Dynamics and Control*, *37*(3), 543–560.

Benati, L. (2007). *The great moderation in the United Kingdom*. Working paper series: vol. 769. European Central Bank.

Chen, C. (1998). A Bayesian analysis of generalized threshold autoregressive models. *Statistics and Probability Letters*, *40*(1), 15–22.

Chen, C., & Lee, J. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis*, *16*(5), 483–492.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, *90*(432), 1313–1321.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, *86*(2), 221–241.

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association*, *96*(453), 270–281.

Clark, T. E., & McCracken, M. W. (2009a). Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review*, *50*(2), 363–395.

Clark, T. E., & McCracken, M. W. (2009b). Tests of equal predictive ability with real-time data. *Journal of Business and Economic Statistics*, *27*(4), 441–454.

Clark, T. E., & McCracken, M. W. (2011). Testing for unconditional predictive ability. In D. Hendry & M. Clements (Eds.), *Oxford handbook on economic forecasting* (pp. 415–440) Oxford University Press.

Clements, M. P., & Hendry, D. F. (1995). Forecasting in cointegration systems. *Journal of Applied Econometrics*, *10*(2), 127–146.

Cogley, T., Primiceri, G.E., & Sargent, T.J. (2008). *Inflation-gap persistence in the US*. NBER working paper: vol. 13749. National Bureau of Economic Research, Inc.

Cogley, T., & Sargent, T. J. (2002). Evolving post world war II US inflation dynamics. In B. S. Bernanke & K. Rogoff (Eds.), *NBER macroeconomics annual 2001*. Cambridge, US: MIT Press.

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*, 262–302.

Cogley, T., & Sargent, T. J. (2008). Anticipated utility and rational expectations as approximations of Bayesian decision making. *International Economic Review*, *49*(1), 185–221.

Dacco, R., & Satchell, S. (1999). Why do regime-switching models forecast so badly? *Journal of Forecasting*, *18*(1), 1–16.

D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, *28*, 82–101.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–263.

Eickmeier, S., Lemke, W., & Marcellino, M. (2011). *Classical time-varying FAVAR models — estimation, forecasting and structural analysis*. Discussion paper series 1: economic studies 2011,04. Deutsche Bundesbank Research Centre.

Eklund, J., Kapetanios, G., & Price, S. (2010). *Forecasting in the presence of recent structural change*. Bank of England working paper: vol. 406.

Eklund, J., & Karlsson, S. (2007). Forecast combination and model averaging using predictive measures. *Econometric Reviews*, *26*(2–4), 329–363.

Faust, J., & Wright, J. (2011). *Forecasting inflation*. Technical report. John Hopkins University.

Ferrara, L., Marcellino, M., & Mogliani, M. (2012). *Macroeconomic forecasting during the Great Recession: The return of non-linearity?* Technical report.

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, *25*(4), 595–620.

Goncalves, S., & White, H. (2002). The bootstrap of the mean for dependent heterogeneous arrays. *Econometric Theory*, *18*, 1367–1384.

Goncalves, S., & White, H. (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*, *119*, 199–219.

Groen, J.J.J., & Mumtaz, H. (2008). *Investigating the structural stability of the Phillips curve relationship*. Bank of England working paper: vol. 350.

Lopes, H. F., & Salazar, E. (2006). Bayesian model uncertainty in smooth transition autoregressions. *Journal of Time Series Analysis*, *27*(1), 99–117.

Mumtaz, H., & Sunder-Plassmann, L. (2013). Time-varying dynamics of the real exchange rate: an empirical analysis. *Journal of Applied Econometrics*, *28*, 498–525.

Newey, W., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.

Pesaran, H., Pettenuzzo, D., & Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, *73*(4), 1057–1084.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, *72*(3), 821–852.

Rossi, B., & Sekhposyan, T. (2010). Have economic models' forecasting performance for US output growth and inflation changed over time, and when? *International Journal of Forecasting*, *26*(4), 808–835.

Sims, C. A., & Zha, T. (2006). Were there regime switches in U.S. monetary policy? *American Economic Review*, *96*(1), 54–81.

Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, *71*(3), 331–355.

Stock, J. H. (2001). Evolving post-world war II U.S. inflation dynamics: comment. *NBER Macroeconomics Annual*, *16*, 379–387.

Stock, J. H., & Watson, M. M. (2003). How did leading indicator forecasts perform during the 2001 recession? *Economic Quarterly (Summer)*, 71–90.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.

Stock, J. H., & Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*(s1), 3–33.