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, IMFS, Kiel Institute and ifo

November 13, 2022

Preliminary - Do not cite

Abstract

Macroeconomists have been criticized 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-sprea¡d 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. We extend the analysis to two additional recessions. For the 2001 recession we do not find systematic improvements in forecasting accuracy based on post-crisis models. Regarding the 2020 recession, we find that models that focus on labor market dynamics deliver the most accurate forecasts, while there is no systematic difference between pre- and post-crisis models. Hence, these results indicate that post-crisis models improve forecasting accuracy during recessions caused by financial crises, but not during other recessions.

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
† matyas.farkas@ecb.europa.eu
‡ 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; Blanchard, 2016; Romer, 2016).

In an earlier model forecast comparison study (Wieland and Wolters, 2011, 2012), we assessed whether this critique 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 examine the forecast performance of models built after the Global Financial Crisis, contrasting their propagation mechanisms and real-time forecast performance to those of 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).

We study post-crisis models with different types of financial frictions, observables and information sets to predict the Great Recession. In particular, we estimate 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); unemployment augmented version of the Smets-Wouters model, following Gali et al. (2012); and varying the real-time information content to compare their impact on forecast performance. To check whether the restrictions imposed by the structure of DSGE models in general is 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 third alternative, we estimate a Bayesian VAR model with the popular GLP prior by Giannone et al. (2015a) on the data-set of the small- and medium-scale NK-DSGE models with and without financial variables and apply the same real-time conditioning using entropic tilting. 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.

To allow for a direct comparison with 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 available to 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 account for informational advantage of professional forecasters compared to standard DSGE 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 most important finding is that NK-DSGE models with financial market frictions and employing 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. Our results confirm that DSGE models developed prior to the Global Financial Crisis were not able to predict the downturn in 2008 or 2009. In contrast we document a remarkable progress in structural macroeconomic modeling enabling to detect the Great Recession at its onset.

Our results also reveal that it is not sufficient to include financial market frictions in NK-DSGE models to obtain accurate real-time forecasts: the choice of observables used to inform the model about financial market distress is pivotal. NK-DSGE models involving long-term 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 either focus on housing-market collateral constraints or are informed by other financial market data do not improve upon the pre-crisis NK-DSGE models.

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 NK-DSGE 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.

We repeat the forecasting exercise for the 2001 and 2020 recessions. During these, forecasts based on post-crisis models would not have been more accurate than those based on pre-crisis models. This is not too surprising, because financial frictions were not the main drivers of these recessions. For the 2020 recession we rather find that the modeling of labor market frictions and the inclusion of labor market variables improves forecasting accuracy. This seems plausible due to the large adverse labor market effects of the COVID-19 containment policy measures. Hence, the usage of post-crisis models seems to be particularly appealing to forecast recessions during financial crisis periods, while during other times there are no systematic advantages compared to pre-crisis models. 18

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 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 regarding the Great Recession, while in Section 5 we present results regarding the 2001 and 2020 recessions. Finally, Section 6 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.

Regarding models that have been developed before the Global Financial Crisis, we use two small-scale, three 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 by incorporating nominal and real 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: the real GDP growth, the GDP deflator 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.

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, 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, inflation, interest rate
Smets and Wouters (2007)	SW07	medium-scale DSGE model with nominal and real frictions and seven structural shocks	7: output, consumption, investment, inflation, wages, hours, interest rate
Fratto and Uhlig (2020)	FU20	medium-scale DSGE model that is only different from the SW07 model in its shock processes	7: output, consumption, investment, inflation, hours, wages, interest rate
Edge et al. (2008)	FRBEDO08	medium-scale DSGE-model developed at the Federal Reserve. Two sectors with different technology growth rates, demand side disaggregated into different consumption and investment components	11: output, 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
Gali et al. (2012)	GSW12	medium-scale DSGE model similar to Smets and Wouters (2007) + labor market dynamics	8: output, consumption, investment, inflation, wages, employment level, unemployment rate, interest rate
Fair (2004)	Fair	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, inflation, interest rate, investment, credit spread
Iacoviello and Neri (2010)	IN10	medium-scale DSGE model similar to Smets and Wouters (2007) + financing frictions + housing sector	10: consumption, residential investment, business investment, inflation, hours in the housing sector, hours in the goods sector, house prices, interest rate, wages in the housing sector, wages in the goods sector
Christiano et al. (2014)	CMR14	medium-scale DSGE model similar to Smets and Wouters (2007) + financial accelerator + fluctuations in idiosyncratic uncertainty	12: output, consumption, investment, inflation, relative price of investment goods, wages, hours, interest rate, credit growth, credit spread, term spread, net worth
Del Negro et al. (2015)	DNGS15	medium-scale DSGE model similar to Smets and Wouters (2007) + financial accelerator	8: output output, consumption, investment, inflation, wages, hours, interest rate, credit spread
Kolasa and Rubaszek (2015)	KR15_FF	medium-scale DSGE model (Del Negro et al., 2007) + financial accelerator	9: output, consumption, investment, inflation, wages, hours, interest rate, credit spread, loan growth
Kolasa and Rubaszek (2015)	KR15_HH	medium-scale DSGE model (Del Negro et al., 2007) + financing frictions + housing sector	11: output, consumption, investment, 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	5: output, inflation, interest rate, unemployment rate, bank lending tightness
		Bayesian VARs	
Giannone et al. (2015b)	GLP3v, GLP5v, GLP8v	Bayesian VARs with optimal shrinkage prior estimated on the same observables as the above models	3 (same as the DS04 model), 5 (same as the NKBGG model), or 8 (same as the DNGS15 model)

The three medium scale models include physical capital in the production function and account for endogenous capital formation. Labor supply is modelled explicitly. Nominal frictions include sticky prices and wages, as well as 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 growth, consumption growth, investment growth, output deflator, the federal funds rate, hours and wage growth—and includes seven structural shocks. To confirm the robustness of the results, the second model we consider is from Fratto and Uhlig (2020), which has the same model structure as the original SW07 model with some adjustments on its shock processes. The third model is by Edge et al. (2008). Following these authors we refer to it as the FRBEDO08 model reflecting that this is one of the models used at the Federal Reserve Board. 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, growth of consumption of non-durables and services, growth of consumption of durables, growth of residential investment, growth of business investment, hours, wage growth, inflation for consumer non-durables and services and inflation for consumer durables.

Another model we consider is from Gali et al. (2012). This model does not bring in any new type of frictions, but rather reformulates the Smets-Wouters model to allow for involuntary unemployment. The positive unemployment rate in this model is caused by positive wage markups, which implies the presence of monopolistic power in the labor market. The model measures labor market conditions by the unemployment rate and the growth of employment level.

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 growth, output deflator, the federal funds rate, fixed private investment growth 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 three medium-scale DSGE models with the financial accelerator. The first one is from Christiano et al. (2014), which incorporates the financial accelerator in a standard medium-scale DSGE model. They use this model framework to show that the fluctuations in idiosyncratic uncertainty is the most important shock that drives the US business cycles. Four financial variables, which measure the credit spread, the term spread, the credit growth, and the net worth growth, are added as observables. The second one by 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 third 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 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 investigate two models of this kind. The first one is from Iacoviello and Neri (2010), which includes the housing sector and financing constraints in an otherwise standard medium-scale model. The model is informed by residential investment and business investment data separately rather than using an aggregate investment series as observable. Further, data on hours and wages is split up into data covering the consumption and the housing sector. Another distinct feature is that output is not an observable in this model, and thus we instead obtain the forecasts of output as the summation of forecasts of consumption, residential investment, and business investment. The second model is 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 modelled 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 the macroeconomic models by basing them on the same respective observables. We use the GLP prior by Giannone et al. (2015b) 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. We adopt a rolling window estimation strategy to fix the number of quarters in each sample to 100, and we re-estimate the models by shifting both the starting and ending quarters, to ensure that the sample size remains constant.

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 at the deadline for professional forecasters to submit their forecasts to the SPF. These deadlines, which are also data vintages we employ, are August 7th in 2008Q3, November 10th in 2008Q4, February 10th in 2009Q1, and May 12th in 2009Q2. Hence, the model-based forecasts are based on the same data vintage as available to professional forecasters in real time. This guarantees that DSGE models have no informational advantage compared with professional forecasters when generating forecasts.

In addition to a detailed forecasting performance analysis during the Great Recession, we investigate the forecasting performance around two other recessions, the 2001 recession and the 2020 COVID-19 crisis. Regarding the former, we study forecasts with starting points form 2001Q1 to 2001Q4. The SPF deadlines are February 14th, May 12th, August 15th, and November 14th of 2001, respectively, and we use the latest data vintages available at these dates for estimating the models. Regarding the ongoing COVID-19 crisis, we study forecasts with starting points from 2020Q1 to 2020Q3. The SPF deadlines or equivalently the vintages at which we collect data are February 11th, May 12th, and August 12th of 2020, respectively.

Key macroeconomic time series used for estimating the models are published on a quarterly frequency and are subject to a publication lag of one quarter. Thus, when computing forecasts starting in middle of the current quarter, only data vintages based on the advance estimate, i.e. the first estimate, for the previous quarter for variables like GDP growth are 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, leading indicators, financial market variables, as well as macroeconomic variables that are published on a monthly or even higher frequency. 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 time series used in the estimation of the different macroeconomic models. The only exceptions are those series that are published very late in the quarter after the one they refer to. Since professional forecasters cannot observe the previous quarter values for these variables at the SPF deadlines in the middle of the quarter, for these series we only use observation up to two quarters before the forecast starting point. Fortunately, only two variables fall into this category, which are all from the Financial Accounts of the United States, which is usually released by the Federal Reserve Board towards the end of a quarter.

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, treasury yields, credit spread, term spread), but also for macroeconomic data published on a monthly frequency, especially for those series representing labor market conditions such as hours worked and the unemployment rate, which are included in the Employment Situation provided by the Bureau of Labor Statistics in each month. 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.

The appendix includes detailed information about the raw data that we collect, the data transformations applied before using the time series as observables for estimating models and the usage of current quarter information or SPF nowcasts for information scenarios 2, 3 and 4 as outlined above.

All models are estimated using the same observables, measurement equations and priors as proposed by the original authors. We adopt a standard Bayesian methodology to estimate models and generate forecasts. That is, 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 subdraws and compute point forecasts as the mean of the density forecasts.

4 Results: Forecasting the Great Recession

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 whether the specific theoretical transmission channels in DSGE models are needed for producing accurate forecasts.

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 seven pre-crisis models as described in Table 1 and the four different conditioning assumptions (from left to right). SPF forecasts are shown for comparison.

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

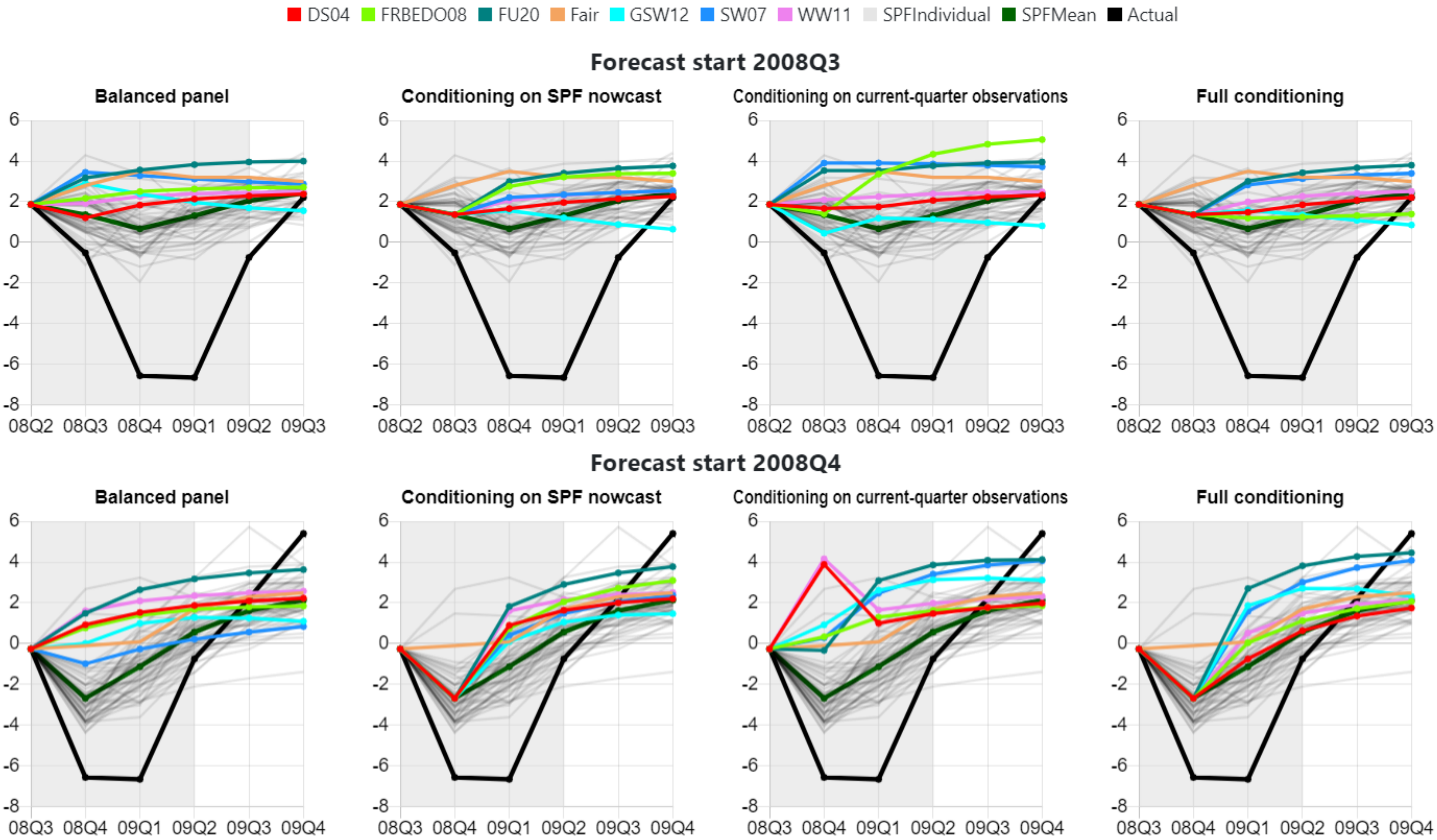
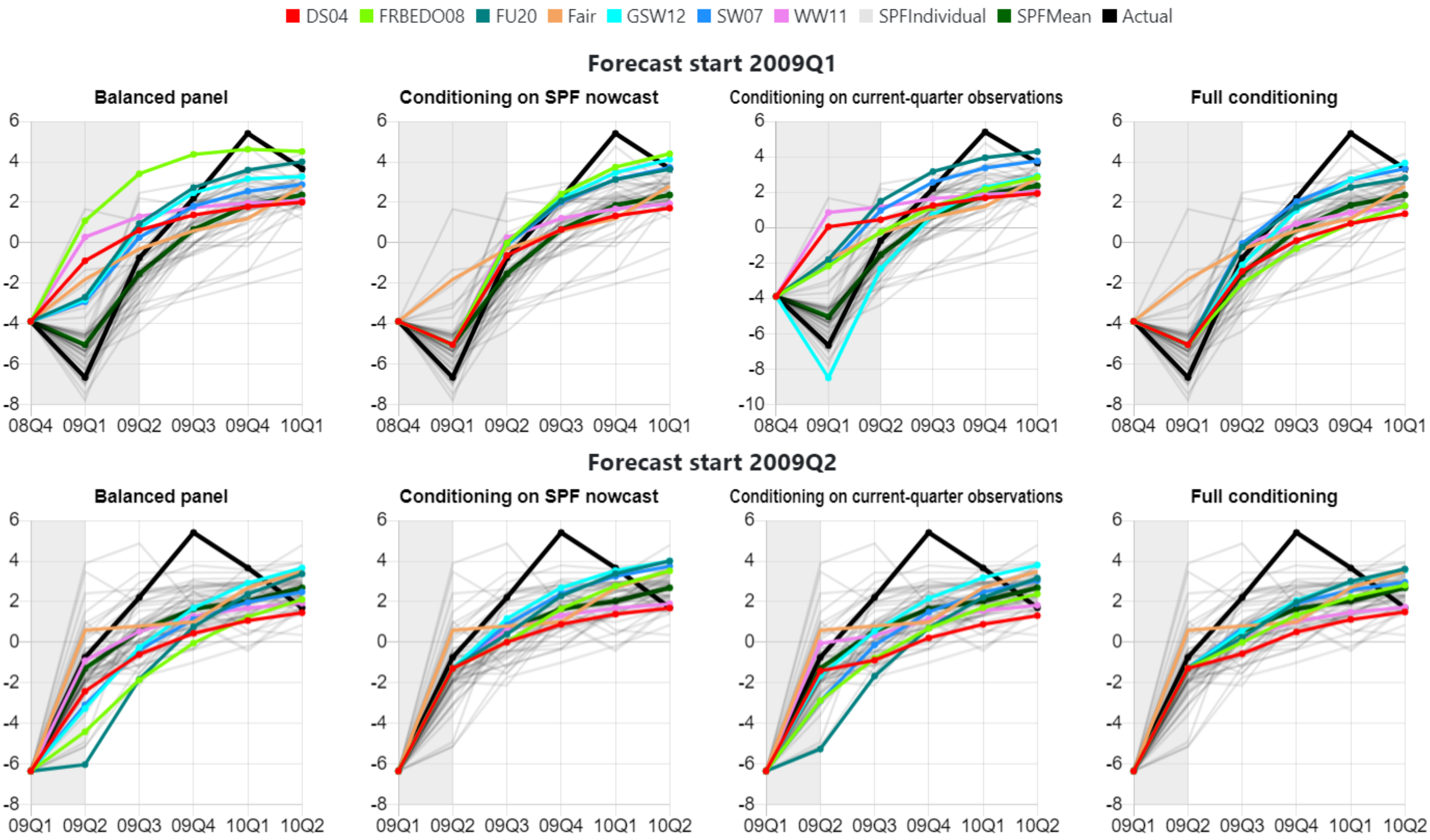


Figure 2: GDP Growth Forecasts in 2009:I–2009:II:Pre-Crisis Models



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. Finally, the shaded areas correspond to the periods which are defined as recessions by the NBER.

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). Only the DS04 model provides a slightly lower nowcast of GDP growth compared to the mean SPF. 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. The only exceptions are the FRBEDO08 model and the GSW12 model. The former is the largest of the considered pre-crisis DSGE models and 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. The GSW12 model enriches a standard DSGE model with frictions in the labor market and includes data on the unemployment rate and on changes in employment. Both models predict a lower GDP growth rate in the third scenario when conditioning on current quarter data, compared with their predictions in the balanced panel scenario, but are far from predicting a deep recession.

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. The inclusion of current-quarter information about the drop in the federal funds rate in the third scenario also provides no help in forecasting the recession. In fact, conditioning on the low nominal interest rates in 2008Q4, the DS04 and WW11 models even predict a economic boom in the fourth quarter of 2008. 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.

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, when conditioning on current-quarter data, the extremely low nominal rates leads to positive nowcasts of the GDP growth rates in the DS04 and WW11 models.

The important observation from the above forecasts is that pre-crisis macroeconomic models cannot predict the downturn in the GDP growth rates during the Great Recession, including those medium-scaled DSGE models of the type used by central banks at least before the Global Financial Crisis. Professional forecasters also failed in predicting the Great Recession in advance. Their performance in detecting the recession at its onset in 2008Q4 is better than that of the DSGE models, but they systematically underestimated the depths and length of the Great Recession. Regarding the prediction of the recovery, there are no systematic differences between the models and the professional forecasters. Both get the timing of the recovery roughly right.

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 or 2009Q2 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 explaining the Global Financial Crisis in the post-crisis models lead to better recession forecasts. Figures 3 and 4 show forecasts for the seven 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

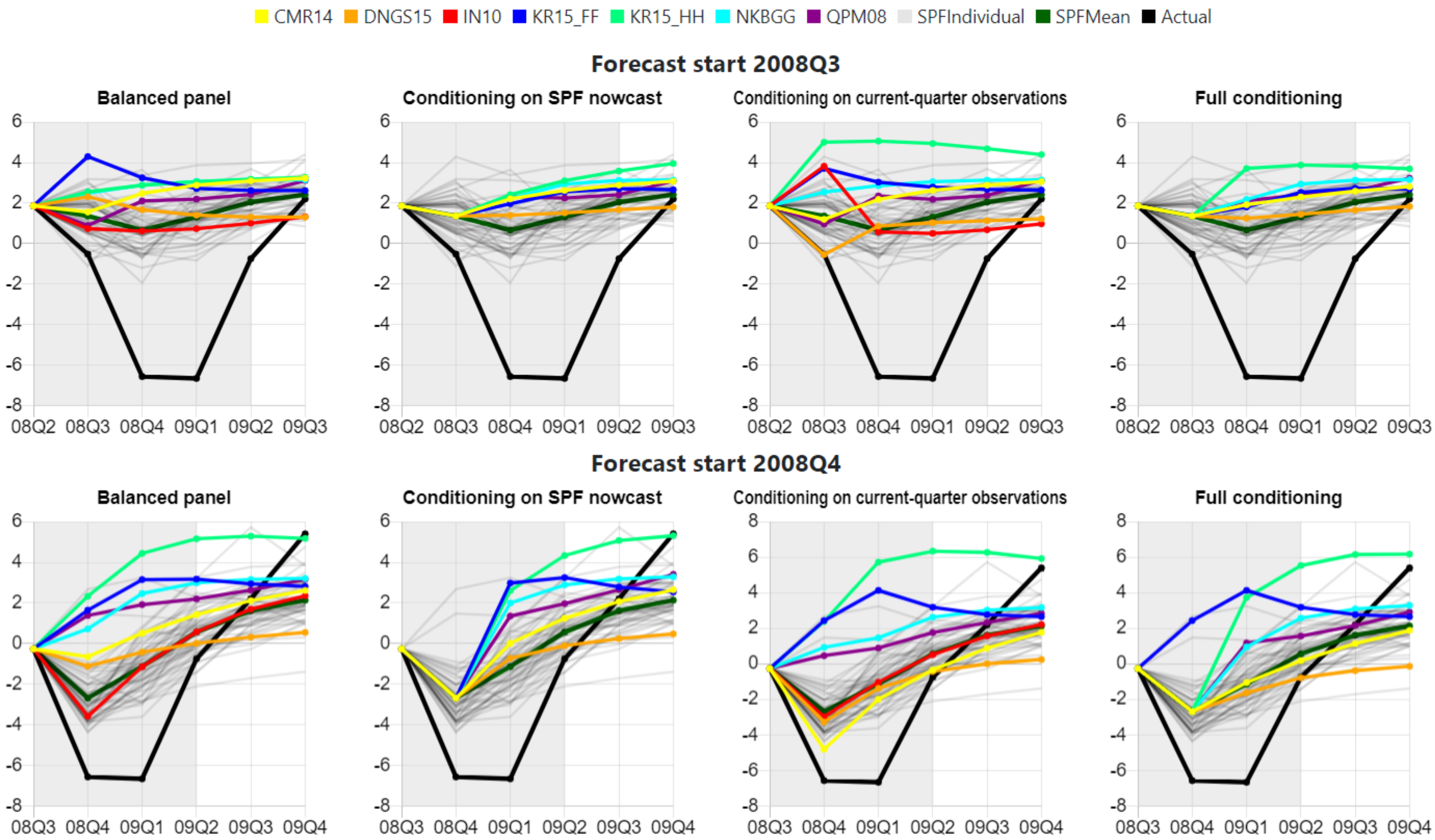
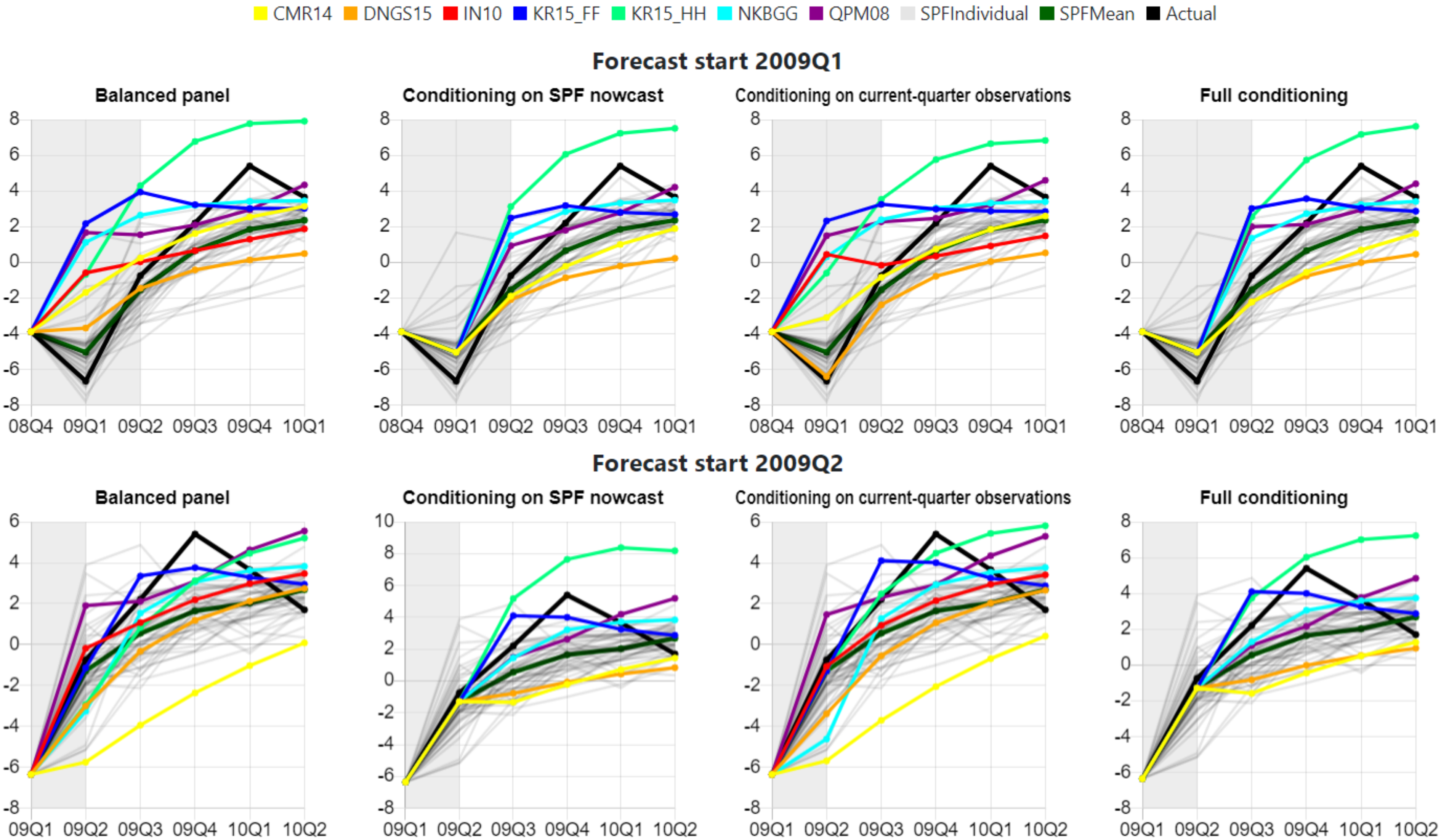


Figure 4: GDP Growth Forecasts in 2009:I–2009:II:Post-Crisis Models



Most of the forecasts of the post-crisis models starting in 2008Q3 look similar to those of the pre-crisis models, although some forecasts are somewhat closer to the mean SPF forecast, so that the overprediction of growth is somewhat smaller. None of the models foresees the upcoming recession, though. This result does not change when conditioning the forecasts on the SPF nowcast.[2] When conditioning on the data available for the current quarter, the nowcasts from some models turn out to be somewhat lower. In particular, the medium scale models that include a financial accelerator (CMR14 and DNGS15) or a measure of bank lending tightness (QPM08) yield GDP growth forecasts that are lower than the mean SPF nowcast, due to worsening financial conditions during 2008Q3. In particular, the DNGS15 model yields a highly precise nowcast that is much lower than the SPF nowcast. There have already been some signals indicating the crisis on August 7th, 2008, which is the SPF deadline in 2008Q3. The federal funds rate dropped, the credit spread somewhat widened, and the labor market weakened for the eight consecutive quarters since 2007Q1. The DNGS15 model combines these latest signals from the financial market and the real economy with its financial accelerator mechanism that amplifies the unfavorable conditions in the data, which results in the negative nowcast of the GDP growth in 2008Q3, even before the largest drop in output in 2008Q4. The nowcasts from the CMR14 model that also features the financial accelerator are slightly lower than the mean SPF nowcast, but the model still predicts a positive growth rate for 2008Q3. The reason is that due to the difference in observables only observations for two out of four financial variables of the CMR14 model for 2008Q3 are available in the third scenario. The nowcasts of other models even increase when being conditioned on current quarter observations, due to the decrease in the federal funds rate.

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), most of the post-crisis models do not perform any better than the pre-crisis models. The exception is the IN10 model: The worsening of the housing market conditions over 2008Q3 leads to the most pessimistic GDP growth nowcast of all models for 2008Q4. This nowcast in more precise than the mean SPF nowcast and the vast majority of individual nowcasts from the SPF. However, even the IN10 model fails in predicting the continued strong decline of GDP growth in 2009Q1, but rather predicts only a slighty negative growth rate for 2009Q1.

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 similar or less negative growth rates in the next quarter and a return to positive growth rates directly afterwards similar to the pre-crisis models.

[2]The IN10 model does not contain any observables for which SPF nowcasts are available, and therefore we do not show forecasts under scenarios 2 (conditioning on SPF nowcast) and 4 (full conditioning) for this model.

When conditioning on current quarter data (column 3), the forecasts of models that include a financial accelerator mechanism and use a credit spread as an observable change quite drastically. These models endogenously generate highly negative GDP growth rates for the fourth quarter of 2008. The most accurat nowcast is produced by the CMR14 model (-4.79 percent) followed by the DNGS15 model (-3.25 percent).[3]. Moreover, the nowcast of the CMR14 model is even more accurate than the nowcast provided by the most pessimistic professional forecaster. As the deterioration of the financial market increased throughout 2008Q4, the financial accelerator mechanism in these two models enables them to detect the crisis through the drastic changes of the asset prices in the current quarter. This is reflected in the elevated credit spread, which is an observable for the DNGS15 and CMR14 models. Besides, the CMR14 model can additionally takes the information from another observable, the slope of the term structure, into account.[4]

Is it sufficient to include a financial accelerator mechanism in a medium-scale DSGE model to to detect the Great Recession at its onset? The forecasts of the KR15_FF, another medium-scale DSGE model with the BGG-type financial accelerator, shows that the specific choice of observables is crucial, too. The KR1_FF does not predict the large downturn, despite being very 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 as shown in Figure 5. The fourth quarter of 2008 was a five sigma event for the long-term financing credit spread, occurring 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 occurring 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 of the credit spread increase on GDP growth to some extent.

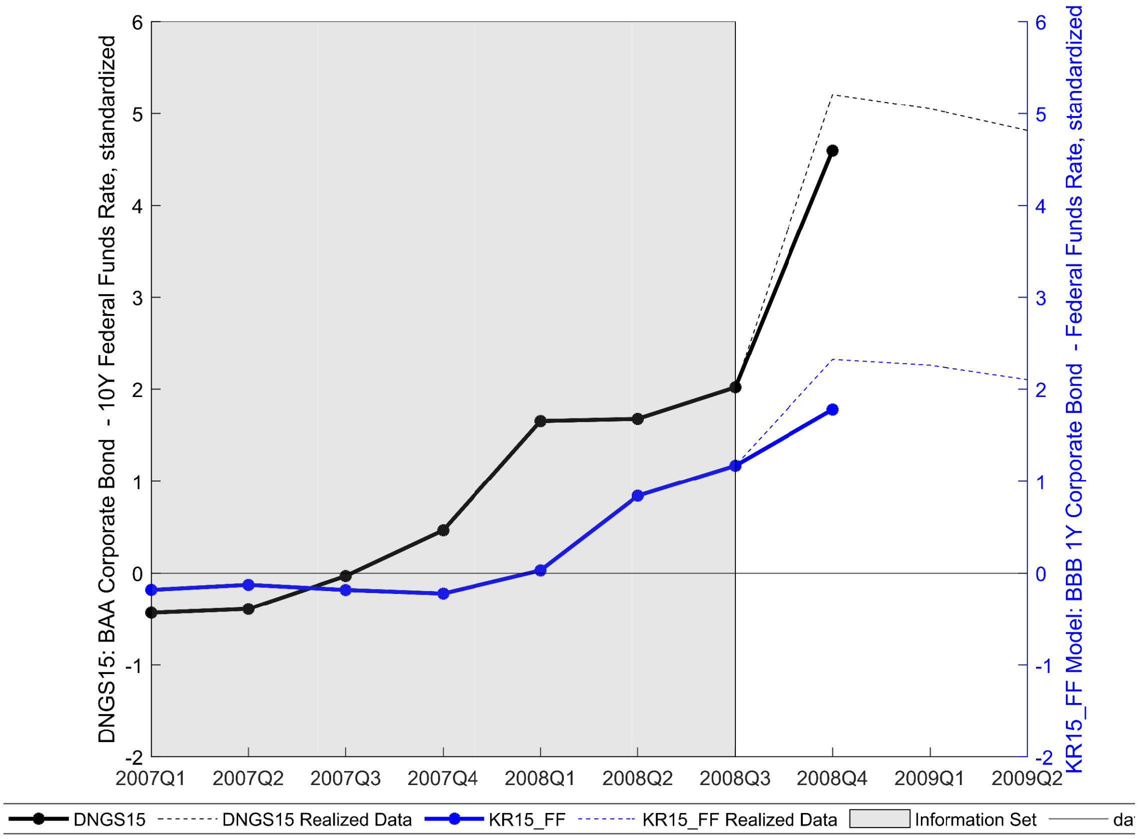
The smaller NKBGG model also includes the financial accelerator mechanism, but the lack of other frictions of medium-scale DSGE models (such as habit formation and investment adjustment costs) prevents it from successfully forecasting a recession

Another model that detects the recession at its onset is the IN10 model. This model does not have a financial accelerator mechanism, but includes financing constraints in the housing sector. Together with observables capturing developments in the housing sector these apear useful for detecting the recession early on. In contrast, the KR15_HH model, which has similar structures as the IN10 model, does not succeed in forecasting the decline of the GDP growth. Although the IN10 and KR15_HH model have essentially the same core mechanism, that is the inclusion of the housing market and a collateral constraint,

[3]The GDP growth nowcast of the DNGS15 model for 2008Q4 is about -7 percent in the original paper by Del Negro et al. (2015), which is more accurate than what we obtain. The difference can be explained by the different samples. First, their sample starts from 1964Q1, while our sample starts from 1984Q1. Second, they only take the values of the federal funds rate and the credit spread in 2008Q4 into account when generating the nowcast, while we consider the 2008Q4 values of all the observables in the data, which means that we additionally include the value of the hours worked in 2008Q4. Finally and most importantly, while we include data up to the SPF deadline on November 10th, Del Negro et al. (2015) include data until December 31th. If we use credit spread data for the whole fourth quarter, the GDP growth nowcast is very similar to the one in Del Negro et al. (2015).

[4]Stock returns are also an observable in the CMR14 model. Although the stock index itself is updated on a daily frequency, we cannot obtain the current-quarter value for this observable, because to construct it, we need other variables that are updated on a quarterly frequency and thus do not have current-quarter values. In addition, the current quarter condition of the credit growth, another financial variable in the CMR14 model, is also not updated at the SPF deadline, as the Federal Reserve usually release this data at the end of the quarter following the quarter the data refers to.

Figure 5: Credit Spread Variables in 2008Q4



Notes: The figure shows the real-time observable credit spread variable standardized with the real-time in sample mean and standard deviation. That is both real-time observables have a zero mean and unit variance. The resulting difference reflects their real-time information content. The credit spread of the DNGS15 is in black, measured on the left axis, and that of KR15_FF model is in blue. The grey shaded area indicates the information content when the forecasts were made without external information, that is only data up until and including 2008Q3 is used. The solid lines indicate the current vintage values, that is the currently observable credit spread. Lastly the dashed lines show the standardized realized values.

they are also different in many perspectives. One important difference in the data is that the IN10 model not only contains residential investment and house prices, which also appears in the KR15_HH model, but also the wage growth and hours for the housing sector as observables, and the decline of these variables are deeper than their counterparts for the whole economy.

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. The QPM model and the GSW12 model are the only models that include 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, so that also the GSW12 model fails in predicting a 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 CMR14 and DNGS15 models 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, in the balanced panel scenario none of the models predicts a further deepening of the recession for 2009Q1. The range of forecasts regarding the following quarters is very wide. It ranges from much too pessimistic forecasts regarding the return to positive growth rates based on the DNGS15 model to predicting a large boom based on the KR15_HH model. Conditioning on the SPF nowcast does not change this wide range of forecasts substantially. When conditioning on current-quarter observations, the nowcast from the DNGS15 model is the most accurate one (-6.39). The nowcast from the CMR14 model also becomes slightly lower in this scenario (-3.08) compared to its value in the first scenario, but is already much less accurate than the nowcasts from the DNGS15 model or the mean SPF nowcast. One explanation is that the credit spread contains more information about the ongoing financial market turbulence than the term spread, where the latter observable only appears in the CMR14 model. The nowcasts and forecasts of all other models predict a return to positive growth rates in 2009Q1 and a stable recovery afterwards, rather than a deepening of the recession.

The above results show that a number of post-crisis macroeconomic models that contain financial market frictions outperform pre-crisis models and professional forecasters in predicting the downturn in the GDP growth rates, before (IN10), at the beginning (IN10, CMR14, DNGS15), or in the middle (DNGS15, CMR14) of the recession. In particular, the performance of the DNGS15 and the CMR14 model across the different information scenarios, shows the importance of using current-quarter data that informs the models about the latest developments of the financial market and real economy in the estimation. The differences in the forecasts that come from models having similar mechanisms but different observables imply that estimating models using suitable observables is crucial to achieve an accurate crisis forecasts.

In the second quarter of 2009, most post-crisis models predict a quick return to positive growth rates as the pre-crisis models and the SPF. The KR15_FF provides the most accurate nowcast in almost all the scenarios. Moreover, the subsequent forecasts from the KR15_FF model replicate the hump-shaped dynamics of the actual GDP growth out of the recession, although the predicted peak is one quarter earlier than the actual peak.

While the financial accelerator mechanism combined with credit spread data in the DNGS15 and CMR14 model turned out to bee highly useful for achieving precise recession nowcast, regarding recovery forecasts these turn out to be detrimental. These two models yield the most pessimistic forecasts and the recovery turned out to be much faster. Hence, there seems to be some state dependence in the usefulness of the financial accelerator for obtaining precise forecasts. It is very useful for obtaining precise short-term forecasts during the acute phase of a crisis, but not for medium term forecasts once the acute phase of a crisis is over.

4.3 Forecasting the Great Recession: Data versus Model Structure

In the previous section we showed that the inclusion of financial frictions in DSGE models in combination with timely information about financial market distress leads to accurate GDP nowcasts. Now we compare forecasts from the small-scale (NKBGG) and the medium-scale (DNGS15) models with financial frictions, 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 empirical model is sufficient or whether modeling financial frictions is important to generate precise forecasts.

Figures 6 to 9 show forecasts for the four 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 GLP5v is its empirical counterpart. The DNGS15 model is estimated on the same eight time series as the BVAR labeled GLP8v.

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

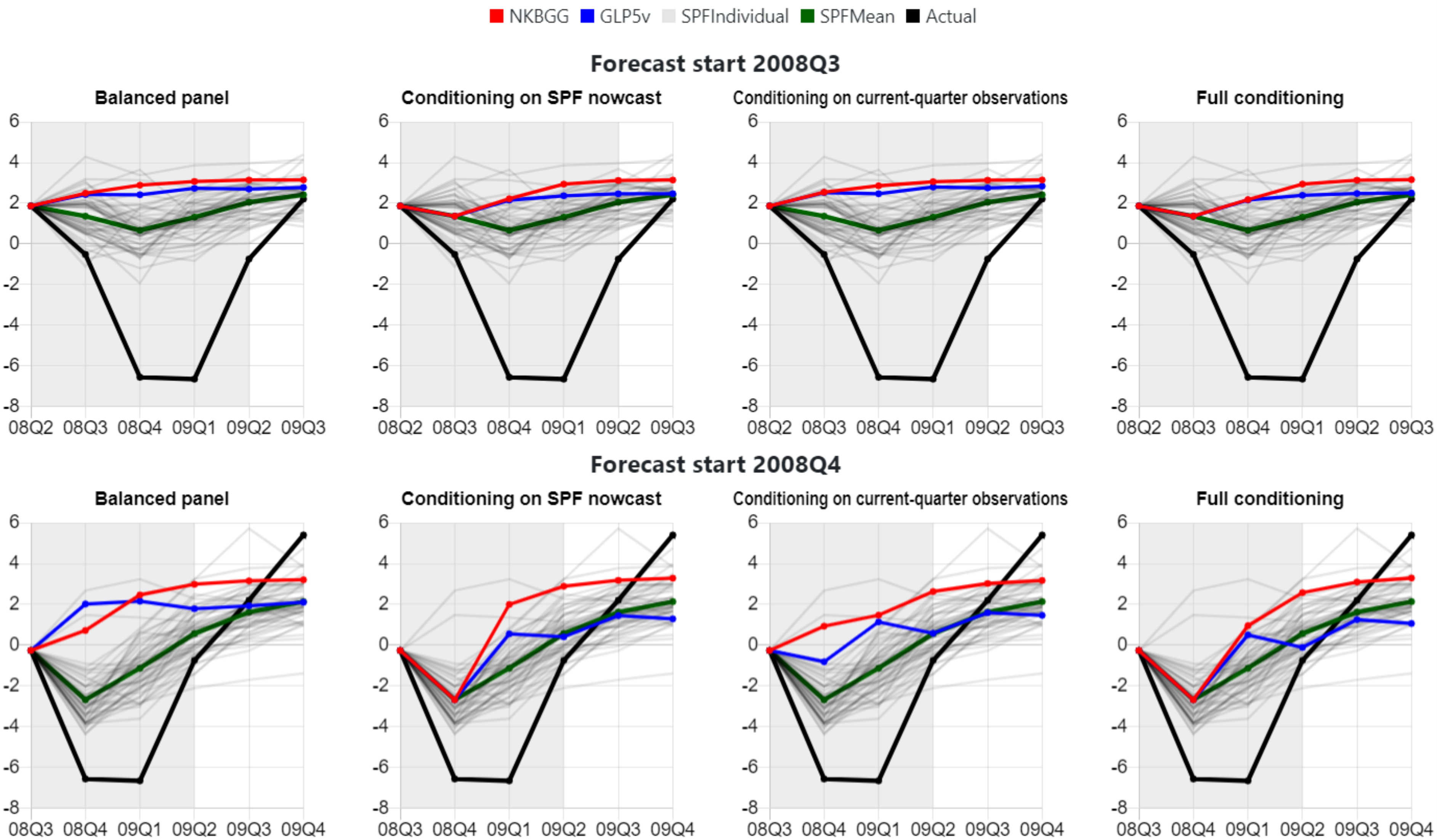


Figure 7: GDP Growth Forecasts in 2009:I–2009:II: 5-Variable NK Model versus 5-Variable Bayesian VAR Model

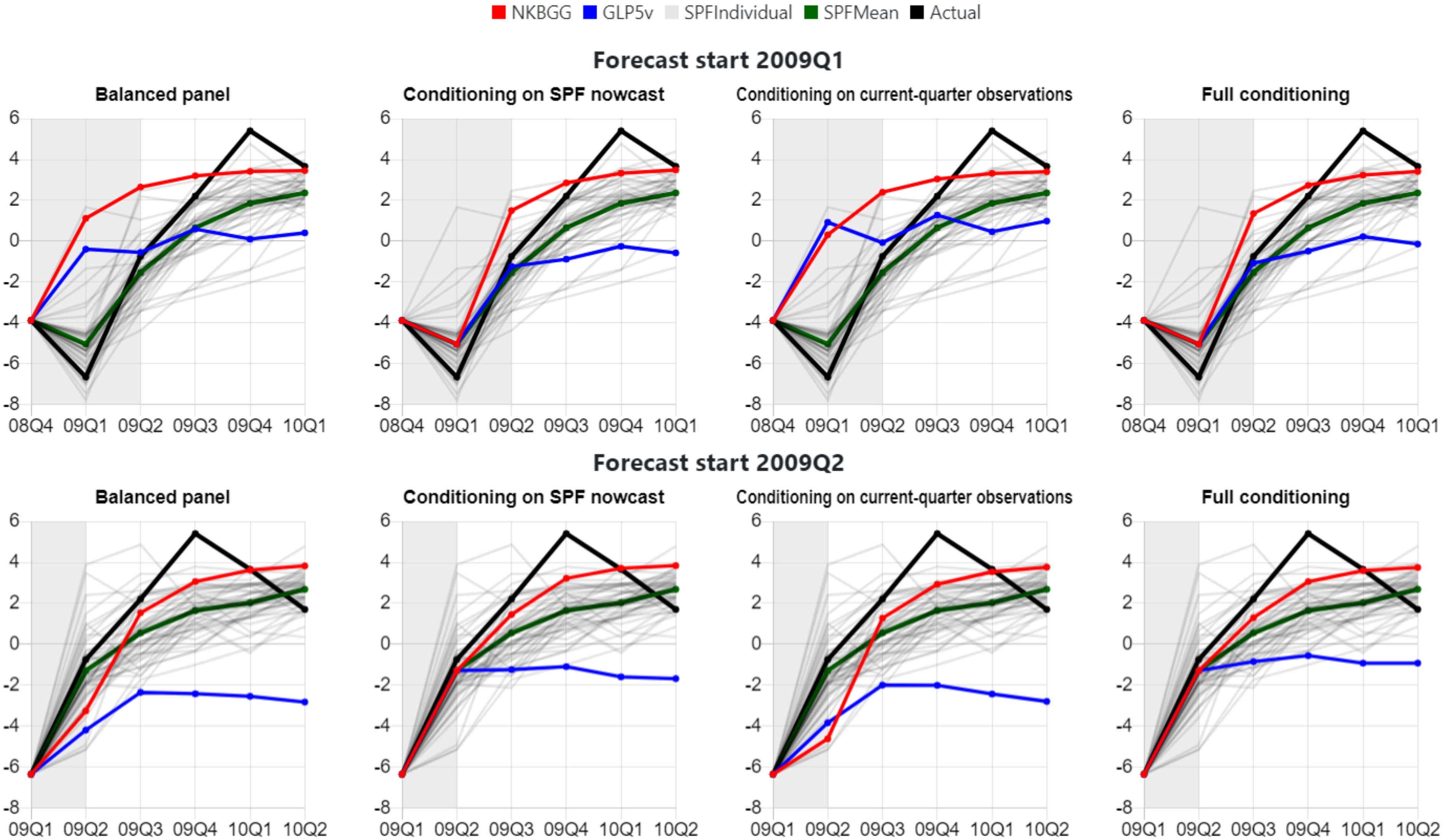


Figure 8: GDP Growth Forecasts in 2008:III–2008:IV: 8-Variable NK Model versus 8-Variable Bayesian VAR Model

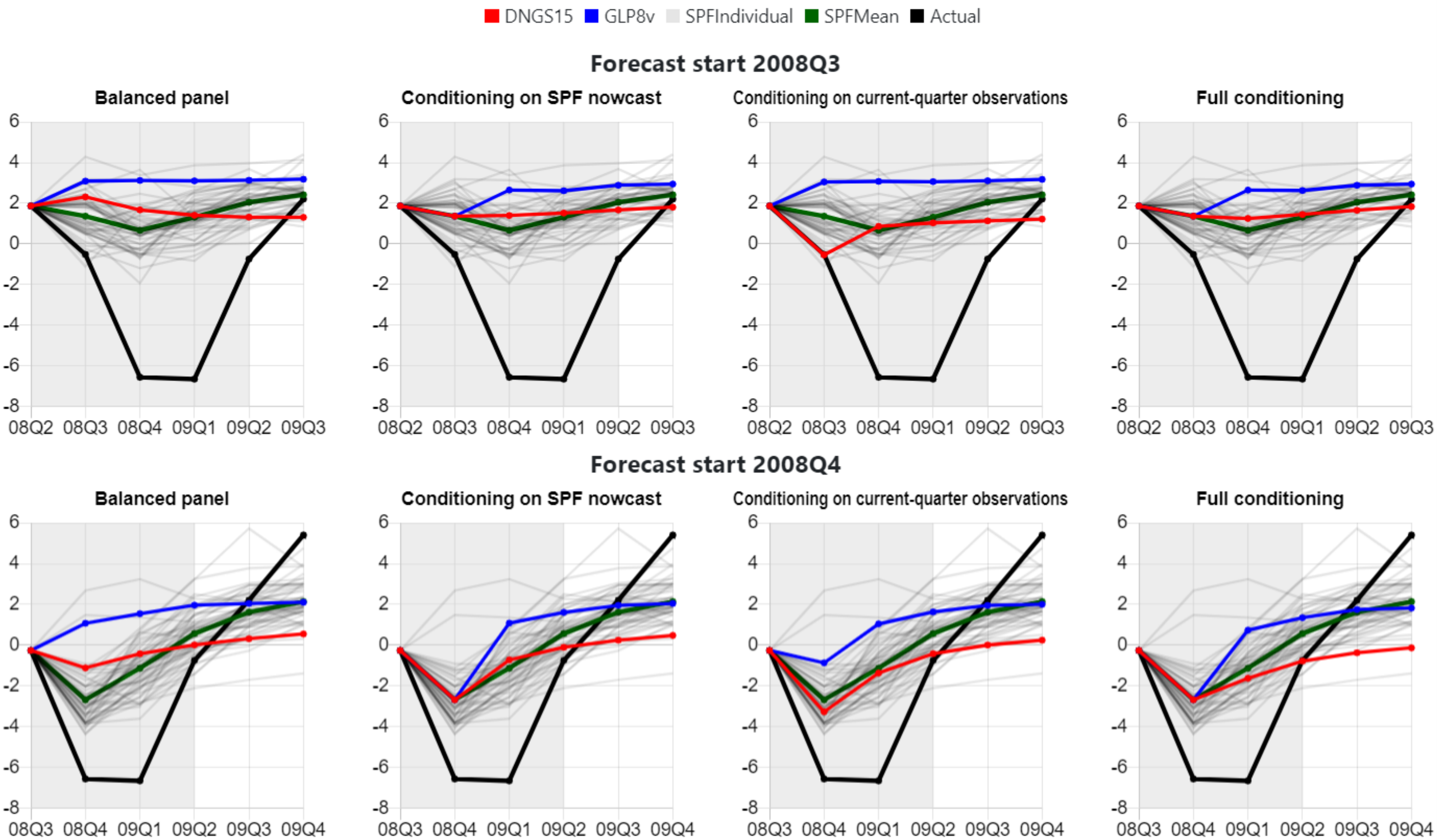
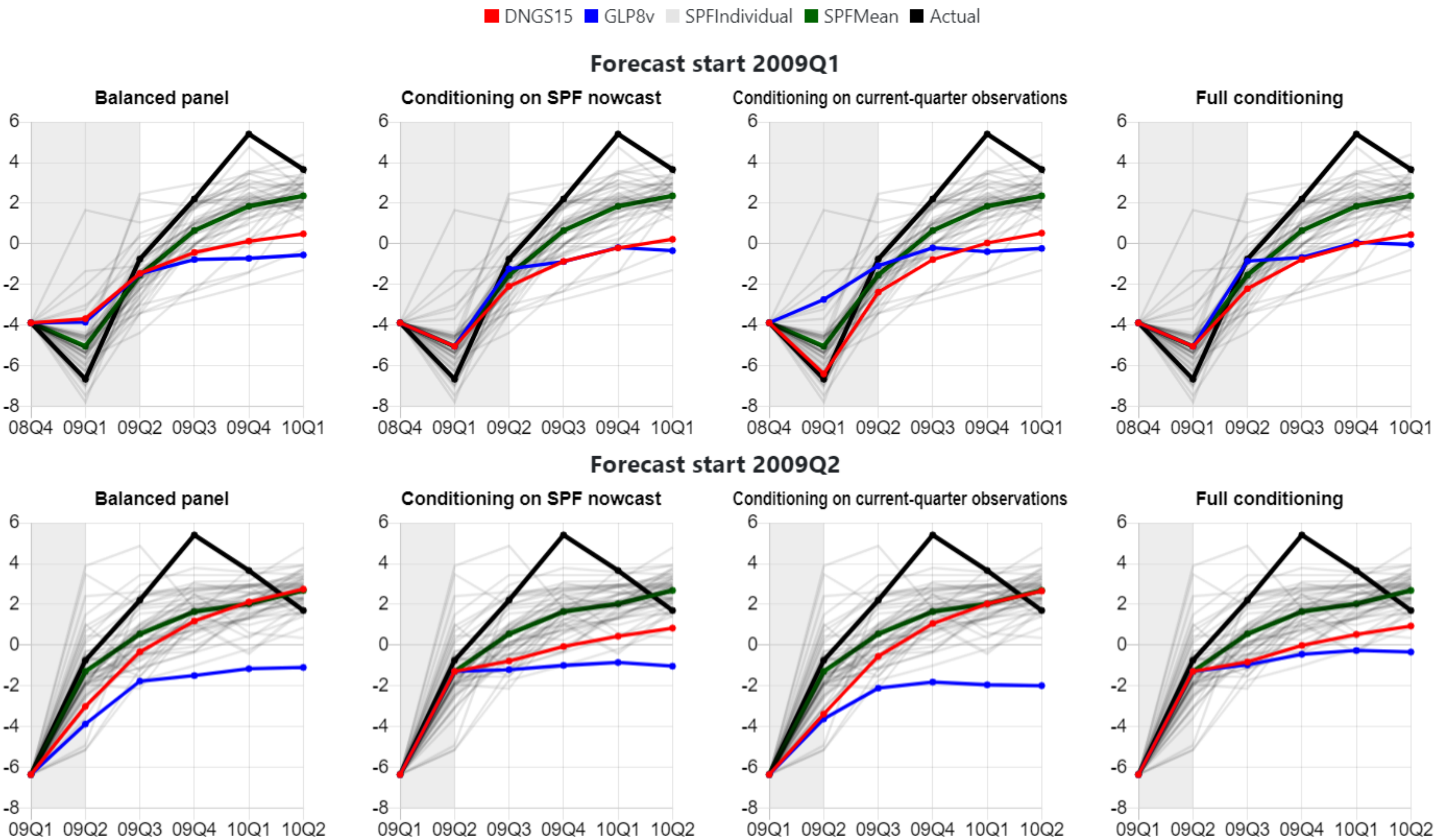


Figure 9: GDP Growth Forecasts in 2009:I–2009:II: 8-Variable NK Model versus 8-Variable Bayesian VAR Model



Starting from the third quarter of 2008, both the GLP5v and GLP8v forecasts remain at high levels, which are slightly lower than the NKBGG forecasts, but higher than the DNGS15 forecasts. Irrespective of the conditioning this pattern holds across all four scenarios (columns 1-4). The forecasts from BVARs merely extrapolate the relatively high growth numbers of the most recent quarters into the future. The cross-equation restrictions imposed by economic theory are beneficial in improving the accuracy of forecasts for the medium-scale DSGE model, but are of little help for the small-scale model. A comparison of the nowcast from the DNGS15 and GLP8v models suggests that, early signs of crisis in the financial and labor markets are only helpful for detecting the recession at its onset when modeling financial frictions rather than using a reduced form model.

Regarding the forecasts starting in 2008Q4, the forecasts from both the GLP5v and GLP8v models display relatively similar dynamics across the four scenarios. When conditioning on the SPF nowcast, both forecasts return back to positive levels in the quarter afterwards. When conditioning on current-quarter observations, both VARs generate a slightly negative GDP growth nowcast, but predict a quick return to positive growth rates afterwards. The GLP5v forecasts are systematically lower and thus more accurate than the NKBGG forecasts, with the only exception being the nowcast in 2008Q4 under the balanced panel scenario. The GLP8v forecasts are systematically higher and thus less accurate than the DNGS15 forecasts with no exception.

Starting from 2009Q1, the GLP5v model forecasts a quick return up to zero percent growth, but not a full recovery to positive growth rates afterwards. The GLP8v model generates a similar forecast, thought it is even a little bit lower over the medium term than the one of the GLPv model. The NKBGG model yields medium term forecasts that are roughly in line with the actual recovery, while the DNGS15 model yields forecasts that are too pessimistic and that are similar to the ones of the GLP8v counterpart.

Consulting the forecasts starting in the second quarter of 2009, the BVAR forecasts are more pessimistic compared to their DSGE counterparts and do predict that growth rates remain negative over the medium term. 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. The forecast of the DNGS15 model is more optimistic than the one of the GLP8v counterpart, but still too low, while the NKBGG forecasts is more precise.

Overall, the theoretical structure of the DSGE models turns out to be important in improving forecasting accuracy. Only using an additional credit spread time series in an otherwise standard BVAR model is not sufficient to generate accurate forecasts. Driven by the GLP prior's unit root property the BVARs simply 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.

In the previous section we showed that the inclusion of financial market information increases forecasting accuracy for DSGE models only during the acute phase of the crises, but decreases forecasting accuracy during the recovery. This also holds for BVARs, though the underestimation of growth rates during the recovery is much more extreme compared to structural models.

4.4 Systematic Forecast Evaluation

In this section we study the forecast performance of the models more systematically by comparing root mean squared prediction errors (RMSE), though we refrain from conducting tests of equal forecasting accuracy due to the low number of observations.

The RMSE 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 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+h) and T_1 the end (2009Q1+h) of the evaluation period. The RMSE is the optimal measure for forecast performance if a quadratic loss function is assumed.

Table 6 shows the RMSEs for GDP growth. The rows show the different forecast horizons starting from horizon 0 (nowcast) up to horizon 4, while the columns show the different information scenarios (BP: balanced panel, SPF: conditining on SPF nowcast, CQ: conditioning on current quarter information, FC: full conditioning on SPF nowcast and current quarter information). RMSEs are reported relative to the RMSE of the forecasts computed with the Smets and Wouters model (SW07), for which absolute RMSEs are reported in the last column of the last panel. This model is one of the most well-known DSGE models. Many models used at central banks are variants of it. Further, some of the post-crisis models directly build on the Smets and Wouters model. For example, the DNGS15 model that has been shown to yield quite accurate recession forecasts during the acute phase of the Great Recession consists of the Smets and Wouters model (with very small adjustments) plus a financial accelerator mechanism. A relative RMSE that is lower (higher) than one suggests that the forecast based on a particular model is more (less) accurate than the forecast from the Smets and Wouters model for a specific forecast horizon and information scenario, which is indicated by using green (red) color. A comparison to the average RMSE of all structural models is shown in the Appendix and yields overall similar results. Further, a comparison to the RMSE of the mean SPF is also reported in the Appendix.

The first panel shows the relative RMSEs of the pre-crisis models, the second panel shows those of the post-crisis models (continued in the third panel), and the third panel shows those relative RMSEs for three Bayesian VAR counterparts and the mean SPF forecast. The mean SPF forecasts is shown relative to scenario CQ (conditioning on current quarter information) of the Smets and Wouters model. This scenario is most closely comparable between the SPF and models as both have access to within quarter information in this scenario. The (absolute) RMSEs of Smets and Wouters model forecasts are shown in the last column of the third panel.

Table 2: RMSE relative to RMSE of Smets and Wouters model, 2008:III–2009:II

Note: The table shows the RMSEs relative to average the RMSEs of all the macroeconomic models for the GDP growth forecasts on five horizons in four scenarios. Numbers that are lower than one are displayed in green and numbers that are higher than one are displayed in red. The (absolute) average RMSEs of the SW07 models in four scenarios are shown in the last column.

Source		DS04				WW11				FU20				GSW12				FRBEDO08				Fair
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	
	0	1.20	-	1.32	-	1.35	-	1.40	-	1.36	-	1.04	-	1.07	-	0.82	-	1.42	-	0.91	-	1.08
	1	1.01	1.00	0.83	0.82	1.05	1.07	0.87	0.90	1.20	1.14	1.04	1.07	1.00	0.94	0.87	0.94	1.11	1.07	0.92	0.84	1.01
Horizon	2	0.98	1.04	0.87	0.92	0.97	1.06	0.87	0.93	1.13	1.14	1.03	1.05	0.90	0.87	0.79	0.85	1.05	1.12	1.02	0.85	1.04
	3	1.03	1.39	1.03	1.18	0.97	1.34	0.97	1.10	1.02	1.29	1.01	1.12	0.67	0.67	0.70	0.61	0.82	1.16	1.26	1.02	1.11
	4	0.76	1.02	1.59	1.97	0.69	0.93	1.45	1.73	0.64	0.88	1.11	1.24	1.01	1.32	1.44	1.77	0.78	0.89	1.93	1.86	0.76

Source		NKBGG				QPM08				DNGS15				KR15_FF				KR15_HH				CMR14			
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC
	0	1.39	-	1.19	-	1.46	-	1.16	-	0.88	-	0.44	-	1.13	-	1.40	-	1.13	-	1.27	-	1.16	-	0.69	-
	1	1.13	1.11	0.91	0.94	1.03	1.06	0.85	0.96	0.88	0.92	0.68	0.78	1.19	0.92	1.02	1.05	1.14	0.92	1.06	0.83	1.09	1.02	0.81	0.87
Horizon	2	1.00	1.08	0.89	0.94	0.90	1.00	0.80	0.90	0.89	1.06	0.78	0.92	1.02	0.97	0.92	0.90	1.12	1.00	1.15	1.03	1.17	1.16	1.00	1.00
	3	0.85	1.14	0.83	0.90	0.79	1.05	0.72	0.83	1.17	1.82	1.17	1.43	0.80	1.05	0.82	0.78	0.96	1.51	1.32	1.23	1.28	1.63	1.27	1.33
	4	0.68	0.86	1.32	1.43	0.97	1.13	1.90	1.93	1.26	1.66	2.55	2.97	0.64	0.87	1.32	1.45	1.07	2.10	2.25	3.04	0.72	0.92	1.68	1.90

Source		IN10				3vBVAR				5vBVAR				8vBVAR				SPF	SW07			
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC		BP	SPF	CQ	FC
	0	0.85	-	0.95	-	1.47	-	1.27	-	1.42	-	1.09	-	1.16	-	0.87	-	0.48	4.07	-	4.78	-
	1	0.76	-	0.65	-	1.06	1.03	0.89	0.92	1.11	1.04	0.89	0.93	1.10	1.10	0.92	0.97	0.65	6.03	5.68	7.09	6.30
Horizon	2	0.78	-	0.68	-	1.08	1.10	0.93	0.95	1.17	1.19	1.01	1.01	1.18	1.22	1.05	1.04	0.76	5.36	4.87	5.98	5.52
	3	0.87	-	0.89	-	1.36	1.74	1.25	1.29	1.69	2.13	1.60	1.53	1.66	2.04	1.66	1.51	0.9	2.62	1.97	2.69	2.51
	4	0.85	-	1.81	-	0.94	1.50	1.77	2.45	1.37	1.85	2.71	2.92	1.29	1.62	2.65	2.57	1.51	2.37	1.83	1.21	1.09

The absolute RMSEs of the Smets and Wouters model 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 relative performance of the different models differs with respect to short-term (0 to 2 quarters ahead) and medium term (3 to 4 quarters ahead) forecasts. Regarding short-term forecasts, survey based benchmarks outperform most DGSE and BVAR based forecasts, while DSGE models do improve upon the BVARs providing evidence in support for theoretical restrictions. Among the DSGE models, in most cases post-crisis models outperform the pre-crisis benchmark Smets and Wouters model, in particular when the forecasts are conditioned on current quarter information (columns CQ and FC). Among the pre-crisis models the picture is more mixed with some models delivering more accurate and some less accurate short-term forecasts than the Smets and Wouters model. The most precise pre-crisis model forecasts are obtained by the GSW12 model with its focus on labor market dynamic. Among the post-crisis models, the very low nowcast RMSE of the DNGS15 model when being conditioned on current quarter information stands out. These nowcasts are even more accurate than the ones from the SPF reflecting that in 2008Q3 and 2009Q1 the model very precisely predicts current GDP growth. These highly accurate nowcasts in turn lead to highly accurate one and two quarter ahead forecasts. Only those from the IN10 model are similarly accurate. The success of these two models highlights the importance of choosing an appropriate model structure and observables that inform the model in a timely manner about financial market disruptions for accurate predictions during financial crises.

The less accurate forecasts from some other post-crisis models implies that having a superior model structure is not enough for obtaining accurate forecasts. According to a special survey conducted by the Philadelphia Fed, professional forecasters usually combine mathematical models plus their subjective considerations when making projections (Stark, 2013). The subjective judgement of the forecasters might explain the superior performance of short-run forecasts of professional forecasters (SPF) compared to most model forecasts. There are reasons to believe that a series of extreme events, including the subprime mortgage crisis, the collapse of the Lehman Brothers and the turbulence in the financial markets, have all left deep impression in forecasters' minds, which could result it rather pessimistic forecasts in the second half of 2008. However, these extreme events cannot be fully reflected in the data that we use to estimate DSGE models, in particular in aggregate data series that are updated only once or twice in a quarter.

Regarding medium-term forecasts, those from professional forecasters are not more precise than model-based forecasts. GDP growth is a stationary process that has the tendency of reverting back to its steady-state value. Therefore, the information advantage of professional forecasters might become a drag rather than a help, as they might place too much weight on those signals, which are helpful in predicting GDP growth in the short run, but cannot reflect long-run trends. It is also worth noting that, all post-crisis models that excel in generating accurate short-term forecasts, cannot maintain their performance in the medium run, as their forecasts of the speed of recovery are all too pessimistic. This might be caused by the nominal and real rigidities of those models that lead to a too persistent impact of adverse financial shocks.

Overall, the pre-crisis models deliver better long-horizon forecasts, while the post-crisis models deliver better short-term forecasts. This reflects that modeling and including measures of financial conditions is in particular relevant for forecasting recessions during financial crises, while during other times models without financial frictions might be more suitable.

5 Forecast Performance During two Additional Recessions

The results in the above section show that models with financial market frictions are able to detect the 2008/09 Great Recession at its onset. In addition, we highlight the importance of employing appropriate model structures, as well as using data that contain latest financial markets developments for generating accurate forecasts during the Great Recession. To check whether these findings also hold for other recession, we investigate the forecasting performance of macroeconomic models during the 2001 and the 2020 recessions.

5.1 Forecasting the 2001 Recession

For the 2001 recession we compute forecasts with starting points from 2001Q1 to 2001Q4. Figures 10 and 11 show the forecasts from the pre-crisis models and Figures 12 and 13 show the forecasts from the post-crisis models. The 2001 recession was a very mild one. GDP growth started decreasing in 2001Q2 and it turned negative in 2001Q2 before returning to positive growth in 2001Q3 followed by strong recovery afterwards.

In 2001Q1 at the onset of the 2001 recession, regarding the pre-crisis model only the GSW12 model that focuses on frictions in the labor market predicts a growth slowdown to zero percent over the next year. The other models are divided into two groups. One group including all medium scale models except the FRBEDO model predicts a continuation of low growth around one percent, while the other group consisting of the FRBEDO model and two small-scale models predicts a quick return to steady state growth rates. These results are similar across the four information scenarios. The forecasts from the post-crisis models are not systematically different from those of the pre-crisis models. Among the post-crisis models two medium-scale models, the CMR14 and the DNGS15 model, yield similar low growth forecasts compared to the medium-scale pre-crisis models and the IN10 model even predicts negative growth rates. While this is the only model that predicts a recession, its forecasts have nevertheless unfavorable properties. The model predicts throughout all the forecast starting points that we analyse for the 2001 recession a very long recession, returning to steady growth only after 40 quarters, so that these forecasts are not very plausible.[5] The other post-crisis models yield quite optimistic forecasts similar to those of the more optimistic pre-crisis models. While none of the models predicts the upcoming recession (except for the implausible predictions of the IN10 model), professional forecaster did neither, confirming the results in (Wieland and Wolters, 2012).

Moving to the second quarter of 2001, the mean SPF predicts a growth slowdown that turned out to be a too optimistic prediction. Afterwards, the professional forecasters falsely predict that the economy directly starts recovering. Among the pre-crisis models, the SW07 and the FU20 model predict a decline for the current quarter and are more pessimistic than the mean SPF forecast over the medium run, though they still predict positive growth rates, while actual growth rates turned negative in 2001Q3. These two models are very similar and their low growth forecasts hold across all four information scenarios. All other pre-crisis models predict a return to steady state growth rates rather than a deepening of the recession. There are two exception though. When being conditioned on current quarter data, the GSW12 model and the FRBEDO08 model predict a rather precise nowcast that is much closer to the actual data than the SPF mean nowcast.

[5] We checked in detail the estimation to figure out whether there is a problem with the mode or the convergence of the MH-algorithm of the IN10 model. Since we are using a Monte-Carlo based optimization routine to obtain a suitable starting point for the MH-algortihm (mode computation algorithm 6 in the Dynare package), and we run 1.000.000 draws and have checked convergence statistics, the posterior distribution should be fine. Nevertheless, from a practitioner view, forecasts that need 40 quarters to return back to steady state are not very plausible.

The GSW12 model is informed by employment and unemployment data, while for the FRBEDO08 model the relative detailed information on investment and consumption in different sectors might be useful. Again, there are no systematic differences between pre-crisis and post-crisis models, but for both model types there are models that yield low growth forecasts and some that yield more optimistic forecasts. Among the post-crisis models, again the CMR14, the DNGS15 and the IN10 model yield the most pessimistic forecasts that are even slightly more accurate than the most accurate pre-crisis model forecasts over the short run. When being conditioned on current-quarter information, the nowcasts of the CMR14 and the DNGS15 model are quite precise, similar to those from the FRBEDO08 and the GSW12 model, while the one of the IN10 model is even too pessimistic. The relatively precise short-term forecasts come, however, at the cost of being way too pessimistic over the medium term. While there was a very strong recovery in 2002Q1, these model continue instead predicting very low growth rates due to the highly persistent forecasts. The other post-crisis models predict a quick return to steady state growth rates rather than a deepening of the recession, similar to the group of overly optimistic pre-crisis models.

Regarding forecasts starting in 2001Q3, neither the pre-crisis models nor the mean SPF foresee the continued decline of GDP growth to negative growth rates. Among the pre-crisis models, the most pessimistic forecast is produced by the GSW12 model as it accounts for worsening labor market conditions. When being conditioned on current quarter information, the nowcast of this model is, however, puzzling, as it switches to highly positive growth rates.[6] Among the post-crisis models, the CMR14, the DNGS15 and the IN10 models yield again the most pessimistic forecasts, while the other models predict a relatively quick recovery. Hence, again the divide is not between pre- and post-crisis models, but within both model classes there are models that produce rather pessimistic and models that produce rather optimistic forecasts. As the recovery started in 2001Q4, the pessimistic forecasts are more accurate for the current quarter, while the optimistic forecasts are more accurate for the following quarters.

In 2001Q4 GDP growth turned positive with very high growth rates of almost 5 percent following in the quarter afterwards. Interestingly, the mean SPF nowcast did not predict this recovery, but a further deepening of the recession to -2.2 percent growth in 2001Q4 and a stagnation of GDP for the following quarter. None of the models predicts the very high growth rates of 2002Q1, though some of the post-crisis models yield quite optimistic forecasts. When being conditioned on current quarter information, the FRBEDO08, the GSW12, the DNGS15 and the CMR14 model continue to predict negative growth rates for the current quarter, though the nowcasts are not as pessimistic as the SPF mean nowcast except for the ones of the DNGS15 and the IN10 model.

To sum up, there is no systematic difference between pre- and post-crisis model forecasting performance during the 2001 recession. In both groups of models there are some models yielding rather pessimistic forecasts, though hardly forecasts of negative growth rates, while other yield rather optimistic forecasts. The former have some advantages with regards to short-term forecasting accuracy, while the others perform better for medium-term forecasts, because the recession was relatively short with a strong recovery following afterwards. The 2001 recession was not accompanied by substantial turmoil in the financial or real estate sector, so that credit spreads did not increase much. Hence, it is not too surprising that there is no systematic difference in forecasting accuracy between pre- and post-crisis models. Further, a number of both pre-crisis and post-crisis models surpass the mean SPF in forecasting accuracy throughout the 2001 recession. This finding is rather reasonable, because when the objective is to forecast a relatively mild recession, the information advantage that professional forecasters have are no longer essential for producing

[6]Similar to the analysis done for the puzzling forecasts of the IN10 model, we checked the estimation results for the GSW12 model in detail and did not find any convergence problems.

accurate forecasts. On the contrary, some information might have even been misleading. For example, there are reasons to believe that the negative nowcast from the professional forecasters in 2001Q4 is affected by worries that the turbulence caused by the 9-11 terrorist attacks might have severe consequence for the US economy.

We also compare the forecasts from structural and time-series models during the 2001 recession. Figures 14 to 17 display the forecasts from two structural models (NKBGG and DNGS15) and two time-series models (GLP5v and GLP8v) that have the same observables, respectively. The forecasts from the NKBGG model are systematically lower than those from the GLP5v model. As a result, the NKBGG performs better than the GLP5v model during the first three quarters of the recession when GDP growth kept declining, while it performs worse during the last quarter when GDP growth started to increase. Similarly, the forecasts from the DNGS15 model are also lower than that from the GLP8v model throughout almost all forecast staring points and information scenarios. These findings suggest that the theoretical restrictions help the macroeconomic models in forecasting the 2001 recession.

Finally, we evaluate the forecasting performance more systematically based on RMSEs. Table 3 shows RMSEs relative to those from the Smets and Wouters model. The table is structured in the same way as the RMSE table for the 2008/09 recession in the previous section. The numbers confirm that there is no advantage for post-crisis compared to pre-crisis models during the 2001 recession. Overall, the post-crisis models perform even worse (see, e.g., the RMSEs for the DNGS15 model). On the one hand, even for those forecast starting points where some of the post-crisis models yield quite accurate short-term forecasts, this comes at the cost of inaccurate medium-term forecasts. On the other hand, even the nowcasting and short-term forecasting accuracy of the post-crisis models is worse than the one of the Smets and Wouters model (and some other pre-crisis models) due to the highly inaccurate nowcasts for 2001Q4 when GDP was recovering, but the models predicted a deepening of the recession. For the same reason the SPF short-term nowcasts turn out to be much less accurate than those of the Smets and Wouters model. Similar to the 2008/09 recession, the usage of structural models seems to be advantageous compared to using Bayesian VARs for forecasting the 2001 recession.

Figure 10: GDP Growth Forecasts in 2001:I–2001:II: Pre-Crisis Models

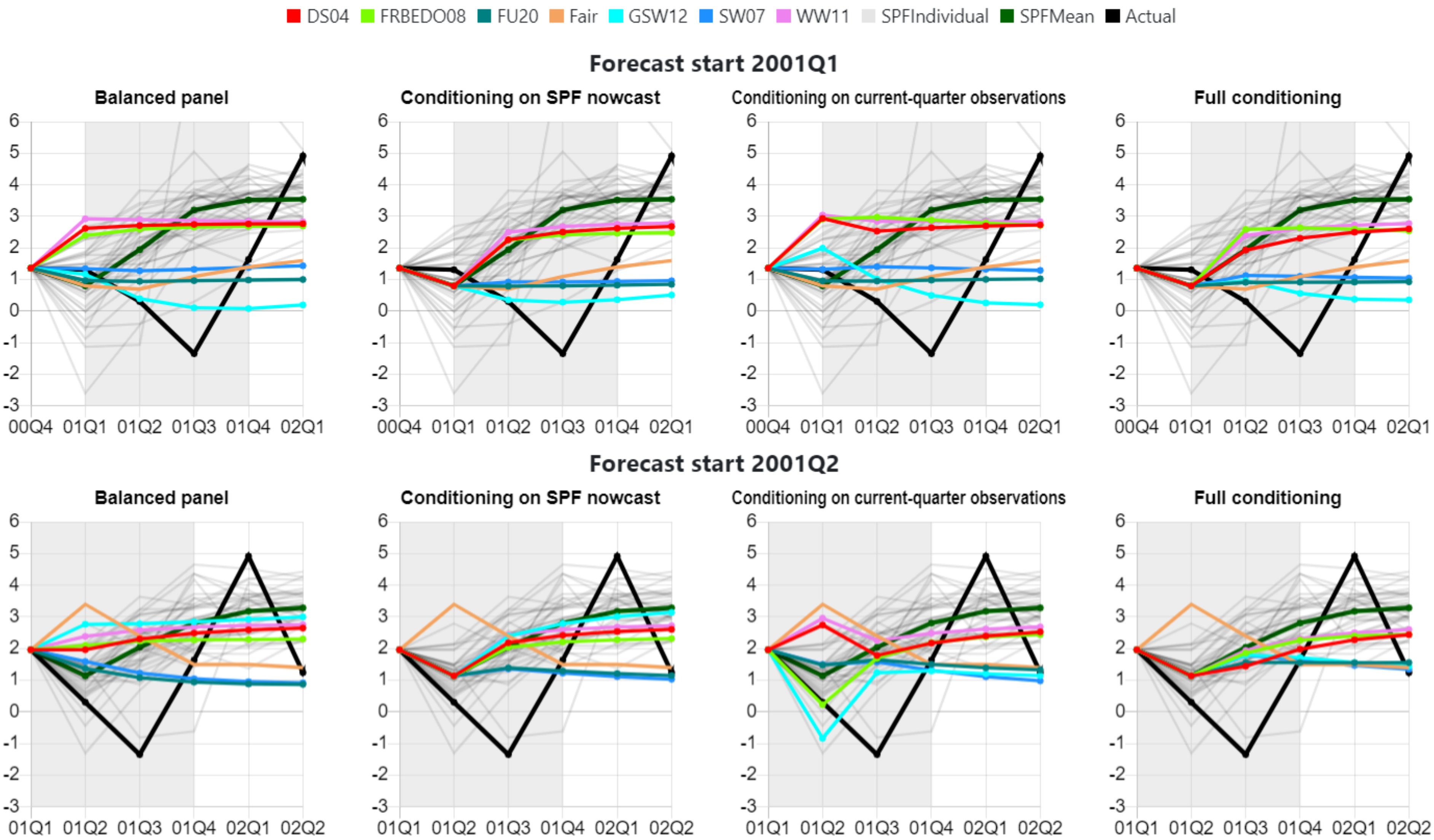


Figure 11: GDP Growth Forecasts in 2001:III–2001:IV:Pre-Crisis Models

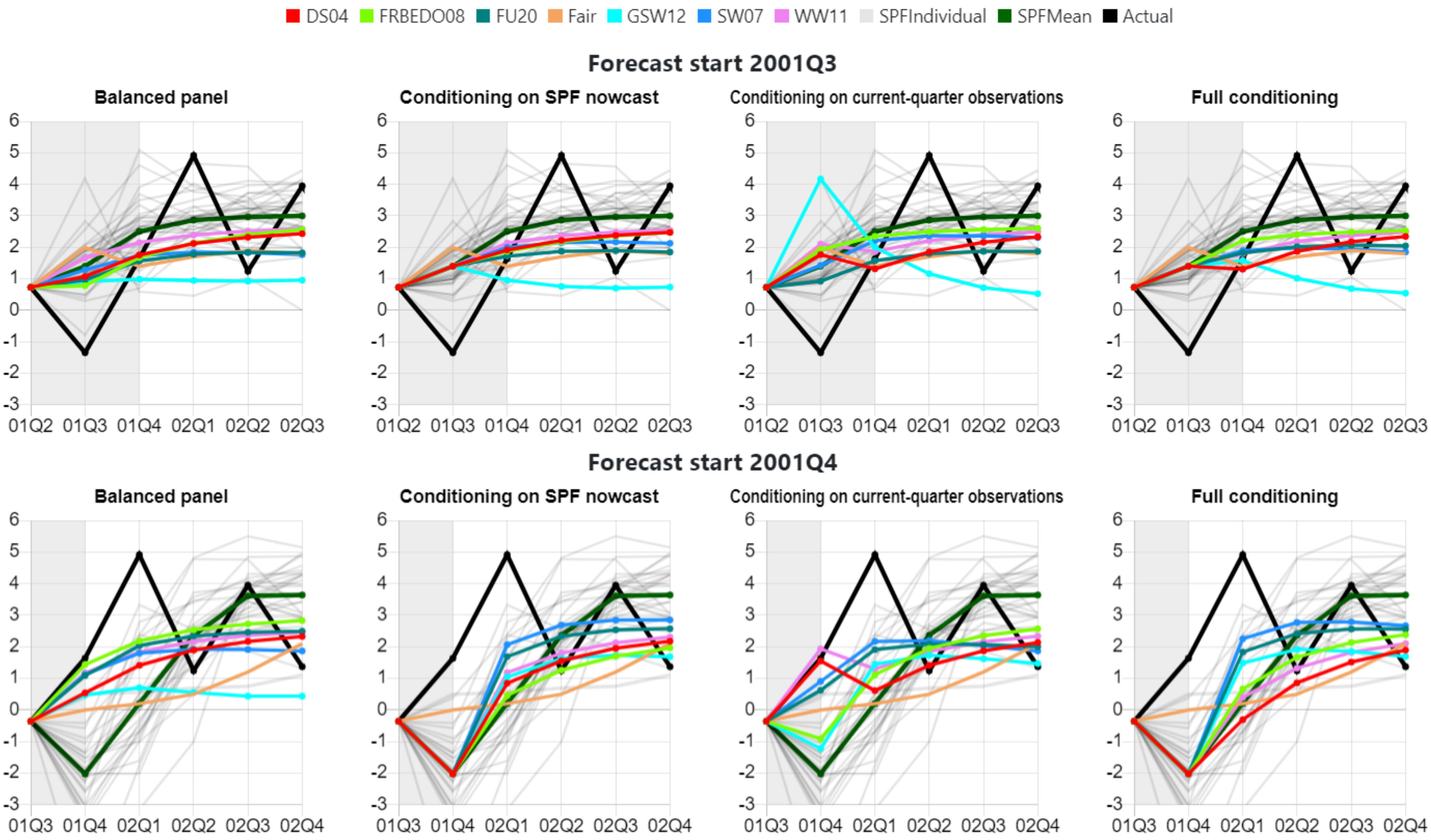


Figure 12: GDP Growth Forecasts in 2001:I–2001:II: Post-Crisis Models

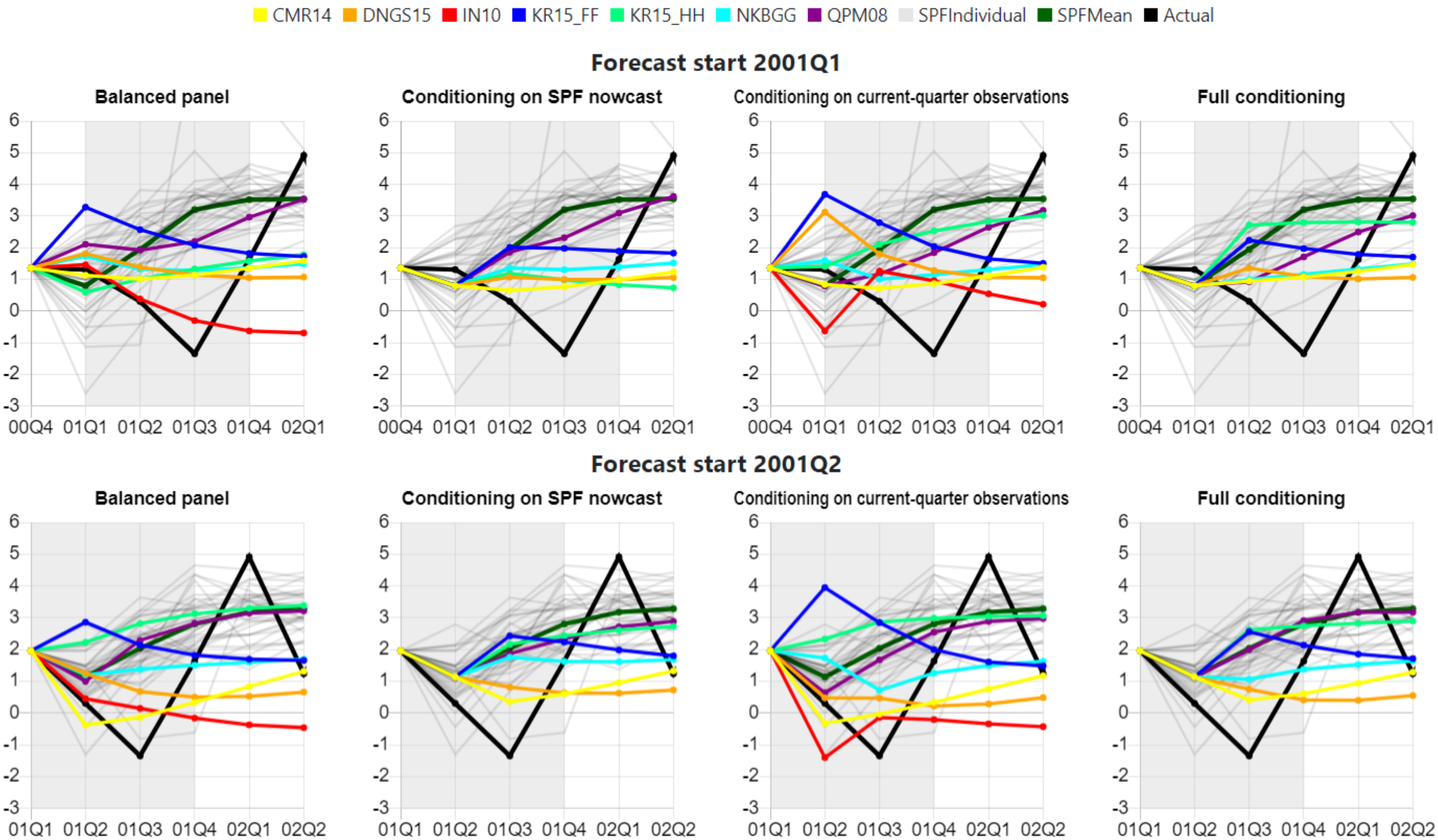


Figure 13: GDP Growth Forecasts in 2001:III–2001:IV:Post-Crisis Models

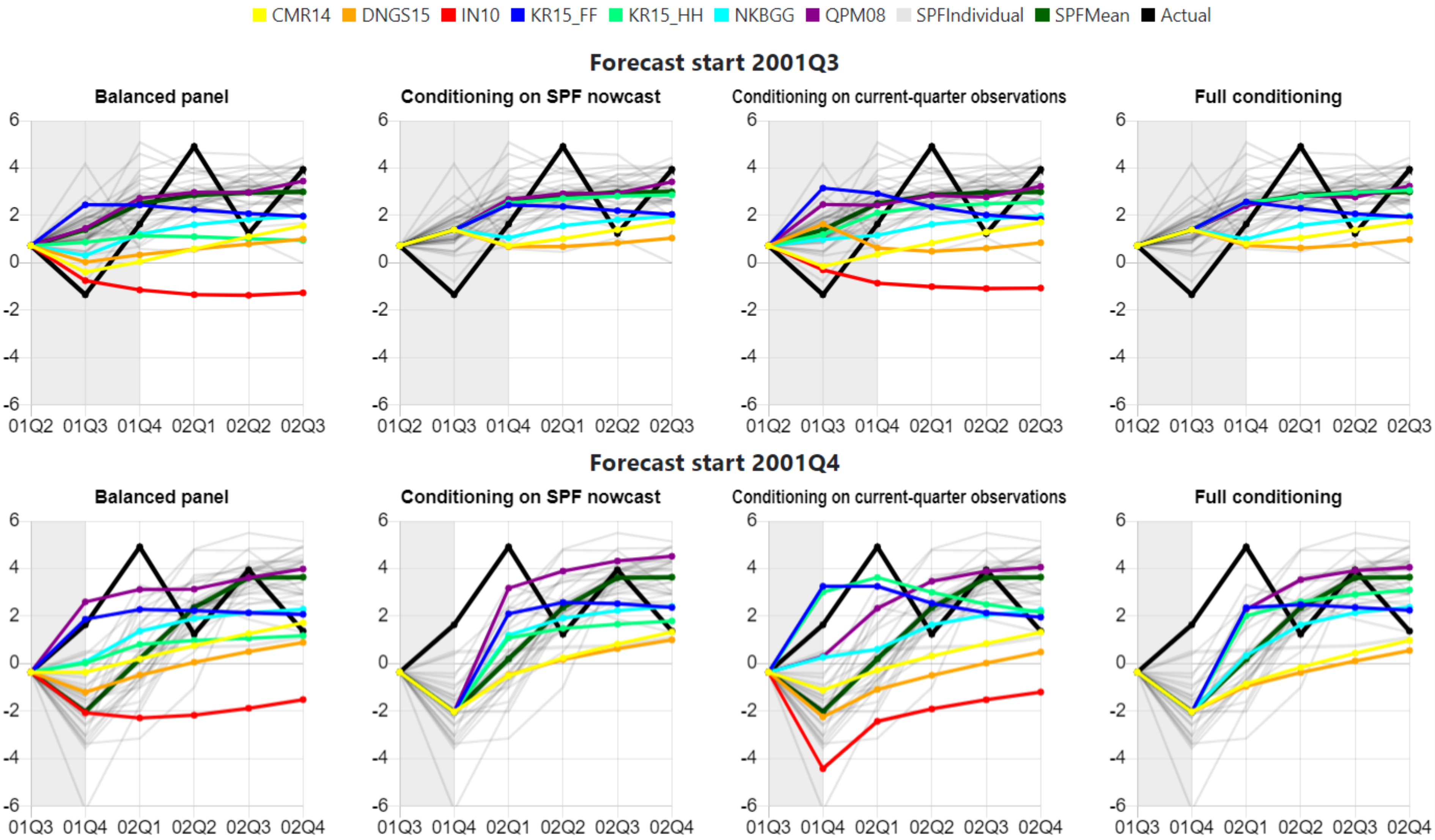


Figure 14: GDP Growth Forecasts in 2001:I–2001:II: 5-Variable NK Model versus 5-Variable Bayesian VAR Model

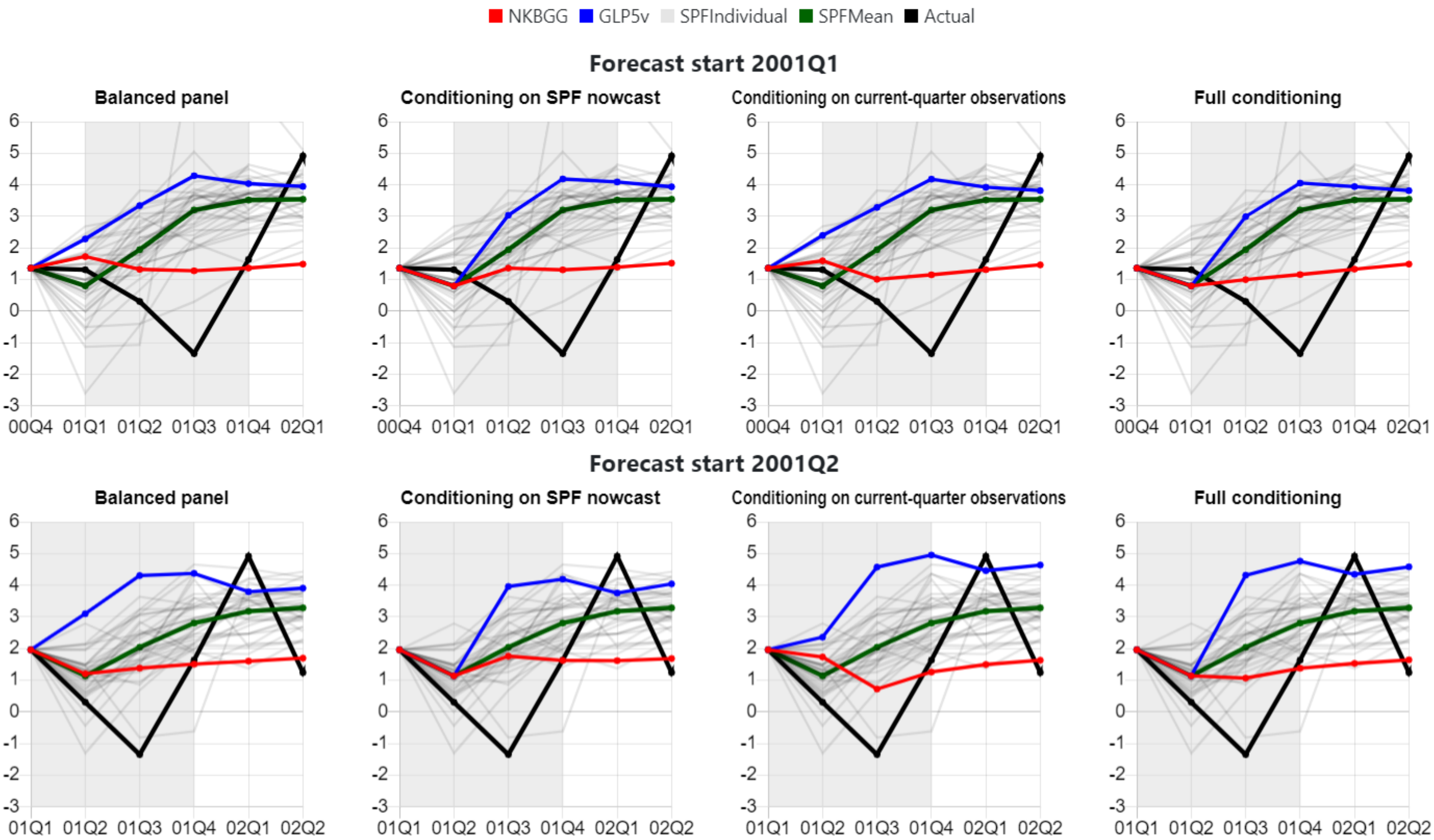


Figure 15: GDP Growth Forecasts in 2001:III–2001:IV: 5-Variable NK Model versus 5-Variable Bayesian VAR Model

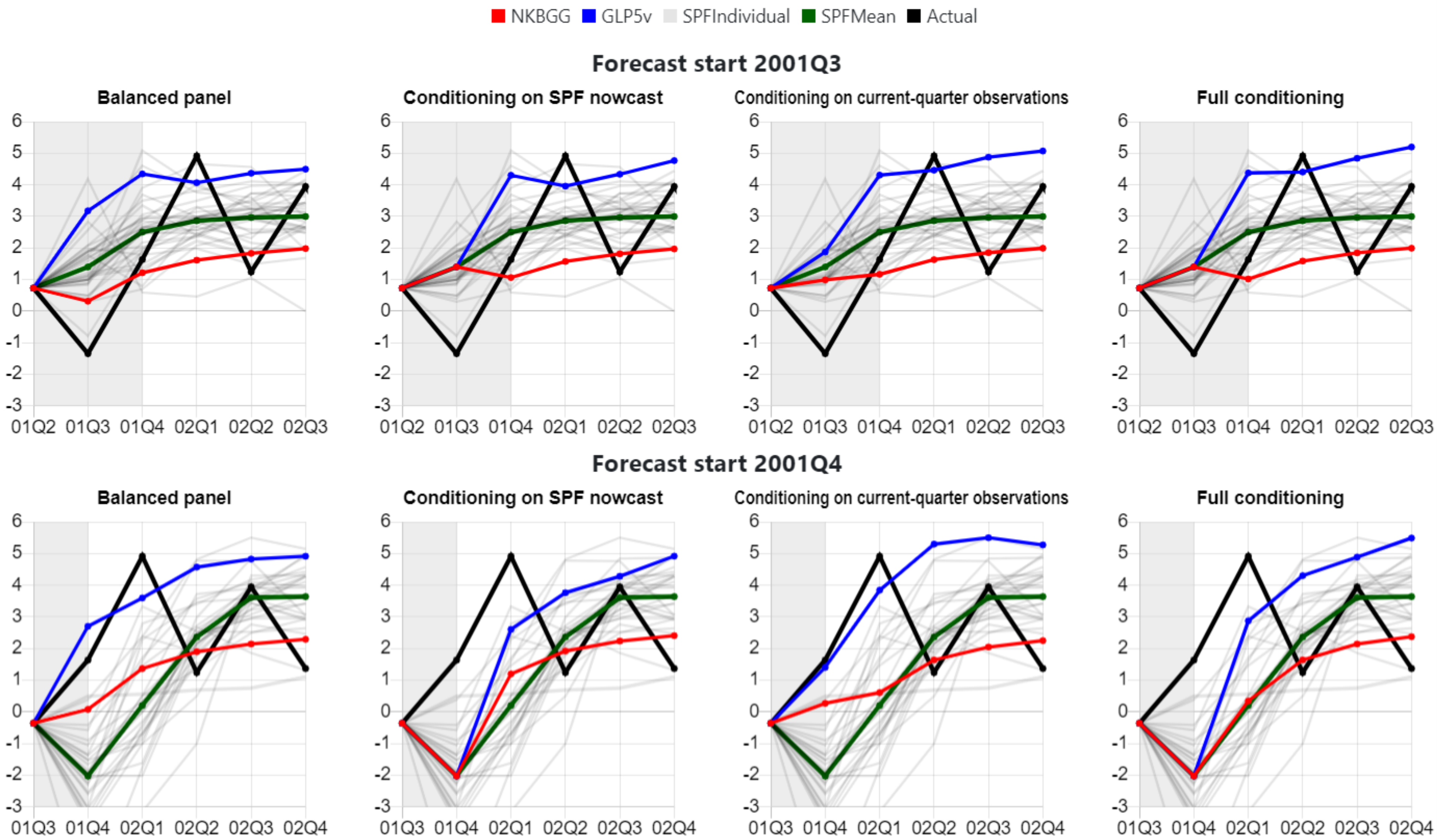


Figure 16: GDP Growth Forecasts in 2001:I–2001:II: 8-Variable NK Model versus 8-Variable Bayesian VAR Model

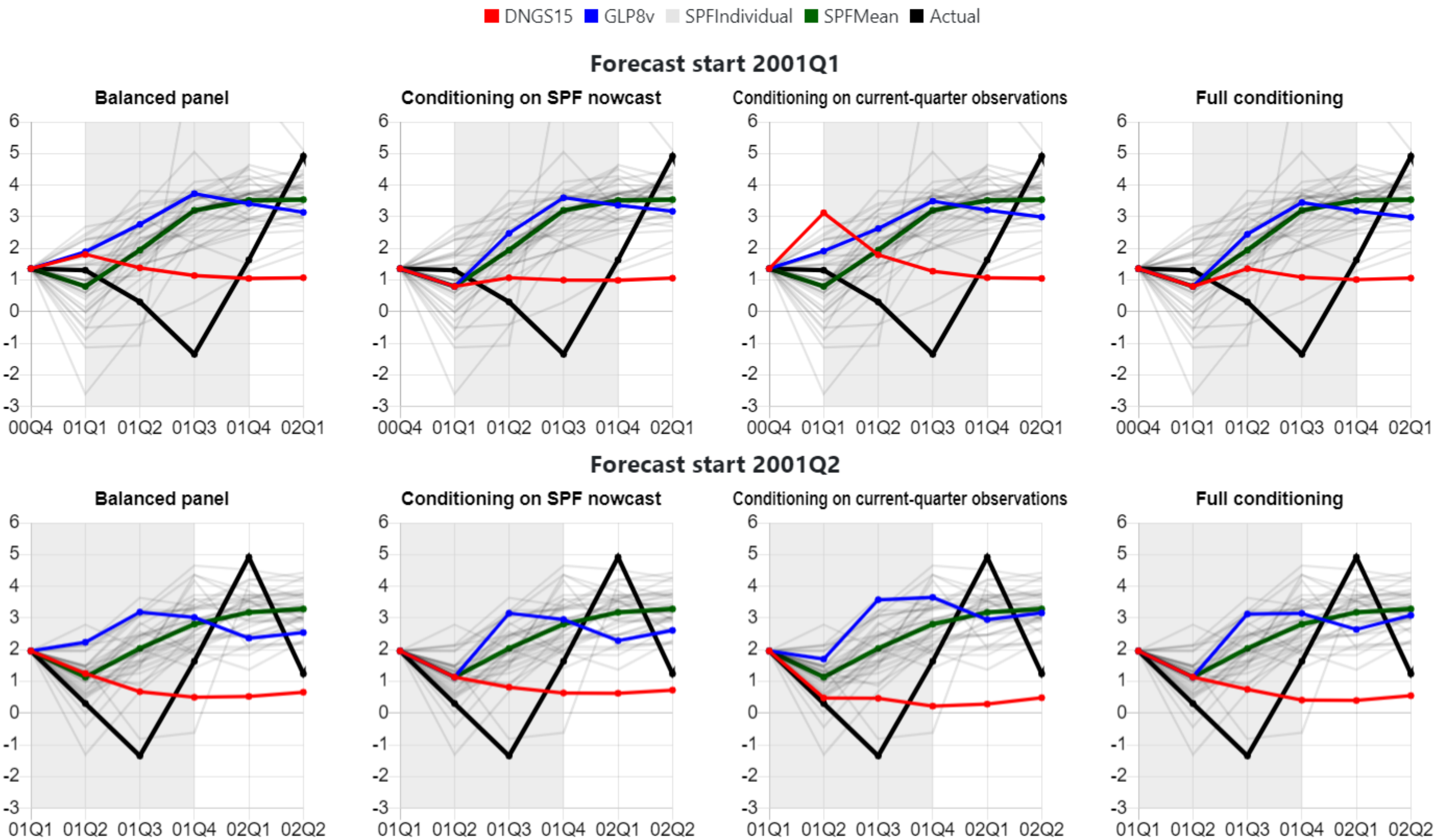


Figure 17: GDP Growth Forecasts in 2001:III–2001:IV: 8-Variable NK Model versus 8-Variable Bayesian VAR Model

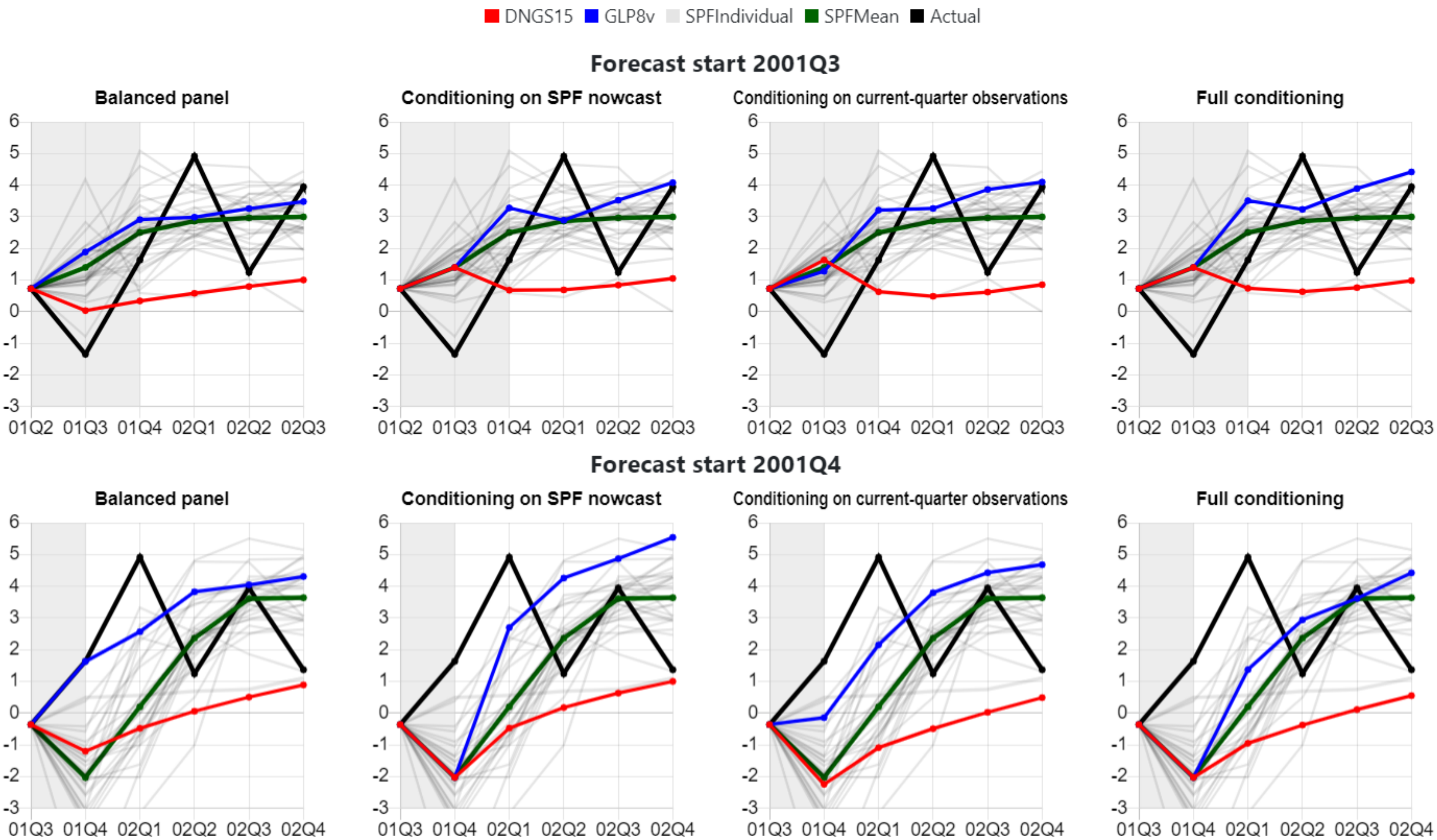


Table 3: Relative Root Mean Squared Errors (RMSE) for the GDP Growth Forecast in 2001:I–2001:IV

Note: The table shows the RMSEs relative to average the RMSEs of all the macroeconomic models for the GDP growth forecasts on five horizons in four scenarios. Numbers that are lower than one are displayed in green and numbers that are higher than one are displayed in red. The (absolute) average RMSEs of the SW07 models in four scenarios are shown in the last column.

Source		DS04				WW11				FU20				GSW12				FRBEDO08				Fair
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	
	0	1.14	-	1.38	-	1.35	-	1.51	-	0.88	-	0.89	-	1.19	-	2.05	-	1.01	-	1.43	-	1.64
	1	1.35	1.44	1.38	1.46	1.36	1.44	1.36	1.42	0.92	1.07	1.02	1.02	1.43	1.36	1.05	1.12	1.22	1.48	1.33	1.38	1.45
Horizon	2	1.22	1.23	1.31	1.16	1.23	1.26	1.32	1.18	0.99	1.00	1.03	0.94	1.07	1.19	1.10	1.07	1.22	1.21	1.28	1.16	0.99
	3	0.74	0.84	0.81	1.02	0.72	0.81	0.76	0.96	0.98	0.99	0.93	1.01	0.96	0.78	1.05	1.12	0.73	0.91	0.79	0.93	0.99
	4	0.75	0.68	0.77	0.68	0.75	0.67	0.77	0.66	1.11	1.03	1.11	1.00	1.43	1.26	1.45	1.24	0.77	0.67	0.77	0.69	0.97

Source		NKBGG				QPM08				DNGS15				KR15_FF				KR15_HH				CMR14			
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC
	0	0.83	-	0.99	-	1.06	-	1.32	-	1.12	-	1.68	-	1.12	-	2.08	-	1.12	-	1.14	-	0.78	-	1.00	-
	1	1.11	1.25	1.16	1.25	1.08	1.03	0.99	1.04	1.44	1.49	1.56	1.52	1.45	1.85	1.59	1.76	1.47	1.85	1.03	1.76	1.24	1.44	1.32	1.45
Horizon	2	1.02	1.11	1.08	1.02	1.11	1.29	1.17	1.10	1.26	1.30	1.45	1.29	1.15	1.12	1.34	1.10	1.11	0.91	1.35	1.18	1.24	1.20	1.27	1.18
	3	0.85	0.91	0.90	1.04	0.63	0.77	0.62	0.66	1.25	1.33	1.39	1.60	0.82	0.78	0.78	0.89	0.77	0.93	0.59	0.86	1.09	1.23	1.18	1.42
	4	0.98	0.89	1.02	0.89	0.87	0.83	0.92	0.84	1.18	1.06	1.27	1.08	0.90	0.82	0.97	0.83	1.18	0.97	0.82	0.70	0.99	0.93	1.04	0.90

Source		IN10				3vBVAR				5vBVAR				8vBVAR				SPF	SW07			
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC		BP	SPF	CQ	FC
	0	1.26	-	2.15	-	2.25	-	1.75	-	1.85	-	1.28	-	1.28	-	1.13	-	1.51	1.49	-	1.55	-
	1	1.89	-	1.88	-	1.64	1.65	1.73	1.65	1.70	1.74	1.72	1.70	1.40	1.43	1.50	1.52	1.45	2.08	2.00	2.10	2.10
Horizon	2	1.78	-	1.90	-	1.66	1.72	1.93	1.67	1.71	1.72	1.98	1.69	1.48	1.62	1.57	1.35	1.36	2.08	1.94	1.93	2.06
	3	1.91	-	1.81	-	0.95	0.99	1.06	1.16	0.93	1.00	1.04	1.18	0.82	0.97	0.83	1.02	0.7	2.25	2.06	2.21	1.87
	4	2.01	-	1.87	-	1.12	1.06	1.39	1.18	1.10	1.02	1.35	1.21	0.89	1.03	1.07	0.89	0.87	2.07	2.30	2.01	2.30

5.2 Forecasting the 2020 Recession

In this section, we focus on the forecasting performance of the models in the ongoing recession that is caused by the the COVID-19 pandemic. Since this time the economic disruption originates from the spread of the virus, which is not modelled in any of the macroeconomic models introduced in this paper, it is unrealistic to except that these models are able to predict the contraction before or at the beginning of the crisis. A relatively accurate nowcast might come from models that take current-quarter information into account. Further, the models could potentially be suited to forecast the recovery after the recessionary shocks have been observed. Given this, we will mainly examine (1) whether including current-quarter observations in the data help in forecasting the recession, and (2) how the different models predict the speed of temporary recovery in 2020Q3.

Figures 18 to 21 show forecasts starting in 2020Q1, Q2 and Q3 for the pre-crisis and post-crisis models, respectively. Unsurprisingly, for the forecasts starting from 2020Q1, neither pre-crisis models, post-crisis models, nor professional forecasters predict a deep recession over the next quarters. Including current-quarter observations does not change this result.

As the severity of the crisis became obvious during spring 2020, the nowcast from the mean SPF drops to -38.21 percent in 2020Q2. This nowcast is extremely accurate, given that the actual GDP growth in the second quarter turned out to be -38.13 (as of August 27th, 2020). Under the balanced panel information scenario, no pre-crisis nor post-crisis model can predict the severe deepening of the recession, but the forecasts from these models simply revert back to the steady state over the following quarters. When conditioning on the SPF nowcast in the second and fourth scenarios, the forecasts from the pre-crisis models and most post-crisis over the following quarters turn out to be too pessimistic. Having observed highly adverse shocks, the estimated high persistence of most shock processes leads to negative growth rate predictions in some cases even beyond 2021Q2. There is one exceptions though. The QPM08 model predicts a quick recovery similar to the SPF mean forecast for 2020Q3 and the quarters afterwards. While the QPM08 forecast for 2020Q3 turned out to be highly accurate, the SPF forecast was too pessmistics, because actual GDP growth turned out to be even higher.

When conditioning on current-quarter observations, there are five models that yield extremely pessimistic nowcasts: FRBEDO08, QPM08, IN10, DNGS15, and GSW12. Only these models include current quarter information on the labor market. The U.S. labor market was highly adversely influenced by both the fast spread of the disease as well as subsequent countermeasures, such as travel restrictions, facility closures and other containment policies. As a consequence, the unemployment rate, an observable in the QPM08 and GSW12 models, rose from 4.4 percent in March to 14.7 percent in April 2020. Hours and employment as included in the observables in the FRBEDO08, IN10, GSW12, and DNGS15 models also dropped extremely. The resulting nowcasts are all lower than the actual drop in 2020Q2 OF -31.4 percent, ranging from -42.7 in the FRBEDO08 model to -87.15 in the GSW12 model. As opposed to the nowcasts from these models, the nowcasts from all the other models that do not include current-quarter labor market data, are much too optimistic. For example, the CMR14 model, which provided accurate nowcasts for the second half of 2008, predicts an ordinary recession only, but no extreme crisis. Obtaining precise nowcasts for 2020Q2 comes, however, at the cost of too pessimistic forecasts over the following quarters. Only the QPM08 model provides a highly negative nowcast for 2020Q2 and predicts with a growth rate of 26.97 percent a strong recovery for 2020Q3. The other models that predict highly negative growth rates for 2020Q2, continue to do so also for the following quarters.

Moving to the third quarter of 2020, the mean SPF predicts that GDP growth with 17.64 percent in the current quarter, while actual growth turned out to be even higher. Afterwards, the SPF predicts that growth rates gradually return back to steady state. Under the balanced panel assumpton, only the forecasts from the QPM08 model show similar dynamics, while the forecasts of all other models—pre- and post-crisis ones—are much more pessimistic. When being conditioned on the mean SPF nowcast almost all models predict a gradual return to average growth rates over the next quarters similar to the dynamics predicted by the SPF. When conditioning forecasts on current-quarter information, interestingly the same five models that showed highly negative nowcasts based on data for 2020Q2, predict a strong recovery when being conditioned on data for 2020Q3. The reason is again that these models account for current quarter labor market dynamics that early on indicated a strong recovery in 2020Q3. The nowcasts from the QPM08 (29.20), IN10 (21.20), and DNGS15 (18.45) models are higher than the one the mean SPF, while the nowcasts from the GSW12 (16.85) and FRBEDO08 (7.80) are lower than the one from the mean SPF which is lower than actual GDP growth (33.1). The other models are much more pessimistic with nowcasts ranging from a stagnation of GDP to negative growth rates up to -20 percent.

How do forecasts from BVARs compare to those from structural models? In Figures 22 to 25 we compare againg the NKBGG model to the GLP5v model and the DNGS15 model to GLP8v model. For forecasts staring from 2020Q1, the BVAR and DSGE models based forecasts show very similar estimates. Not surprisingly, both classes of models cannot predict the upcoming recession. For forecasts starting in 2020Q2, the BVAR and DSGE models predict a somewhat similar recovery path, although the DSGE models predict more optimistic recovery in most scenarios. It should be noted that, when conditioning on current-quarter labor market observations, only the DNGS15 but not the GLP8v can predict a crisis, as the hours worked series in the GLP8v is defined in the same way as the hours worked in the SW07 model, and its current-quarter value cannot be observed. Finally, for forecasts starting in 2020Q3, because of the presence of the extreme values in the data, the forecasts from the GLP5v model under one scenario and GLP8v under all scenarios do not converge, while the forecasts from the NKBGG and DNGS15 eventually return back to the steady states.

The above findings suggest that conditioning on labor market indicators is important in obtaining accurate nowcasts throughout the COVID-19 crisis, though all models except the QMP08 model face substantial difficulties in dealing with such a large shock. The models include highly persistent shock processes, so that a highly adverse shock in a given quarter leads to pessimistic forecasts for the following quarters rather than the prediction of a quick recovery as occurred during 2020Q3 when containment policies were eased. The forecasts from the QMP08 model are a remarkable exception. It is the only model that provides a highly negative nowcast in 2020Q3 and predicts a highly positive growth rate for 2020Q3, i.e. the growth pattern that actually occurred.

Figure 18: GDP Growth Forecasts in 2020:I–2020:II: Pre-Crisis Models

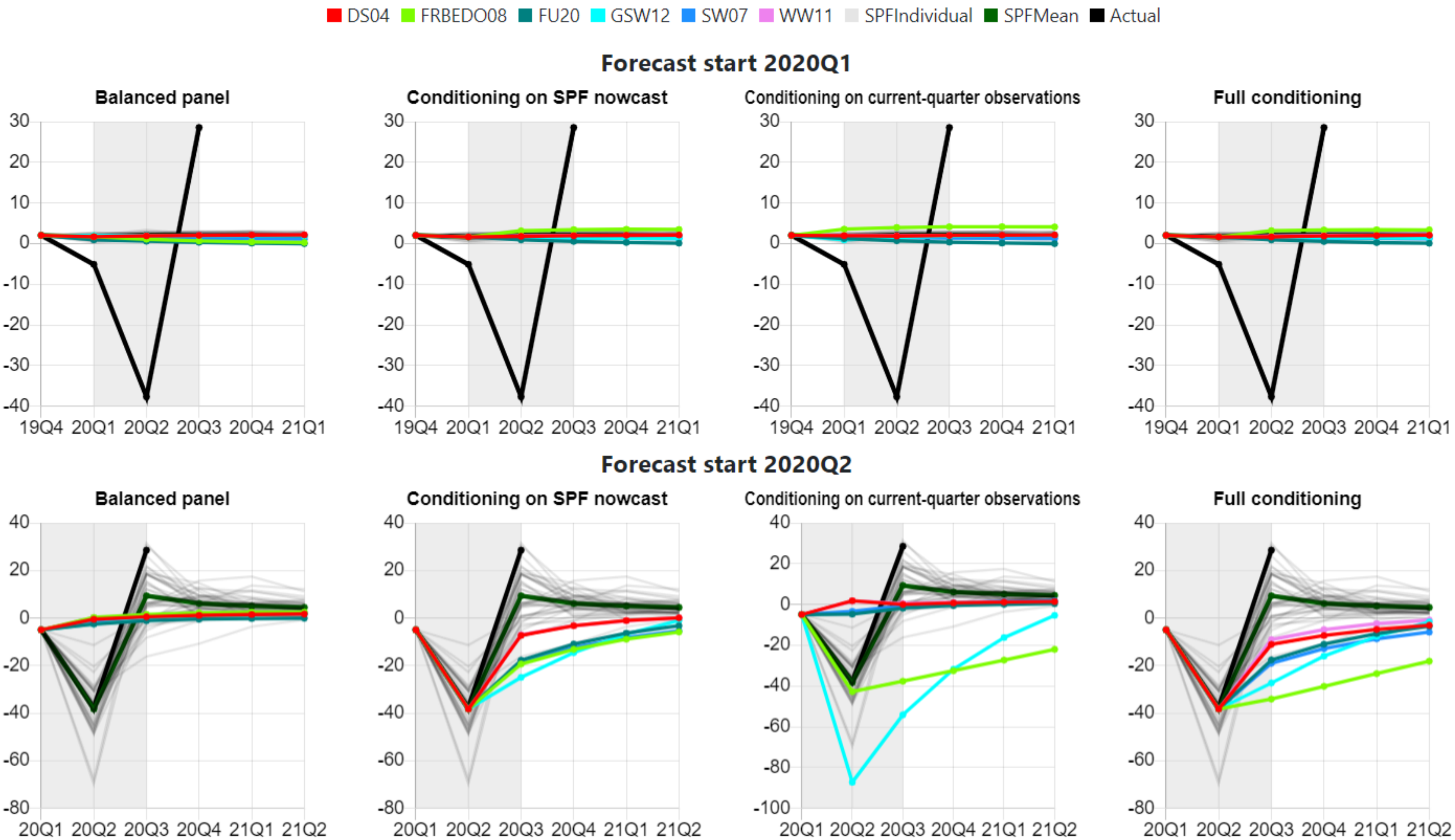


Figure 19: GDP Growth Forecasts in 2020:III: Pre-Crisis Models

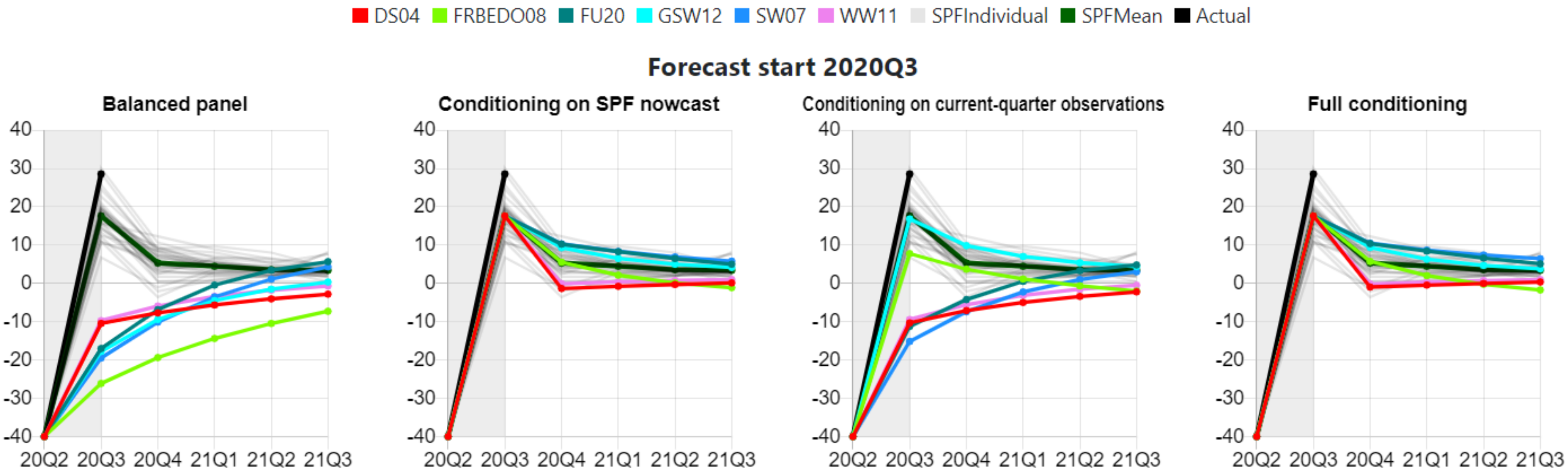


Figure 20: GDP Growth Forecasts in 2020:I–2020:II: Post-Crisis Models

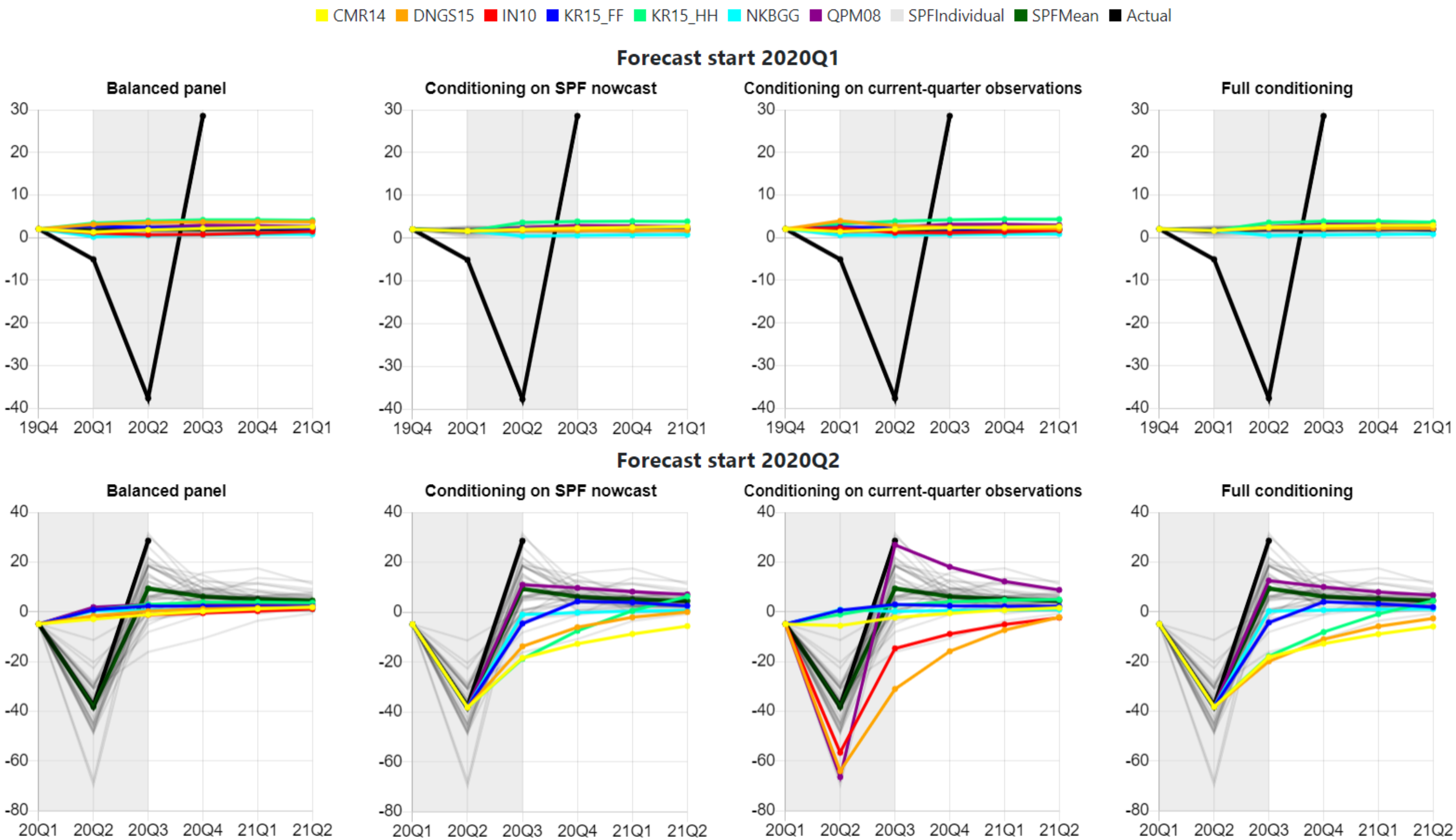


Figure 21: GDP Growth Forecasts in 2020:III: Post-Crisis Models

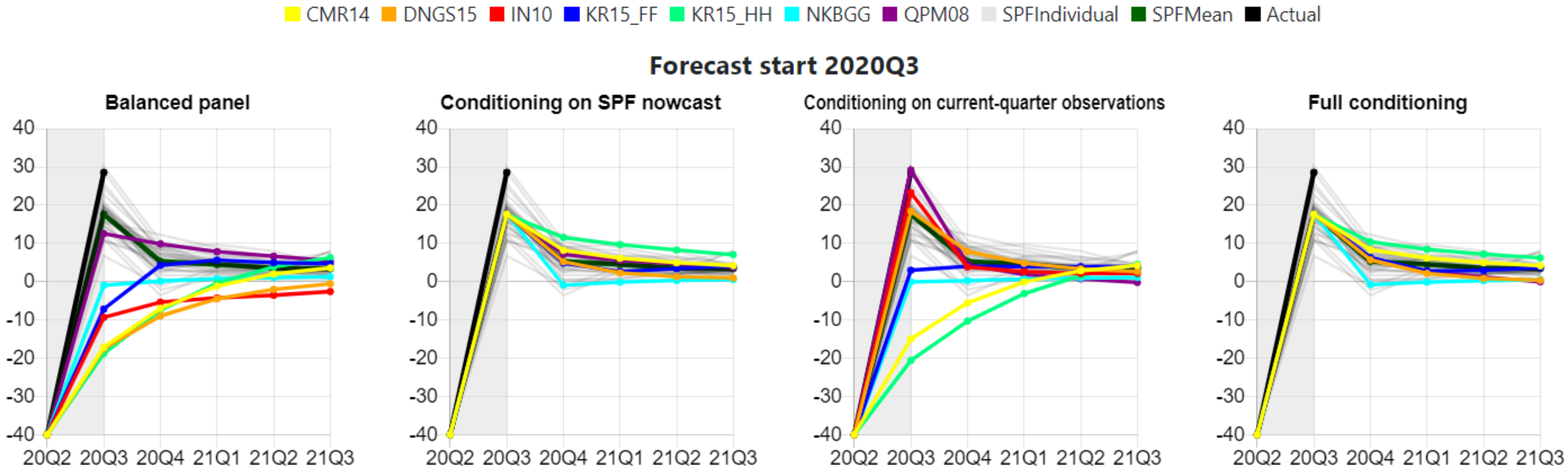


Figure 22: GDP Growth Forecasts in 2020:I–2020:II: 5-Variable NK Model versus 5-Variable Bayesian VAR Model

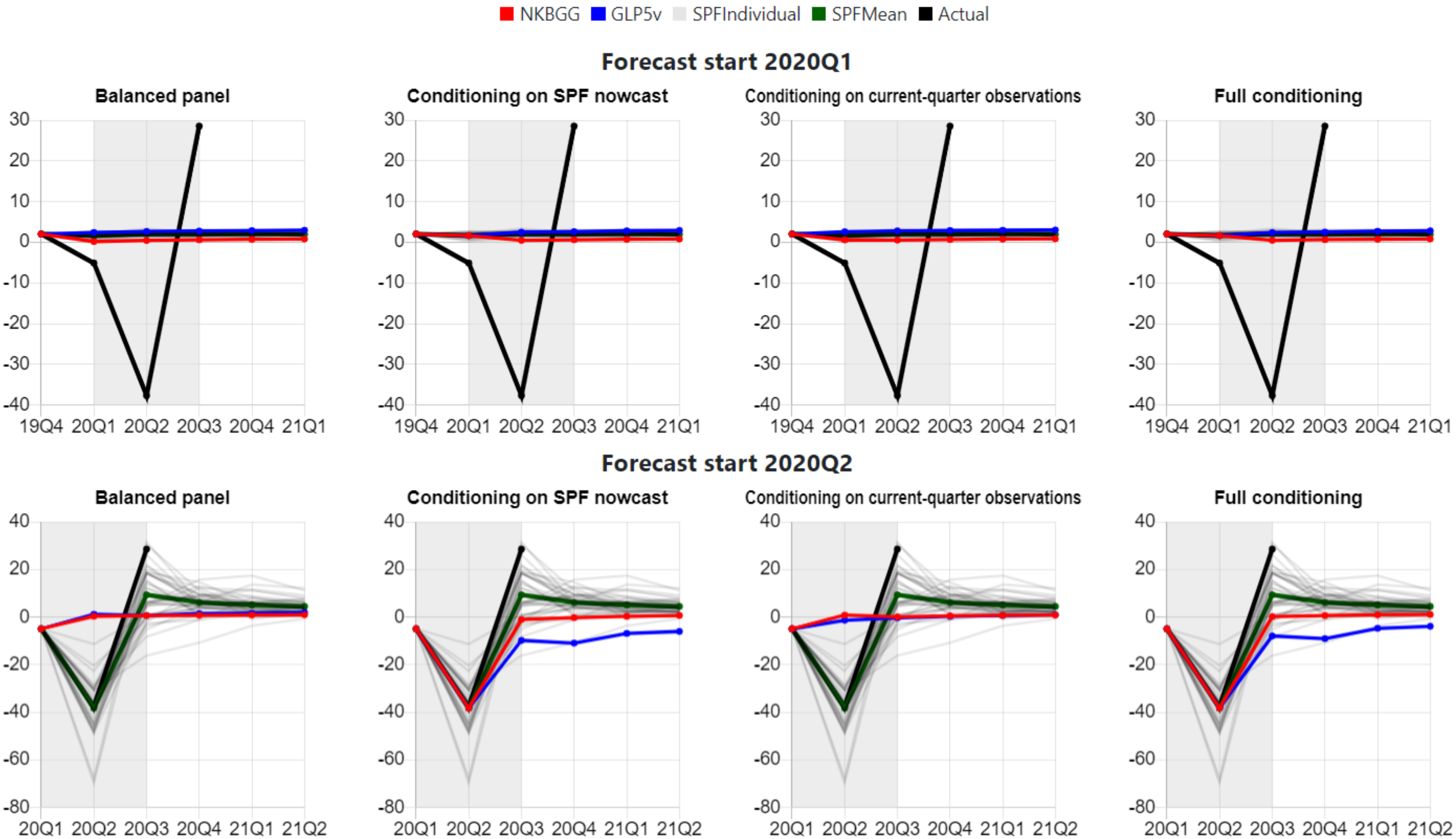


Figure 23: GDP Growth Forecasts in 2020:III: 5-Variable NK Model versus 5-Variable Bayesian VAR Model

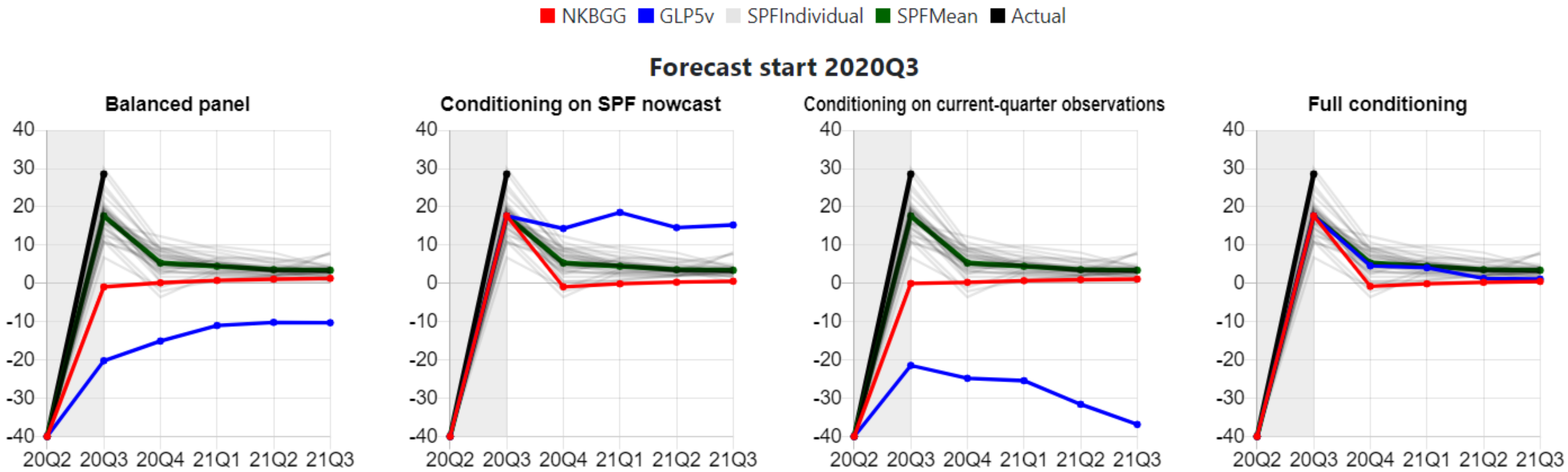


Figure 24: GDP Growth Forecasts in 2020:I–2020:II: 8-Variable NK Model versus 8-Variable Bayesian VAR Model

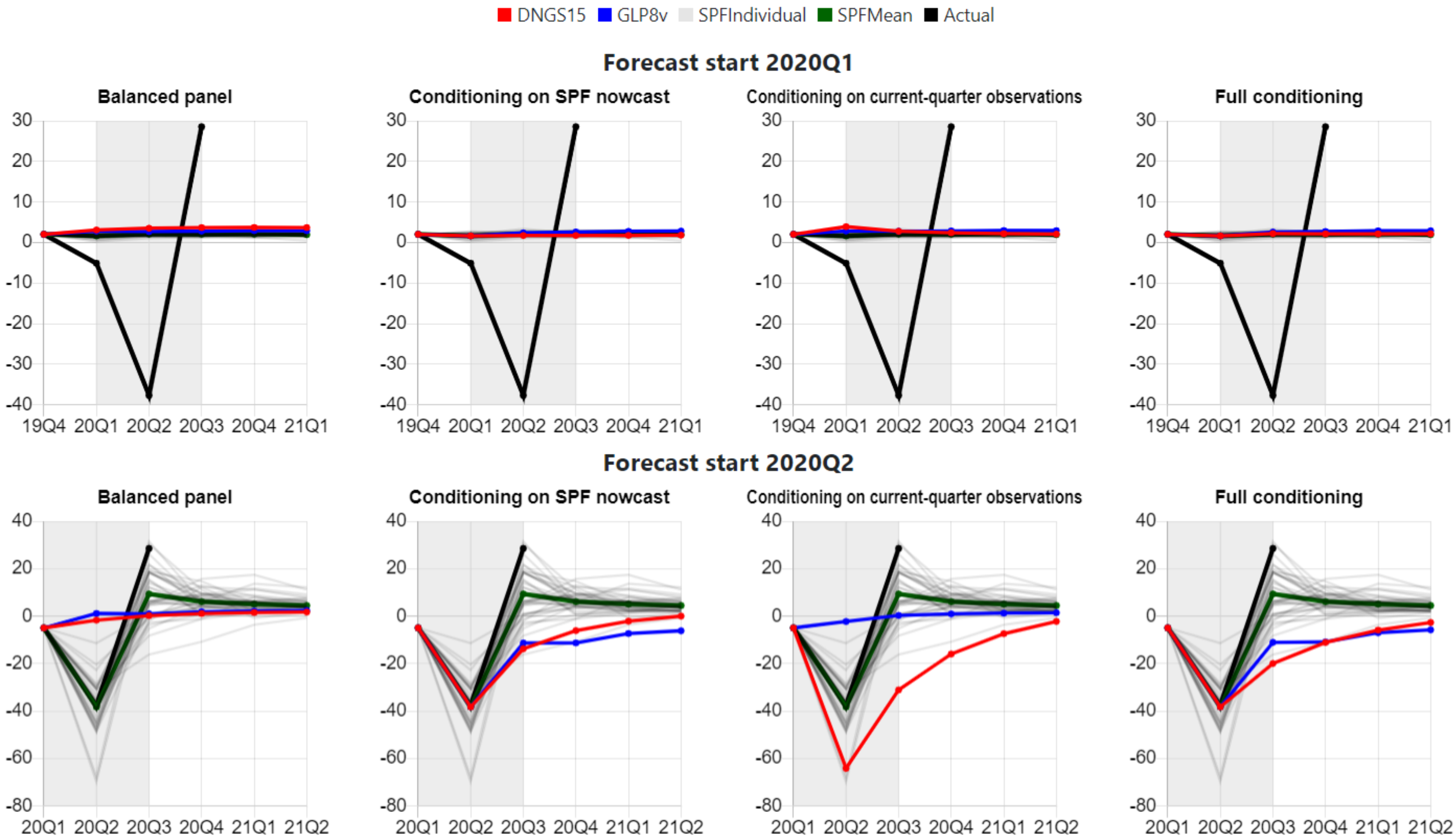
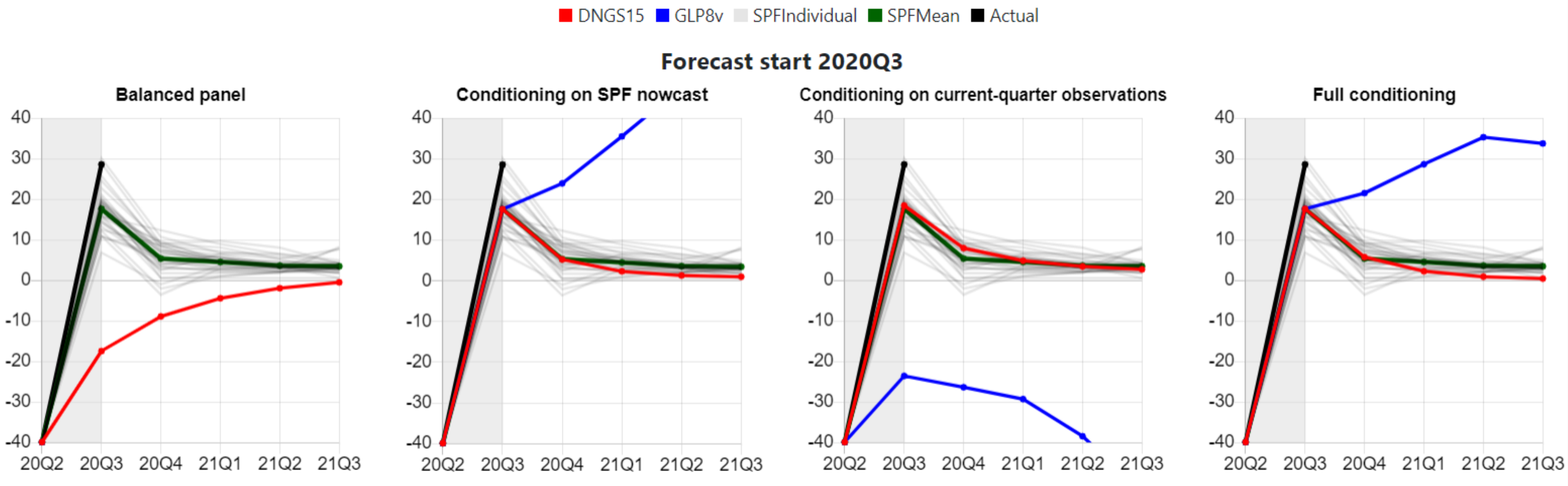


Figure 25: GDP Growth Forecasts in 2020:III: 8-Variable NK Model versus 8-Variable Bayesian VAR Model



6 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, the results regarding the 2001 and 2020 recession show that the advances in including financial frictions in macroeconomic models to not help in improving forecasting during recessions that are not caused by financial crises, but could in some cases even lead to a deteriorating forecasting performance. Hence, further efforts are needed to inform models about financial market distortions and to model financial frictions and financial shocks in a way that they are used for explaining the data during financial crisis, but are kept muted during other times.

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.

Blanchard, O. (2016). Do dsge models have a future? *Policy Brief - Peterson Institute for International Economics PB 16-11*, 1–4.

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., R. Motto, and M. Rostagno (2014). Risk Shocks. *American Economic Review 104*(1), 27–65.

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.

Fratto, C. and H. Uhlig (2020). Accounting for post-crisis inflation: A retro analysis. *Review of Economic Dynamics 35*, 133–153.

Gali, J., F. Smets, and R. Wouters (2012). Unemployment in an estimated new keynesian model. *NBER macroeconomics annual 26*(1), 329–360.

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 (2015a). Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics 97*(2), 436–451.

Giannone, D., M. Lenza, and G. Primiceri (2015b). 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.

Iacoviello, M. and S. Neri (2010). Housing market spillovers: evidence from an estimated dsge model. *American Economic Journal: Macroeconomics 2*(2), 125–64.

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 Comparison of forecasts from the original DNGS15 model and a modified DNGS15 model

DNGS15_nofa is the variant of DNGS15 model with the limited degree of nominal rigidity. There are 6 parameters associated with nominal rigidity: price/wage Calvo parameters, price/wage indexation, second derivate investment adjustment cost, and capital utilization cost. Calvo parameters and price indexations are given by 0.5 in DNGS15 model. To alleviate the degree of price adjustment in each period and the dependence to the past inflation, they are adjusted to 0.1 in DNGS15_nofa model. In addition, by decreasing the value of the second derivative of investment adjustment cost from 4 to 1, DNGS15_nofa mitigates the relationship between the value of capital and the level of investment. Finally, the capital utilization cost in DNGS15 leads to a discrepancy between the rate of utilization and the rental rate of capital. The parameter is adjusted so that the discrepancy disappears in DNGS15_nofa.

Figure 26: GDP Growth Forecasts in 2008:III–2008:IV: Original DNGS15 Model versus Modified DNGS15 Model

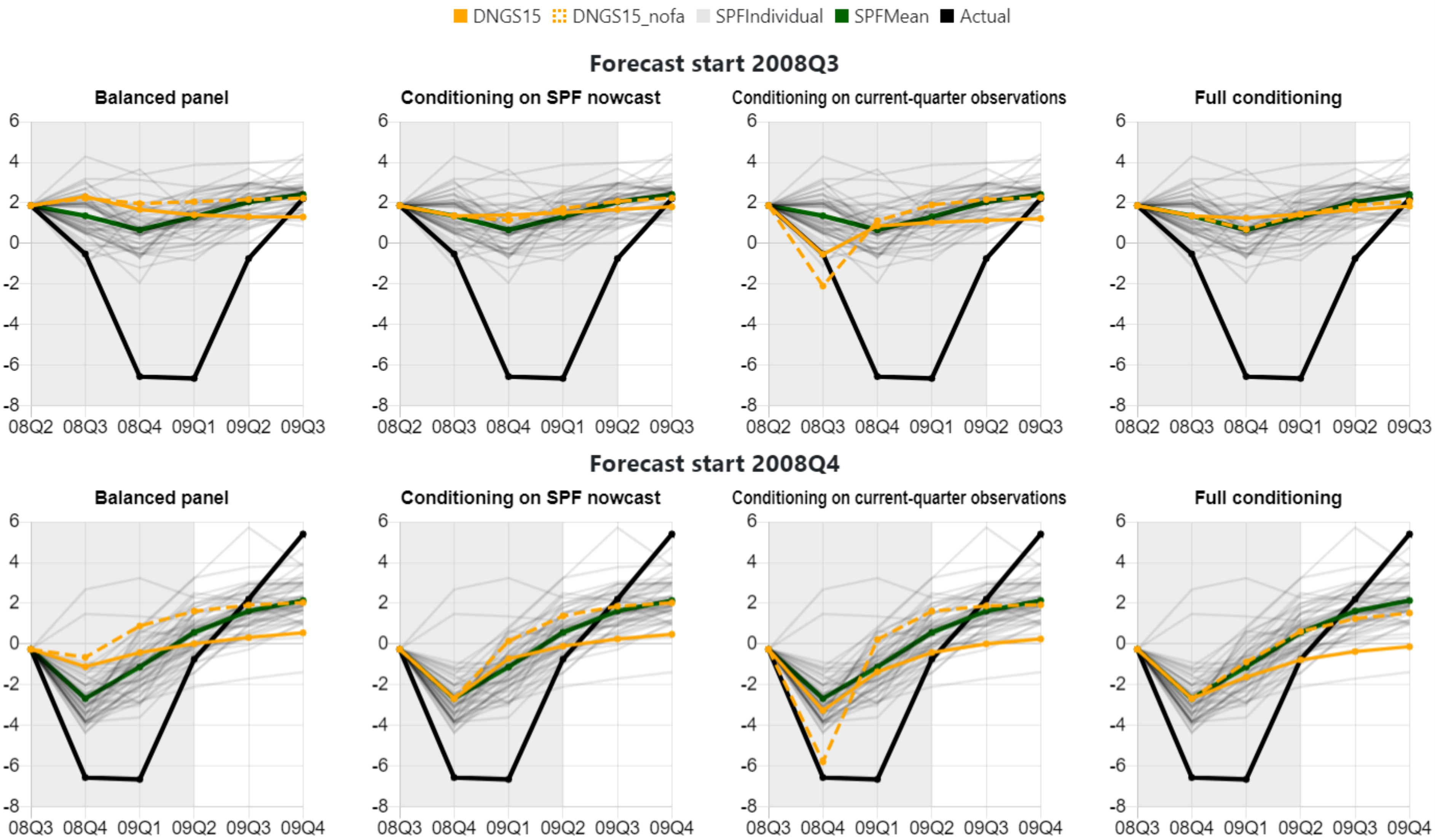


Figure 27: GDP Growth Forecasts in 2009:I–2009:II: Original DNGS15 Model versus Modified DNGS15 Model

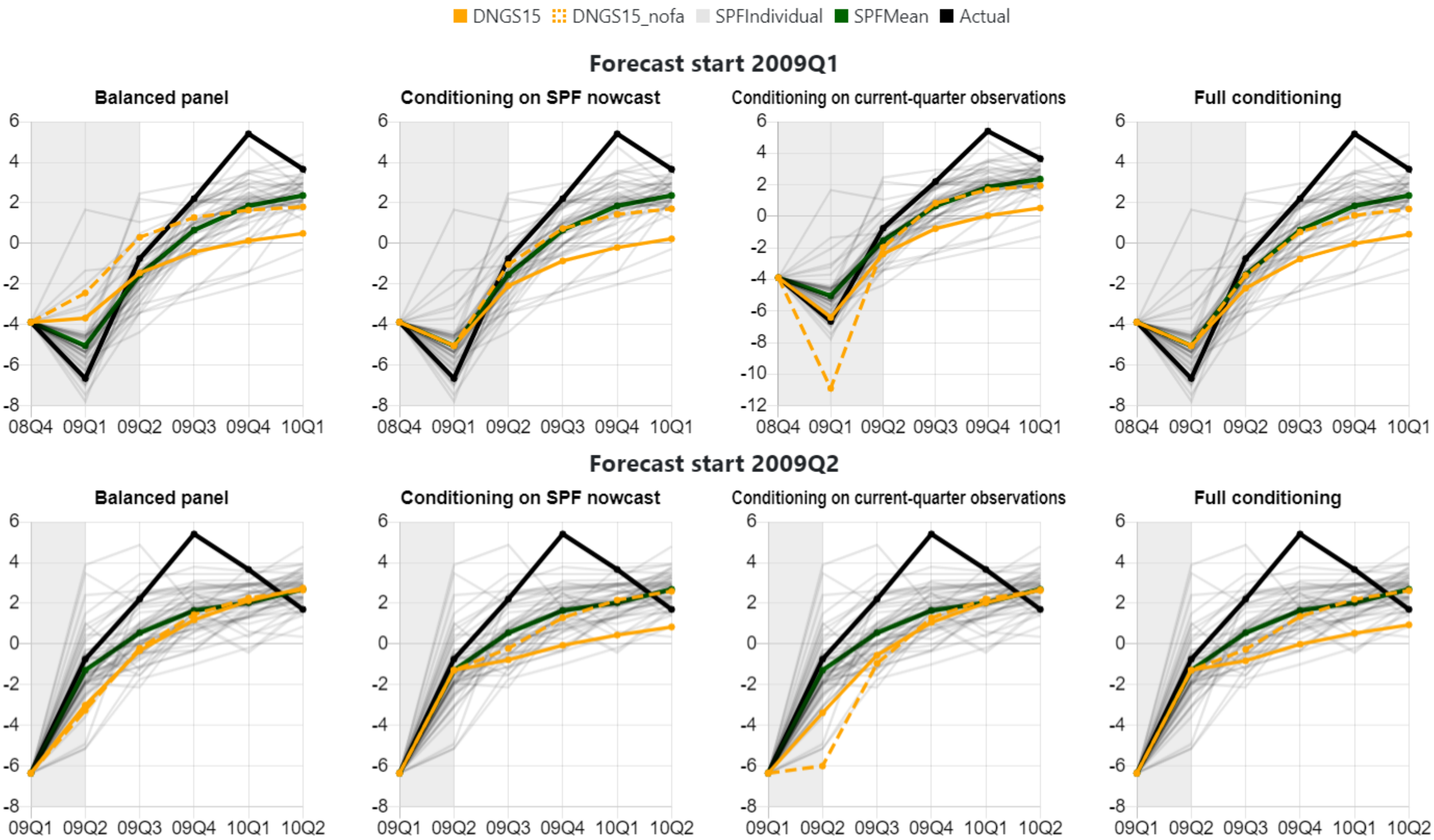


Table 1: Relative Root Mean Squared Errors (RMSE) for the GDP Growth Forecast in 2008:III–2009:II

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.

Source		DNGS15				DNGS15_nofa				SPF
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	
Horizon	0	1.55		0.91		1.50		1.76		2.32
	1	1.15	1.13	1.04	1.07	1.17	1.04	1.26	1.14	4.64
	2	1.05	1.14	1.03	1.13	1.09	1.03	1.09	1.07	4.52
	3	1.27	1.48	1.30	1.48	1.02	1.06	1.03	1.06	2.42
	4	1.63	1.66	1.69	1.76	1.09	1.22	1.08	1.10	1.83

B Comparison of forecasts from models estimated based on rolling-window and expanding-window specifications

The main difference between rolling window and expanding window specification lies in the number of data point we used to estimate the parameters. In rolling window specification, the size of each sample is fixed to 100, while in the expanding window specification the start of in-sample period is fixed to 1964Q1. The comparison of both specifications could serve as a robustness check for the forecasting predictions.

In general, the results from expanding window are more optimistic that they predict weaker crisis (except for 2008Q4) and stronger recovery path. For the cases that the forecasts start from all the quarters in year 2001 as shown in figure 28, the estimation using expanding window systematically more optimistic than using rolling window. Models of both specifications could not predict the economic downturn in 2001, but using expanding window predicts a recovery path that is closer to the actual data. The results are similar for the predictions in 2008 Q3/Q4 and 2009 Q1/Q2, except for the forecast starting from 2008Q4 conditioning on current-quarter observations, where the predicted economic downturn in expanding window specification is lower than that of in rolling window specification. The result is understandable as the parameters estimated with longer data series would tend to wash out the influences of short or medium run data. We could observe it especially in the forecasting starting from 2020Q2, expanding window specification predicts much stronger recovery than that of using rolling window. For the current crisis with full conditioning, the DNGS15 model under expanding window specification could even generate forecasts that follow the SPF mean closely. To understand which specification performs better, we look at table 2 and compare the RMSE. We can see that in both forecasting period in 2001Q1-2001Q4 and 2008Q3-2009Q2, the RMSE of rolling window specification for shorter forecasting horizon (1-2 quarters) is lower, while that of longer forecasting horizon (3-4 quarters) is higher. As a result, in general it is hard to say which specification outperforms the other, but we can conclude that rolling window specification is better in predicting a crisis and the expanding window specification is better in predicting the recovery path.

Figure 28: GDP Growth Forecasts in 2001:I–2001:II: Rolling-Window and Expanding-Window Specifications

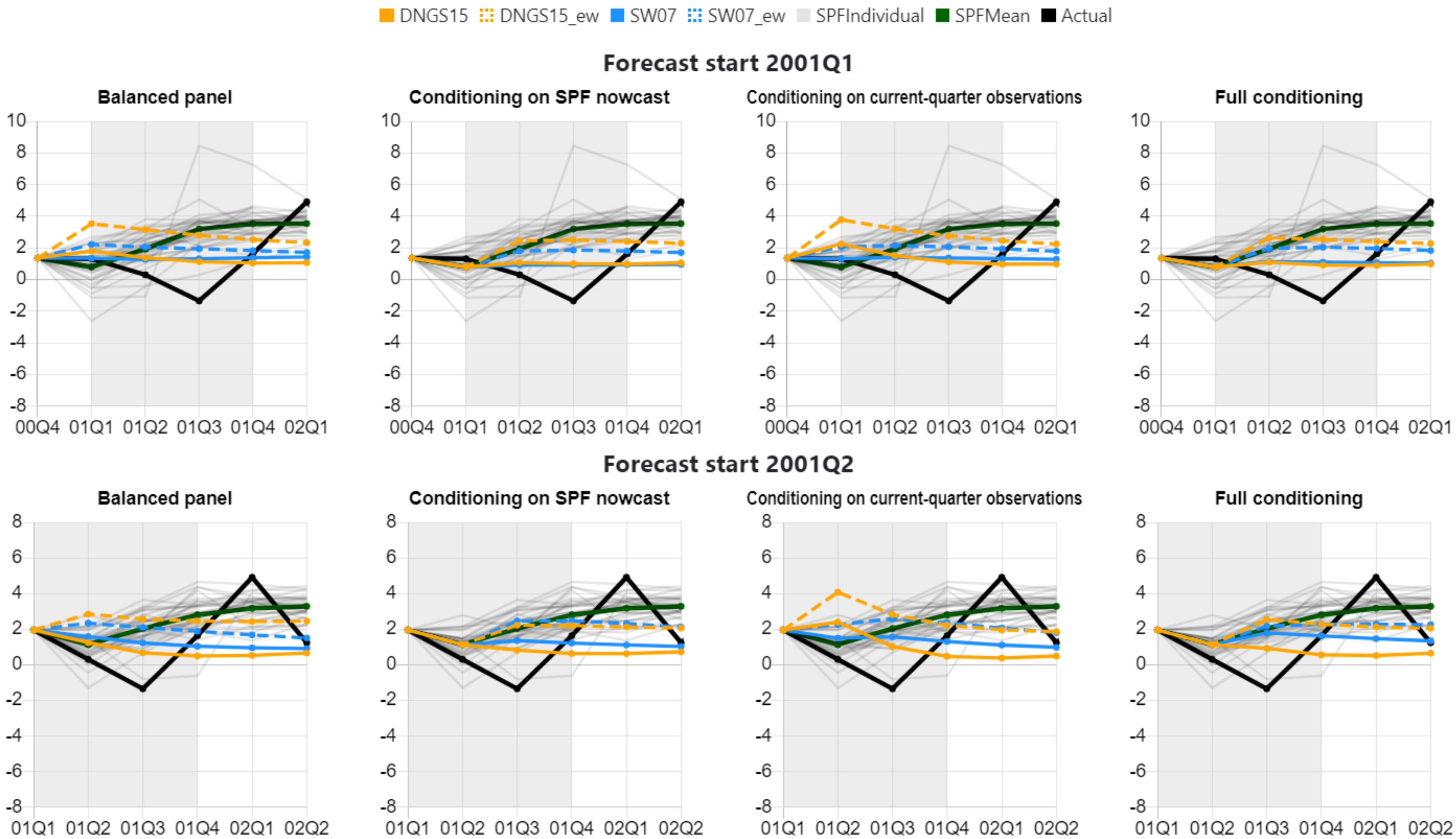


Figure 29: GDP Growth Forecasts in 2001:III–2001:IV: Rolling-Window and Expanding-Window Specifications

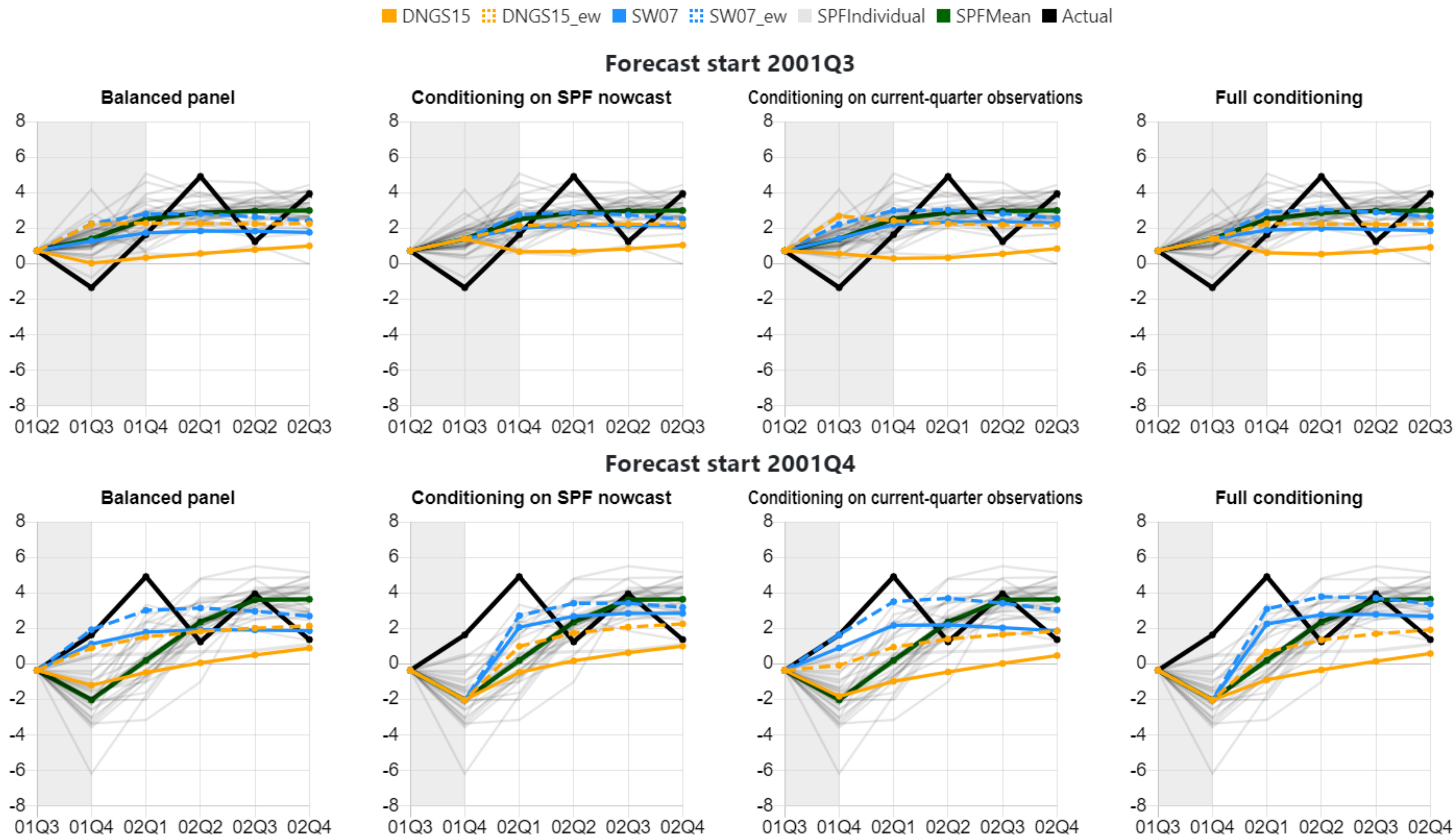


Figure 30: GDP Growth Forecasts in 2008:III–2008:IV: Rolling-Window and Expanding-Window Specifications

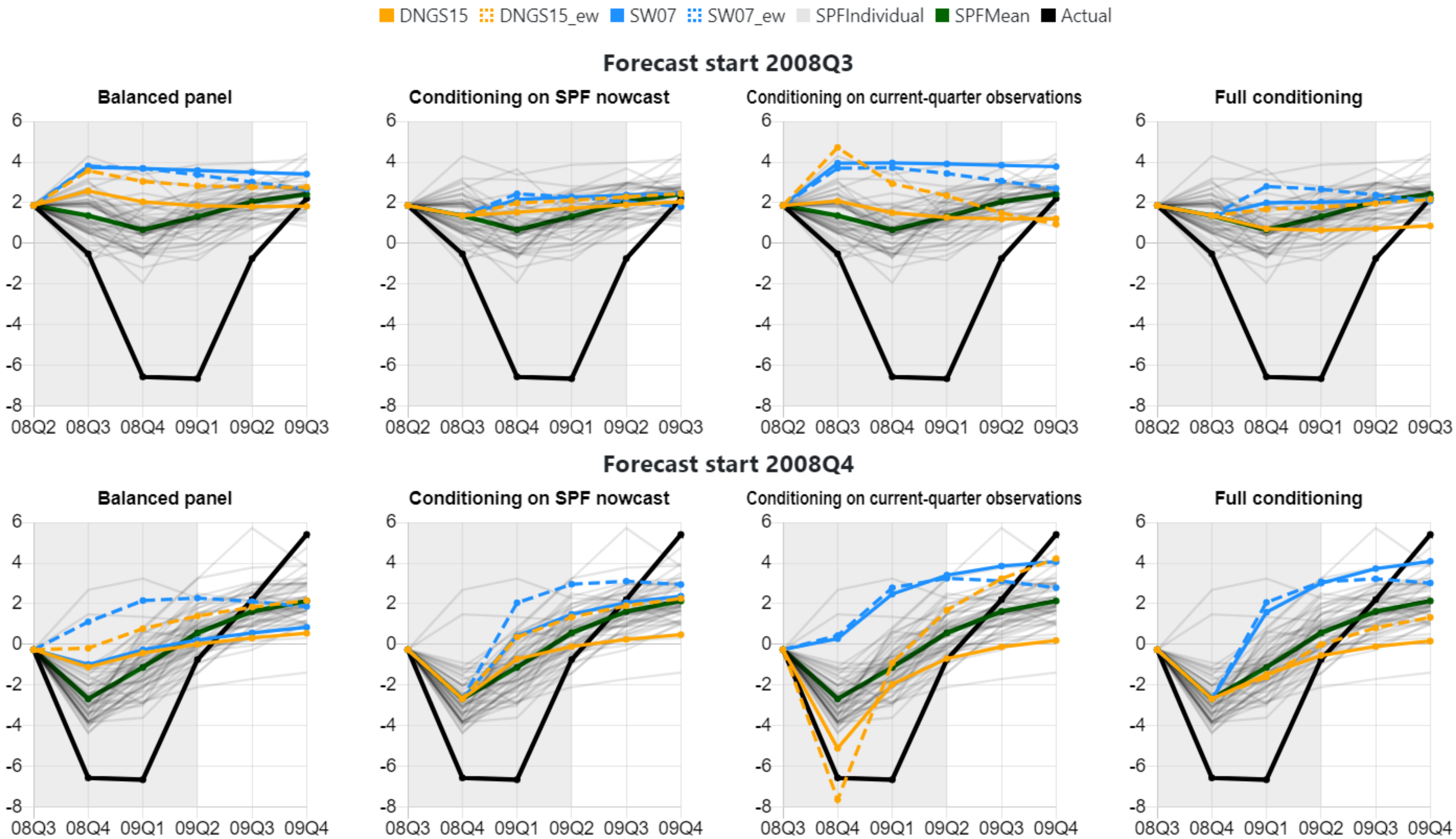


Figure 31: GDP Growth Forecasts in 2009:I–2009:II: Rolling-Window and Expanding-Window Specifications

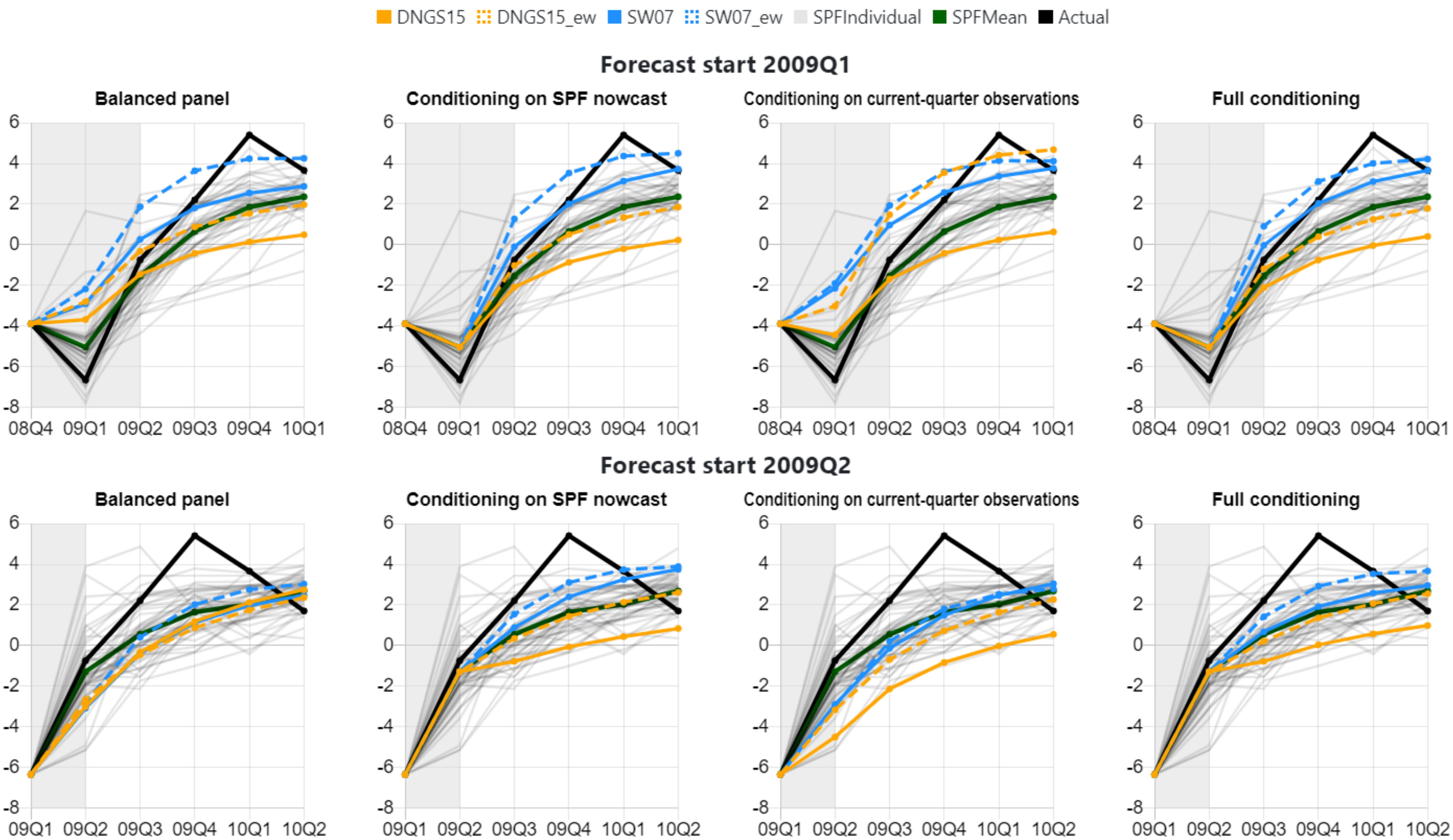


Figure 32: GDP Growth Forecasts in 2020:I–2020:II: Rolling-Window and Expanding-Window Specifications

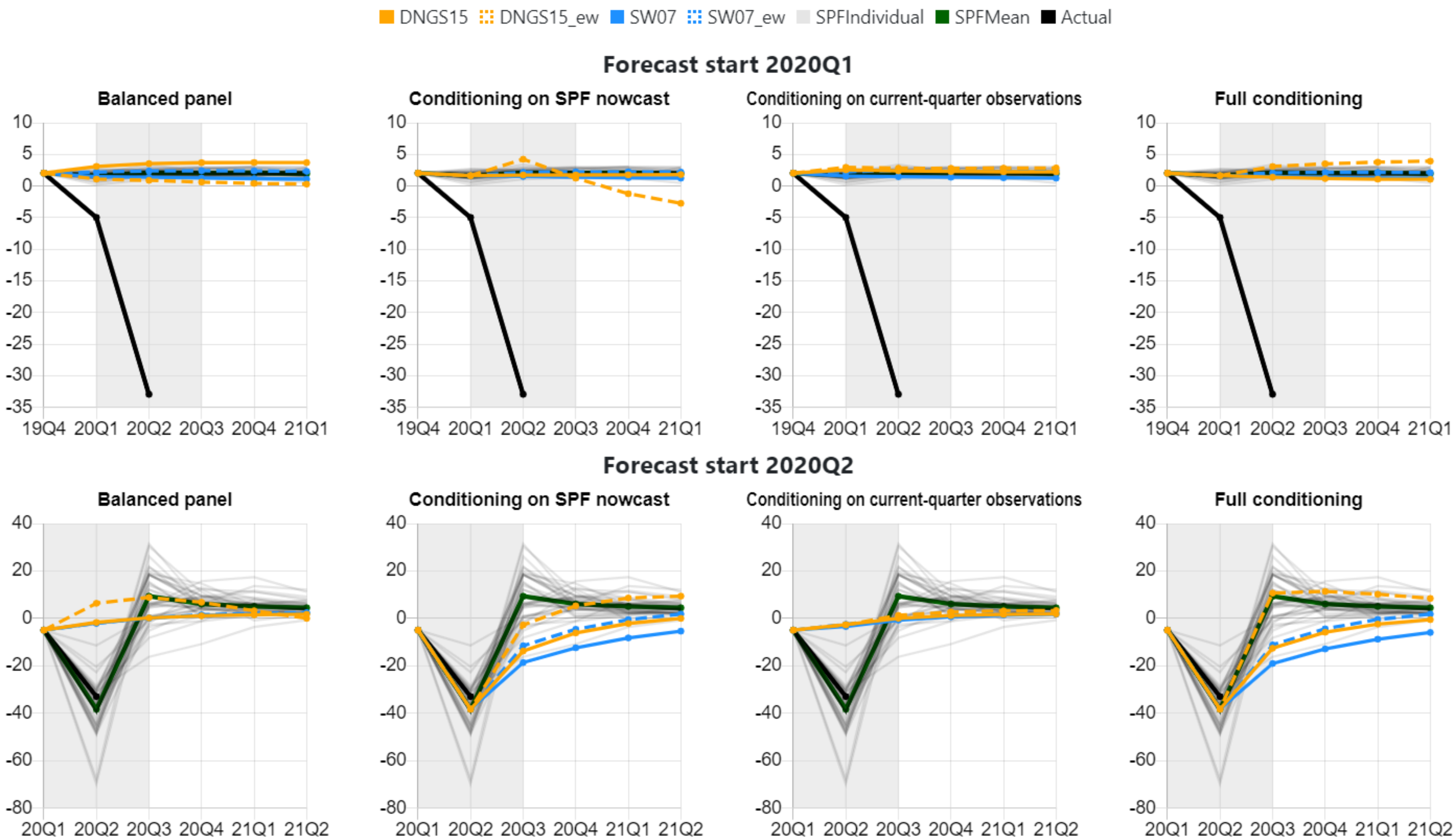


Table 2: Relative Root Mean Squared Errors (RMSE) for the GDP Growth Forecast in 2001:I–2001:IV and 2008:III-2009:II

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.

Forecast period: 2001:I - 2001:IV

Source		DNGS15				DNGS15_ew				SW07				SW07_ew				SPFM
Scenario		NoExt	SPF	CQ	FV	NoExt	SPF	CQ	FV	NoExt	SPF	CQ	FV	NoExt	SPF	CQ	FV	
	0	0.71		0.98		1.06		1.34		0.64		0.66		0.91		0.88		2.34
	1	0.99	0.97	1.08	1.05	0.98	0.94	1.07	1.03	0.68	0.66	0.69	0.69	0.73	0.78	0.78	0.74	3.05
Forecast Horizon	2	1.00	0.96	1.07	1.01	0.96	0.91	0.94	0.91	0.79	0.74	0.74	0.78	0.83	0.85	0.89	0.89	2.62
	3	1.81	1.76	1.95	1.89	1.10	1.15	1.27	1.22	1.45	1.33	1.42	1.20	1.18	0.98	1.07	1.01	1.55
	4	1.41	1.40	1.48	1.45	0.98	0.96	0.94	0.95	1.19	1.32	1.15	1.32	1.09	1.16	1.10	1.16	1.74

Forecast period: 2008:III - 2009:IV

Source		DNGS15				DNGS15_ew				SW07				SW07_ew				SPFM
Scenario		NoExt	SPF	CQ	FV	NoExt	SPF	CQ	FV	NoExt	SPF	CQ	FV	NoExt	SPF	CQ	FV	
	0	1.57		1.14		1.88		1.49		1.79		2.07		2.18		2.09		2.32
	1	1.18	1.14	1.11	1.03	1.34	1.21	1.26	1.07	1.34	1.22	1.53	1.29	1.50	1.37	1.55	1.39	4.64
Forecast Horizon	2	1.09	1.16	1.15	1.05	1.20	1.11	1.16	1.06	1.23	1.07	1.33	1.12	1.23	1.11	1.27	1.15	4.52
	3	1.31	1.50	1.45	1.41	1.15	1.10	0.69	1.11	1.16	0.80	1.12	0.85	0.84	0.64	0.88	0.74	2.42
	4	1.61	1.66	1.70	1.74	1.03	1.03	0.57	1.25	1.33	1.00	0.67	0.50	1.05	0.93	0.80	0.86	1.83

C Bayesian VAR Model Description

C.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, ϕ_0; ϕ_1; ϕ_2 are hyperparameters and $\frac{\sigma_j}{\sigma_i}$ is a scaling factor. The Minnesota prior's hyperparameters are the following:

1. ϕ_0 represents the tightness on the variance of the first lag,
2. ϕ_1 is the parameter governing the relative tightness on other variables,
3. ϕ_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 ϕ_3 is positive.

For the standard errors σ_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.)

C.2 GLP prior

Giannone et al. (2015b) adopt a hierarchical model to "make inference about the informativeness of the prior distribution" (Giannone et al., 2015b, 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., 2015b, 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 λ, μ, δ, and ψ which they treat as additional parameters. The equivalent notation for these parameters is the following:

- From the Minnesota prior they focus on the hyperparameter ϕ_0, denoted by λ. 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 (ϕ_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:

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

  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., 2015b, p.440.) The hyperparameter μ 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., 2015b, p.440.)

- The "dummy-initial-observation" prior is implemented by following using the dummy observation:

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

  This prior implies that a martingale property for the asymptotic of the sample. The hyperparameter δ 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., 2015b, p.440.)

The GLP hyperpriors for λ, μ, δ 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 Σ, 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.[7]

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., 2015b, p.442.)

C.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 [8] 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 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

[7] 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.

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

$(\pi_1^*, \pi_2^*, ...\pi_N^*)$ 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:

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

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 $g : \mathscr{R}^p \to \mathscr{R}^{\mathrm{m}}$, 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.

D RMSEs relative to SPF and average RMSE of all macroeconomic models

Table 3: RMSE relative to SPF, 2008:III–2009:II

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. Numbers that are lower than one are displayed in green and numbers that are higher than one are displayed in red. The (absolute) RMSEs of the SPF Mean are shown in the last column.

Source		SW07				DS04				WW11				FU20				GSW12				FRBEDO08				Fair
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	
	0	1.76		2.06		2.10		2.72		2.37		2.88		2.38		2.15		1.88		1.69		2.50		1.88		1.91
	1	1.30	1.22	1.53	1.36	1.31	1.23	1.27	1.12	1.37	1.30	1.34	1.23	1.55	1.39	1.59	1.46	1.31	1.15	1.33	1.28	1.45	1.31	1.41	1.14	1.32
Horizon	2	1.19	1.08	1.32	1.22	1.16	1.12	1.15	1.12	1.15	1.14	1.15	1.14	1.34	1.23	1.37	1.28	1.06	0.94	1.04	1.04	1.24	1.21	1.35	1.04	1.24
	3	1.08	0.82	1.11	1.04	1.11	1.13	1.14	1.22	1.05	1.09	1.08	1.14	1.11	1.06	1.12	1.16	0.73	0.55	0.77	0.63	0.89	0.94	1.40	1.06	1.21
	4	1.30	1.00	0.66	0.59	0.98	1.02	1.06	1.17	0.89	0.93	0.96	1.03	0.83	0.88	0.73	0.74	1.31	1.33	0.95	1.06	1.02	0.89	1.28	1.11	0.99

Source		NKBGG				QPM08				DNGS15				CMR14				KR15_FF				KR15_HH			
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC
	0	2.45		2.45		2.57		2.39		1.55		0.91		2.04		1.42		1.98		2.89		1.98		2.62	
	1	1.46	1.35	1.39	1.27	1.34	1.30	1.30	1.30	1.15	1.13	1.04	1.07	1.41	1.25	1.24	1.18	1.55	1.13	1.56	1.43	1.49	1.13	1.63	1.13
Horizon	2	1.19	1.16	1.17	1.15	1.06	1.08	1.05	1.09	1.05	1.14	1.03	1.13	1.38	1.25	1.33	1.23	1.21	1.05	1.21	1.09	1.32	1.07	1.52	1.26
	3	0.92	0.93	0.93	0.94	0.86	0.86	0.80	0.86	1.27	1.48	1.30	1.48	1.39	1.33	1.41	1.38	0.87	0.85	0.91	0.80	1.04	1.23	1.46	1.28
	4	0.88	0.86	0.87	0.85	1.26	1.14	1.26	1.15	1.63	1.66	1.69	1.76	0.94	0.93	1.11	1.13	0.83	0.87	0.87	0.86	1.38	2.11	1.49	1.81

Source		IN10				VI16_BGG				VI16_GK				3vBVAR				5vBVAR				8vBVAR				SPF
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	Mean
	0	1.49		1.96		2.19		2.09		2.20		2.09		2.58		2.61		2.49		2.26		2.03		1.79		2.32
	1	0.99		0.99		1.47	1.11	1.48	1.13	1.42	1.06	1.45	1.06	1.38	1.26	1.37	1.26	1.44	1.28	1.36	1.26	1.44	1.35	1.41	1.32	4.64
Horizon	2	0.92		0.90		1.60	1.26	1.57	1.26	1.54	1.14	1.54	1.13	1.28	1.19	1.23	1.16	1.39	1.28	1.34	1.24	1.39	1.32	1.39	1.28	4.52
	3	0.94		0.99		2.50	2.13	2.44	2.09	2.28	1.83	2.25	1.80	1.47	1.42	1.39	1.33	1.83	1.73	1.78	1.59	1.80	1.66	1.85	1.56	2.42
	4	1.11		1.20		2.86	2.60	2.82	2.60	2.60	2.26	2.55	2.30	1.22	1.51	1.17	1.46	1.78	1.86	1.80	1.74	1.67	1.62	1.75	1.53	1.83

Table 4: RMSE relative to average RMSE of all macroeconomic models, 2008:III–2009:II

Note: The table shows the RMSEs relative to average the RMSEs of all the macroeconomic models for the GDP growth forecasts on five horizons in four scenarios. Numbers that are lower than one are displayed in green and numbers that are higher than one are displayed in red. The (absolute) average RMSEs of all the macroeconomic models are shown in the last column.

Source		SW07				DS04				WW11				FU20				GSW12				FRBEDO08				Fair
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	
	0	0.83		0.98		0.99		1.29		1.12		1.36		1.13		1.02		0.89		0.80		1.18		0.89		0.90
	1	1.00	0.94	1.18	1.05	1.01	0.94	0.98	0.86	1.05	1.00	1.03	0.94	1.20	1.07	1.22	1.12	1.01	0.88	1.02	0.99	1.12	1.01	1.08	0.88	1.01
Horizon	2	0.99	0.90	1.11	1.02	0.97	0.94	0.97	0.94	0.97	0.96	0.96	0.95	1.12	1.03	1.15	1.07	0.89	0.79	0.87	0.87	1.04	1.01	1.13	0.87	1.04
	3	0.90	0.67	0.92	0.86	0.92	0.94	0.94	1.01	0.87	0.90	0.89	0.94	0.91	0.87	0.92	0.95	0.60	0.45	0.64	0.52	0.73	0.78	1.15	0.87	1.00
	4	0.99	0.77	0.51	0.46	0.75	0.79	0.81	0.90	0.68	0.71	0.74	0.79	0.64	0.68	0.56	0.57	1.01	1.02	0.73	0.81	0.78	0.68	0.98	0.85	0.76

Source		NKBGG				QPM08				DNGS15				CMR14				KR15_FF				KR15_HH			
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC
	0	1.16		1.16		1.21		1.13		0.73		0.43		0.96		0.67		0.94		1.37		0.94		1.24	
	1	1.13	1.04	1.07	0.98	1.03	1.00	1.00	1.00	0.88	0.87	0.80	0.82	1.09	0.97	0.96	0.91	1.19	0.87	1.20	1.10	1.14	0.87	1.25	0.87
Horizon	2	0.99	0.97	0.98	0.97	0.89	0.90	0.88	0.92	0.88	0.96	0.86	0.94	1.16	1.05	1.11	1.03	1.02	0.88	1.02	0.92	1.11	0.90	1.28	1.06
	3	0.76	0.77	0.76	0.77	0.71	0.71	0.66	0.71	1.05	1.22	1.08	1.22	1.14	1.10	1.16	1.14	0.72	0.70	0.75	0.66	0.86	1.02	1.21	1.05
	4	0.67	0.66	0.67	0.65	0.96	0.87	0.97	0.88	1.25	1.27	1.30	1.35	0.72	0.71	0.85	0.87	0.63	0.67	0.67	0.66	1.06	1.62	1.14	1.39

Source		IN10				VI16_BGG				VI16_GK				3vBVAR				5vBVAR				8vBVAR				Models
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	Mean
	0	0.70		0.93		1.04		0.99		1.04		0.99		1.22		1.24		1.18		1.07		0.96		0.85		4.90
	1	0.76		0.76		1.13	0.85	1.14	0.87	1.09	0.82	1.12	0.82	1.06	0.97	1.05	0.97	1.11	0.98	1.05	0.97	1.11	1.04	1.09	1.01	6.02
Horizon	2	0.77		0.76		1.34	1.06	1.32	1.05	1.29	0.96	1.29	0.95	1.08	0.99	1.03	0.97	1.17	1.07	1.12	1.04	1.17	1.10	1.16	1.07	5.39
	3	0.78		0.82		2.06	1.75	2.01	1.73	1.88	1.51	1.86	1.49	1.22	1.17	1.15	1.10	1.51	1.43	1.47	1.31	1.48	1.37	1.52	1.29	2.93
	4	0.85		0.92		2.19	1.99	2.16	2.00	1.99	1.73	1.95	1.76	0.93	1.16	0.90	1.12	1.36	1.43	1.38	1.33	1.28	1.24	1.35	1.17	2.38

Table 5: Relative Root Mean Squared Errors (RMSE) for the GDP Growth Forecast in 2008:III–2009:II

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. Numbers that are lower than one are displayed in green and numbers that are higher than one are displayed in red. The (absolute) RMSEs of the SPF Mean are shown in the last column.

Source		SW07				DS04				WW11				FU20				GSW12				FRBEDO08				Fair
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	
	0	0.64		0.66		0.72		0.91		0.86		1.00		0.56		0.59		0.76		1.36		0.64		0.95		1.04
	1	0.68	0.66	0.69	0.69	0.92	0.94	0.95	1.01	0.93	0.95	0.94	0.98	0.63	0.70	0.70	0.70	0.97	0.89	0.72	0.77	0.83	0.97	0.92	0.96	0.99
Horizon	2	0.79	0.74	0.74	0.78	0.97	0.91	0.96	0.91	0.97	0.93	0.97	0.93	0.78	0.74	0.76	0.74	0.85	0.88	0.81	0.84	0.97	0.89	0.94	0.91	0.78
	3	1.45	1.33	1.42	1.20	1.07	1.11	1.14	1.22	1.04	1.07	1.08	1.15	1.41	1.32	1.32	1.22	1.40	1.04	1.49	1.35	1.06	1.20	1.12	1.12	1.43
	4	1.19	1.32	1.15	1.32	0.90	0.89	0.89	0.89	0.89	0.89	0.88	0.87	1.32	1.36	1.28	1.32	1.70	1.66	1.67	1.64	0.91	0.89	0.89	0.91	1.16

Source		NKBGG				QPM08				DNGS15				CMR14				KR15_FF				KR15_HH			
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC
	0	0.53		0.66		0.67		0.87		0.71		1.11		0.50		0.66		0.71		1.38		0.71		0.76	
	1	0.76	0.82	0.80	0.86	0.74	0.67	0.68	0.72	0.99	0.97	1.07	1.05	0.85	0.95	0.91	1.00	0.99	1.21	1.10	1.21	1.00	1.21	0.71	1.21
Horizon	2	0.81	0.82	0.79	0.80	0.88	0.95	0.86	0.86	1.00	0.96	1.07	1.02	0.99	0.89	0.93	0.93	0.91	0.83	0.98	0.86	0.88	0.67	0.99	0.93
	3	1.23	1.21	1.28	1.25	0.91	1.02	0.88	0.80	1.81	1.76	1.97	1.92	1.57	1.63	1.68	1.71	1.19	1.04	1.11	1.07	1.12	1.24	0.83	1.03
	4	1.17	1.17	1.17	1.17	1.03	1.10	1.07	1.11	1.41	1.40	1.46	1.43	1.18	1.23	1.20	1.18	1.07	1.08	1.12	1.10	1.41	1.28	0.95	0.92

Source		IN10				VI16_BGG				VI16_GK				3vBVAR				5vBVAR				8vBVAR				SPF
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	Mean
	0	0.80		1.42		1.28		1.36		1.41		1.46		1.43		1.16		1.18		0.85		0.81		0.75		2.34
	1	1.29		1.30		1.60	1.27	1.62	1.27	1.67	1.31	1.68	1.33	1.12	1.08	1.19	1.14	1.16	1.14	1.19	1.17	0.95	0.94	1.03	1.05	3.05
Horizon	2	1.41		1.40		1.84	1.43	1.87	1.45	1.91	1.46	1.92	1.50	1.31	1.27	1.42	1.31	1.36	1.27	1.45	1.33	1.18	1.20	1.16	1.06	2.62
	3	2.76		2.58		3.58	3.14	3.67	3.20	3.68	3.16	3.72	3.28	1.38	1.32	1.50	1.39	1.35	1.33	1.48	1.42	1.19	1.29	1.18	1.23	1.55
	4	2.40		2.16		3.18	2.88	3.23	2.93	3.21	2.90	3.23	2.96	1.34	1.41	1.61	1.55	1.31	1.35	1.55	1.60	1.06	1.36	1.23	1.17	1.74

Table 6: Relative Root Mean Squared Errors (RMSE) for the GDP Growth Forecast in 2001:I–2001:IV

Note: The table shows the RMSEs relative to average the RMSEs of all the macroeconomic models for the GDP growth forecasts on five horizons in four scenarios. Numbers that are lower than one are displayed in green and numbers that are higher than one are displayed in red. The (absolute) average RMSEs of all the macroeconomic models are shown in the last column.

Source		SW07				DS04				WW11				FU20				GSW12				FRBEDO08				Fair
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	
	0	0.71		0.74		0.81		1.02		0.96		1.12		0.63		0.66		0.85		1.52		0.71		1.06		1.16
	1	0.70	0.67	0.70	0.70	0.94	0.96	0.97	1.03	0.95	0.96	0.95	1.00	0.64	0.71	0.72	0.72	0.99	0.91	0.74	0.79	0.85	0.99	0.94	0.97	1.01
Horizon	2	0.79	0.74	0.73	0.78	0.96	0.91	0.96	0.91	0.97	0.93	0.97	0.92	0.78	0.74	0.75	0.73	0.84	0.88	0.80	0.83	0.97	0.89	0.94	0.91	0.78
	3	0.90	0.83	0.88	0.75	0.66	0.69	0.71	0.76	0.64	0.67	0.67	0.72	0.88	0.82	0.82	0.76	0.87	0.65	0.93	0.84	0.66	0.75	0.70	0.70	0.89
	4	0.81	0.90	0.79	0.90	0.61	0.61	0.61	0.61	0.61	0.61	0.60	0.59	0.90	0.93	0.88	0.90	1.16	1.14	1.14	1.12	0.62	0.61	0.61	0.62	0.79

Source		NKBGG				QPM08				DNGS15				CMR14				KR15_FF				KR15_HH			
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC
	0	0.59		0.73		0.75		0.98		0.80		1.25		0.56		0.74		0.79		1.55		0.79		0.84	
	1	0.77	0.83	0.81	0.88	0.75	0.69	0.69	0.73	1.00	0.99	1.09	1.07	0.86	0.96	0.92	1.02	1.01	1.24	1.12	1.24	1.02	1.24	0.72	1.24
Horizon	2	0.81	0.82	0.79	0.79	0.87	0.95	0.85	0.86	1.00	0.96	1.06	1.01	0.98	0.88	0.93	0.92	0.91	0.82	0.98	0.86	0.88	0.67	0.98	0.92
	3	0.77	0.75	0.80	0.78	0.57	0.63	0.55	0.50	1.13	1.10	1.23	1.20	0.98	1.02	1.04	1.06	0.74	0.65	0.69	0.67	0.70	0.77	0.52	0.64
	4	0.80	0.80	0.80	0.80	0.71	0.75	0.73	0.76	0.96	0.96	1.00	0.98	0.81	0.84	0.82	0.81	0.73	0.74	0.76	0.75	0.96	0.88	0.65	0.63

Source		IN10				VI16_BGG				VI16_GK				3vBVAR				5vBVAR				8vBVAR				Models
Scenario		BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	BP	SPF	CQ	FC	Mean
	0	0.90		1.59		1.43		1.53		1.57		1.64		1.60		1.30		1.32		0.95		0.91		0.84		2.09
	1	1.32		1.32		1.63	1.30	1.65	1.30	1.70	1.34	1.71	1.36	1.14	1.11	1.21	1.16	1.19	1.16	1.21	1.19	0.97	0.95	1.05	1.07	2.99
Horizon	2	1.41		1.39		1.83	1.42	1.86	1.44	1.90	1.45	1.91	1.49	1.31	1.27	1.41	1.30	1.35	1.26	1.45	1.32	1.17	1.19	1.15	1.06	2.64
	3	1.72		1.60		2.23	1.96	2.28	1.99	2.29	1.97	2.31	2.04	0.86	0.82	0.93	0.87	0.84	0.83	0.92	0.88	0.74	0.80	0.73	0.77	2.50
	4	1.64		1.48		2.17	1.97	2.21	2.00	2.19	1.98	2.21	2.02	0.91	0.96	1.10	1.06	0.90	0.92	1.06	1.09	0.73	0.93	0.84	0.80	2.55

E DSGE Model Descriptions

E.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[9].

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.

E.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.

[9] 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.

E.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-modelled 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).

E.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.

E.5 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 (E.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).

E.6 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.

E.7 The FU20 Model

Fratto and Uhlig investigate on the missing deflation puzzle by estimating versions of the Smets and Wouters (2007) on different samples of US data that include or exclude the years after the Financial Crisis. They find that markup shocks account for the almost all of the variation in inflation before and after the crisis. In the MMB, we parametrize the model according to the estimates on 1984-2015 data.

- Aggregate Demand: Households maximize their lifetime utility, where the utility function is nonseparable in consumption and leisure, subject to an intertemporal 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 autoregressive 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 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: A total factor productivity shock, a risk premium shock, an investment-specific technology shock, a wage and a price mark-up shock and two policy shocks: the common fiscal policy shock entering the government spending equation and the common monetary policy shock.

- Estimation: The model is estimated for the U.S. with Bayesian techniques for the period 1984Q1-2015Q4 using seven key macroeconomic variables: real GDP, consumption, investment, the GDP deflator, real wages, employment and the nominal short-term interest rate. The replication package additionally contains the baseline version of the model estimated on a shorter sample (1984Q1-2007Q4).

E.8 The GSW12 Model

Gali et al. (2012) is based on Smets and Wouters (2007) and differs from the latter in the following ways:

- Labor decision on the extensive margin (whether to work or not) rather than the intensive margin (how many hours to work), unemployment is included as an observable variable

- Logarithmic consumption utility, the utility function is separable in consumption and leisure

- the error term in the wage equation captures only the wage markup shock and not the preference shock (as in SW07)

- Dixit-Stigliz type aggregator functions for aggregate labor demand and aggregate nominal wage (SW07 uses Kimball).

- Shocks: A total factor productivity shock, a risk premium shock, an investment-specific technology shock, a labor supply shock, a wage and a price mark-up shock and two policy shocks: the common fiscal policy shock entering the government spending equation and the common monetary policy shock.

- Estimation: The model is estimated for the euro area with Bayesian techniques for the period 1985:1-2009:4. In addition to SW07, unemployment is used as an observable variable for the estimation of parameters.

E.9 The IN10 Model

The IN10 model is based on various dynamic equilibrium models with neoclassical core and real and nominal rigidities (e.g. Smets and Wouters (2007) model). The main goal of this model is to explain the development of the price and the quantity side of the housing market and to examine the spillovers from the housing market to the rest of the economy. The model features a multi-sector structure with housing and non-housing goods, financial frictions in the household sector (introduced through a collateral constraint imposed on a fraction of households), and rich set of shocks which take the model to the data.

- Aggregate demand: There are two types of households according to their discount factors: patient (lenders) and impatient (borrowers). A representative household within each group obtain utility from consumption and housing and disutility from supplying labor in an additively-separable way. Habit formation and balanced growth in consumption are considered. Imperfect labor mobility is introduced across sectors. Patient households accumulate housing and capital, make loans to impatient households, rent capital and land to firms, and choose the capital utilization rate. Impatient households work, consume, accumulate housing and borrow against the value of their housing. Impatient households accumulate housing and borrow the maximum possible amount against its collateral value in equilibrium.

- Aggregate Supply: Wholesale firms consists of two production units. Housing sector produces new houses (using capital, labour, land and intermediate goods). Non-housing sector produces consumption goods, investment goods and intermediate goods (using capital and labour). It is allowed for price rigidities in the consumption sector and for wage rigidities in both the consumption and housing sectors, but there are no price rigidities in the housing market.

- Shocks: An intertemporal preference shock, a labor supply shock, a housing preference shock, a cost-push shock, a monetary policy shock, a shock on the central banks inflation target, sectoral productivity shocks (housing, consumption and non-residential sector)

- Calibration/Estimation: The model is estimated with Bayesian methods using ten US observables over the period 1965:Q1 to 2006:QIV.

E.10 The CMR14 Model

Christiano et al. (2014) augment a standard DSGE model such as Smets and Wouters (2007) with a financial accelerator mechanism as in Bernanke et al. (1999). In particular, the return on capital of individual entrepreneurs is subject to idiosyncratic uncertainty. The model is fitted to US data, while modeling aggregate risk as the innovation to the variance of the distribution determining the return on capital. The papers main-finding is that fluctuations in risk are the most important shock driving the business cycle.

- Aggregate Demand: Households maximize expected lifetime utility by choosing consumption of final goods, labor supply and investment. They obtain funds from supplying labor, purchasing long- and short-term bonds, building and selling raw capital, as well as from various lump-sum transfers. Further, each household is subject to taxes on consumption and labor income.

- Aggregate Supply: Competitive final-goods producers purchase and combine intermediate goods from monopolistic intermediate-goods producers. These produce by employing labor and renting capital while subject to Calvo-style rigidities. Homogenous labor units are produced by perfectly competitive labor contractors which aggregate differentiated household labor services purchased from monopolistic unions that set wages subject to Calvo-style frictions. Households build raw capital subject to capital-adjustment costs and sell it to entrepreneurs, which they own.

- Financial Sector: Risk-neutral entrepreneurs finance their purchases of capital through their net worth and loans from competitive mutual funds. The loan contract between entrepreneurs and mutual funds is as in Bernanke et al. (1999). However, the authors introduce a shock to the variance of idiosyncratic productivity that influences individual entrepreneurs return on capital. It is referred to as a risk shock. With an agency problem between entrepreneurs and mutual funds, a positive risk shock increases the required return on borrowing, that is, the external finance premium.

- Shocks: The model includes shocks to the following 12 variables: price markup, price of investment goods, government consumption, technology growth persistence, technology (transitory), risk, consumption preference, marginal efficiency of investment, term structure, equity, monetary policy, and the inflation target.

- The model is estimated by Bayesian techniques using 12 quarterly observables covering the period 1985:Q1 to 2010:Q2. The data set includes 8 macroeconomic and 4 financial variables.

E.11 The QPM08 Model

Carabenciov et al. (2008) design and estimate two versions of a small projection model for the U.S. economy: one with financial real linkages and one without, the latter is the model employed in our investigation. These models are part of the IMF research agenda in developing a Small Quarterly Global Projection Model (GMP) which consists of many small country models integrated into a single global market. Both versions of the model consist of few behavioral equations, focusing on the joint determination of output, unemployment, inflation and the federal funds rate.

- Aggregate Demand: The behavioral IS curve relates the output gap to its past and expected future value, to the past value of the short interest rate gap and to a disturbance term. This specification allows for inertia and persistent effects of the shocks. In the model with financial linkages, the output gap is a function of a financial variable as well, constructed using information from FEDs quarterly Senior Loan Officer Opinion Survey on Bank Lending Practices. This variable enters in the form of a shock and it is supposed to reflect the bank lending conditions (tightening or loosening). Thus, if lending conditions are tighter than anticipated, the effect will be a lower output gap and a weaker economy.

- Aggregate Supply: In the Phillips curve equation, inflation is linked to its past and expected future values, to the lagged output gap and a disturbance term. This representation reflects the way agents set their prices: a share of them uses indexation to past inflation and others are forward looking. These expectations are based on model-consistent estimates of future inflation.

- Shocks: A shock to the level and the growth rate of potential output, a shock to the level and the growth rate of the equilibrium rate of unemployment, a shock to the equilibrium real interest rate. In the model with financial linkages, a financial shock is introduced in addition and cross correlations of the error terms between certain shocks are allowed.

- Estimation: Both models are estimated with Bayesian techniques, using U.S. quarterly data over the period 1994:Q12008:Q1.

E.12 The DS04 Model

Del Negro and Schorfeide (2004) employ Bayesian estimation methodologies to estimate a small-scale New Keynesian model with the standard three-equations. The economy depicted in DS04 include a forward-looking representative household, a continuum of monopolistically competitive firms, a central bank and a fiscal authority which collects lump-sum tax. The log-linearized equations resulting from the household and firm problems solution correspond to the IS-curve and the Phillips curve, the third equation to close the model is the monetary policy set by the central bank. The model presents microecononomic foundations, as it accounts for the developments in designing policies related parameters following the famous Lucas critique. Nominal frictions enter the model through price rigidities and the aforementioned monopolistic competition. The model is estimated using data for output, inflation and interest rates.

- The demand side is represented by the IS curve derived from the log-linearization of the Household maximization problem and imposing the market clearing condition. In the model specification the IS-equation exhibits a looking-forward behaviour for output formation. The IS-equation involves two exogenous shocks, both following an AR(1) process, one for government spending and one for technology progress.

- Aggregate Supply: The continuum of monopolistically competitive firms faces demand curves derived from a Dixit-Stiglitz final good aggregator. The firms maximization problem leads to a Phillips curve where inflation is linked to its expected future values.

- Shocks: The model present three shocks. The first one is a government spending shock which follows a AR(1) process and is linked to output and inflation; the second one is a technology shock which also follows a AR(1) process and is linked to the output; the third one is a monetary policy shock and is iid distributed.

- Estimation: DS04 is estimated through Bayesian techniques. The observables variables included in the measurement equations correspond to quarterly output growth, quarterly inflation, and the quarterly federal funds rate. The data sample ranges from 1955Q3 to 2001Q3.

E.13 The WW11 Model

Wieland and Wolters (2011) augment a standard monetary three-equations NK model such as DS04 with additional economic shocks. In particular, a preference shock and a mark-up shock are the additional innovation processes integrate to the standard framework. Thus, the model is able to display richer fluctuations in output and inflation dynamics. The three-equations characterising the model correspond to forward looking IS and Phillips Curve and the monetary rule set by the central bank; the latter is the same of DS04.

- Aggregate Demand: The demand side is represented by the usual IS curve derived from the log-linearization of the Household maximization problem and imposing the market clearing condition. As in Del Negro and Schorfeide (2004), the IS-equation exhibits a forward-looking behaviour for output formation. Besides the government spending shock and the technology shock, both part of the IS-curve derived in DS04, also the preference shock enters the demand side, following a AR(1) process.

- Aggregate Supply: The firms maximization problem leads to a Phillips curve where inflation is linked to its expected future values. Both the preference and the mark-up shocks enter the New-Keynesian Phillips curve. The mark-up shock has a direct effect on inflation. The preference shock has an indirect effect on inflation as it is linked to marginal costs. Both shocks follow a AR(1) process.

- Shocks: The model present five shocks. In particular, all the three shocks embedded in DS04 are present and two more shocks are included, a preference shock and a mark-up shock.

- Estimation: WW11 is estimated using Bayesian techniques, using quarterly real-time data for real output, the output deflator and the effective federal funds rate. The model is estimated for the period 1980Q1-2009Q4.

E.14 The Fair Model

The Cowles Commission type model by Fair (2004) is based on economic theory and is a multi-country model, such that it is possible to identify a part of the model related to the US economy, and another part describing the economies of Rest of the World (ROW), which includes 38 additional countries. The model consists of 30 behavioral equations and 97 identities for what regards the part related to the US economy, while 15 stochastics equations and 22 additional identities for each foreign country. Concerning the level of disaggregation of the model, Fair (2004) considers six different sectors: Household sector, production sector, financial sector, federal government sector, state and local government sector and foreign sector. Furthermore, the model is consistent with the Lucas critique, since it stresses microfoundations; besides, the model departs from the rational expectations hypothesis, given the difficulties arising testing it.

- Estimation: The data employed in the estimation process include almost 100 observables collected for the period that goes from 1954 to 2002. The model is estimated equation-by-equation through 2SLS.

F 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 more extensive treatment of the topic we instruct the reader to consult the excellent works of 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.

F.1 Specification

A unique stable solution of the DSGE model in its most general specification is a difference equation [10]:

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

where S_t is the (endogenous) state vectors of length n, ε_t is the vector of exogenous shocks, i.e. innovations. The disturbances ε_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 ξ_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 Ω_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

[10] 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[11] we can define the likelihood of the model, i.e. the probability distribution of the data given parameters of the model:

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

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

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

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

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

F.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:

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

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

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

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.

[11] 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 $0 = T(\theta) P_{0|0} T(\theta) - P_{0|0} + RR'$.

G Data Collection

Table 7 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*, *BOGZ1FL141104005Q*, 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 7: Description of Raw Data

Name	Description	Units	Seasonally Adjusted	Source	Vintage Dates 01:I	01:II	01:III	01:IV	08:III	08:IV	09:I	09:II	20:I	20:II	20:III
Variables updated in daily frequency															
DFF	Effective Federal Funds Rate			FRED											
DBAA	Moody's Seasoned BAA Corporate Bond Yield			FRED											
DGS10	10-Year Treasury Constant Maturity Rate	%		FRED	Same as the SPF deadlines[1]										
DTB3	3-Month Treasury Bill: Secondary Market Rate	Index		FRED											
WILL5000IND	Wilshire 5000 Total Market Index			ALFRED											
C0091Y	Industrial BBB Corporate Bond Yield			BB											
Variables updated in monthly frequency															
CE16OV	Employment Level	Thousands of Persons	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
CNP16OV	Population Level	Thousands of Persons		ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
AWHNONAG	Average Weekly Hours of Production and Nonsupervisory Employees	Hours	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
UNRATE	Unemployment Rate	%	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
CPIAUCSL	Consumer Price Index for All Urban Consumers	Index 1982-1984=100	•	ALFRED	01-17	04-17	07-18	10-19	07-16	10-16	01-16	04-15	02-11	05-12	08-12
PCEC96	Real Personal Consumption Expenditures	Billions of Dollars	•	ALFRED	02-01	04-30	07-31	11-01	08-04	10-31	02-02	04-30	01-31	04-30	07-31
PCE	Personal Consumption Expenditures	Billions of Dollars	•	ALFRED	02-01	04-30	07-31	11-01	08-04	10-31	02-02	04-30	01-31	04-30	07-31
PCENDC96	Real Personal Consumption Expenditures: Nondurable Goods	Billions of Dollars	•	ALFRED	02-01	04-30	07-31	11-01	08-04	10-31	02-02	04-30	01-31	04-30	07-31
PCEDG	Personal Consumption Expenditures: Durable Goods	Billions of Dollars	•	ALFRED	02-01	04-30	07-31	11-01	08-04	10-31	02-02	04-30	01-31	04-30	07-31
PCEDGC96	Real Personal Consumption Expenditures: Durable Goods	Billions of Dollars	•	ALFRED	02-01	04-30	07-31	11-01	08-04	10-31	02-02	04-30	01-31	04-30	07-31
PCES	Personal Consumption Expenditures: Services	Billions of Dollars	•	ALFRED	02-01	04-30	07-31	11-01	08-04	10-31	02-02	04-30	01-31	04-30	07-31
PCESC96	Real Personal Consumption Expenditures: Services	Billions of Dollars	•	ALFRED	02-01	04-30	07-31	11-01	08-04	10-31	02-02	04-30	01-31	04-30	07-31
CLF16OV	Civilian Labor Force Level	Thousands of Persons	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
BAA	Moody's Seasoned Baa Corporate Bond Yield	%		FRED	02-06	05-08	08-07	11-01	08-04	11-03	02-02	04-30	02-03	05-01	08-03
TB3MS	3-Month Treasury Bill: Secondary Market Rate	%		FRB	02-06	05-08	08-07	11-06	08-04	11-03	02-02	05-04	02-03	05-01	08-03
PAYEMS	All Employees, Total Nonfarm	Thousands of Persons	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
USCONS	All Employees, Construction	Thousands of Persons	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees	Hours	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
AWHAECON	Average Weekly Hours of All Employees, Construction	Hours	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
CES2000000008	Average Hourly Earnings of Production and Nonsupervisory Employees	Dollars per Hour	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
CPILFESL	Consumer Price Index for All Urban Consumers	Index 1982-1984=100	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
EMRATIO	Employment-Population Ratio	%	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07

Name	Description	Units	Seasonally Adjusted	Source	Vintage Dates										
					01:I	01:II	01:III	01:IV	08:III	08:IV	09:I	09:II	20:I	20:II	20:III
CIVPART	Labor Force Participation Rate	%	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
JTSJOL	Job Openings: Total Nonfarm	Thousands of Dollars	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
PCEND	Personal Consumption Expenditures: Nondurable Goods	Billions of Dollars	•	ALFRED	02-01	04-30	07-31	11-01	08-04	10-31	02-02	04-30	01-31	04-30	07-31
CES2000000007	Average Weekly Hours of Production and Nonsupervisory Employees	Hours	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
AHETPI	Average Hourly Earnings of Production and Nonsupervisory Employees	Dollars per Hour	•	ALFRED	02-02	05-04	08-03	11-02	08-01	11-07	02-06	05-08	02-07	05-08	08-07
MORTRATE	Effective Interest Rate on Conventional Single Family Mortgages PLUS MIRS Transition Index	%		FHFA	N/A										
Variables updated in quarterly frequency															
FPI	Fixed Private Investment	Billions of dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
GDPC1	Real Gross Domestic Product	Billions of dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PCEC	Personal Consumption Expenditures	Billions of dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PRFI	Private Residential Fixed Investment	Billions of dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PNFI	Private Nonresidential Fixed Investment	Billions of dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PCND	Personal Consumption Expenditures: Nondurable Goods	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PCESV	Personal Consumption Expenditures: Services	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PCDG	Personal Consumption Expenditures: Durable Goods	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PCNDGC96	Real Personal Consumption Expenditures: Nondurable Goods	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PCESVC96	Real Personal Consumption Expenditures: Services	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PCDGCC96	Real Personal Consumption Expenditures: Durable Goods	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
A006RD3Q086SBEA	Gross private domestic investment (implicit price deflator)	Index 2012=100	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
GDPCTPI	Gross Domestic Product: Chain-type Price Index	Index 2012=100	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
A007RD3Q086SBEA	Gross private domestic investment: Fixed investment	Index 2012=100	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
GPDIC1	Real Gross Private Domestic Investment	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
GPDI	Gross Private Domestic Investment	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
HOANBS	Nonfarm Business Sector: Hours of All Persons	Index 2012=100	•	ALFRED	02-07	05-08	08-07	11-07	06-04	06-11	02-05	05-07	02-06	05-07	06-04
NETEXP	Net Exports of Goods and Services	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
NETEXC	Real Net Exports of Goods and Services	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
TOTLQ	Hours of Wage and Salary Workers on Nonfarm Payrolls: Total	Billions of Hours	•	ALFRED	N/A										
PNFIC1	Real Private Nonresidential Fixed Investment	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
PRFIC1	Real Private Residential Fixed Investment	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
IPDNBS	Nonfarm Business Sector: Implicit Price Deflator	Index 2012=100	•	ALFRED	02-07	05-08	08-07	11-07	06-04	06-11	02-05	05-07	02-06	05-07	06-04
DNDGRD3Q086SBEA	Personal consumption expenditures: Nondurable goods	Index 2012=100	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
DSERRD3Q086SBEA	Personal consumption expenditures: Services	Index 2012=100	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30

Name	Description	Units	Seasonally Adjusted	Source	Vintage Dates										
					01:I	01:II	01:III	01:IV	08:III	08:IV	09:I	09:II	20:I	20:II	20:III
GCEC1	Real Government Consumption Expenditures and Gross Investment	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
COMPRNFB	Nonfarm Business Sector: Real Compensation Per Hour	Index 2012=100	•	ALFRED	02-07	05-08	08-07	11-07	06-04	06-11	02-05	05-07	02-06	05-07	06-04
PRS85006023	Nonfarm Business Sector: Average Weekly Hours	Index 2012=100	•	ALFRED	02-07	05-08	08-07	11-07	06-04	06-11	02-05	05-07	02-06	05-07	06-04
PRS85006141	Nonfarm Business Sector: Implicit Price Deflator	Percent Change From Quarter One Year Ago	•	ALFRED	N/A										
DDURRD3Q086SBEA	Personal consumption expenditures: Durable goods	Index 2012=100	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
GDP	Gross Domestic Product	Billions of Dollars	•	ALFRED	01-31	04-27	07-27	10-31	07-31	10-30	01-30	04-29	01-30	04-29	07-30
CBHPI	Census Bureau House Price Index	N/A		N/A	01-31	04-25	07-27	10-26	07-25	10-26	01-29	04-24	01-27	04-23	07-24
COMPNFB	Compensation Per Hour of Nonfarm Business Sector	Index 2012=100	•	ALFRED	02-07	05-08	08-07	11-07	06-04	06-11	02-05	05-07	02-06	05-07	06-04
BOGZ1FL144104005Q	Credit Market Liabilities of Nonfinancial Business Sector	Millions of Dollars		ALFRED	N/A										
HMLBSHNO	Home Mortgage Liabilities of Households and Nonprofit Organizations	Billions of Dollars		FRED	N/A										
FGS	Financing Gap Share	%	•	COMP	02-01	05-01	08-01	11-01	07-01	11-01	02-02	05-01	02-01	05-01	08-01
BLT	Bank Lending Tightening Index (Carabenciov et al., 2008)	%		FRB	02-05	02-05	05-17	11-13	08-11	11-03	02-02	05-04	02-03	05-04	08-03
				SPFdeadline:	02-14	05-12	08-15	11-14	08-07	11-10	02-10	05-12	02-11	05-12	08-12

H Observed Variables

Table 8 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 8: Description of Observed Variables

Name	Description	Construction	Has SPF Nowcast Value	Has Current-Quarter Value	DS04	WW11	SW07	FRBEDO	NKBGG	QPM08	DNGS15	KR15FF	KR15HH	CMR14	GSW12	IN10	FU20	3vBVAR	5vBVAR	8vBVAR
Common variables																				
gdp_rgd_obs	Real GDP growth	ΔLN(GDPC1)*100	•		•	•	•	•	•		•	•	•	•	•		•	•	•	•
gdpdef_obs	GDP deflator growth	ΔLN(GDPCTPI)*100	•		•	•	•	•	•		•	•	•	•	•		•	•	•	•
ffr_obs	Fedeal funds rate	DFF/4		•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•
ifi_rgd_obs	Real invest. growth: fixed invest.	ΔLN(FPI/GDPCTPI)*100					•		•		•	•			•		•		•	•
c_rgd_obs	Real cons. growth	ΔLN(PCE/GDPCTPI)*100					•				•	•	•		•		•			•
wage_rgd_obs	Real wage growth	ΔLN(COMPNFB/GDPCTPI)*100					•	•			•	•	•	•	•		•			•
baag10_obs	Credit spread	(DBAA-DGS10)/4		•					•		•			•					•	•
gdpl_rgd_obs	Real GDP level	LN(GDPC1)*100	•							•						•				
cpil_obs	CPI level	LN(CPIAUCSL)*100	•							•										
ir_nom_obs	Nominal invest. (residential) growth	ΔLN(PRFI)*100 SPF through PRFIC1 and GDPCTPI	•					•					•							
inr_nom_obs	Nominal invest. (non-residential) growth	ΔLN(PNFI)*100 SPF through PNFIC1 and GDPCTPI	•					•					•							
Model-specific variables																				
blt_obs	Bank lending tightening index	BLT		•						•										
emp_obs	Total employment	ΔLN(CE16OV)*100													•					
unr_obs	Unemployment rate	UNRATE	•							•					•					
hours_dngs15_obs	Hours in the DNGS15 model	LN(AWHNONAG*CE16OV/100/(CNP16OV/3))*100	•								•									•
hours_sw07_obs	Hours in the SW07 model	LN(PRS85006023*(CE16OV/118753) /(CNP16OV/193024.333))*100 - its mean					•										•			
cnds_nom_obs	Nominal cons. (non-durables/services) growth	ΔLN(PCEND+PCES)*100						•												
cd_nom_obs	Nominal cons. (durables) growth	ΔLN(PCEDG)*100						•												
cnds_def_obs	Cons. deflator (non-durables/services) growth	ΔLN((PCEND+PCES)/(PCENDC96+PCESC96))*100						•												
cd_def_obs	Cons. deflator (durables) growth	ΔLN(PCEDG/PCEDGC96)*100						•												
hours_frbedo08_obs	Hours in the FRBEDO08 model	Divide by mean: AWHNONAG*CE16OV/CNP16OV	•	•				•												
mortffr_obs	Mortgage spread	(MORTRATE-DFF)/4		•									•							
bbb1yffr_obs	Loan spread	(C0091Y-DFF)/4		•								•								
hp_nom_obs	House price growth	ΔLN(CBHPI)*100											•							
credit_nom_obs	Nominal credit growth	ΔLN(BOGZ1FL144104005Q)*100										•								

Name	Description	Construction	Has SPF Nowcast Value	Has Current-Quarter Value	Model															
					DS04	WW11	SW07	FRBEDO	NKBGG	QPM08	DNGS15	KR15FF	KR15HH	CMR14	GSW12	IN10	FU20	3vBVAR	5vBVAR	8vBVAR
mortgage_nom_obs	Nominal mortgage growth	ΔLN(HMLBSHNO)*100											•							
hours_kr15_obs	Hours in the KR15_FF and KR15_HH models	LN(12*PRS85006023*CE16OV/CNP16OV)*100										•	•							
g10ffr_obs	Term premium	(DGS10-DFF)/4		•										•						
credit_rgd_obs	Real credit growth	ΔLN(BOGZ1FL144104005Q/GDPCTPI)*100												•						
hours_cmr14_obs	Hours in the CMR14 model	LN(HOANBS/CNP16OV)												•						
cnds_rim_obs	Real cons. growth: nondurable/services	ΔLN(PCESC96+PCENDC96)*100												•						
igid_rim_obs	Real invest. growth: durable/gross invest.	ΔLN(PCEDGC96+GPDIC1)*100												•						
igiddef_rgd_obs	Real invest. deflator growth	ΔLN(((GPDI+PCEDG)/(GPDIC1+PCEDGC96)) /GDPCTPI)*100												•						
networth_rgd_cmr14_obs	Real net worth growth in the CMR14 model	EXP(ΔLN(WILL5000IND/GDPCTPI) - its mean)												•						
rc_obs	Real consumption	LN(PCEC96/CNP16OV) - its first value														•				
pi_dm_obs	Inflation	ΔLN(IPDNBS) - its mean														•				
rri_obs	real residential investment	LN(PRFIC1/CNP160V) - its first value														•				
rbi_obs	real business investment	LN(PNFIC1/CNP160V) - its first value														•				
hwc_pd_obs	hours worked in the goods sector	LN((PAYEMS-USCONS)*AWHMAN/CNP16OV) - its mean		•												•				
hwr_pd_obs	hours worked in the residential sector	LN(USCONS*CES2000000007/CNP16OV) - its mean		•												•				
hp_r_obs	real house prices	LN(CBHPI/IPDNBS) - its first value														•				
i_nom_obs	Nominal interest rate	TB3MS/400 - its mean		•												•				
c_winf_obs	Wage inflation in consumption-good sector	ΔLN(AHETPI) - its mean		•												•				
h_winf_obs	Wage inflation in housing sector	ΔLN(CES2000000008) - its mean		•												•				

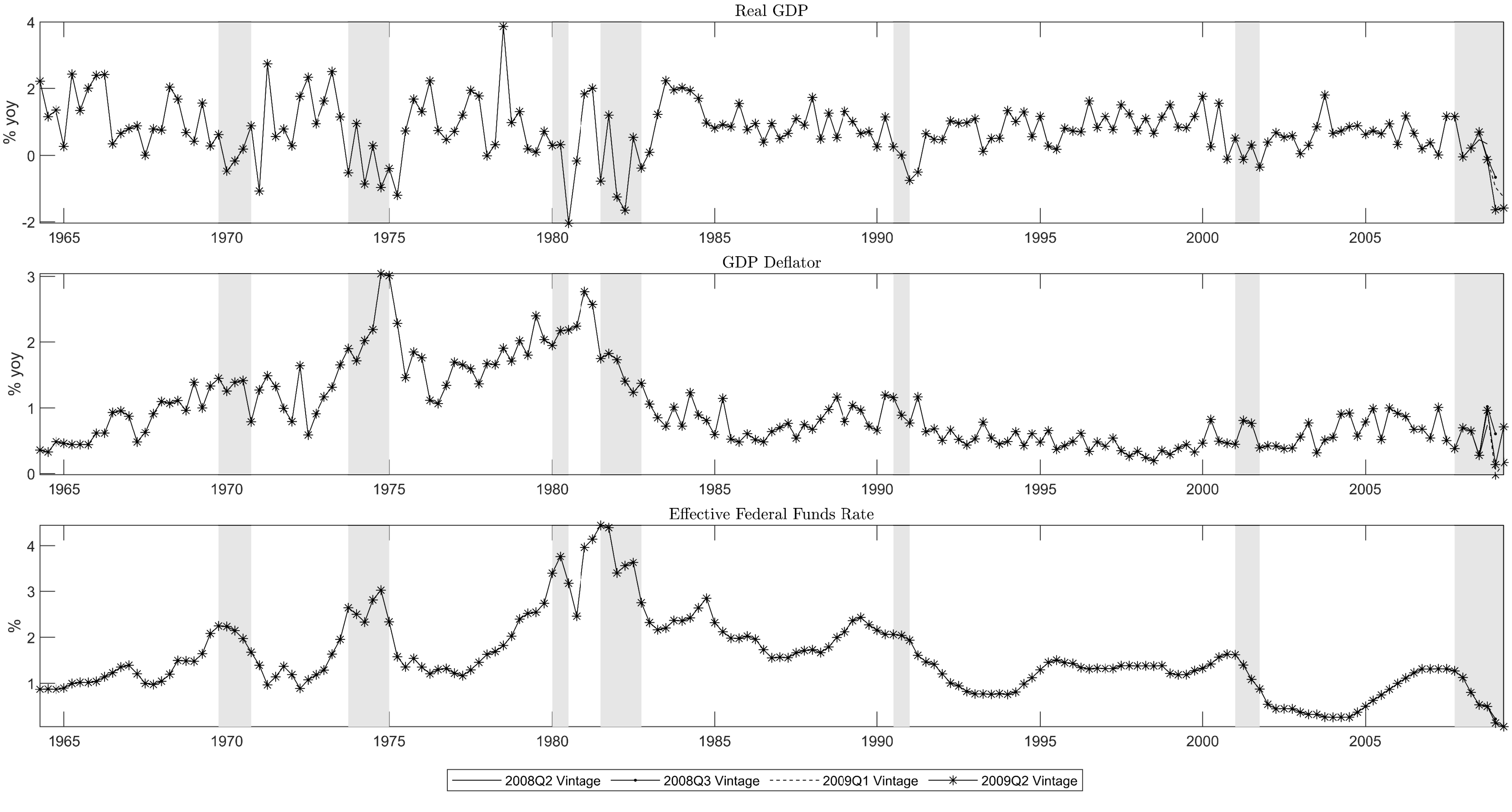


Figure 33: Real GDP, Inflation and Federal Funds Rate

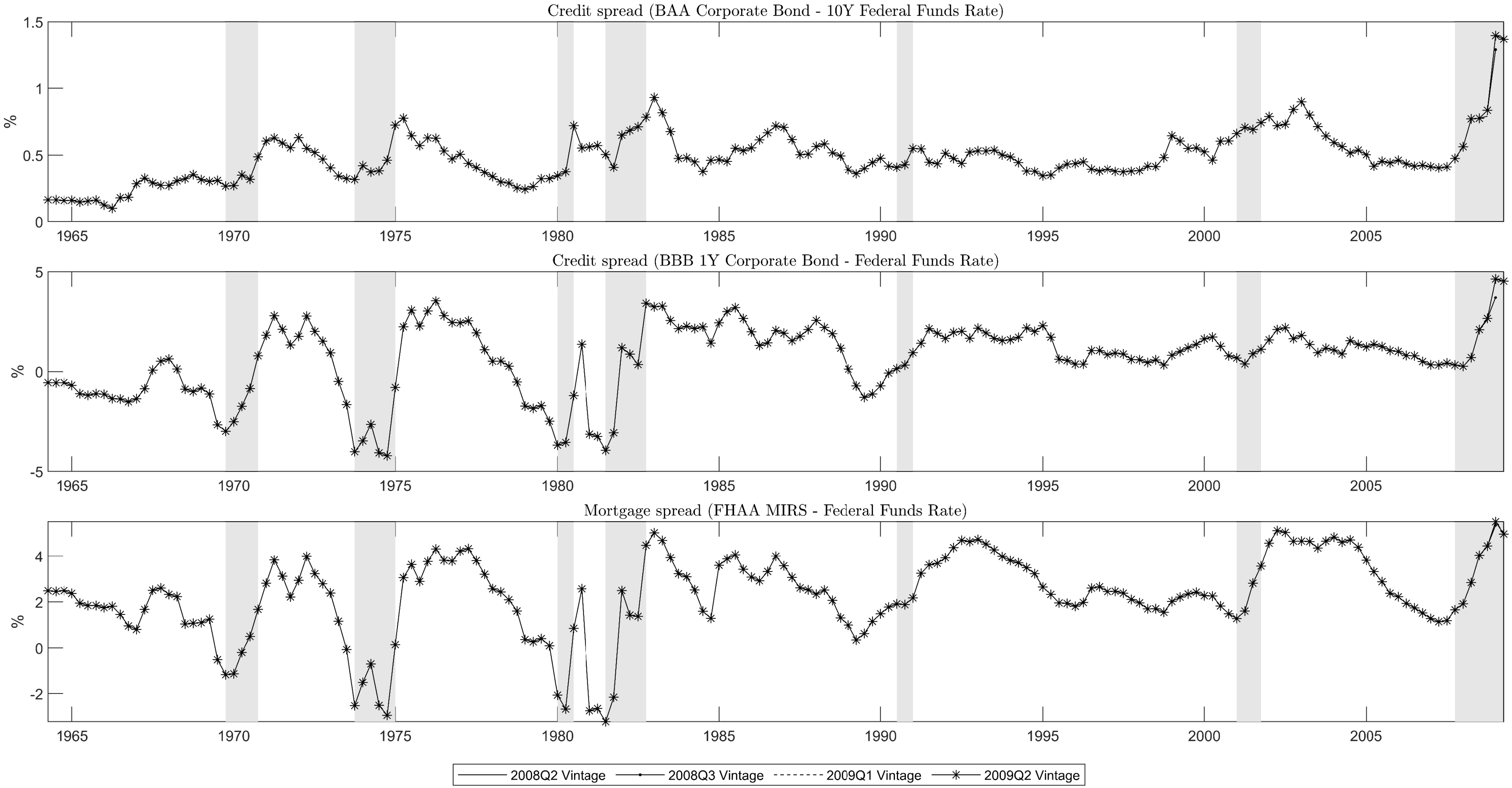


Figure 34: Credit Spread Variables

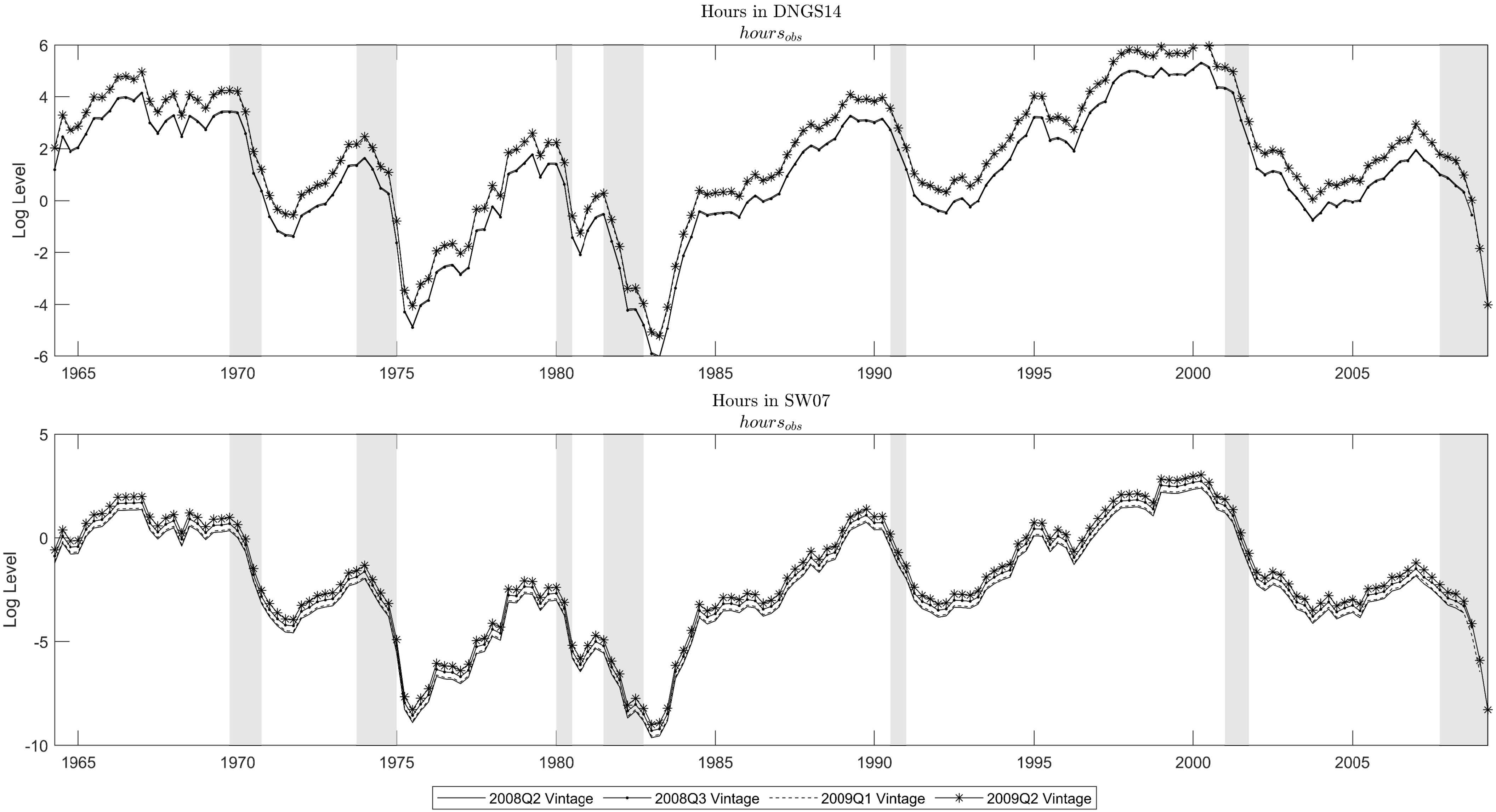


Figure 35: Comparison of Hours Worked Series Definitions