Forecasting Macroeconomic Variables Under Model Instability

Davide Pettenuzzo
Brandeis University*

Allan Timmermann
UCSD, CEPR, and CREATES†

May 9, 2015

Abstract

We compare different approaches to accounting for parameter instability in the context of macroeconomic forecasting models that assume either small, frequent changes versus models whose parameters exhibit large, rare changes. An empirical out-of-sample forecasting exercise for U.S. GDP growth and inflation suggests that models that allow for parameter instability generate more accurate density forecasts than constant-parameter models although they fail to produce better point forecasts. Model combinations deliver similar gains in predictive performance although they fail to improve on the predictive accuracy of the single best model which is a specification that allows for time-varying parameters and stochastic volatility.

Key words: Time-varying parameters; regime switching; change point models; stochastic volatility; GDP growth forecasts; inflation forecasts.

JEL classification: C22, C53

*Brandeis University, Sachar International Center, 415 South St, Waltham, MA, Tel: (781) 736-2834. Email: dpettenu@brandeis.edu.

†University of California, San Diego, 9500 Gilman Drive, MC 0553, La Jolla CA 92093. Tel: (858) 534-0894. Email: atimmerm@ucsd.edu.

1 Introduction

Parameter instability is pervasive, affecting models used to predict many commonly studied macroeconomic variables (Stock and Watson, 1996; Rossi, 2013). Although many empirical studies of simple forecasting models have found that parameters change over time, little is known about how best to incorporate such evidence of instability into the model specification in order to improve on forecasting models that assume constant parameters.

Since many different methods exist for addressing model instability it is particularly important to address if (and how) the assumed form of instability affects the models' ability to generate accurate forecasts. A key question is whether it is best to assume frequent, but small changes to model parameters or, conversely, to allow for rare, but large, shifts. The familiar time-varying parameter (TVP) model of Cooley and Prescott (1973) assumes that the parameters are subject to small shocks every period, converging either to a steady state value (mean-reverting process) or drifting over time (random walk process). The Markov switching (MS) model of Hamilton (1989) assumes that model parameters switch between a small set of repeated values (states). Detectable regime switches are typically large but do not occur every period. The change point (CP) model of Chib (1998) also allows for regime switches but dispenses with the assumption that regimes repeat, instead allowing the parameters within each regime to be unique.

Evaluating the impact of parameter instability on forecasting performance is important in part because it is difficult to accurately determine the nature of such instability. As pointed out by Elliott and Müller (2006), standard tests for model instability have power in multiple directions and so one can generally not infer from a rejection of the null of stable parameters, which type of model instability (e.g., drifting parameters versus regime switching) characterizes a particular variable. However, whether one faces multiple small breaks versus occasional large breaks could potentially have large consequences for many economic decisions. For example, the effect on economic welfare of a government's policy decisions may depend on whether shifts in the underlying GDP growth rate occur suddenly or more gradually through time.

This paper evaluates the importance for predictive performance of how parameter instability is modeled. Our evaluation considers the accuracy of both point and density forecasts. In an empirical analysis we apply a range of models to quarterly inflation and real GDP growth in the U.S. Both of these series have been widely studied; see Chauvet and Potter (2013) and Faust and Wright (2013) for recent reviews. We consider a TVP-stochastic volatility (TVP-SV) model along with MS models with two or three regimes and CP models with up to four different regimes. Using a mean squared error loss function, we find modest evidence that models that allow for parameter instability can produce better out-of-sample point forecasts of inflation while they do not seem to generate notable gains for the real GDP series. In contrast, we find strong evidence that allowing for parameter instability can greatly improve on the accuracy of the density forecasts associated with a constant-coefficient, homoskedastic model. Moreover, this improvement is mainly due to the ability of time-varying parameter models to generate more accurate density forecasts in the post-1984 Great Moderation sample. The best performance is observed for the models with stochastic volatility followed by MS and CP models with three states. Moreover, decompositions of the TVP-SV model's performance into separate TVP and SV components suggest that it is the ability of the models to account for time-varying volatility dynamics that leads to the improvements over the linear, homoskedastic benchmark.

In a recursive combination analysis that combines forecasts from the individual models we find that equal-weighted combination, Bayesian Model Averaging and the optimal prediction pool of Geweke and Amisano (2011) produce density forecasts that are superior to those generated by the benchmark linear models. However, the model combinations do not perform as well as the TVP-SV model. Plots of the recursively computed combination weights tell a clear story. Prior to 1985, the linear and CP models receive most of the weights in the combination. The TVP-SV model rapidly increases in importance after the emergence of the Great Moderation, however, and receives a weight above 80% towards the end of the sample for both the inflation and real GDP series. These results suggest that a model that allows for gradual changes to the model parameters performs better both in-sample and out-of-sample and highlight the importance of allowing for time-varying volatility.

Other papers have studied the effect of structural breaks on predictability of macroeconomic time series. Bauwens et al. (2014) provide a comprehensive analysis of the forecasting performance of two types of change point models for a range of macroeconomic series but do not compare TVP, MS and CP models as we do here. Giacomini and Rossi (2009) analyze the detection and prediction of breakdowns in forecasting models, whereas Rossi and Sekhposyan (2014) propose new regression-based tests for forecast optimality under model instability. Rossi (2013) provides an extensive comparison of the performance of different ways to account for model instability.

The remainder of the paper proceeds as follows. Section 2 introduces the benchmark (constant coefficient), TVP-SV, MS and CP models considered in our study and explains how we estimate the models. Section 3 introduces the data on inflation and real GDP growth and presents empirical results for the out-of-sample forecasting experiment. Section 4 discusses different model combination schemes while Section 5 concludes.

2 Models

This section introduces the different model specifications considered in our study and explains how they are estimated and used to generate forecasts.

Our benchmark specification is a linear model with constant coefficients. To capture time-variation in model parameters we consider three different specifications: (i) a model with time varying parameters and stochastic volatility (TVP-SV); (ii) a Markov switching (MS) model; and (iii) a change point (CP) model. These specifications are all common ways to account for parameter instability and represent very different ways to approach the problem. Whereas the TVP-SV model lets the parameters of both the first and second moments change every period, the MS and CP models typically identify discrete shifts in the parameters which occur infrequently. The MS model assumes that a small number of regimes repeat whereas the CP model assumes that the regimes are historically unique. Both of these models allow for regime switching in the parameters governing first and second moments. Ultimately, it is an empirical question which of these models will perform best as their performance depends on the nature of any instabilities in the data generating process.

2.1 Linear model

Suppose we are interested in predicting a univariate variable, y_{t+1}, given a set of predictors known at time t, $\mathbf{x}_t$. As the benchmark specification, we consider a standard linear forecasting model with constant regression coefficients and constant volatility:

$$y_{\tau+1} = \mu + \boldsymbol{\beta}'\mathbf{x}_\tau + \varepsilon_{\tau+1}, \quad \varepsilon_{\tau+1} \sim N(0, \sigma_\varepsilon^2), \quad \tau = 1, ..., t-1. \tag{1}$$

Here $\boldsymbol{\beta}$ and $\mathbf{x}_\tau$ are $k \times 1$ vectors of regression coefficients and predictors that are specific to each empirical application.

We assume that the parameters of (1), along with those of its competitors, are estimated using Bayesian methods. Following standard practice in the Bayesian literature (e.g., Koop, 2003), the priors for the parameters μ and $\boldsymbol{\beta}$ in (1) are assumed to be normal and independent of σ_ε^2

$$\begin{bmatrix} \mu \\ \boldsymbol{\beta} \end{bmatrix} \sim N\left(\underline{\mathbf{b}}, \underline{\mathbf{V}}\right), \tag{2}$$

where all elements of $\underline{\mathbf{b}}$ are set to zero, except for the term corresponding to the first lag of $y_{\tau+1}$, which is set to 1. As for the variance-covariance matrix $\underline{\mathbf{V}}$, we set aside an initial training sample of t_0 observations to calibrate its parameters and use a g-prior (see

Zellner, 1986):

$$\underline{\mathbf{V}} = \underline{\psi}^2 \left[s_{y,t_0}^2 \left(\sum_{\tau=1}^{t_0-1} \mathbf{x}_\tau \mathbf{x}_\tau' \right)^{-1} \right], \tag{3}$$

where

$$s_{y,t_0}^2 = \frac{1}{t_0 - 2} \sum_{\tau=1}^{t_0-1} \left(y_{\tau+1} - \underline{\mathbf{b}}' \mathbf{x}_\tau \right)^2 .$$

The approach of calibrating some of the prior hyperparameters using statistics computed over an initial training sample is quite standard in the Bayesian literature; see, e.g., Primiceri (2005), Clark (2011), Clark and Ravazzolo (2014), and Banbura et al. (2010).

In (3), the scalar $\underline{\psi}$ controls the tightness of the prior. $\underline{\psi} \to \infty$ corresponds to a diffuse prior on μ and $\boldsymbol{\beta}$. Our baseline results set $\underline{\psi} = 10$ for the inflation application, and $\underline{\psi} = 25$ for the GDP growth rate application. The larger value of $\underline{\psi}$ used for the GDP growth rate data reflects that this series is more volatile than the inflation data.

A standard gamma prior is assumed for the error precision of the return innovation, σ_ε^{-2}:

$$\sigma_\varepsilon^{-2} \sim \mathcal{G}\left(s_{y,t_0}^{-2}, \underline{v}_0 \left(t_0 - 1 \right) \right), \tag{4}$$

where $\underline{v}_0$ is a prior hyperparameter that controls the degree of informativeness of this prior, with $\underline{v}_0 \to 0$ corresponding to a diffuse prior on σ_ε^{-2}. We set $\underline{v}_0 = 0.01$ in the inflation application and $\underline{v}_0 = 0.005$ for the GDP growth rate application.

We estimate the model in (1) using a Gibbs sampler which allows us to draw from the posterior distributions of μ, $\boldsymbol{\beta}$, and σ_ε^{-2}, given the information set available at time t, $\mathcal{Y}^t = \{\mathbf{x}_\tau, y_\tau\}_{\tau=1}^t$. These draws are then used to compute a predictive density for y_{t+1}:

$$p\left(y_{t+1} \middle| \mathcal{Y}^t \right) = \int p\left(y_{t+1} \middle| \mu, \boldsymbol{\beta}, \sigma_\varepsilon^{-2}, \mathcal{Y}^t \right) p\left(\mu, \boldsymbol{\beta}, \sigma_\varepsilon^{-2} \middle| \mathcal{Y}^t \right) d\mu d\boldsymbol{\beta} d\sigma_\varepsilon^{-2}. \tag{5}$$

We refer the reader to an online appendix for more details on the Gibbs sampler and computation of the integral in (5).

2.2 Time-varying parameter, stochastic volatility model

Next, we modify the constant-coefficient model in (1) to allow for continuous changes in the regression coefficients and volatility:

$$y_{\tau+1} = \left(\mu + \mu_{\tau+1} \right) + \left(\boldsymbol{\beta} + \boldsymbol{\beta}_{\tau+1} \right)' \mathbf{x}_\tau + \exp\left(h_{\tau+1} \right) u_{\tau+1}, \quad u_{\tau+1} \sim \mathcal{N}(0,1), \tag{6}$$

where $h_{\tau+1}$ denotes the log-volatility at time $\tau + 1$. We assume that the time-varying parameters $\boldsymbol{\theta}_{\tau+1} = \left(\mu_{\tau+1}, \boldsymbol{\beta}_{\tau+1}' \right)$ follow a zero-mean, stationary process

$$\boldsymbol{\theta}_{\tau+1} = \boldsymbol{\gamma}_{\boldsymbol{\theta}}' \boldsymbol{\theta}_\tau + \boldsymbol{\eta}_{\tau+1}, \quad \boldsymbol{\eta}_{\tau+1} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q} \right), \tag{7}$$

where $\boldsymbol{\theta}_1 = \mathbf{0}$ and the elements in $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ are restricted to lie between -1 and 1. The log-volatility $h_{\tau+1}$ is also assumed to follow a stationary and mean reverting process:

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \quad \xi_{\tau+1} \sim \mathcal{N}\left(0, \sigma_\xi^2\right), \tag{8}$$

where $|\lambda_1| < 1$ and u_τ, $\boldsymbol{\eta}_t$ and ξ_s are mutually independent for all τ, t, and s.

Our choices of priors for $(\mu, \boldsymbol{\beta}')$ are the same as those in (2). The time-varying parameter, stochastic volatility (TVP-SV) model in (6)-(8) also requires eliciting priors for the sequence of time-varying parameters, $\boldsymbol{\theta}^t = \{\boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_t\}$, the variance covariance matrix $\mathbf{Q}$, the sequence of volatilities, $h^t = \{h_1, ..., h_t\}$, the error precision σ_ξ^{-2}, and the SV parameters $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$, λ_0, and λ_1. Using the decomposition $p\left(\boldsymbol{\theta}^t, \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \mathbf{Q}\right) = p\left(\boldsymbol{\theta}^t \middle| \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \mathbf{Q}\right) p\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}}\right) p\left(\mathbf{Q}\right)$, we note that (7) along with the assumption that $\boldsymbol{\theta}_1 = \mathbf{0}$ implies

$$p\left(\boldsymbol{\theta}^t \middle| \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \mathbf{Q}\right) = \prod_{\tau=1}^{t-1} p\left(\boldsymbol{\theta}_{\tau+1} \middle| \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \boldsymbol{\theta}_t, \mathbf{Q}\right), \tag{9}$$

with $\boldsymbol{\theta}_{\tau+1} | \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \boldsymbol{\theta}_\tau, \mathbf{Q} \sim \mathcal{N}\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}}' \boldsymbol{\theta}_\tau, \mathbf{Q}\right)$, for $\tau = 1, ..., t-1$. To complete the prior elicitation for $p\left(\boldsymbol{\theta}^t, \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \mathbf{Q}\right)$, we specify priors for $\mathbf{Q}$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ as follows. We choose an Inverted Wishart distribution for $\mathbf{Q}$:

$$\mathbf{Q} \sim \mathcal{IW}\left(\underline{\mathbf{Q}}, \underline{v}_{\mathbf{Q}}\left(t_0 - 1\right)\right), \tag{10}$$

with

$$\underline{\mathbf{Q}} = \underline{k}_{\mathbf{Q}} \underline{v}_{\mathbf{Q}} \left(t_0 - 1\right) \underline{\mathbf{V}}. \tag{11}$$

$\underline{k}_{\mathbf{Q}}$ controls the degree of variation in the time-varying regression coefficients $\boldsymbol{\theta}_\tau$, with larger values of $\underline{k}_{\mathbf{Q}}$ implying greater variation in $\boldsymbol{\theta}_\tau$. Our analysis sets $\underline{k}_{\mathbf{Q}} = \left(\underline{\psi}/100\right)^2$ and $\underline{v}_{\mathbf{Q}} = 10$. These are more informative priors than the earlier choices and limit the changes to the regression coefficients to be $\underline{\psi}/100$ on average.

We specify the elements of $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ to be a priori independent of each other with generic element $\gamma_{\boldsymbol{\theta}}^i$

$$\gamma_{\boldsymbol{\theta}}^i \sim \mathcal{N}\left(\underline{m}_{\gamma_{\boldsymbol{\theta}}}, \underline{V}_{\gamma_{\boldsymbol{\theta}}}\right), \quad \gamma_{\boldsymbol{\theta}}^i \in (-1, 1), \quad i = 1, ..., k. \tag{12}$$

where $\underline{m}_{\gamma_{\boldsymbol{\theta}}} = 0.8$, and $\underline{V}_{\gamma_{\boldsymbol{\theta}}} = 1.0e^{-6}$, implying relatively high autocorrelations.

Next, consider the sequence of log-volatilities, h^t, the error precision, σ_ξ^{-2}, and the parameters λ_0 and λ_1. Decomposing the joint probability of these parameters $p\left(h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}\right) = p\left(h^t \middle| \lambda_0, \lambda_1, \sigma_\xi^{-2}\right) p\left(\lambda_0, \lambda_1\right) p\left(\sigma_\xi^{-2}\right)$ and using (8), we have

$$\begin{aligned} p\left(h^t \middle| \lambda_0, \lambda_1, \sigma_\xi^{-2}\right) &= \prod_{\tau=1}^{t-1} p\left(h_{\tau+1} \middle| \lambda_0, \lambda_1, h_\tau, \sigma_\xi^{-2}\right) p\left(h_1\right), \\ h_{\tau+1} | \lambda_0, \lambda_1, h_\tau, \sigma_\xi^{-2} &\sim \mathcal{N}\left(\lambda_0 + \lambda_1 h_\tau, \sigma_\xi^2\right). \end{aligned} \tag{13}$$

To complete the prior elicitation for $p\left(h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}\right)$, we choose priors for λ_0, λ_1, the initial log volatility h_1, and σ_ξ^{-2} from the normal-gamma family:

$$h_1 \sim \mathcal{N}\left(\ln\left(s_{y,t_0}\right), \underline{k}_h\right), \tag{14}$$

$$\begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \underline{m}_{\lambda_0} \\ \underline{m}_{\lambda_1} \end{bmatrix}, \begin{bmatrix} \underline{V}_{\lambda_0} & 0 \\ 0 & \underline{V}_{\lambda_1} \end{bmatrix}\right), \quad \lambda_1 \in (-1, 1), \tag{15}$$

and

$$\sigma_\xi^{-2} \sim \mathcal{G}\left(1/\underline{k}_\xi, \underline{v}_\xi\left(t_0 - 1\right)\right). \tag{16}$$

We set $\underline{k}_\xi = 1.0e^{-04}$, $\underline{v}_\xi = 10$, and $\underline{k}_h = 0.1$. These choices restrict changes to the log-volatility to be only 0.01 on average and place a relatively diffuse prior on the initial log-volatility state.

Following Clark and Ravazzolo (2014) we set the hyperparameters to $\underline{m}_{\lambda_0} = 0$, $\underline{m}_{\lambda_1} = 0.9$, $\underline{V}_{\lambda_0} = 0.25$, and $\underline{V}_{\lambda_0} = 1.0e^{-4}$. This corresponds to setting the prior mean and standard deviation of the intercept to 0 and 0.5, respectively, and represents uninformative priors on the intercept of the log volatility specification and a prior mean of the AR(1) coefficient, λ_1, of 0.9 with a standard deviation of 0.01. This is a more informative prior that matches persistent dynamics in the log volatility process.

To estimate the model in (6)-(8), again we use a Gibbs sampler that lets us compute posterior draws for the model parameters μ, $\boldsymbol{\beta}$, $\boldsymbol{\theta}^t$, h^t, $\mathbf{Q}$, σ_ξ^{-2}, $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$, λ_0, and λ_1. These draws are used to compute density forecasts for y_{t+1}:

$$\begin{aligned} p\left(y_{t+1} \middle| \mathcal{Y}^t\right) &= \int p\left(y_{t+1} \middle| \boldsymbol{\theta}_{t+1}, h_{t+1}, \Theta, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t\right) \\ &\quad \times p\left(\boldsymbol{\theta}_{t+1}, h_{t+1} \middle| \Theta, \boldsymbol{\theta}^t, h^t, \mathcal{Y}^t\right) \\ &\quad \times p\left(\Theta, \boldsymbol{\theta}^t, h^t \middle| \mathcal{Y}^t\right) d\Theta d\boldsymbol{\theta}^{t+1} dh^{t+1}. \end{aligned} \tag{17}$$

$\Theta = \left(\mu, \beta, \mathbf{Q}, \sigma_\xi^{-2}, \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \lambda_0, \lambda_1\right)$ contains the time-invariant parameters. We refer the reader to an online appendix for more details on the Gibbs sampler and computations of the integral in (17).

As special cases of the general TVP-SV model we also consider models with time-varying parameters, but constant volatility (TVP) and a model with constant mean parameters and stochastic volatility (SV). These specifications allow us to identify whether changes in forecasting performance are mainly due to the TVP or SV components of the model.

2.3 Markov switching model

The Markov switching (MS) regression models allow both the regression coefficients and volatility parameters to change across a finite set of recurring regimes (states) $s_{\tau+1} \in$

$\{1, ..., K\}$:

$$y_{\tau+1} = \mu_{s_{\tau+1}} + \boldsymbol{\beta}'_{s_{\tau+1}} \mathbf{x}_{\tau} + \sigma_{s_{\tau+1}} u_{\tau+1}, \quad u_{\tau+1} \sim \mathcal{N}(0, 1). \tag{18}$$

The state transition probabilities are given by

$$\Pr\left(s_{\tau} = j | \, s_{\tau-1} = i \right) = p_{ij}, \quad i, j \in \{1, ..., K\}, \tag{19}$$

where $\sum_{j=1}^{K} p_{ij} = 1$ and $p_{ij} \geq 0$ for all $i, j \in \{1, ..., K\}$. Transition probabilities p_{ij} are collected in a $(K \times K)$ matrix $\mathbf{P}$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{1K} \\ p_{12} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ldots & \vdots \\ p_{1K} & p_{2K} & \cdots & p_{KK} \end{bmatrix}. \tag{20}$$

Turning to our choice of priors for this MS specification, let $\boldsymbol{\theta}_i = (\mu_i, \boldsymbol{\beta}'_i)$ be the regression coefficients in regime i, for $i = 1, ..., K$. Also, let $\mathbf{p}_{i,\cdot}$ be the $i-$th row of $\mathbf{P}$. Finally, collect all state-dependent regression coefficients and volatilities in $\boldsymbol{\Xi} = \left(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_K, \sigma_1^{-2}, ..., \sigma_K^{-2} \right)$. Following standard practice, we assume that the state-specific regression parameters and error precisions, $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_K, \sigma_1^{-2}, ..., \sigma_K^{-2}$, are a priori independent of the transition matrix $\mathbf{P}$:

$$p\left(\boldsymbol{\Xi}, \mathbf{P} \right) = p\left(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_K, \sigma_1^{-2}, ..., \sigma_K^{-2} \right) p\left(\mathbf{P} \right), \tag{21}$$

with

$$p\left(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_K, \sigma_1^{-2}, ..., \sigma_K^{-2} \right) = \prod_{i=1}^{K} p\left(\boldsymbol{\theta}_i, \sigma_i^{-2} \right). \tag{22}$$

In a straightforward extension of (2) and (4) we assume that, for each regime i, the prior distribution for the vector of regression parameters $\boldsymbol{\theta}_i$ is normal and independent of the error precision σ_i^{-2}, whose prior distribution is a standard gamma:

$$p\left(\boldsymbol{\theta}_i, \sigma_i^{-2} \right) = p\left(\boldsymbol{\theta}_i \right) p\left(\sigma_i^{-2} \right), \tag{23}$$

with

$$\boldsymbol{\theta}_i \sim \mathcal{N}\left(\underline{\mathbf{b}}, \underline{\mathbf{V}} \right), \quad i = 1, ..., K, \tag{24}$$

and

$$\sigma_i^{-2} \sim \mathcal{G}\left(s_{y,t_0}^{-2}, \underline{v}_0 \left(t_0 - 1 \right) \right), \quad i = 1, ..., K. \tag{25}$$

Finally, we assume that the individual rows of $\mathbf{P}$ are independent and follow a Dirichlet distribution:

$$\mathbf{p}_{i,\cdot} \sim \mathcal{D}\left(e_{i1}, ..., e_{iK} \right), \quad i = 1, ..., K. \tag{26}$$

Following Frühwirth-Schnatter (2001), we specify a prior that is invariant to relabeling the states by setting $e_{ii} = e^{\underline{\varkappa}}$ and $e_{ij} = e^{\underline{\varrho}}$, for all $i \neq j$. We choose $\underline{\varkappa} = 2$ and $\underline{\varrho} = 1/(K-1)$. Our choices for $\underline{\varkappa}$ and $\underline{\varrho}$ guarantee that the Markov switching model is bounded away from a finite mixture model. See also Frühwirth-Schnatter (2006).

To estimate the model in (18)-(20), we use a Gibbs sampler which provides a sequence of posterior draws for the parameters of the model $\mathbf{\Xi}, \mathbf{P}$, as well as the sequence of hidden states, $s^t = (s_1, ..., s_t)$. These draws are then used to form a density forecast for y_{t+1}:

$$\begin{aligned} p\left(y_{t+1} \middle| \mathcal{Y}^t\right) &= \int p\left(y_{t+1} \middle| s_{t+1}, s^t, \mathbf{\Xi}, \mathbf{P}, \mathcal{Y}^t\right) p\left(s_{t+1} \middle| s^t, \mathbf{\Xi}, \mathbf{P}, \mathcal{Y}^t\right) \\ &\quad \times p\left(s^t, \mathbf{\Xi}, \mathbf{P} \middle| \mathcal{Y}^t\right) ds^{t+1} d\mathbf{\Xi} d\mathbf{P}. \end{aligned} \tag{27}$$

$p\left(s_{t+1} \middle| s^t, \mathbf{\Xi}, \mathbf{P}, \mathcal{Y}^t\right)$ is the one-step-ahead predicted probability for the hidden Markov chain. We refer the reader to an online appendix for more details on the Gibbs sampler for this model.

2.4 Change-point model

Finally, we consider a Change Point (CP) regression model that allows both the regression coefficients and volatility parameters to change across non-recurring regimes $s_{\tau+1} \in \{1, 2, ..., M\}$:

$$CP : y_{\tau+1} = \mu_{s_{\tau+1}} + \boldsymbol{\beta}'_{s_{\tau+1}} \mathbf{x}_t + \sigma_{s_{\tau+1}} u_{\tau+1}, \quad u_{\tau+1} \sim \mathcal{N}(0,1). \tag{28}$$

Following Chib (1998), shifts to the regression coefficients and error term volatility are captured through the integer-valued state variable s_t that tracks the underlying regime. For example, a change from $s_\tau = k$ to $s_{\tau+1} = k+1$ indicates that a break has occurred at time $\tau + 1$. The transition probability is designed so that at each point in time s_τ can either remain in the current state or jump to the subsequent state:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & 0 & \cdots & 0 \\ 0 & p_{22} & p_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \vdots & 0 & p_{M-1,M-1} & p_{M-1,M} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}, \tag{29}$$

where $p_{k,k+1} = \Pr\left(s_\tau = k+1 \middle| s_{\tau-1} = k\right) = 1 - p_{k,k}$ is the probability of moving to regime $k+1$ at time τ given that the regime at time $\tau - 1$ is k.

Analogous to the MS model, let $\boldsymbol{\theta}_i = (\mu_i, \boldsymbol{\beta}'_i)$ be the regression coefficients in regime i, for $i = 1, ..., M$, and collect the state-dependent regression coefficients and volatilities in $\mathbf{\Xi} = \left(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_M, \sigma_1^{-2}, ..., \sigma_M^{-2}\right)$. Again, assume that the regression parameters and volatilities $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_M, \sigma_1^2, ..., \sigma_M^2$, are independent of the transition matrix $\mathbf{P}$

$$p(\Xi, \mathbf{P}) = p\left(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_M, \sigma_1^{-2}, ..., \sigma_M^{-2}\right) p(\mathbf{P}), \tag{30}$$

with

$$p\left(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_M, \sigma_1^{-2}, ..., \sigma_M^{-2}\right) = \prod_{i=1}^{M} p\left(\boldsymbol{\theta}_i, \sigma_i^{-2}\right). \tag{31}$$

Similarly, assume that the mean and variance parameters have normal-inverse Gamma priors, respectively:

$$\begin{aligned} p\left(\boldsymbol{\theta}_i, \sigma_i^{-2}\right) &= p(\boldsymbol{\theta}_i)\, p\left(\sigma_i^{-2}\right), \\ \boldsymbol{\theta}_i &\sim \mathcal{N}\left(\underline{\mathbf{b}}, \underline{\mathbf{V}}\right), \\ \sigma_i^{-2} &\sim \mathcal{G}\left(s_{y,t_0}^{-2}, \underline{v}_0 (t_0 - 1)\right), \quad i = 1, ..., M. \end{aligned} \tag{32}$$

Only the diagonal elements of $\mathbf{P}$ need to be specified for the CP model. The closer p_{ii} is to 1, the longer the expected duration of regime i. We assume that each of the diagonal elements of $\mathbf{P}$ follows an independent Beta distribution

$$p_{ii} \sim \mathcal{B}\left(\underline{a}_p, \underline{b}_p\right), \quad i = 1, ..., M - 1. \tag{33}$$

Specifically, we set $\underline{a}_p = 0.1\,(t/M)$ and $\underline{b}_p = 0.1$. These choices reflect the belief that a priori each regime should have the same duration which is approximately equal to t/M.

We use a Gibbs sampler to estimate the CP model (28)-(29). This yields posterior draws for the parameters $\Xi, \mathbf{P}$ as well as the sequence of hidden states, $s^t = (s_1, ..., s_t)$. These draws can be used to compute a density forecast for y_{t+1} conditional on M states occurring up to time $t + 1$:

$$p\left(y_{t+1} \middle| s_{t+1} = M, \mathcal{Y}^t\right) = \int p\left(y_{t+1} \middle| s_{t+1} = M, s^t, \Xi, \mathbf{P}, \mathcal{Y}^t\right) \times p\left(s^t, \Xi, \mathbf{P} \middle| \mathcal{Y}^t\right) ds^t d\Xi d\mathbf{P}. \tag{34}$$

By using the predictive density $p\left(y_{t+1} | s_{t+1} = M, s^t, \Xi, \mathbf{P}, \mathcal{Y}^t\right)$ we implicitly assume that no breaks will occur between the end of the estimation sample t and the end of the forecasting horizon, $t + 1$. This assumption is justified by the fact that our focus is on one-step-ahead forecasts. We refer the reader to an online appendix for more details on the Gibbs sampler for the CP model. See Pesaran et al. (2006) for a more general setup that allows for breaks occurring at longer forecast horizons.

3 Empirical results

This section introduces the data on inflation and real output growth considered in our empirical application. Our focus on these particular variables is motivated in part by the work of Stock and Watson (1999, 2003, 2007), who find strong evidence of instability in predictive relations for both output growth and inflation.

Next, we analyze the full-sample performance of the models described in Section 2 for the inflation and real GDP growth rate data. Finally, we turn to the out-of-sample predictive accuracy of these models, applying a range of measures to evaluate the accuracy of the point and density forecasts.

3.1 Data

Let $\pi_t = 400 \times \ln(P_t/P_{t-1})$ denote the annualized quarterly inflation rate, with P_t being the quarterly price index for the GDP deflator. We model next quarter's change in the inflation rate using a backward-looking Phillips curve

$$\Delta\pi_{t+1} = \mu + \beta(L)u_t + \lambda(L)\Delta\pi_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \tag{35}$$

where $\Delta\pi_{t+1} = \pi_{t+1} - \pi_t$ is the quarter-on-quarter change in the annualized inflation rate, and u_t is the quarterly unemployment rate. μ is a constant while $\beta(L)$ and $\lambda(L)$ are lag polynomials written in terms of the lag operator L. See Stock and Watson (2007) for a similar specification.

Both the GDP deflator and the unemployment rate data are seasonally adjusted series obtained from the Federal Reserve of St. Louis' FRED database. In particular, P_t is the quarterly Gross Domestic Product - Implicit Price Deflator series, and we construct a quarterly unemployment rate series by retaining the total civilian unemployment rate during the last month within each quarter. Our sample ranges from 1950:Q1 to 2013:Q4.

Turning to the GDP series, let $z_t = 400 \times \ln(Q_t/Q_{t-1})$ denote the annualized quarterly U.S. real GDP growth rate, where Q_t is the quarterly U.S. real GDP series. We model next quarter's GDP growth rate using an autoregressive specification

$$z_{t+1} = \mu + \beta(L)z_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right). \tag{36}$$

Our quarterly GDP growth rate series is constructed using seasonally adjusted quarterly U.S. real GDP data from the Federal Reserve of St. Louis database FRED and the sample ranges from 1950:Q1 to 2013:Q4.

3.2 Full sample estimates and model comparisons

We first compare the fit of the different models over the full sample, 1950:Q1-2013:Q4, using the first twenty years of data, 1950:Q1-1969:Q4, to train the prior hyperparameters (that is, $t_0 = 80$). A natural approach to model selection in a Bayesian setting is to compute the Bayes factor $B_{1,0}$ of the null model M_0 versus an alternative model, M_1. Higher Bayes factors suggest higher posterior odds in favor of M_1 against M_0. Specifically, we report twice the natural log of the Bayes factors, $2\ln(B_{1,0})$. To interpret the strength of the evidence, we follow studies such as Kass and Raftery (1995) and note that if $2\ln(B_{1,0})$ is below zero, the evidence supports M_0 over M_1. For values of $2\ln(B_{1,0})$ between 0 and 2, there is "weak evidence" that M_1 is a better model than M_0, whereas values of $2\ln(B_{1,0})$ between 2 and 6, 6 and 10, and higher than 10, can be viewed as "some evidence," "strong evidence," and "very strong evidence", respectively, in support of M_1 versus the null model, M_0.

Table 1 reports the Bayes factors from such model comparisons. For both the inflation rate and real GDP growth series there is compelling evidence against the linear model: in each case the models that allow for parameter instability easily beat their linear, constant-coefficient counterpart. The strongest results are observed for the SV and TVP-SV models, with the three-state MS model and the four-state CP model ranked second and third, respectively. Although clearly better than the linear model, there is less support for the two-state MS and CP models and the TVP model is only marginally better than the linear benchmark. These results suggest that it is the ability of the models to capture time-varying second moment dynamics that lead them to outperform the linear, homoskedastic model.

To understand how different models handle parameter instability, we next present time-series plots of the key regression coefficients and volatility parameters. Figure 1 displays the coefficient estimates for the intercept, AR(1) and lagged unemployment rate in the inflation rate model (35). The TVP-SV estimates of these three parameters are relatively stable over time, although they do respond to events such as the Great Recession in 2008-2009–a period during which inflation becomes less sensitive to the unemployment rate. We see much greater variations over time in the parameters of the three-state CP model in particular. For this model, the inflation rate is initially not very sensitive to variations in the unemployment rate. This changes dramatically between 1952 and 1985, during which we see a strongly negative relation between the lagged unemployment rate and subsequent changes in inflation. After 1985 the sensitivity of the inflation rate with regard to the lagged unemployment rate is again close to zero. Interestingly, the sensitivity of inflation to its own lagged value (shown in the middle panel) also changes significantly two years into the sample. The amount of parameter instability uncovered by the MS model falls somewhere in between that uncovered by the CP and TVP-SV models.

Turning to the full-sample estimates of inflation volatility, Figure 2 shows large differences across the three models. Though all three models identify the first two years of the sample as a period with high conditional volatility, their subsequent volatility estimates are very different. Specifically, the TVP-SV model suggests that the sixties and nineties were periods with relatively low volatility, whereas volatility was somewhat higher during the seventies, peaking around 1975, and again around 2007. The three-state CP model classifies the 30-year period 1953-1983 as a medium-volatility period, with low volatility ensuing after 1985. Finally, the two-state MS model detects spikes in the conditional volatility in 1950-52, 1956 and 1975.

The filtered state probabilities plotted in Figures 3 and 4 help explain the differences in the dynamics of the two-state MS and three-state CP models. Unsurprisingly the two-state MS model assigns a probability close to one to the high-volatility regime at the beginning of the sample (1950-52) and, again, in 1975, with a probability near one-half assigned to this state around 1987. The remainder of the time the probability assigned to the low volatility state (state 1) is close to one. The three-state CP model assigns a probability close to one for the first state from 1950 to 1952, and zero thereafter. It assigns a probability close to one for the second regime in the mid fifties and again during 1975-1985. Finally, the filtered probability of being in the third state is close to one after 1995. Overall, the regimes are quite well identified by the data for both of these models.

Turning to the GDP growth rate variable, Figure 5 shows that the intercept and AR(1) coefficient estimates are relatively stable under the TVP-SV model, whereas both the MS and CP models suggest higher variability in these coefficients. The conditional volatility plot in Figure 6 shows that the three-state CP model identifies a short-lived volatility spike in the early part of the sample followed by a long-lived regime with medium volatility from 1951-1983 and a regime with low volatility after 1984 which coincides with the Great Moderation. Unsurprisingly, the two-state MS model that allows for recurring states identifies more short-run dynamics in the conditional volatility than the three-state CP model which rules out such transitions.

Figure 7 shows that the filtered state probabilities for the two-state MS model are quite noisy for the GDP series up to around 1984. After 1984 the model mostly stays in the low-volatility regime, main exceptions occurring around 2000 and, more pronounced, in 2008, where volatility spiked. Clearer separation between states is obtained by the three-state CP model. Figure 8 shows that an early high-volatility state occurs in the first part of the sample. This state is followed by a medium-volatility state that lasts from 1952 through 1984. After 1984 the model transitions to a low volatility state.

3.3 Out-of-sample forecasting performance

We use the first twenty years of data from 1950Q1 to 1969Q4 to obtain initial parameter estimates which are then used to predict outcomes in 1970Q1. The next period we include data for 1970:Q1 in the training sample and use the resulting estimates to predict the outcome in 1970Q2. We proceed recursively in this fashion until the last observation in the sample, producing a time series of one-step-ahead forecasts from 1970QI to 2013Q4.

3.3.1 Measures of Predictive Accuracy

We summarize the precision of the point forecasts of model m, relative to that from the linear model, by means of the ratio of RMSFE values

$$RMSFE_m = \frac{\sqrt{\frac{1}{\bar{t}-\underline{t}+1}\sum_{\tau=\underline{t}}^{\bar{t}} e_{m,\tau}^2}}{\sqrt{\frac{1}{\bar{t}-\underline{t}+1}\sum_{\tau=\underline{t}}^{\bar{t}} e_{LIN,\tau}^2}}, \tag{37}$$

where $e_{LIN,\tau}$ and $e_{m,\tau}$ are the forecast errors associated with the linear (LIN) and alternative $(m \in TVP-SV, MS, CP)$ model, respectively, and $\underline{t}$ =1970Q1 and $\bar{t}$ =2013Q4 denote the beginning and end of the evaluation sample. The point forecasts used to compute the forecast errors are obtained by averaging over the draws from the predictive density $p\left(y_\tau | \mathcal{Y}^{\tau-1}\right)$. Values of $RMSFE_m$ below one suggest that model m produces more accurate point forecasts than the linear, homoskedastic benchmark.

To study the evolution in the accuracy of the point forecasts, we plot the Cumulative Sum of Squared prediction Error Difference (CSSED) proposed by Welch and Goyal (2008)

$$CSSED_{m,t} = \sum_{\tau=\underline{t}}^{t}\left(e_{LIN,\tau}^2 - e_{m,\tau}^2\right). \tag{38}$$

Positive and rising values of this measure indicate that the point forecasts generated by model m are more accurate than those produced by the linear, constant coefficient model.

One of the advantages of our Bayesian framework is its ability to generate predictive distributions, rather than simple point forecasts, in a manner that accounts for parameter estimation errors. To compare the performance of two density forecasts, following Amisano and Giacomini (2007), Geweke and Amisano (2010), and Hall and Mitchell (2007), we consider the average log score differential,

$$LSD_m = \frac{1}{\bar{t}-\underline{t}+1}\sum_{\tau=\underline{t}}^{\bar{t}}\left(LS_{m,\tau} - LS_{LIN,\tau}\right), \tag{39}$$

where $LS_{m,\tau}$ ($LS_{LIN,\tau}$) denotes model m's (LIN's) log predictive score computed at time τ, i.e., the log of the outcome evaluated at the posterior predictive density. Positive values of LSD_m indicate that on average model m produces more accurate density forecasts than the benchmark model (LIN). We also study the cumulative log score differential between the mth model and the LIN benchmark:

$$CLSD_{m,t} = \sum_{\tau=\underline{t}}^{t} \left(LS_{m,\tau} - LS_{LIN,\tau} \right). \tag{40}$$

Positive and rising values of $CLSD$ again indicate periods where model m produces more accurate density forecasts than the linear, homoskedastic benchmark.

3.3.2 Empirical Results

Table 2 reports estimates of these predictive accuracy measures for different sub-samples 1970-1983, 1984-2013, and for the full sample 1970-2013. These two subsamples are chosen to coincide with the transition to the Great Moderation period which is associated with a substantial reduction in the volatility of many macroeconomic variables, see McConnell and Perez-Quiros (2000). The first row reports the RMSFE values for the linear model whereas the subsequent rows report RMSFE ratios measured relative to this model. First consider the RMSFE ratio for the inflation series. Over the full sample 1970-2013 all models that allow for instability produce more accurate point forecasts than those generated by the linear model with RMSFE ratios ranging from 0.965 to 0.999, suggesting modest gains in RMSFE performance of 1-3%, none of which is significantly different from zero. The sub-sample results show that these improvements are all due to the better performance, measured relative to the linear model, in the 1970-1983 sample whereas the models that allow for instability actually slightly underperform over the sample 1984-2013.

Much greater, and significant, improvements are observed for the LSD estimates over the sample 1970-2013. Interestingly, and in contrast to the RMSFE results, the improvements in the LSD measure are mostly due to the post-1983 sample–a period for which all of the LSD estimates are significantly positive–although smaller improvements are also obtained in the 1970-1983 sample period.

These seemingly contradictory results for the RMSFE measure (suggesting improvements in predictive accuracy from 1970-1983 but not after 1984) versus for the LSD measure (suggesting the reverse) are explained by the fact that the RMSFE measure only focuses on predicting the mean whereas the LSD measure accounts for the models' ability to predict the full outcome distribution. In fact, the improved density forecasts after 1984 result from the models' ability to account for the reduced volatility in the period after the Great Moderation. In contrast, the linear model fails to capture this decline in volatility. This is consistent with a decomposition of the Bayes factor for the TVP-SV model into its TVP and SV parts which shows that the improvement over the linear, homoskedastic model predominantly arises from the SV component rather than from the TVP component of this model.

Turning to the real GDP series, only the SV and TVP-SV models manage to reduce the RMSFE of the linear constant-coefficient model although the modest gain (1.1%) is deemed to be statistically significant using a Diebold-Mariano type test statistic. Stronger results are observed for the LSD measure which shows that all of the models that allow for parameter instability produce significantly more accurate density forecasts than the constant-coefficient, homoskedastic benchmark. Once gain these gains in accuracy for the density forecasts are obtained during the 1984-2013 sample.

The first three windows in Figure 9 report the cumulative SSE differentials for the TVP–SV model, the MS models and the CP models applied to the inflation rate series. In each case we find strong forecasting performance up to around 1985 at which point the line flattens. All models see a deterioration in their performance relative to the linear benchmark model in 2008/09.

Figure 10 shows a very different path for the cumulative sum of log-score differentials for the quarterly inflation rate. In this case the models that allow for instability perform slightly worse or as well as the constant-coefficient benchmark up to around 1984. After this, the instability models start performing much better up to around 2006 at which point the accuracy of their density forecasts is about as good as that of the linear benchmark.

Figure 11 shows that the TVP–SV model fitted to real GDP growth manages to produce slightly more accurate point forecasts than the linear benchmark throughout most of the sample. In contrast, the sample period 1970-2013 sees steady underperformance of the MS and CP models relative to the constant-coefficient benchmark's RMSFE performance. Very different results are obtained under the LSD measure. The cumulative sum of log-score differentials depicted in Figure 12 shows superior density forecasting performance of the instability models from 1985 onwards, only interrupted by brief spells of underperformance in 1999 and, again, in 2008.

A popular approach for dealing with model instability is to use a rolling estimation window. We also consider this approach, using a 20-year estimation window corresponding to 80 quarterly observations. The results for the rolling window approach (shown in the last line of Table 2) show no evidence of overall improvements in the full-sample RMSFE performance for the inflation rate series. The rolling window approach generates some improvements in the LSD measure (relative to the linear model) although clearly less than the improvements seen for the better of the more formal modeling approaches. Similar results are found for the GDP growth rate series.

3.4 Results for longer forecast horizons

So far we have focused on forecasting results for the one-quarter-ahead horizon. However, our approach is easily generalized to allow for an arbitrary forecast horizon, H, by using lagged predictors $\mathbf{x}_{t-H}$ to predict y_t. To see how sensitive our results are to changing the forecast horizon, Table 3 shows results for $H = 4$ and $H = 8$, corresponding to one- and two-year forecast horizons. At the four-quarter horizon, the full-sample point forecasts are a little better (relative to the linear, homoskedastic benchmark) for the inflation data, but a little worse for real GDP growth. Once again, improvements in the point forecasts are mainly due to the 1970-1983 period. Greater, and often significant, improvements in performance are observed for the four-quarter-ahead density forecasts evaluated using the LSD measure, although, as for the one-quarter-ahead forecasts, here the improvements occur after 1984.

At the eight-quarter forecast horizon there is little evidence of systematic improvements over the linear, homoskedastic model, except for the density forecasting results for the SV and TVP SV models from 1984 onwards.

These results suggest that the forecasting performance of the models that account for parameter instability and second moment dynamics is best at relatively short forecast horizons of up to one year. These results are perhaps unsurprising given that the empirical evidence of non-linearities associated with regime switching or time-varying volatility tends to be weaker at longer horizons.

4 Model Combinations

Rather than attempting to identify one particular model specification that allows for changing parameters, an alternative strategy for dealing with model instability is to fit a variety of models and then use model combination to generate an averaged forecast.

Model combinations form weighted averages of individual prediction models using weights that can reflect the individual models' historical performance. The better a model's fit relative to its complexity, the larger its weight. We consider three commonly used combination schemes which we next describe.

4.1 Combination Schemes

The equal-weighted pool (EWP) puts equal weights on each of the N models, M_i :

$$p\left(y_{t+1}\middle|\, \mathcal{Y}^t\right) = \frac{1}{N}\sum_{i=1}^{N} p\left(y_{t+1}\middle|\, M_i, \mathcal{Y}^t\right), \tag{41}$$

where $\{p(y_{t+1}|M_i, \mathcal{Y}^t)\}_{i=1}^N$ denotes the predictive densities specified in (5), (17), (27), and (34).

Bayesian model averaging (BMA) uses weights that are proportional to the posterior model probabilities:

$$p\left(y_{t+1}|\mathcal{Y}^t\right) = \sum_{i=1}^{N} \Pr\left(M_i|\mathcal{Y}^t\right) p\left(y_{t+1}|M_i, \mathcal{Y}^t\right), \tag{42}$$

where $\Pr(M_i|\mathcal{Y}^t)$ denotes the posterior probability of model i, relative to all models under consideration, computed using information available at time t, $\mathcal{Y}^t$:

$$\Pr\left(M_i|\mathcal{Y}^t\right) = \frac{\Pr(\mathcal{Y}^t|M_i)\Pr(M_i)}{\sum_{j=1}^{N}\Pr(\mathcal{Y}^t|M_j)\Pr(M_j)}. \tag{43}$$

$\Pr(\mathcal{Y}^t|M_i)$ and $\Pr(M_i)$ denote the marginal likelihood and prior probability for model i, respectively. We assume that all models are equally likely a priori so that $\Pr(M_i) = 1/N$. Following Geweke and Amisano (2010), we compute the marginal likelihoods by cumulating each model's predictive log score model after conditioning on an initial warm-up sample.

Finally, we computes the weighted average of the predictive densities using the optimal predictive pool (OW) proposed by Geweke and Amisano (2011):

$$p\left(y_{t+1}|\mathcal{Y}^t\right) = \sum_{i=1}^{N} \omega_{t,i}^* \times p\left(y_{t+1}|M_i, \mathcal{Y}^t\right). \tag{44}$$

The $(N \times 1)$ vector of model weights $\omega_t^* = \left[\omega_{t,1}^*, ..., \omega_{t,N}^*\right]$ is determined by solving the following maximization problem:

$$\omega_t^* = \arg\max_{\omega_t} \sum_{\tau=1}^{t} \log\left[\sum_{i=1}^{N} \omega_{it} \times S_{\tau,i}\right], \tag{45}$$

where $LS_{\tau,i}$ is the recursively computed log-score for model i at time τ, $S_{\tau,i} = \exp(LS_{\tau,i})$, and $\omega_t^* \in [0,1]^N$.

By recursively updating the combination weights in (42) and (45), the BMA and OW combination methods accommodate time variations in the underlying data generating process. This is empirically important as we shall see.

4.2 Empirical results for model combinations

Figures 13 and 14 plot the time-series evolution in the combination weights in the optimal prediction pool. These are similar to recursively computed Bayes factors and turn out to be highly informative about shifts in different models' performance over the sample. First consider the weights on the inflation models (Figure 13). The CP model receives a high weight in the first couple of quarters and again between 1972 and 1995 with weights generally less than 20% thereafter. The linear model receives up to 70% of the probability weight in the mid-seventies, but its weight declines to 15% or less after 1985. The TVP-SV model picks up most of the remaining probability mass and so this model accounts for more than 80% of the total weight at the end of the sample. The MS models receive no weight during the sample.

For the real GDP growth series (Figure 14) the linear model is dominant most of the time between 1972 and the late eighties, at times receiving a weight of 100%. Starting in 1985 the TVP-SV model begins to dominate the optimal prediction pool with a weight greater than 70% in 1990, rising to 85% at the end of the sample. The CP models get a weight around 20% from 1990-2008 and the MS models get a weight around 10% in the last five years of the sample.

Table 4 reports out-of-sample forecasting results for the different model combination schemes. First consider the results for the inflation rate shown in Panel A. All three combination schemes manage to reduce the RMSFE of the linear model by 2-3 percent with the improvements once again resulting from the 1970-1983 subperiod. Moreover, as for the individual models, the forecasting performance of the model combinations is much stronger based on the predictive density LSD measure. Improvements are particularly strong in the period after the beginning of the Great Moderation and occurs for all three combination schemes.

Turning to the results for real GDP growth, Panel B of Table 4 shows that none of the model combination schemes manages to improve on the accuracy of the point forecasts generated by the linear model. In contrast, all three approaches produce better density forecasts than the linear model in the full sample (1970-2013) as well as in the post-1983 sample.

In general we find that the model combination results are close to, but slightly weaker than, those generated by the best of the individual models (the TVP-SV model) at the one-quarter horizon. This is unsurprising given that this model's performance is so much better than that of the other individual models and given that we are not averaging over a very large set of models which reduces the scope of gains achievable from model combination.

At the longer forecast horizons, $H = 4, 8$ quarters, the forecast combinations continue to produce density forecasts of inflation and GDP growth that are superior relative to the linear, homoskedastic benchmark. In some cases the forecast combinations now perform better than the best of the individual models (such as the TVP-SV model) that enter into the combination.

The cumulative sum of squared forecast error differential plots for the model combinations, shown in the last windows in Figures 9 and 10, show that the model combinations perform well up to around 1985, at which point they cease producing gains in predictive accuracy relative to the constant-coefficient model.

Figures 11 and 12 show that the model combinations initially produce poor out-of-sample point forecasts but generate more accurate density forecasts than the constant coefficient benchmark from 1985 onwards.

5 Conclusions

This paper compares the predictive performance of three popular approaches to account for model instability that make very different assumptions about the nature of parameter instability. In applications to inflation and real GDP growth forecasting we find that accounting for parameter instability and, notably, second moment dynamics has the potential to produce sizeable gains in the accuracy of the density forecasts but only modest gains in the accuracy of the point forecasts.

Our empirical results suggest that incorporating time-varying volatility is important for quarterly macroeconomic time-series that are affected by important shifts such as the Great Moderation which led to a substantial reduction in macroeconomic volatility. Although limited to two (key) macroeconomic variables, the results also suggest that SV or TVP-SV specifications that allow for gradual (frequent) shifts in model parameters perform better than alternative regime switching models that assume more sudden shifts in model parameters with or without recurring states.

References

Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics 25*(2), 177–190.

Banbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Bauwens, L., G. Koop, D. Korobilis, and J. V. Rombouts (2014). The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics*, In Press.

Chauvet, M. and S. Potter (2013). Forecasting output. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Part A, pp. 141 – 194. Elsevier.

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

Clark, T. E. (2011). Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics 29*(3), 327–341.

Clark, T. E. and F. Ravazzolo (2014). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, in press.

Cooley, T. F. and E. C. Prescott (1973). An adaptive regression model. *International Economic Review 14*(2), pp. 364–371.

Elliott, G. and U. K. Müller (2006). Efficient tests for general persistent time variation in regression coefficients. *The Review of Economic Studies 73*(4), 907–940.

Faust, J. and J. H. Wright (2013). Forecasting inflation. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Part A, pp. 2 – 56. Elsevier.

Frühwirth-Schnatter, S. (2001). Markov chain monte carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*(453), pp. 194–209.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

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

Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics 164*(1), 130 – 141.

Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies 76*(2), 669–705.

Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting 23*(1), 1 – 13.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), pp. 357–384.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), pp. 773–795.

Koop, G. (2003). *Bayesian Econometrics*. John Wiley & Sons, Ltd.

McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review 90*(5), 1464–1476.

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

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

Rossi, B. (2013). Advances in forecasting under instability. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Part B, pp. 1203 – 1324. Elsevier.

Rossi, B. and T. Sekhposyan (2014). Forecast optimality tests in the presence of instabilities, with applications to federal reserve and survey forecasts. Manuscript, Pompeu Fabra and Texas A&M.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

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

Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature 41*(3), 788–829.

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

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g prior distributions. In P. Goel and A. Zellner (Eds.), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pp. 233–243. North-Holland.

Table 1. Bayes factors for different model specifications

M_1 vs. *M_0*	*Inflation rate*	*Real GDP*
TVP vs. LIN	2.278	0.840
SV vs. LIN	94.322	48.860
TVP-SV vs. LIN	93.361	48.528
MS, K=2 vs. LIN	66.402	29.862
MS, K=3 vs. LIN	78.648	32.072
CP, K=2 vs. LIN	67.920	21.233
CP, K=3 vs. LIN	71.954	31.475
CP, K=4 vs. LIN	72.770	29.128

This table reports pairwise model comparisons using twice the natural logarithm of the Bayes factor, $2 \times (\ln B_{1,0})$, where $B_{1,0}$ denotes the Bayes factor obtained from comparing model M_1 to model M_0

$$B_{1,0} = \frac{Pr\left(\mathcal{Y}^t \middle| M_1\right)}{Pr\left(\mathcal{Y}^t \middle| M_0\right)}$$

Pairwise model comparisons are listed in the first and second columns for the inflation rate and growth in real GDP, respectively. Kass and Raftery (1995) suggest interpreting the results as follows: when $2 \times (\ln B_{1,0})$ is negative, the evidence supports M_0 over M_1. For $2 \times (\ln B_{1,0})$ between 0 and 2, there is "weak evidence" that M_1 is a more likely characterization of the data than M_0. Values of $2 \times (\ln B_{1,0})$ between 2 and 6, 6 and 10, and higher than 10, can be viewed as "some evidence," "strong evidence," and "very strong evidence" in support of M_1 relative to M_0.

Table 2. Out-of-sample forecasting performance: RMSFE and LSD values

Panel A: Inflation rate						
Models	**RMSFE ratio**			**LSD**		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
LIN	1.593	0.759	1.096	-1.949	-1.545	-1.674
TVP	0.999	0.998	0.999	0.004**	0.008***	0.006***
SV	0.947*	1.007	0.967	-0.012	0.408***	0.274***
TVP-SV	0.946	1.003	0.965	-0.055	0.425***	0.272***
MS, K=2	0.953*	1.019	0.975	0.046	0.261***	0.192***
MS, K=3	0.946*	1.016	0.969	0.061	0.305***	0.227***
CP, K=2	0.950	1.029	0.977	0.080	0.251***	0.196***
CP, K=3	0.963	1.026	0.984	0.075	0.270***	0.208***
CP, K=4	0.964	1.021	0.983	0.075	0.274***	0.210***
ROL	0.963	1.033	0.987	0.049	0.261***	0.193***
Panel B: Real GDP growth						
Models	**RMSFE**			**LSD**		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
LIN	4.919	2.233	3.331	-3.023	-2.464	-2.642
TVP	1.002	0.998	1.001	-0.000	0.004**	0.002*
SV	0.990**	0.989***	0.990**	-0.068	0.238***	0.140***
TVP-SV	0.990*	0.987***	0.989**	-0.121	0.261***	0.140***
MS, K=2	1.010	0.998	1.006	-0.017	0.133***	0.085***
MS, K=3	1.009	1.004	1.008	-0.014	0.141***	0.091***
CP, K=2	1.003	0.998	1.001	-0.018	0.098**	0.061**
CP, K=3	1.012	1.033	1.018	-0.031	0.146***	0.090**
CP, K=4	1.021	1.048	1.029	-0.042	0.142**	0.083**
ROL	1.018	1.012	1.017	-0.051	0.076**	0.036*

The left panels of this table report the ratio between the RMSFE of model i and the RMSFE of the linear (LIN) model, computed as

$$RMSFE_i = \frac{\sqrt{\frac{1}{\bar{t}-\underline{t}+1}\sum_{\tau=\underline{t}}^{\bar{t}} e_{i,\tau}^2}}{\sqrt{\frac{1}{\bar{t}-\underline{t}+1}\sum_{\tau=\underline{t}}^{\bar{t}} e_{LIN,\tau}^2}},$$

where $e_{i,\tau}^2$ and $e_{LIN,\tau}^2$ are the squared forecast errors at time τ generated by model i and the LIN model, respectively, and i denotes any of the models described in section 2. For the linear model in the first row, the actual RMSFE-value is reported. Values below one for $RMSFE_i$ indicate that model i produces more accurate point forecasts than the LIN model. The right panels in the table report the average log-score (LS) differential, $LSD_i = \sum_{\tau=\underline{t}}^{\bar{t}} \left(LS_{i,\tau} - LS_{LIN,\tau}\right)$, where $LS_{i,\tau}$ ($LS_{LIN,\tau}$) denotes the log-score of model i (LIN) computed at time τ and i denotes any of the models described in section 2. For the linear model in the top row, the mean of the actual log score is reported. Positive values of LSD_i indicate that model i produces more accurate density forecasts than the LIN model. All forecasts are generated out-of-sample using recursive estimates of the models. Stars refer to Diebold-Mariano p-values for the null that a particular model generates the same predictive performance as the benchmark LIN model. p-values are based on one-sided t-tests computed using a serial correlation robust variance and the pre-whitened quadratic spectral estimates of Andrews and Monahan (1992). The out-of-sample period starts in 1970:I and ends in 2013:IV. *significance at 10% level; **significance at 5% level; ***significance at 1% level.

Table 3. Out-of-sample forecasting performance, $H = 4$ and $H = 8$

Panel A: Inflation rate, $H = 4$

Models	RMSFE ratio			LSD		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
LIN	2.658	0.997	1.710	-2.407	-1.878	-2.046
TVP	1.000	1.001	1.000	-0.004	0.007***	0.003***
SV	0.905*	1.010	0.930	-0.187	0.449***	0.247**
TVP-SV	0.907*	1.015	0.933	-0.214	0.457***	0.243**
MS, K=2	0.923	0.976	0.936	-0.137	0.391***	0.223**
MS, K=3	0.907	1.083	0.951	-0.081	0.380***	0.233***
CP, K=2	0.904	1.069	0.945	-0.156	0.257***	0.126
CP, K=3	0.897	1.044	0.933	-0.015	0.280***	0.186***
CP, K=4	0.937	1.041	0.962	-0.087	0.347***	0.209***
ROL	0.888	1.047	0.927	-0.152	0.255***	0.125*

Panel B: Inflation rate, $H = 8$

Models	RMSFE ratio			LSD		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
LIN	3.344	1.356	2.193	-2.675	-2.011	-2.222
TVP	1.003	0.996	1.001	-0.000	0.004***	0.003***
SV	0.986	0.980	0.984	-0.335	0.318***	0.110
TVP-SV	0.988	0.977	0.985	-0.382	0.322***	0.098
MS, K=2	1.015	1.002	1.012	-0.191	0.207***	0.081
MS, K=3	1.033	1.137	1.061	-0.203	0.100	0.003
CP, K=2	1.011	0.974	1.001	-0.328	0.135***	-0.012
CP, K=3	1.009	0.964	0.997	-0.429	0.113*	-0.060
CP, K=4	1.020	0.968	1.007	-0.228	0.204***	0.067
ROL	1.002	0.940	0.986	-0.200	0.114**	0.014

Panel C: Real GDP growth, $H = 4$

Models	RMSFE			LSD		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
LIN	5.651	2.579	3.833	-3.167	-2.664	-2.824
TVP	1.001	1.000	1.001	-0.005	0.003***	0.000
SV	0.975***	0.982***	0.977***	0.022	0.258***	0.183***
TVP-SV	0.971***	0.981***	0.974***	0.016	0.276***	0.194***
MS, K=2	1.036	1.051	1.041	-0.017	0.128***	0.082***
MS, K=3	1.042	1.064	1.049	-0.027	0.120***	0.073***
CP, K=2	1.043	1.077	1.053	-0.040	0.071	0.036
CP, K=3	1.062	1.111	1.077	-0.080	0.079	0.028
CP, K=4	1.073	1.139	1.094	-0.095	0.104*	0.041
ROL	1.051	1.102	1.067	-0.074	0.032	-0.002

Panel D: Real GDP growth, $H = 8$

Models	RMSFE			LSD		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
LIN	6.118	2.859	4.181	-3.245	-2.657	-2.844
TVP	0.999	1.002	1.000	0.006	-0.000	0.002
SV	0.962***	0.974**	0.966***	-0.033	0.185**	0.115**
TVP-SV	0.952***	0.968**	0.957***	-0.048	0.195**	0.118**
MS, K=2	1.018	1.046	1.027	-0.008	0.013	0.006
MS, K=3	1.015	1.049	1.026	-0.007	0.006	0.002
CP, K=2	1.008	1.001	1.006	-0.013	0.019**	0.009
CP, K=3	1.019	0.991*	1.010	-0.040	0.026*	0.005
CP, K=4	1.043	1.024	1.037	-0.078	0.014	-0.016
ROL	1.029	1.098	1.051	-0.053	-0.056	-0.055

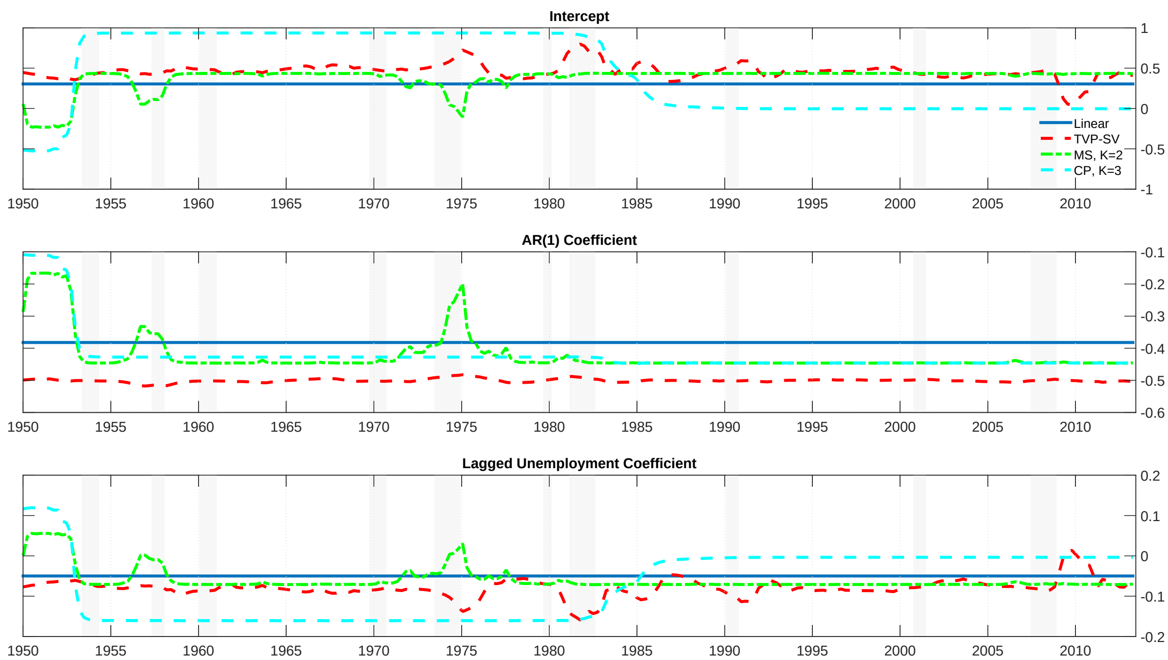
The left panels of this table report the ratio between the RMSFE of model i and the RMSFE of the linear (LIN) model, computed as $RMSFE_i = \sqrt{\frac{1}{\bar{t}-\underline{t}+1}\sum_{\tau=\underline{t}}^{\bar{t}} e^2_{i,\tau}} \Big/ \sqrt{\frac{1}{\bar{t}-\underline{t}+1}\sum_{\tau=\underline{t}}^{\bar{t}} e^2_{LIN,\tau}}$, where $e^2_{i,\tau}$ and $e^2_{LIN,\tau}$ are the squared forecast errors at time τ generated by model i and the LIN model, respectively, and i denotes any of the models described in section 2. For the linear model in the first row, the actual RMSFE-value is reported. Values below one for $RMSFE_i$ indicate that model i produces more accurate point forecasts than the LIN model. The right panels in the table report the average log-score (LS) differential, $LSD_i = \sum_{\tau=\underline{t}}^{\bar{t}} (LS_{i,\tau} - LS_{LIN,\tau})$, where $LS_{i,\tau}$ ($LS_{LIN,\tau}$) denotes the log-score of model i (LIN) computed at time τ and i denotes any of the models described in section 2. For the linear model in the top row, the mean of the actual log score is reported. Positive values of LSD_i indicate that model i produces more accurate density forecasts than the LIN model. All forecasts are generated out-of-sample using recursive estimates of the models. Stars refer to Diebold-Mariano p-values for the null that a particular model generates the same predictive performance as the benchmark LIN model. p-values are based on one-sided t-tests computed using a serial correlation robust variance and the pre-whitened quadratic spectral estimates of Andrews and Monahan (1992). The out-of-sample period starts in 1970:I and ends in 2013:IV. *significance at 10% level; **significance at 5% level; ***significance at 1% level.

Table 4. Out-of-sample forecasting performance for the model combinations

Panel A: Inflation rate, $H=1$						
Combination scheme	**RMSFE ratio**			**LSD**		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
Equal weighted combination	0.953*	1.006	0.971*	0.081*	0.241***	0.190***
Bayesian model averaging	0.972	1.005	0.983	0.061	0.347***	0.256***
Optimal prediction pool	0.963	1.002	0.976	0.012	0.362***	0.251***
Panel B: Inflation rate, $H=4$						
Combination scheme	**RMSFE ratio**			**LSD**		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
Equal weighted combination	0.924*	1.000	0.942	0.036	0.281***	0.203***
Bayesian model averaging	0.995	1.010	0.999	-0.011	0.344***	0.231***
Optimal prediction pool	0.948*	1.005	0.962	0.001	0.364***	0.248***
Panel C: Inflation rate, $H=8$						
Combination scheme	**RMSFE ratio**			**LSD**		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
Equal weighted combination	1.005	0.986	1.000	-0.041	0.163***	0.098**
Bayesian model averaging	1.008	1.016	1.010	-0.012	0.133**	0.087**
Optimal prediction pool	1.001	0.993	0.999	-0.035	0.206***	0.129***
Panel D: Real GDP growth, $H=1$						
Combination scheme	**RMSFE**			**LSD**		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
Equal weighted combination	1.003	0.997	1.001	-0.016	0.133***	0.086***
Bayesian model averaging	1.005	0.996	1.002	-0.010	0.190***	0.126***
Optimal prediction pool	1.003	0.987*	0.998	-0.044	0.208***	0.128***
Panel E: Real GDP growth, $H=4$						
Combination scheme	**RMSFE**			**LSD**		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
Equal weighted combination	1.025	1.030	1.026	-0.004	0.130***	0.088***
Bayesian model averaging	1.000	0.982***	0.994	-0.008	0.271***	0.182***
Optimal prediction pool	0.985*	0.982***	0.984**	-0.000	0.270***	0.184***
Panel E: Real GDP growth, $H=8$						
Combination scheme	**RMSFE**			**LSD**		
	1970-1983	**1984-2013**	**1970-2013**	**1970-1983**	**1984-2013**	**1970-2013**
Equal weighted combination	1.004	1.005	1.004	-0.005	0.055***	0.036***
Bayesian model averaging	1.007	0.986**	1.000	-0.010	0.165**	0.110**
Optimal prediction pool	0.988**	0.979***	0.985***	-0.035	0.183***	0.114***

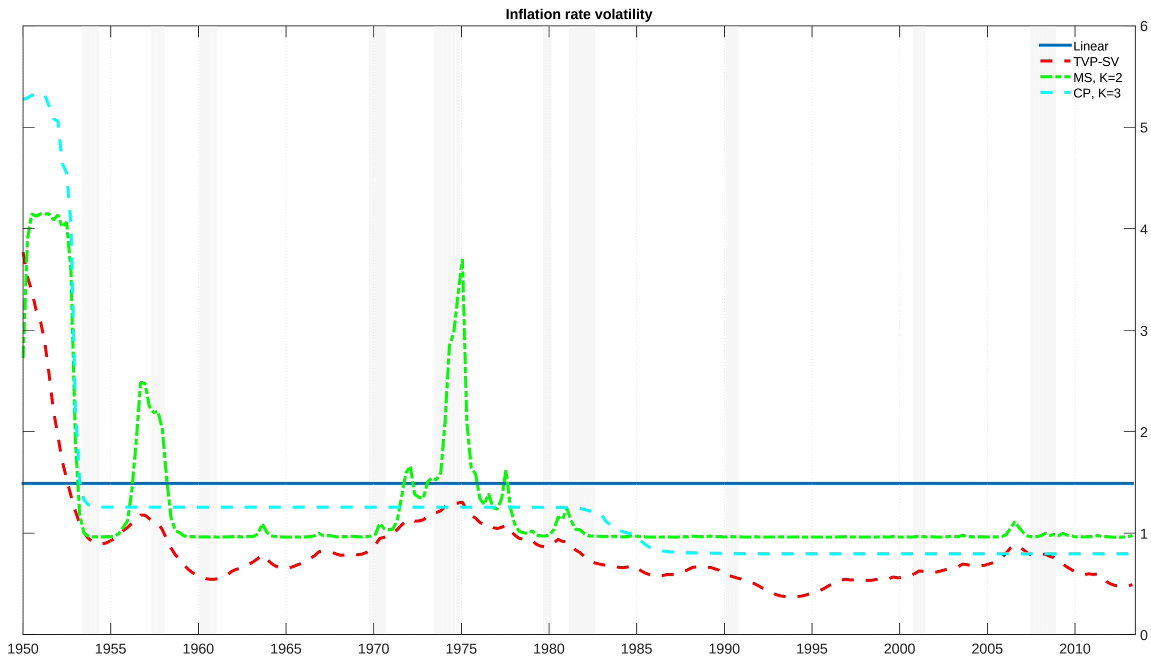
Left panels in this table report the RMSFE of the equal-weighted model combination scheme, the optimal predictive pool of Geweke and Amisano (2011), or Bayesian Model Averaging computed relative to the RMSFE for the linear (LIN) model: $RMSFE_i = \sqrt{\frac{1}{\bar{t}-\underline{t}+1}\sum_{\tau=\underline{t}}^{\bar{t}} e_{i,\tau}^2} \Big/ \sqrt{\frac{1}{\bar{t}-\underline{t}+1}\sum_{\tau=\underline{t}}^{\bar{t}} e_{LIN,\tau}^2}$, where $e_{i,\tau}^2$ and $e_{LIN,\tau}^2$ are the squared forecast errors at time τ generated by model combination i and the LIN model, respectively, and i refers to one of the model combination schemes. Values of the $RMSFE_i$ ratio below one indicate that model combination i produces more accurate point forecasts than the LIN model. The right panels report the average log-score (LS) differential, $LSD_i = \sum_{\tau=\underline{t}}^{\bar{t}} \left(LS_{i,\tau} - LS_{LIN,\tau}\right)$, where $LS_{i,\tau}$ ($LS_{LIN,\tau}$) denotes the log-score of model combination i (the LIN model), computed at time τ, and i denotes either the equal-weighted model combination scheme, the optimal predictive pool of Geweke and Amisano (2011), or Bayesian Model Averaging. Positive values of LSD_i indicate that model combination i produces more accurate density forecasts than the LIN model. All forecasts are generated out-of-sample using recursive estimates of the models and combination weights. Stars refer to Diebold-Mariano p-values for the null that a particular model combination generates the same predictive performance as the benchmark LIN model. p-values are based on one-sided t-tests computed with a serial correlation robust variance, using the pre-whitened quadratic spectral estimates of Andrews and Monahan (1992). The out-of-sample period starts in 1970:I and ends in 2013:IV. *significance at 10% level; **significance at 5% level; ***significance at 1% level.

Figure 1. **Coefficient estimates for the quarterly inflation rate**



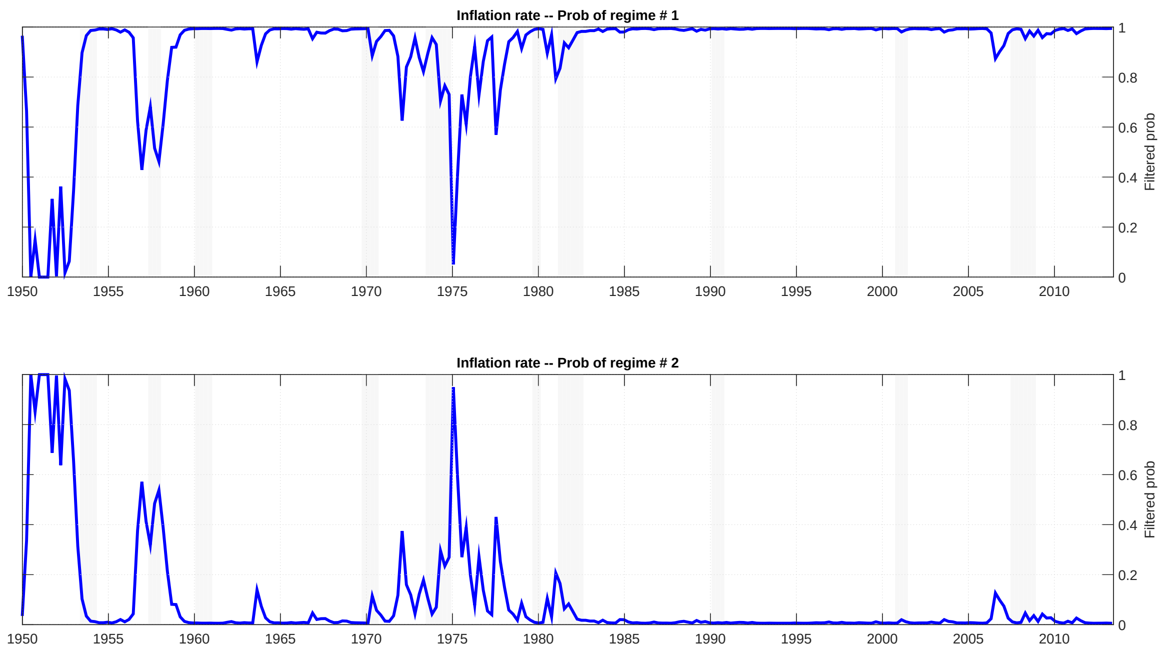
This figure plots the estimated coefficients for the inflation rate model over the sample 1950Q1-2013Q4. The top panel plots estimates of the intercept, the middle panel plots estimates of the AR(1) term, and the bottom panel plots estimates of the coefficient on the lagged unemployment rate. The solid blue line tracks the linear model, the red dashed line tracks the TVP-SV model, the green dashed/dotted line tracks the Markov switching model with two regimes, and the dashed light blue line tracks the change point model with three regimes.

Figure 2. **Volatility estimates for quarterly inflation**



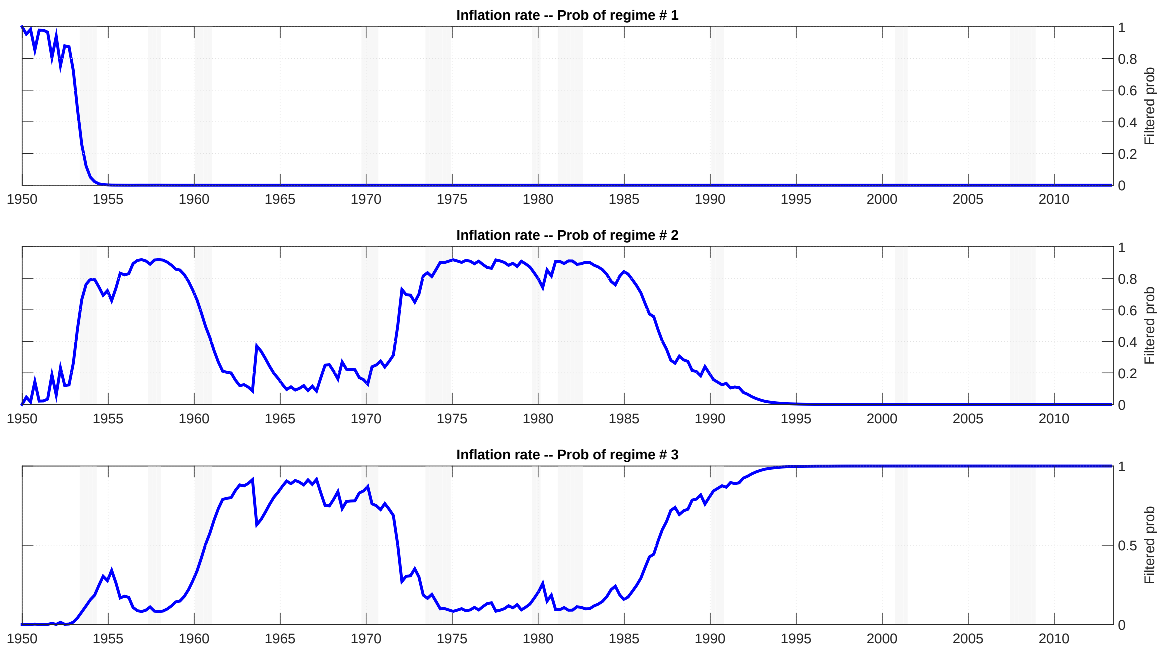
This figure plots volatility estimates for the inflation series over the sample 1950Q1-2013Q4. The Blue solid line tracks the linear model, the red dashed line tracks the TVP-SV model, the green dashed/dotted line tracks the MS model with two regimes, and the dashed light blue line tracks the CP model with three regimes.

Figure 3. **Filtered state probabilities for the two-state MS model fitted to quarterly inflation**



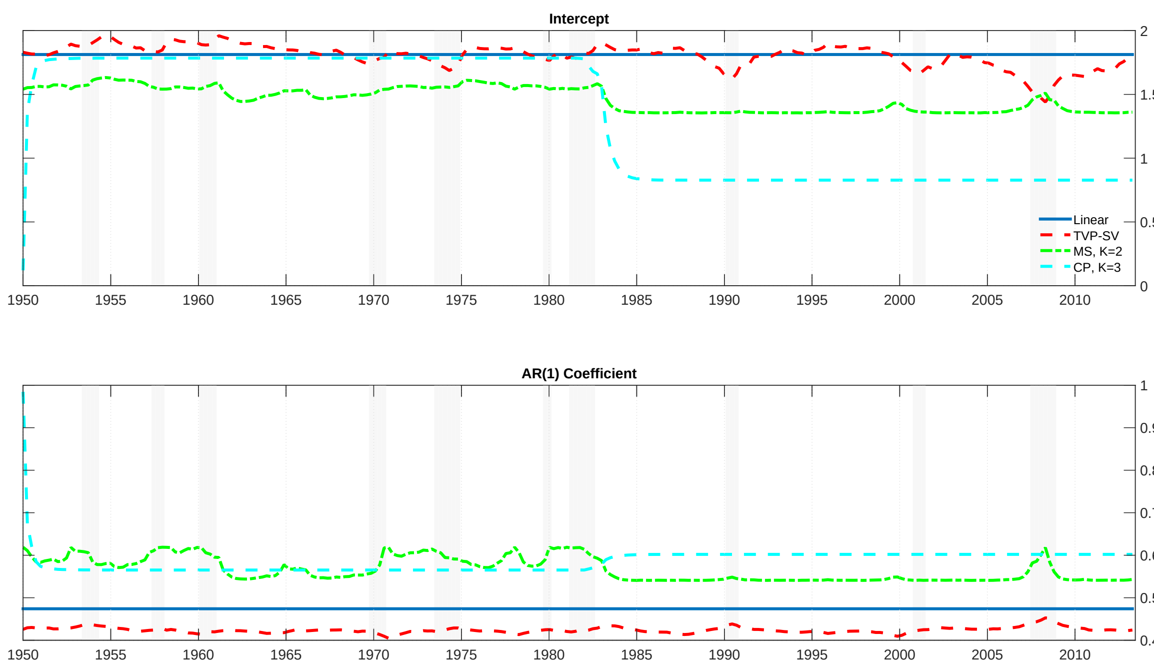
This figure plots the time series of filtered state probabilities for the two-state Markov switching model fitted to the quarterly inflation series over the sample 1950Q1-2013Q4.

Figure 4. **Filtered state probabilities for the three-state CP model fitted to quarterly inflation**



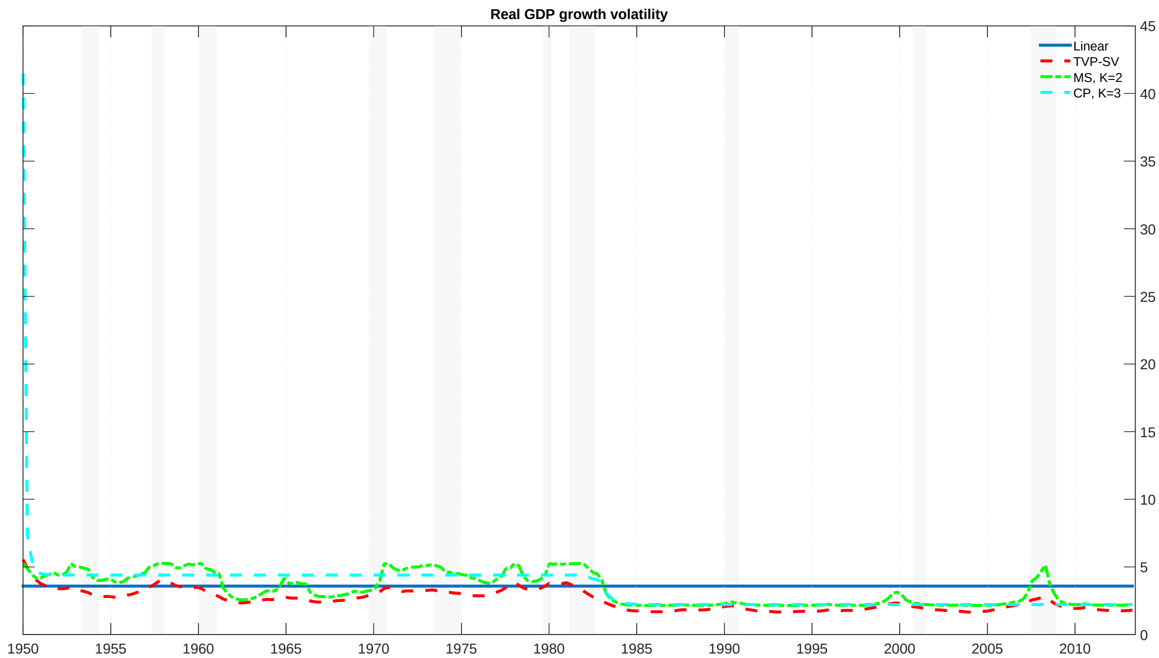
This figure plots the time series of filtered state probabilities for the three-state change point model fitted to the quarterly inflation series over the sample 1950Q1-2013Q4.

Figure 5. **Coefficient estimates for quarterly growth in real GDP**



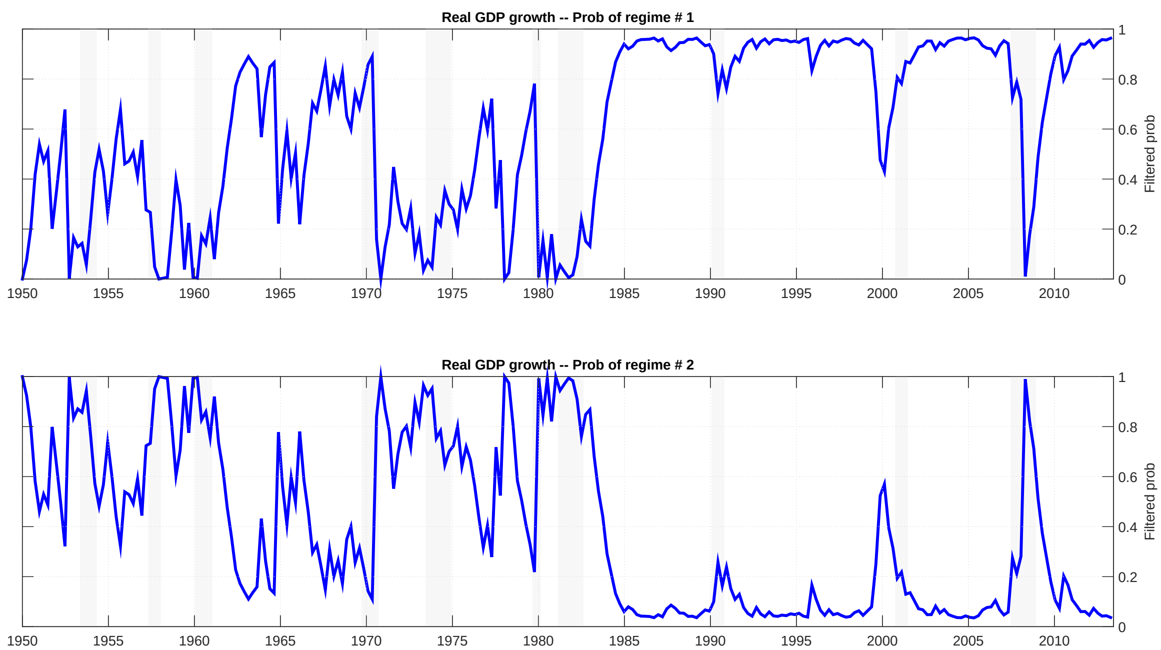
This figure plots the estimated coefficients for real GDP growth over the sample 1950Q1-2013Q4. The top panel plots estimates of the intercept and the bottom panel plots estimates of the AR(1) term. The solid blue line refers to the linear model, the dashed red line refers to the TVP-SV model, the dashed/dotted green line refers to the Markov switching model with two regimes, and the dashed light blue line refers to the change point model with three states.

Figure 6. **Volatility estimates for real GDP growth**



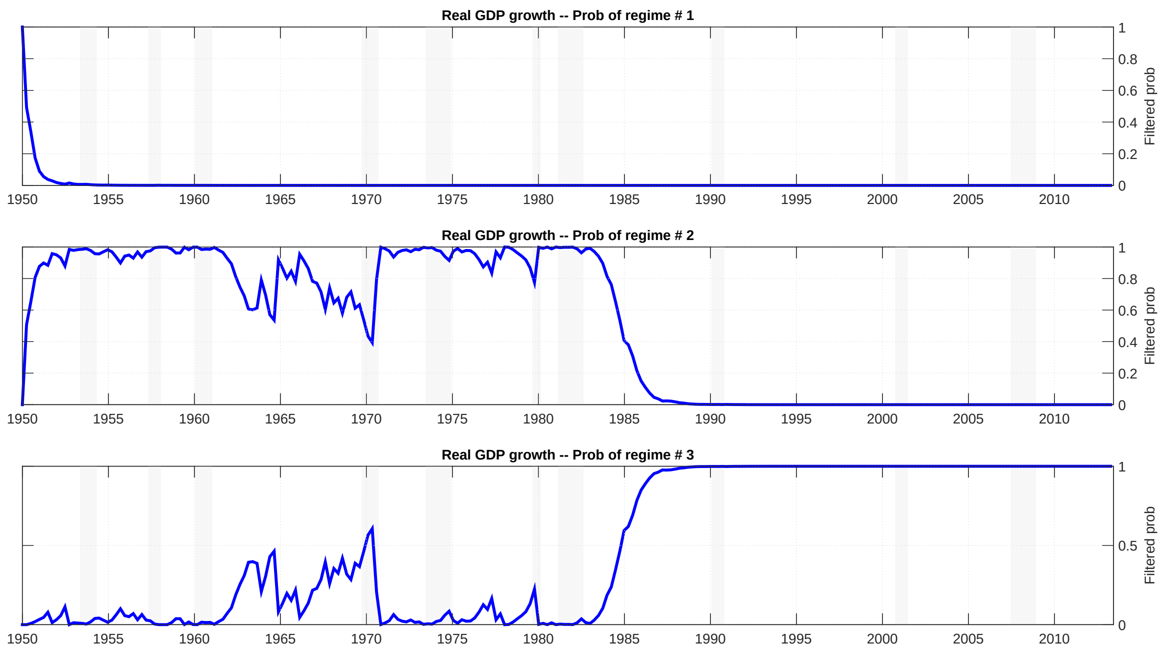
This figure plots volatility estimates for quarterly growth in real GDP over the sample 1950Q1-2013Q4. The Blue solid line tracks the linear model, the red dashed line tracks the TVP-SV model, the green dashed/dotted line tracks the MS model with two regimes, and the dashed light blue line tracks the CP model with three regimes.

Figure 7. **Filtered state probabilities for the two-state MS model fitted to real GDP growth**



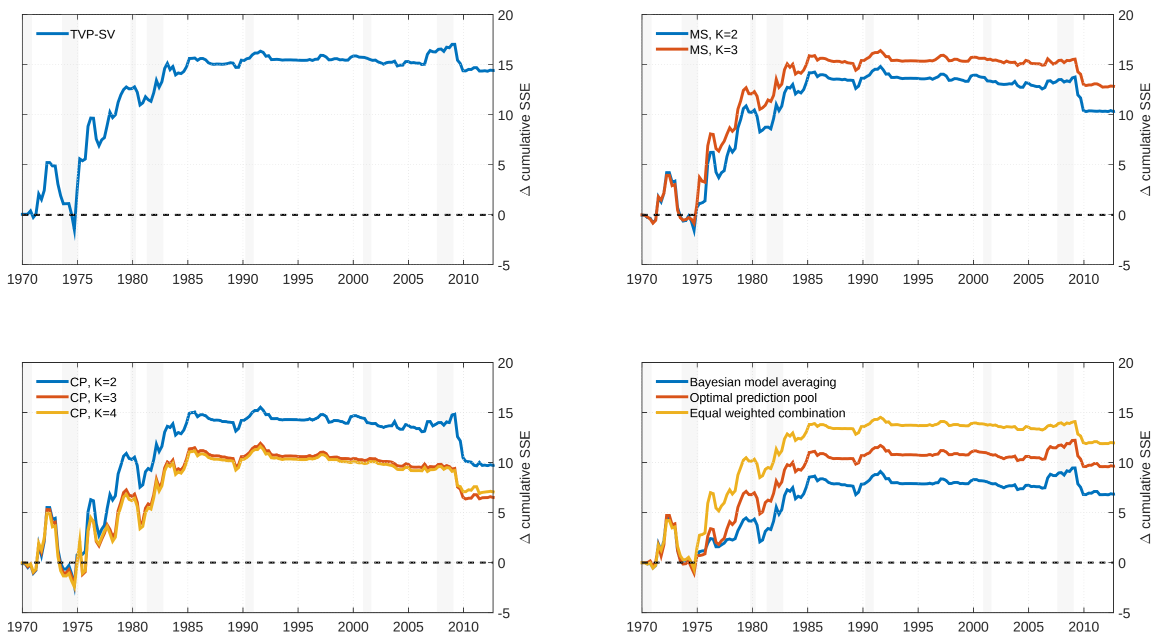
This figure plots the time series of filtered state probabilities for the two-state Markov switching model fitted to real GDP growth over the sample 1950Q1-2013Q4.

Figure 8. **Filtered state probabilities for the three-state CP model fitted to real GDP growth**



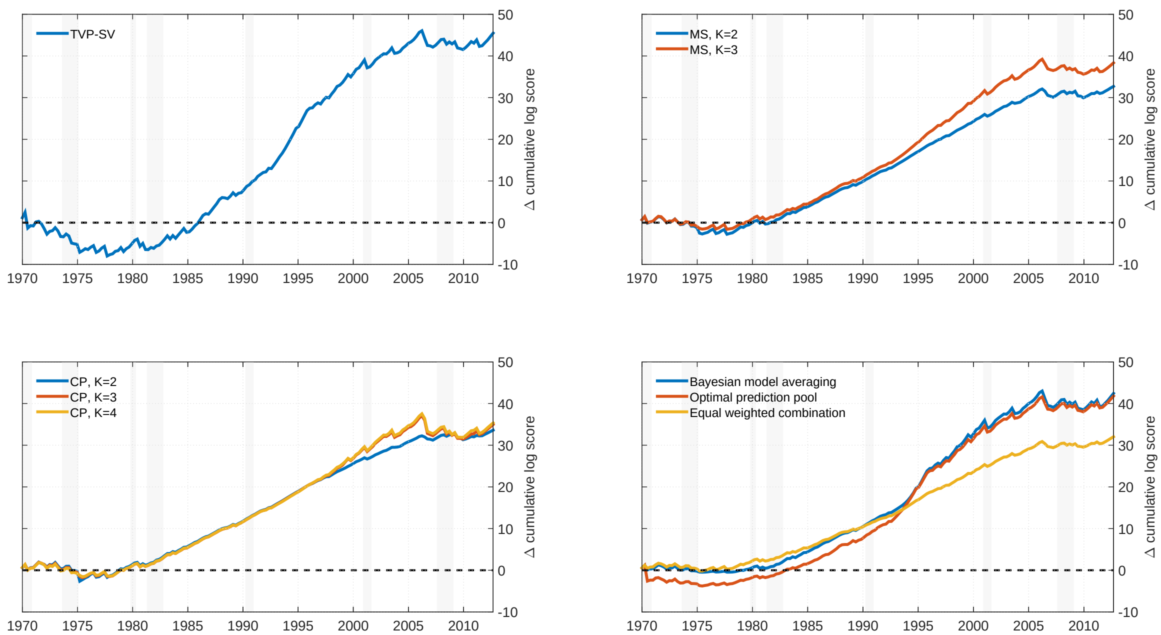
This figure plots the time series of filtered state probabilities for the three-state change point model fitted to real GDP growth over the sample 1950Q1-2013Q4.

Figure 9. **Cumulative sum of squared forecast error differentials: Quarterly inflation**



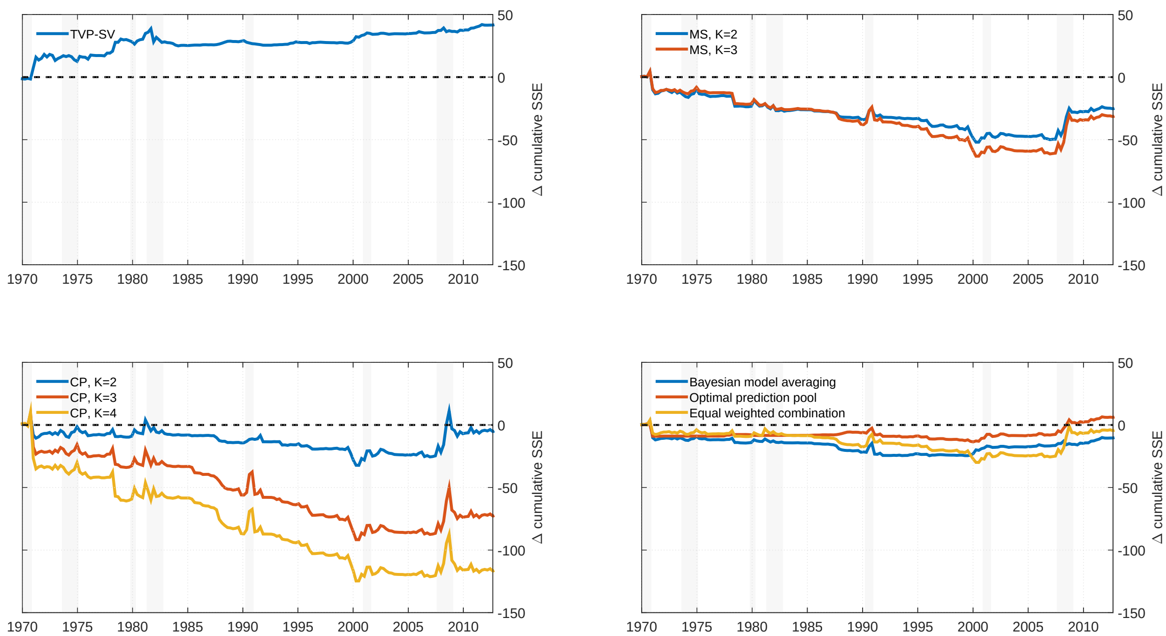
This figure shows the sum of squared forecast errors generated by the linear model model minus the sum of squared forecast errors generated by a set of alternative models, $CSSED_{i,t} = \sum_{\tau=\underline{t}}^{t}\left(e^2_{LIN,\tau} - e^2_{i,\tau}\right)$. Values above zero indicate that a model generates better performance than the linear benchmark, while negative values suggest the opposite. Each panel displays results for different types of models, with TVP-SV models in the top left panel, MS models in the top right panel, CP models in the bottom left panel, and the model combinations in the bottom right panel. Shaded areas indicate NBER-dated recessions.

Figure 10. **Cumulative sum of log-score differentials: Quarterly inflation**



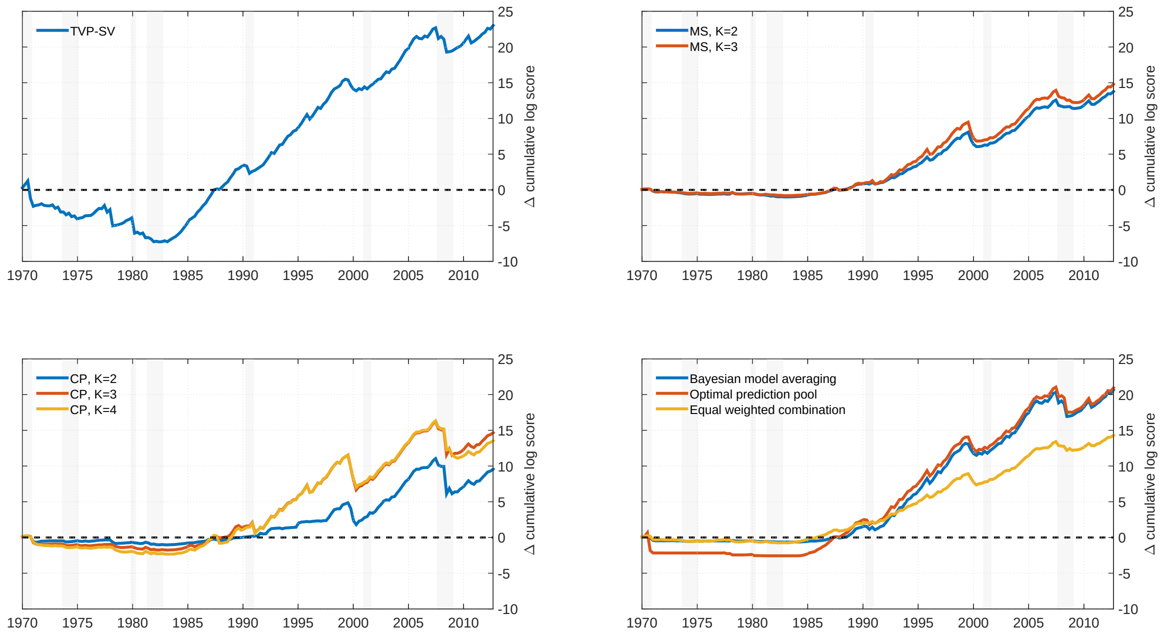
Notes: This figure shows the sum of log predictive scores generated by a set of alternative model specifications minus the sum of log predictive scores generated by the linear model, $CLSD_{i,t} = \sum_{\tau=\underline{t}}^{t} \left(LS_{i,\tau} - LS_{LIN,\tau} \right)$. Values above zero indicate that a model generates more accurate forecasts than the linear benchmark, while negative values suggest the opposite. Each panel displays results for different types of models, with TVP-SV models in the top left panel, MS models in the top right panel, CP models in the bottom left panel, and model combinations in the bottom right panel. Shaded areas indicate NBER-dated recessions.

Figure 11. **Cumulative sum of squared forecast error differentials: Real GDP growth**



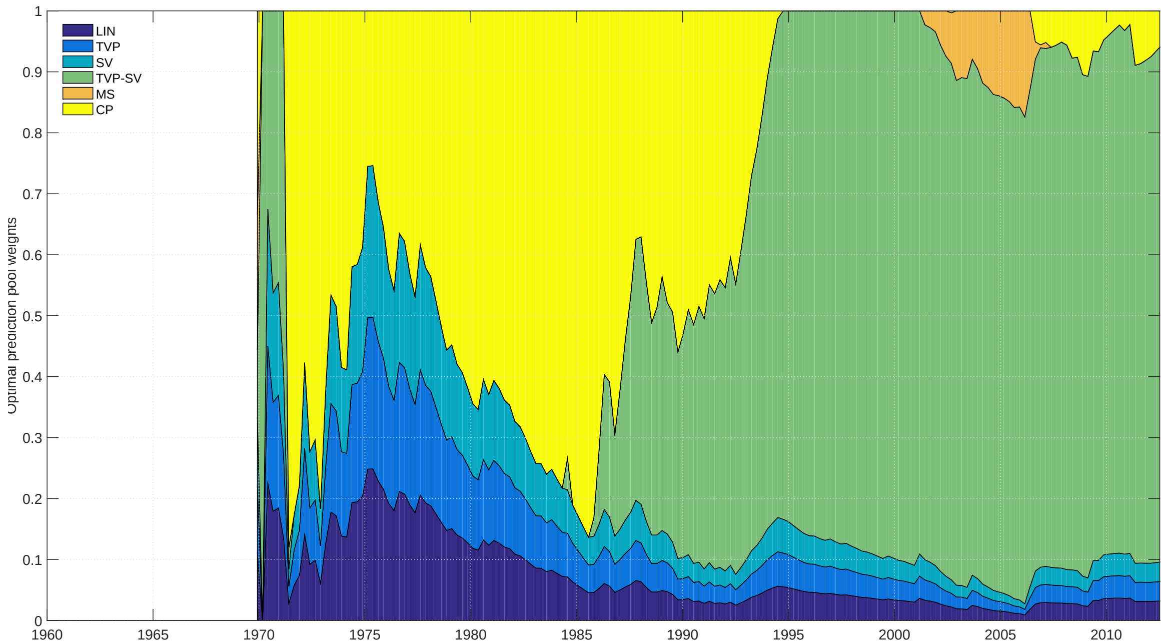
This figure shows the sum of squared forecast errors generated by the linear model model minus the sum of squared forecast errors generated by a set of alternative models, $CSSED_{i,t} = \sum_{\tau=\underline{t}}^{t} \left(e_{LIN,\tau}^2 - e_{i,\tau}^2 \right)$. Values above zero indicate that a model generates better performance than the linear benchmark, while negative values suggest the opposite. Each panel displays results for different types of models, with TVP-SV models in the top left panel, MS models in the top right panel, CP models in the bottom left panel, and the model combinations in the bottom right panel. Shaded areas indicate NBER-dated recessions.

Figure 12. **Cumulative sum of log-score differentials: Real GDP growth**



Notes: This figure shows the sum of log predictive scores generated by a set of alternative model specifications minus the sum of log predictive scores generated by the linear model, $CLSD_{i,t} = \sum_{\tau=\underline{t}}^{t} \left(LS_{i,\tau} - LS_{LIN,\tau} \right)$. Values above zero indicate that a model generates more accurate forecasts than the linear benchmark, while negative values suggest the opposite. Each panel displays results for different types of models, with TVP-SV models in the top left panel, MS models in the top right panel, CP models in the bottom left panel, and model combinations in the bottom right panel. Shaded areas indicate NBER-dated recessions.

Figure 13. **Probability weights on different classes of models in the optimal prediction pool: Quarterly inflation**

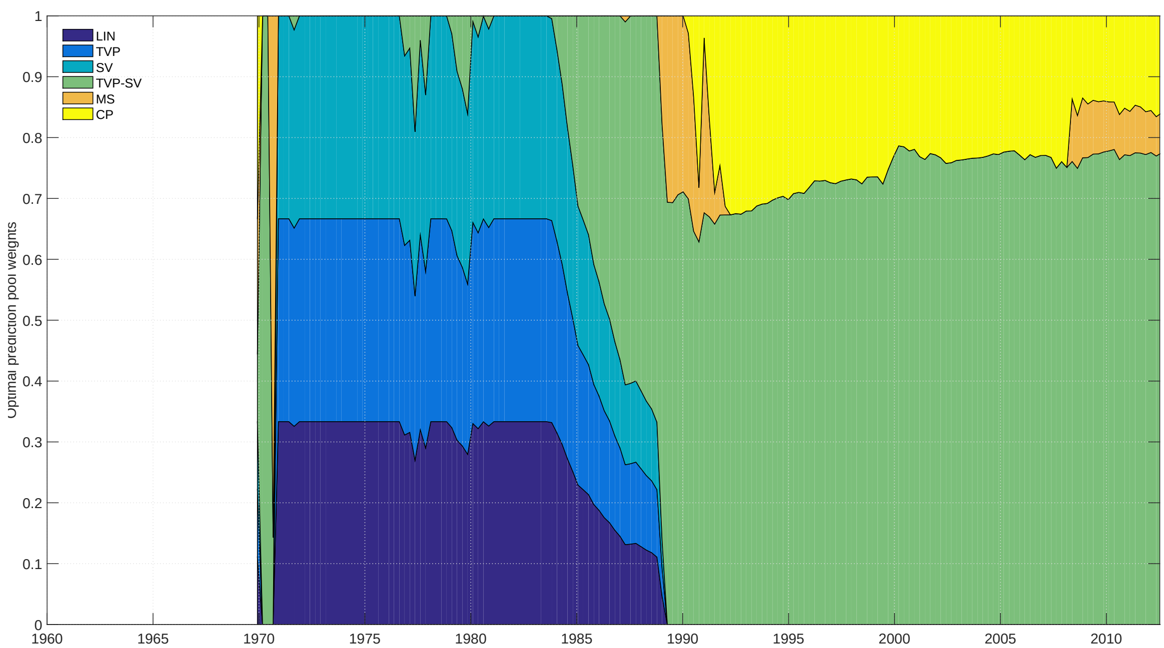


This figure plots recursively calculated weights on different classes of models in the predictive pool for the quarterly inflation series. The weights are computed by solving the minimization problem

$$\mathbf{w}_t^* = \arg\max_{\mathbf{w_t}} \sum_{\tau=1}^{t-1} \log\left[\sum_{i=1}^{N} w_{it} \times S_{\tau+1,i}\right]$$

where $N = 7$ different models are considered, and the solution is found subject to $\mathbf{w}_t^*$ belonging to the $N-$dimensional unit simplex. $S_{\tau+1,i}$ denotes the time $\tau+1$ recursively computed log score for model i, i.e., $S_{\tau+1,i} = \exp\left(LS_{\tau+1,i}\right)$. Dark blue bars show the weights on the linear model in the optimal prediction pool, light blue bars show the weights assigned to the TVP-SV model, yellow bars show the weights on the MS models, and maroon bars show the weights assigned to the CP models.

Figure 14. **Probability weights on different classes of models in the optimal prediction pool: Real RGDP growth**



This figure plots recursively calculated weights on different classes of models in the predictive pool for real GDP growth. The weights are computed by solving the minimization problem

$$\mathbf{w}_t^* = \arg\max_{\mathbf{w_t}} \sum_{\tau=1}^{t-1} \log \left[\sum_{i=1}^{N} w_{it} \times S_{\tau+1,i} \right]$$

where $N = 7$ different models are considered, and the solution is found subject to $\mathbf{w}_t^*$ belonging to the $N-$dimensional unit simplex. $S_{\tau+1,i}$ denotes the time $\tau+1$ recursively computed log score for model i, i.e., $S_{\tau+1,i} = \exp\left(LS_{\tau+1,i}\right)$. Dark blue bars show the weights on the linear model in the optimal prediction pool, light blue bars show the weights assigned to the TVP-SV model, yellow bars show the weights on the MS models, and maroon bars show the weights assigned to the CP models.