working paper

14 11R

Have Standard VARs Remained Stable since the Crisis?

Knut Are Aastveit, Andrea Carriero,
Todd E. Clark, and Massimiliano Marcellino

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 14-11R April 2015*

Have Standard VARs Remained Stable since the Crisis?

Knut Are Aastveit, Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino

Small or medium-scale VARs are commonly used in applied macroeconomics for forecasting and evaluating the shock transmission mechanism. This requires the VAR parameters to be stable over the evaluation and forecast sample, or to explicitly consider parameter time variation. The earlier literature focused on whether there were sizable parameter changes in the early 1980s, in either the conditional mean or variance parameters, and in the subsequent period till the beginning of the new century. In this paper we conduct a similar analysis but focus on the effects of 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 then discuss and evaluate alternative methods to handle parameter instability in a forecasting context. While none of the methods clearly emerges as best, some techniques turn out to be useful to improve the forecasting performance.

Keywords: Bayesian VAR, Forecasting, Time-varying parameters, Stochastic volatility.

JEL classication code: E17, C11, C33, C53.

Suggested citation: Aastveit, Knut Are, Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino, 2015. "Have Standard VARs Remained Stable since the Crisis?" Federal Reserve Bank of Cleveland, working paper no. 14-11R.

Knut Are Aastveit is at Norges Bank (knut-are.aastveit@norges-bank.no), Andrea Carriero is at Queen Mary, University of London (a.carriero@qmul.ac.uk), Todd E. Clark is at the Federal Reserve Bank of Cleveland (todd.clark@clev.frb.org), and Massimiliano Marcellino is at Bocconi University, IGIER, and CEPR, (massimiliano.marcellino@unibocconi.it). The authors gratefully acknowledge helpful comments from 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.

*First version posted September 2014.

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.[2]

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.

Recent practical experience in macroeconomic forecasting with common VARs in U.S.

[1] Barnett, 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.

[2] Mian and Sufi (2010) and Sahin, et al. (2012) also find that recoveries after financial driven recessions are slow, due to post-crisis de-leveraging in the former and regional or industry job mismatch in the latter.

data also suggests the possibility of a break in model coefficients during the severe recession of 2007-2009 and subsequent slow recovery.[3] In particular, practical experience has highlighted some dramatic sensitivity of GDP growth and unemployment forecasts to the period over which the model is estimated. Consider, for example, forecasts of GDP growth for 2012 and 2013 (measured on a Q4/Q4 basis) made with data as of 2011:Q4, using a Bayesian VAR in GDP growth, the unemployment rate, inflation ex food and energy, and the federal funds rate.[4] Using coefficients estimated with data for 1961-2011 yields GDP growth forecasts of 4.8 percent in 2012 and 4.4 percent in 2013. However, using coefficients estimated with data for 1961-2008 yields a much sharper bounce-back in GDP growth, with forecasts of 6.8 percent in 2012 and 5.7 percent in 2013. Yet using coefficients estimated with data for 1985-2011 (a shorter sample motivated by the potential of a break in VAR dynamics associated with the early 1980s shift in monetary policy behavior) pushes the forecasts down sharply, to 3.5 and 3.8 percent. Actual GDP growth has come in well below these forecasts, at 1.6 percent in 2012 and 3.1 percent in 2013.[5]

Accordingly, in this paper we examine the stability of common VARs in the period since the sharp recession of 2007-2009. Does the evolution of macro variables for the U.S. over the 2007-2013 period represent a break in VAR dynamics or large shocks? Focusing on small and medium size VARs, we use a variety of approaches to assess the stability of VAR specifications commonly used for forecasting and policy analysis.

We first use a variety of approaches to assess model stability around the Great Recession, all of which suggest significant instabilities. Specifically, we show that VARs produce large forecast errors during and after the crisis, even when conditioning on the actual evolution of GDP growth.[6] The large forecast errors can be due to parameter changes and/or mistakes in the forecasts of the future values of the conditioning variables. By using the actual future values of the conditioning variables we are able to disentangle the two effects and show that most of the forecast errors can be attributed to parameter changes. Moreover, when parameter time variation is allowed, there is substantial variation in the time series of coefficient estimates. More formal (Bayesian) analysis provides additional evidence of VAR

[3]Parameter stability can be assessed by looking directly at the model parameters or indirectly by studying the forecasts. The latter can provide an indication of the extent and economic relevance of the parameter changes, which cannot emerge by simply looking at the single model parameters (in)stability.

[4]The model is estimated using a Normal-inverted Wishart prior, with the Normal prior on the VAR coefficients taking a Minnesota form.

[5]These growth rates are based on the vintage of GDP data available in late August 2014.

[6]Alessi, et al. (2014) examine the performance of central bank forecasts during the crisis, highlighting large errors.

parameter instability.

We then examine the efficacy of a range of forecasting methods that can be used to deal with structural change.[7] Specifically, we consider forecasts from VARs with time varying parameters and volatility (TVP-SV VARs) and VARs estimated with different samples: recursive, starting in 1961; 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, but accounting for time variation turns out to be useful to improve the forecasting performance both in terms of root mean squared forecasting errors (RMSEs) and the average continuous ranked probability score (CRPS). However, in the absence of a strong 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 a long sample may be conceptually preferable, because it does not hinge on the ability to identify a specific break date.

In the context of the broader literature, our analysis can be seen as extending previous work on (1) the stability of VAR forecasting models and efficacy of various methods for accommodating parameter change and (2) the stability of factor models since the Great Recession. We extend the first line of work by focusing on the period since the Great Recession, after presenting some evidence of a break in VAR dynamics around the Great Recession. We extend the second line of work, on factor models, by considering small and medium-scale VAR models commonly used in forecasting.

The remainder of the paper is organized as follows. Section 2 presents the models and conditional forecast methodology used in our analysis. Section 3 details the data. Section 4 provides evidence of instability after 2007 in terms of Bayesian analysis of breaks in coefficients, time variation in coefficients, and unconditional and conditional forecasts. Section 5 reports results from a real-time out-of-sample forecast comparison of alternative approaches to handle instabilities. Section 6 concludes.

[7] Giraitis, Kapetanios, and Price (2012) develop an approach for optimizing the rate of discount used in discounted least squares estimation of forecasting models, which they apply to forecasts over a sample ending in mid-2008.

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 (TVP), as in such studies as Cogley and Sargent (2005), Primiceri (2005), D'Agostino, Gambetti, and Giannone (2013), and Koop and Korobilis (2013). This section first details the models used in our analysis and then presents the approaches 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 $B_i, i = 1, \ldots, p$, 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:

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

We base most of our results on a baseline VAR(4) in GDP growth, the unemployment rate, core PCE inflation, and the federal funds rates.[8] 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 spread between the BAA corporate bond rate and the 10-year Treasury yield. We also include some results for a 13-variable BVAR(5) in levels or log levels in our out-of-sample forecasting exercise, 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.[9] 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. We describe below and in the appendix a large number of alternative model specifications we have considered to verify robustness.

We estimate the 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-

[8] We 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.

[9] The forecasting performance of large BVARs is comparable to that of factor models (e.g., Banbura, Giannone, and Reichlin 2010).

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):

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

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

For the four and five variable models (which use growth rates of trending variables), prior means are set to zero for all coefficients. At each forecast origin, 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. Using the common, fixed setting of 0.2 yields essentially the same results.

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 and Sims and Zha (1998). At each forecast origin, 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.[10]

2.2 Time-varying parameter VAR specifications

To directly accommodate structural change, we also consider both four and 13 variable VARs with TVP and time-varying volatility, using the sets of variables detailed above.[11]

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

[10]However, 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 (2013) found the forecast accuracy payoff to optimization to be fairly small.

[11]We do not consider models with Markov switching. Switching VARs are difficult to estimate, which has limited their use (see, for example discussions in Hubrich and Tetlow (2012) and Bognanni (2013)). Models with TVP are easier to estimate and generally capable of capturing sharp breaks like those of interest with switching models (see, e.g., the discussion in Koop and Potter (2007) and Baumeister and Benati (2012)).

in Primiceri (1995):

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

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

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

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 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 of the vector of innovations to volatility is set to $0.01 \times I$, with fixed degrees of freedom equal to 5 (a prior deliberately more generous than in Cogley and Sargent, in light of considerable evidence of time-varying volatility).

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.

2.3 Conditional forecast methodology

As section 4 will explain in more detail, one of the tools we use to examine the stability of models 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 subsequent out-of-sample (real-time) forecast analysis, we also use conditional forecasts during the recent period under which the federal funds rate has been constrained by the zero lower bound on nominal rates (from 2009:Q1 through 2013:Q4, the end of our sample). In this part of the analysis, we condition the forecasts of other variables on a path of the federal funds rate that holds the rate fixed at 15 basis points. We explain the rationale in more detail in section 4.

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., Clarida and Coyle (1984) or Banbura, Giannone, and Lenza (2014)). 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). We use the Kalman filter/smoother implementation of Banbura, Giannone, and Lenza (2014) to produce our conditional forecasts. Doan, Litterman, and Sims (1984) developed an alternative approach to the conditional forecasting problem, which consists of solving a least squares problem to pick the shocks needed to satisfy the conditions. In the context of conditioning on the path of actual GDP, this approach to conditional forecasting can be seen as consisting of the following: determining the set of shocks to the VAR that, by a least squares metric, best meet the conditions on GDP. Note that, 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. 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 approach of Banbura, Giannone, and Lenza (2014) to compute the conditional forecasts of interest. 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.

However, with our baseline model, we have verified that taking the conditions into account in model estimation — under the Waggoner and Zha (1998) approach — yields extremely similar results. Waggoner and Zha develop a Gibbs sampling algorithm that provides the exact finite-sample distribution of the conditional forecasts. Our reasons for abstracting from their extension are primarily computational. Using their algorithm would greatly add to the time required to produce all of the forecasts needed with the Pesaran-Pick average window approach. Moreover, with the size of the large model we use, their Gibbs sampling algorithm would be extremely slow, due to computations of an extremely large VAR coefficient variance-covariance matrix.

In contrast to the minimum-MSE approach we use for the conditional forecast results in the paper, it is more standard with DSGE-based forecast models to feed in structural shocks to hit the path of interest. In particular, the common DSGE model approach to achieving conditions on the policy path rests on feeding in structural shocks to monetary policy needed to hit the policy path. Under this approach, the scheme for identifying policy shocks matters for the conditional forecasts.[12]

3 Data

In our formal assessment of break probabilities, time variation in model estimates, and conditional forecasts, we use quarterly data for 1959:Q1-2013:Q4 obtained in early August

[12] In a supplementary appendix available upon request, we provide real-time out-of-sample forecast results that use this structural policy shock approach to conditioning on a slightly positive funds rate path over the 2009-2012 period. These results are qualitatively similar to the results we provide below.

2014 from the Federal Reserve Board's FAME database.

In the analysis of real-time out-of-sample forecasts that concludes the paper, we use real time data vintages from 1996:Q1 through 2014:2 (with data ending in 2013:Q4 or earlier). We obtained the real-time data vintages from the Federal Reserve Bank of Philadelphia's Real-Time Dataset for Macroeconomists (RTDSM), described in Croushore and Stark (2001). We don't use earlier vintages of data because they are not available for core PCE inflation. 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, nonfarm payroll employment, and headline PCE inflation. For the other variables, for which data are either not revised or only slightly revised, we rely on just currently available time series (these series are unemployment, the federal funds rate, the credit spread, and stock prices). 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, as discussed by Romer and Romer (2000) and Croushore (2006), evaluating the accuracy of real-time forecasts requires a difficult decision on what to use as the actual data when calculating forecast errors. 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 and then shows time series of model estimates. The section subsequently reports unconditional and conditional forecasts for the period 2008:Q1-2013:Q4 for a set of key variables.

4.1 More formal 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(4) 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 the beginning of 2008 (as well as at some other dates).[13] Let T_B denote the date of the possible break. We specify a VAR with coefficients having

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

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.

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, Minnesota-type shrinkage controlled by a hyperparameter λ_B, and tighter shrinkage on longer lags than shorter lags. Because we don't have much data for estimating a break that could have begun in 2008, this prior on the break is necessarily informative. Note, however, that the prior means of zero on the post-break coefficient changes mean 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, we search over a grid of values for the overall shrinkage hyperparameter and the break shrinkage hyperparameter λ to maximize the marginal likelihood. In the second, we fix the hyperparameters at 0.2 for overall shrinkage and 0.1 for the break shrinkage. In each case, we compute break probabilities from marginal likelihoods (posterior odds ratios) in the usual way, from a model without a break and a model with a break at the indicated, single date. The marginal likelihoods are computed with the analytical solution available for the Normal-inverted Wishart prior and posterior.

We begin by using the full sample of data to evaluate the probability of a 2008:Q1 break in VAR coefficients (i.e., the probability of a break that began with the most recent recession). The estimates in the above table put the probability at 100 percent (with both an optimized prior and fixed prior).

Of course, there are other possible break dates; for example, Strahan and van Dijk (2013) find evidence of a break in VAR coefficients in 1984. The second row of the table corroborates their finding in the full sample of 1961-2013 data, putting the 1984 break probability at essentially 100 percent (with both a fixed prior and an optimized prior). However, when the prospect of a break in 2008 is eliminated by ending the estimation

Table 1: **Probabilities of break in coefficients of 4-variable BVAR(4)**

estimation start	estimation end	break date	break probability (%)
optimized prior			
1961:Q1	2013:Q4	2008:Q1	100.0
1961:Q1	2013:Q4	1984:Q1	99.8
1961:Q1	2007:Q4	1984:Q1	90.8
1985:Q1	2013:Q4	2008:Q1	100.0
fixed prior: overall tightness = 0.2, break tightness = 0.1			
1961:Q1	2013:Q4	2008:Q1	100.0
1961:Q1	2013:Q4	1984:Q1	100.0
1961:Q1	2007:Q4	1984:Q1	5.8
1985:Q1	2013:Q4	2008:Q1	100.0

Note: The probability of a break is computed as $\frac{r}{1+r}$, where r, the posterior odds ratio, is the marginal likelihood of the model with a break divided by the marginal likelihood of the model without a break.

sample in 2007, the evidence of a 1984 break is more mixed: the probability of a 1984 break in the 1961-2007 estimates is 90.8 percent with the optimized prior but only 5.8 percent with the fixed prior. When the prospect of a 1984 break is accommodated by shortening the estimation sample to start in 1985, the estimates point to a very high probability of a break in 2008: the last rows of the table panels put the probability of a 2008 break in 1985-2013 data at 100 percent with both of the priors considered.[14]

Moreover, while this analysis has assumed the most recent break date to be 2008:Q1, in line with the timing of the Great Recession, there is considerable uncertainty about the exact timing of any break. We have repeated the break analysis of this section for break dates of 2007:Q1 and 2009:Q1 and obtained similar results.[15]

To provide some sense of where the most important breaks may lay in the reduced form VAR specification, Table 2 reports the posterior mean (and standard deviation) estimates of coefficient changes in a BVAR that allows for a coefficient break in 2008:Q1. For brevity, we report only results based on the marginal likelihood-optimized prior. For the model estimated with data starting in 1961, there is a substantial change in the unemployment

[14] We obtain similar results for a model augmented to include the spread between the BAA corporate bond rate and the 10-year Treasury bond and for an alternative VAR specification including the GDP gap, unemployment gap and the inflation gap (defined as inflation less trend inflation measured by a long-run survey forecast).

[15] However, in the case of fixed (rather than optimized) hyperparameters, the evidence of a break in 2007:Q1 is weaker than the evidence for 2008:Q1 or 2009:Q1.

Table 2: **Posterior mean estimates of coefficient changes in 4-variable BVAR(4) with optimized prior**

right-hand side variables	GDP growth	Unemployment rate	Core PCE inflation	Federal funds rate
		equation for:		
1961-2013 sample, break in 2007				
GDP growth	0.067 (0.159)	-0.022 (0.012)	-0.004 (0.045)	-0.028 (0.048)
Unemployment rate	-0.463 (0.177)	0.029 (0.014)	0.048 (0.052)	0.063 (0.056)
Core PCE inflation	-0.326 (0.612)	0.014 (0.049)	-0.131 (0.174)	-0.131 (0.190)
Federal funds rate	-0.470 (0.432)	0.016 (0.035)	-0.036 (0.123)	-0.189 (0.130)
Intercept	-0.071 (0.478)	0.007 (0.039)	-0.007 (0.138)	-0.010 (0.147)
1985-2013 sample, break in 2007				
GDP growth	0.062 (0.200)	-0.033 (0.016)	-0.023 (0.058)	-0.041 (0.034)
Unemployment rate	-0.328 (0.203)	0.031 (0.016)	0.010 (0.060)	0.036 (0.035)
Core PCE inflation	-0.265 (0.770)	0.022 (0.063)	-0.203 (0.226)	-0.099 (0.132)
Federal funds rate	-0.481 (0.491)	-0.025 (0.041)	0.106 (0.144)	-0.208 (0.084)
Intercept	-0.044 (0.450)	0.002 (0.037)	-0.001 (0.133)	-0.001 (0.079)

Note: Reported results are posterior means of sums of coefficients for lags of each variable in each equation. Numbers in parentheses are posterior standard deviations.

rate coefficients of the GDP growth equation as well as a change in the unemployment rate coefficients of the unemployment rate equation itself. For the model estimated with data starting in 1985, there are pretty sizable changes in both the unemployment rate and interest rate coefficients of the GDP equation, but there is little that is large relative to the standard deviations. Despite the high likelihood of a break in the VAR's parameters, the post-break sample is short enough to make it difficult to estimate the coefficient changes with much precision.

Overall, this break analysis shows evidence of VAR model instabilities over the 2008-2013 period. While 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 zero lower bound on interest rates, in reduced form estimates the instability is mostly evident in the relationship between GDP growth and the unemployment rate.

4.2 Time series of coefficient estimates

As a further check of model stability, we consider time series of coefficient estimates for two different models. The first is the baseline constant parameter BVAR estimated with rolling 20 year windows of observations, with prior optimized for each sample to maximize the marginal likelihood.[16] The second is the VAR with TVP and stochastic volatility detailed above. To streamline presentation, we report just the sum of coefficients (across lags) for each variable in each equation, along with the mean of each variable implied by the VAR estimates at each point in time. For all results we report 70 percent probability bands (posterior credible sets).

Figure 1 reports results for the rolling window estimates. Note that the date on the horizontal axis refers to the end point of the 20 year rolling window of data used to obtain the reported estimate (the posterior median and credible set). The estimates suggest significant instability over time, including, most importantly for our purposes, following the 2007-2009 recession (because the estimates are based on 20 year windows of data, we shouldn't expect them to pinpoint the break in 2007 suggested by the more formal break analysis of the preceding section). In the GDP growth equation, the sum of coefficients on lagged GDP has been trending up since 1985. The same is true for the sum of coefficients on the federal funds rate in the GDP equation. In the same equation, the sum of coefficients on unemployment fell sharply after the 2007-2009 recession, by an amount comparable to the decline that occurred in the early 1990s. In the unemployment rate equation, any changes in coefficients mostly look to be quantitatively small. The same mostly applies to the inflation equation, except that there is a more meaningful rise and fall in the sum of coefficients on lagged inflation, with a significant fall in the sum of coefficients in the late 1990s and post-recession. Also, with the rolling window approach, the coefficients on unemployment in the inflation equation materially rise from the early part of the chart sample to the present. In the federal funds rate equation, the coefficients on unemployment show a more meaningful rise, concentrated in the most recent recession and recovery. Finally, while we omit a chart in the interest of brevity, the rolling window estimates yield considerable variation in implied means of GDP growth, the unemployment rate, inflation, and the funds rate.

Figures 2 and 3 show that estimates of the VAR-TVP-SV specification yield similarly broad evidence of coefficient change over time, but less evidence of any change since 2007. (The limited evidence of a change since 2007 is perhaps not surprising in light of typical end-of-sample filtering challenges and the effects of Kalman smoothing; unreported estimates of the filtered, but not smoothed, coefficients suggest more change around 2007, in line with the rolling window BVAR estimates of Figure 1.) In the GDP growth equation, the coefficients

[16] Our use of a 20-year rolling window follows such studies as Del Negro and Schorfheide (2004) and Gurkaynak, et al. (2013).

on GDP and the funds rate have trended up significantly over time, while the coefficients on unemployment have trended down. In the unemployment equation, coefficient movements have mostly been relatively small. The same is true for the funds rate equation. In the inflation equation, the sum of coefficients on past inflation has declined materially (even if not necessarily statistically significantly, given fairly wide credible sets), as has the sum of coefficients on the funds rate. Furthermore, the estimates show a steady downward trend in mean GDP growth and an upward and then downward trend in inflation and the funds rate. Finally, the estimates of the residual standard deviations show considerable variation over time, with some fairly significant co-movement. The volatility estimates for all four variables rise sharply during the mid 1970s and early 1980s recessions and then decline sharply during the mid 1980s and remain low until the 2007-2009 recession. During the 2007-2009 recession the volatility estimates rise sharply again, but quantitatively by a smaller amount than during the mid 1970s and early 1980s recessions.

In summary, we highlight three implications from the time series of coefficient estimates shown above. (1) The rolling coefficient estimates look to be consistent with breaks in the mid-1980s and sometime around 2007. In particular, for our purposes, these estimates reveal material changes in the GDP and unemployment coefficients of the GDP equation over the last several years, consistent with some kind of recent break in the GDP–unemployment relationship. The TVP-SV show clear shifts in many sets of VAR parameters that begin in the early or mid-1980s, but less clear evidence of a break in 2007 or shortly thereafter. (2) While there is good reason to expect some shift in policy parameters in recent years, due to the zero lower bound, the rolling coefficient and TVP-SV estimates differ somewhat. The changes in the funds rate parameters look pretty small in the TVP-SV estimates. However, the changes are quite large in the rolling coefficient estimates. At a minimum, it is safe to say that a researcher using rolling BVAR estimates would likely see some changes in policy aspects of his/her model. (3) Mean shifts have also been pretty dramatic, largest for inflation and the funds rate, smaller but still material for GDP growth, and evident, although not necessarily all that large or important, for unemployment. Other studies, including Clark (2011), Wright (2013), and Stock and Watson (2012), have highlighted the broader historical importance of mean shifts.

4.3 Unconditional forecasts

As a further check of model stability, we study the performance of unconditional forecasts from the 4-variable BVAR in the period 2008:Q1-2013:Q4 — the period following the start of the recession with the NBER peak in 2007:Q4 — with models estimated using data samples ending with 2007:Q4. To provide basic checks of model stability, we consider forecasts produced with models estimated with 1961-2007 data and 1985-2007 data. The use of a sample starting in 1985 is motivated by Strahan and van Dijk's (2013) finding of a break in VAR dynamics in 1984. Large forecast errors are likely to either reflect instabilities in the dynamic system or that one or several large shocks have occurred.

The upper panel of Figure 4 provides unconditional forecasts obtained with a model estimated using 1961-2007 data. For this model specification, the actual paths of the unemployment rate and the federal funds rate fall well outside the forecast confidence bands. In particular, the actual unemployment rate differs greatly from the unconditional forecasts. Moreover, the actual path of GDP growth also falls well outside the forecast confidence bands for the second half of 2008 and first half of 2009, but lays well inside the bands for the 2010-2013 period. Finally, the forecasted path of inflation lays above the actual path, but still inside the confidence bands.

The lower panel of Figure 4 provides similar forecasts obtained with a model estimated using 1985-2007 data. In general, unconditional forecasts are very similar to the ones based on the longer estimation sample. However, consistent with the prospect of a model break in the early 1980s, the confidence bands are tighter for the forecasts obtained using the shorter estimation sample. If anything, the actual values falls even further outside the forecast confidence bands when using the shorter estimation sample.

Overall, the unconditional forecasting results corroborate our preceding evidence of VAR instabilities over the period 2008-2013 based on break analysis and rolling window and TVP-SV estimates of coefficients.

4.4 Conditional forecasts

As a final check of model stability we study conditional forecasts from various BVAR models over the period 2008:Q1-2013:Q4. 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. Recent studies by Giannone, Lenza, Pill, and Reichlin (2012), Giannone, Lenza, and Reichlin (2012), and Stock and Watson (2012) have used conditional forecasts to study stability in various economic relations during the Great Recession.[17] In the appendix, we provide 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. To illustrate the power of using conditional forecasts for detecting instabilities, we first provide results for a Monte Carlo simulation study.

4.4.1 Monte Carlo analysis of conditional forecast power

Suppose that the data-generating process (DGP) is a bivariate zero-mean stationary VAR(1) taking the form

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_t \\ v_t \end{pmatrix},$$

with i.i.d. N(0,1) errors with contemporaneous correlation ρ.

We produce both unconditional forecasts and conditional forecasts, where the latter are conditioned on the actual path of variable x over the forecast horizon, here specified as 12 quarters. For a given data set, we use Monte Carlo simulations to obtain forecast confidence bands, at significance levels of 30 and 10 percent. We then compare the actual path of the variable y to the forecast confidence band. In particular, in our context, we want to know if, when there is a break in the data generating process, are actual outcomes likely to fall outside the confidence bands (do the comparisons have power)? If instead the model is stable, are the actual outcomes likely to lay within the confidence bands (do the comparisons have the intended size)? In these exercises, we use an artificial (quarterly) data sample of 1985-2013.

In the size experiment (DGP1), we assume the model is constant over 1985-2013 sample with $a = 0.5$, $b = 0.1$, $c = 0.8$ and $\rho = -0.5$. For the break experiment, we consider two different simulation experiments. In first break experiment (DGP2), there is a break in just

[17] 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.

one coefficient, with b rising from 0.1 to 0.4 starting with 2008:Q1. In the second break experiment (DGP3), there is both a similar break in b (from 0.1 to 0.4) and a break in the correlation between innovations to y and x, from a pre-break value of -0.5 to a post-break value of 0.2. For each experiment we generate 5000 artificial data sets, produce forecast confidence bands for each data set, and then track the rates at which the outcome for y falls outside the forecast confidence band.[18] Finally, to help figure out the drivers of differences for unconditional and conditional forecasts, we compare the width of confidence intervals. In each Monte Carlo data set, for a given variable and forecast horizon, we compute the widths of the 70 and 90 percent confidence intervals.

Table 3 provides rejection rates (1 minus coverage rates) and confidence band spreads (averaged across Monte Carlo trials) for forecasts of y obtained in three experiments. In the size experiment with DGP1, we would like to see rejection rates close to the nominal rates of 30 and 10 percent. In the power experiments, we would like to see rejection rates above the nominal sizes of 30 and 10 percent. In the results, in the stable DGP1, the rejection rates for both unconditional and conditional forecasts are close to the nominal rates. As expected, on average, confidence bands are somewhat narrower for conditional forecasts than unconditional. In addition, the simulation results indicate that both unconditional forecasts and conditional forecasts have power. Moreover, the conditional forecast comparison has better power than the unconditional forecast comparison. The differences are fairly small for DGP2, where there is just a break in the slope coefficient while the conditional forecast comparison has a bigger forecast advantage when there is also a break in the error correlation, as in DGP3. Results for the confidence band comparison shows that the conditional forecast confidence bands are narrower than the unconditional confidence bands, consistent with the analytical results provided in the appendix.

4.4.2 Break evidence based on conditional forecasts

The Monte Carlo simulations above indicate that conditional forecasts have power in terms of detecting instabilities. We therefore begin by checking the consistency of the evolution of the economy in the period 2008:Q1-2013:Q4 — the period following the start of the recession with the NBER peak in 2007:Q4 — with models estimated using data samples ending with 2007:Q4. In particular, we compare the actual evolution of unemployment,

[18]Note that the rejection rates are computed on a variable-by-variable and horizon-by-horizon basis, at a maximum horizon of 12 periods.

Table 3: **Results for DGP1 and DGP2 (just variable y)**

Measure	$h=1$	$h=2$	$h=4$	$h=8$	$h=12$
DGP1 (no breaks)					
reject rate, 30% signif.: uncond.	0.315	0.322	0.302	0.313	0.328
reject rate, 30% signif.: cond.	0.300	0.321	0.315	0.328	0.336
reject rate, 10% signif.: uncond.	0.111	0.115	0.109	0.107	0.120
reject rate, 10% signif.: cond.	0.104	0.116	0.116	0.119	0.123
band spread, 30% signif.: uncond.	2.110	2.261	2.306	2.316	2.318
band spread, 30% signif.: cond.	1.886	2.027	2.063	2.067	2.075
band spread, 10% signif.: uncond.	3.361	3.606	3.678	3.697	3.700
band spread, 10% signif.: cond.	3.004	3.231	3.291	3.298	3.310
DGP2 (break in b, constant ρ)					
reject rate, 30% signif.: uncond.	0.306	0.311	0.324	0.320	0.329
reject rate, 30% signif.: cond.	0.330	0.374	0.398	0.403	0.396
reject rate, 10% signif.: uncond.	0.103	0.113	0.115	0.110	0.127
reject rate, 10% signif.: cond.	0.123	0.159	0.179	0.182	0.187
band spread, 30% signif.: uncond.	2.114	2.267	2.311	2.322	2.324
band spread, 30% signif.: cond.	1.890	2.029	2.068	2.078	2.085
band spread, 10% signif.: uncond.	3.368	3.614	3.688	3.707	3.709
band spread, 10% signif.: cond.	3.011	3.235	3.299	3.316	3.327
DGP3 (break in both b and ρ)					
reject rate, 30% signif.: uncond.	0.306	0.323	0.348	0.361	0.368
reject rate, 30% signif.: cond.	0.449	0.489	0.511	0.516	0.513
reject rate, 10% signif.: uncond.	0.103	0.121	0.137	0.141	0.156
reject rate, 10% signif.: cond.	0.222	0.263	0.291	0.300	0.298
band spread, 30% signif.: uncond.	2.114	2.267	2.311	2.322	2.324
band spread, 30% signif.: cond.	1.893	2.037	2.087	2.113	2.117
band spread, 10% signif.: uncond.	3.368	3.614	3.688	3.707	3.709
band spread, 10% signif.: cond.	3.015	3.248	3.330	3.372	3.379

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 size of the shocks that would have caused the recent recession if it were due to the shocks that have typically generated recessions. 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 2008-2013 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 consider forecasts produced with models estimated with 1961-2007 data and 1985-2007 data. We will report here results for a range of specifications; the appendix describes robustness in still more specifications.

The upper panel of Figure 5 provides conditional forecasts obtained with a model estimated using 1961-2007 data. For this model specification, the actual paths of unemployment, inflation, and the funds rate fall well outside the conditional forecast confidence intervals. The forecasted path of inflation generally lays well above the actual path, probably reflecting a previously documented tendency of models that do not in some way take account of mean shifts in inflation to yield upward biased forecasts (e.g., Clark (2011) and Wright (2013)). The same is true for the federal funds rate. Finally, and what will turn out to be most significantly 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.

The lower panel of Figure 5 provides similar forecasts obtained with a model estimated using 1985-2007 data. Consistent with the prospect of a model break in the early 1980s, the actual paths of unemployment, inflation, and the funds rate are more consistent with the conditional forecasts from this version of the model than with the forecasts from the longer-sample version of the model. In particular, using a shorter sample lowers the forecast paths of inflation and the funds rate, such that the actual paths generally lay within the forecast confidence bands, particularly over the recovery period. However, even with this model, it remains the case that the actual path of unemployment lays well outside the conditional forecast confidence interval.

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 6, 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 remains the case that the actual path of unemployment is far outside the conditional forecast bands.

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 (2014) 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 zero lower bound. 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).[19] The results are reported in Figure 7. 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. Interestingly, conditioning on the spread and the funds rate improves the forecast of inflation; similarly, conditioning on just GDP growth and the spread improves forecasts of both inflation and the funds rate. This pattern is broadly consistent with findings in Christiano, Eichenbaum and Trabandt (2014), Del Negro, Giannoni, and Schorfheide (2014) and Gilchrist, et al. (2014) that financial constraints during the Great Recession influenced the response of inflation.

To provide a further check on coefficient change, we consider another conditional forecasting exercise, using the baseline model: we estimate the model with samples including data up through the end of 2013 and construct pseudo-forecasts conditional on the actual path of GDP. These results are shown in Figure 8. When the estimation sample begins with 1961:Q1, extending the data sample though 2013 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 the estimation sample ends in 2007:Q4). But the actual evolution of unemployment and the funds rate still falls fairly well outside the forecast bands. When the estimation sample begins with 1985:Q1, consistency between the actual paths and conditional forecasts improves further. Based on the 1985-2013 model estimates, the actual paths lay within the conditional forecast bands. The contrast with the results for models estimated with data through 2007 suggests a material change in coefficients between samples ending in 2007 versus 2013 and, in turn, a post-2007 coefficient break.

One might also wonder if the recent disconnect between GDP growth and unemployment

[19] When conditioning on GDP growth and the spread (not the funds rate, in this previous analysis), we obtained similar results with alternative credit condition indicators, including housing prices, the GZ spread used by Gilchrist and Zakrajsek (2012) and the Chicago Fed's index of financial conditions or stress.

has a precedent in other recent recessions and recoveries. To provide some check on this, we use the version of the model augmented to include employment growth to produce conditional forecasts for the 24 quarters (a duration corresponding to the one we use for the most recent recession) following the previous business cycle peaks of 1990:Q3 and 2000:Q1. We include employment growth in the model in light of conventional wisdom that views the recoveries following the 1990 and 2000 recessions as different, "jobless" recoveries. We estimate models with data samples ending at each of these points in time and produce forecasts conditional on the path of actual GDP growth. In the interest of brevity, we report in 9 results for just the 2000 recession; results for the 1990 recession are qualitatively similar.

For the early 2000s, we observe again 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. A similar pattern emerges following the recession that began in 2000:Q1, in particular when the estimation sample starts in 1961Q1. Hence, the GDP-unemployment relationship observed in the most recent recession and recovery is indeed different: there appears to be a break that did not emerge following the preceding two recessions.

To summarize, the results above from conditional forecasts are consistent with breaks in VAR dynamics in the mid-1980s and in 2007 or shortly thereafter. In the 2007 case, there is clearly a shift in the GDP–unemployment relationship. While this 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 zero lower bound and the Great Recession may have contributed in ways that are difficult to disentangle in reduced form VARs. The 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 in at least the early or mid-1980s and sometime around 2007-08. 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 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) and VARs estimated with different samples: recursive, starting in 1961; 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. 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. 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.

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 2014), 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.

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. We focus on the recovery period of

2009:Q3-2013:Q4 that followed the Great Recession, but in charts we provide results for longer samples going farther back in time, to 1996:Q1.

We first consider the accuracy of point forecasts (defined as posterior medians), using RMSEs. We then consider density forecasts, using the CRPS, suggested by Gneiting and Raftery (2007) and Gneiting and Ranjan (2011). The CRPS, defined such that a lower number is a better score, is given by

$$CRPS_t(y^o_{t+h}) = \int_{-\infty}^{\infty} \left(F(z) - 1\{y^o_{t+h} \leq z\}\right)^2 dz = E_f|Y_{t+h} - y^o_{t+h}| - 0.5E_f|Y_{t+h} - Y'_{t+h}|, \quad (4)$$

where F denotes the cumulative distribution function associated with the predictive density f, $1\{y^o_{t+h} \leq z\}$ denotes an indicator function taking value 1 if $y^o_{t+h} \leq z$ 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 in the Federal Reserve and in other central banks' fan charts. Up until 2009, the forecasts are unconditional. Starting with forecasts generated in 2009:Q1, 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) Blue Chip Survey forecasts of the federal funds rate and Survey of Professional Forecasters projections of the 3-month Treasury bill rate. As detailed in section 2, we generate the conditional forecasts using the Kalman filter approach to computing the minimum-MSE forecast described in Banbura, Giannone, and Lenza (2014).[20]

As a benchmark model we consider a constant parameter model estimated recursively with data starting in 1961. 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.

[20]For a given model and estimation approach, when we compare the efficacy of the DLS method for conditional forecasting to the structural policy shock approach, the results are mixed. Late in the recession and perhaps early in the recovery, forecasts based on policy shock conditioning are more accurate than forecasts based on DLS conditioning. But for the bulk of the recovery, the reverse is true.

5.2 Point forecasts

In Figure 10 our forecasts start with the 1996:Q1 origin, using data through the preceding quarter to estimate models and form forecasts. For each date t shown in the chart, we compute a shrinking window of RMSEs, from period t through 2013:Q4. As time moves forward, the RMSE is based on fewer and fewer observations, until the last observation in the chart, which is based on RMSEs for the eight forecast observations of 2012:Q1 through 2013:Q4 (dates refer to the date of the forecast outcome). This approach is useful for isolating the performance of each method in the period since the crisis.

In general the figure shows that allowing for time variation of the parameters increases forecast accuracy. We begin by briefly considering the full evaluation period 1996Q1-2013Q4, captured by the first (in time) observation in the chart. Over this period, all of the methods generally improve on the recursive, 1961 start baseline. Consistent with the results of D'Agostino, Gambetti, and Giannone (2013), the VAR-TVP-SV specification performs best or about the best for all variables except inflation. For inflation, most of the methods considered offer modest gains over the constant parameter, recursive sample baseline.

For the purpose of evaluating methods that worked relatively well in the recovery following the 2007-2009 crisis, we are most interested in the last (in time) several observations in each chart panel. Interestingly, there are substantial changes in the relative performance of the different methods during and after the 2007-2009 crisis, where forecast accuracy, relative to the baseline model, increases for GDP growth and inflation and decreases for the unemployment rate and the interest rate. To help clearly results that isolate the recovery period, we provide forecast RMSE ratios for 2009:Q3-2013:Q4 in Table 4.[21]

For forecasting GDP growth, all of the methods considered for allowing variation in parameters greatly improves the accuracy of forecasts of GDP growth. The BVAR-TVP-SV model works best, while approaches of using either a 20 year rolling window or a sample starting in 1985 also work relatively well. For forecasting the unemployment rate, all of the methods considered increase 4-step ahead forecast accuracy relative to the recursively generated forecast from the baseline model, but have mixed effects on 1-step ahead forecast accuracy. The same results also apply for forecasting inflation. The only exception is that forecasts produced by a model using a 20 year rolling window are slightly worse than the

[21] With just 18 observations of forecast errors for this period, it would be very difficult to establish statistical significance, so we don't report significance indicators.

Table 4: **RMSEs, 2009:Q3-2013:Q4 forecasts from 4-variable BVAR**

model or estimation approach	GDP growth	Unemployment rate	Core PCE inflation
1-quarter horizon			
recursive, 1961 start	2.287	0.277	0.579
recursive, 1985 start	0.728	0.953	0.928
rolling, 80 obs. window	0.740	1.020	1.037
avg. rolling window	0.809	1.012	0.976
TVP-SV	0.594	0.986	0.986
4-quarter horizon			
recursive, 1961 start	3.393	1.496	0.667
recursive, 1985 start	0.631	0.831	0.844
rolling, 80 obs. window	0.685	0.852	0.880
avg. rolling window	0.746	0.872	0.870
TVP-SV	0.598	0.840	0.818

Note: RMSE levels for baseline forecast, ratios for all others.

baseline model.

5.3 Density forecasts

Central banks and other forecasters are increasingly interested in various aspects of density forecasts. Several studies have shown that allowing for time-varying parameters and volatility materially improves the real-time accuracy of density forecasts (e.g., Clark 2011 and Clark and Ravazzolo 2014). In Figure 11, we report density accuracy computed with a shrinking window of average CRPSs. For each date t shown in the chart, we compute the forecast average CRPS from t through 2013:Q4. As time moves forward, the average CRPS is based on fewer and fewer observations, until the last observation in the chart, which is based on the average CRPS for the eight forecast observations of 2012:Q1 through 2013:Q4. All results are shown as relative to the baseline of a model with constant parameters estimated recursively with data starting in 1961.

Broadly, the CRPS results are quite similar to the RMSE results. Figure 11 shows that allowing for time variation of the parameters, in general, increases the accuracy of density forecasts (full sample results can be read from the first observation of Figure 11. However, for each of the variables there are large changes in the relative performance of the different methods during and after the 2007-2009 crisis. There are considerable gains in terms of increased accuracy of density forecasts from the different methods, relative to the baseline

Table 5: **CRPSs, 2009:Q3-2013:Q4 forecasts from 4-variable BVAR**

model or estimation approach	GDP growth	Unemployment rate	Core PCE inflation
1-quarter horizon			
recursive, 1961 start	1.310	0.161	0.338
recursive, 1985 start	0.724	0.973	0.910
rolling, 80 obs. window	0.743	1.039	1.036
avg. rolling window	0.794	1.023	0.963
TVP-SV	0.658	0.984	1.007
4-quarter horizon			
recursive, 1961 start	2.170	0.801	0.382
recursive, 1985 start	0.573	0.823	0.819
rolling, 80 obs. window	0.611	0.858	0.889
avg. rolling window	0.660	0.880	0.856
TVP-SV	0.499	0.834	0.843

Note: CRPS levels for baseline forecast, ratios for all others.

model, when forecasting GDP growth. In particular, the BVAR-TVP-SV model seems to provide accurate forecasts. While there are also gains in terms of more accurate forecasts from the different models relative to the baseline model for the other variables, the relative performance decreases during the crisis and recovery period. Table 5 provides the relative CRPS results for each model over the 2009:Q3-2013:Q4 recovery period.[22]

5.4 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, core PCE inflation, the unemployment rate and the funds rate. We take a constant parameter 13-variable model estimated recursively with data starting in 1961 as our baseline model. 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 Figure 12 (while we omit a figure of CRPS results in the interest of brevity, they are similar to the RMSE

[22] We also obtained very similar results when measuring density forecast accuracy using log scores instead of CRPS. For brevity, the log score results are not reported here.

Table 6: **RMSEs, 2009:Q3-2013:Q4 forecasts from 13-variable BVAR**

model or estimation approach	GDP growth	Unemployment rate	Core PCE inflation
1-quarter horizon			
recursive, 1961 start	1.614	0.293	0.484
recursive, 1985 start	0.867	1.030	1.065
rolling, 80 obs. window	0.853	1.053	1.168
avg. rolling window	0.867	1.029	1.127
TVP-KK	0.947	1.071	1.282
4-quarter horizon			
recursive, 1961 start	1.825	1.229	0.788
recursive, 1985 start	1.124	0.792	0.906
rolling, 80 obs. window	1.080	0.820	1.003
avg. rolling window	1.003	0.782	0.790
TVP-KK	0.811	1.004	0.922

Note: RMSE levels for baseline forecast, ratios for all others

results) and Tables 6 and 7.

In the case of the model with 13 variables, in general, it is much more difficult to use a shorter sample for estimation and materially improve on the accuracy of a model estimated recursively with data starting in 1961. This probably has to do with model size and precision of parameter estimates.[23] The full 1996-2013 sample results captured by the first observation in Figure 12 show that, with a larger model, the baseline approach of recursive estimation with data starting in 1961 is hard to beat for GDP growth and the unemployment rate. However, the approaches considered for accommodating parameter change work much better for forecasting inflation and somewhat better for forecasting the funds rate at short horizons.

However, around the last recession and recovery, there were some significant shifts in relative performance of the different methods. The results near the end of the period covered in Figure 12 and for the post-recession sample covered in Table 6 show that, at the one quarter horizon for the recovery period of 2009Q3-2013Q4, all of the approaches improve on the baseline forecast in the case of GDP growth forecasts, but yield less accurate inflation and unemployment rate forecasts. At the 4-quarter horizon, the picture is the opposite, with most methods providing more accurate forecasts for inflation and the unemployment

[23] Note that the 13 variable baseline model provides superior forecasts compared to the 4 variable baseline model for GDP growth. For the unemployment rate and core inflation the forecast accuracy are more similar.

Table 7: **CRPSs, 2009:Q3-2013:Q4 forecasts from 13-variable BVAR**

model or estimation approach	GDP growth	Unemployment rate	Core PCE inflation
1-quarter horizon			
recursive, 1961 start	0.991	0.165	0.299
recursive, 1985 start	0.839	1.104	1.006
rolling, 80 obs. window	0.848	1.113	1.092
avg. rolling window	0.841	1.083	1.059
TVP-KK	0.924	1.074	1.174
4-quarter horizon			
recursive, 1961 start	1.039	0.648	0.435
recursive, 1985 start	1.027	0.834	0.925
rolling, 80 obs. window	1.024	0.869	1.031
avg. rolling window	0.914	0.824	0.802
TVP-KK	0.795	1.227	0.962

Note: CRPS levels for baseline forecast, ratios for all others

rate than the baseline model. Again, results for density forecasts as captured in CRPS comparisons are similar to the point forecast results. Table 7 provides CRPS comparisons for the 2009-2013 sample.

5.5 Hybrid survey-model forecasts

As a number of studies have shown, forecasts obtained from surveys of professional forecasters are often more accurate than forecasts obtained from time series models (e.g., Faust and Wright (2009, 2013)). While models designed for nowcasting can perform comparably to surveys in short-term forecasting, surveys can be difficult to match around business cycle turning points (see, e.g., Carriero, Clark, and Marcellino 2014), probably reflecting the benefits of judgment. At medium-term forecast horizons, survey-based forecasts can have a number of advantages over common model-based methods: the survey-based forecasts often have access to information more timely than that used in constructing model forecasts; the survey-based forecasts can be based on a wider set of variables; and survey-based forecasts can incorporate subjective judgment that may be helpful for forecasting.

Of these differences, the timeliness of information may be most important (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, 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 forecast 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 (using the Doan, Litterman, and Sims 1984 approach). 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 4.

Results for point forecasts from our baseline model and the hybrid specifications are provided in Table 8 and Figure 13. At the 1-quarter horizon, conditioning the model forecasts on the survey-based forecast yields — by construction — the same (survey-based) forecasts and forecast accuracy. At the 4-quarter horizon, using the survey-based current-quarter forecast as the jumping-off point for model forecasts significantly improves the accuracy of the model forecasts: RMSE ratios are lower in Table 8 and Figure 13 than in the pure model-based results provided in Table 4 and Figure 10. However, using the survey-based current-quarter forecast as the jumping-off point for model forecasts does not seem to noticeably reduce the changes in relative forecast accuracy that occur over time: the shrinking window RMSE ratios for hybrid forecasts in Figure 13 move around just about as much as the RMSE ratios of pure model-based forecasts in Figure 10.

Table 8: **RMSEs, 2009:Q3-2013:Q4 forecasts, VAR-SPF hybrids**

model or estimation approach	GDP growth	Unemployment rate	Core PCE inflation
	1-quarter horizon		
rec., 1961 start	2.287	0.277	0.579
hybrid, 1961 start	0.419	0.113	0.701
hybrid, 1985 start	0.419	0.113	0.701
hybrid, rolling, 20y	0.419	0.113	0.701
hybrid, avg. window	0.419	0.113	0.701
hybrid, TVP-SV	0.419	0.113	0.701
	4-quarter horizon		
rec., 1961 start	3.393	1.496	0.667
hybrid, 1961 start	0.756	0.737	0.761
hybrid, 1985 start	0.487	0.625	0.747
hybrid, rolling, 20y	0.541	0.645	0.732
hybrid, avg. window	0.594	0.662	0.770
hybrid, TVP-SV	0.427	0.620	0.724

Note: RMSE levels for baseline VAR forecast, ratios for all others. The other forecasts are from 4-variable VAR specifications (corresponding to those in Table 1), obtained by conditioning on SPF forecasts for the 1-quarter ahead horizon.

5.6 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. For small models, there seem to be fairly consistent advantages to using a model with TVP or either allowing a break in 1985, using a 20 year rolling window, or taking an average of forecasts from models using different rolling windows. It seems hard to say that one of these approaches is clearly better than the other. However, in the absence of a strong 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 a long sample may be conceptually preferable, because it does not hinge on the ability to identify a specific break date.

With large models, it is harder to establish consistent advantages to either allowing a break in 1985, using a 20 year rolling window, taking an average of forecasts from models using different rolling windows, or allowing TVP as in Koop and Korobilis (2013). However, in general, the larger information set makes the baseline constant parameter model more competitive. It either provides some insulation from instability in forecasting (even if the conditional forecast results of section 4 using the large model clearly point to instabilities in the large model similar to those of the small model) or, due to the number of parameters to be estimated, makes it difficult to use TVP or simple approaches such as rolling window estimation to improve forecast accuracy through some accommodation of structural change.

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 sharp recession of 2007-2009.

We first use a variety of approaches to assess model stability around the Great Recession, all of which suggest significant instabilities. Specifically, we show that VARs produce large forecast errors during and after the crisis, even when conditioning on the actual evolution of GDP growth. Moreover, when parameter time variation is allowed, there is substantial variation in the time series of coefficient estimates. More formal (Bayesian) analysis provides additional evidence against VAR parameter stability.

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 1961; 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, but accounting for time variation turns out to be useful to improve the point and density forecasting performance of the models. The gains are larger and systematic in smaller VAR models but remain also in larger VARs.

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 zero 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 zero lower bound, extended forward guidance from central banks, and government bond purchases (see, e.g., Chen, Curdia, and Ferraro (2012)).

References

[1] Aastveit, K., Gerdrup, K., Jore, A., and Thorsrud, L., 2014. Nowcasting GDP in Real Time: A Density Combination Approach, *Journal of Business and Economic Statistics* 32(1), 48-68

[2] Alessi, L., Ghysels, E., Onorante, L., Peach, R., and Potter, S., 2014. Central Bank Macroeconomic Forecasting During the Global Financial Crisis, ECB Working Paper No. 1688.

[3] Banbura, M., Giannone, D., and Reichlin, L., 2010. Large Bayesian Vector Autoregressions, *Journal of Applied Econometrics* 25(1), 71-92.

[4] Banbura, M., Giannone, D., and Lenza, M., 2014. Conditional Forecasts and Scenario Analysis with Vector Autoregressions for Large Cross-Sections, *International Journal of Forecasting*, forthcoming.

[5] Barnett, A., Mumtaz, H., and 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(1), 129-143.

[6] Baumeister, C., and Benati, L., 2012. Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound, *International Journal of Central Banking* 9(2), 165-212.

[7] Bognanni, M., 2013. An Empirical Analysis of Time-Varying Fiscal Multipliers, manuscript, University of Pennsylvania.

[8] Carriero, A., Clark, T., and Marcellino, M., 2013. Bayesian VARs: Specification Choices and Forecasting Performance, *Journal of Applied Econometrics*, forthcoming.

[9] Carriero, A., Clark, T., and Marcellino, M., 2014. "Real-Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility," *Journal of the Royal Statistical Society, Series A*, forthcoming.

[10] Chen, H., Curdia, V., and Ferraro, A., 2012. The Macroeconomic Effects of Large-Scale Asset Purchase Programs, *The Economic Journal* 122(564), F289-F315.

[11] Cheng, X., Liao, Z., and Schorfheide, F., 2014. Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities, NBER Working Paper 19792.

[12] Christiano L., Eichenbaum, M., and Evans, C., 1996. The Effects of Monetary Policy Shocks: Evidence from the Flows of Funds, *Review of Economics and Statistics* 78 (1), 16-34.

[13] Christiano L., Eichenbaum, M., and Trabandt, M., 2014. Understanding the Great Recession, manuscript.

[14] Clarida R., and Coyle, D., 1984. Conditional Projection by Means of Kalman Filtering, NBER Technical Working Paper 0036.

[15] Clark T., 2011. Real-Time Density Forecasts from BVARs with Stochastic Volatility, *Journal of Business and Economic Statistics* 29, 327-341.

[16] Clark T., and 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.

[17] Clark T., and McCracken, M., 2014. Evaluating Conditional Forecasts from Vector Autoregressions, manuscript.

[18] Clark T., and Ravazzolo, F., 2014. Macroeconomic Forecasting Performance Under Alternative Specifications of Time-Varying Volatility, *Journal of Applied Econometrics*, forthcoming.

[19] Cogley, T., and Sargent, T., 2005. Drifts and Volatilities: Monetary Policies and Outcomes in the Post-WWII US, *Review of Economic Dynamics* 8, 262-302.

[20] Croushore, D., 2006. Forecasting with Real-Time Macroeconomic Data, in *Handbook of Economic Forecasting*, Elliott, G., Granger, C., and Timmermann, A. (eds). North Holland: Amsterdam.

[21] Croushore, D., and Stark, T., 2001. A Real-Time Data Set for Macroeconomists, *Journal of Econometrics* 105, 111-30.

[22] Daly, M., Hobijn, B., Sahin, A., and 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.

[23] D'Agostino, A., Gambetti, L. and Giannone, D., 2013. Macroeconomic Forecasting and Structural Change, *Journal of Applied Econometrics* 28(1), 82-101.

[24] Del Negro, M., Giannoni, M and Schorfheide, F., 2014. Inflation in the Great Recession and New Keynesian Models, *American Economic Journal: Macroeconomics*, forthcoming.

[25] Del Negro, M., and Schorfheide, F., 2004. Priors from General Equilibrium Models for VARs, *International Economic Review* 45, 643-673.

[26] Doan, T., Litterman, R., and Sims, C., 1984. Forecasting and Conditional Projection Using Realistic Prior Distributions, *Econometric Reviews* 3, 1-100.

[27] Erceg, C., and Levin, A., 2013. Labor Force Participation and Monetary Policy in the Wake of the Great Recession, *CEPR Discussion Paper* DP9668.

[28] Faust, J., and Wright, J., 2009. Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset, *Journal of Business and Economic Statistics* 27(4), 469-479.

[29] Faust, J., and Wright, J., 2013. Inflation Forecasting, in *Handbook of Economic Forecasting, Volume 2*, Elliott, G., and Timmermann, A. (eds). Amsterdam: North Holland.

[30] Francis, N., Jackson, L., and Owyang, M., 2014. How Has Empirical Monetary Policy Analysis Changed After the Financial Crisis? Federal Reserve Bank of St. Louis Working Paper 2014-019A.

[31] Gali, J., Smets, F., and Wouters, R., 2012. Slow Recoveries: A Structural Interpretation, *Journal of Money, Credit, and Banking* 44(2), 9-30.

[32] Giannone, D., Lenza, M., and Primiceri, G., 2012. Prior Selection for Vector Autoregressions, *Review of Economics and Statistics*, forthcoming.

[33] Giannone, D., Lenza, M., and Reichlin, L., 2012. Money, Credit, Monetary Policy and the Business Cycle in the Euro Area, manuscript.

[34] Giannone, D., Lenza, M., Pill, H., and Reichlin, L., 2012. The ECB and the Interbank Market, *Economic Journal* 122, F467-F486.

[35] Giannone, D., Reichlin, L., and Small, D., 2008. Nowcasting: The Real-Time Informational Content of Macroeconomic Data, *Journal of Monetary Economics* 55(4), 656-676.

[36] Gilchrist S., and Zakrajsek, E., 2012. Spreads and Business Cycle Fluctuations, *American Economic Review* 102(4), 1692-1720.

[37] Gilchrist S., Schoenle, R., Sim, J., and Zakrajsek, E., 2014. Inflation Dynamics During the Financial Crisis, manuscript.

[38] Giraitis, L., Kapetanios, G. and Price, S., 2012. Adaptive Forecasting in the Presence of Recent and Ongoing Structural Change, CAMA Working Papers 2012-14.

[39] Gneiting T., and Raftery A., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association* 102, 359-378.

[40] Gneiting T., and Ranjan, R., 2011. Comparing Density Forecasts Using Threshold and Quantile Weighted Proper Scoring Rules, *Journal of Business and Economic Statistics* 29, 411-422.

[41] Gurkaynak, R., Kisacikoglu, B., and 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*, *Advances in Econometrics* 32, Fomby, T., Kilian, L., and Murphy, A. (eds.)

[42] Hall, Robert E., 2010. Why Does the Economy Fall to Pieces after a Financial Crisis? *Journal of Economic Perspectives* 24, 3-20.

[43] Hubrich, K., and Tetlow, R., 2012. Financial Stress and Economic Dynamics: The Transmission of Crises," manuscript.

[44] Koop, G., 2013. Forecasting with Medium and Large Bayesian VARs, *Journal of Applied Econometrics* 28(2), 177-203.

[45] Koop, G., and Korobilis, D., 2013. Large Time-Varying Parameter VARs, *Journal of Econometrics* 177(2), 185-198.

[46] Koop, G., and Potter, S., 2007. Estimation and Forecasting in Models with Multiple Breaks, *Review of Economic Studies* 74, 763-789.

[47] Litterman, R., 1986. Forecasting with Bayesian Vector Autoregressions-Five Years of Experience, *Journal of Business and Economic Statistics* 4, 25-38.

[48] Mian, A., and Sufi, A., 2010. Household Leverage and the Recession of 2007-09, *IMF Economic Review* 58(1), 74-117.

[49] Ng, S. and Wright, J., 2013. Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling, *Journal of Economic Literature* 51(4), 1120-1154

[50] Pesaran, M., and Pick, A., 2011. Forecast Combination Across Estimation Windows, *Journal of Business and Economic Statistics* 29, 307-318.

[51] Primiceri, G., 2005. Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies* 72, 821-852.

[52] Ravenna, F., and Walsh, C., 2013. Slow Recoveries, Worker Heterogeneity, and the Zero Lower Bound, manuscript.

[53] Reinhart, C., and Rogoff, K., 2009. *This Time is Different: Eight Centuries of Financial Folly*, Princeton University Press.

[54] Romer, C., and Romer, D., 2000. Federal Reserve Information and the Behavior of Interest Rates, *American Economic Review* 90, 429-457.

[55] Sahin, A., Song, J., Topa, G., and Violante, G., 2012. Mismatch Unemployment, Federal Reserve Bank of New York Staff Report 566.

[56] Sims, C., 1993. A Nine-Variable Probabilistic Macroeconomic Forecasting Model, in *Business Cycles, Indicators and Forecasting*, Chicago: University of Chicago.

[57] Sims, C., and Zha, T., 1998. Bayesian Methods for Dynamic Multivariate Models, *International Economic Review* 39(4), 949-968.

[58] Stock, J., and Watson, M., 1996. Evidence of Structural Instability in Macroeconomic Time Series Relations, *Journal of Business and Economic Statistics* 14(1), 11-30.

[59] Stock, J., and Watson, M., 2012. Disentangling the Channels of the 2007-2009 Recession, *Brookings Papers on Economic Activity*, Spring, 81-130.

[60] Strahan, R., and van Dijk, H., 2013. Evidence on Features of a DSGE Business Cycle Model from Bayesian Model Averaging, *International Economic Review* 54, 385-402.

[61] Van Zandweghe, W., 2012. Interpreting the Recent Decline in Labor Force Participation, *Economic Review, Federal Reserve Bank of Kansas City*, 5-34.

[62] Waggoner, D., and Zha, T., 1999. Conditional Forecasts in Dynamic Multivariate Models, *The Review of Economics and Statistics* 81(4), 639-651

[63] Wright, J., 2013. Evaluating Real-Time VAR Forecasts with an Informative democratic Prior, *Journal of Applied Econometrics* 28(5), 762-776.

[64] Xia, F., and Wu, C., 2013. Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound, manuscript.

Appendix

In this appendix, we first provide simple analytical results to establish that the approach of comparing conditional forecasts to actual variable paths can provide evidence of parameter breaks. We then summarize the variety of robustness checks we conducted in the conditional forecast analysis.

A Conditional Forecasts

Suppose that the data-generating process is a bivariate a zero-mean stationary VAR(1) taking the form

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_t \\ v_t \end{pmatrix},$$

with i.i.d. N(0,1) errors with contemporaneous correlation ρ.

Suppose we produce forecasts at a two-step ahead horizon using estimated parameters, denoted, e.g., $\hat{a}$. The question is what we can learn about possible parameter instability from different types of forecast errors. Over the forecast horizon, the data will be determined by true values of parameters (without hats) that may have shifted from the time of estimation to the forecast period. We are particularly interested in determining whether a shift in the relationship between y and x as reflected in the coefficient b or the error correlation ρ can be detected from the behavior of forecasts.

First consider the unconditional forecast. At the two-step horizon, the unconditional forecast of y and the associated error are:

$$\begin{aligned} \hat{y}^u_{t+2} &= \hat{a}^2 y_t + \hat{b}(\hat{a} + \hat{c}))x_t \\ u^u_{t+2} &= u_{t+2} + au_{t+1} + bv_{t+1} + (a^2 - \hat{a}^2)y_t + (ab - \hat{a}\hat{b})x_t + (bc - \hat{b}\hat{c})x_t. \end{aligned}$$

Now consider the conditional forecast of y_{t+2} in period t (two steps ahead) obtained by conditioning on the actual values of x_{t+1} and x_{t+2}, using the approach of Doan, Litterman, and Sims (1984). As noted in Clark and McCracken (2014), the conditional forecast is

$$\hat{y}^c_{t+2} = \hat{y}^u_{t+2} + \hat{\rho}(\hat{x}^c_{t+2} - \hat{x}^u_{t+2}) + \left(\hat{b} + \hat{\rho}\,(a - \hat{c})\right)(\hat{x}^c_{t+1} - \hat{x}^u_{t+1}),$$

where

$$(\hat{x}^c_{t+2} - \hat{x}^u_{t+2}) = x_{t+2} - \hat{c}^2 x_t = v_{t+2} + cv_{t+1} + (c^2 - \hat{c}^2)x_t$$

and

$$(\hat{x}^c_{t+1} - \hat{x}^u_{t+1}) = x_{t+1} - \hat{c}x_t = v_{t+1} + (c - \hat{c})x_t.$$

The conditional forecast error is then

$$\hat{u}^c_{t+2} = \hat{u}^u_{t+2} - \hat{\rho}(\hat{x}^c_{t+2} - \hat{x}^u_{t+2}) - \left(\hat{b} + \rho\left(\hat{a} - \hat{c}\right)\right)(\hat{x}^c_{t+1} - \hat{x}^u_{t+1}),$$

which, substituting terms, can be rewritten as

$$\hat{u}^c_{t+2} = \hat{u}^u_{t+2} - \hat{\rho} v_{t+2} - \left(\hat{b} + \hat{\rho}\hat{a}\right) v_{t+1} - \hat{\rho}\left(c^2 - \hat{c}^2\right) x_t - \left(\hat{b} + \hat{\rho}\left(\hat{a} - \hat{c}\right)\right)\left(c - \hat{c}\right) x_t.$$

From the above solutions, can we figure out what a conditional forecast can tell us about structural breaks that an unconditional forecast cannot? In particular, if we were to compare forecast errors for the two step horizon against confidence bands that would be based on simulating history, how informative about breaks would the unconditional and conditional forecasts be? To assess this question, it is To simplify expressions, suppose that the AR coefficient for y is 0: $a = \hat{a} = 0$. Also suppose c is known, such that its estimated value is replaced by its true value: $\hat{c} = c$. With these simplifications, the unconditional and conditional forecast errors become:

$$\begin{aligned}
\hat{u}^u_{t+2} &= u_{t+2} + b v_{t+1} + c(b - \hat{b}) x_t \\
\hat{u}^c_{t+2} &= u_{t+2} + b v_{t+1} + (b - \hat{b}) c x_t - \hat{b} v_{t+1} - \hat{\rho} v_{t+2} = u_{t+2} + (b - \hat{b}) c x_t + (b - \hat{b}) v_{t+1} - \hat{\rho} v_{t+2}.
\end{aligned}$$

Further suppose that the estimation sample is large enough to permit treating parameter estimation uncertainty as small. Now consider forming confidence bands around the unconditional and conditional forecasts. As long as we had a fairly long estimation sample, b (the pre-break value) would be estimated fairly precisely. In the case of the unconditional forecast, the confidence band around the forecast error would be driven by $\text{var}(\hat{u}^u_{t+2}) \approx \sigma^2_u + b^2\sigma^2_v$ $= 1 + b^2$. In the case of the conditional forecast, we give the forecast even more information about future x, reducing the estimated forecast uncertainty. Based on pre-break history and a fairly long estimation sample, both b and ρ (the pre-break values) would be estimated fairly precisely. Our confidence band around the forecast error would be driven by $\text{var}(\hat{u}^c_{t+2}) \approx \sigma^2_u + \hat{\rho}^2\sigma^2_v - 2\hat{\rho}\rho \approx 1 - \rho^2$. This forecast error band will be the tighter than the unconditional forecast error band.

This analysis indicates that by conditioning, we are tightening the historical confidence band around the forecast, making breaks easier to see. The conditional forecast will make breaks in b more evident than in the unconditional case. The conditional forecast will also likely reveal shifts in ρ, which the unconditional forecast will not. On the other hand, when b is stable, the conditional forecast is less likely than the unconditional to be pushed outside historical norms by an unusually big shock to x_{t+1}, because the associated coefficient on v_{t+1} is $(b - \hat{b})$. It may also be less sensitive to big shocks to v_{t+2}. A big shock to v_{t+2} will be associated with a big shock to u_{t+2}, for which the conditional error will make some correction, reflected in the subtraction of $\hat{\rho} v_{t+2}$.

B Robustness checks of conditional forecasts

We have verified the robustness of the conditional forecast results presented above in a range of other model specifications. We summarize our robustness checks, as follows.

- To ensure robustness in larger VAR models specified in the levels and log levels formulation that some researchers prefer (e.g., Sims and Zha (1998), Giannone, et al. (2012)), we also produced conditional forecasts for the 13 variable model described in section 2. These forecasts yield results very similar to those from the baseline specification.

- To ensure that the instability that seems to follow the 2007-2009 recession is not due to previous breaks in coefficients, we also produced conditional forecasts from the four-variable VAR with time-varying parameters and stochastic volatility (estimated through 2007:Q4). Having TVP improves the conditional forecasts of inflation and the funds rate, but still provides forecasts of the unemployment rate that are well below the actual unemployment rate.

- We also generated conditional forecasts from a VAR with a steady state prior (see, e.g., Clark (2011) and Villani (2009)), using the same variables as in the baseline model. This specification yielded results consistent with the baseline.

- As the business cycle indicator or factor, with the baseline model, we treated the unemployment rate (instead of GDP growth) as the business cycle factor and produced forecasts of GDP growth, inflation, and the interest rate conditioning on the actual unemployment rate path. Again, the results point to significant model instabilities. In this case, there is less evidence of a break in the relationship between unemployment and GDP growth: the path of actual GDP growth mostly falls within the conditional forecast band. However, the forecasts of inflation and the funds rate fall further outside their conditional forecast bands than in the case of forecasts conditioned on the path of GDP growth.

- As the business cycle indicator or factor, we modified the baseline model by replacing GDP growth with the Chicago Fed's index of national economic activity and generated forecasts conditional on the activity index. In this case, the actual paths of unemployment were slightly more consistent with the conditional forecast paths, most noticeably with the model estimated for 1985-2007. This finding, too, is suggestive of some break in the business cycle dynamics of the labor market, particularly since labor market indicators have a very large weight in the activity index.

- Motivated by other work on changes in labor turnover (e.g., Daly et al. (2012), Erceg and Levin (2013), and Van Zandweghe (2012)), to further assess whether there is some simple explanation for what looks like a shift in the relationship between unemployment and GDP growth we added the labor force participation rate to the baseline model and produced forecasts conditional on both GDP growth and the participation rate. This exercise yields results very similar to the baseline results presented above.

- For similar reasons, we added the job finding rate to the baseline model and produced forecasts conditional on both GDP growth and the job finding rate. This exercise yields results very similar to the baseline results presented above.

- In light of other work that has distinguished long-term and short-term unemployment, we considered a model including rates of both short-term (26 weeks or less) and long-term unemployment (27 weeks or more). In this case, results are qualitatively similar to those of the baseline case, with the bulk of the unusual behavior of the overall unemployment rate seemingly associated with unusual behavior of the long-term unemployment rate.

- In light of the potential for the zero lower bound on the federal funds rate to create instabilities, we produced conditional forecasts (conditioned on the actual path of GDP growth) from a model in which we spliced the federal funds rate to the shadow rate estimate of Xia and Wu (2013). Their shadow rate estimate is obtained from a term structure model that allows for an unobserved shadow short rate that is zero, even when actual short-term rates are constrained to be at least 0. We replace the funds rate with the shadow rate for the portion of the 2009-2012 period in which the shadow rate is less than zero. In our VAR specifications, using the shadow rate in lieu of the federal funds rate does not reduce the evidence of instabilities, consistent with the findings of Francis, Jackson, and Owyang (2014). We obtained results very similar to those in the baseline specification.

- In light of the debate that fiscal policy has been less expansive in the recent crisis compared to previous recessions, we considered a model including a simple fiscal policy measure, the ratio of the budget deficit to GDP. We have produced forecasts conditional on just GDP and conditional on both GDP and the deficit to GDP ratio. These forecasts yield results very similar to those from the baseline specification.

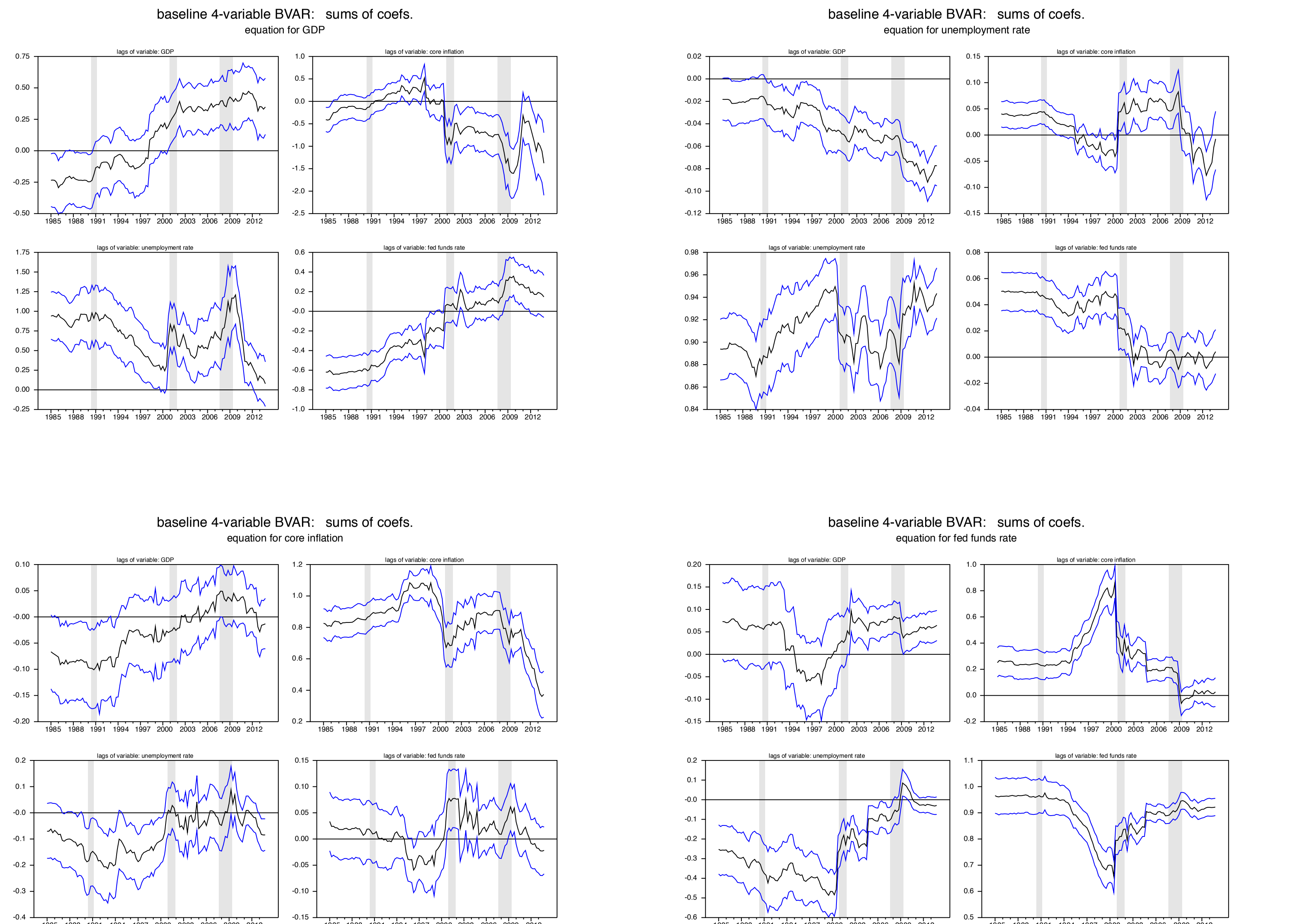


Figure 1: Estimates of 4-variable BVAR coefficients obtained with rolling sample of 80 observations (solid line = posterior median, dotted = 70% posterior credible set)

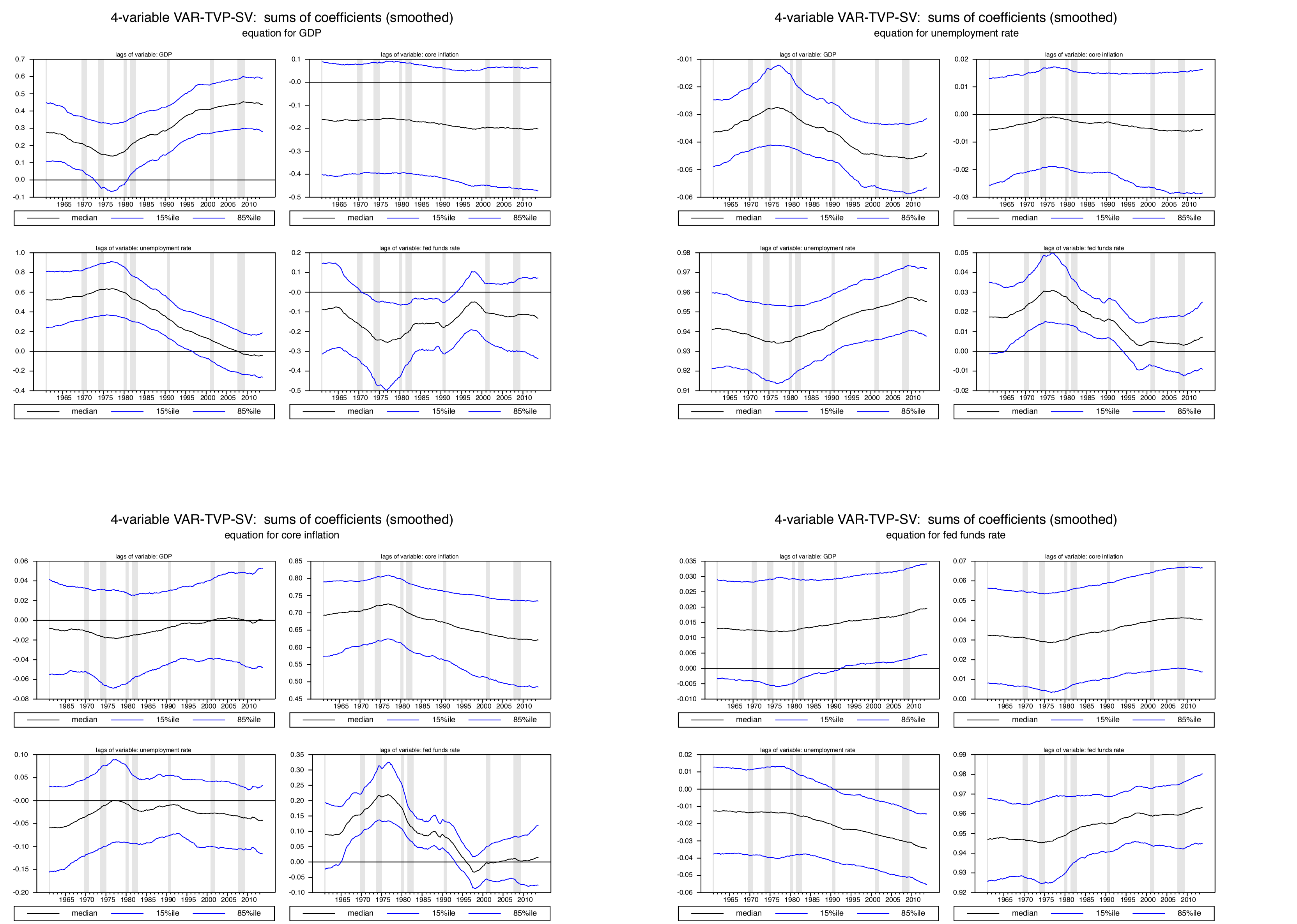


Figure 2: Estimates of 4-variable VAR-TVP SV coefficients (solid line = posterior median, dotted = 70% posterior credible set)

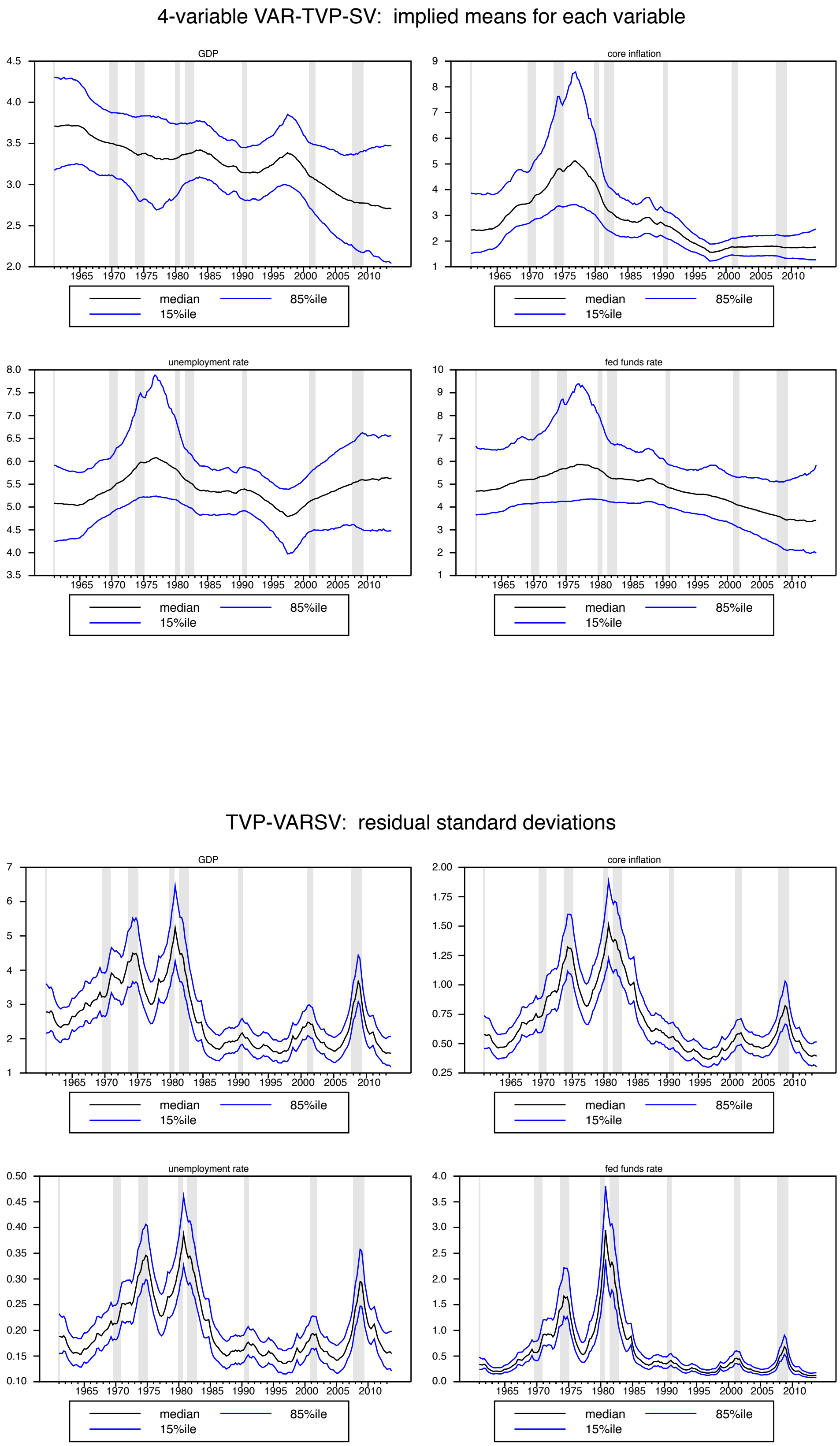
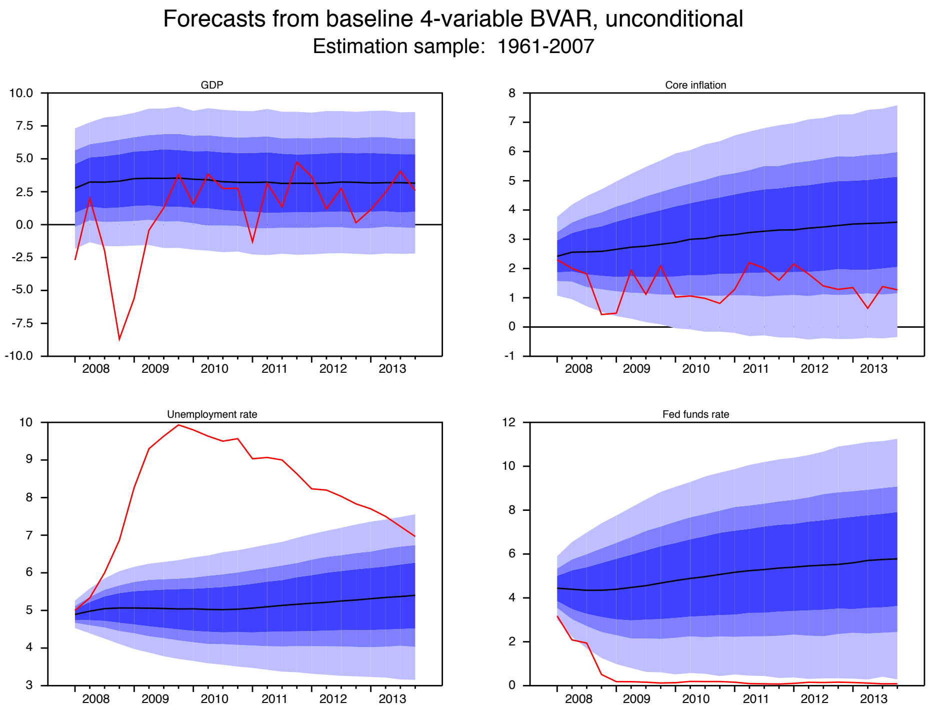
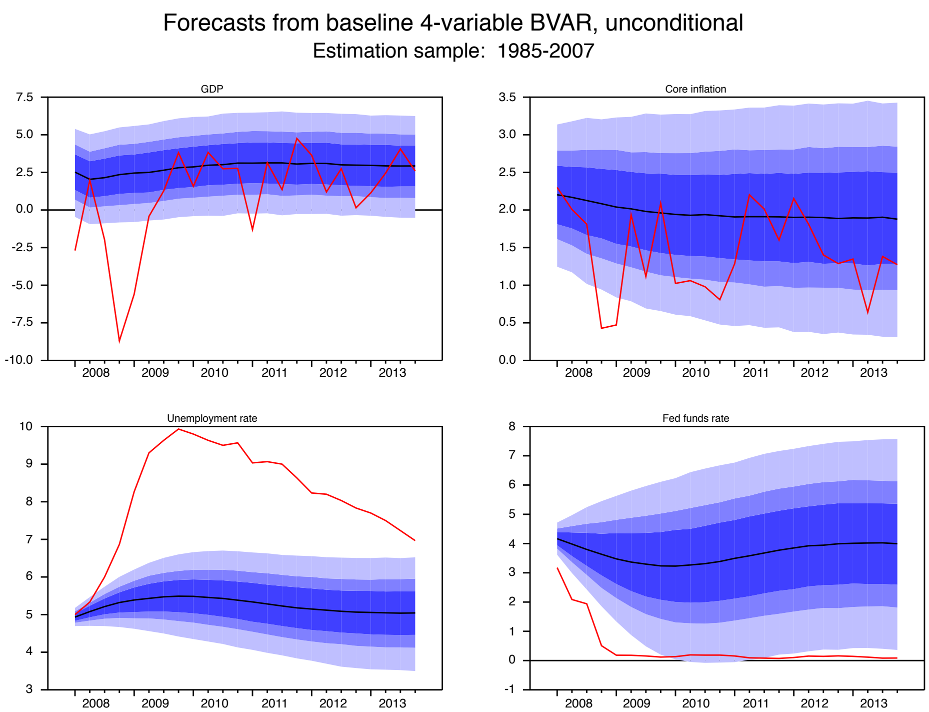


Figure 3: Estimates of implied means and residual volatilities from 4-variable VAR-TVP-SV model

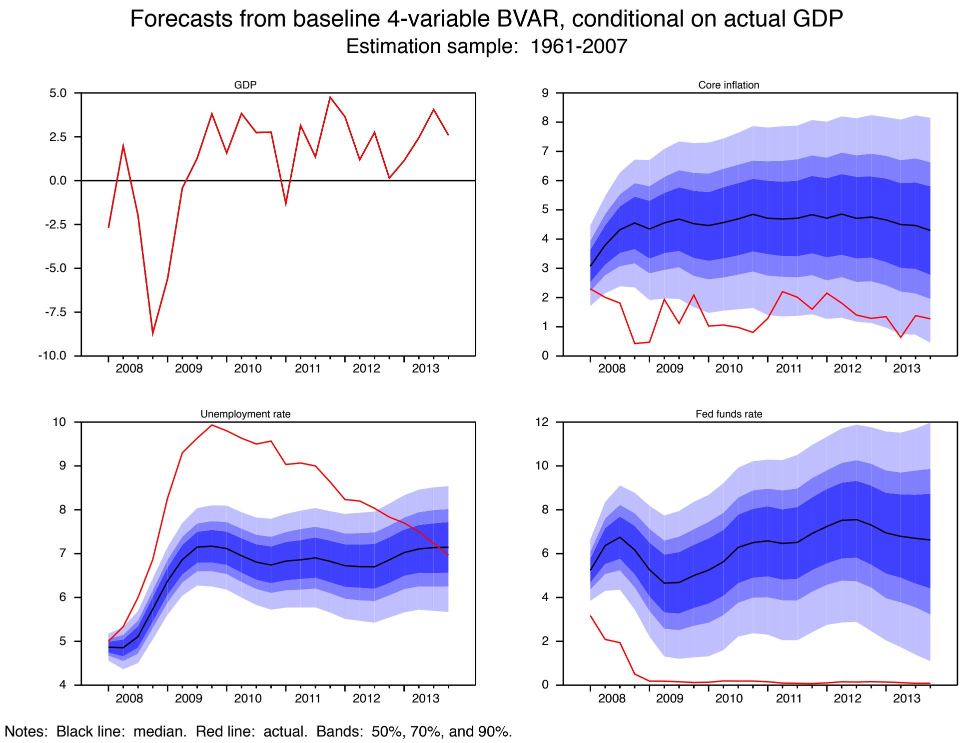


Notes: Black line: median. Red line: actual. Bands: 50%, 70%, and 90%.



Notes: Black line: median. Red line: actual. Bands: 50%, 70%, and 90%.

Figure 4: 2008-2013 unconditional forecasts from 4-variable BVAR, estimated with 1961-2007 and 1985-2007 samples



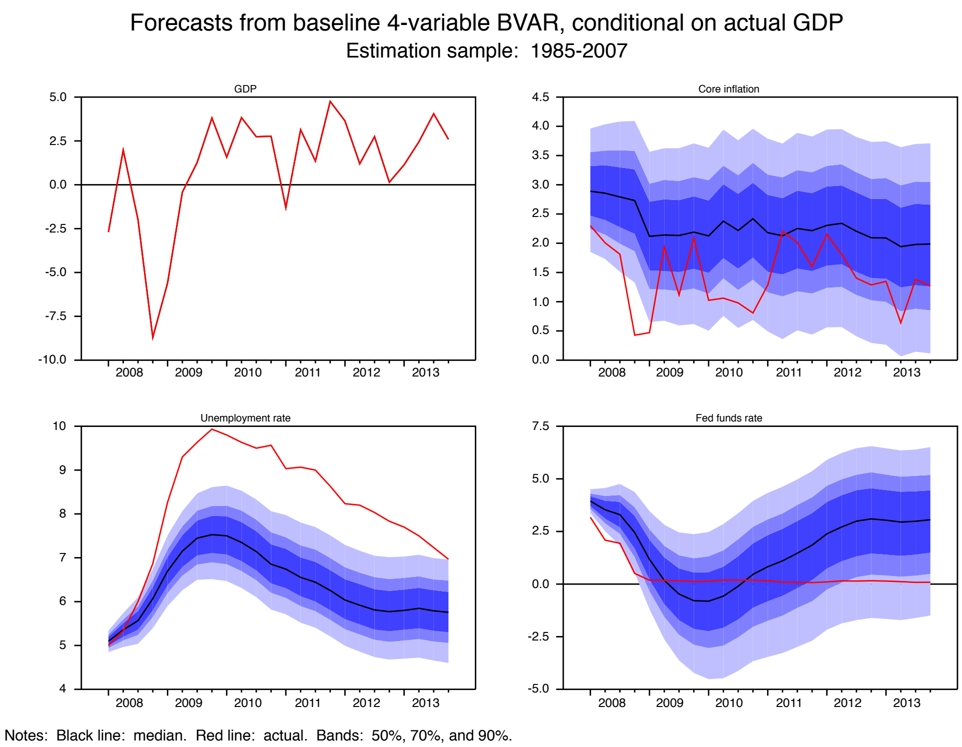
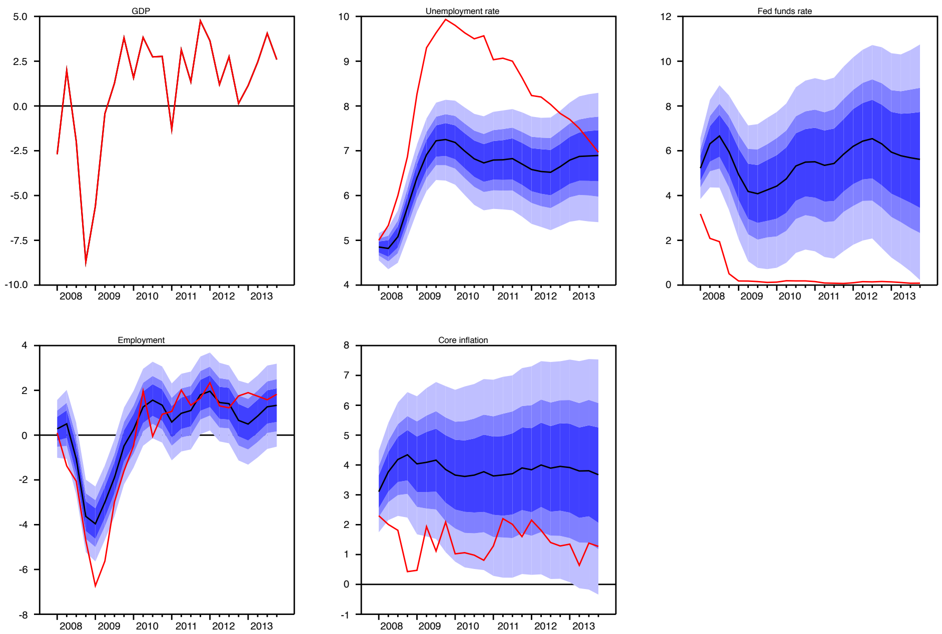


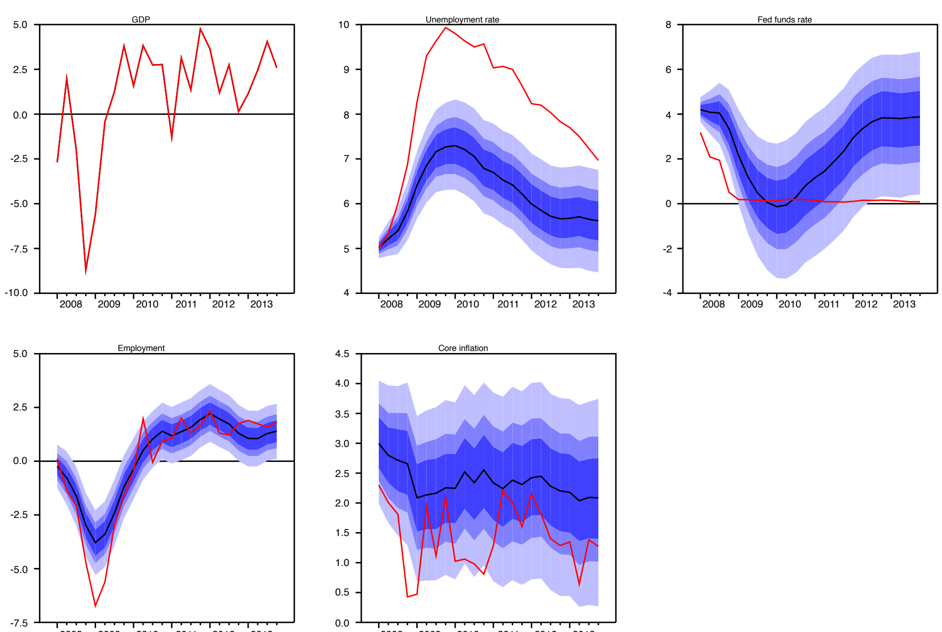
Figure 5: 2008-2013 conditional forecasts from 4-variable BVAR, estimated with 1961-2007 and 1985-2007 samples

Forecasts from 5-variable BVAR with employment, conditional on actual GDP
Estimation sample: 1961-2007



Notes: Black line: median. Red line: actual. Bands: 50%, 70%, and 90%.

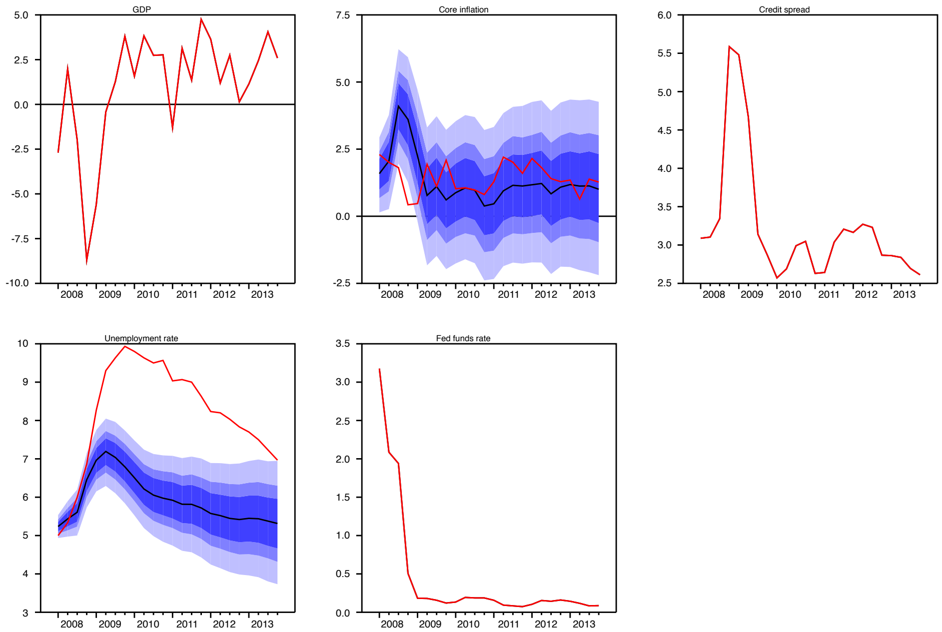
Forecasts from 5-variable BVAR with employment, conditional on actual GDP
Estimation sample: 1985-2007



Notes: Black line: median. Red line: actual. Bands: 50%, 70%, and 90%.

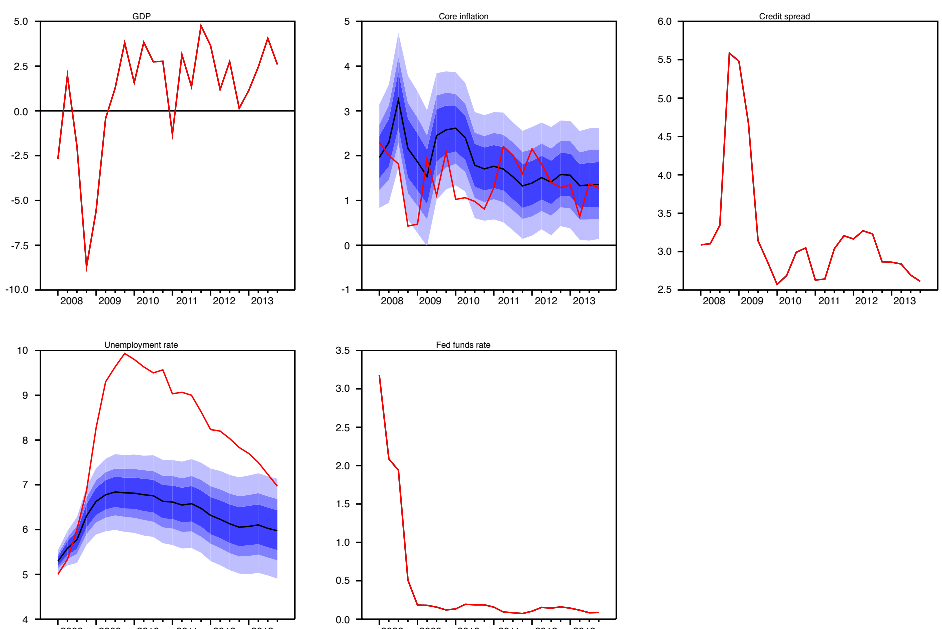
Figure 6: 2008-2013 conditional forecasts from 5-variable BVAR with employment, estimated with 1961-2007 and 1985-2007 samples

Forecasts from 5-variable BVAR with credit spread, condit. on actual GDP, FFR, and credit spread
Estimation sample: 1961-2007



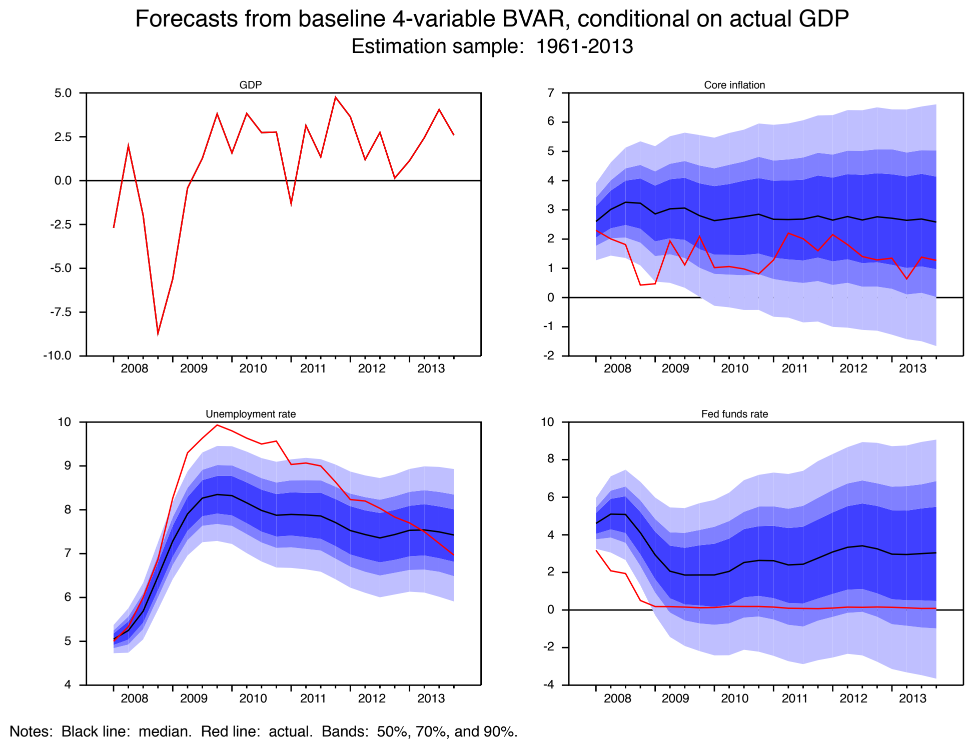
Notes: Black line: median. Red line: actual. Bands: 50%, 70%, and 90%.

Forecasts from 5-variable BVAR with credit spread, condit. on actual GDP, FFR, and credit spread
Estimation sample: 1985-2007



Notes: Black line: median. Red line: actual. Bands: 50%, 70%, and 90%.

Figure 7: 2008-2013 conditional forecasts from 5-variable BVAR with credit spread, estimated with 1961-2007 and 1985-2007 samples



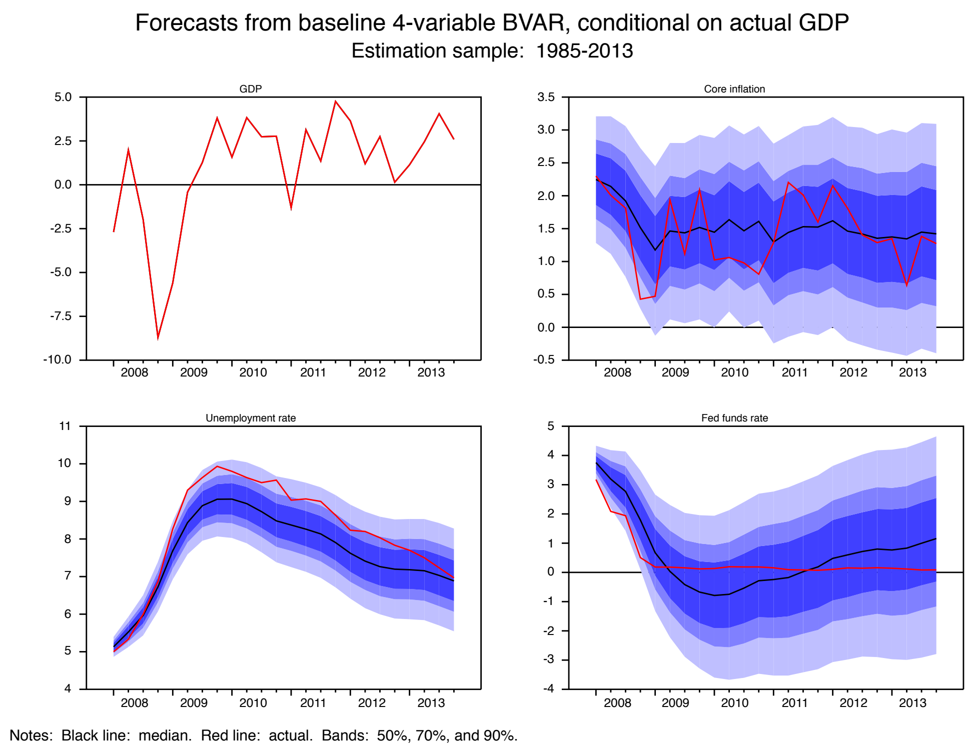
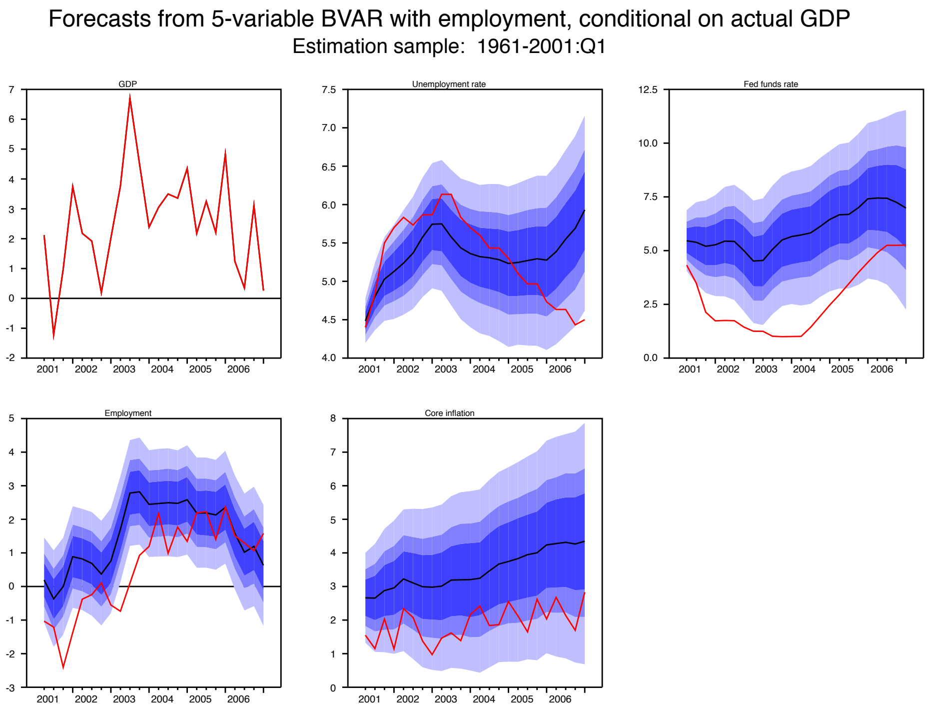
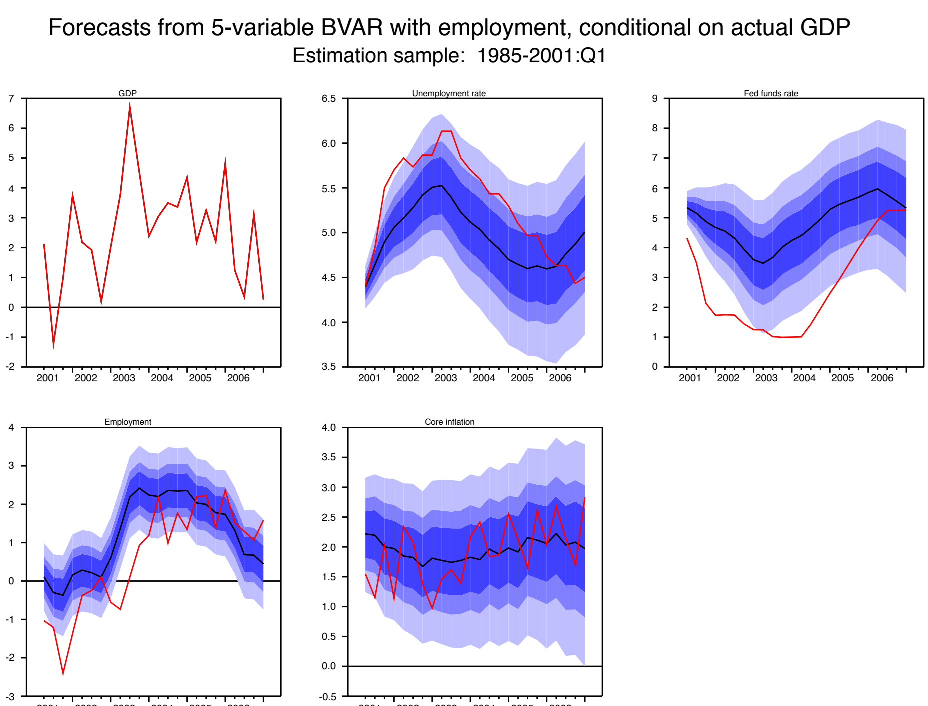


Figure 8: 2008-2013 conditional forecasts from 4-variable BVAR, estimated with 1961-2013 and 1985-2013 samples

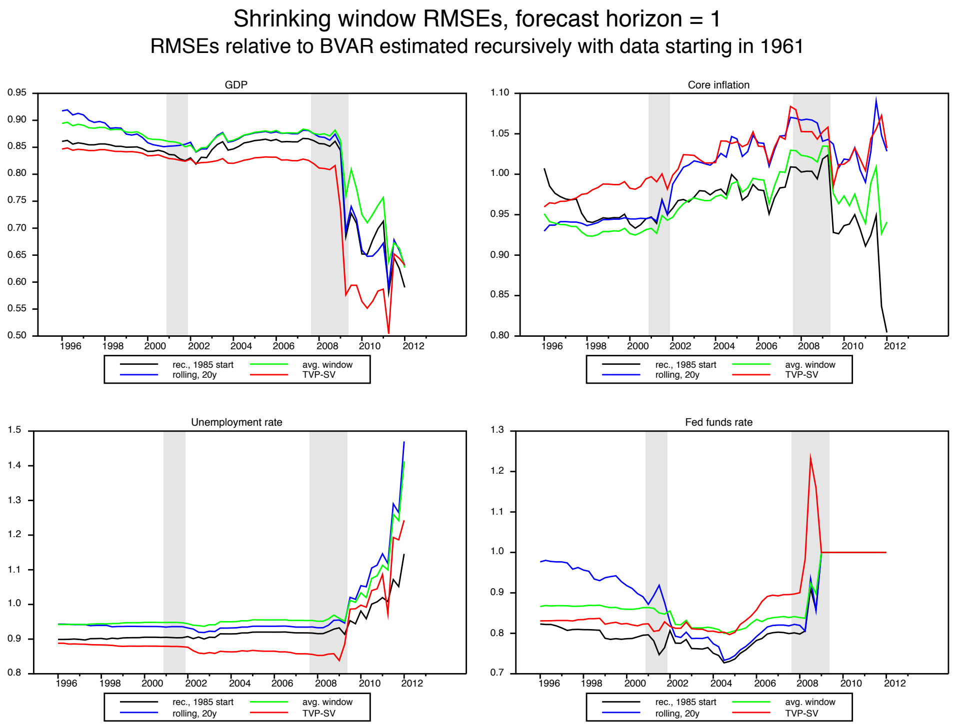


Notes: Black line: median. Red line: actual. Bands: 50%, 70%, and 90%.



Notes: Black line: median. Red line: actual. Bands: 50%, 70%, and 90%.

Figure 9: 2001-2005 conditional forecasts from 5-variable BVAR, estimated with 1961-2000:Q1 and 1985-2000:Q1 samples



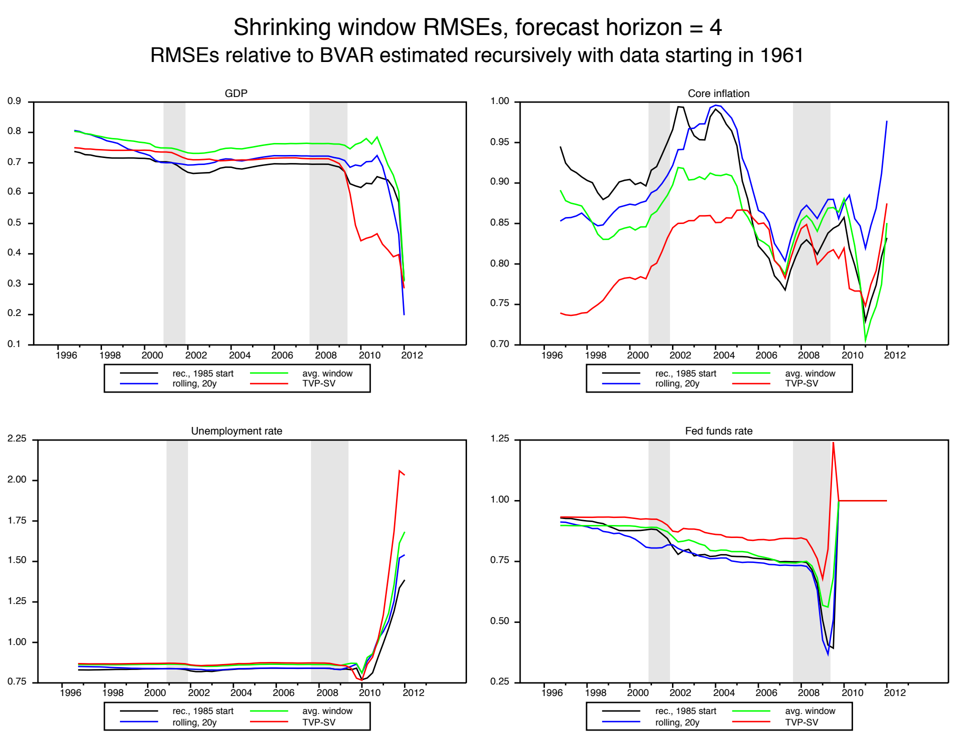
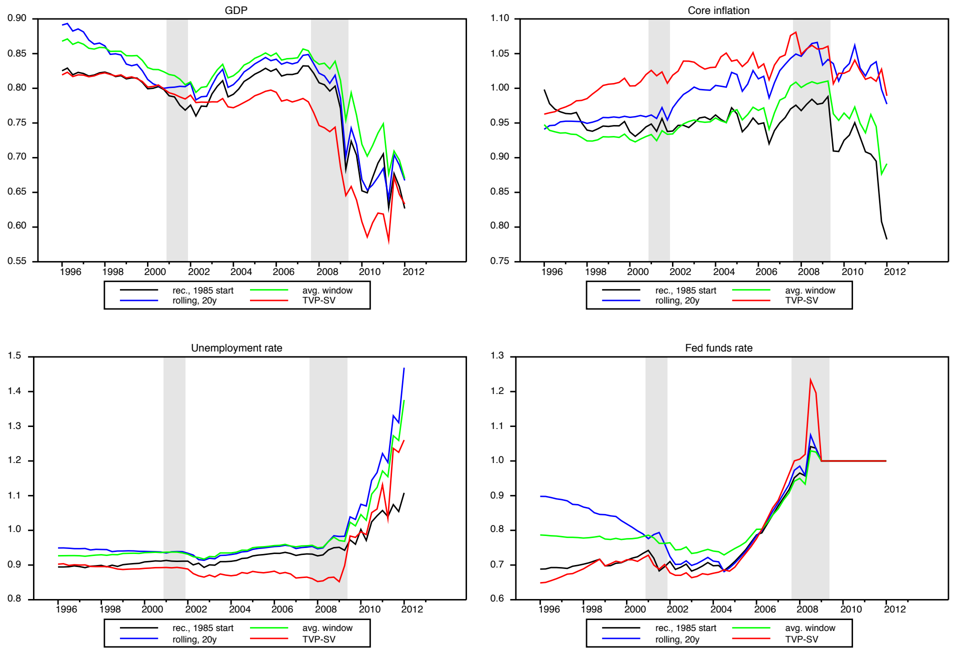


Figure 10: Shrinking window time series of RMSE ratios from 4-variable BVARs, real-time data

Shrinking window CRPSs, forecast horizon = 1
CRPSs relative to BVAR estimated recursively with data starting in 1961



Shrinking window CRPSs, forecast horizon = 4
CRPSs relative to BVAR estimated recursively with data starting in 1961

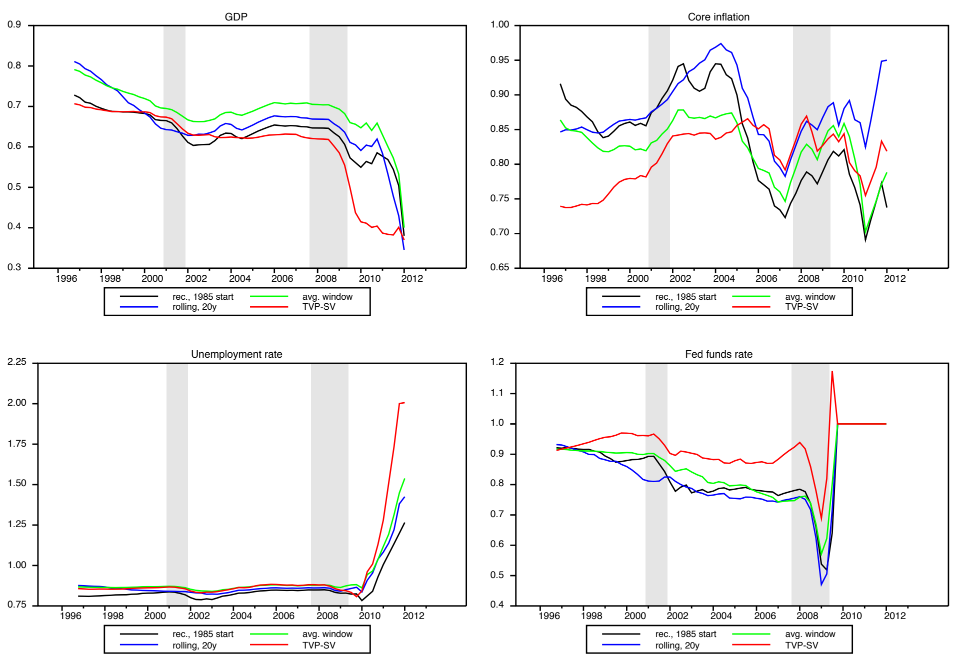
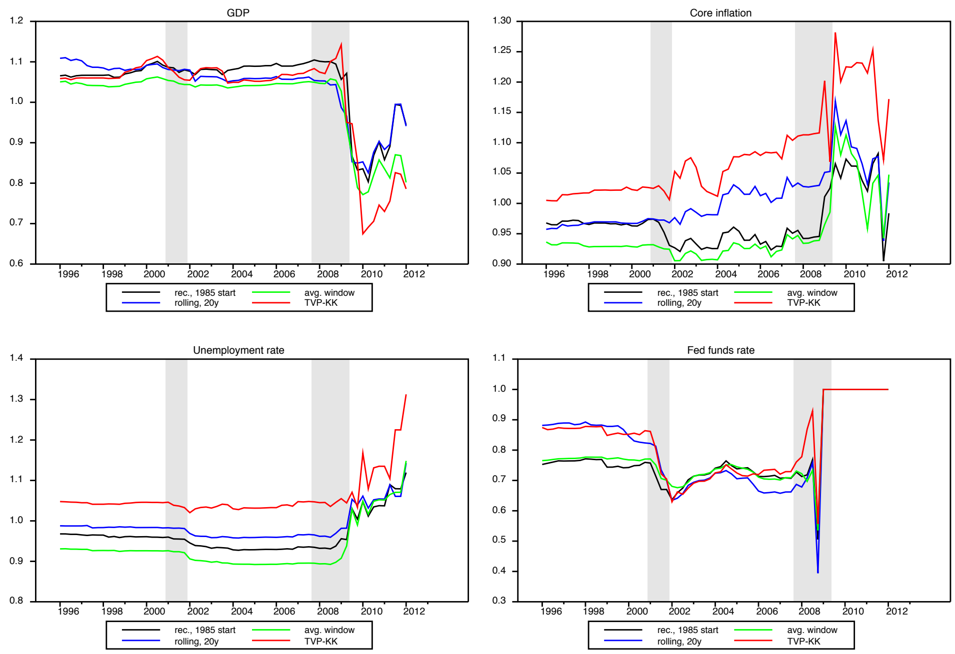


Figure 11: Shrinking window time series of CRPS ratios from 4-variable BVARs, real-time data

Shrinking window RMSEs, forecast horizon = 1
RMSEs relative to BVAR estimated recursively with data starting in 1961



Shrinking window RMSEs, forecast horizon = 4
RMSEs relative to BVAR estimated recursively with data starting in 1961

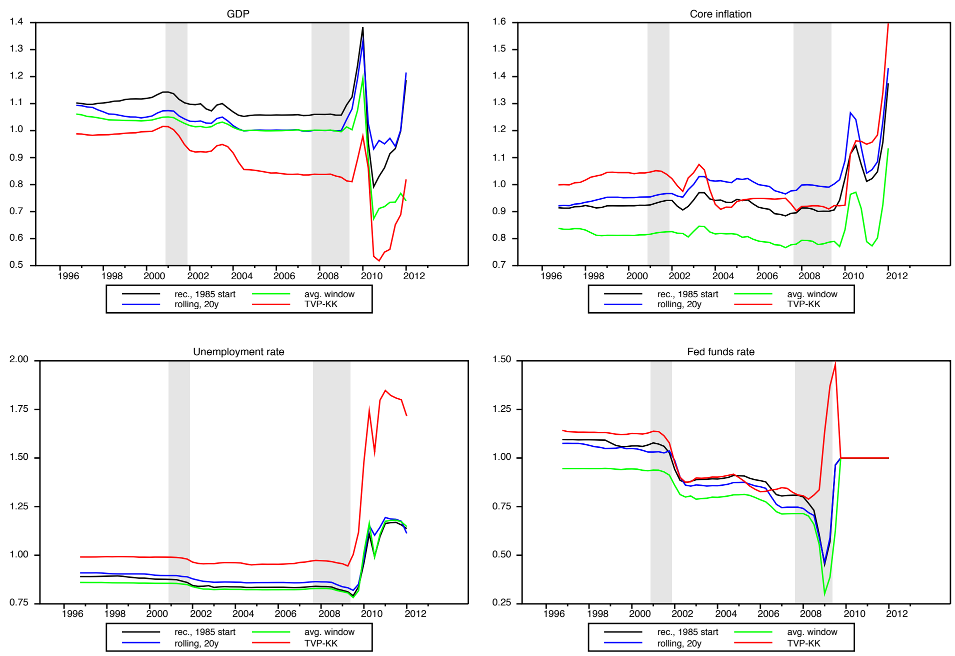


Figure 12: Shrinking window time series of RMSE ratios from 13-variable BVARs, real-time data

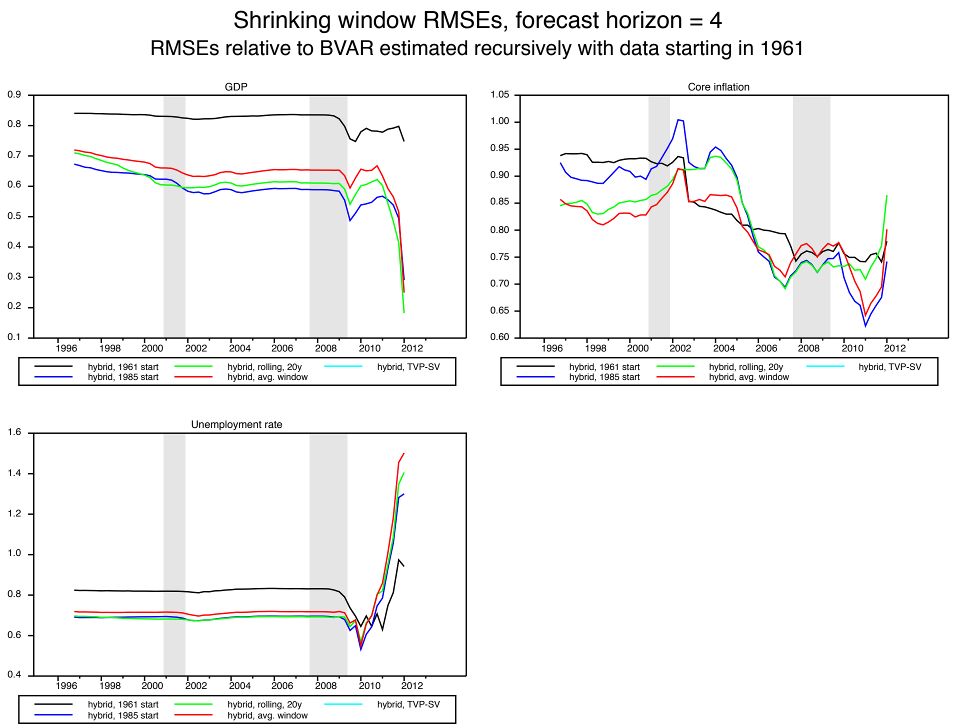


Figure 13: Shrinking window time series of RMSE ratios from 4-variable hybrid BVARs, real-time data