

Have Standard VARs Remained Stable Since the Crisis?*

Knut Are Aastveit
Norges Bank
knut-are.aastveit@norges-bank.no

Andrea Carriero
Queen Mary, University of London
a.carriero@qmul.ac.uk

Todd E. Clark
Federal Reserve Bank of Cleveland
todd.clark@clev.frb.org

Massimiliano Marcellino
Bocconi University, IGIER and CEPR
massimiliano.marcellino@unibocconi.it

August 2016

Abstract

Small VARs are commonly used in macroeconomics for forecasting and evaluating shock transmission. This requires VAR parameters to be stable over the evaluation and forecast sample or modeled as time-varying. Prior work has considered whether there were sizable parameter changes in the early 1980s and in the subsequent period until the beginning of the new century. This paper conducts a similar analysis focused on period since the recent crisis. Using a range of techniques, we provide substantial evidence against parameter stability. The evolution of the unemployment rate seems particularly different relative to its past behavior. We also evaluate alternative methods to handle parameter instability in a forecasting context.

Keywords: Model stability, structural breaks, forecasting

JEL classification code: E17, C11, C33, C53

*We gratefully acknowledge helpful comments from Editor Jonathan Wright, two anonymous referees, Sandra Eickmeier, Domenico Giannone, Mike McCracken, Haroon Mumtaz, and seminar participants at Norges Bank, the 7th CSDA International Conference on Computational and Financial Economics, the 20th International Conference on Computing in Economics and Finance, the International Association for Applied Econometrics 2014 Annual Conference, the 34th International Symposium on Forecasting and the 25th EC² conference on advances in forecasting. The views expressed herein are solely those of the authors and do not necessarily reflect the views of Norges Bank, the Federal Reserve Bank of Cleveland or the Federal Reserve System.

1 Introduction

The economies of the United States and many other industrialized countries experienced a number of sharp changes in the post-war period. In particular, the 1970's were characterized by large supply shocks and wide economic fluctuations, with a sharp reduction in volatility of most macroeconomic variables between the mid-1980's and up to the end of 2007, often referred to as the *Great Moderation*. Reflecting the influences of these changes, studies of the period up through 2007, including Doan, Litterman, and Sims (1984), Stock and Watson (1996), Clark and McCracken (2008), Clark (2011), and D'Agostino, Gambetti, and Giannone (2013), have emphasized the importance of allowing for parameter changes in either the conditional mean or variance parameters of standard vector autoregressive (VAR) models.¹

The recent Great Recession (the most severe recession in the United States' postwar period), financial crisis, and sluggish recovery could represent another event that causes the parameters of standard VAR models to shift. One possible reason, articulated by Reinhart and Rogoff (2009) and Hall (2010), is that financial recessions and recoveries are different from "normal" recessions and recoveries. Ng and Wright (2013) provide an overview of the facts and explanations of recent recessions from the perspective of macro-econometricians and reach a similar conclusion: recessions that originate with financial market dislocations are distinctively different from those in which financial markets play a passive role, and recoveries are typically slow when the preceding recessions have financial origins.

Recent studies of the stability of factor models provide mixed evidence on the notion that the recent recession and recovery may have been different enough to create a break in time series models. The evidence in Stock and Watson (2012) suggests the stability of a factor model for U.S. data: Stock and Watson find that the 2007-2009 recession was the result of one or more large shocks with no evidence of changes in the response of macroeconomic variables. However, Cheng, Liao, and Schorfheide (2014) find significant instabilities (in a factor model) associated with the most recent recession and recovery, based on new methods for testing the stability of factor models in the period since 1985.

Building on this past work, in this paper we examine the stability of common VARs in

¹Barnett, Mumtaz and Theodoridis (2014) emphasize the importance of allowing for parameter changes based on the evaluation of VAR forecasts for the UK in data through 2010. Pettenuzzo and Timmermann (2015) focus on the forecasting performance of univariate models that accommodate structural change, using U.S. data and a sample ending with 2013.

the period since the Great Recession, focusing on small and medium-size VARs of the type commonly used for forecasting and policy analysis. Does the evolution of macro variables for the U.S. over the 2008-2015 period represent a break in VAR dynamics?

We first use a variety of approaches to assess model stability around the Great Recession, all of which suggest significant instabilities. First, we conduct formal Bayesian inference for a single break in data since 1985, obtaining strong evidence of parameter instability in about 2008. Second, building on the Sims and Zha (2006) analysis of changes in macroeconomic dynamics, we estimate a regime-switching VAR, similar in spirit to the models in Pesaran, Pettenuzzo and Timmermann (2006) and Barnett, Mumtaz and Theodoridis (2014). The regime-switching estimates provide additional evidence of a break in the VAR coefficients around 2008. Finally, we examine BVAR forecasts conditioned on the actual path of GDP — conditioning intended to control for the normal effects of the business cycle — and find large forecast errors that appear consistent with a break in the model’s parameters.

We then examine the efficacy of a range of forecasting methods that can be used to deal with structural change.² Specifically, we consider forecasts from VARs with time varying parameters and volatility (TVP-SV VARs) and VARs estimated with different samples: recursive, starting in 1985; rolling, 20 year window; and a Pesaran and Pick (2011)–type average of forecasts computed over a range of rolling window sizes of 8, 10, 12, 14, . . . , 30 years. We gauge efficacy on the basis of the accuracy of both point and density forecasts. Overall, none of the methods clearly emerges as best. Among the methods that accommodate structural change, allowing for time-varying parameters (TVP-SV) seems to perform best with a small model, whereas with the large model, the average rolling window approach offers the most consistent gains on the baseline. However, in the absence of a strong, consistent forecast accuracy advantage of one method over another, one might argue that, in the presence of a series of past and possibly future instabilities, the approach of incorporating TVP in a model estimated over the available history of data may be conceptually preferable, because it does not hinge on the ability to identify a specific break date.

Our analysis can be seen as extending previous work on the stability of VAR forecasting models and efficacy of various methods for accommodating parameter change. We extend this line of work by focusing on the period since the Great Recession, presenting some evidence of a break in VAR dynamics around the Great Recession. Moreover, our analysis

²Alessi, et al. (2014) examine the performance of central bank forecasts during the crisis.

complements recent studies by Stock and Watson (2012) and Cheng, Liao and Schorfheide (2014) on macroeconomic stability since the Great Recession. While these papers rely on a factor model framework, we consider small and medium-scale VAR models commonly used for forecasting in policy institutions.

The remainder of the paper is organized as follows. Section 2 presents the models and conditional forecast methodology. Section 3 details the data. Section 4 provides evidence of instability around 2008 in terms of Bayesian analysis of breaks in coefficients and conditional forecasts. Section 5 reports results from a real-time out-of-sample forecast comparison of alternative approaches to handling instabilities. Section 6 concludes.

2 Models and Forecast Methodology

Reflecting the models most commonly used in forecasting and policy analysis, we focus on linear BVARs. To more directly accommodate structural change, we also consider VARs with time-varying parameters, as in such studies as Cogley and Sargent (2005), Primiceri (2005), D’Agostino, Gambetti, and Giannone (2013), and Koop and Korobilis (2013), and regime-switching VARs, as in such studies as Sims and Zha (2006), Pesaran, Pettenuzzo and Timmermann (2006) and Barnett, Mumtaz and Theodoridis (2014). This section first details the models used in our analysis and then presents the approach to generating the conditional forecasts we use in some of the analysis.

2.1 Constant parameter VAR specifications

Let y_t denote the $k \times 1$ vector of model variables of interest, B_0 contain a $k \times 1$ vector of intercepts, and $B_i, i = 1, \dots, p$, denote a $k \times k$ matrix of coefficients on lag i . The VAR(p) model with a constant variance-covariance matrix of shocks takes the form:

$$y_t = B_0 + \sum_{i=1}^p B_i y_{t-i} + v_t, \quad v_t \sim N(0, \Phi). \quad (1)$$

We base most of our results on a baseline VAR in GDP growth, the unemployment rate, core PCE inflation, and the federal funds rate.³ This type of specification is common in analysis of small macroeconomic VARs. For some results, we augment the model to include another variable, such as growth in payroll employment or a credit spread defined as the

³We use the core PCE measure of inflation rather than alternatives such as inflation in the GDP deflator because the Federal Reserve focuses on PCE prices.

BAA corporate bond rate less the 10-year Treasury yield. Unless otherwise noted, the small BVARs include four lags.

We also include some results for a 13-variable BVAR in levels or log levels, specified to include GDP, consumption, business fixed investment, residential investment, industrial production, capacity utilization, employment, unemployment, headline PCE inflation, core PCE inflation, federal funds rate, credit spread, and S&P 500 index of stock prices. In this specification, patterned after those used in such studies as Sims and Zha (1998) and Banbura, Giannone, and Reichlin (2010), all variables except capacity utilization, the unemployment rate, funds rate, and credit spread are specified in log levels. This larger BVAR includes five lags.

We estimate constant parameter BVARs of the form (1) using a normal-inverted Wishart prior and posterior. The basic prior on the VAR coefficients takes the Minnesota-style form described in sources such as Sims and Zha (1998), without cross-variable shrinkage (note that i and j refer to the row and column of B_l):

$$\underline{\mu}_B \text{ such that } E[B_l^{(ij)}] = 0 \quad \forall i, j, l \quad (2)$$

$$\underline{\Omega}_B \text{ such that } V[B_l^{(ij)}] = \begin{cases} \frac{\theta^2}{l^2} \frac{\sigma_i^2}{\sigma_j^2} & \text{for } l > 0 \\ \varepsilon^2 \sigma_i^2 & \text{for } l = 0 \end{cases} \quad (3)$$

For the four and five variable models (which use growth rates of trending variables), prior means are set to zero for all coefficients. In all of our forecasting results, we set the hyperparameter for overall shrinkage at the value that maximizes the marginal likelihood, based on a search across values of 0.1, 0.2, . . . , 0.9, 1.0. In most samples, the optimal shrinkage hyperparameter is 0.4. In some of our other results, we also fix the shrinkage hyperparameter at the common value of 0.2.

For the 13-variable model specified in levels and log levels, we set the prior means to imply random walks for all variables, putting a mean of 1 on the coefficient $B_1^{(ii)}$ and 0's on all other coefficients. We also supplement the usual Minnesota prior with the “sum of coefficients” and “dummy initial observation” priors proposed in Doan, Litterman, and Sims (1984) and Sims (1993), respectively. Both these priors can be implemented by augmenting the VAR system with dummy observations, as detailed in such sources as Sims and Zha (1998). In all of our forecasting results, we set the hyperparameters governing overall shrinkage, the tightness on the sum of coefficients prior, and the tightness on the cointegration prior at the set of values that maximize the marginal likelihood, based on a

search across a grid that included all combinations of values of 0.1, 0.2, ..., 0.9, 1.0 for each of the three hyperparameters.⁴

2.2 Time-varying methods

To directly accommodate structural change, we also consider both four and 13 variable VARs with TVP and time-varying volatility and four variable regime-switching VARs (RS-VAR), using the sets of variables detailed above.

2.2.1 Time-varying VAR

In the more tractable case of the small model, our specification corresponds to that of Cogley and Sargent (2005), modified to allow innovations to volatility to be correlated, as in Primiceri (2005):

$$\begin{aligned}
 y_t &= B_{0,t} + \sum_{i=1}^p B_{i,t} y_{t-1} + v_t \\
 B_t &= B_{t-1} + n_t, \quad \text{var}(n_t) = Q \\
 v_t &= A^{-1} \Lambda_t^{0.5} \epsilon_t, \quad \epsilon_t \sim N(0, I_k) \\
 \Lambda_t &= \text{diag}(\lambda_{1,t}, \dots, \lambda_{k,t}) \\
 \log(\lambda_{i,t}) &= \log(\lambda_{i,t-1}) + \nu_{i,t}, \\
 \nu_t &\equiv (\nu_{1,t}, \nu_{2,t}, \dots, \nu_{k,t})' \sim N(0, \Phi),
 \end{aligned}$$

where B_t denotes a vector containing the coefficients of $B_{0,t}$ and $B_{i,t}$, $i = 1, \dots, p$. We include two lags in the small model, following studies such as Cogley and Sargent (2005).

In the case of our 13 variable model, a fully proper Bayesian approach is not computationally feasible. Instead, we rely on the specification of Koop and Korobilis (2013) (hereafter, K-K), which introduces shortcuts to make computation tractable. In this case, the model takes the form

$$\begin{aligned}
 y_t &= B_{0,t} + \sum_{i=1}^p B_{i,t} y_{t-1} + v_t, \quad \text{var}(v_t) = \Sigma_t \\
 B_t &= B_{t-1} + n_t, \quad \text{var}(n_t) = Q_t.
 \end{aligned}$$

To facilitate computations associated with time-varying parameters, K-K rely on forgetting factors, which simplify Kalman filtering by replacing the usual formulae for the state

⁴However, to streamline computations in the construction of the average window-based forecasts from the 13 variable model, we fixed the hyperparameters at values of 0.2 for overall tightness and 1 for the other two hyperparameters, the settings used by Sims and Zha (1998). Carriero, Clark, and Marcellino (2015a) found the forecast accuracy payoff to optimization to be fairly small.

variance with $V_{t|t-1} = \frac{1}{\lambda}V_{t-1|t-1}$, which eliminates the need to estimate or simulate the innovation variance matrix Q_t . For forecasting, K-K abstract from Kalman smoothing. To streamline computations associated with stochastic volatility, K-K use an exponentially-weighted moving average to model time variation in the variance of innovations v_t to the VAR:

$$\hat{\Sigma}_t = \kappa\hat{\Sigma}_{t-1} + (1 - \kappa)\hat{\epsilon}_t\hat{\epsilon}_t'$$

where the innovation $\hat{\epsilon}_t = y_t - X_t'B_{t|t}$ is obtained with the Kalman filter.

In the implementation, for the small model we mostly follow the prior specification of Cogley and Sargent (2005). The period 0 mean and variance of the coefficient vector is set on the basis of OLS estimates for a training sample. The prior on the variance-covariance matrix of innovations to coefficients is set at 0.0001 times the training sample OLS variance matrix. The prior mean for the variance-covariance matrix of the vector of innovations to volatility is set to $0.01 \times I$, with fixed degrees of freedom equal to 5.

For the large model, we follow some, but not all, of the specification choices of K-K. As in K-K, we specify a Minnesota-style prior on the period 0 (initial) mean and variance of the coefficient vector. The prior takes a form similar to that described above for the 13 variable model with constant parameters. However, in the TVP-KK implementation, for simplicity we abstract from sums of coefficients and cointegration priors, and we instead add, as a partial substitute, cross-variable shrinkage (Litterman (1986)-style). The prior means are set to impose unit root priors for all variables. The hyperparameters governing overall and cross-variable shrinkage are each set to 0.2. Finally, following the baseline settings of K-K, we fix the forgetting factor λ at 0.99 and the volatility weighting coefficient κ at 0.96.

2.2.2 Regime-switching VAR

Finally, we also consider a four variable RS-VAR. Since we are primarily interested in using the RS-VAR model to detect possible break points, we follow Pesaran, Pettenuzzo and Timmermann (2006) and Barnett, Mumtaz and Theodoridis (2014) and consider a RS-VAR model with nonrecurrent states:

$$y_t = B_{0,S_t} + \sum_{i=1}^p B_{i,S_t}y_{t-i} + \Omega_{H_t}^{1/2}\varepsilon_t, \quad (4)$$

where B_S and Ω_H are regime dependent autoregressive coefficients and reduced-form covariance matrices. The RS-VAR model allows for M breaks at unknown dates, where these

are modeled via the latent state variable S_t for the VAR coefficients, and H_t for the error covariance matrix. 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 $p_{ij} = p(S_t = j | S_{t-1} = i)$ and $q_{ij} = p(H_t = j | H_{t-1} = i)$. The transition probability matrices are defined as:

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

Note that in this model transitions are only allowed in a sequential manner. To move from regime 1 to regime 3, the process has to visit regime 2, and transitions to past regimes are not allowed. However, this structure is not necessarily more restrictive than a standard Markov switching-VAR model. It simply implies that any new regimes are given new labels, rather than being linked to past states explicitly. This feature offers a clear computational advantage (relative to a regime-switching VAR with unrestricted transition probabilities).

We consider three versions of the RS-VAR model: (1) the general switching model of equation (4) which allows for independent breaks in the VAR coefficients and the error variance (RS-CV); (2) a version of the RS-VAR where only the VAR coefficients can break (RS-C); and (3) a version of the RS-VAR with only breaks in the error variance (RS-V). We set the lag length to two (which suffices to yield serially uncorrelated residuals) and consider specifications with two and three regimes. We follow Barnett, Mumtaz and Theodoridis (2014) and estimate the models using a Gibbs sampling algorithm, where we employ a normal-inverse Wishart prior on the VAR parameters in each regime. In the implementation, we adopt the prior specification and algorithm of Barnett, Mumtaz and Theodoridis (2014); we refer the reader to their paper and appendix for details. Under their specification, the prior on the VAR coefficients takes a Minnesota-type form similar to the one we use with our constant parameter VAR; the prior tightness is set to render the prior distributions non-informative.

2.3 Conditional forecast methodology

As section 4 will explain in more detail, one of the tools we use to examine stability following the Great Recession is conditional forecasting. In most cases, we use a given VAR

specification to produce forecasts of model variables conditional on the actual path of GDP following the onset of the Great Recession. In our out-of-sample forecast analysis, we also use conditional forecasts during the recent period under which the federal funds rate has been constrained by the effective lower bound on nominal rates (from 2009:Q1 through the end of our sample). Specifically, we condition the forecasts of other variables on a path of the federal funds rate that holds the rate fixed at 15 basis points.

In both cases, to produce these conditional forecasts, we use the minimum-MSE approach that is standard in VAR forecasting. This standard is based on the textbook problem of conditional projection, as can be handled with a state space formulation of the VAR and the Kalman filter and smoother (see, e.g., Banbura, Giannone, and Reichlin (2015)). The conditions on the variables of interest are contained in the measurement vector and equation; the data vector of the VAR is the state vector of the transition equation. The projection problem is one of predicting the state vector given the measurements (conditions). Doan, Litterman, and Sims (1984) developed an alternative algorithm, which consists of solving a least squares problem to pick the shocks needed to satisfy the conditions. Conditioning the forecasts on the path of actual GDP can be seen as consisting of the following: determining the set of shocks to the VAR that, by a least squares metric, best meets the conditions on GDP. Under the minimum-MSE approach, the conditional forecasts are not dependent on the identification of structural shocks in the VAR.

In our implementation, as is common, we form the posterior distribution of VAR parameters without taking the conditions to be imposed into account.⁵ For each model, we use Monte Carlo simulations to obtain 5000 draws of the BVAR coefficients and the error variance matrix from the standard posterior. For each draw, we use the Kalman filter to compute the conditional forecasts. In the case of models with time-varying parameters and stochastic volatility, to simplify calculations of conditional forecasts we hold the parameters and volatilities constant (over the forecast horizon) at their end-of-sample values.

3 Data

In our formal assessment of break probabilities, regime switches, and conditional forecasts, we use quarterly data for 1959:Q1-2015:Q2 obtained in early October 2015 from the Federal

⁵Waggoner and Zha (1998) develop a Gibbs sampling algorithm that provides the exact finite-sample distribution of the conditional forecasts. With our baseline model, we have verified that using the Waggoner and Zha (1998) approach to accounting for conditions in the model estimation yields extremely similar results.

Reserve Board’s FAME database.

In the analysis of real-time out-of-sample forecasts, we use real time data vintages through 2015:Q3, obtained from the Federal Reserve Bank of Philadelphia’s Real-Time Dataset for Macroeconomists (RTDSM), described in Croushore and Stark (2001). In the out-of-sample forecast analysis, real-time data are used for GDP, core PCE prices, consumption, business fixed investment, residential investment, industrial production, capacity utilization in manufacturing, non-farm payroll employment, and headline PCE inflation. For the other variables (unemployment, the federal funds rate, the credit spread, and stock prices), for which data are either not revised or only slightly revised, we rely on just currently available time series. In constructing forecasts at each point in time, we use only the data (for the model variables in use) that would have been available at the time the models would have been estimated and forecasts would have been constructed. Finally, to measure the forecast accuracy of the different models, we follow Romer and Romer (2000), among many others, and use the 2nd available (in the RTDSM) estimate as actuals.

4 Evidence of instability

This section first reports a formal break point analysis, based on a specification allowing a single break and a model allowing multiple change-points. In light of extant evidence of instabilities in the early 1980s — such as the VAR coefficient breaks documented in Strahan and van Dijk (2013) and widespread evidence of the Great Moderation in volatility — we focus on data samples starting in 1985. Cheng, Liao, and Schorfheide (2014) adopt the same sample starting point for their analysis of factor model stability.⁶

4.1 Break analysis in constant parameter VARs

We start by formally assessing the possibility of a coefficient break in our baseline constant parameter model (a BVAR using GDP growth, unemployment, core PCE inflation, and the federal funds rate), using Bayesian methods to compute the probability of a shift in all VAR coefficients occurring at specific dates.⁷ We include two lags in the model; using four lags yields the same baseline results. To ensure that our results are not driven by some large

⁶In results reported in the supplementary appendix, applying the discrete break analysis we detail below to the period of 1961-2007, we find strong evidence of a break in about 1980. We also include in the supplementary appendix results for the regime-switching BVAR with three regimes estimated for 1961-2015-Q2; these results also consistently point to a coefficient break in 2008.

⁷In a study written concurrently with this one, Francis, Jackson, and Owyang (2014) use Bayesian methods to assess stability in VARs associated with monetary policy.

shocks, we also consider a constant parameter VAR model using the fat tails formulation of Jacquier, Polson, and Rossi (2004). We assume fat tails in the form of t-distributed errors with either 5 or 10 degrees of freedom. The supplementary appendix provides more detail on the model and Gibbs sampler used for estimation.

With the break date treated as unknown, we consider a single break at a range of possible dates. Letting T_B denote the date of the possible break, we specify a VAR with coefficients having one value from observation 1 through $T_B - 1$ and potentially a different value from T_B through the end of the sample. This model includes as regressors the usual intercept and lags of endogenous regressors as well as terms interacting a dummy variable (with value 1 from T_B through the end of the sample and 0 otherwise) with the intercept and lags of the endogenous variables. In this implementation, the coefficients on the dummy-interacted terms represent changes in VAR coefficients occurring at the specified break date.

In our Bayesian framework, each break date constitutes a different model. With N_B break dates allowed, the total number of models is $N_B + 1$, with the additional model one with no break. Our prior treats each model (each break date and the no-break case) as equally likely. We compute break probabilities from marginal likelihoods obtained from the one model without a break and the N_B models that each incorporate a break at single, specific date. Letting ML_i denote the marginal likelihood for model i , its posterior probability is $ML_i / \sum_{i=1}^{N_B+1} ML_i$. The marginal likelihoods are computed with the analytical solution available for the Normal-inverted Wishart prior and posterior and, for the model with fat tails, the Gibbs sampler-based approach of Chib (1995).

The prior on the pre-break coefficients takes the usual Minnesota form, with prior means of 0 on all coefficients, an overall shrinkage hyperparameter λ , and tighter shrinkage on longer lags than shorter lags. The prior on post-break coefficient changes takes a similar form, with prior means of 0 on all coefficient changes and Minnesota-type shrinkage controlled by a hyperparameter λ_B that applies to all coefficient changes. Because we don't have much data for estimating a break that could have begun towards the end of the sample, this prior on the break is necessarily informative. Note, however, that the prior means of zero on the post-break coefficient changes imply our prior is a no-break prior, not a prior that supposes a break.

In using this framework to assess break probabilities for the baseline VAR, we consider two different prior specifications. In the first (base prior), we fix the hyperparameters at

0.2 for overall shrinkage and 0.1 for the break shrinkage. In the second (alternative prior), we search over a grid of values for the overall shrinkage hyperparameter λ and the break shrinkage hyperparameter λ_B to maximize the marginal likelihood. However, for the model with fat tails, the shrinkage is fixed in the second case as well as the first, due to the computational burden of computing the marginal likelihood by MCMC; in the second case, the prior hyperparameters are set at values similar to the typical optimized settings for the constant parameter model.

With a data sample of 1985:Q1-2015:Q2, we consider the probabilities of a single break in the VAR's coefficients at dates of 2000:Q1 through 2002:Q4 (dates chosen to be around the 2001 recession) and 2005:Q1 through 2009:Q4. We don't consider all possible dates simply to limit the computational burden of MCMC estimation for the model with fat tails. However, for the baseline model, we obtain the same results when we consider all possible dates between 1990 and 2009.) In addition, we only consider coefficient changes that all occur at the same time. In the interest of brevity, in Table 1 we omit most of the dates, for which the posterior probabilities are zero to two decimal places, and we report the probabilities for breaks in 2007 or 2008 and for no break. The supplementary appendix provides complete results, including the underlying marginal likelihoods.

These results provide strong evidence against stability, with a break most likely in about 2008. The posterior probability of the no-break model is essentially zero in all cases (models and priors) considered. For the baseline model with constant variances, the posterior probability of a break in 2008:Q2 or 2008:Q3 is 97 percent under the baseline, fixed prior and 100 percent under the alternative, optimized prior. For the model with fat tails, the evidence is quite similar, except that in one specification, some of the probability mass on a break in 2008 is moved to 2007. For the fat tails specification with 5 degrees of freedom, the posterior probability of a break is 82 percent for 2008 and 17 percent for 2007 under the baseline prior and 100 percent for 2008 under the alternative prior.⁸ Results for the fat tails specification with 10 degrees of freedom are qualitatively similar.

To complement this analysis of a single break, Table 2 reports estimated break dates from three different regime-switching VARs, described in section 2.2.2. We estimate the models over the sample 1985:Q1-2015:Q2, allowing for 2 regimes.⁹ If the inferred probability of

⁸As reported in the supplementary appendix, the estimated latent volatility states of the fat tails model with a break in 2008 show unusually large increases in the 2007-2009 period. The data indicate a significant break in the VAR's coefficients despite large shocks captured by the latent states of the fat tails specification.

⁹The results are based on 55,000 iterations with a burn-in of 50,000 draws. From the remaining 5000

being in a particular state exceeds 0.5, we take that as evidence of a shift in regime.¹⁰ In these estimates, both the RS-C and RS-CV regime-switching specifications suggest a break in coefficients in 2008:Q4. For the error covariance, RS-V finds a break in 2009:Q3. The more flexible specification that allows for independent breaks in the VAR coefficients and the error variance, RS-CV, finds a break in the error covariance in 2013:Q1, in addition to the coefficient break in 2008. Overall, the evidence from the regime switching models points to a coefficient break in 2008; with both coefficient and variance breaks allowed, the RS-VAR puts a variance break in 2013.

Overall, this break analysis shows significant evidence of VAR model instabilities around the time of the Great Recession. More precisely, the analysis points towards a break in VAR coefficients sometime around 2008 as well as a possible break in the error covariance in 2013.

4.2 Conditional forecasts

Following some other recent studies — Giannone, Lenza, Pill, and Reichlin (2012), Giannone, Lenza, and Reichlin (2012), and Stock and Watson (2012) — we consider conditional forecasts as an additional check of model stability.¹¹ Conditional forecasts are projections of a set of variables of interest on future paths of some other variables. The prior knowledge, albeit imperfect, of the future evolution of some economic variables may carry information for the outlook of other variables. Significant differences between expected and observed developments may signal that either historically (highly) unusual shocks have occurred or the relationships among variables have changed during the crisis.¹²

We study the performance of conditional forecasts from various BVARs in the period 2008:Q3-2015:Q2 — in line with the outcome of the formal break analysis — with models estimated using data samples ending with 2008:Q2.¹³ (Of course, the exact break date is

draws, we use every 5th draw for inference.

¹⁰As reported in the supplementary appendix, the inferred probabilities are usually high, and therefore provide strong evidence about the likelihood of being in a particular state. So for instance using a threshold of 0.8 instead of 0.5 yields similar results. Moreover, the transition into a state is usually very brief.

¹¹For example, Giannone, Lenza, Pill, and Reichlin (2012) study the interaction between money, credit and the business cycle, in normal times and during the financial crisis for the euro area. They compare the realized path of the variables of interest with forecasts that are conditional on the actual path of the variables capturing economic activity in the model. Giannone, Lenza and Reichlin (2012) conduct a similar type of exercise, comparing conditional forecasts for Eurosystem intermediation (conditioning on the actual path of economic activity variables) with the observed series.

¹²The supplementary appendix presents simple analytical results based on a bivariate VAR showing that conditioning on variables can tighten the historical confidence bands relative to unconditional forecast bands, making breaks easier to see. It also contains the results of a Monte Carlo simulation study on the size and power of the conditional forecasts-based procedure to detect coefficient instabilities.

¹³As shown in results in the supplementary appendix, using model estimates for the 1961-2008:Q2 sample

not known with certainty; it could be a little earlier or a little later. Using a date slightly earlier or later than 2008:Q2 would not materially affect the conditional forecast results.) In particular, we compare the actual evolution of unemployment, inflation, and the funds rate with forecasts conditional on the path of actual GDP. In this exercise, we mean to treat GDP as the business cycle factor, and we view the forecasts of the other variables as paths implied by the business cycle factor and the model's parameters. That is, by conditioning on real GDP we make sure that we capture the main business cycle influence of the period since mid-2008. If the actual paths of these variables lay materially outside conditional forecast bands, the evidence will be taken as suggesting some change in model dynamics over the 2008-2015 period.

An alternative interpretation is that historically (highly) unusual shocks caused departures from normal business cycle patterns. To be robust to such an interpretation, we study conditional forecasts from various BVAR models. This kind of exercise with VARs is similar to the Stock and Watson (2012) exercise based on a FAVAR. We will report here results for a range of specifications; the appendix describes robustness in still more specifications. In light of questions about the roles of large shocks versus coefficient instabilities, one specification of particular interest allows fat tails in the BVAR's errors. Results under this specification are very similar to those we report below for the baseline case. In addition, results for a 13-variable BVAR in levels are very similar to the baseline, small model case.

Figure 1 provides conditional forecasts obtained with a model estimated using 1985:Q1-2008:Q2 data. For this model specification, the actual paths for inflation and the funds rate, at most horizons, lay within but in the lower part of the conditional forecast confidence intervals. In contrast, and what will turn out to be most significant for our purposes, the actual unemployment rate differs greatly from the path forecast conditional on GDP growth throughout the period. At first, the actual unemployment rate rises far more than the model projects given GDP. Later, the actual unemployment rate declines far faster than the model projects.

To provide a further check of the consistency of labor market outcomes with standard VARs, we also consider a version of the model augmented to include growth in payroll employment. As shown in Figure 2, over the course of the recovery, the path of employment growth is generally consistent with the model and the path of GDP growth. But it re-yields even stronger evidence of instabilities in conditional forecasts.

mains the case that the actual path of unemployment is far outside the conditional forecast bands.¹⁴

Since the most recent recession is widely known to have involved financial stress of historic proportions, one might wonder if some of the difficulty of the baseline model in capturing the evolution of unemployment given GDP growth could be due to financial developments. Christiano, Eichenbaum, and Trabandt (2015) construct a structural model and argue that the bulk of movements in aggregate real economic activity during the Great Recession was due to financial frictions interacting with the effective lower bound on interest rates. Accordingly, we also consider a BVAR augmented to include the spread between the BAA corporate bond rate and the 10-year Treasury bond, and we construct forecasts of unemployment, inflation, and the funds rate conditioned on the actual paths of GDP growth, the funds rate, and the spread (just conditioning on GDP growth and the spread yields qualitatively similar results). The results are reported in Figure 3. Conditioning on the funds rate and credit spread in addition to GDP growth doesn't change the baseline picture much: the evolution of unemployment still remains far outside the conditional forecast bands.

To provide a further check on coefficient change, we consider another conditional forecasting exercise, using the full-sample version of the baseline (4-lag) model that permits a break in 2008:Q2, consistent with the formal break analysis above. The break is estimated using the same approach described above, with the baseline, fixed prior settings. By full-sample, we mean that we estimate the model with the sample including data up through the second quarter of 2015 and construct pseudo-forecasts conditional on the actual path of GDP. These results are shown in Figure 4. Extending the estimation sample though 2015:Q2 and allowing the break in 2008:Q2 improves the consistency of the actual paths of unemployment, inflation, and the funds rate with forecasts conditional on actual GDP growth (improves compared to the case in which model parameters are constant and the estimation sample ends in 2008:Q2). Based on the 1985:Q1-2015:Q2 model estimates, the actual paths lay within the conditional forecast bands. The contrast with the results

¹⁴To check whether the recent disconnect between GDP growth and unemployment has a precedent in other recent recessions and recoveries, we use the version of the model augmented to include employment growth to produce forecasts conditioned on actual GDP growth following the previous business cycle peaks of 1990:Q3 and 2000:Q1. We estimate models with data samples ending at each of these points in time. In both cases, we find a substantial over-prediction of the interest rate, but the forecasts of the unemployment rate were rather accurate, with some mild under-prediction at shorter horizons but over-prediction at longer horizons. In the interest of brevity, we report the detailed results in the supplementary appendix.

for models estimated with data through 2008:Q2 suggests a material change in coefficients between samples ending in mid-2008 versus mid-2015 and, in turn, a coefficient break in mid-2008. If instead the path of the unemployment rate were driven by extremely large shocks, the actual path of unemployment would remain outside the conditional forecast confidence bands from the model featuring a break in coefficients estimated with data through mid-2015, in contrast to what we find.

To summarize, the results from conditional forecasts are consistent with breaks in VAR dynamics in about 2008. There is clearly a shift in the GDP–unemployment relationship. While such a breakdown contrasts recent findings by Ball, et al. (2013) and Daly, et al. (2014) of a stable Okun’s law relationship, it supports findings of a shift in labor market dynamics, documented by Kroft, et al. (2016) and Diamond and Sahin (2015). These authors point out that the labor market during the Great Recession and subsequent recovery has been characterized by a dramatic rise in long-term unemployment and unemployment duration, marked declines in the labor force participation rate and employment-to-population ratio, as well as shifts in the Beveridge curve that reflect both the effects of changing demographics and the business cycle. While our finding of a break in the GDP–unemployment relationship is strongly suggestive of a shift in labor market dynamics, shifts in the behavior of monetary policy associated with the lower bound on interest rates and the Great Recession may have contributed in ways that are difficult to disentangle in reduced form VARs.

Our finding of a break in model dynamics following the Great Recession is different from the finding in Stock and Watson (2012), but in line with results in Cheng, Liao, and Schorfheide (2014). Stock and Watson (2012) find that the 2007-2009 recession was the result of one or more large shocks with no evidence of changes in the response of macroeconomic variables. On the contrary, results in Cheng, Liao, and Schorfheide (2014) indicate that the factor loadings changed drastically during the Great Recession. The difference in the results in these two studies can be ascribed to differences in normalization. Stock and Watson (2012) normalize the size of the loadings rather than the variance of the factors, as in Cheng, Liao, and Schorfheide (2014). The change in loadings in Cheng, Liao, and Schorfheide (2014) therefore mirrors the increase in factor volatility in the Stock and Watson (2012) analysis. An advantage with our study using VARs is that we do not need to rely on such normalization restrictions.

5 Comparing methods for managing instabilities in out-of-sample forecasting

So far we have provided evidence of instabilities in common VAR models sometime around 2008; previous studies such as Strahan and van Dijk (2013) have also provided evidence of a break in the early 1980s. These instabilities likely pose considerable challenges to forecast accuracy. There are a range of methods one might use to forecast in the face of potential instabilities. In this section we use the period following the Great Recession to consider some of the leading possible approaches, drawing in part on what has been shown to work in previous studies of forecast samples that ended before the most recent crisis and recovery.

Specifically, we consider VARs that allow for time varying parameters and volatility (TVP-SV and TVP-KK VARs, estimated with a sample starting with 1961:Q1) and VARs estimated with different samples: recursive, starting in 1985; rolling, 20 year window; and a Pesaran and Pick (2011) type average of forecasts computed over a range of rolling window sizes of 8, 10, 12, 14, \dots , 30 years.¹⁵ The specification with constant parameters and a recursive estimation sample starting in 1985:Q1 is consistent with the aforementioned evidence of a break in the early 1980s. In the rolling window case, our use of a 20-year rolling window follows such studies as Del Negro and Schorfheide (2004) and Gurkaynak, et al. (2013). Whereas other studies of VAR forecasting have used shorter windows of 10 or 15 years, our Pesaran and Pick (2011) average forecast covers such shorter estimation windows, as well as longer windows. Finally, we focus on TVP models over regime-switching specifications in light of the Pettenuzzo and Timmermann (2015) conclusion that the former generally perform better than the latter in forecasting.

In light of the advantages in short-term prediction that survey forecasts tend to have around business cycle turning points (see, e.g., the discussion in Carriero, Clark, and Marcellino 2015b), we also consider hybrid forecasts that use forecasts from the Survey of Professional Forecasters as jumping-off points for model-based forecasts. Faust and Wright (2009, 2013) have found that using survey forecasts as jumping-off points can substantially improve purely model-based forecasts.

¹⁵Reflecting these specifications, the different forecasts are produced using different start dates for model estimation. At a given forecast origin $t - 1$ for forecasting from t through $t + 3$, we use the following estimation samples. (1) Recursive, 1985 start: 1985:Q1 through $t - 1$. (2) Rolling, 80 obs. window: period $t - 80$ through $t - 1$. (3) Avg. rolling window: for each different rolling window of $R = 32, 40, 48, \dots, 120$ observations, we use a (correspondingly different) estimation period of $t - R$ through $t - 1$. (4) TVP-SV and TVP-KK: 1961:Q1 through $t - 1$.

We compare the efficacy of these approaches on the basis of the accuracy of both point and density forecasts. As noted above, we obtain the forecasts by simulating from the appropriate posterior distributions.

5.1 Empirical exercise and forecast metrics

We perform a real-time out-of-sample forecasting exercise for GDP growth, core PCE inflation, the unemployment rate, and the funds rate.¹⁶ We focus on the recovery period of 2009:Q3-2015:Q1 that followed the Great Recession.¹⁷ Even though the evidence above points to a coefficient break in 2008, we deliberately omit the (NBER-dated) recession observations of 2008:Q1-2009:Q2 by starting the forecast evaluation with 2009:Q3, so that the recession observations do not unduly influence the results of the exercise.¹⁸

We first consider the accuracy of point forecasts (defined as posterior medians), using RMSEs. We then consider the accuracy of density forecasts, using the CRPS, suggested by Gneiting and Raftery (2007). The CRPS, defined such that a lower number is a better score, is given by

$$CRPS_t(y_{t+h}^o) = \int_{-\infty}^{\infty} (F(z) - 1\{y_{t+h}^o \leq z\})^2 dz = E_f |Y_{t+h} - y_{t+h}^o| - 0.5 E_f |Y_{t+h} - Y'_{t+h}|, \quad (5)$$

where F denotes the cumulative distribution function associated with the predictive density f , y_{t+h}^o the realization of the forecasted variable, $1\{y_{t+h}^o \leq z\}$ an indicator function taking value 1 if $y_{t+h}^o \leq z$ and 0 otherwise, and Y_{t+h} and Y'_{t+h} are independent random draws from the posterior predictive density.

The forecast horizons are 1 quarter and 1 year. At the 1 year horizon, the growth and inflation forecasts are aggregated to be 4-quarter averages (quarter on 4-quarter ago growth rates), in keeping with the way things are commonly reported by the Federal Reserve and other central banks. The forecasts are conditional on a federal funds rate of 0.15 percentage point each quarter. In the absence of conditioning, the models would sometimes predict very negative funds rates in the 2009-10 period. By early 2009, verbal forward guidance from the Federal Reserve had led financial markets to expect the federal funds rate to remain near zero for at least a year, as evidenced in (early 2009 and subsequent) projections of

¹⁶Our study differs from some recent analyses of performance in the face on instabilities, such as Pettenuzzo and Timmermann (2015), in that we use real-time and not final revised data.

¹⁷The supplementary appendix provides results for a sample starting in 1996.

¹⁸However, some of the recession period is deliberately included in the conditional forecast analysis presented above, which uses conditions on actual GDP growth to capture the main business cycle influences of the period.

short-term interest rates from Blue Chip and the Survey of Professional Forecasters.

We take the benchmark forecast to be the constant parameter model estimated recursively with data starting in 1985. To facilitate the reading of results from tables, we present the RMSEs, and the CRPS, for this benchmark model and results for all other models or approaches relative to measures of RMSEs and CRPS from the baseline model. We gauge the statistical significance of each method relative to the benchmark with t -tests (see, e.g., Diebold and Mariano (1995) and West (1996)) of equal MSE and equal CRPS. The tests are two-sided such that a positive rejection indicates the alternative forecast is more accurate than the recursive benchmark and negative indicates the reverse. Motivated by simulation evidence in Clark and McCracken (2013), we employ a rectangular kernel truncated at lag $h - 1$ in the variance estimator which enters the test statistic, with the finite sample correction due to Harvey, Leybourne, and Newbold (1997), and compare the test statistic to standard normal critical values.

5.2 Results with a four variable BVAR model

We start by focusing on the results for point forecasts, the RMSEs, provided in Table 3. For forecasting GDP growth, the baseline model (the specification of which allows a break in the early 1980s but does not model a break around 2008) is not easily beaten. Only the TVP-SV model improves on the baseline forecast at both forecast horizons. However, the gains are not statistically significant, perhaps due to power limitations associated with the short forecast sample. The 20 year rolling window forecast is better than the baseline at the shorter horizon and worse at the longer horizon, again not by amounts that are statistically significant. The average rolling window forecast is consistently worse than the baseline. For the unemployment rate, the baseline forecast is even more difficult to beat; it is in fact the most accurate point forecast, although, again, not by margins that generally appear statistically significant. Finally, for core inflation, results differ by horizon: at the 1-step horizon, the baseline is best; at the 4-step horizon, the TVP-SV model performs best, although not by a margin (relative to the baseline) that appears statistically significant.

The results in Table 3 for CRPS-measured density accuracy are quite similar to the RMSE results. Broadly, the baseline specification of the constant parameter model estimated recursively with data starting in 1985 is difficult to beat. Among the approaches that accommodate structural change, the TVP-SV model beats the baseline most often, although not uniformly and not by margins that are statistically significant.

5.3 Results with a medium size BVAR model

Some recent research has found that larger BVARs tend to forecast more accurately than smaller BVARs and that the forecasting performance of large and medium sized BVARs is comparable to that of factor models (e.g., Banbura, Giannone, and Reichlin (2010) and Koop (2013)). In this section, we therefore report the results obtained with the 13-variable BVAR detailed in section 2. In presenting the results, we still focus on the forecasting performance for GDP growth, the unemployment rate and core PCE inflation. As explained in section 2, in the time-varying parameter and volatility version of the model, in light of computational constraints we use the specification of Koop and Korobilis (2013). Results are reported in Table 4.

In the case of the models with 13 variables, forecast performance varies some by the forecast horizon. At the 1-step horizon, the differences in forecast accuracy are fairly modest, with some exceptions for inflation. For growth and unemployment, the RMSE and CRPS ratios of the different methods are generally close to 1, sometimes a little above and other times a little below. At the 4-step horizon, the baseline is sometimes beaten by larger margins. The average rolling window approach consistently beats the baseline, offering RMSE and CRPS gains of more than 10 percent for growth and inflation (although only statistically significant for inflation). The TVP-KK specification performs best for GDP growth, with RMSE gains of more than 20 percent, but fares relatively poorly for unemployment.

5.4 Hybrid survey-model forecasts

Survey-based forecasts can have a number of advantages over common model-based methods; perhaps most importantly, the survey-based forecasts often have access to information more timely than that used in constructing model forecasts (see e.g. Giannone, Reichlin and Small (2008) and Aastveit, et al. (2014) for the importance of using timely information for nowcasting) In our context, at each forecast origin t , which in RTDSM timing corresponds to roughly the middle of the quarter, we use quarterly information through $t - 1$ to estimate each VAR model and forecast. The corresponding Survey of Professional Forecasters is the one published in about the middle of the quarter. At the time the survey is conducted, following the release of the initial estimate of GDP for the previous quarter, respondents have available interest rates and other financial indicators for the first month of the quarter, as well as (normally) readings on some important indicators of economic

activity, including employment and unemployment and the purchasing managers index for manufacturing. This more timely information likely gives the survey forecast an important advantage over model forecasts constructed as they are in this paper (and in much of the forecasting literature).

Accordingly, in this section we consider the accuracy of forecasts we characterize as hybrids of model and survey-based forecasts. Following Faust and Wright (2009, 2013), we construct hybrid forecasts by using the survey-based forecasts for the 1-quarter ahead horizon as jumping off points for model-based forecasts for horizons of 2-4 quarters. Formally, we use the survey forecasts for 1 quarter ahead as conditions on the VAR forecasts. Given the timing underlying our forecast analysis, this means we are giving the models the current-quarter forecast obtained from a survey. This approach yields VAR forecasts that are the same as the SPF forecasts for the 1 quarter horizon but determined by VAR dynamics and the 1-quarter ahead conditions at subsequent horizons. This approach serves to adjust for the timing advantage of the survey over the models and for — to some degree — some of the wider information set and judgment underlying the survey. We apply this survey-forecast conditioning to the same set of forecast specifications covered in Table 3.

Results for point forecasts from our baseline model and the hybrid specifications are provided in Table 5. At the 1-quarter horizon, conditioning the model forecasts on the survey-based forecasts yields — by construction — the same (survey-based) forecasts and forecast accuracy, which are all substantially better than the baseline forecast that does not include the SPF conditions. At the 4-quarter horizon, it continues to be the case that the conditioning on the SPF 1-step ahead forecasts improves the accuracy of the models' forecasts. In broad terms, the methods perform comparably. The hybrid TVP-SV forecast is best for GDP growth and inflation, although not all that different from the other forecasts.

5.5 Summary of out-of-sample forecast results

Putting all of this together, it is hard to say that a single approach works best in out-of-sample forecasting. This may not be surprising given the short evaluation period available following the circa-2008 break in VAR dynamics identified in other sections of our analysis. The baseline of a constant parameter model is difficult to beat in out-of-sample forecasting. Among the methods that accommodate structural change, allowing time-varying parameters (TVP-SV) seems to perform best with a small model, whereas with the large model, the average rolling window approach offers the most consistent gains on the baseline.

6 Conclusions

Building on previous work on the stability of factor models since the Great Recession and on the stability of VAR models in forecasting in data preceding the Great Recession, in this paper we examine the stability of common VARs in the period since the Great Recession.

We first use a variety of approaches to assess model stability around the Great Recession, all of which suggest significant instabilities. Bayesian analysis of a model with a discrete break and a regime-switching model provides additional evidence against VAR parameter stability, with a break most likely around 2008. Using conditional forecasts, we show that VAR specifications assumed stable through mid-2008 produce large forecast errors during and after the crisis, even when conditioning on the actual evolution of GDP growth.

We then examine the efficacy of a range of forecasting methods that can be used to deal with structural change. Specifically, in addition to models that allow for time varying parameters and volatility, we consider forecasts from VARs estimated with different samples: recursive, starting in 1985; rolling, 20 year window; and a Pesaran and Pick (2011) type average of forecasts computed over a range of rolling window sizes of 8, 10, 12, 14, \dots , 30 years. Overall, the baseline of a constant parameter model is difficult to beat in out-of-sample forecasting. This may not be surprising given the short evaluation period available following the circa-2008 break in VAR dynamics we have found. Among the methods that accommodate structural change, allowing time-varying parameters (TVP-SV) seems to perform best with a small model, whereas with the large model, the average rolling window approach offers the most consistent gains over the baseline.

As we noted above, while our reduced form evidence points most clearly to instability in the relationship between GDP growth and the unemployment rate, in a more structural sense it is difficult to disentangle instabilities that could truly be due to either shifts in labor market dynamics or the behavior of monetary policy associated with the effective lower bound on interest rates. Over time, further research on possible structural changes associated with the Great Recession and subsequent recovery may point to model specifications that yield further gains in forecast accuracy. For example, there is considerable structural work on the macroeconomics of labor markets that examines whether the most recent recession was fundamentally different from previous recessions, with conflicting findings (see, e.g., Gali, Smets, and Wouters (2012) and Ravenna and Walsh (2013)). There is also a growing body of structural work on modeling monetary policy since the Great Recession,

to capture the effects of the effective lower bound, extended forward guidance from central banks, and government bond purchases (see, e.g., Chen, Curdia, and Ferraro (2012)).

7 References

- Aastveit K, Gerdrup K, Jore A, Thorsrud L. 2014. Nowcasting GDP in real time: a density combination approach. *Journal of Business and Economic Statistics* **32**: 48-68. DOI: 10.1080/07350015.2013.844155
- Alessi L, Ghysels E, Onorante L, Peach R, Potter S. 2014. Central bank macroeconomic forecasting during the global financial crisis. *Journal of Business and Economic Statistics* **32**: 483-500. DOI: 10.1080/07350015.2014.959124
- Ball L, Leigh D, Loungani P. 2013. Okun's Law: fit at fifty? NBER Working Paper 18668.
- Banbura M, Giannone D, Reichlin L. 2010. Large Bayesian vector autoregressions. *Journal of Applied Econometrics* **25**: 71-92. DOI: 10.1002/jae.1137
- Banbura M, Giannone D, Reichlin L. 2015. Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting* **31**: 739-756. DOI: 10.1016/j.ijforecast.2014.08.013
- Barnett A, Mumtaz H, Theodoridis K. 2014. Forecasting UK GDP growth and inflation under structural change: a comparison of models with time-varying parameters. *International Journal of Forecasting* **30**: 129-143. DOI: 10.1016/j.ijforecast.2013.06.002
- Carriero A, Clark T, Marcellino M. 2015a. Bayesian VARs: specification choices and forecasting performance. *Journal of Applied Econometrics* **30**: 46-73. DOI: 10.1002/jae.2315
- Carriero A, Clark T, Marcellino M. 2015b. Real-time nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society, Series A* **178**: 837-862. DOI: 10.1111/rssa.12092
- Chen H, Curdia V, Ferraro A. 2012. The macroeconomic effects of large-scale asset purchase programs. *The Economic Journal* **122**: F289-F315. DOI: 10.1111/j.1468-0297.2012.02549.x
- Cheng X, Liao Z, Schorfheide F. 2014. Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies*, forthcoming.
- Chib S. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**: 1313-1321. DOI: 10.2307/2291521
- Christiano L, Eichenbaum M, Trabandt M. 2015. Understanding the Great Recession. *American Economic Journal: Macroeconomics* **7**: 110-167. DOI: 10.1257/mac.7.1.110

- Clark T. 2011. Real-time density forecasts from BVARs with stochastic volatility. *Journal of Business and Economic Statistics* **29**: 327-341. DOI: 10.1198/jbes.2010.09248
- Clark T, McCracken M. 2008. Forecasting with small macroeconomic VARs in the presence of instability. In *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, Rapach D, Wohar M (eds). Emerald Group Publishing: Bingley.
- Clark T, McCracken M. 2013. Advances in forecast evaluation. In *Handbook of Economic Forecasting, Volume 2*, Elliott G, Timmermann A (eds). North Holland: Amsterdam.
- Cogley T, Sargent T. 2005. Drifts and volatilities: monetary policies and outcomes in the Post-WWII US. *Review of Economic Dynamics* **8**: 262-302. DOI: 10.1016/j.red.2004.10.009
- Croushore D, Stark T. 2001. A real-time data set for macroeconomists. *Journal of Econometrics* **105**: 111-30. DOI: 10.1016/S0304-4076(01)00072-0
- Daly M, Hobijn B, Sahin A, Valletta R. 2012. A search and matching approach to labor markets: did the natural rate of unemployment rise? *Journal of Economic Perspectives* **26**: 3-26. DOI: 10.1257/jep.26.3.3
- Daly M, Fernald J, Jorda O, Nechio F. 2014. Interpreting deviations from Okun's Law. Federal Reserve Bank of San Francisco *Economic Letter*. 2014-12.
- D'Agostino A, Gambetti L, Giannone D. 2013. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* **28**: 82-101. DOI: 10.1002/jae.1257
- Del Negro M, Schorfheide F. 2004. Priors from general equilibrium models for VARs. *International Economic Review* **45**: 643-673. DOI: 10.1111/j.1468-2354.2004.00139.x
- Diamond P, Sahin A. 2015. Shifts in the Beveridge Curve. *Research in Economics* **69**: 12-18. DOI: 10.1016/j.rie.2014.10.004
- Diebold F, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253-263. DOI: 10.1198/073500102753410444
- Doan T, Litterman R, Sims C. 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* **3**: 1-100.
- Faust J, Wright J. 2009. Comparing Greenbook and reduced form forecasts using a large realtime dataset. *Journal of Business and Economic Statistics* **27**: 469-479. DOI: 10.1198/jbes.2009.07214
- Faust J, Wright J. 2013. Inflation forecasting. In *Handbook of Economic Forecasting, Volume 2*, Elliott G, Timmermann A (eds). North Holland: Amsterdam.
- Francis N, Jackson L, Owyang M. 2014. How has empirical monetary policy analysis

- changed after the financial crisis? Federal Reserve Bank of St. Louis Working Paper 2014-019A.
- Gali J, Smets F, Wouters R. 2012. Slow recoveries: a structural interpretation. *Journal of Money, Credit, and Banking* **44**: 9-30. DOI: 10.1111/j.1538-4616.2012.00552.x
- Giannone D, Lenza M, Reichlin L. 2012. Money, credit, monetary policy and the business cycle in the Euro Area, manuscript.
- Giannone D, Lenza M, Pill H, Reichlin L. 2012. The ECB and the interbank market. *Economic Journal* **122**: F467-F486. DOI: 10.1111/j.1468-0297.2012.02553.x
- Giannone D, Reichlin L, Small D. 2008. Nowcasting: the real-time informational content of macroeconomic data. *Journal of Monetary Economics* **55**: 656-676. DOI: 10.1111/j.1468-0297.2012.02553.x
- Gneiting T, Raftery A. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**: 359-378. DOI: 10.1198/016214506000001437
- Gurkaynak R, Kisacikoglu B, Rossi B. 2013. Do DSGE models forecast more accurately out-of-sample than reduced-form models? In *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims, Advances in Econometrics* 32, Fomby T, Kilian L, Murphy A (eds). Emerald: Bingley, UK.
- Hall R. 2010. Why does the economy fall to pieces after a financial crisis? *Journal of Economic Perspectives* **24**: 3-20. DOI: 10.1257/jep.24.4.3
- Harvey D, Leybourne S, Newbold P. 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* **13**: 281-91. DOI: 10.1016/S0169-2070(96)00719-4
- Jacquier E, Polson N, Rossi P. 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* **122**: 185-212. DOI: 10.1016/j.jeconom.2003.09.001
- Koop G. 2013. Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics* **28**: 177-203. DOI: 10.1002/jae.1270
- Koop G, Korobilis D. 2013. Large time-varying parameter VARs. *Journal of Econometrics* **177**: 185-198. DOI: 10.1016/j.jeconom.2013.04.007
- Kroft K, Lange F, Notowidigdo M, Katz L. 2016. Long-term unemployment and the Great Recession: the role of composition, duration dependence, and non-participation. *Journal of Labor Economics* **34, No. S1, Part 2**: S7-54. DOI: 10.1086/682390

- Litterman R. 1986. Forecasting with Bayesian vector autoregressions-five years of experience. *Journal of Business and Economic Statistics* **4**: 25-38. DOI: 10.1080/07350015.1986.10509491
- Ng S, Wright J. 2013. Facts and challenges from the Great Recession for forecasting and macroeconomic modeling. *Journal of Economic Literature* **51**: 1120-1154. DOI: 10.1257/jel.51.4.1120
- Pesaran H, Pettenuzzo D, Timmermann A. 2006. Forecasting time series subject to structural breaks. *Review of Economic Studies* **73**: 1057-1084. DOI: 10.1111/j.1467-937X.2006.00408.x
- Pesaran H, Pick A. 2011. Forecast combination across estimation windows. *Journal of Business and Economic Statistics* **29**: 307-318. DOI: 10.1198/jbes.2010.09018
- Pettenuzzo D, Timmermann A. 2015. Forecasting macroeconomic variables under model instability. *Journal of Business and Economic Statistics*, forthcoming.
- Primiceri G. 2005. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* **72**: 821-852. DOI: 10.1111/j.1467-937X.2005.00353.x
- Ravenna F, Walsh C. 2013. Slow recoveries, worker heterogeneity, and the zero lower bound, manuscript.
- Reinhart C, Rogoff K. 2009. *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press: Princeton, NJ.
- Romer C, Romer D. 2000. Federal Reserve information and the behavior of interest rates. *American Economic Review* **90**: 429-457. DOI: 10.1257/aer.90.3.429
- Sims C. 1993. A nine-variable probabilistic macroeconomic forecasting model. In *Business Cycles, Indicators and Forecasting*, Stock J, Watson M (eds). Chicago: University of Chicago.
- Sims C, Zha T. 1998. Bayesian methods for dynamic multivariate models. *International Economic Review* **39**: 949-968. DOI: 10.2307/2527345
- Sims C, Zha T. 2006. Were there regime switches in U.S. monetary policy? *American Economic Review* **96**: 54-81. DOI: 10.1257/000282806776157678
- Stock J, Watson M. 1996. Evidence of structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* **14**: 11-30.
- Stock J, Watson M. 2012. Disentangling the channels of the 2007-2009 recession. *Brookings Papers on Economic Activity*, Spring, 81-130.
- Strahan R, van Dijk H. 2013. Evidence on features of a DSGE business cycle model from Bayesian model averaging. *International Economic Review* **54**: 385-402. DOI:

10.1111/j.1468-2354.2012.00737.x

Waggoner D, Zha T. 1999. Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics* **81**: 639-651. DOI: 10.1162/003465399558508

West K. 1996. Asymptotic inference about predictive ability. *Econometrica* **64**: 1067-1084.
DOI: 10.2307/2171956

Table 1: **Break probabilities, BVAR(2), 1985:Q1-2015:Q2**

Break date	Base prior			Alternative prior		
	BVAR	Fat tails df=5	Fat tails df=10	BVAR	Fat tails df=5	Fat tails df=10
2007:01	0.00	0.00	0.00	0.00	0.00	0.00
2007:02	0.00	0.00	0.01	0.00	0.00	0.00
2007:03	0.01	0.13	0.05	0.00	0.00	0.00
2007:04	0.02	0.04	0.35	0.00	0.00	0.01
2008:01	0.00	0.00	0.00	0.00	0.00	0.00
2008:02	0.83	0.68	0.58	0.34	0.02	0.51
2008:03	0.14	0.14	0.01	0.66	0.98	0.48
2008:04	0.00	0.00	0.00	0.00	0.00	0.01
no break	0.00	0.01	0.00	0.00	0.00	0.00

Note: Letting ML_i denote the marginal likelihood for model (break date) i , its posterior probability is $ML_i / \sum_{i=1}^{N_B+1} ML_i$.

Table 2: **Break dates from a RS-VAR(2), 1985:Q1-2015:Q2**

Model	break coefficients	break variance
RS-C	2008:Q4	NA
RS-V	NA	2009:Q3
RS-CV	2008:Q4	2013:Q1

Note: RS-C refers to a RS-VAR with only breaks in the coefficients, RS-V refers to a RS-VAR with only breaks in the error variance, and RS-CV refers to a RS-VAR which allows for independent breaks in the VAR coefficients and the error covariance. All of the models allow for two regimes. We identify a break in the coefficients or the error covariance as the date at which the inferred probability of being in a particular state exceeds 0.5.

Table 3: Accuracy of forecasts from 4-variable BVAR, 2009:Q3-2015:Q1

model or estimation approach	GDP growth		Unemployment rate		Core PCE inflation	
	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS
<i>1-quarter horizon</i>						
recursive, 1985 start	2.105	1.175	0.252	0.146	0.510	0.289
rolling, 80 obs. window	0.965	0.967	1.075	1.080	1.141 *	1.181 **
avg. rolling window	1.017	1.021	1.058	1.052	1.053	1.070
TVP-SV	0.917	0.955	1.052	1.045	1.021	1.064
<i>4-quarter horizon</i>						
recursive, 1985 start	1.914	1.062	1.169	0.636	0.522	0.286
rolling, 80 obs. window	1.083	1.064	1.029	1.045 **	1.003	1.013
avg. rolling window	1.179	1.144	1.048 **	1.063 **	0.989	0.983
TVP-SV	0.947	0.858	1.043	1.116	0.916	0.945

Note: RMSE and CRPS levels for baseline (recursive, 1985 start) forecasts, ratios for all others. Statistical significance based on two-sided t-tests for equal MSE or equal CRPS indicated by ** for 5 percent and * for 10 percent.

Table 4: Accuracy of forecasts from 13-variable BVAR, 2009:Q3-2015:Q1

model or estimation approach	GDP growth		Unemployment rate		Core PCE inflation	
	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS
<i>1-quarter horizon</i>						
recursive, 1985 start	1.802	1.020	0.282	0.166	0.489	0.284
rolling, 80 obs. window	0.991	1.005	1.006	0.987	1.091 *	1.085
avg. rolling window	0.993	0.995	0.991	0.974	1.055	1.055
TVP-KK	1.001	1.022	1.052	1.000	1.179	1.142
<i>4-quarter horizon</i>						
recursive, 1985 start	1.860	0.949	0.923	0.519	0.647	0.356
rolling, 80 obs. window	0.957	0.993	0.998	0.979	1.077	1.046
avg. rolling window	0.883	0.881	0.985	0.985	0.859 **	0.843 **
TVP-KK	0.716	0.769 **	1.297 **	1.502 **	1.023	1.044

Note: RMSE and CRPS levels for baseline (recursive, 1985 start) forecasts, ratios for all others. Statistical significance based on two-sided t-tests for equal MSE or equal CRPS indicated by ** for 5 percent and * for 10 percent.

Table 5: **Accuracy of forecasts from 4-variable hybrid BVAR, 2009:Q3-2015:Q1**

model or estimation approach	GDP growth	Unemployment rate	Core PCE inflation
<i>1-quarter horizon</i>			
rec., 1985 start	2.105	0.252	0.510
hybrid, 1985 start	0.783 **	0.632 **	0.883
hybrid, rolling, 20y	0.783 **	0.632 **	0.883
hybrid, avg. window	0.783 **	0.632 **	0.883
hybrid, TVP-SV	0.783 **	0.632 **	0.883
<i>4-quarter horizon</i>			
rec., 1985 start	1.914	1.169	0.522
hybrid, 1985 start	0.769	0.768	0.858 *
hybrid, rolling, 20y	0.852	0.791	0.851
hybrid, avg. window	0.939	0.811	0.879
hybrid, TVP-SV	0.667 *	0.796	0.819

*Note: RMSE levels for baseline (recursive, 1985 start) VAR forecasts, ratios for all others. The other forecasts are from 4-variable VAR specifications (corresponding to those in Table 3), obtained by conditioning on SPF forecasts for the 1-quarter ahead horizon. Statistical significance based on two-sided t-tests for equal MSE or equal CRPS indicated by ** for 5 percent and * for 10 percent.*

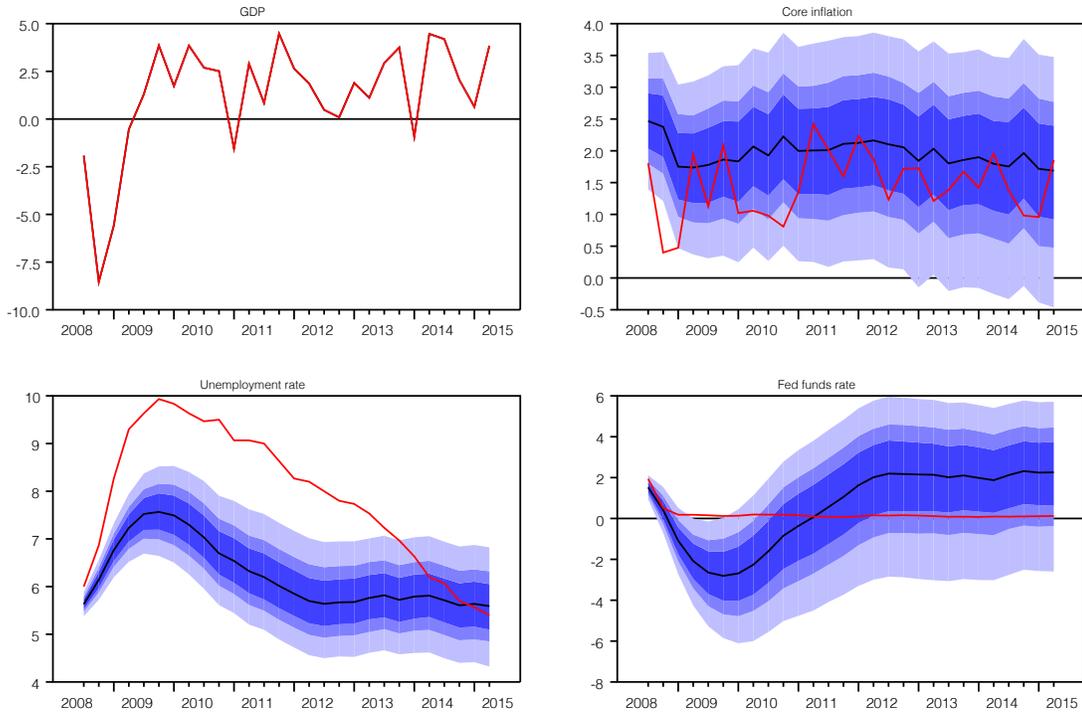


Figure 1: Conditional forecasts for 2008:Q3-2015:Q2 from a 4-variable BVAR, estimated with 1985:Q1-2008:Q2 sample. Conditioning on GDP growth.

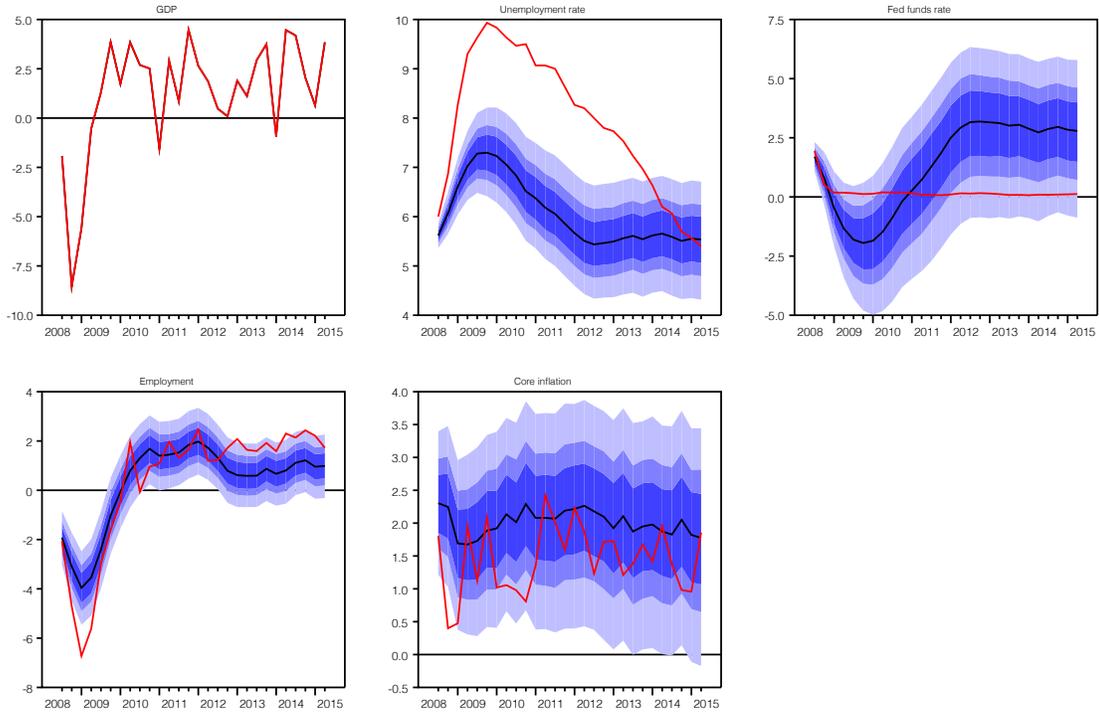


Figure 2: Conditional forecasts for 2008:Q3-2015:Q2 from a 5-variable BVAR with employment, estimated with 1985:Q1-2008:Q2 sample. Conditioning on GDP growth.

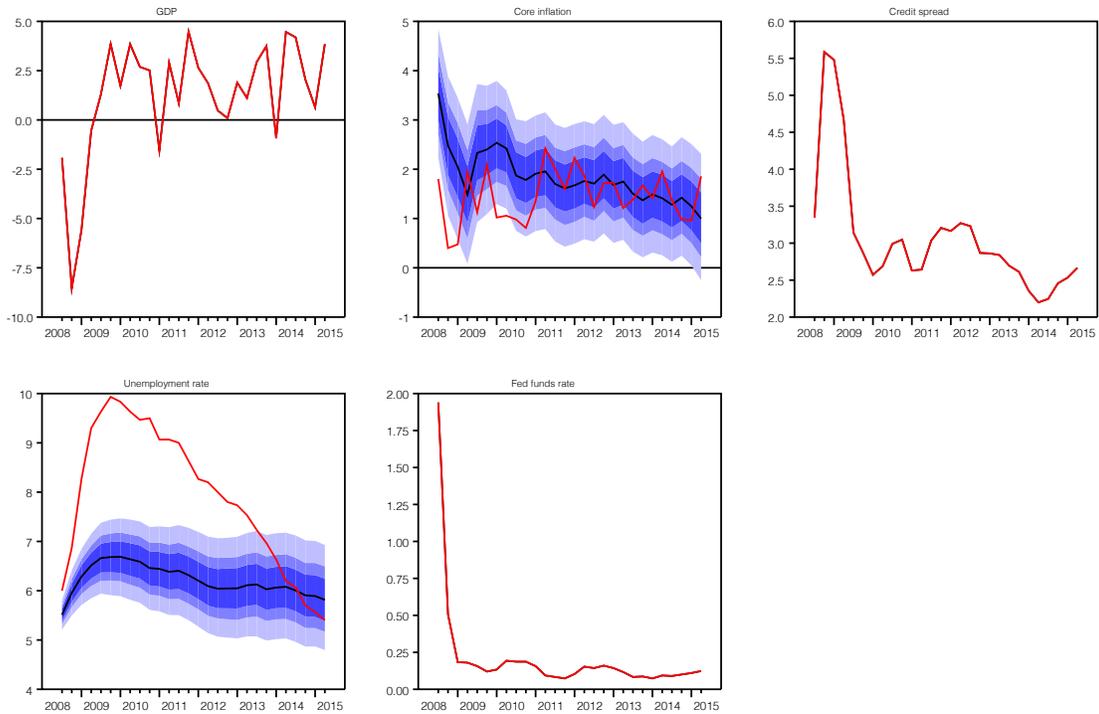


Figure 3: Conditional forecasts for 2008:Q3-2015:Q2 from a 5-variable BVAR with credit spread, estimated with 1985:Q1-2008:Q2 sample. Conditioning on GDP growth, federal funds rate, and credit spread.

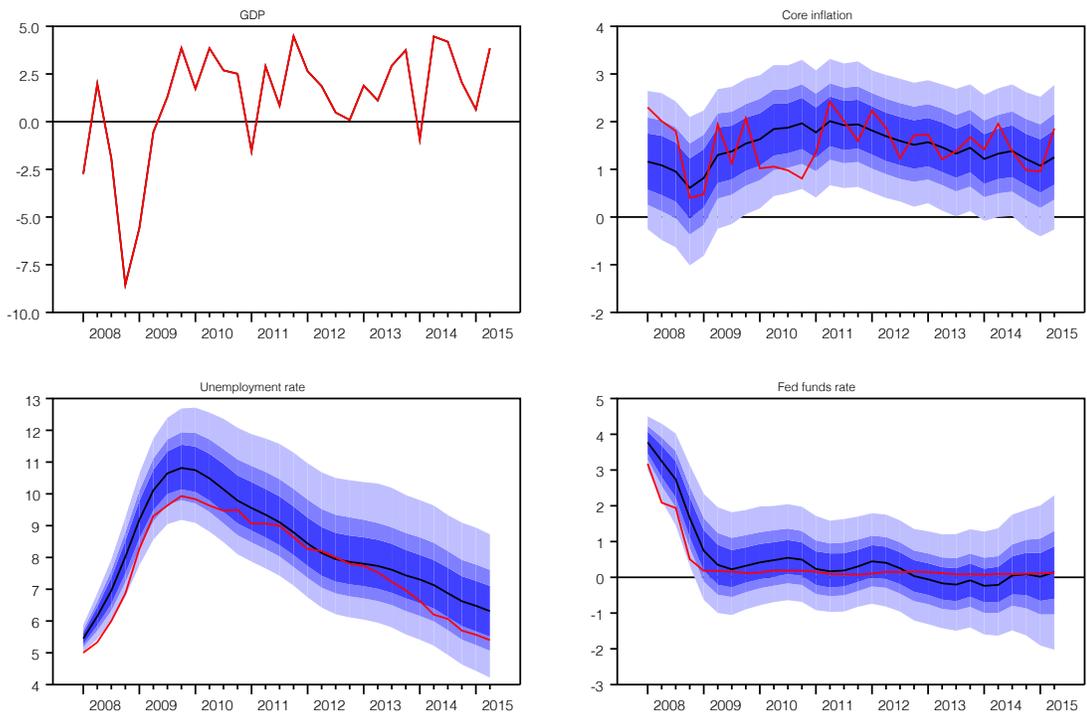


Figure 4: Conditional forecasts for 2008:Q3-2015:Q2 from a 4-variable BVAR allowing 2008:Q2 break, estimated with 1985:Q1-2015:Q2 sample. Conditioning on GDP growth.