# Forecasting the Great Recession in the United States: First Results from a Model Comparison Exercise


Michael Binder*
IMFS and Goethe-University Frankfurt

Mátyás Farkas†
European Central Bank

Zexi Sun‡
IMFS and Goethe-University Frankfurt

John B. Taylor§
Hoover Institution, Stanford University

Volker Wieland¶
IMFS and Goethe-University Frankfurt

Maik H. Wolters‖
University of Würzburg, Kiel Institute for the World Economy and IMFS


June 10, 2020

*Preliminary - Do not cite*


**Abstract**

Macroeconomists have been criticised for failing to predict the massive macroeconomic effects of the global financial crisis. As part of this criticism, the usage of DSGE models without financial frictions has been questioned. In response, during the last decade a good many new models with financial frictions have been developed. In this paper, we compare forecasts of the Great Recession based on a range of post-crisis NK-DSGE models embedding financial frictions with forecasts from professional forecasters as well as forecasts based inter alia on NK-DSGE models developed prior to the crisis, a Cowles Commission model and Bayesian VARs. A forecasting experiment based on recursive estimation using real-time data vintages provides evidence that NK-DSGE models embedding a financial accelerator and information provided by higher-frequency data, specifically on credit spreads, produce high-quality GDP nowcasts at the onset of the Great Recession. These models can also detect the beginning of this recession earlier than pre-crisis models as well as post-crisis models that do not embed the same higher-frequency information. Furthermore, forecasts from the pre-crisis models and those from professional forecasters tend to strongly underpredict the extent of the Great Recession. The post-crisis NK-DSGE models that make use of higher-frequency credit-spread information in addition yield better forecasts than similarly informed unrestricted Bayesian VARs. Nonetheless, like the professional forecasters, not even these post-crisis models succeed in predicting the Great Recession prior to its onset.



*Keywords:* Global Financial Crisis, Great Recession, Forecasting, Model Uncertainty, Financial Frictions, Real-Time Data

*JEL-Codes:* C53, E32, E37, E44



*mbinder@wiwi.uni-frankfurt.de
†matyi.farkas@gmail.com
‡zsun@wiwi.uni-frankfurt.de
§johntayl@stanford.edu
¶wieland@wiwi.uni-frankfurt.de
‖maik.wolters@uni-wuerzburg.de

# 1 Introduction

Macroeconomists have been criticised for failing to predict the Great Recession of 2008 and 2009 or at least failing to provide adequate warning that financial disruptions could trigger such a massive contraction. This in turn led to a wave of criticism of the state of macroeconomic modeling and forecasting, though forecasters have long been aware that the timing of recessions is rather difficult to predict (for recent work on this, see, e.g., Dovern and Jannsen, 2017). Nonetheless, the fact that well into the fall of 2008 adequate warnings about a looming severe recession were generally lacking provided evidence that pre-crisis assessments that "the state of macro is good" (Blanchard, 2009) were too general. A good bit of the post-crisis criticism has been directed at the Dynamic Stochastic General Equilibrium (DSGE) modeling paradigm and it has been prominently argued that macroeconomists in academia and at policy institutions rely too heavily on this modeling paradigm (see, e.g., Buiter, 2009; Krugman, 2009; Stiglitz, 2015; Romer, 2016).

In an earlier model forecast comparison study (Wieland and Wolters, 2011, 2012), we assessed whether this criticism was warranted and studied the forecasting performance of a variety of DSGE models during the time period of the Great Recession. We adduced evidence that DSGE models that were developed prior to the global financial crisis indeed could not forecast the Great Recession. However, a comparison with forecasts from the Survey of Professional Forecasters showed that professional forecasters also failed to predict the Great Recession. Given this failure, the widespread criticism of the state of macroeconomic forecasting applies to macroeconomic models originating in academic-oriented research as well as to professional forecasters, and in part reflects the general difficulty of predicting (the timing of) recessions. Nevertheless, the pre-crisis generation of DSGE models usually abstracts from financial market frictions, so that these models clearly are not suitable for predicting a recession driven by a financial crisis.

Over the last decade, many new structural macroeconomic models have been developed in response to the global financial crisis and its aftermath. Many of these models incorporate financial market frictions that can generate sizable declines of output in response to adverse financial market shocks, thereby addressing some of the earlier criticism. Some of these new models are now regularly used by central banks and other policy institutions. It is thus a good time to assess what progress macroeconomists have made in forecasting. To this end, we aim to examine how propagation mechanisms in models built after the global financial crisis differ from those in models developed prior to the crisis, and in particular to which extent this has changed the profession's ability to predict recessions and overall business cycle dynamics, in absolute terns and relative to the pre-crisis models. In considering a range of model structures, our approach is related to our earlier work in which we proposed a comparative approach to macroeconomic policy analysis that is open to competing modeling paradigms (Wieland et al., 2012; Wieland and Taylor, 2012).

To be more specific, among post-crisis models we consider a variety of models with different types of financial frictions, different observables and different information sets to forecast the Great Recession. In particular, we use small- and medium-scale New Keynesian (NK) - DSGE models including the financial accelerator of Bernanke et al. (1999) and/or allowing for frictions in the housing market, following Iacoviello (2005). To check whether the usage of DSGE models in general may be problematic, we also consider alternatives. As a first alternative, we include the IMF's small quarterly projection model (Carabenciov et al., 2008), that has no model-internal microeconomic foundations. Compared to DSGE models, the model specification aims to take advantage of the resultant higher degree of flexibility as to how to include forward- and backward-looking variables in the structural equations, and the model also allows for relatively flexible changes of the long-run trends. In regards to macro-financial linkages, the model in its IS-equation involves a bank-lending-tightness shock that has a direct effect on output. As a second alternative, we also include forecasts from a traditional Cowles Commission-type model, specifically the model by Fair (2018), to probe whether the use of more traditional Keynesian models results in improved crisis forecasts, as has been conjectured by Krugman (2009) and Buiter (2009). We do not re-estimate this model ourselves, but rely on the forecasts computed by Ray Fair in real time throughout the time period of the global financial crisis, based on the then-applicable model vintages. As a measure of the contributions of financial frictions in the NK-DSGE models, we also include small and medium-scale pre-crisis NK-DSGE models without financial frictions in our model comparison set. To disentangle to which extent the micro-theoretical foundations of the NK-DSGE models strengthen their forecasting performance, we furthermore compute forecasts based on reduced-form Bayesian VARs. For these VAR models, we consider specifications including measures of financial market distress, and specifications without such measures, yielding models that involve the same range of observables as the structural models do.

In part to allow a direct comparison to the forecasts from the Survey of Professional Forecasters and the Greenbook projections, we estimate all models recursively and based on real-time data vintages. In addition, we consider different specifications of the information set that can be employed by the forecaster: For the baseline information set, we use a balanced set of observations extending to one quarter before the quarter of forecast computation. This accounts for publication lags for macroeconomic variables as is experienced by the professional forecasters. To partially reflect that professional forecasters can also make use of observables that are not part of our model specifications, we consider two types of augmentation of this baseline information set: (i) augmentations involving higher-frequency financial market data as well as nowcasts of macroeconomic variables based on other higher-frequency data, all as actually available at the time of forecast computation, and (ii) augmentations involving nowcasts of macroeconomic variables based on the Survey of Professional Forecasters, again as actually available at the time of forecast computation.

Our perhaps most important finding is that NK-DSGE models with financial market frictions and employing also higher-frequency data, particularly data on financial market distress, would have been able to accurately predict the extent of the Great Recession at its onset. In particular, a medium-scale NK-DSGE model with financial accelerator for which also higher-frequency data, specifically on credit spreads, as available at the time of forecast computation is used, can predict the largest drop of GDP within the Great Recession in the same quarter as it actually occurs, and thus an accurate nowcast of output growth for the fourth quarter of 2008 can be obtained. A comparison with forecasts from the Survey of Professional Forecasters and the Greenbook projections shows that this model-based prediction would have been a more timely prediction of the extent of the Great Recession than the predictions actually made by the professional forecasters. In this sense, then, there has been quite remarkable progress in structural macroeconomic modeling, keeping in mind also that corresponding models as developed prior to the global financial crisis were not able to predict any downturn in 2008 or 2009.

To be clear, our results also reveal that it is not sufficient to include financial market frictions in NK-DSGE models to obtain accurate forecasts during the time period of the Great Recession, but that the choice of observables used to inform the model about financial market distress is important, too: NK-DSGE models involving credit-spread data use the sudden increase in spreads towards the end of 2008 to deliver accurate short-term forecasts. Other post-crisis NK-DSGE models that focus on housing-market collateral constraints and are fed with higher-frequency data on the mortgage and housing market yield forecasts that do not anticipate the depth of the Great Recession at its onset, as is true for the pre-crisis NK-DSGE models. This forecasting performance of the models with housing-market collateral constraints is due to the fact that in the relevant aggregate housing market data no jumps of similar magnitude occurred as did occur for the credit spreads.

Another important aspect of our results is that the precise timing of the data included in the information set on the basis of which the forecasts are computed can be crucial. A standard approach for NK-DSGE models estimated on quarterly data is to use a balanced set of observations extending to one quarter before the quarter within which the forecast computation occurs. This puts the models at a disadvantage compared to professional forecasters, as the latter can use within-quarter information derived from higher-frequency data also. For example, for a forecast starting in November 2008, a quarterly model would usually only include data up to the third quarter of 2008, when the widening of credit spreads was still modest. When including fourth-quarter credit-spread data that are available at the time of forecast computation and estimating missing observations of the macroeconomic observables for the fourth quarter via a state-space model based Kalman filter, the forecast accuracy improves substantially.

A comparison to the forecasts based on Bayesian VARs shows that the parameter restrictions that are part of the NK-DSGE models are helpful in generating more accurate forecasts. The forecasts from these models under different metrics perform notably better than those from Bayesian VARs estimated on the same data series.

Finally, the use of a traditional Cowles Commission-type model would not have resulted in better forecasts compared to those from pre-crisis NK-DSGE models or those from professional forecasters. The Cowles Commission-type model's forecasts are similar to those from the pre-crisis NK-DSGE models. This is not surprising as the model that we employ does not include any channels that can capture the distress in credit markets that was observed in 2008 and 2009. This does not mean that Cowles Commission-type models may not be generally useful for macroeconomic forecasting, but that for purposes of predicting the Great Recession they need to be augmented with frictions capturing credit distress effects — just as the pre-crisis NK-DSGE models.

While the literature on the forecast evaluation of macroeconomic models with financial frictions is very small, there are a few papers closely related to our work. In most of these a standard medium-sized NK-DSGE model is augmented with the financial accelerator mechanism. Del Negro and Schorfheide (2013) and Del Negro et al. (2015) show that once forecasts are conditioned on short-term interest-rate and credit-spread data for the current quarter, such a model can predict a sizable downturn of output growth in the fourth quarter of 2008, the forecast being similar to the Blue Chip for that quarter. These improvements are, however, restricted to crisis periods, while in non-crisis times a model without financial frictions performs better. Using a similar model, Kolasa and Rubaszek (2015) find an improvement in output growth forecasts during the Great Recession only for medium-term forecasts, while the short-term forecasts are worse than those from a model without financial frictions. They also consider a model specification in which financial frictions in the housing sector are captured and higher-frequency observables capturing mortgage market information and housing prices are added. This model yields substantially better output growth forecasts for all forecast horizons during the Great Recession, but forecasts that are substantially worse outside time periods of financial crises. Finally, Christiano et al. (2011) find no improvement in the accuracy of output growth forecasts when augmenting a model for the Swedish economy with the financial accelerator based on an evaluation sample from 2005 to 2010. Forecasting accuracy improves for inflation and interest rates, though they do not report whether these improvements are statistically significant.

Overall, these papers indicate that whether the inclusion of financial market frictions improves forecast accuracy of macroeconomic models depends on the type of financial friction, the variables to be predicted, the timing of the evaluation sample (in particular, whether it is restricted to deep crisis periods or includes other recessions as well), the specification of those observables that inform the model about financial market distress, and the specification of the forecaster's information set (in particular, whether this includes higher-frequency measures of financial market distress). The existing studies yield seemingly contradictory results, however, and do not disentangle which of the various elements entering the forecast specification are crucial to yield accurate forecasts. Our approach employs a variety of models, different sets of observables and entails a comparison as to how differing specifications of the information set affect the forecast accuracy. Based on such a rather broad comparative approach, we can systematically evaluate which of the elements of the forecast specification are important so as to achieve (more) accurate forecasts for a range of evaluation samples. Further, in contrast to the earlier papers, all of the estimation is based on real-time data vintages, which allows a direct comparison to professional and academic forecasts that were published before and during the global financial crisis.

The remainder of the paper is organized as follows. In Section 2 we describe the various models, Section 3 explains the set-up of the forecasting experiment and in Section 4 we present and discuss our results. Finally, Section 5 provides concluding remarks.

## 2 Forecasting Models

We consider standard small and medium scale DSGE models with and without financial frictions, a more flexible structural model without microeconomic foundations, several Bayesian VARs and a traditional Cowles Commission type model. Except for the latter, similar versions of the models considered here are regularly used by central banks or international policy institutions. Table 1 summarizes the most important model features of the models considered in this paper.

Table 1: Model Overview

| Name/Reference | Short Name | Description | Observable Variables |
|---|---|---|---|
| | | Pre-crisis models | |
| Del Negro and Schorfheide (2004) | DS04 | standard 3-equation New Keynesian model with forward looking IS- and Phillips curve with government spending, technology and monetary policy shocks | 3: output growth, inflation, interest rate |
| Wieland and Wolters (2011) | WW11 | standard 3-equation New Keynesian model with forward looking IS- and Phillips curve with government spending, technology, monetary policy, preference and markup shocks | 3: output growth, inflation, interest rate |
| Smets and Wouters (2007) | SW07 | medium-scale DSGE models with many nominal and real frictions and 7 structural shocks | 7: output growth, consumption growth, investment growth, inflation, wages, hours, interest rate |
| Edge et al. (2008) | FRBEDO08 | medium-scale DSGE-model developed at he Federal Reserve. Two sectors with different technology growth rates, demand side disaggregated into different consumption and investment components | 11: output growth, inflation, interest rate, consumption of non-durables and services, consumption of durables, residential investment, business investment, hours, wages, inflation for consumer non-durables and services, inflation for consumer durables |
| Giannone et al. (2015) | BVAR3, BVAR7, BVAR11 | Bayesian VARs with optimal shrinkage prior estimated on the same observables as the above models | 3, 7 or 11 |
| Fair (2004) | Fair04 | large-scale Cowles-Commission type model with 25 stochastic equations + about 100 identities, large degree of disaggregation | more than 100 |
| | | Post-crisis models with Financial Frictions | |
| Bernanke et al. (1999) | NKBGG | small New Keynesian model with financial accelerator, estimated version of Bernanke et al. (1999) with small extensions | 5: output growth, inflation, interest rate, investment, credit spread (long-term financing) |
| Del Negro and Schorfheide (2013), Del Negro et al. (2015) | DNGS15 | medium-scale DSGE model similar to Smets and Wouters (2007) + financial accelerator | 8: output growth, consumption growth, investment growth, inflation, wages, hours, interest rate, credit spread (long-term financing) |
| Kolasa and Rubaszek (2015) | KR15_FF | medium-scale DSGE model (Del Negro et al., 2007) + financial accelerator | 9: output growth, consumption growth, investment growth, inflation, wages, hours, interest rate, credit spread (short-term financing), loan growth |
| Kolasa and Rubaszek (2015) | KR15_HH | medium-scale DSGE model (Del Negro et al., 2007) + collateral constraints in the housing sector | 11: output growth, consumption growth, investment growth, inflation, wages, hours, interest rate, residential investment, mortgage loans, house prices, mortgage loan spread |
| Carabenciov et al. (2008) | QPM08 | IMF Quarterly Projection model without microeconomic foundations, hybrid IS- and Phillips curve, flexible long-run equilibrium | 6: unemployment rate, output growth, inflation, interest rate, bank lending tightness |
| | | Bayesian VARs | |
| Giannone et al. (2015) | FF-BVAR5-CS, FF-BVAR5-BLT, FF-BVAR8, FF-BVAR9, FF-BVAR11 | Bayesian VARs with optimal shrinkage prior estimated on the same observables as the above models. There are two versions with 5 variables, one with credit spread (CS) and one with bank lending tightness (BLT) | 5, 8, 9 or 11 |

Regarding models that have been developed before the Global Financial Crisis, we use two small-scale and two medium-scale New Keynesian models and a Cowles Commission type model. The New Keynesian models are derived based on optimization problems of households and firms, incorporate nominal rigidities and are solved under the assumption of rational expectations. The small scale models include an IS-equation, a Phillips curve and a monetary policy rule and are estimated on three key variables: output growth, inflation and the federal funds rate. Specifically, we use the models by Del Negro and Schorfheide (2004) and Wieland and Wolters (2011). While the former includes a government spending, a technology and a monetary policy shock, the latter includes in addition shifts in preferences as a more general demand shock and a markup-shock. Otherwise the two models are very similar.

The two medium scale models include physical capital in the production function and account for endogenous capital formation. Labor supply is modeled explicitly. Nominal frictions include sticky prices and wages and price and wage indexation. Real frictions include consumption habit formation, investment adjustment costs and variable capital utilization. The first model is the one by Smets and Wouters (2007), probably the most well-known medium-scale DSGE model. It is estimated on seven observables—output, consumption, investment, inflation, the federal funds rate, hours and wages—and includes seven structural shocks. The second model is by Edge et al. (2008). Following these authors we refer to it as the FRB-EDO model reflecting that this is one of the models used at the Fed. It features two production sectors, which differ in their pace of technological progress. This structure can capture the different growth rates and relative prices observed in the data. Accordingly, the expenditure side is disaggregated as well. It is divided into business investment and three categories of household expenditure: consumption of non-durables and services, investment in durable goods and residential investment. The data used in estimation covers output growth, inflation, the federal funds rate, consumption of non-durables and services, consumption of durables, residential investment, business investment, hours, wages, inflation for consumer non-durables and services and inflation for consumer durables.

The Cowles Commission type model by Fair (2004) is based on economic theory including insights from solving intertemporal microeconomic optimization problems, but equations are not strictly derived from such microeconomic optimization problems and there is no rational expectations assumption. The model is larger than typical DSGE models. It consists of 25 stochastic equations and about 100 identities. The model is divided into six sectors (households, firms, financial, federal government, state and local government, foreign) and is estimated equation-by-equation using two-stage least squares. The model includes stock prices and house prices and their implied effects for household wealth and the effect of capital gains or losses on the accumulation of capital in the firm sector. Hence, the model already included before the Global Financial Crisis some macro-financial linkages, while pre-crisis DSGE models did not. However, the model does not include frictions in credit markets that are important for explaining the Global Financial Crisis. These are emphasized in the post-crisis DSGE models described next.

The post-crisis DSGE models capture principal-agent problems in the credit market. Many of them include the financial accelerator of Bernanke et al. (1999) that focuses on frictions in the financing of investment in firm's capital. Risk-neutral entrepreneurs manage the capital stock. In addition to using their private wealth, they borrow funds from households via a financial intermediary. The return to capital is subject to idiosyncratic shocks that can only be observed after the credit contract has been signed. As a result, the entrepreneurs' net worth determines the external finance premium. The external finance premium varies countercyclically as the net worth varies procyclically. For example, in a recession usually net worth falls so that the risk of default increases. Hence, the external finance premium increases and this increase in the cost of credit further decreases investment, which deepens the recession.

The original financial accelerator model by Bernanke et al. (1999) is based on a calibrated small-scale New Keynesian model. We estimate the model based on five time series: output, inflation, interest rate, investment and a credit spread measure. Therefore, we augment the original model that contains three shocks with two additional shocks: an investment specific technology shock and a risk premium shock. Further, we consider two medium-scale DSGE models with financial accelerator. The first one by Del Negro and Schorfheide (2013) and Del Negro et al. (2015) includes the financial accelerator into a slightly changed version of the Smets-Wouters model and adds a measure of the credit spread as an additional observable. The Federal Reserve Bank of New York uses this model and labels it as the FRBNY DSGE Model. The second model is based on Kolasa and Rubaszek (2015). They augment the model by Del Negro et al. (2007), which is very similar to the Smets-Wouters model, with the financial accelerator. There are some differences to the FRBNY DSGE model: the model parameters describing the financial sector are estimated directly, rather than implicit functions of them, the nominal loan growth rate is in addition to the credit spread included to inform the model about financial market dynamics and shocks to the survival probability of entrepreneurs are included in addition to a riskiness shock. Finally, the credit spread measure focuses on short-term financing, rather than on long-term financing as in Del Negro and Schorfheide (2013) and Del Negro et al. (2015).

An alternative model of financial frictions during the Global Financial Crisis is based on the roots of the crisis in the housing market. Iacoviello (2005) incorporates the collateral constraint model of Kiyotaki and Moore (1997) into a model of the housing market. Impatient households borrow from patient household, but face collateral constraints and their borrowing is thus constraint by the value of their housing stock. Both household types derive utility from housing. Banks intermediating between the two household types work under monopolistic competition as in Gerali et al. (2010), so that their lending rate is higher than the deposit rate. We use the version by Kolasa and Rubaszek (2015) who incorporate the housing sector and the collateral constraints in the model by Del Negro et al. (2007), include four shocks regarding the housing sector (housing weight in the utility function, the loan-to-value ratio, relative price of residential investment, and the markup in the banking sector) and use four observables to capture dynamics in the housing market: residential investment, mortgage loans, house prices and the spread on mortgage loans.

We also use the IMF's quarterly projection model (Carabenciov et al., 2008), a model without strict microeconomic foundations. Similar models calibrated for different countries are used at several country desks at the IMF to help structure the dialogue with member countries. The model includes an IS-equation and a Phillips curve with forward and backward looking elements. Further, the model includes a version of Okun's law relating unemployment to the output gap. The model is more flexible than standard DSGE models in the sense that various equilibrium values are modeled as stochastic processes. For example, potential output is driven by permanent level shocks as well as highly persistent shocks to its growth rate. Regarding macro-finance linkages, output in the IS-equation is affected by bank lending conditions. Banks are assumed to adjust their lending practices around an equilibrium value depending on their expectations about the real economy four quarters ahead and a financial shock. The equilibrium value of bank lending conditions follows a random walk. Empirically, bank lending conditions are measured based on the survey answers regarding financial conditions from the Federal Reserve Boards quarterly Senior Loan Officer Opinion Survey on Bank Lending Practices. The other observables used to estimate the model are the unemployment rate, output growth, inflation and the federal funds rate.

Finally, we use BVAR counterparts to all macroeconomic models by basing them on the same respective observables. We use the prior by Giannone et al. (2015) which shrinks the parameters towards a random walk. This results in a reduction in estimation uncertainty and improves out-of-sample forecasts substantially compared to an unrestricted VAR. The prior is centered around the Minnesota prior, but the degree of shrinkage is chosen optimally given the marginal likelihood of the data, by treating it as an additional hyperparameter, whereas in the original Minnesota prior (Doan et al., 1984) it is set ad hoc. The GLP prior also optimizes the hyperparameters of the sum-of-coefficients prior and the dummy-initial observations prior affecting the long-run properties.

# 3 Forecasting Methodology

We use U.S. real-time data vintages to estimate the various macroeconomic models. Hence, for a given forecast starting date, we estimate each of the models on the basis of the most recent data vintage that would have been available at that time. The data vintages that we use are those published close to the middle of each quarter, so that the information set is aligned with the timing of the Survey of Professional forecasts. We focus on forecasts around the Global Financial Crisis, so that the first forecasts are computed for 2008Q3 onwards and the last ones for 2009Q2 onwards. We consider forecasts for horizons $h = 0, \dots, 4$, i.e. we start with the nowcast and consider forecasts up to 4 quarters ahead and we focus on output growth forecasts. All estimation samples start in 1964Q1 and we re-estimate the models quarter-by-quarter with the arrival of a new data vintage, i.e. we extend the samples recursively.

To make sure that the the information set used for the generation of forecasts are perfectly aligned with the information available to professional forecasters, we estimate the models in each quarter based on the data vintage being published about 10 days before the deadline for professional forecasters to submit their forecasts to the SPF. These deadlines are August 7th in 2008Q3, November 10th in 2008Q4, February 10th in 2009Q1, and May 12th in 2009Q2. The corresponding data vintages were published on July 31st 2008, October 30th 2008, January 30th 2009, and April 29th 2009. Hence, the model-based forecasts are based on the same data vintage as available to professional forecasters in real time.

Most macroeconomic data series are subject to a publication lag of one quarter. Thus, for these data vintages, only the first estimate for key macroeconomic variables like GDP growth, is available. While the models use quarterly data, professional forecasters can use information until the forecast starting point (deadline of the SPF survey) based on data available on a higher frequency. This includes survey data, other leading indicators as well as financial market data. While many papers studying DSGE model-based forecasts focus on a balanced panel of data ending one quarter before the forecast starting point, using the Kalman filter it is straightforward to include observations for the current quarter for those variables for which they are available, i.e. to make efficient use of the ragged edge data property that forecasters are facing. The Kalman filter treats the observations for the variables for which no data is available for the current quarter as missing data points that need to be estimated. For example, the Kalman filter could use financial market data for the current quarter to infer the most likely current quarter value of GDP growth, i.e. to compute a GDP growth nowcast. For BVARs the Kalman filter can be used as well to nowcast missing observations at the sample end.

We consider four different scenarios regarding the data included in DSGE models in the current quarter:

1. Balanced Panel: As a benchmark scenario we do not use any data for the current quarter, but only include observations until the previous quarter for all time series used in the estimation of the different macroeconomic models.

2. Conditioning on SPF nowcast: To mimic the information set of professional forecasters that includes a variety of time series of different frequencies, but possibly also information not fully captured in

any data series, we append the mean of the SPF nowcasts as an additional observation to the data used for the model estimations. SPF nowcasts are available for the most frequently used observables: output growth, inflation, unemployment rate, non-residential investment, residential investment. We treat the current quarter observation for the remaining observables as missing and estimate it using the Kalman filter.

3. Conditioning on current quarter data: In this scenario observations for the current quarter are included for the data series that are not subject to publication lags. This is in particular the case for financial market data (interest rates, credit spread, mortgage spread), but also for macroeconomic data published on a monthly frequency (hours and bank lending conditions). For financial market data, we use the average of the daily data for the days until the forecast starting point in the current quarter and treat it as a quarterly observation for the whole quarter. For monthly data, we use the value for the first month of the quarter and also treat it as a quarterly observation for the whole quarter to avoid mixed-frequency estimation. In this scenario, the models do not rely on SPF nowcasts, but compute an own nowcast for output growth based on all the information about the respective observables available until the forecast starting point.

4. Full information conditioning: conditioning on SPF and current quarter data. We combine the information from scenarios 2. and 3. to include as much information about the current quarter as possible.

All models are estimated using the same observables, measurement equations and priors as proposed by the original authors. We maximize the posterior mode and then run the Metropolis-Hastings (MH) algorithm with 1,000,000 replications to simulate the posterior density. In the MH algorithm, the scale factor of the proposal distributions covariance matrix is set individually for each model to ensure the resulting acceptance rate is between 20% and 40%. In addition, we discard the first 30% of the samples to enhance the stability of posterior means. Finally, we compute density forecasts based on the posterior and compute point forecasts as the mean of the density forecasts.

# 4 Results

First, we analyze forecasts of the Great Recession based on pre-crisis models to document the state of macroeconomic modeling before the Global Financial Crisis. Afterwards, we repeat the forecasting exercise based on models developed after the Global Financial Crisis to see to which extent the progress in macroeconomic modeling over the last decade would have improved forecasts of the Great Recession. Finally, we compare the forecasts from the structural models to forecasts from Bayesian VARs to understand to which extent the usage of financial market variables that capture information about the crisis is sufficient to generate an accurate forecast or whether in addition the specific theoretical transmission channels in DSGE models are needed.

## 4.1 Forecasting the Great Recession: Pre-crisis models

Figures 1 and 2 show forecasts for annualized quarterly real output growth starting in 2008Q3, 2008Q4, 2009Q1 and 2009Q2 (from top to bottom) for the five pre-crisis models as described in Table 1 and the four different conditioning assumptions (from left to right).

Figure 1: GDP Growth Forecasts in 2008:III–2008:IV: Pre-Crisis Models

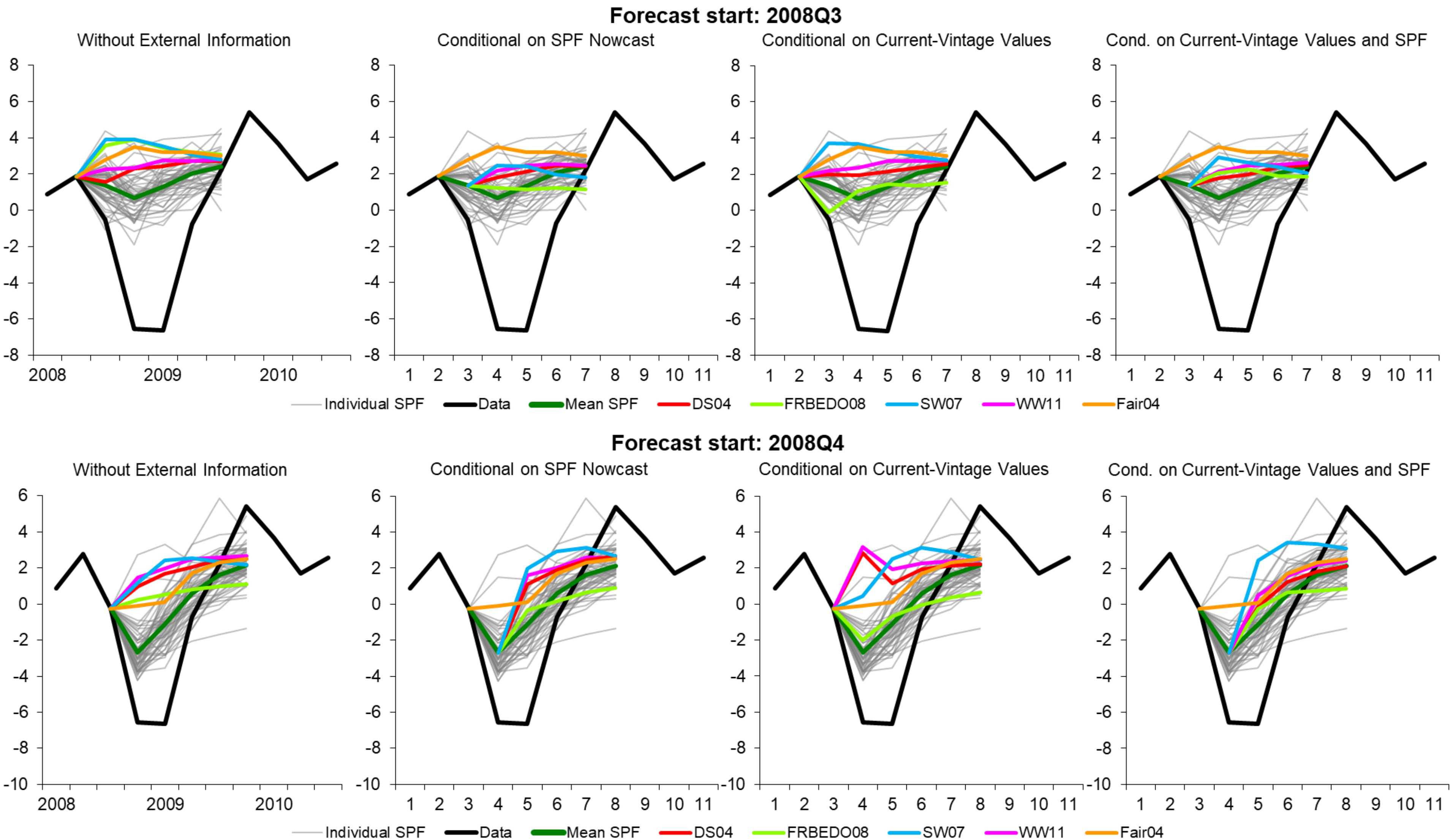
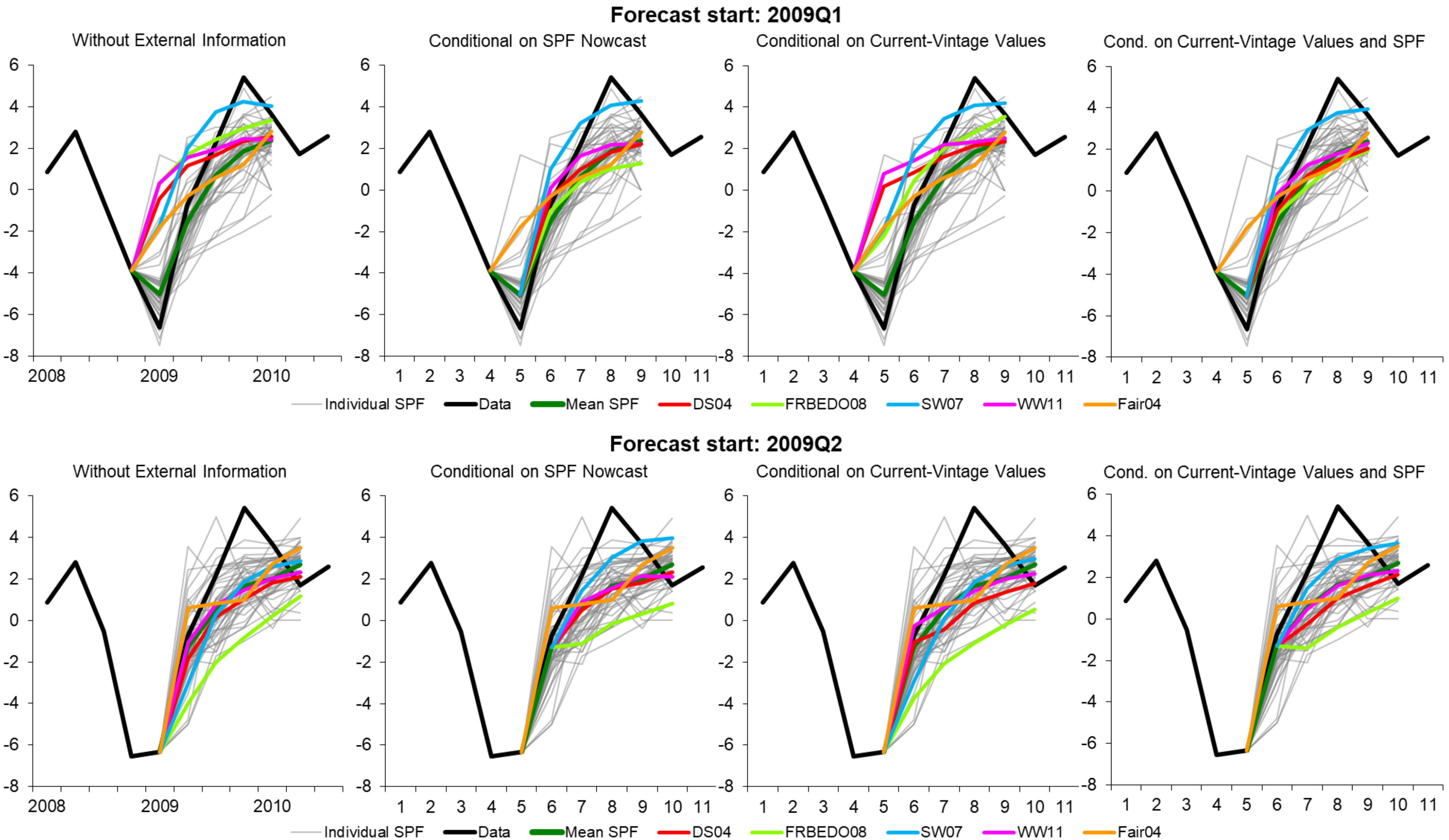


Figure 2: GDP Growth Forecasts in 2009:I–2009:II: Pre-Crisis Models



The black line shows real-time data until the forecast starting point and revised data afterwards. The models cannot predict changes in data revisions, so that we do not aim at forecasting final revised data that includes benchmark revisions. Instead, we use the data point in the vintage that was released two quarters after the quarter to which the data refer to as revised data. This data vintage includes the most important initial revisions of GDP, but excludes later benchmark revisions. The gray lines show forecasts collected in the SPF and the dark green line shows their mean. The professional forecasts are independent of the four conditioning assumptions for the model forecasts, so that respective forecasts are the same in all four columns. The remaining lines show the various model forecasts and the legend includes the short names introduced in the model overview in Table 1.

The forecasts shown in the top row start in the third quarter of 2008 and have been computed before the collapse of Lehman brothers. It is apparent that almost all professional forecasters failed to foresee the downturn. There are only few exceptions, but even these forecasters merely predicted a growth slowdown or a mild rather than a Great Recession. Accordingly, the mean SPF forecast indicates a slowdown of growth to about 0.7 percent followed by a return to growth rates above 1 percent in the first quarter of 2009 and above 2 percent afterwards.

The model-based forecasts would not have performed any better and predict even higher growth rates than most professional forecasters when looking at the scenario without conditioning on any current quarter information (column 1). Conditioning on SPF nowcasts (column 2), on current quarter data (column 3) or both (column 4) does not improve most model forecasts. As professional forecasters did not predict a downturn and financial market distress did not increase substantially before the fourth quarter of 2008, there is simply no additional information in these scenarios that would lead to much more pessimistic model-based forecasts.

There is one exception though. The FRBEDO08 model, the largest of the considered pre-crisis DSGE models, disaggregates investment into non-residential and residential investment and includes data for both variables in the estimation. While the model does not include any financial market frictions, financial turmoil during the housing crisis can be captured by the model to the extent that it affects residential investment. When conditioning the model forecast on data that is available for the current quarter (column 3), the model predicts the stagnation of GDP in the third quarter of 2008 with high accuracy due to the drop in residential investment during that quarter. The model, however, fails in predicting a prolonged recession as there is no mechanism in the model that amplifies the effects of negative residential investment growth rates, so that the effects on GDP growth are rather short-lived.

Moving forward one quarter, the plots in row 2 show forecasts starting in the fourth quarter of 2008. The professional forecasters predicted on average negative growth of about $-2\%$ for this quarter. The most pessimistic forecaster predicted a downturn of $-4\%$, but GDP turned out to be even lower. Further, professional forecasters wrongly predicted that the most negative growth rates would occur during this quarter and that afterwards growth rates would be less negative and quickly reach positive values again.

The models predict again on average higher growth rates than the professional forecasters in all four scenarios. Even when conditioning the forecasts on the negative growth rates of the mean SPF nowcast, the models predict positive growth rates already for the next quarter. During the fourth quarter of 2008 there were clear signals of severe financial turmoil, like a strong increase in credit spreads, but the pre-crisis models cannot capture this as they do not use such information in the set of observable nor do they include transmission mechanisms that would create recessionary dynamics in the next quarters. Again, the forecasts of the FRBEDO08 model differ from the other pre-crisis models. As the strong decrease of residential investment continued, the model predicts lower growth rates than the other models. When conditioning on the SPF nowcast, the model even predicts negative growth rates for the next three quarters, though it strongly underestimates the actual depth of the recessions. When conditioning on data for the current quarter, the GDP nowcast of the model is not very far from the one based on the mean of the SPF, though actual GDP growth was much lower.

Moving to forecasts starting in 2009Q1 and 2009Q2, almost all models get the speed of the recovery roughly right, in particular when conditioned on the mean SPF nowcast. This is due to the strong tendency of the models to revert rather quickly back to steady state. Again, the foreasts from the FRBEDO08 are more pessimistic, in particular those staring in 2009Q2, due to the inclusion of residential investment. While this increased the accuracy of forecasts at height of the crisis, the forecast of the recovery turns out to be too pessimistic.

An important finding from this section is the difference in the accuracy of the FRBEDO08 model and the other pre-crisis models. This shows that the inclusion of variables that capture some of the crisis dynamics increases the forecasting accuracy during the largest downturn of the U.S. economy. The FRBEDO08 model achieves a similar forecasting accuracy as the SPF. However, the lack of appropriate transmission channels prevent the model from predicting a deepening of the recession and the actual depths of the recession.

Would have using a more traditional Cowles Commision type of model resulted in more accurate forecasts? To check whether this proposal that has been put forward for example by Krugman (2009) and Buiter (2009) would have resulted in more promising recession predictions, we use the model by Fair (2018). We do not estimate this model ourselves, but rely on the forecasts computed by Ray Fair in real time throughout the Global Financial Crisis based on earlier model vintages. Hence, we cannot condition the model on the four different scenarios, so that the four forecasts shown in the four columns in Figures 1 and 2 are the same for this model.

In 2008Q3 the forecasts from this model are among the most optimistic ones. The forecast starting in 2008Q4 predicts a stagnation of GDP for the next two quarters, but no recession. Finally, starting from 2009Q1 the model shows a similar tendency to revert back to steady state as the other pre-crisis models. Ray Fair discusses these forecasts in detail in forecast memos.[1] Regarding the 2008Q4 forecast, he explains that the negative wealth effect from the fall in stock prices is a major reason for the no growth prediction over the next two quarters, but that the model cannot capture any credit crunch effects due to financial distress nor is there any information in the initial conditions and the assumed path for exogenous variables (mainly fiscal variables) that would lead to the prediction of a recession. The same effect explains why the model predicts negative growth for the two quarters after 2009Q1, but not a further decrease of growth.

## 4.2 Forecasting the Great Recession: Post-crisis models

Next, we analyse whether the inclusion of financial market frictions that are important for explainig the Global Financial Crisis in the post-crisis models lead to better recession forecasts. Figures 3 and 4 show forecasts for the five post-crisis models described in Table 1. The figures are structured exactly as the previous ones for the pre-crisis models.

[1] See https://fairmodel.econ.yale.edu/memo/memo083.htm, https://fairmodel.econ.yale.edu/memo/memo084.htm, https://fairmodel.econ.yale.edu/memo/memo091.htm, and https://fairmodel.econ.yale.edu/memo/memo092.htm.

Figure 3: GDP Growth Forecasts in 2008:III–2008:IV: Post-Crisis Models

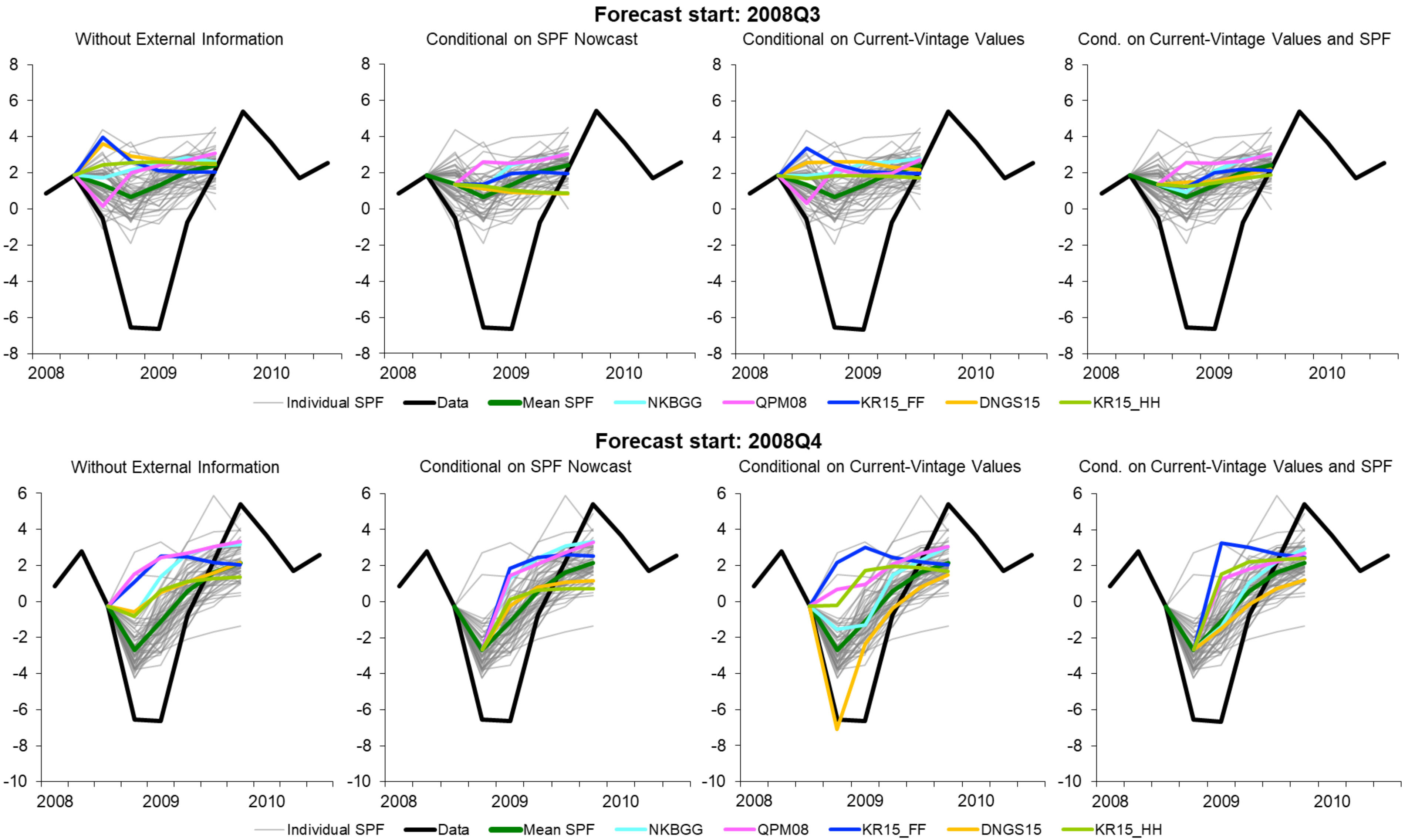
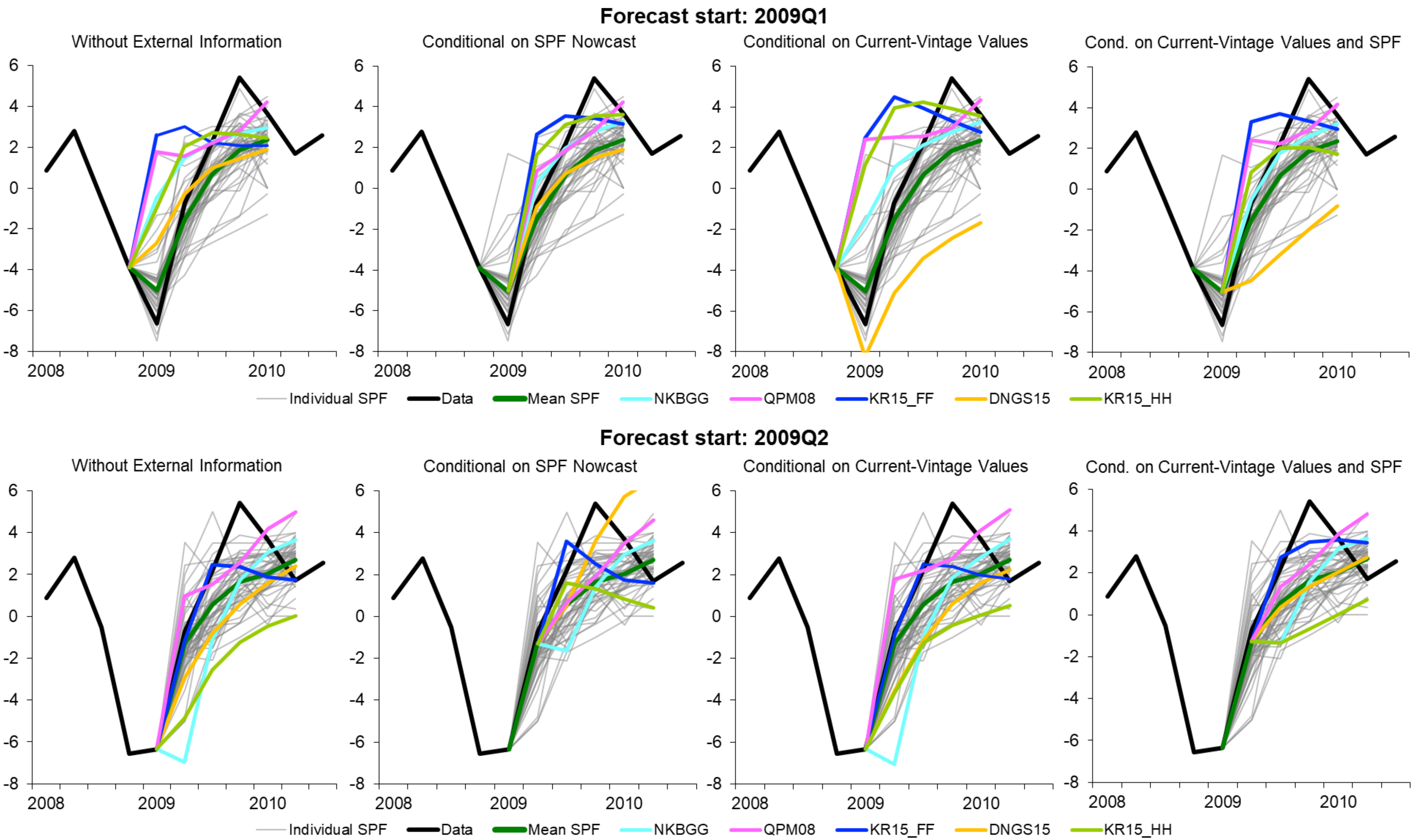


Figure 4: GDP Growth Forecasts in 2009:I–2009:II: Post-Crisis Models



The forecasts of the post-crisis models starting in 2008Q3 do not look systematically different from those of the pre-crisis models. They are somewhat closer to the mean SPF than the pre-crisis models, so that the overprediction of growth is somewhat smaller compared to most pre-crisis models, but none of the post-crisis models predicts an upcoming recession. This result does not change when conditioning the forecasts on the SPF nowcast or on the data available for the current quarter. This is not entirely surprising, as financial market indicators, like for example credit spreads, did not change drastically until the fourth quarter of 2008. As long as there is no information in the data about financial distress, the inclusion of financial frictions does not lead to a prediction of a recession.

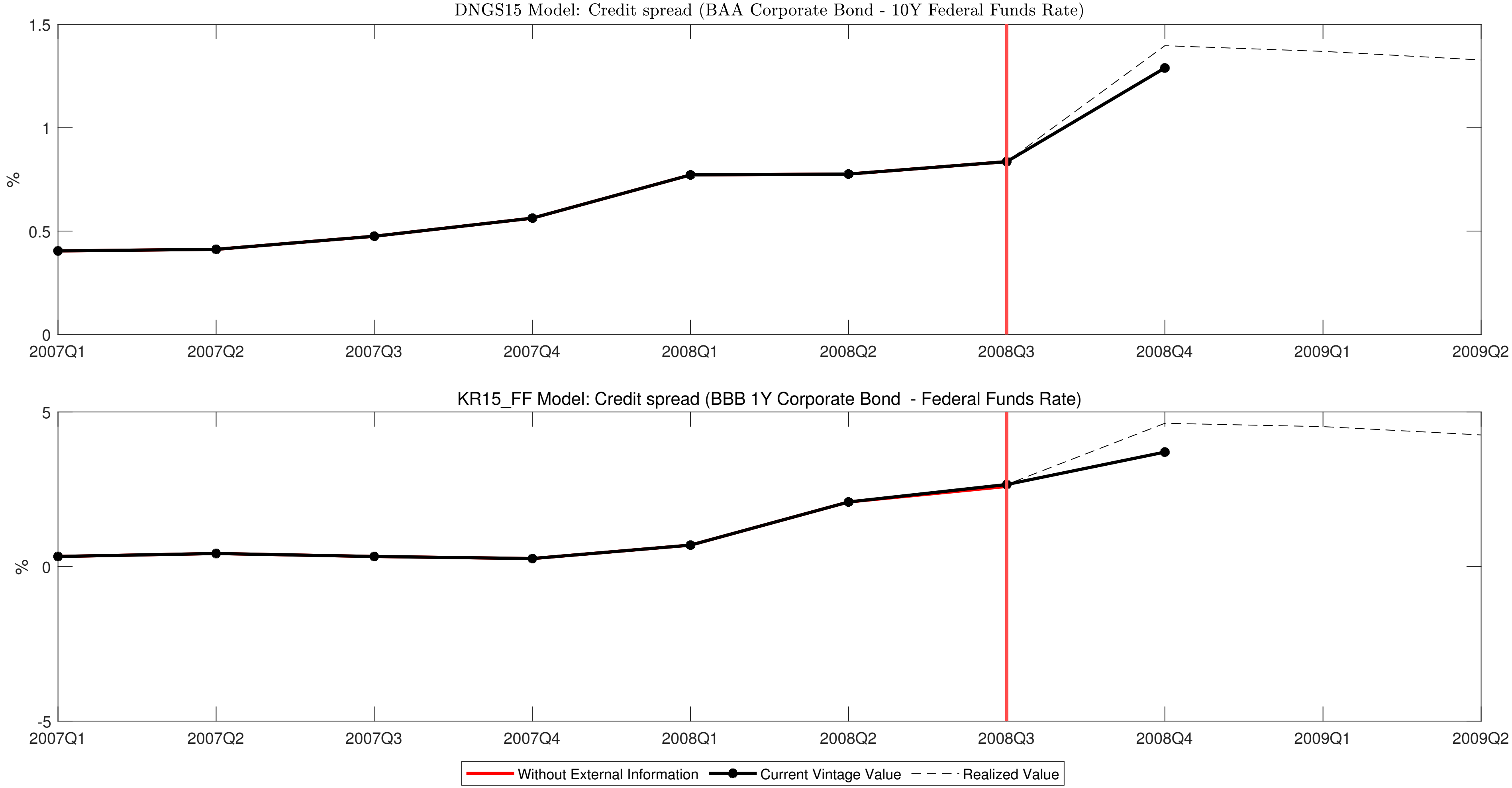
During the fourth quarter of 2008 and the first quarter of 2009 the largest contraction of the US economy took place. However, at the beginning of the fourth quarter, forecasters hardly predicted a recession, while at the end of the quarter it was quite clear that a large contraction was taking place. For example, the Greenbook projections (not shown in the figure) from October 22 predicted a mild downturn only, while the ones from December 10 predict a strong decrease of GDP for the fourth quarter of 2008 and the first quarter of 2009 that turned out to be quite accurate. Hence, the exact timing and the corresponding information available to forecasters is important. The survey for the SPF was sent out on October 31 and answers were collected until November 10, i.e. in the first half of the quarter. Hence, professional forecasters were not yet aware that a severe crisis was taking place.

Accordingly, also for the model forecasts starting in the fourth quarter of 2008, the assumptions regarding the inclusion of information about the current quarter is crucial. If no data for the current quarter is included (column 1) the post-crisis models do not perform any better than the pre-crisis models. Conditioning the forecasts on the mean SPF nowcast (column 2) leads to a more pessimistic view on the current quarter as professional forecasters revised their assessment of current economic conditions downwards compared to the previous quarter. Nevertheless, they substantially underestimated the downturn. The model forecasts for the next quarters are, however, not much affected by the more pessimistic nowcast. All post-crisis models predict less negative growth rates in the next quarter and a return to positive growth rates directly afterwards similar to the pre-crisis models.

When conditioning on current quarter financial market data (column 3), the forecasts of models that include a financial accelerator mechanism and use a credit spread as an observable change drastically. These models endogenously generate highly negative GDP growth rates for the fourth quarter of 2008. The nowcast of the DNGS15 model is very exact and much more accurate than the mean SPF nowcast and even the most pessimistic SPF nowcast. Hence, one can conclude that there has been indeed substantial progress in macroeconomic modeling. While the models cannot predict a crisis in advance, the financial accelerator mechanism enables them to detect the crisis, once credit spreads increase as also documented in Del Negro and Schorfheide (2013) and Del Negro et al. (2015). The smaller NKBGG model also includes the financial accelerator mechanism, and although it does not predict such a large downturn as the DNGS15 model, it also endogenously forecasts a recession with depth and length similar to the mean SPF predictions. The difference to the DNGS15 model shows that the combination of a model with many nominal and real frictions, the financial accelerator and the conditioning on the current quarter credit spread is important to achieve an accurate crisis forecast.

However, the combination of the specific observables chosen is crucial, too. The KR15_FF, another medium-scale DSGE model with the BGG-type financial accelerator, does not predict the large downturn, despite being similar the DNGS15 model. An important difference between the two lies in the credit spread data used. We follow the original authors and use the spread between the BAA 10Y corporate bond and the 10Y treasury constant maturity rate for the DNGS15 model, and the spread between the BBB 1Y corporate bond and the Federal Funds Rate for the KR15_FF model. Hence, for the DNGS15 model a measure of long-term financing risk premia is used, while the focus of the KR15_FF model is on shorter horizon financing risk premia. The resulting difference is large for the 2008/2009 crisis period. The fourth quarter of 2008 was a five sigma event for the long-term financing credit spread, occuring with extremely low probability. This extremely large financial shock leads to the highly negative GDP growth nowcast in the DNGS15 model. The short-term financing credit spread 2008Q4 was only 1.8 sigma larger than the historical mean, i.e. the shock was not too large occuring with about 8 percent probability. Further, the delayed drop in the loan growth rate that is included as an observable in the KR15_FF, but not in the DNGS15 model, might offset the large negative effect off the credit spread increase on GDP growth to some extent.

Figure 5: Credit Spread Variables in 2008Q4




Notes: The figure shows the real time observable credit spread variable and the difference in their definition in the DNGS15, top panel, and in the KR15_FF model, lower panel. The red vertical line indicates the information content when the forecasts were made without external information, that is only data up until and including 2008Q3 is used. The black line indicates the current vintage values, that is the currently observable credit spread. Lastly the black dashed line shows the realized values. The scaling of the vertical axis reflects historical variation of the series.

The other post-crisis models do not predict a recession, even when using information about the current quarter. The KR15_HH model does not include a financial accelerator, but focuses on collateral constraints in the housing market. The model is informed about financial conditions in the housing market via data on residential investment, mortgage loans, house prices and the spread on mortgage loans. In contrary to the credit spread that informs the DNGS15 and the NKBGG model, these variables do not show large changes in the fourth quarter of 2008. The QPM model does not predict a recession either. This model is informed by a measure of bank lending tightness about financial market distress. This measure increased throughout 2008 to unprecedented values, but there are no sudden changes towards the end of 2008. Further, there is no data on bank lending tightness for the current quarter. On the other hand, this is the only model that includes the unemployment rate as observable. The increase in the unemployment rate in 2008 was modest, so that this does not help in predicting a large recession.

When conditioning on SPF nowcasts and current quarter financial market data (column 4), the forecasts of the post-crisis models look again similar to those of the pre-crisis models. Even the DNGS15 and the NKBGG model cannot predict the full downturn, because the output growth nowcast is restricted to be equal to the mean SPF output growth nowcast.

Moving to the forecasts starting in 2009Q1 and 2009Q2, almost all post-crisis models predict a quick return to positive growth rates as the pre-crisis models and the SPF. The only exception is the DNGS15 model when being conditioned on current quarter data. In this case, the elevated credit spread leads to a more pessimistic nowcast and a forecast that in retrospect turns out to be too pessimistic.

### 4.3 Forecasting the Great Recession: Data versus Model Structure

In the previous section it turned out that the inclusion of the financial accelerator in DSGE models leads in combination with information about the credit spread for long-term financing to accurate GDP nowcasts. Now we will compare forecasts from the small-scale (NKBGG) and the medium-scale (DNGS15) financial accelerator models, to Bayesian VAR counterparts estimated on the same time series. In this way, we can analyse to which extent the inclusion of appropriate financial market indicators into an atheoretical empirical model is sufficient or whether modeling the financial accelerator is important to generate precise forecasts.

Figures 6 to 9 show forecasts for the three forecast starting points and the four scenarios for the two DSGE models and the BVAR counterparts. The NKBGG model is estimated on five time series including a credit spread, so that the BVAR labeled 5vGLP is its atheoretical counterpart. The DNGS15 model is estimated on the same eight time series as the BVAR labeled 8vGLP.

Figure 6: GDP Growth Forecasts in 2008:III–2008:IV: 5-Variable NK Model versus 5-Variable Bayesian VAR Model

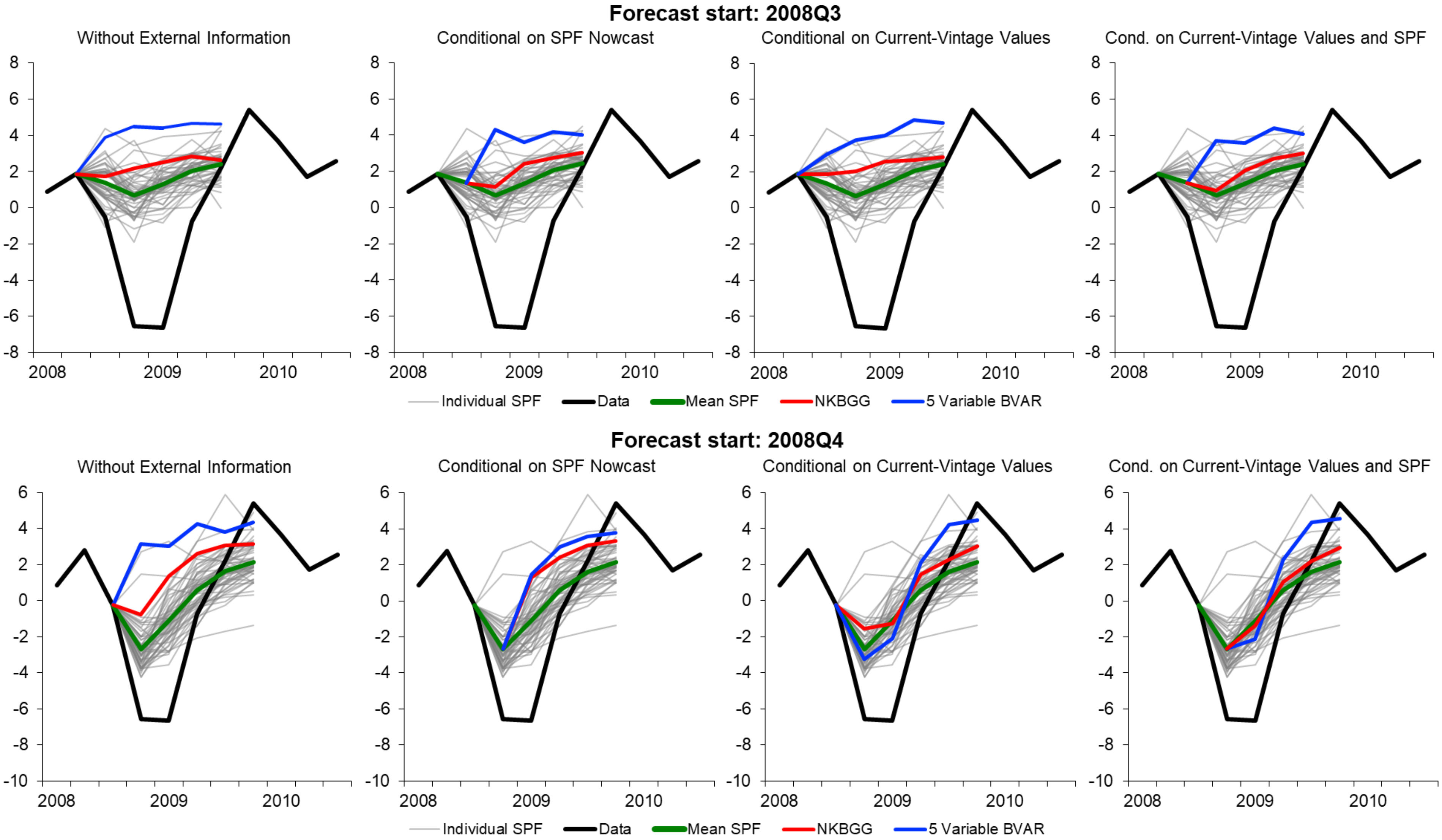


Figure 7: GDP Growth Forecasts in 2009:I–2009:II: 5-Variable NK Model versus 5-Variable Bayesian VAR Model

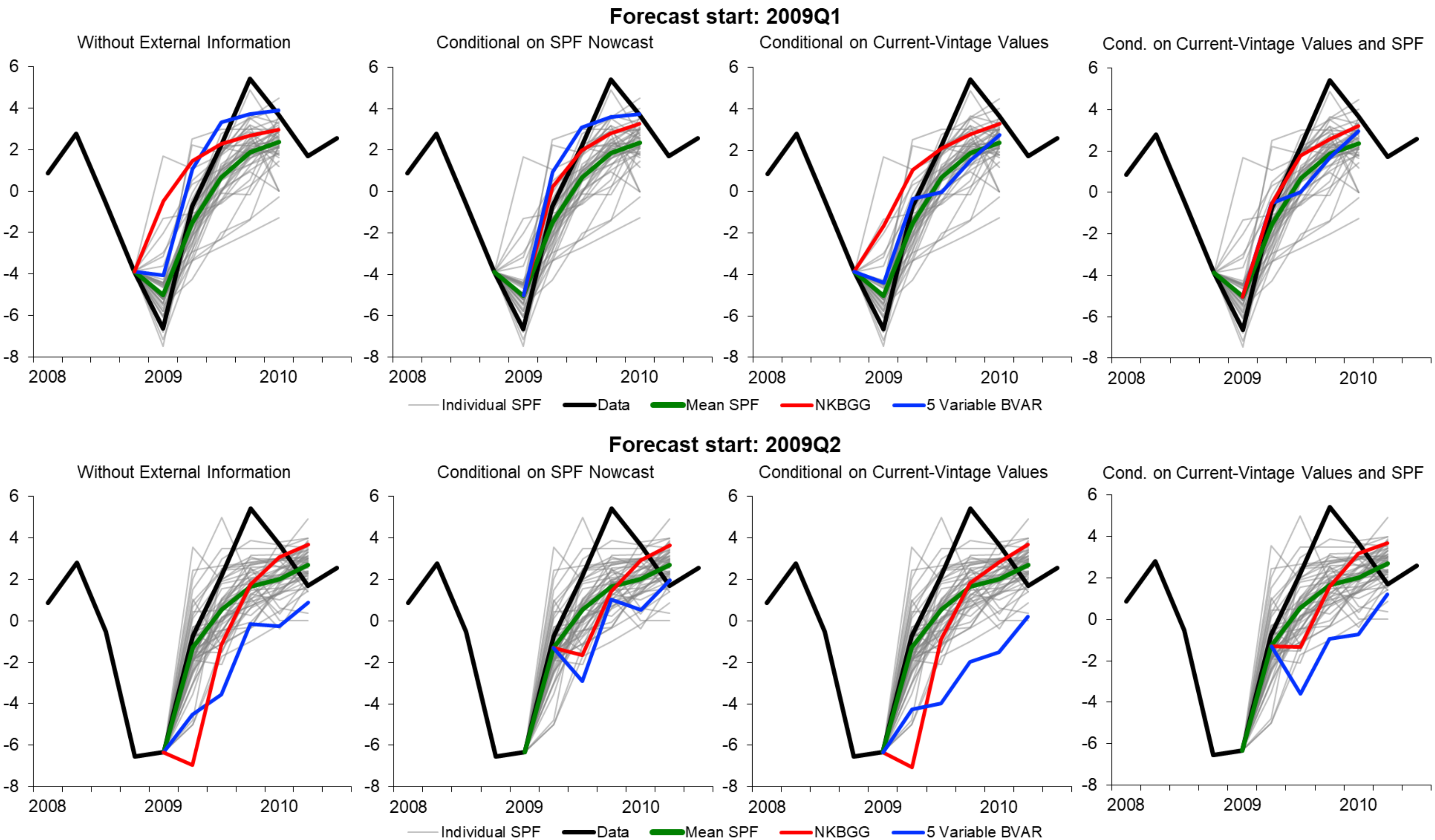


Figure 8: GDP Growth Forecasts in 2008:III–2008:IV: 8-Variable NK Model versus 8-Variable Bayesian VAR Model

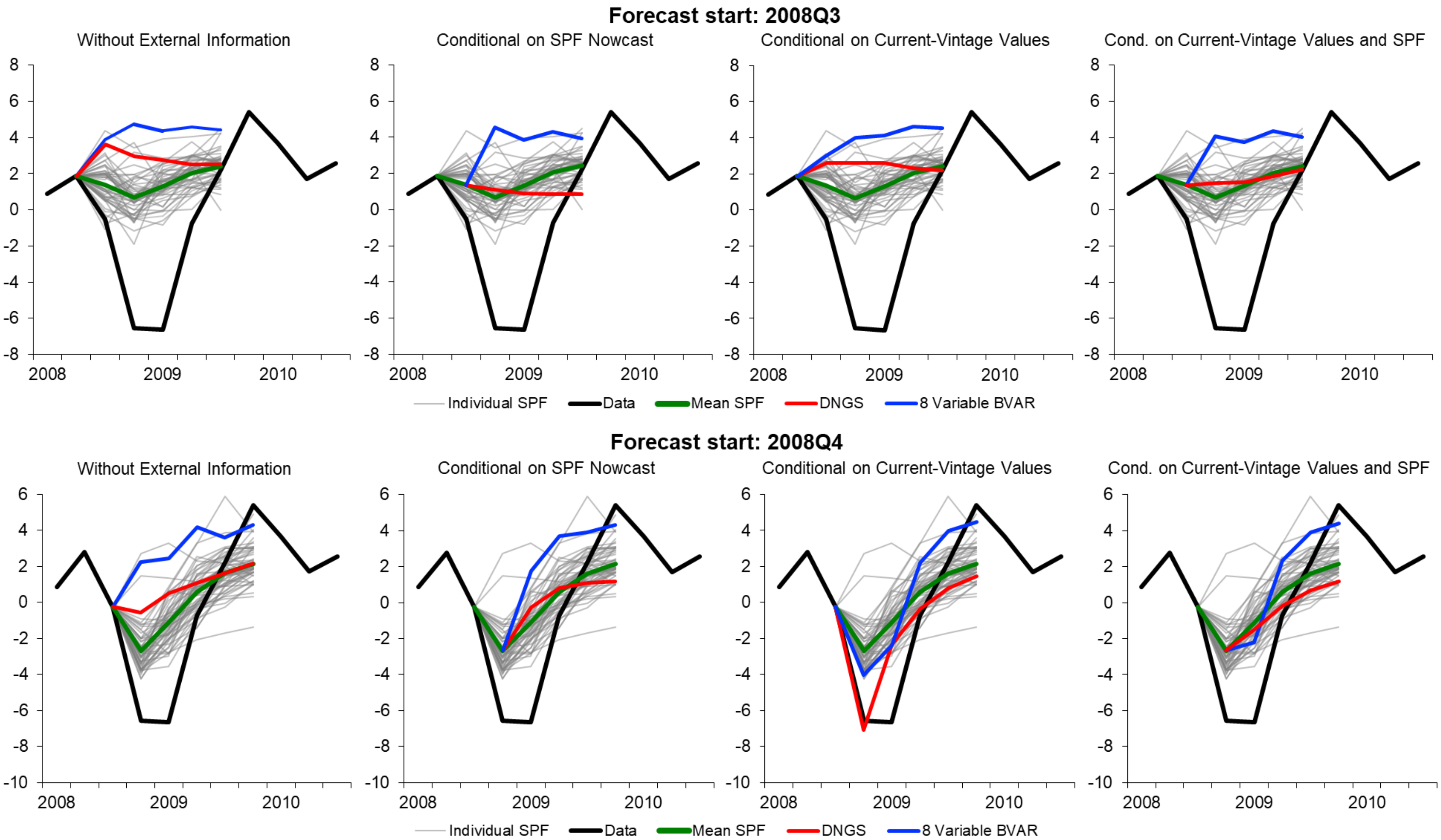
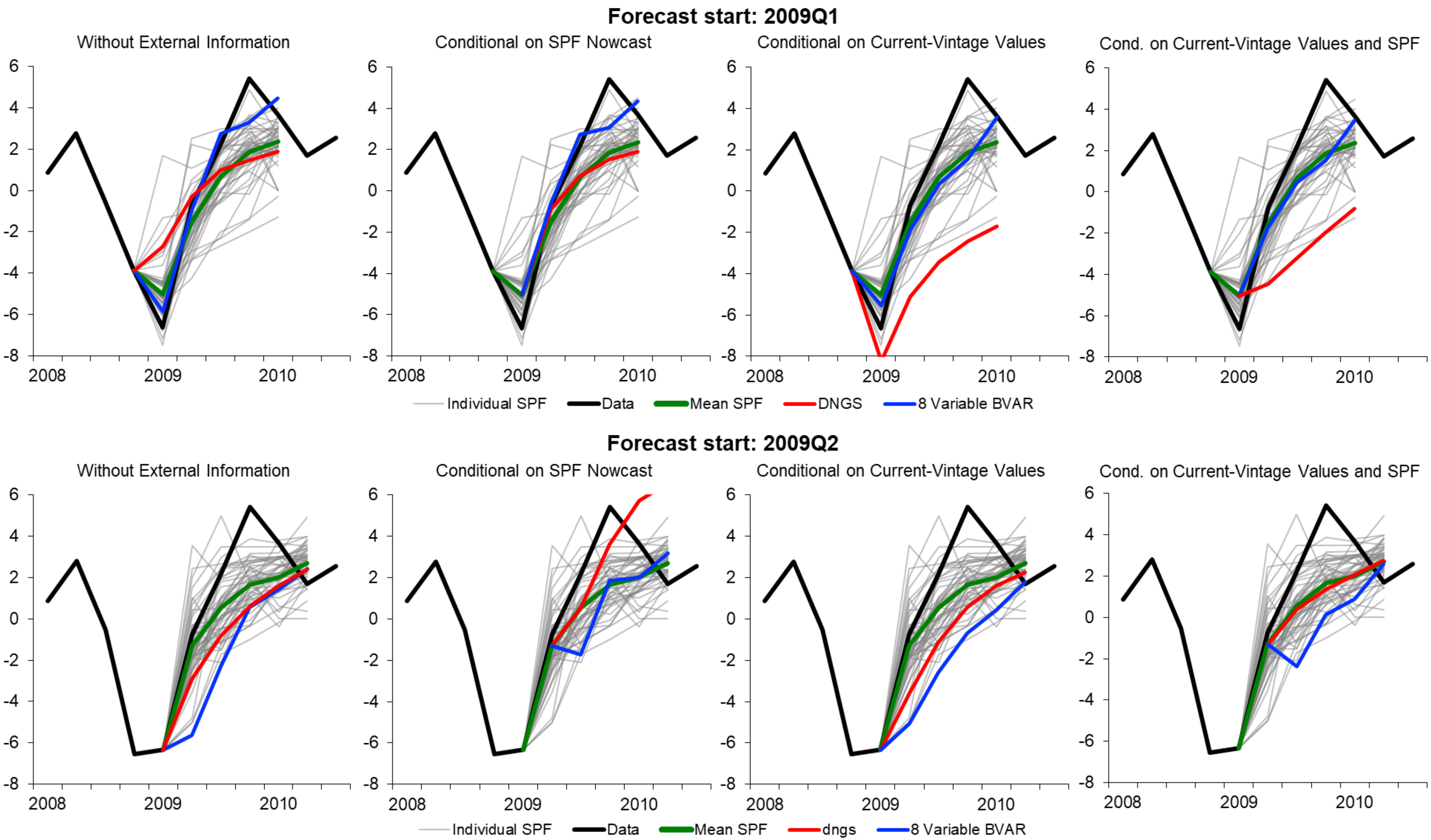


Figure 9: GDP Growth Forecasts in 2009:I–2009:II: 8-Variable NK Model versus 8-Variable Bayesian VAR Model



The BVAR forecasts are systematically higher than the DSGE counterparts for the forecasts starting in 2008Q3. They merely extrapolate the relatively high growth numbers of the most recent quarters into the future. The cross-equation restrictions imposed by economic theory are helpful to improve the accuracy of forecasts, though the DSGE models are all too optimistic, too. Irrespective of the conditioning this pattern holds across all four scenarios (columns 1-4).

Regarding the forecasts starting in 2008Q4, the results are very different for the different forecasting scenarios. Without conditioning on information from the SPF or data for the current quarter (column 1) or when conditioning on the SPF nowcast (column 2), the BVAR forecasts are higher than the DSGE forecasts and therefore much too optimistic. Conditional on additional information for the current quarter (columns 3-4) the forecasts from the 5-variable BVAR and its DSGE counterpart are very similar. The 8-variable BVAR and its DSGE counterpart predict a large downturn, when being conditioned on current quarter data including data on the credit spread data (column 3). While the nowcast of the DNGS15 is very precise, the 8-variable BVAR nowcast is still too optimistic, though it is very similar to the most pessimistic forecasts from the SPF.

For the forecasts starting from 2009Q1 the BVAR and DSGE model based forecasts show overall similar dynamics, predicting the return to average growth rates quite accurately. An exception is the DNGS15 model when being conditioned on current quarter data (column 3) or in addition on SPF nowcasts (column 4). The information on the credit spread that was strongly elevated in 2009Q1 leads to a too pessimistic forecast. Comparing the BVAR 8-variables model with the DNGS15 shows that the theoretical restrictions a DSGE model represents make GDP forecasts more sensitive to conditioning on financial conditions compared to the agnostic BVAR.

Consulting the forecasts starting in the second quarter of 2009, the BVAR forecasts are more pessimistic compared to their DSGE counterparts. The BVAR models extrapolate the highly negative growth rates of the previous quarters, so that the forecasts turn out very pessimistic, while the theoretical restrictions imposed by the DSGE models prove to be helpful to predict the speed of the recovery with higher accuracy. Both BVARs and both DSGE models are, however, too pessimistic compared to the actual path of GDP growth that was temporarily higher than its historical mean.

Overall, the theoretical structure of the DSGE models shows to be very important in achieving precise forecasts. Only using an additional credit spread time series in an otherwise standard BVAR model leads in most cases to less accurate forecasts, that driven by the GLP prior's unit root property extrapolate recent GDP growth observations into the future. The structure of the DSGE models is important when accounting for financial conditions' impact on GDP, as comparing the conditioning's impact with the BVAR model they restrict the predicted joint dynamics of the model variables in a way that forecasts become much more accurate. It is important to highlight the different impact of conditioning on BVARs across horizons. The agnostic conditioning employed by the entropic tilting has a short-lived impact on forecasts and leaves longer-term forecasts unchanged. At the same time the case of the DNGS15 highlights that theoretical restrictions can amplify the impact of conditioning.

### 4.4 Systematic Forecast Evaluation

In this section we study the point forecast performance of the models using root means squared prediction errors (RMSE). It is defined for the target variable, GDP growth $y$, as follows:

$$RMSE_{j,h} = \sqrt{\frac{1}{T_1 - T_0 - h + 1} \sum_{T=T_0+h-1}^{T_1} \left(E[y_{j,T+h}|I_T] - y_{T+h}\right)^2} \tag{1}$$

where $E[y_{j,T+h}|I_T]$ is the forecast of the model $j$ estimated conditional on information set $I_T$, that accounts for differences in conditioning, for forecast horizon $h$. $y_{T+h}$ denotes the data realization $h$ periods ahead. We use the data realization in the data vintage that was released two quarters after the quarter to which the data refer to as revised data, as in Wieland and Wolters (2011). This data vintage includes the most important initial revisions of GDP, but excludes later benchmark revisions. $T_0$ denotes the start (2008Q3) and $T_1$ the end (2009Q1) of the evaluation period. The RMSE of the mean is the optimal measure for forecast performance if a quadratic loss function is assumed.

Table 2 shows the RMSEs for GDP growth starting from horizon 0 (nowcast) up to horizon 4, i.e. forecasts for GDP growth four quarters ahead. RMSEs are reported relative to those of the RMSE of the mean SPF forecast, for which absolute RMSEs are reported in the last column of the last panel. Hence, a number that is lower (higher) than one suggests that the forecast based on a particular source is more (less) accurate than the SPF mean forecast on the corresponding horizon.

The first two panels presents the results of the structural models, while the third panel presents the RMSEs for the BVAR counterparts and professional forecasts (mean forecast of the Survey of Professional Forecasters, SPFM, and the forecasts from Federal Reserves Greenbook, GB). The Greenbook forecasts include the forecast presented in the first (GB1) and second meetings (GB2) of the FOMC in a quarter. Results are presented for all four different data scenarios: balanced panel without external information (NoExt), conditioning on nowcast data from the SPF (SPF), conditioning on current quarter data (CQ), full information conditioning (FV). We refrain from conducting statistical tests for difference in the RMSEs as the number of observations is very small.

The SPF RMSEs show that the forecasting performance is worse for short-term forecasts compared to long-term forecasts. This is due to the specific choice of the evaluation period. Short-term forecasts are evaluated against data from the midst of the Great Recession, while long-term forecasts are evaluated against data from the recovery period, during which GDP growth was closer to its historical mean. The SPF nowcast is an exception with a forecasting accuracy similar to the long-term forecast. Here, the additional within quarter information from other indicators increases the nowcasting precision. The Greenbook forecast show a similar forecasting accuracy as the SPF. The only difference is that for the Greenbook forecasts from the first (second) FOMC meeting in a quarter the nowcasting accuracy is lower (higher) compared to the SPF, which reflects the earlier (later) forecast publication date compared to the SPF.

Regarding short-term forecasts, survey based benchmarks outperform DGSE and BVAR based forecasts, while DSGE models do improve upon the BVARs providing evidence in support for theoretical restrictions. According to a special survey conducted by the Philadelphia Fed, professional forecasters usually combine mathematical models plus subjective considerations in reporting their projection (Stark, 2013). The subjective judgement of the forecasters might be reason for the better performance of the mean SPF in forecasting GDP growth in short horizons. It is hard to imagine that, any forecasters who experienced the subprime mortgage crisis, the collapse of the Lehman Brothers and the turbulence in the financial markets, would still believe that the economy would stay on its normal track in the coming quarters. While forecasters can directly take such extreme disturbances into their subjective considerations, it is more difficult for linear models to document and reflect such a non-linearity in their forecasts. Even so, the similarity between the performance of the mean SPF and the DNGS15 model in the third scenario suggests that, as long as we inform the structural model of the financial market disturbance in a timely manner, the resulting nowcast will become much closer to that of professional forecasters or can be even more precise.

In contrast, professional forecasts shows no general advantage of forecasting the GDP growth in longer horizons compared with structural and time-series models. Particularly, they are surpassed by four out of five pre-crisis models, as well as four out of five post-crisis models, in the three-step-ahead forecast. GDP growth is a stationary process that has the tendency of reverting back to its steady-state value gradually. In such case, the information advantage of the professional forecasters becomes a drag rather than a help, as they might place too much weight on those signals, which are only helpful in predicting the short-run GDP growth, but cannot reflect the long-run trend. This also explains why the second-release of the Greenbook forecast, which is obtained based on a richer information set than any other sources, is much better in forecasting short-run but not long-run GDP growth than other models or the mean SPF. For the four-quarter-ahead horizon, the post-crisis models' forecasting performance deteriorates compared to the SPF. This reflects the too pessimistic forecasts regarding the recovery, caused by a persistent increase in credit spreads.

Among the structural models, the relative forecasting performance depends strongly on the assumed information scenario. Without current quarter information (column 1, NoExt), there are not many systematic differences across the models. Overall, the pre-crisis models deliver better long-horizon forecasts, while the post-crisis models deliver better short-term forecasts. This reflects that including measures of financial conditions is in particular relevant for forecasting the recession, while it in other times models without financial frictions might perform better (Kolasa and Rubaszek, 2015; Del Negro and Schorfheide, 2013; Del Negro et al., 2015).

Exploring the impact of SPF nowcast conditioning (second column, SPF) shows that not only short-term forecast performance (1 quarter ahead) improves, but long-term forecasts as well. This illustrates that getting the depth of the recession more accurate helps in predicting the recovery. Nevertheless the improvement is not enough to make the forecasts as accurate as the survey based forecasts. The BVARs also show improvement in forecast performance albeit the least. This can be accounted to the entropic tilting conditioning being more flexible compared to the Kalman filter based nowcast conditioning of the DSGE that has to satisfy the cross equation restrictions strictly.

Current quarter conditioning (column 3, CQ) does not improve the forecasts to the same extent as the SPF conditioning for most models. This highlights that SPF nowcasts included more information than the current observables of financial variables. Not surprising conditioning on current quarter information improves the forecasts the most if the model features financial variables (NKBGG, DNGS15, KR15_FF, KR15_HH, BVARv5 and BVARv8). The highly precise nowcast of the DNGS15 model is reflected in a large decrease of the RMSE for horizon 0 compared to the other models.

Lastly full information conditioning (column 4, FV) improves the forecast performance the most for the less complex small DSGE models (DS04, WW11, NKBGG, QPM), while it deteriorates the forecast performance in larger models compared to conditioning only on SPF nowcasts. A potential explanation could be that the information of the SPF nowcasts are at odds with the current period financial variables, thus their joint conditioning deteriorates the forecast performance.

Table 2: Relative Root Mean Squared Errors (RMSE) for the GDP Growth Forecast

Note: The table shows the RMSEs relative to the RMSEs of the SPF Mean for the GDP growth forecasts on five horizons in four scenarios. The only exception is the last column, which shows the (absolute) RMSEs of the SPF Mean.

| Model | | DS04 | | | | WW11 | | | | SW07 | | | | FRBEDO | | | | Fair04 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario | | NoExt | SPF | CQ | FV | NoExt | SPF | CQ | FV | NoExt | SPF | CQ | FV | NoExt | SPF | CQ | FV | |
| Forecast Horizon | 0 | 2.17 | | 2.59 | | 2.37 | | 2.72 | | 2.27 | | 2.11 | | 2.39 | | 1.51 | | 1.91 |
| | 1 | 1.34 | 1.24 | 1.29 | 1.17 | 1.37 | 1.31 | 1.36 | 1.23 | 1.54 | 1.36 | 1.52 | 1.43 | 1.46 | 1.14 | 1.15 | 1.22 | 1.32 |
| | 2 | 1.16 | 1.11 | 1.14 | 1.11 | 1.19 | 1.14 | 1.17 | 1.13 | 1.26 | 1.12 | 1.25 | 1.16 | 1.32 | 1.08 | 1.15 | 1.21 | 1.24 |
| | 3 | 1.03 | 1.07 | 1.04 | 1.13 | 1.00 | 1.00 | 1.02 | 1.05 | 0.86 | 0.65 | 0.86 | 0.77 | 1.22 | 1.25 | 1.13 | 1.25 | 1.21 |
| | 4 | 0.87 | 0.87 | 0.95 | 1.02 | 0.85 | 0.89 | 0.89 | 0.92 | 0.96 | 0.99 | 0.90 | 0.83 | 1.21 | 1.45 | 1.35 | 1.35 | 0.99 |

| Model | | NKBGG | | | | QPM | | | | DNGS15 | | | | KR15_FF | | | | KR15_HH | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario | | NoExt | SPF | CQ | FV | NoExt | SPF | CQ | FV | NoExt | SPF | CQ | FV | NoExt | SPF | CQ | FV | NoExt | SPF | CQ | FV |
| Forecast Horizon | 0 | 2.33 | | 2.12 | | 2.55 | | 2.57 | | 1.86 | | 1.03 | | 2.77 | | 2.87 | | 2.06 | | 2.31 | |
| | 1 | 1.35 | 1.27 | 1.16 | 1.06 | 1.37 | 1.34 | 1.30 | 1.29 | 1.32 | 1.09 | 1.24 | 1.12 | 1.46 | 1.31 | 1.53 | 1.43 | 1.39 | 1.15 | 1.43 | 1.29 |
| | 2 | 1.15 | 1.15 | 1.12 | 1.07 | 1.12 | 1.13 | 1.04 | 1.04 | 1.19 | 0.89 | 1.31 | 1.18 | 1.09 | 1.07 | 1.10 | 1.08 | 1.28 | 0.98 | 1.20 | 1.17 |
| | 3 | 0.96 | 0.93 | 0.90 | 0.93 | 0.91 | 0.89 | 0.77 | 0.81 | 1.15 | 1.00 | 1.82 | 1.68 | 0.97 | 0.81 | 0.80 | 0.75 | 1.24 | 0.84 | 0.97 | 1.14 |
| | 4 | 1.14 | 1.11 | 1.17 | 1.21 | 1.49 | 1.39 | 1.56 | 1.55 | 1.40 | 2.53 | 2.47 | 2.32 | 1.37 | 1.08 | 1.29 | 1.29 | 1.69 | 1.86 | 1.46 | 1.38 |

| Model | | 3vBVAR | | | | 5vBVAR | | | | 8vBVAR | | | | Greenbook | | SPFM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario | | NoExt | SPF | CQ | FV | NoExt | SPF | CQ | FV | NoExt | SPF | CQ | FV | GB1 | GB2 | |
| Forecast Horizon | 0 | 3.07 | | 2.98 | | 2.54 | | 1.39 | | 2.43 | | 1.36 | | 1.21 | 0.48 | 2.27 |
| | 1 | 1.68 | 1.53 | 1.69 | 1.53 | 1.71 | 1.57 | 1.39 | 1.36 | 1.64 | 1.56 | 1.34 | 1.34 | 0.96 | 0.87 | 4.64 |
| | 2 | 1.40 | 1.26 | 1.46 | 1.34 | 1.48 | 1.30 | 1.49 | 1.39 | 1.44 | 1.32 | 1.42 | 1.35 | 1.01 | 1.05 | 4.52 |
| | 3 | 1.22 | 1.14 | 1.31 | 1.25 | 1.46 | 1.29 | 1.82 | 1.65 | 1.30 | 1.25 | 1.57 | 1.49 | 1.07 | 1.34 | 2.42 |
| | 4 | 1.06 | 1.05 | 0.99 | 1.00 | 1.08 | 0.94 | 1.24 | 0.86 | 1.05 | 1.01 | 0.97 | 0.88 | 1.82 | 1.91 | 1.83 |

# 5 Conclusion

The real-time forecasting experiment shows that some macroeconomic models that have been developed after the Global Financial Crisis, could have detected the Great Recession at its onset, i.e. they provide a precise nowcast, while pre-crisis models do not. In this sense, there has been substantial progress in macroeconomic modeling, yielding models that provide more accurate forecasts than professional forecasters. Our results show, however, that not all post-crisis models yield accurate forecasts, but these are restricted to medium-scale DSGE models that include the financial accelerator, while models focusing on collateral constraints in the housing market do not perform better than pre-crisis models. Further, the data used to inform the models about financial distress is very important. Only in a scenario, in which the models are conditioned on current quarter information about the credit spread, a precise nowcast is achieved, while starting from data that ends with the previous quarter or conditioning model-based forecasts on SPF nowcasts results in much too optimistic forecasts. A comparison with forecasts from Bayesian VARs shows that the theoretical restrictions of DSGE models help in increasing the accuracy of forecasts during the Great Recession. While these results are encouraging, all models still fail in predicting the Great Recession in advance. Hence, for future research more work on appropriate transmission mechanims as well as on linking the models to data series that contain information about financial distress early on is important. Further, it needs to be tested whether medium-scale financial accelerator models also generate accurate forecasts outside of financial crises. The additional parameters to be estimated might increase estimation uncertainty which could potentially worsen forecasting accuracy outside of financial crises, when the financial accelerator mechanism is not important for predicting GDP accurately.

# References


Altavilla, C., R. Giacomini, and G. Ragusa (2013). Anchoring the Yield Curve Using Survey Expectations. Technical Report 9738, C.E.P.R. Discussion Papers.

Bernanke, B. S., M. Gertler, and S. Gilchrist (1999). The financial accelerator in a quantitative business cycle framework. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1, Chapter 21, pp. 1341–1393. Elsevier.

Blanchard, O. (2009). The state of macro. *Annual Review of Economics 1*(1), 209–228.

Buiter, W. (2009). The unfortunate uselessness of most 'state of the art' academic monetary economics. originally published in the Financial Times, 3 March, now available at: `https://mpra.ub.uni-muenchen.de/58407/1/MPRA_paper_58407.pdf`.

Canova, F. (2011). *Methods for applied macroeconomic research.* Princeton university press.

Carabenciov, I., I. Ermolaev, C. Freedman, M. Juillard, O. Kamenik, D. Korshunov, and D. Laxton (2008). A small quarterly projection model of the u.s. economy. IMF Working Paper No. 08/278.

Christiano, L., M. Eichenbaum, and C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy 113*(1), 1–45.

Christiano, L. J., M. Trabandt, and K. Walentin (2011). Introducing financial frictions and unemployment into a small open economy model. *Journal of Economic Dynamics and Control 35*(12), 1999–2041.

Del Negro, M., M. P. Giannoni, and F. Schorfheide (2015). Inflation in the Great Recession and New Keynesian models. *American Economic Journal: Macroeconomics 7*, 168–196.

Del Negro, M. and F. Schorfheide (2004). Priors from general equilibrium models for VARS. *International Economic Review 45(2)*, 643–673.

Del Negro, M. and F. Schorfheide (2013). DSGE model-based forecasting. In G. Elliott and A. Timmerman (Eds.), *Handbook of Economic Forecasting*, Volume 2, Chapter 2, pp. 57–140. Elsevier.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007). On the fit of new keynesian models. *Journal of Business and Economic Statistics 25*(2), 123–143.

Doan, T., R. Litterman, and C. Sims (1983). Forecasting and Conditional Projection Using Realistic Prior Distributions. Technical Report 1202, National Bureau of Economic Research, Inc.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews 3*, 1 – 100.

Dovern, J. and N. Jannsen (2017). Systematic errors in growth expectations over the business cycle. *International Journal of Forecasting 33*(4), 760–769.

Edge, R. M., M. T. Kiley, and J.-P. Laforte (2008). Natural rate measures in an estimated DSGE model of the U.S. economy. *Journal of Economic Dynamics and Control 32*, 2512–2535.

Fair, R. (2004). *Estimating how the macroeconomy works*. Harvard University Press.

Fair, R. (2018). The us model workbook. available at: `https://fairmodel.econ.yale.edu/wrkbook/zabwrk.pdf`.

Fair, R. C. (2019). Inflation in the great recession and new keynesian models: Comment.

Fisher, J. D. (2015). On the structural interpretation of the smets–wouters risk premium shock. *Journal of Money, Credit and Banking 47*(2-3), 511–516.

Gerali, A., S. Neri, L. Sessa, and F. M. Signoretti (2010). Credit and banking in a dsge model of the euro area. *Journal of Money, Credit and Banking 42*(s1), 107–141.

Giannone, D., M. Lenza, and G. Primiceri (2015). Prior selection for vector autoregressions. *The Review of Economics and Statistics 97*(2), 436–451.

Giffin, A., R. Urniezius, A. Giffin, and R. Urniezius (2014, 2). The Kalman Filter Revisited Using Maximum Relative Entropy. *Entropy 16*(2), 1047–1069.

Hamilton, J. D. (1994). *Time series analysis*, Volume 2. Princeton University Press.

Herbst, E. P. and F. Schorfheide (2015). *Bayesian estimation of DSGE models*. Princeton University Press.

Iacoviello, M. (2005). House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review 95*(3), 739–764.

Kimball, M. (1995). The quantitative analytics of the basic Neomonetarist model. *Journal of Money, Credit and Banking 27(4)*, 1241–1277.

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of Political Economy 105*(2), 211–248.

Kolasa, M. and M. Rubaszek (2015). Forecasting using DSGE models with financial frictions. *International Journal of Forecasting 31*(1), 1–19.

Krüger, F., T. Clark, and F. Ravazzolo (2015). Using Entropic Tilting to Combine BVAR Forecasts with External Nowcasts. Technical report, Verein für Socialpolitik / German Economic Association.

Krugman, P. (2009). How did economists get it so wrong? available at: `http://www.nytimes.com/2009/09/06/magazine/06Economic-t.html`.

Litterman, R. (1986). Forecasting with Bayesian Vector Autoregressions-Five Years of Experience. *Journal of Business & Economic Statistics 4*(1), 25–38.

Robbins, H. (1956). An Empirical Bayes Approach to Statistics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 157–163.

Roberts, G. O., A. Gelman, and W. R. Gilks (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab. 7*(1), 110–120.

Robertson, J., E. Tallman, and C. Whiteman (2005). Forecasting Using Relative Entropy. *Journal of Money, Credit and Banking 37*(3), 383–401.

Romer, P. (2016). The trouble with macroeconomics. Working Paper, available at: https://paulromer.net/wp-content/uploads/2016/09/WP-Trouble.pdf.

Smets, F. and R. Wouters (2007). Shocks and frictions in US business cycles: a Bayesian DSGE approach. *American Economic Review 97*(3), 586–606.

Stark, T. (2013). Spf panelists' forecasting methods: A note on the aggregate results of a november 2009 special survey. Working Paper, available at: https://www.philadelphiafed.org/-/media/research-and-data/real-time-center/survey-of-professional-forecasters/spf-special-survey-on-forecast-methods.pdf.

Stiglitz, J. E. (2015). Reconstructing macroeconomic theory to manage economic policy. In E. Laurent and J. L. Cacheux (Eds.), *Fruitful Economics: Papers in Honor of and by Jean-Paul Fitoussi*, pp. 20–49. Basingstoke, UK: Palgrave Macmillan.

Wieland, V., T. Cwik, G. J. Müller, S. Schmidt, and M. Wolters (2012). A new comparative approach to macroeconomic modeling and policy analysis. *Journal of Economic Behavior & Organization 83*(3), 523–541.

Wieland, V. and J. B. Taylor (2012). Surprising comparative properties of monetary models: Results from a new monetary model base. *Review of Economics and Statistics 94*(3), 800–816.

Wieland, V. and M. H. Wolters (2011). The diversity of forecasts from macroeconomic models of the u.s. economy. *Economic Theory 47*(2-3), 247–292.

Wieland, V. and M. H. Wolters (2012). Macroeconomic model comparisons and forecast competitions. available at https://voxeu.org/article/failed-forecasts-and-financial-crisis-how-resurrect-economic-modelling.

# Appendix

## A Bayesian VAR Model Description

### A.1 The Minnesota Prior

The Minnesota, or the Litterman prior, was introduced by Doan, Litterman and Sims 1983 and Litterman 1986. It is a conjugate normal prior, with a prior structure specifying the shape of the multivariate normal distribution of the parameters. The Minnesota prior sets a unit root structure for the univariate series, by centering the prior distribution of the own first lag coeffcient equal to one, while the prior mean of the other variables' lags are set to zero. It sets the prior covariance matrix of the coeffcients as diagonal, with the prior standard deviations having the following structure:

$$\sigma_{ij,l} = \begin{cases} \frac{\phi_0}{h(l)} & ,\text{if } i = j \ \forall\, l. \\ \phi_0 * \frac{\phi_1}{h(l)} * \left(\frac{\sigma_j}{\sigma_i}\right)^2 & ,\text{otherwise when } i \neq j,\ j \text{ is endogenous, } \forall\, l. \\ \phi_0 * \phi_2 & ,\text{for } j \text{ being exogenous.} \end{cases} \tag{2}$$

Where $h(l)$ is a deterministic function of the lag, $\phi_0$; $\phi_1$; $\phi_2$ are hyperparameters and $\frac{\sigma_j}{\sigma_i}$ is a scaling factor. The Minnesota prior's hyperparameters are the following:

1. $\phi_0$ represents the tightness on the variance of the first lag,

2. $\phi_1$ is the parameter governing the relative tightness on other variables,

3. $\phi_2$ represents the relative tightness of the exogenous variables,

4. $h(l)$ sets the relative tightness of the variance on lags other than the first lag. Usually harmonic decay is assumed like $h(l) = l^{\phi_3}$, where $\phi_3$ is positive.

For the standard errors $\sigma_i$ consistent estimates are used. (Canova, 2011, p. 355-356.)

The parameter governing the relative tightness of the own lagged variances introduces a shrinkage in to the Minnesota prior, i.e. the prior the variance of lagged coefficients around the zero mean decreases as the lag length increases. This shrinkage property combined with the relative tightness that makes lags of other variables to contain less information than own lags, i.e. $\phi \leq 1$. This specifies the whole distribution of lagged variables. The Minnesota prior is a normal conjugate prior, that can be implemented using dummy observations.

Multiple authors have shown that VARs with a Minnesota prior produce better forecasts than univariate ARIMA models or traditional multivariate simultaneous equations. "Therefore, it is not surprising that BVARs are routinely used for short-term macroeconomic forecasting in Central Banks and international institutions." (Canova, 2011, p. 358.)

### A.2 GLP prior

Giannone et al. (2015) adopt a hierarchical model to "make inference about the informativeness of the prior distribution" (Giannone et al., 2015, p. 437.) of the BVAR . They argue maximizing the posterior of the hyperparameters of a model with conjugate priors captured by the likelihood function $p(y|\theta,\gamma)$ , prior $p(\theta|\gamma)$, and prior over hyperparameters $p(\gamma)$, corresponds to maximizing the one-step-ahead out-of-sample forecasting ability of the model.

In a hierarchical model, they use a combination of the conjugate priors most commonly used in the literature. They combine the Minnesota, sum-of-coeffcients and dummy-initial-observation priors. (Giannone et al., 2015, p.440.) This methodology is also referred to Empirical Bayesian approach, estimating the hyperparameters by maximizing the marginal likelihood (ML) (Robbins, 1956, p.3.). Giannone, Lenza and Primiceri focus on the set of hyperparameters that govern the informativeness of the combination of the three priors to create, by maximizing the posterior of the hyperparameters, a procedure that automatically selects the optimal tightness to optimize the one-step-ahead out-of-sample forecasting ability of the BVAR.

They collect the set of hyperparameters $\lambda$, $\mu$, $\delta$, and $\psi$ which they treat as additional parameters. The equivalent notation for these parameters is the following:

- From the Minnesota prior they focus on the hyperparameter $\phi_0$, denoted by $\lambda$. For the lag penalty $h(l)$ a quadratic function is chosen, i.e. $h(l) = l^2$ . In their representation of the Minnesota prior, the relative tightness on other variables ($\phi_1$) is set to 1.
- For the "sum-of-coeffcients" prior that was originally proposed by Doan, Litterman, and Sims 1983, they calculate the dummy observations as follows:

$$\begin{aligned} \underset{n\times n}{y^+} &= diag\left(\frac{\bar{y_0}}{\mu}\right) \\ \underset{n\times(1+np)}{x^+} &= \left[\underset{n\times 1}{0}, y^+, ..., y^+\right] \end{aligned} \tag{3}$$

  where $\bar{y_0}$ is the vector containing the average of the first $p$ observations for each variable. "The prior implied by these dummy observations is centered at 1 for the sum of coefficients on own lags for each variable, and at 0 for the sum of coeffcients on other variables' lags." (Giannone et al., 2015, p.440.) The hyperparameter $\mu$ is of interest as it controls the variance, if $\mu \leftarrow \infty$ is equivalent to an uninformative prior, while "$\mu \leftarrow 0$ implies the presence of a unit root in each equation and rules out cointegration." (Giannone et al., 2015, p.440.)
- The "dummy-initial-observation" prior is implemented by following using the dummy observation:

$$\begin{aligned} \underset{1\times n}{y^{++}} &= \frac{\bar{y_0}'}{\delta} \\ \underset{n\times(1+np)}{x^{++}} &= \left[\frac{1}{\delta}, y^{++}, ..., y^{++}\right] \end{aligned} \tag{4}$$

  This prior implies that a martingale property for the asymptotic of the sample. The hyperparameter $\delta$ of interest, as it controls the informativeness of the prior. Similarly to the "sum-of-coefficients" prior as $\delta \leftarrow \infty$ the prior becomes uninformative. In contrast to the previous case, as "$\delta \leftarrow 0$, all the variables of the VAR are forced to be at their unconditional mean." (Giannone et al., 2015, p.440.)

The GLP hyperpriors for $\lambda$, $\mu$, $\delta$ are chosen to be Gamma densities with mode equal to 0.2, 1 and 1 and standard deviations equal to 0.4, 1 and 1 respectively. Finally, for the prior mean of the main diagonal of $\Sigma$, the hyperprior for $\frac{\psi}{d-n-1}$ Giannone, Lenza and Primiceri choose an Inverse-Gamma with scale and shape equal to $0.02^2$.

They show that the ML can be expressed a sum of two terms, that capture the trade-off between model fit and complexity.

The first term, the degrees of freedom weighted difference between the log-determinant of the prior and posterior mode (or mean) of the residual covariance matrix, captures the in-sample fit of the model. The in-sample fit increases as the informativeness of the priors decreases.

The second term, the difference between the log-determinant of the prior and posterior variance, induces a penalty for model complexity. A less informative prior penalizes the ML as the distance between the prior and posterior variance of the coefficients increases.

The GLP is implemented using the authors' code, that relies on a Markov chain Monte Carlo algorithm (MCMC) with Metropolis Hastings algorithm for the joint posterior density simulation from a Gaussian proposal distribution.[2]

Giannone, Lenza and Primiceri document that based on US data "the forecast accuracy of the BVARs does not deteriorate when increasing the scale of the model and sometimes even improves substantially (as it is the case for inflation)." (Giannone et al., 2015, p.442.)

### A.3 BVAR Conditioning - Entropic tilting

Conditioning enables the researcher to incorporate information about future variables into the forecasts. We employ entropic tilting to combine information content of the model to forecasters. In comparison to standard conditioning techniques [3] relative entropy based entropy tilting provides important benefits. First, it is agnostic about model identification, allowing to account for model and parameter uncertainty. Second, instead of forcing unconditional forecast to meet their fixed conditioning values, it delivers a distribution centered around the conditioned value. Third, it enables to obtain a new predictive, tilted forecast density that is as close as possible to the unconditional forecast density while satisfying the conditioning restriction. Finally, most importantly for our methodology, the measure of distance, relative entropy, that is minimized provides an information content based interpretation,and is linked to logarithmic scoring rule based forecast density evaluation. Entropic tilting distorts the model's unconditional predictive density to meet the moment condition while maximizing the information content between the conditional and unconditional forecasts.

Robertson, Tallman, and Whiteman Robertson et al. (2005) first introduced entropic tilting into macroeconomic forecasting, by conditioning short term policy rates from a small scale BVAR.

Altavilla et al. (2013) made use of conditioning to anchor yield curve forecasts to survey-based forecasts of short-term interest rates, finding that expectation formation significantly improves the model fit.

Krüger et al. (2015) studied the impact of nowcast conditioning on BVAR forecast performance, finding that it improves forecast performance at two and three quarters ahead, and has limited impact beyond. Tilting towards not only mean but variances of forecasts yields small gains in density forecast accuracy. Furthermore the authors document that conditional forecasts on nowcasts of all variable versus on each variable separately perform comparably, implying that a tilting at the system level does not come at the cost in terms of forecast performance. In our setup we could also verify this finding. Furthermore we found while nowcast conditioning alleviates the forecast accuracy of the short end Following their results we condition on nowcasts in our model, however instead of conditioning on higher moments of forward looking variables, we study the impact of conditioning has on higher moments of the conditioned forecast density.

In the following we briefly introduce entropic tilting used for forecasting. Consider a model forecast density $f(\hat{y}|y)$, for which the analytically form of the density function might not be known, but it is possible to sample from if numerically on the computer. Therefore the finite approximation to the density can be characterized wit a series of $N$ draws and weights $(\pi_1, \pi_2, ...\pi_N)$ attributed to each draw. The idea

[2] An important detail for our approach is the calibration of the jump size. It is desired for a asymptotically consistent sampler over a multivariate normal distribution to have an acceptance rate around 23.4 percent in order to explore the posterior properly (Roberts et al., 1997). Since our goal is the uncertainty captured by the forecast density, we pay careful attention to the calibration, as a too high acceptance rate would undersample the tails, while a too low would require substantially longer chains to be proper.

[3] like the one proposed by Waggoner and Zha (1999)

behind entropic tilting is the following: in order to incorporate additional information into the unconditional forecast density $f(\hat{y}|y)$ one needs to a new distribution $f^{*}(\hat{y}|y)$, the conditional density. With the entropic tilting $f^{*}(\hat{y}|y)$ will based on the same draws of the predictive density $\hat{y^{(i)}}_{i=1}^{N}$ but attributes alternative weights $(\pi_1^*, \pi_2^*, ...\pi_N^*)$ to them.The optimal wights are those that minimize the Kullback-Leibler divergence between the unconditional and conditional forecast density, where the latter incorporates new information in the form of constraints on the family of allowed posteriors. Formally:

$$\begin{aligned} minKL(f^{*}(\hat{y}|y) \rightarrow f(\hat{y}|y)) &= \int f^{*}(\hat{y}|y) \cdot log_2 \left( \frac{f^{*}(\hat{y}|y)}{f(\hat{y}|y)} \right) d\hat{y} \\ \text{subject to: } \mathbb{E}g(\hat{y}) &= \bar{g} \end{aligned} \tag{5}$$

where $f$ is the unconditional, $f^{*}$ is the conditional forecast distribution for $\hat{y} \in \mathscr{R}^{p}, p \geq 1$ over horizons $T+1 : T+h$, furthermore $g : \mathscr{R}^{p} \rightarrow \mathscr{R}^{\mathrm{m}}$, i.e. a function, mapping the for variables to conditions, and $\bar{g} \in \mathscr{R}^{\mathrm{m}}$ representing the conditions.

Giffin and Urniezius (2014) show explicitly and analytically, that in a linear state space system with multivariate Normal innovations maximizing the relative entropy, i.e. minimizing the Kullback-Leibler divergence, leads to the Kalman filter. This result is pivotal, as it highlights that the Kalman filter is the dual of the entropy filter. It also implies that the Kalman filter will deliver optimal entropic tilting for the GLP prior based BVAR, as it has a multivariate normal posterior. With the latter we have established that not only the prior selection, but the forecast conditioning is designed to incorporate the maximum information, thus it is informationally consistent.

# B DSGE Model Descriptions

## B.1 The SW07 Model

The Smets and Wouters (2007) model is a medium-scale closed economy DSGE-Model estimated for the US economy with Bayesian techniques. The model features a deterministic growth rate driven by labor-augmenting technological progress. Following the work of Christiano et al. (2005) it contains both nominal an real frictions. The primary frictions in the model are the nominal frictions affecting labour and goods market and the real frictions of the capital markets in form of investment adjustment costs.

Households make consumption and savings decisions given investment adjustments costs. Households maximize expected utility over an infinite horizon, given their habit formation. Capital faces capital utilization costs that will affects its use of intensity. Intermediate firms produce differentiated goods using labour and capital as input, and face Calvo type nominal rigidities. Labour services are aggregated by a union facing nominal Calvo wage rigidities. Both wage and product pricing is subject to partial indexation to lagged inflation.

- Aggregate Demand: Households maximize their lifetime utility, where the utility function is non-separable in consumption and leisure, subject to an inter-temporal budget constraint. Smets and Wouters (2007) include external habit formation to make the consumption response in the model more persistent. Households own firms, rent capital services to firms and decide how much capital to accumulate given certain capital adjustment costs. They additionally hold their financial wealth in the form of one-period, state-contingent bonds. Exogenous spending follows a first-order auto-regressive process with an iid-normal error term and is also affected by the productivity shock.

- Aggregate Supply: The final goods, which are produced under perfect competition, are used for consumption and investment by the households and by the government. The final goods producer maximizes profits subject to a Kimball (1995) aggregator of intermediate goods, which introduces monopolistic competition in the market for intermediate goods and features a non constant elasticity of substitution between different intermediate goods, which depends on their relative price. A continuum of intermediate firms produce differentiated goods using a production function with Cobb-Douglas technology and fixed costs and sell these goods to the final-good sector. They decide on labor and capital inputs, and set prices according to the Calvo model. Labor is differentiated by a union using the Kimball aggregator, too, so that there is some monopoly power over wages, which results in an explicit wage equation. Labor packers buy the labor from the unions and resell it to the intermediate goods producer in a perfectly competitive environment. Sticky wages à la Calvo are additionally assumed. The Calvo model in both wage and price setting is augmented by the assumption that prices that can not be freely set, are partially indexed to past inflation rates.

- Shocks: The model is subject to seven exogenous shock process, beyond the standard total factor productivity, monetary policy, investment specific technology, exogenous spending, the model features a risk premium shock and wage and price markup shocks with a MA structure. The latter property introduces anticipated, news shocks for both the regular and the wage Phillips curve. Finally, the risk premium shock was introduced to the linearized inter-temporal consumption-Euler equation of the household, and greatly improved model fit. The nature of the shock is subject to interpretation: initially it was considered to capture a preference shock, lately Fisher (2015) showed that it can be reinterpreted as a liquidity shock, i.e. a demand shock for safe and liquid assets.

- Monetary policy is described by a Taylor type rule, with interest rate smoothing and the reaction to inflation- and output gap, the former defined as the deviation from the estimated steady state inflation, the latter as the distance to the flex price economy, following the Taylor principle.

- Original Estimation: The model is estimated for the U.S. with Bayesian techniques for the period 1966:1–2004:4 using seven key macroeconomic variables: real GDP, consumption, investment, the GDP deflator, real wages, employment and the nominal short-term interest rate. Both real consumption and investments are deflated using the GDP deflator. The hours variable is defined as average weekly hours of all persons in the non-farm business sector times total civilian employment[4].

The Smets Wouters model became the standard workhorse model for monetary policy analysis. It served as a basis for newer generations of DSGE models that followed.

## B.2 The DNGS15 Model

Del Negro et al. (2015) build a medium-scale New Keynesian model that can predict a sharp contraction in economic activity along with a protracted but relatively modest decline in inflation, following the Great Recession. They use a standard DSGE model like in Smets and Wouters (2007) enriched with financial frictions and a time-varying target inflation rate. The model is estimated using an endogenous steady state, that is around a balanced growth path subject to the financial friction.

- Aggregate Demand: As in Smets and Wouters (2007), households maximize a nonseperable utility function with two arguments (goods and labor effort) over an infinite life horizon, subject to an intertemporal budget constraint. Preferences for consumption are subject to habit persistence. They supply labor monopolistically and wage stickiness is introduced via the Calvo framework.

- Aggregate Supply: Monopolistically competitive firms produce intermediate goods, which a competitive firm aggregates into a single final good that is used for both consumption and investment. The intermediate goods firms decide on labor and capital inputs, and set prices according to the Calvo model.

- Financial Sector: Building on the work of Bernanke et al. (1999) a financial intermediary, capital producers and entrepreneurs are introduced in the model in addition to the intermediate and final goods firms as in Smets and Wouters (2007). Financial frictions come into play by the presence of entrepreneurs and the financial intermediary. Banks collect deposits from households and lend to entrepreneurs who use these funds as well as their own wealth to acquire physical capital, which is then rented to intermediate goods producers. Entrepreneurs are subject to idiosyncratic disturbances that affect their ability to manage capital which leads to the costly state verification framework as in Bernanke et al. (1999) and gives raise to a spread, above the risk-free rate. This spread may vary as a function of the entrepreneurs leverage and their riskiness.

- Shocks: A preference shock, a financial friction shock, a total factor productivity shock, an investment specific technology shock, a government spending shock, an inflation target shock, a monetary policy shock, a wage and price mark-up shock. Similar to the Smets and Wouters (2007) the mark-up shocks feature a one period ahead anticipated news shock and thus have a VARMA(1,1) structure.

[4] Fair (2019) argues that these definitions might be erroneous, as the nominal series should be deflated by the respective price deflator, while the non-farm hours excludes farm and government workers.

- Original Estimation: The model is estimated using Bayesian methods on quarterly U.S. data for the period 1964:Q1 2008:Q3 using 8 key macroeconomic variables: output growth, consumption growth, investment growth, inflation, wages, hours, interest rate, credit spread.

### B.3 The FRBEDO Model

Edge et al. (2008) estimate a model featuring two production sectors, which differ in their pace of technological progress.

The model is estimated using Bayesian techniques to explain the long-run and cyclical properties of related data in the US.

- Aggregate Demand: There are four components of aggregate demand: consumer non-durable goods and non-housing services (sold to households), consumer durable goods, residential capital goods, and non-residential capital goods. Consumer non-durable goods and non-housing services and residential capital goods are purchased (by households and residential capital goods owners, respectively) from the first of economys two final goods producing sectors.

- Aggregate Supply: The model possesses two final goods, which are produced in two stages by intermediate- and then final-goods producing firms. The first sector represents the economys slow growing sector, it amounts for most of the output as it combines consumption goods and services account for most of its output and it is produced by the business and institutions sector of the economy.

  Consumer durable goods and non-residential capital goods are purchased (by consumer durable and residential capital goods owners, respectively) from the second sector. It presents the economys fast growing sector, so denoted because its output is capital goods and it is produced by the business sector of the economy.

  Intermediate-goods producing firms are monopolistically competitive and the two final goods are aggregated by two competitive firms. The final products of the first sector are purchased by households and residential capital owners. While the output of the second sector are purchased by non-residential capital owners and durable consumer good owners.

  The distinction between the goods originates from the assumption that residential capital and consumer durables capital are rented to households, while non-residential capital is rented to firms. In addition to consuming the non-durable goods and non-housing services that they purchase, households also supply labor to the intermediate goods-producing firms in both sectors of the economy. The model is built around a stationary un-modeled output process following an AR(1) in growth rates.

- Monetary Policy: sets the short term rate in accordance with an Taylor-type interest-rate feedback rule, featuring interest rate smoothing and a target. The central banks target nominal interest rate, depends on GDP growth relative to steady-state growth, the acceleration of GDP growth, GDP deflator relative to target, and the acceleration of GDP deflator.

- Shocks: The model exhibits 14 shocks, and 9 measurement errors. The shocks can be categorized into a monetary policy shock, 3 elasticity of substitution shocks, 2 type of technology shocks, 3 efficiency of investments shocks, 4 preference shock and an exogenous output shock.

The shocks are the following: monetary policy shock, shock to the elasticity of substituion between the differentiated intermediate goods inputs for sector one, shock to the elasticity of substituion between the differentiated intermediate goods inputs for sector two, shock to the elasticity of substitution between the differentitated labor inputs, capital specific technology growth shock, economy-wide technology growth shock, shock to efficiency of investment in non-residential capital, shock to efficiency of investment in residential capital, shock to efficiency of investment in consumer durable goods, shock to preferences over non-durables and non-housing services, shock to preferences over durables, shock to preferences over residential capital, shock to preferences over leisure and finally a shock to unmodelled exogenous output growth.

- Original Estimation: The model is estimated using Bayesian methods on quarterly U.S. data for the period 1983:Q1 2005:Q4 using 11 key macroeconomic variables: Nominal gross domestic product; Nominal consumption expenditure on non-durables and services excluding housing services; Nominal consumption expenditure on durables; Nominal residential investment expenditure; Nominal business investment expenditure, which equals nominal gross private domestic investment minus nominal residential investment; GDP price inflation; Inflation for consumer non-durables and non-housing services; Inflation for consumer durables; Hours, which equals hours of all personsin the non-farm business sectorfrom the Bureau of Labor Statistics; Wage inflation, which equals compensation per hour in the non-farm business sector from the Bureau of Labor Statistics; and the federal funds rate (Edge et al., 2008, p.11).

### B.4 The NKBGG Model

Bernanke et al. (1999) introduce credit market imperfections into an otherwise standard New Keynesian model with capital and show that these financial frictions contribute to propagate and amplify the response of key macroeconomic variables to nominal and real shocks. An agency problem arises due to asymmetries of information in borrower-lender relationships. The economy is inhabited by three types of agents, risk-averse households, risk-neutral entrepreneurs and retail firms.

- Aggregate Demand: Households gain utility from consumption, leisure and real money balances. Household optimization results in a standard dynamic IS equation. Entrepreneurs use capital and labor to produce wholesale goods that are sold to the retail sector. Each period, entrepreneurs have to accumulate capital that becomes available for production in the subsequent period. Entrepreneurs have to borrow from households via a financial intermediary to finance capital purchases. Since the financial intermediary has to pay some auditing costs to observe the idiosyncratic return to capital, an agency problem arises. The optimal contract leads to an aggregate relationship of the spread between the external finance costs and the risk-free rate and entrepreneurs' financial conditions represented by the leverage ratio.

- Aggregate Supply: Retail firms act under monopolistic competition. They buy wholesale goods produced by entrepreneurs in a competitive market and differentiate them at zero cost. Price stickiness is introduced via the Calvo framework. Bernanke et al. (1999) assume that reoptimizing firms have to set prices prior to the realization of shocks in that period, so that previous period's expectations of the output gap and future inflation enter the New Keynesian Phillips curve.

- Shocks: The model exhibits a technology shock, a demand shock and the common monetary policy shock. Since we have no information about the variances of the shock terms, we set all shock variances equal to zero.

- Calibration/Estimation: The model is originally calibrated at quarterly frequency.

## B.5 The DSSW07 Model

Del Negro et al. (2007) models is a slightly modified version of Christiano et al. (2005) model, and thus is very similar to the Smets and Wouters (2007) model. DSSW07 model is originally a presented as a "hybrid model". In essence it is a linear combination of an unrestricted VAR model of the data and the VAR implied by the estimated DSGE model. The parameter $\lambda$ is the scalar that regulates the combination, and is estimated searching for the $\hat{\lambda}$ parameter that results in the highest marginal likelihood for the data. If $\hat{\lambda} \to \infty$ then the DSGE model is preferred, while a low $\hat{\lambda}$ indicates that the unrestricted VAR is a better fit.

The conclusion of the paper are the following: First, the posterior odds of a DSGE versus a posterior of a VAR based on diffuse priors do not provide a robust assessment of fit, small changes in the sample can deliver contrasting results. The hybrid DSGE-VAR($\hat{\lambda}$) method delivers much less sensitive results. Second, the paper documents evidence for misspecification of the DSGE model based on altering specifications. Finally, the paper documents that accounting for misspecification, by optimally relaxing the cross variable restrictions does not alter qualitatively and quantitatively the responses of variables to technology and monetary policy shocks. In terms of forecast comparison Del Negro et al. (2007) documents that the DSGE model performs well for nominal variables, inflation and interest rate, but performs very poorly for output and investments. The long run co-integrating restrictions of the proportional growth rate explains this finding.The DSGE-VAR($\hat{\lambda}$) inherits these long run restrictions and thus performs worse than the unrestricted VAR. On the other hand the unrestricted VAR performs substantially worse than both the DSGE-VAR($\hat{\lambda}$) and the DSGE for inflation.

The DSSW07 model is a simplified version of the Smets and Wouters (2007). It displays all the standard set of nominal and real rigidities, like consumption habits, investment adjustment costs, capacity utilization costs, Calvo wage and price stickiness with partial indexation. As in Smets and Wouters (2007) government spending is exogenous and monetary policy follows a Taylor type rule. The model features seven stochastic disturbances: labor augmenting technology that creates the common underlying cointegration driving long run real wages, consumption, investment, capital and output. The remaining shocks are stationary: time preference shock, relative price of investment, disutility of labor, price markup, government purchases and monetary policy. (Kolasa and Rubaszek, 2015)

The model is estimated on the same set of seveb variables as the Smets and Wouters (2007) model.

## B.6 The KR15_FF Model

The DSSW07FF model is an extension of the DSSW07 model with financial frictions a'la Bernanke et al. (1999) presented in Kolasa and Rubaszek (2015). Please see details of the financial contract in the section describing the FF-BGG99 model (B.4). The main mechanism of the financial accelerator is the costly state verification problem. Thus the DSSW07FF model includes the external finance premium, driven by two additional shocks affecting the standard deviation of idiosyncratic risk facedby entrepreneurs and their survival rate.

- Aggregate Demand: Households maximize their lifetime utility, where the utility function is non-separable in consumption and leisure, subject to an inter-temporal budget constraint and external habit formation. Households own firms, rent capital services to firms and decide how much capital to accumulate given certain capital adjustment costs. Exogenous (government) spending follows a first-order auto-regressive process.

Entrepreneurs operate capital and labor to produce wholesale goods that are sold to the retail sector.

- Aggregate Supply: The final goods, which are produced under perfect competition, are used for consumption and investment by the households and by the government. The final goods producer maximizes profits subject to a Kimball (1995) aggregator of intermediate goods, which introduces monopolistic competition in the market for intermediate goods and features a non constant elasticity of substitution between different intermediate goods, which depends on their relative price. A continuum of intermediate firms produce differentiated goods using a production function with Cobb-Douglas technology and fixed costs and sell these goods to the final-good sector. They decide on labor and capital inputs, and set prices according to the Calvo model. Labor is differentiated by a union using the Kimball aggregator, too, so that there is some monopoly power over wages, which results in an explicit wage equation. Labor packers buy the labor from the unions and resell it to the intermediate goods producer in a perfectly competitive environment. Sticky wages à la Calvo are additionally assumed. The Calvo model in both wage and price setting is augmented by the assumption that prices that can not be freely set, are partially indexed to past inflation rates.
- Financial Friction: Agency problem between entrepreneurs and the financial intermediaries, e.g. banks in form of costly state verification resulting in an external finance premium over the short term rate, that entrepreneurs have to face to acquire financing. Each period, entrepreneurs have to accumulate capital that becomes available for production in the subsequent period. Entrepreneurs have to borrow from households via a financial intermediary to finance capital purchases. Since the financial intermediary has to pay some auditing costs to observe the idiosyncratic return to capital, an agency problem arises. The optimal contract leads to an aggregate relationship of the spread between the external finance costs and the risk-free rate and entrepreneurs' financial conditions represented by the leverage ratio.
- Shocks: A total factor productivity shock, a risk premium shock, an investment-specific technology shock, a wage and a price mark-up shock, two policy shocks: the common fiscal policy shock entering the government spending equation and the common monetary policy shock and finally two shocks affecting the financial friction: the standard deviation of idiosyncratic risk faced by entrepreneurs and their survival rate.
- Calibration/Estimation: The model is estimated on the observables of the Smets and Wouters (2007) model, i.e. output growth, consumption growth, investment growth, inflation,wages, hours, interest rate, augmented with a credit spread and loan growth observables. The latter two are defined as log difference of credit market instruments; liabilities of the non-farm non-financial business sector and as the between the industrial BBB corporate bond yield, back-casted using BAA corporate bond yields, and the federal funds rate (Kolasa and Rubaszek, 2015).

## B.7 The KR15_HH Model

The DSSW07HH model is an extension of the DSSW07 model with financial frictions following Iacoviello (2005), i.e. collateral constraint in the housing market. Debt contracts are written in nominal terms and some agents face collateral constraints tied to housing values. This gives rise to an accelerator effect for demand shocks and a decelerator effect for supply shocks.

- Aggregate Demand: In contrast to the DSSW07 model there are two types of households, the patient and the impatient ones, i.e. they discount the future differently. The two types of agents differ in their rate of time preference: the impatient household discounting the future more heavily. This specification induces the impatient household to face borrowing constraints, consistent with standard lending criteria used in the mortgage market where the borrowing is limited to a fraction of the housing value. Financial intermediaries take deposits from savers and lend them to borrowers. ”The financial intermediation between patient and impatient households is conducted by imperfectly competitive banks, which accept deposits at thepolicy rate and offer loans at a rate reflecting their monopolistic power” (Kolasa and Rubaszek, 2015, p.3.). The interest spread of lending over policy rate depends on loan to value ratios, mark-up charged over funding.

- Aggregate Supply: Entrepreneurs produce a homogeneous intermediate good using a Cobb-Douglas technology with labor from both types of households, capital and real estate as inputs. Housing and variable capital are subject to adjustment costs.

- Shocks: A total factor productivity shock, a risk premium shock, an investment-specific technology shock, a wage and a price mark-up shock, two policy shocks: the common fiscal policy shock entering the government spending equation and the common monetary policy shock and finally four additional shocks affecting the financial friction. These shocks are shocks to the housing weights in utility for each agent, loan-to-value ratio, relative price of residential investment and markups in the banking sector (Kolasa and Rubaszek, 2015, p.3.).

- Calibration/Estimation: The model is estimated on the observables of the Smets and Wouters (2007) model, i.e. output growth, consumption growth, investment growth, inflation,wages, hours, interest rate, augmented with series on residential investment, mortgage loans, house prices and the spread on mortgage loans. The latter four relate to the financial accelerator.

# C DSGE Nowcasting - Using the Kalman filter

This section aims to present a brief discussion of the specification and estimation of a DSGE nowcast based on its linearized state space form. For a detailed discussion of the topic we instruct the reader to consult Hamilton (1994) for the state space models and the Kalman filter, and Herbst and Schorfheide (2015) for the details on its application for Bayesian DSGE estimation.

## C.1 Specification

A unique stable solution of the DSGE model in its most general specification is a difference equation [5]:

$$S_t = T(\theta)S_{t-1} + R(\theta)\varepsilon_t, \tag{6}$$

where $S_t$ is the (endogenous) state vectors of length $n$, $\varepsilon_t$ is the vector of exogenous shocks, i.e. innovations. The disturbances $\varepsilon_t$ are assumed to be serially independent, with unit variances and zero covariance. $T(\theta)$ is the state transition matrix, that is a nonlinear function of the DSGE parameters, $R(\theta)$ is the mean-squared error matrix of the states, once again a nonlinear function of the DSGE parameters. In other words a DSGE's unique stable solution can be written as a first order vector autoregressive model. We refer to this set of equations as the ”state” equations. The state equations are linked through the observation equations to the data:

$$Y_t = H'S_t + \xi_t, \tag{7}$$

where $Y_t$ is the vector of observables, and $\xi_t$ is the measurement error. In practice the measurement error is usually set to zero, and in the paper we also do so, for the current follows we drop it. The matrix $H$ is called the emission matrix, that is usually a matrix of zeros and ones, selecting from the states the variables linked to the data.

We can then define the conditional forecast of the state given information set $\Omega_t$, available at period $t$, that is the expectation of $S_t$ given its filtration:

$$S_{t|t-1} = E\left[S_t|S_{t-1}, S_{t-2}...S_0\right], \tag{8}$$

and the means square error or covariance matrix:

$$P_{t|t-1} = E\left[(S_t - S_{t-1})(S_t - S_{t-1})'|S_{t-1}, S_{t-2}...S_0\right]. \tag{9}$$

Using the state transition equation of the DSGE and the means square error matrix we can write the conditional forecast of the state as:

$$S_{t|t-1} = T(\theta)S_{t-1|t-1}P_{t|t-1}. \tag{10}$$

With the help of the observation equations the one period ahead forecast of the data given the states

[5] For notational purposes we use the Sims form representation, acknowledging that other solution techniques give rise to the same structural form.

follows:

$$Y_{t|t-1} = H' S_{t|t-1}, \tag{11}$$

and its variance:

$$\Sigma_{t|t-1} = E\left[\left(Y_t - Y_{t|t-1}\right)\left(Y_t - Y_{t|t-1}\right)'\right]. \tag{12}$$

Given these definitions and the initialization[6] we can define the likelihood of the model, i.e. the probability distribution of the data given parameters of the model:

$$\mathscr{L}(Y_t|\theta) = (2\pi)^{-n/2} det\left(\Sigma_{t|t-1}^{-1}\right)^{1/2} exp\left[-\frac{1}{2}\left(Y_t - Y_{t|t-1}\right)' \Sigma_{t|t-1}^{-1}\left(Y_t - Y_{t|t-1}\right)\right] \tag{13}$$

We can then use the Kalman filter to compute the nowcasts of the state vector and its means squared error matrix as

$$S_{t|t} = S_{t|t-1} + P_{t|t-1} H \Sigma_{t|t-1}^{-1}\left(Y_t - Y_{t|t-1}\right), \tag{14}$$

$$P_{t|t} = P_{t|t-1} + P_{t|t-1} H \Sigma_{t|t-1}^{-1}\left(Y_t - Y_{t|t-1}\right). \tag{15}$$

Where the term $P_{t|t-1} H \Sigma_{t|t-1}^{-1}$ is usually called the Kalman gain.

## C.2 Nowcasts

With the help of the Kalman filter's nowcast for the current state it follows that the nowcast of the data is:

$$Y_{t|t} = H' S_{t|t} \tag{16}$$

In particular consider the case for the three equation DSGE, where we have also three observables. Conditioning on one of them, e.g. the first, then means that for that variable we have the observation in $Y_t$, and thus the nowcast of the sate is updated given the Kalman filter:

$$Y_{t|t}(1) = H'(n,1)\left(S_{t|t-1} + P_{t|t-1} H \Sigma(1,1)_{t|t-1}^{-1}\left(Y_t(1) - Y_{t|t-1}(1)\right)\right), \tag{17}$$

whereas for the other variables the nowcast will only be affected by the one step ahead forecast error made on the first variable $\left(Y_t(1) - Y_{t|t-1}(1)\right)$:

$$Y_{t|t}(2:3) = H'(n,2:3)\left(S_{t|t-1} + P_{t|t-1} H \Sigma(1,1)_{t|t-1}^{-1}\left(Y_t(1) - Y_{t|t-1}(1)\right)\right). \tag{18}$$

In other words, the nowcast of the not conditioned variables will be the one step ahead prediction, based on the past of the variable and updated with the impact of the forecast error made in the conditioning dimension.

[6] The Kalman filter starts by initializing the states $S_{1|0}$ at 0 and finding the initial conditions for $P_{0|0}$, by solving the Lyapunov equation $0 = T(\theta) P_{0|0} T(\theta) - P_{0|0} + RR'$.

## D Data Collection

Table 1 displays the description of raw data that are collected for obtaining observed variables in different models. The variables are grouped by their updating frequencies. For each variable, the table lists the name, description, units, whether or not being seasonally adjusted, source, and the vintage dates in each quarter (if available). Some variables have similar meanings (e.g., both AWHNONAG and PRS85006023 represent average weekly hours), but they are used to compute different observed variables in different models. For reference, the deadlines for professional forecasters to submit their questionnaires in each quarter are listed on the lower right corner.

Data are collected from the following sources:

- ALFRED: Archival Federal Reserve Economic Data (`https://alfred.stlouisfed.org/`)
- BB: Bloomberg Professional Services (`https://www.bloomberg.com/professional/`)
- CB: Census Bureau (`https://www.census.gov/`)
- FHFA: Federal Housing Finance Agency (`https://www.fhfa.gov/`)
- FRB: Federal Reserve Board (`https://www.federalreserve.gov`)
- FRED: Federal Reserve Economic Data (`https://fred.stlouisfed.org/`)
- RTDSM: Real-Time Data Set for Macroeconomists (`https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/`)

Notes:

1. For variables that are updated in daily frequency, their values from the beginning of a quarter until the SPF deadlines in that quarter will be used.
2. The publication dates of *MortgageRates* are not disclosed by the Federal Housing Finance Agency. Since this series is updated in monthly frequency, we assume that only its value in the first month of a quarter can be observed before the SPF deadline.
3. The base year for *P* is not specified in the RTDSM. Nevertheless, it has no influence on the data transformation process, because only the log-difference of this variable is used in the data transformation process.
4. The exact publication dates of *ROUTPUT* and *P* are not recorded in the RTDSM. According to the user guide, the values of these two variables were known in the middle of each quarter.
5. In 2008:III, the publication date of *COMPNFB* is one day later than the SPF deadline. However, the main results does not change even if we treat the value of *COMPNFB* in the second quarter of 2008 as missing.
6. The publication dates of *PRS85006023*, *BOGZ1FL144104005Q*, and *HMLBSHNO* are not recorded in the ALFRED. We assume that only their values in the previous quarter can be observed before the SPF deadline.
7. The publication dates of *HousePriceIndex* are not disclosed by the Census Bureau. We assume that only its value in the previous quarter can be observed before the SPF deadline.
8. This index is constructed by Carabenciov et al. (2008), based on the Senior Loan Officer Opinion Survey on Bank Lending Practices carried out by the Federal Reserve Board. Notice that the vintage dates in the table are the ”last update” dates but not ”initial release” dates of the Survey, because the latter kind of dates are not disclosed by the FRB.

Table 1: Description of Raw Data

| Name | Description | Units | Seasonally Adjusted | Source | Vintage Dates 2008:III | 2008:IV | 2009:I | 2009:II |
|---|---|---|---|---|---|---|---|---|
| Variables updated in daily frequency | | | | | | | | |
| EFFR | Effective Federal Funds Rate | % | | FRED | Same as the SPF deadlines[1] | | | |
| DBAA | Moody's Seasoned BAA Corporate Bond Yield | % | | FRED | Same as the SPF deadlines[1] | | | |
| DGS10 | 10-Year Treasury Constant Maturity Rate | % | | FRED | Same as the SPF deadlines[1] | | | |
| BBB1Y | 1-Year Industrial BBB Corporate Bond Yield | % | | BB | Same as the SPF deadlines[1] | | | |
| Variables updated in monthly frequency | | | | | | | | |
| CE16OV | Employment Level | Thousands of Persons | • | ALFRED | 08-01 | 11-07 | 02-06 | 05-08 |
| CNP16OV | Population Level | Thousands of Persons | | ALFRED | 08-01 | 11-07 | 02-06 | 05-08 |
| AWHNONAG | Average Weekly Hours of Production and Nonsupervisory Employees | Hours | • | ALFRED | 08-01 | 11-07 | 02-06 | 05-08 |
| UNRATE | Unemployment Rate | % | • | ALFRED | 08-01 | 11-07 | 02-06 | 05-08 |
| CPIAUCSL | Consumer Price Index for All Urban Consumers | Index 1982-1984=100 | • | ALFRED | 07-16 | 10-16 | 01-16 | 04-15 |
| MortgageRate | Interest Rate on Housing Mortgages | % | | FHFA | N/A[2] | | | |
| Variables updated in quarterly frequency | | | | | | | | |
| ROUTPUT | Real Gross Domestic Product | Billions of dollars | • | RTDSM | Middle of the Quarter[4] | | | |
| P | Consumer Price Index | NA[2] | • | RTDSM | Middle of the Quarter[4] | | | |
| PCEC | Personal Consumption Expenditures | Billions of dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| FPI | Fixed Private Investment | Billions of dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| PRFI | Private Residential Fixed Investment | Billions of dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| PNFI | Private Nonresidential Fixed Investment | Billions of dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| PCND | Personal Consumption Expenditures: Nondurable Goods | Billions of Dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| PCESV | Personal Consumption Expenditures: Services | Billions of Dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| PCDG | Personal Consumption Expenditures: Durable Goods | Billions of Dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| PCNDGC96 | Real Personal Consumption Expenditures: Nondurable Goods | Billions of Dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| PCESVC96 | Real Personal Consumption Expenditures: Services | Billions of Dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| PCDGCC96 | Real Personal Consumption Expenditures: Durable Goods | Billions of Dollars | • | ALFRED | 07-31 | 10-30 | 01-30 | 04-29 |
| COMPNFB | Compensation Per Hour of Nonfarm Business Sector | Index 2012=100 | • | ALFRED | 08-08[5] | 11-06 | 02-05 | 05-07 |
| PRS85006023 | Average Weekly Hours of Nonfarm Business Sector | Index 2012=100 | • | ALFRED | N/A[6] | | | |
| BOGZ1FL144104005Q | Credit Market Liabilities of Nonfinancial Business Sector | Millions of Dollars | | ALFRED | N/A[6] | | | |
| HMLBSHNO | Home Mortgage Liabilities of Households and Nonprofit Organizations | Billions of Dollars | | ALFRED | N/A[6] | | | |
| HousePriceIndex | Price Index of New Single-Family Houses Sold | Index 2005=100 | • | CB | N/A[7] | | | |
| BLT | Bank Lending Tightening Index (Carabenciov et al., 2008) | % | | FRB[8] | 08-11[8] | 11-03 | 02-02 | 05-04 |
| | | | | SPFdeadline: | 08-07 | 11-10 | 02-10 | 05-12 |

# E Observed Variables

Table 2 displays the description of observed variables. For each variable, the table shows whether it has SPF nowcast value, whether it has current-quarter value, and which model(s) it is used as an observable.

We let observed variables to be shared by as many models as possible, so that the difference in forecasting performance can be explained more by specific model structure rather than by data collection and transformation procedures. The only exception is hours worked. As it is constructed in very different ways in the (1) DNGS15, (2) FRBEDO and (3) SW07 and KR models, we follow the original papers to obtain three unique hour worked series.

Table 2: Description of Observed Variables

| Name | Description | Has SPF Nowcast Value | Has Current-Quarter Value | Model | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | DS04 | WW11 | SW07 | FRBEDO | NKBGG | QPM08 | DNGS15 | KR15FF | KR15HH | 3vBVAR | 5vBVAR | 8vBVAR |
| Common variables | | | | | | | | | | | | | | | |
| xgdp_q_obs | Real GDP net growth | • | | • | • | • | • | • | | • | • | • | • | • | • |
| pgdp_q_obs | Consumer price index | • | | • | • | • | • | • | | • | • | • | • | • | • |
| rff_q_obs | Federal funds rate (quarterly) | | • | • | • | • | • | • | | • | • | • | • | • | • |
| fpi_q_obs | Real investment net growth | | | | | • | | • | | • | • | | | • | • |
| cp_q_obs | Credit spread | | • | | | | | • | | • | | | | • | • |
| pcer_q_obs | Real consumption net growth | | | | | • | | | | • | • | • | | | • |
| wage_obs | Real wage net growth | | | | | • | | | | • | • | • | | | • |
| per_q_obs | Nominal residential investment gross growth | • | | | | | • | | | | | • | | | |
| penr_q_obs | Nominal non-residential investment gross growth | • | | | | | • | | | | | • | | | |
| hours_ obs | Hours worked in SW07 and two KR models | | | | | • | | | | | • | • | | | |
| Model-specific variables | | | | | | | | | | | | | | | |
| hours_obs_dngs | Hours worked in DNGS15 model | | | | | | | | | • | | | | | • |
| LGDP_US | Real GDP level | • | | | | | | | • | | | | | | |
| BLT_US | Bank lending tightening index | | • | | | | | | • | | | | | | |
| LCPI_US | Consumer price index in QPM08 model | • | | | | | | | • | | | | | | |
| RS_US | Federal funds rate (yearly) | | | | | | | | • | | | | | | |
| UNR_US | Unemployment rate | • | | | | | | | • | | | | | | |
| pecnn_q_obs_frbedo | Nominal nondurables and services cons gross growth | | | | | | • | | | | | | | | |
| pecd_q_obs_frbedo | Nominal durables cons gross growth | | | | | | • | | | | | | | | |
| paipc_q_obs_frbedo | Gross inflation for consumer nondurables and services | | | | | | • | | | | | | | | |
| paipk_q_obs_frbedo | Gross inflation for consumer durables | | | | | | • | | | | | | | | |
| wage_obs_frbedo | Real wage gross growth in FRBEDO model | | | | | | • | | | | | | | | |
| hours_obs_frbedo | Hours worked in FRBEDO model | | • | | | | • | | | | | | | | |
| spread_kr | Firm loan spread | | • | | | | | | | | • | | | | |
| dlndn_kr | Nominal firm loan net growth | | | | | | | | | | • | | | | |
| spreadi_kr | Mortgage loan spread | | • | | | | | | | | | • | | | |
| dlnpo_kr | Nominal housing price net growth | | | | | | | | | | | • | | | |
| dlndin_kr | Nominal mortgage loan net growth | | | | | | | | | | | • | | | |

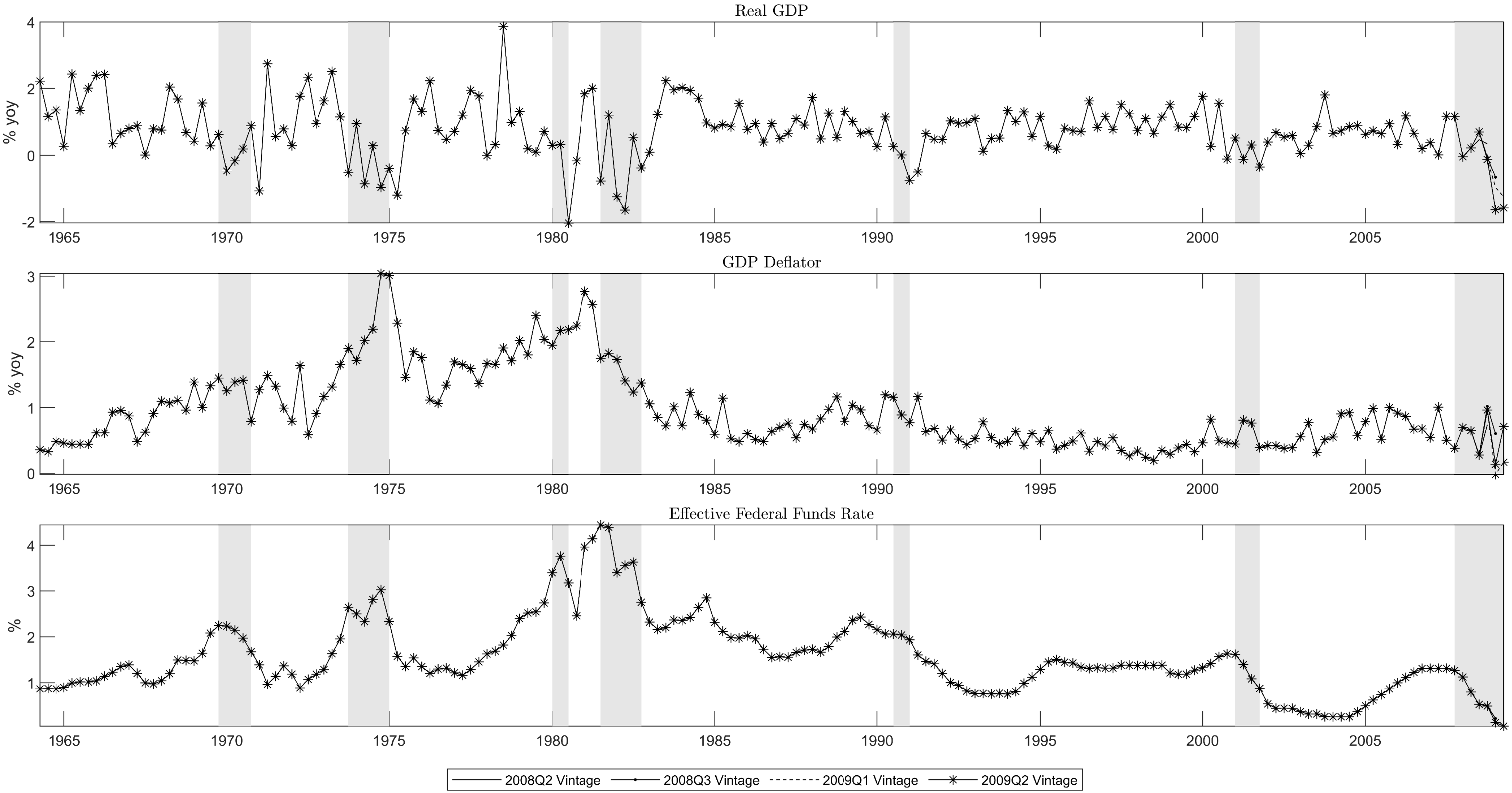



Figure 10: Real GDP, Inflation and Federal Funds Rate

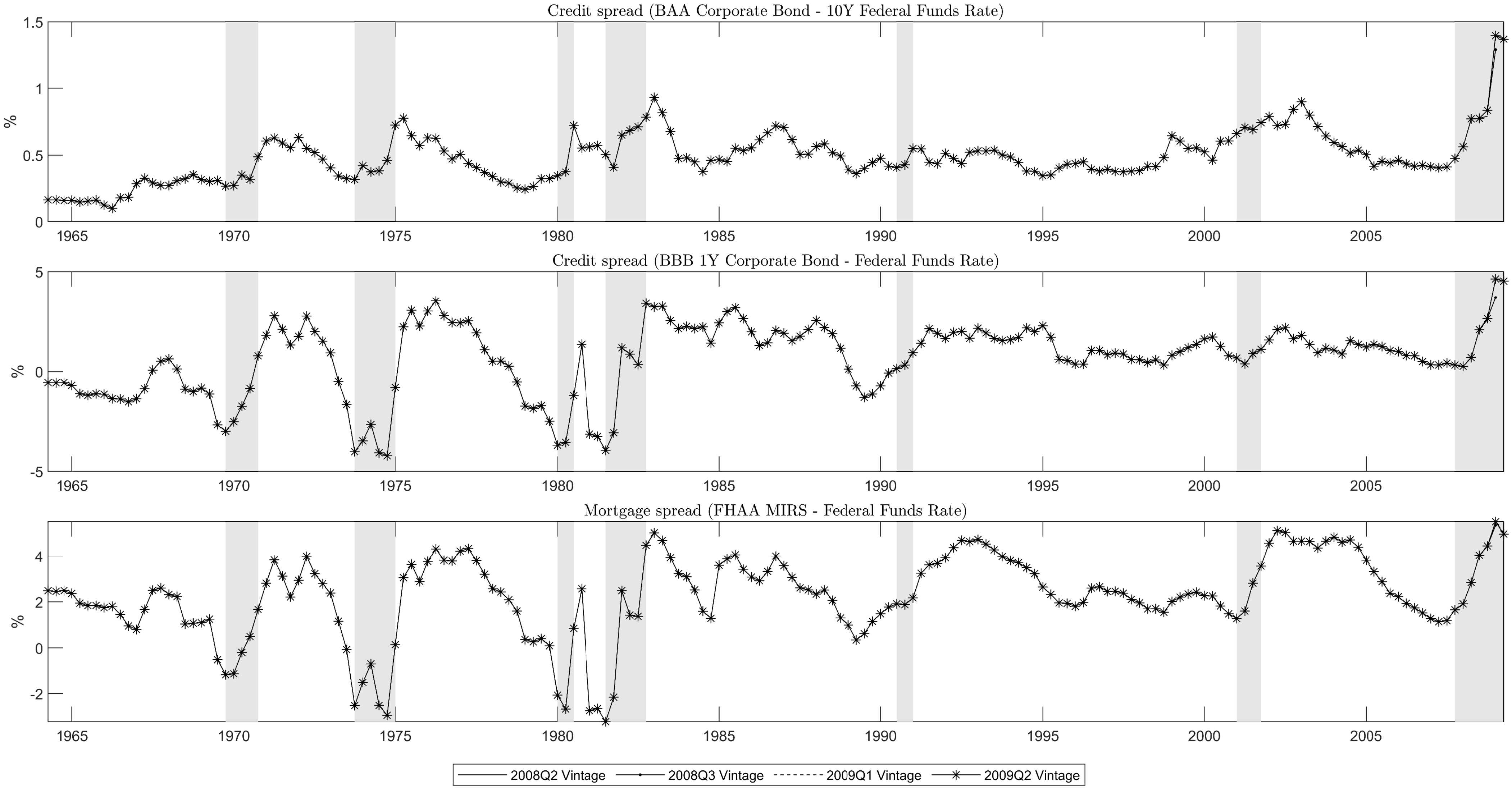



Figure 11: Credit Spread Variables

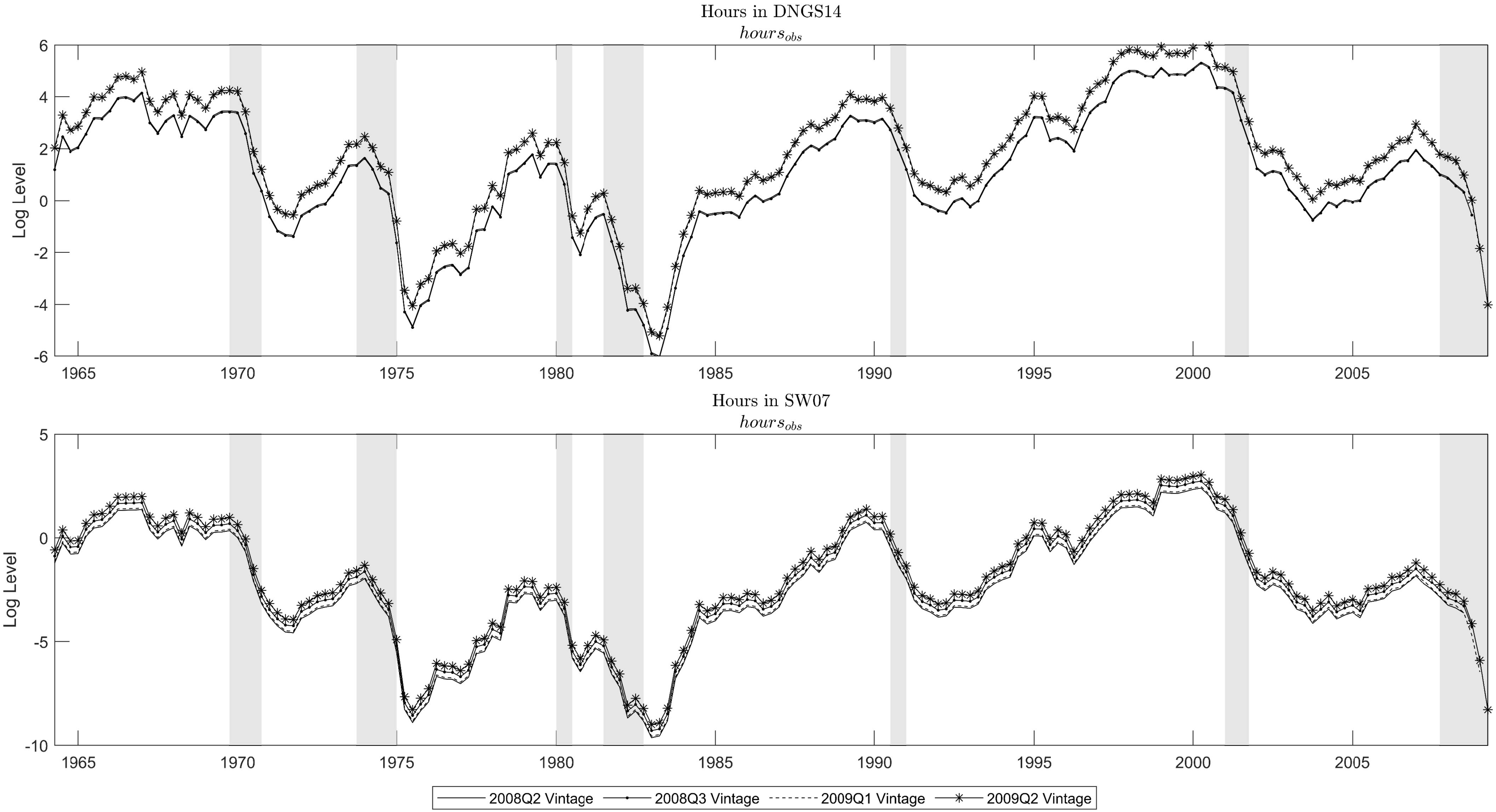



Figure 12: Comparison of Hours Worked Series Definitions