Judging the DSGE Model by Its Forecast

Rochelle M. Edge

Refet S. Gürkaynak

Burçin Kısacıkoğlu

Federal Reserve Board
rochelle.m.edge@frb.gov

Bilkent University and CEPR
refet@bilkent.edu.tr

Johns Hopkins University
bkisaci1@jhu.edu

Abstract

We study the forecasting ability of the standard estimated medium scale dynamic stochastic general equilibrium model. We show that although over the Great Moderation period the model forecasts have good *relative* forecasting ability, in an *absolute* sense their forecasting ability is poor. However, we argue that average forecasting ability during the Great Moderation is not a good metric to judge a model's validity given that this is a period that is well-known to be characterized by a lack of persistent fluctuations in the data. We then consider the forecast performance of the DSGE model prior to the Great Moderation – a period that is known to be characterized by persistent and thereby forecastable fluctuations in the data generating process – and find notably better *absolute* forecasting performance. We then offer alternative ways of using forecasts to judge the empirical validity of the model. In particular, we suggest that looking at whether the model captures well the forecastability versus nonforecastability of the data upon which the model is estimated is a more fitting question than simply whether the model forecasts well. As part of the empirical analysis that we undertake to address our key questions, we also uncover and document the importance of data and sample choices in the model's forecasting ability.

---VERY PRELIMINARY AND VERY INCOMPLETE---
---PLEASE DO NOT CITE---

March 17, 2013

1.   Introduction

The forecasting performance of the DSGE models has been an issue of considerable interest to policymakers and researchers alike. This interest has two dimensions: An interest in the forecasting ability of DSGE models for its own sake, given the possibility of using such models for forecasting at policy institutions, and an interest in the forecasting ability of DSGE models as a way to validate (or invalidate) the model itself. In this paper we argue that these two questions are different and on the second issue that the model's forecasting ability need not necessarily tell us anything about the empirical validity of the DSGE model. We then offer a different way to use the model's forecast performance to consider the model's empirical coherence. In particular, we suggest that rather than asking whether the model forecast well, we should be asking whether the model captures well the forecastability versus nonforecastability of the data upon which it is estimated.[1]

The use of a model's forecasting performance to make inferences about its empirical relevance dates back at least to the influential paper of Atkeson and Ohanian (2001) that documented the inability of Phillips curve models to outperform simple random walk forecasts and thus called into question the ability of Phillips Curve models to explain inflation data well. The analysis of forecast performance has also been used extensively in the DSGE modeling literature. Smets and Wouters (2007), most notably, documented the competitive forecasting performance of their richly-specified DSGE model relative to respectable alternative models (specifically, a Bayesian VAR model) and dramatically altered consensus opinion about the empirical relevance of DSGE models. Subsequently, for central bank modeling teams evaluating the DSGE models that they have developed for practical use, establishing the competitive forecast performance of these models has become standard practice.[2] Associated with this, the benchmark against which DSGE model forecasts are evaluated has broadened so as to include official staff or policy committee macroeconomic forecasts, and in such exercises the use of real-time data has also become more common.

---

[1] We are not the first to argue out that using forecast performance to evaluate a model is a different issue to what of whether a model produces good forecasts and thereby should be used for forecast generation. Clements and Hendry (2005) have also made this point, although they did not highlight the data's being unforecastable as a reason for not using forecast performance to evaluate a model.

[2] For examples of this type of research see Adolfson, Andersson, Linde, Villani, and Vredin (2007) for the Swedish Riksbank DSGE model, Lees, Matheson, and Smith (2007) for the RBNZ's DSGE model, Edge, Kiley, and Laforte (2010) for the Federal Reserve Board's DSGE model, and Christoffel, Coenen, and Warne (forthcoming) for the ECB's New Area Wide DSGE model. To be sure, in some cases the purposes for doing this evaluation is because the model is intended to be used for forecast generation. In other cases, however, evaluation is undertaken to provide validation more broadly for the model.

Overwhelmingly, this literature has found favorable results for DSGE models; that is, in all cases DSGE models have been found to be competitive with alternative models, including official forecasts.

While previous research documented the competitive *relative* forecast performance of DSGE models, Edge and Gürkaynak (2010) also examined *absolute* forecast performance – in this case using the Smets and Wouters (2007) model – and here found more discouraging results. In particular, whereas *relative* forecasting performance of the DSGE model was found to be competitive, the *absolute* forecasting performance was found to be poor. Ultimately, however, this should not have been surprising given the sample period used for evaluating forecast performance. Edge and Gürkaynak's analysis, like all DSGE-model research on forecast performance, was undertaken over the Great Moderation period – specifically, over the period 1992 to 2006 – a period that is known to be characterized by a lack of persistent fluctuations in the data generating process. Given this feature of the data and the fact there is little to be forecast over this period, making inferences about the empirical validity of a model based on its forecasting performance is somewhat counter-intuitive.[3] (In effect, we are asking how well the DSGE model can do something that we know is not possible to do at all.) This point is especially valid for inflation, usually one of the key variables of interest in forecasting exercises.

The fact that DSGE-model forecast performance analysis on Great Moderation data is not amenable to any interpretation regarding the validity of the underlying model was pointed out by Edge and Gürkaynak (2010). However, Edge and Gürkaynak did not pursue some of the questions that immediately follows this observation and that we pick-up in this paper. For example, Edge and Gürkaynak did not consider the performance of DSGE-model forecasts prior to the Great Moderation; that is, over a period when macroeconomic aggregates are known to have been persistently varying and to have, at least, the potential to be able to be forecast by the model. Additionally, they did not consider whether a model's forecasting ability can be used in some other way to assess its empirical validity when it is known that the potential for it being able to forecast key macroeconomic aggregates varies over time.

The first of these questions at least – and it turns out the second as well – require extending our sample period back to before the Great Moderation; specifically, to the early

---

[3] See Stock and Watson (2007) and Tulip (2009) for papers that document the lack of persistent fluctuations in the data in the Great Moderations period and the fact that there is little in the data over this period that can be forecasted.

1970s. This presents one significant constraint on our analysis, which is the absence of a full set of real-time data for model estimation. Thus, for our forecast-performance analysis in the period prior to the Great Moderation we have no choice but to conduct our analysis using, what we will call, the "final" vintage of data; specifically, the vintage data ending 2012:Q2.[4] This switch to final vintage data means that the forecast-performance results from Edge and Gürkaynak, which were undertaken entirely on real-time data, will not be directly comparable with those that we will generate on earlier periods of data. Thus, to ensure the comparability of our results in the Great Moderation and Pre-moderation forecast evaluation periods we first revisit the results of Edge and Gürkaynak and compare the out-of-sample forecasting performance of the model using real-time and final vintage data for the period both are available.

Revisiting the results of Edge and Gürkaynak so as to use final vintage data – rather than real-time data – is not a wholly straightforward exercise, in as much as, more than just the data upon which the model is estimated ends up changing. For example, in Edge and Gürkaynak forecasts generated by the DSGE model (estimated in real-time) were compared with the first-final release of the data series in question. Clearly, this is no longer possible, or sensible, when the model is estimated on final vintage data. Indeed, in this case, it is realizations of the final vintage of data that are the appropriate point of comparison. Since the model's forecast performance over the Great Moderation sample used in Edge and Gürkaynak forms the basis of our subsequent comparisons (with pre-Moderation forecast results), documenting how these results differ from those of Edge and Gürkaynak on real-time data is necessary.

One other issue that arises when pushing the period of our analysis substantially back in time is that of structural change. If we are to estimate our model in the 1970s we may like to use data extending back to the 1950s, since this will give us 20 years of data. But, given structural change, this may not necessarily be the data that we would want to use to estimate the model in later decades. Consequently, rather than using an expanding sample for estimation as was used by Edge and Gürkaynak we use a 20-year rolling window.

Appendix C [to be included in place of much of section two] documents how the forecast performance of the DSGE model changes when the analysis moves to use final

---

[4] Although it is now a few months since our "final" vintage of data and although this data will revise again in July 2013 (when another annual revision takes place), for the years that we are interested in this paper the data will not revise again until the next comprehensive revision. Thus, for the purposes of this paper, we can refer to this vintage it as "final" vintage data.

vintage data and to using rolling window estimation. Section two of the paper will ultimately report only the highlights of these results [though currently it includes all of results]. On balance, our key result that the model forecasts poorly over the Great Moderation period – particularly, for inflation – remains. That being said, our results regarding relative forecasting performance of the DSGE model – particularly, for inflation – become a little less favorable.

There is, however, one change that we make from Edge and Gürkaynak (2010) that is not the result of our move to work with final vintage data or a longer time series. Rather, this change is motivated by a data issue with Edge and Gürkaynak that came to light only quite late in the course of writing that paper and thus could not be changed in time for that paper's publication. As noted previously, Edge and Gürkaynak undertook all of their DSGE model forecast analysis using the Smets and Wouters (2007) model and following as closely possible all of the assumptions and the data definitions used by these authors.[5] This included using the population series used by Smets and Wouters that Edge and Gürkaynak subsequently found to be problematic.[6] As explained in section two, the population series used by Smets and Wouters is calculated on a "best levels" basis, which results in some very large jumps and drops in the series that we know are not true features of the data. In this paper we use a smoothed population series, which is closer to the true population series. As we discuss in section two, we find that using a more appropriate population series does improve the model's forecasting performance of real GDP growth.

The main part of our analysis begins in section three. After establishing that Edge and Gürkaynak's main result – that the DSGE model forecasts poorly in an absolute sense over the period 1992 to 2006 – holds true also in final vintage data, we consider the model's forecasting performance completely prior to the Great Moderation – in this case, 1970 to 1984. Here we document better absolute forecasting performance of our DSGE model in the

---

[5] The Smets and Wouters model is used in Edge and Gürkaynak – as well as in our analysis – in preference to the other variants of Bayesian estimated DSGE models in existence, such as Edge et al (2010) and Gali et al (2010), which are extensions on the canonical Smets and Wouters model. These latter models add more features to the Smets and Wouters model but do not qualitatively change the model's forecasting performance. Consequently, we focus on the benchmark Smets and Wouters model, which is well known and well understood and abstract from the more recent variants of the model, which do not have a material effect on our broad conclusions.

[6] There was one data series that Edge and Gürkaynak used that differed from Smets and Wouters and this was the series for hours. Smets and Wouters used the measure nonfarm business sector average weekly hours as reported by the Labor Productivity and Costs (LPC) release, whereas Edge and Gürkaynak used average weekly hours of production and nonsupervisory employees in private industries as measured by the Employment Situation Survey (ESS). The sole reason for this difference in data series was the availability of real-time data for the ESS series and the lack of real-time data for the LPC series.

Pre-moderation period, when we know macroeconomic aggregates were forecastable. In addition to examining forecast performance over these two 15-year intervals of 1970 to 1984 and then 1992 to 2006, we also document forecast performance over rolling [15-year or 60-quarter] windows between these two intervals, so as to understand how forecast performance evolved. [Currently, 20-year/80-quarter windows are shown.]

Section four [to be completed] then asks whether good or bad *absolute* forecast performance is a feature of the data that the DSGE model can replicate if the model itself is estimated on data that is forecastable or unforecastable. With the Great Moderation and pre-Moderation versions of the model that we estimate in section three, we simulate multiple realizations of data from these models and evaluate the performance of the models' corresponding forecast. Asking how well our stochastic model forecasts under the assumption that the model is the true data generating process allows us to assess what the model judges that its capacity to forecast should be (given the data upon which it was estimated). We can then examine how these estimates of the model's forecasting capacity, given the data upon which it is estimated, lines up with the model's actual forecasting ability of these data. In general we find [to be completed].

If the results of section four find that the DSGE model's capacity to forecast diminished with the onset of the Great Moderations we are well positioned to use our DSGE model – estimated over a wide range of sample periods – to ask why this has been the case. We consider this question in section five very preliminarily by considering how monetary policy has changed across the sample periods that we consider. Here we find some suggestive evidence that more activist monetary policy could have reduced the forecastability of key macroeconomic aggregates over time. Finally, section six concludes.

Appendix A provides a brief qualitative description of the Smets-Wouters DSGE model and the B-VAR model and Appendix B describes the data series used to estimate these models. Note that all of the forecasts and simulated data generated in this paper use the DSGE and B-VAR models as parameterized at their posterior modes.

2.   Changes to data and consequent results

As noted above, Edge and Gürkaynak (2010) use real-time data for all of their analysis, whereas in this paper we use final vintage data. As a result, a requisite step in our analysis is to verify that the results obtained by Edge and Gürkaynak using real-time data hold true when final vintage data (over the same sample period) are instead used. This is what

subsection 2.4 is spent discussing. Before that, however, we make two other changes. The first is to correct for the population series that is used in the estimation of the model. The results from this are given in subsection 2.2. The second is to move from a sample that expands over time – from 1965 to the last date before the forecast begins – to a sample that is a 20-year rolling window. The results from this are given in subsection 2.3. Before presenting these results, however, we recap the results of Edge and Gürkaynak.

2.1   *Reviewing the Edge and Gürkaynak (2010) results*

Figure 1 reports the relative forecasting performance of the Smets-Wouters DSGE model, the (Smets-Wouters) Bayesian VAR model, and a random walk forecast in the Great Moderation sample and in all cases using real-time data. The results and comparisons of Figure 1 are effectively the same as those generated in Edge and Gürkaynak (2010) but with the random walk forecast replacing the Greenbook forecasts as one of the competitor forecasts. The motivation for dropping the Greenbook forecasts (and also Blue-Chip forecasts) for comparison with the DSGE model is that when we move to using final vintage data for model estimation and forecasting it will no longer make sense to compare the DSGE model forecasts with these judgmental forecasts, which are only available as real-time forecasts. As in Atkeson and Ohanian (2001) we therefore use the random walk forecast as the benchmark against which we judge the model's forecasting performance.

The left panels of Figure 1 shows the relative root mean square error of the DSGE model with respect to random walk forecasts, while the right panels show the comparison with respect to Bayesian VAR forecasts. The top panels show the results for inflation and the bottom panels show the results for GDP growth. As can be seen from these figures, the DSGE model either matches or outperforms both the random walk and the Bayesian VAR forecasts for both inflation and real GDP growth.

Table 1, in contrast, shows the absolute forecasting performance of these three forecasting methods, which are evaluated using forecast efficiency (or Mincer-Zarnowitz) regressions. As is evident from these results, all three methods are extremely poor forecasters with essentially no forecasting power (as measured by the R-squareds of the forecast efficiency regressions) in terms of capturing changes in inflation or real GDP growth. This result was the key finding of Edge and Gürkaynak (2010), which should not have been entirely surprising given that the analysis was undertaken over the Great

Moderation period, where there has been much less (relative to earlier periods) to be forecast in time series like inflation and real GDP growth.

Edge and Gürkaynak noted that the result that inflation and real GDP growth cannot be forecast need not be an entirely negative one, given that the result may be a reflection of policy preemptively moving against any *forecastable* changes in the objective variables and leaving fluctuations in these variables to be mainly due to *unforecastable* shocks. The problem with this explanation is that if more active and preemptive monetary policy is the main reason for the lack of forecastability, inflation and real GDP growth should not *both* be unforecastable. To be sure, in variants of models similar to the Smets-Wouters model, optimal policy can simultaneously close the output gap and set inflation equal to target – a property known as the "divine coincidence." However, this is a feature of new Keynesian models with only one source of nominal rigidity (*e.g.*, sticky prices *or* sticky wages) and without any mark-up shocks to the inflation equations. The DSGE model being used in this analysis has both two sources of nominal rigidity (sticky prices *and* sticky wages) and mark-up shocks, and as such does not feature the divine coincidence. Also, the forecasting exercise undertaken in this paper considers real GDP growth – not the output gap – and policy that makes the output gap unforecastable will still leave real GDP growth forecastable. That is, even if policy closed all forecastable inflation gaps then there should be some forecastability in output growth. As it turns out, the result of inflation and real GDP growth *both* being unforecastable reflects somewhat the model's use of the problematic and volatile population series, which is the issue we turn to now.

2.2   *Correcting the model's population series*

Many of the observable variables that are used to estimate the Smets and Wouters model enter the model in normalized per capita value form, which means that the level of population shows up in many parts of the model. The published US population series, shown in growth rates in Figure 2, is calculated on a "best levels" basis. That is, when new information about population becomes available – such as with decennial censuses or annual benchmarking of the Current Population Survey (CPS) – the population in the census year (and thereafter) is adjusted in line with the new data but population in preceding years is not. This leads to very large – and implausible – jumps in the population series, which in–reality is quite smooth. Figure 2 shows a large number of such jumps (as reflected in some extraordinarily large quarterly growth rates), while a smoothed series – which is taken from

the Federal Reserve Board's US (FRB/US) model – displays the expected gentle low frequency movements in population growth.

Population in the model is important in two aspects, first because the variables entering estimation are normalized by the series, second because the model's resulting per capita GDP growth forecasts are then multiplied with the population growth number to make it compatible with the released aggregate GDP growth number and with other forecasts. The erratic population growth series has the potential to damages the model in both dimensions.

A natural first instinct is to use the smooth the population series for all purposes in estimation of the model and generation of the forecasts. This, however, does not wholly solve the problem. The model's employment series, shown in Figure 3, has similar methodology issues to the population series and thus also exhibits spikes associated with decennial censuses and CPS benchmarkings.[7] While population growth can safely be thought of as a smooth series and can therefore be smoothed as shown in Figure 2, a similar smoothing procedure would change the cyclical properties of the employment series. However, employment never appears alone in the model. Rather, what matters is employment divided by population. Having one of the series smoothed and the other one unsmoothed would lead to jumps in the employment rate that are not really there in the data and would likewise hurt the model fit and forecast performance.

We therefore use the *unsmoothed* population series when normalizing employment, leading to a smoothed employment rate, and using *smoothed* population when normalizing other variables such as consumption, investment, and GDP, leading to conceptually appropriate data series used in estimation. We also use the smoothed population series when we convert the per capita GDP growth forecasts into aggregate GDP growth forecasts, which is the series that we ultimately evaluate in our forecast-performance exercises.

Figure 4 and Table 2 show, respectively, the relative and absolute forecasting performance of the Smets-Wouters DSGE model, the (Smets-Wouters) Bayesian VAR model,

---

[7] Note that this erratic behavior is a feature of the employment series reported in the Employment Report's *household* survey. The employment series taken from the Employment Report's *establishment* survey does not exhibit these jumps and neither does the employment series reported in the Labor Productivity and Costs release. However, it is the household survey employment series that Smets and Wouters used in their paper and so as to follow their methodology as close as possible we continue to use this employment series. The coverage of the household survey's employment series is broader than the (more widely quoted) establishment survey in that it includes self-employed workers and agricultural workers. It is possibly for this reason that Smets and Wouters chose the household survey employment series.

and a random walk forecast in the Great Moderation sample – in all cases using real-time data and the corrected population series. As the actual series to be forecasted have not changed (they are still real GDP growth and GDP price inflation), the change in the population series has no implications for the random walk forecast. But this change does help the DSGE model forecasts as expected.

The change in the performance of the DSGE model's inflation forecast is minor but the improvement in the GDP growth forecast is quite notable. This is somewhat to be expected. The population series affects the GDP growth forecasts in two ways – through the estimation of the parameters and in translating (as a multiplicative factor) the GDP per capita growth forecasts generated by the model into GDP growth forecasts. In contrast, the population series only affects the inflation forecast in one way, which is in the estimation stage. It turns out that the Kalman filter in the model correctly attributes most of the spikes in the GDP per capita growth in the raw data to the model's shock processes, rather than to fitted GDP per capita (Figure 5) and as a result the model's parameter estimates are little affected by the spikes in the population series. This makes the model's inflation and GDP per capita growth forecasts ultimately unaffected by the erratic population series. However when the resulting GDP per capita forecasts are multiplied by population growth to obtain GDP growth forecasts, all of the noise is reintroduced. As a result the volatile population growth series does affect the GDP growth forecast but does not affect the inflation forecast all that much.

The mild forecastability of GDP growth and the continued lack of forecastability of inflation are more supportive of the validity of the model since (provided more active and preemptive monetary policy that is the source of the weak forecastability in the Great Moderation period) the model implies there should be some forecastability of at least one of these variables. That being said, we cannot at this point conclusively say that the model's relative weak ability to forecast of the macroeconomy in the Great Moderation is due to more activist monetary policy. We must first establish that the DSGE model could indeed forecast the macroeconomy better prior to the Great Moderation when monetary policy was less activist. We must also establish that the DSGE model – estimated on both pre and post Great Moderation data – also predicts that its forecast performance should be deteriorating with the changes in the model's estimated coefficients associated with the Great Moderation. And finally we must establish that it is more activist monetary policy and not some other

structural change that causes the decline in macroeconomic forecastability. These are all topics that we return to in section four [to be added].

2.3 *Transitioning to rolling-window estimation*

The next change that we make relates to the fact that we will be considering a longer period of data. Up to this point the DSGE and BVAR models have been estimated over a range in which the start of the sample is fixed – specifically, at 1965 – while the endpoint moves forward in time – specifically, from 1992 to 2006.[8] When we extend our analysis back to the period prior to the Great Moderation we will need to be using data further back in time to estimate both the DSGE and BVAR models. With the measure of hours (from the Labor Productivity and Costs release) that we will be using when we undertake our pre-Moderation period analysis, we can in principal start our estimation in 1948. If we think 20 years is a reasonable quantity of data upon which to estimate our models, this will mean that we can start generating forecasts from the late 1960s, which would give us about 15 years of forecasts prior to the Great Moderations. If we were to begin our sample period in 1948, however, by the time we come to generate forecasts starting in 1992 we would be estimating our models with 44 years of data while in generating our forecasts starting in 2006 we would be estimating our models with 58 years of data. Due to structural change in the model parameters starting the estimation in 1948 and keeping that start date constant leads to poor forecasts over the 1992 to 2006 period. Consequently, we limit the length of the data that we use in estimation to 20 years and estimate the DSGE and BVAR models on rolling windows of data.

This, of course, means that our forecasts over the period 1992 to 2006 will change, as is evident from Figure 6 and Table 3. As can be seen from the figure and table changing the sample period in this way over the Great Moderation period does not result in much change to the performance of the DSGE model's real GDP growth forecasts but it does lead to large changes in the performance of the model's inflation forecasts. The DSGE model now forecasts a worse than the random walk. [We are in the process of trying to understand this.

---

[8] In papers that evaluate forecasting performance of the DSGE models and expanding sample seems to be the norm. Adolfson *et al.* (2007), Lees *et al.* (2007), Edge *et al.* (2010), and Christoffel *et al.* (forthcoming) all seem to use expanding samples and so does Kolosa *et al.* (2012). The only paper that we are aware of in the DSGE model forecasting literature that uses rolling windows is Wolters (2012). This may be because – with the exception of Wolters – all of these papers look at fairly recent forecast performance (e.g., from the late 1990s onwards), while their data samples are not especially long either. Nonetheless given that our analysis considers forecast performance over a long period of time – over which there has been structural change – we have no choice but to use a rolling window.

At present it seems that the shorter sample has more problems with the structural break in inflation that occurs around 1983.]

2.4   *Transitioning to final (2012:Q2) vintage data*

As noted, extending our sample period back to before the Great Moderation prevents us from using real-time data in our analysis, where, more specifically, it is the absence of real-time compensation per hour data that is the constraint in our analysis. We therefore use the 2012:Q2 vintage of data (which we consider to be the final vintage) for all data series in evaluating the forecasting performance of the model over the longer sample.

One immediate implication of using final vintage data to estimate the DSGE and BVAR models is that, in considering forecast accuracy, it is no longer sensible to compare the forecasts generated by the models with their actual realizations as measured by the *first-final* release of the data. When real-time data were being used to estimate the DSGE and BVAR models, comparing forecasts to the first-final estimates of their realizations data was reasonable since the first-final estimates of the forecasts' realizations was in almost all cases from a similar vintage of data as was being used to estimate the model. But clearly, this would no longer be the case if the final vintage of data were being used to estimate the model. As an initial step therefore in transitioning to using final vintage data we examine how forecast performance changes when the forecasts generated by the DSGE and BVAR models *estimated on real-time data* perform when *compared against their final-vintage realizations*.[9] We generate the random walk forecast in the same way that we generate the model forecasts; that is, we continue to generate the random walk forecasts using real-time data but when we evaluate their performance we do so against their realized values as given by the final vintage of data. The relative and absolute forecast performance results for the DSGE model, the BVAR model, and the random walk forecasts are presented in Figure 7. As can be seen, there is not all that much difference between the results in Figure 6 and those in Figure 7. That is, DSGE model forecasts both real GDP growth better than the random walk forecast but forecasts inflation worse.

---

[9] While it does not make much intuitive sense to compare forecasts generated by a model estimated using final vintage data with first-final estimates of realized data, comparing forecasts generated by a model estimated using real-time data with final vintage estimates of realized data does have an intuitive and potentially interesting interpretation. Comparing the forecasts generated by a model estimated on real-time data with final vintage estimates of the realizations of the data asks how well the model forecasts what are currently our best – or even our truest – estimates of GDP growth or inflation in that quarter. By contrast, comparing the forecasts generated by models estimated on real-time data with first-final estimates of realization of the data asks how the model forecasts the Bureau of Economic Analysis' estimates of the data.

Figure 8 shows one small intermediate step in our analysis. Rather than generate the random-walk forecast using real-time data we generate it using final vintage data. The DSGE model and the BVAR model are still, however, estimated using real-time data. What we notice from this is how much the random walk forecast improves for inflation. This reflects the fact that data that has been through one or more annual rebenchmarkings or one or more comprehensive revisions is smoother and has more persistent changes, relative to data that is for the same quarters but are towards the end of the sample period for a real-time vintage and have likely not been through even one annual rebenchmarking. As a result the random walk forecast performs a lot better in final vintage data.

The last step in the transition is to using final vintage data is to estimate the models that are then used to generate the forecasts using this data. Of course, despite the fact that a full time series of data is available, the models are all estimated only up to the quarter before the forecast period begins. As was the case above the forecasts generated by the models estimated on final vintage data are compared to their realized values given in the same final vintage of data. The relative and absolute forecast performance results from the DSGE model, the BVAR model, and the random walk forecasts are presented in Figure 9 and Table 4. The results of Figure 9 – in terms of relative forecast performance – are not vastly different to those of Figure 8. In short, the biggest difference between going between real-time and final-vintage data is the behavior of the random walk forecast, which improves considerably.

There is one additional change that we need to make in light of the next step in this analysis of considering an earlier sample period as well as a longer sample period. And this is to change a particular data series that we use to measure one of the variables – specifically, hours – in the model. As was noted in footnote 6, Edge and Gürkaynak (2010) used a measure of hours taken from the Employment Situation Survey (ESS) in their analysis, in preference to the measures for hours used by Smets and Wouters (2007) and taken from the Labor Productivity and Costs (LPC) release. This was for the simple reason that real-time vintages for the ESS hours series were readily available (extending back to 1970) whereas for the LPC's measure of hours real-time vintages are not available (real-time data for this series in the Alfred database are only available starting in 2011). The problem with the ESS's measure of hours, however, is that while real-time vintages of it are available extending quite far back, the series only begins at the start of 1964. For the analysis conducted in Edge and Gürkaynak – which focused on forecast performance over the 1992 to 2009 period – this did

not pose any constraints since the ESS's measure of hours had close to thirty years of data prior to when the forecasts generated in Edge and Gürkaynak began. For considering the Pre-moderation period, however, the EES's measure of hours beginning in 1964 is more of a constraint and it is for this reason that for the analysis that considers forecast performance in the Pre-moderation period we switch to using the LPC's measure of hours, which extends back to the late 1940s. Figure *** and Table *** show how our results about forecast performance over the period 1992 to 2006 are influenced by changes in the hours data. [To be completed. I think Figure 9 and Table 4 already has this change so we need to split out this step in the next draft.]

2.5 *Summary of the Great Moderation results on final (2012:Q2) vintage data*

The results reported in Figure 9 and Table 4 are our final results for our forecast comparison analysis over the Great Moderation period and it is these results that we will be comparing with those generated over the Pre-moderation period. The key features of these results are that for real GDP growth, the DSGE model forecasts competitively relative to the BVAR and random walk forecast but for inflation, the DSGE model forecasts competitively relative to the BVAR but not relative to the random walk. Interestingly, the worsening of the DSGE model's forecasting performance relative to the random walk, which happens when current (that is, 2012:Q2) data is used in the analysis, reflects a notable improvement in the performance in the random walk forecast that arises when final vintage data is used. This reflects the fact that inflation data that has been through one revision round at least once (and likely many more times) is much smoother than data that has not. The worsening of the DSGE model's forecasting performance relative to the random walk also reflects our switch to rolling window estimation.

3. Results in the Pre-moderation period

We now turn to consider forecast performance in the Pre-moderation sample; starting with forecast performance over the period 1970 to 1984. As was the case with the results for 1992 to 2006 reported in Figure 9 and Table 4, the DSGE and BVAR model used to generate results over the period 1970 to 1984 are estimated using rolling windows of data. Thus for the forecasts generated starting in 1970 the estimation sample of the DSGE and BVAR model is 1950 to 1969, while for the forecasts generated starting in 1984 the estimation sample is 1964 to 1983. In this respect the result generated for the Pre-moderation period are done so in

exactly the analogous way to how they were generated for the Great Moderation period, as shown in Figure 9 and Table 4. Figure 10 and Table 5, show the forecasting ability of the DSGE model and its competitors before the Great Moderation. While the DSGE model's forecasting performance relative to the random walk forecast is a bit better in the Pre-moderation period (relative to the Great Moderation period), the most notable change in moving to this period is the model's much higher absolute ability to forecast in the Pre-moderation period. In a sense, this result is – *ex post* – not especially surprising given previous authors' findings that macroeconomic aggregates were more persistent prior to the Great Moderation and were thereby were "easier" to forecast. That being said, finding that the DSGE model's forecasting ability to be notably better in an absolute sense over this period is re-assuring. Indeed, were we to find that the DSGE model's absolute forecast performance were as poor over the pre-Moderation period as over the Great Moderation period we might be inclined to doubt its connection to the data.

In terms of relative RMSE, the random walk and the DSGE model have about the same average forecast error in one quarter ahead forecasts, with the random walk outperforming as horizon increases. Unlike the Great Moderation samples, however, this time the comparison is between two good forecasts.

In addition to examining forecast performance over the two 15-year intervals of 1970 to 1984 and then 1992 to 2006, we also document forecast performance over rolling [15-year] windows between these two intervals, so as to understand how forecast performance evolved. This entails us estimating 148 versions of the DSGE model – each on a different sample period – and then generating 148 sets of forecasts. We then evaluate the performance of these forecasts – relative to their realized observations – in 15-year or 60-quarter continuous blocks. These results are shown in Figures 8 and 9 with the date for each observation corresponding to the last date of the 15-year/60-quarter block. [Currently the chart shows 20-year/80-quarter blocks.] Figure 8 shows the rolling forecasting ability of the DSGE model relative to the random walk forecast – as measured by relative RMSEs – and Figure 9 shows the corresponding rolling R^2 values of forecasting regressions for the DSGE model and random walk forecasts. These charts show that the forecastability results presented for 1970 to 1984 and 1992 to 2006 are not due to temporary spikes in the series in the two subsamples, and there are secular differences between the Pre-moderation and Great Moderation samples in terms of forecastability.

Although we have emphasized that it is not really appropriate to make comparison between forecast performance results based on forecasts generated on final vintage data with forecast performance results based on real-time forecasts, we think it is somewhat interesting to mention some of the results of previous authors using real-time forecasts, so as to at least allow some qualitative comparisons can be made. Romer and Romer (2000) – working with real-time forecasts – find the R-squareds of forecast efficiency regressions based on 4-quarter ahead DRI and SPF inflation forecasts and estimated over the period 1970 to 1991 or 1968 to 1991 to be on the order of 0.23 or 0.16. [10] We can compare these qualitatively with the R-squareds that we calculated over the 20-year period shown in Figure 9 ending in 1991:Q4 and here we find that the R-squareds implied by the DSGE model are greater than calculated by Romer and Romer for the DRI and SPF forecasts. Again, we would emphasize that this is not an "apple to apples" comparison – not only is there a real-time/final vintage discrepancy but the intervals over which the forecast evaluation measures are being compared vary slightly. But the comparison is nonetheless indicative of the DSGE model being able to generate competitive forecasts in a relative sense (in addition to being able to generate good forecasts in an absolute sense).

4.   Data forecastability and model forecast capacity [to be completed]

We now ask whether good *absolute* forecast performance is a feature of the data that the DSGE model can replicate if it is estimated on data that is forecastable and likewise for bad *absolute* forecast performance if the data is unforecastable. To do this we simulate realized data from the Great Moderation and pre-Moderation versions of the DSGE model estimated in sections two and three and evaluate how well straight projections of real GDP growth and inflation from these models forecast the simulated data. Among the estimated parameters for each version of the model are parameters that govern the distribution of the model's shock processes and it is from these parameterized distributions of the shock processes that we draw the shocks that are then used to generated the simulated data.

We examine what the DSGE model says about its own capacity to forecast should be given the data upon which the model is estimated starting with the Great Moderation period. In particular, for each of the (sixty) DSGE models that we estimated in section two for the

---

[10] We discuss the DRI and SPF forecasts here for the main reason that the sample over which they were evaluated can be roughly compared to one of the ranges shown in Figure 9. The Blue Chip forecast is considered over a much shorter sample while the Greenbook forecast is considered over a longer sample.

Great Moderation period, we draw a sequence of shocks for the six quarters following the last period of the sample and obtain a pure model-simulated realized path of data. With these pure model-simulated realized paths of real GDP growth and inflation and with the forecasts corresponding to each of these realized paths, we can estimate a forecast efficiency (or Mincer-Zarnowitz) regression and from that obtain estimated values of the slope and constant coefficients as well as the regression's R-squared. We can also calculate an estimate of the RMSE associated with these forecasts and simulated realized paths of data. We can do this exercise 1000 times (or more) and obtain distributions of the coefficients from the forecast efficiency regressions, a distribution of the regression's R-squared, and a distribution of the RMSE (and relative RMSE). We can then ask how the actual estimated coefficients from the forecast efficiency regression, the actual R-squared, and the actual RMSE estimated over the Great Moderation period compare with the distributions of these variables calculated based on the pure model-simulated data (where the models being used to simulate the data were estimated on Great Moderation data).

We can do the same exercise for the (sixty) DSGE model that we estimated in section three for the pre-Moderation period. We can then ask how the actual estimated coefficients from the forecast efficiency regression, the actual R-squared, and the actual RMSE estimated over the pre-Moderation period compare with the distributions of these variables calculated based on the pure model-simulated data (where the models being used to simulate the data were estimated on pre-Moderation data).

The basic finding of this analysis is … [to be completed]

5.   Possible extensions: Understanding diminishing model forecast capacity

If the results of section four find that the DSGE model's capacity to forecast diminished with the onset of the Great Moderations we are well positioned with our DSGE model to examine why this is the case. One possible explanation that we have mentioned is that monetary policy, by becoming more active since the 1980s, may have served to counteract a significant portion any forecastable changes in monetary policy's objective variables and as such left most of the fluctuations in these variables to be due to unforecastable shocks. However, this is not the only reason. Other structural parameters in the model might have also changed so as to imply that the effect of any shock on real GDP growth and inflation is less persistent and thus less forecastable. Additionally, the relative importance of various shock processes

in the economy might have changed such that shocks that have persistent and thus forecastable consequences for key macroeconomic aggregates might have over time come to account for a smaller share of disturbances to the macroeconomy, while shocks that have less persistent and thus unforecastable consequences might have come to account for a larger share.

Pursuing the question of why forecast capacity has diminished since the Great Moderation is an issue we leave future work. Properly addressing this question would require us to undertake a thorough review of how all of the estimated structural (including policy) parameters in the model evolved over time and would also requires us to perform a large number of simulations with selected sets of parameters re-calibrated to different values so as to see what, given these parameters, the model predicts about its own capacity to forecast. (These simulations would be similar to those performed in section four.)

For the purposes of this paper we do not take a view on the role of monetary policy in the reduction in macroeconomic forecastability – a question that is closely related the good luck versus good policy debate associated with the Great Moderation.[11] All we note here is that the policy rule does become more aggressive to inflation during the Great Moderation, which would be consistent with the explanation of the economy becoming less forecastable as a result of more activist monetary policy. For example, a quick look in Figure 10 at the rolling estimates of some of the Taylor rule coefficients suggests some interesting broad trends, which could be consistent with more activist monetary policy having reduced the forecastability of key macroeconomic aggregates over time. In the top panel of Figure 10 we see some increase over time in the coefficient on inflation (albeit not an especially large increase), while in the lower panel we see a decline in the coefficient on the output gap (although again not notable) starting in the early 1980s.

It is worth noting that what one finds regarding the reasons for the decline in forecastability of the macroeconomy has implications for the usefulness of macroeconomic forecasting in the monetary policy process. If it is the case that reduced forecastability of the macroeconomy is because policy has become more activist then the forecasting and policy process are operating exactly as they are intended to. That is, the forecasting process is highlighting emerging developments that might conflict with stable levels of activity and inflation while the policy process is responding preemptively to counteract these

---

[11] D'Agostino, Giannone, and Surico (2006) emphasize that any theory aimed at explaining the Great Moderation must also be capable of also accounting for the historical decline in the ability to predict inflation and real activity.

developments. In this case, forecasting has a critical role in the monetary policy process and central banks should be generating heavily resourced and deliberated forecasts. If, however, it is the case the structure of the economy has changed to make the macroeconomy unforecastable the role for forecasting in the monetary policy process is less clear. Indeed, in this case, the need for heavily resourced and deliberated forecasts is less obvious since forecasts that are essentially equal to steady-state growth and steady-state inflation would likely represent equally good forecasts.

6.   Conclusions

[To be completed.]

Appendix A – Forecasting Models Appendix

In this appendix we briefly describe the models that are used to generate the DSGE model forecasts and the Bayesian VAR model forecasts. The text used in this appendix is the same as that of section two of Edge and Gürkaynak (2010)

A.1 DSGE Model

The DSGE model that we use in this paper is exactly that of Smets and Wouters (2007) and the description of the model given here quite closely follows the description presented in section one of Smets and Wouters (2007) and section two of Smets and Wouters (2003).

The Smets and Wouters model is an application of a real business cycle model (in the spirit of King, Plosser, and Rebelo, 1988) to an economy with sticky prices and sticky wages. In addition to nominal rigidities, the model also contains a large number of real rigidities-specifically habit formation in consumption, costs of adjustment in capital accumulation, and variable capacity utilization-that ultimately appear to be necessary to capture the empirical persistence of U.S. macroeconomic data.

The model consists of households, firms, and a monetary authority. Households maximize a non-separable utility function with goods and labor effort as its arguments over an infinite life horizon. Consumption enters the utility function relative to a time-varying external habit variable and labor is differentiated by a union. This assumed structure of the labor market enables the household sector to have some monopoly power over wages. This implies a specific wage-setting equation that in turn allows for the inclusion of sticky nominal wages, modeled following Calvo (1983). Capital accumulation is undertaken by households, who then rent capital to economy's firms. In accumulating capital households face adjustment costs, specifically investment adjustment costs. As the rental price of capital changes, the utilization of capital can be adjusted, albeit at an increasing cost.

The firms in the model rent labor and capital from households (in the former case via a union) to produce differentiated goods for which they set prices, with Calvo (1983) price stickiness. These differentiated goods are aggregated into a final good by different (perfectly competitive) firms in the model and it is this good that is used for consumption and accumulating capital.

The Calvo model in both wage and price setting is augmented by the assumption that prices that are not reoptimized are partially indexed to past inflation rates. Prices are

therefore set in reference to current and expected marginal costs but are also determined, via indexation, by the past inflation rate. Marginal costs depend on the wage and the rental rate of capital. Wages are set analogously as a function of current and expected marginal rates of substitution between leisure and consumption and are also determined by the past wage inflation rate due to indexation. The model assumes a variant of Dixit-Stiglitz aggregation in the goods and labor markets following Kimball (1995). This aggregation allows for time-varying demand elasticities, which allows more realistic estimates of price and wage stickiness.

Finally, the model contains seven structural shocks, which is equal to the number of observables used in estimation. The model's observable variables are the log difference of real per capita GDP, real consumption, real investment, real wage, log hours worked, log difference of the GDP deflator, and the federal funds rate. These series and their data sources, are discussed in detail below.

In estimation, the seven observed variables are mapped into 14 model variables by the Kalman filter. Then 36 parameters (17 of which belong to the seven ARMA shock processes in the model) are estimated via Bayesian methods (while 5 parameters are calibrated). It is the combination of the Kalman filter and Bayesian estimation which allows this large (although technically called a medium scale) model to be estimated rather than calibrated. In estimation we use exactly the same priors as Smets and Wouters (2007) as well as using the same data series. Once the model is estimated for a given data vintage, forecasting is done by employing the posterior modes for each parameter. The model can produce forecasts for all model variables but we only use the GDP growth, inflation and interest rate forecasts.

A.2 Bayesian VAR

The Bayesian VAR is, in its essence, a simple forecasting VAR(4). The same seven observable series that are used in the DSGE model estimation are used in the VAR. Having seven variables in a four lag VAR leads to a large number of parameters to be estimated which leads to over-fitting and poor out-of-sample forecast performance problems. The solution is the same as for the DSGE model. Priors are assigned to each parameter (and the priors we use are again those of Smets and Wouters, 2007) and the data are used to update these in the VAR framework. Similar to the DSGE model, the BVAR is estimated at every forecast date using real time data and forecasts are obtained by utilizing the modes of the posterior densities for each parameter.

Both the judgmental forecast and the DSGE model have an advantage over the purely statistical model, the BVAR, in that the people who produce the Greenbook and Blue Chip forecasts obviously know a lot more than seven time series and the DSGE model was built to match the data that is being forecast. That is, judgment also enters the DSGE model in the form of modeling choices. To help the BVAR overcome this handicap it is customary to have a training sample-to estimate the model with some data and use the posteriors as priors in the actual estimation. Following Smets and Wouters (2007) we also "trained" the BVAR with data from 1955-1965 but, in an a sign of how different the early and the late parts of the sample are, found that the performance of the trained and untrained BVAR are comparable. We therefore report results from the untrained BVAR only.

Appendix B – Data Appendix

All of the data series used in this paper are the same as what Edge and Gürkaynak (2010) used with two exceptions; the data series used to measure population growth and the data series used to measure hours, which for subsequent analysis is changed. The data series used to measure hours is also the only data series used by Edge and Gürkaynak that differed from Smets and Wouters (2007) and the motivation for this difference was the availability of real-time data. Note that since real-time data is not a major part of this paper we do not describe the construction of Edge and Gürkaynak's real-time datasets underlying the results of subsections 2.1 and 2.1. For a discussion of this we refer the reader to Appendix A of that paper.

Since much of the analysis used in section two uses the same data series that were used in Edge and Gürkaynak, we begin by documenting these series. All of these series, with the exception of hours, continue to be used in the papers remaining analysis. Four series used in our estimation are taken from the national income and product accounts (NIPA). These accounts are produced by the Bureau of Economic Analysis and are constructed at quarterly frequency. The four series are real GDP (GDPC), the GDP price deflator (GDPDEF), nominal personal consumption expenditures (PCEC), and nominal fixed private investment (FPI). The variable names that we use, except that for real GDP, are also the same as those used by Smets and Wouters. We use a different name for real GDP because whereas Smets and Wouters define real GDP in terms of chained 1996 dollars, in our analysis the chained dollars for which real GDP is defined change with the data's base year. (In fact, the GDP price deflator also changes with the base year, since it is usually set to 100 in the base year.)

Another series used in our estimation is compensation per hour in the nonfarm business sector (PRS85006103), taken from the Bureau of Labor Statistics' quarterly Labor Productivity and Costs (LPC) release. Three additional series used in our estimation are taken from the Employment Situation Summary (ESS), which contains the findings of two surveys: the Household Survey and the Establishment Survey. These three series, which are produced by the Bureau of Labor Statistics and constructed at monthly frequency, are average weekly hours of production and nonsupervisory employees for total private industries (AWHNONAG), civilian employment (CE16OV), and civilian noninstitutional population (LNSINDEX). Since our model is quarterly, we calculate simple quarterly averages of the monthly data. The first of these series is from the Establishment Survey and

the other two are from the Household Survey. As discussed in section two civilian employment and civilian noninstitutional population as measured by the Household Survey are done so on a best levels basis and as such display some sharp jumps and plunges when new data regarding their levels become available. As also discussed in section two while from subsection 2.2 onwards civilian employment is normalized by civilian noninstitutional population as measured by the Household Survey, all other variables in the paper (such as, real GDP, real consumption, and real investment) are normalized by the smoothed civilian noninstitutional population (LNSINDEX.SM).[12] This is a series that is taken from one of the databases of the FRB/US model.

The final series in our model, the federal funds rate, is obtained from the Federal Reserve Board's H.15 release. This data source is published every business day, and our quarterly series is simply the averages of these daily data.

The (transformed) data series, therefore, used for much of the analysis used in section 2 are as follows:

output = $\ln(\text{GDPC/LNSINDEX.SM}) \times 100$
consumption = $\ln[(\text{PCEC/GDPDEF})/\text{LNSINDEX.SM}] \times 100$
investment = $\ln[(\text{FPI/GDPDEF})/\text{LNSINDEX.SM}] \times 100$
hours = $\ln[(\text{AWHNONAG} \times \text{CE16OV}/100)/\text{LNSINDEX} \times 100$
inflation = $\ln(\text{GDPDEF/GDPDEF-1}) \times 100$
real wage = $\ln(\text{PRS85006103/GDPDEF}) \times 100$
interest rate = federal funds rate $\div$ 4.

As noted the data series used to measure hours in Edge and Gürkaynak (2010) was the only one to differ from the series used by Smets and Wouters (2007). Whereas Edge and Gürkaynak used average weekly hours of production and nonsupervisory employees for total private industries (AWHNONAG) as measured by the Household Survey of the Employment Situation Survey (ESS), Smets and Wouters used nonfarm business sector average weekly hours (PRS85006023) as measured by the Labor Productivity and Costs (LPC) release. Edge and Gürkaynak used the ESS's measure of hours (AWHNONAG) for the simple reason that real-time vintages for this series were readily available (extending back to 1970) whereas for

---

[12] This description of population normalization does not apply to the Edge and Gürkaynak results described in subsection 2.1. There all of the per capita series are calibrated using the unsmoothed version of civilian noninstitutional population.

the LPC's measure of hours (PRS85006023) real-time vintages are not available (real-time data for this series start only in 2011).

However, while real-time vintages of the ESS's measure of hours (AWHNONAG) are available extending quite far back, the series only begins at the start of 1964. For the analysis conducted in Edge and Gürkaynak – which focused on forecast performance over the 1992 to 2009 period – this did not pose any constraints since the ESS's measure of hours had more than thirty years of data prior to when the forecasts generated in Edge and Gürkaynak were to begin. For considering the pre-Great Moderation period, however, the EES's measure of hours beginning in 1964 is more of a constraint. It is for this reason that we switch in this paper to use the LPC's measure of hours (PRS85006023), which extends back to the late 1940s, for the analysis that considers forecast performance in the pre-Great Moderation period. Note that this is the series for hours that was used by Smets and Wouters (2007).

References

1. Adolfson, M., M. Andersson, J. Linde, M. Villani, and A. Vredin. 2007. "Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks," *International Journal of Central Banking* 3(4), 111-144.

2. Calvo, G. 1983. "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics* 12(3), 383-398.

3. Clements, M., and D. Hendrey, 2005. "Evaluating a Model by Forecast Performance," *Oxford Bulletin of Economics and Statistics* 67 (Supplement), 931-956.

4. Atkeson, A., and L. Ohanian. 2001."Are Phillips curves useful for forecasting inflation?," Quarterly Review, Federal Reserve Bank of Minneapolis (Winter), 2-11.

5. Christoffel, K., G. Coenen, and A. Warne. 2010. "Forecasting with DSGE Models," in M. Clements and D. Hendry (eds), *Handbook of Forecasting*, Oxford University Press, Oxford, forthcoming.

6. D'Agostino, A. and D. Giannone. 2006. "Comparing alternative predictors based on large-panel factor models," ECB Working Paper No. 680.

7. D'Agostino, A., D. Giannone, and P. Surico. 2006. "(Un)Predictability and Macroeconomic Stability," ECB Working Paper No. 605.

8. Edge, R. and R. Gürkaynak. 2010. "How Useful Are Estimated DSGE Model Forecasts for Central Bankers?," *Brookings Papers on Economic Activity* 41(2), 209-259.

9. Edge, R., M. Kiley, and JP. Laforte. 2007. "Documentation of the Research and Statistics Division's Estimated DSGE Model of the U.S. Economy: 2006 Version," FEDS Working Paper 2007-53. Washington: Board of Governors of the Federal Reserve System.

10. Edge, R., M. Kiley, and JP. Laforte. 2010. "A Comparison of Forecast Performance between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model," *Journal of Applied Econometrics*, forthcoming.

11. Fuhrer, J., G. Olivei, and G. Tootell. 2009. "Empirical Estimates of Changing Inflation Dynamics," FRB Boston Working Paper No. 09-4.

12. Gali, J., F. Smets, and R. Wouters. 2010. "Unemployment in an Estimated New Keynesian Model," Working Paper, CREI.

13. Goodfriend, M. and R. King. 2009. "The Great Inflation Drift," NBER Working Paper No. 14862

14. Gürkaynak, R., E. Swanson and B. Sack. 2005. "Market-Based Measures of Monetary Policy Expectations," *Journal of Business and Economic Statistics* 25(2), 201-12.

15. Gürkaynak, R. and J. Wolfers. 2007. "Macroeconomic Derivatives: An Initial Analysis of Market-Based Macro Forecasts, Uncertainty, and Risk," *NBER International Seminar on Macroeconomics 2005* (2), 11-50.

16. King, R., C. Plosser, and S. Rebelo. 1988. "Production, Growth and Business Cycles I," *Journal of Monetary Economics* 21(2-3),195-232.

17. Kolasa, M., M. Rubaszek, and P. Skrzypczynski. 2012. "Putting the New Keynesian DSGE model to the Real-time Forecasting Test." *Journal of Money, Credit, and Banking* 44 (7), 1301-1323.

18. Lees K., T. Matheson, and C. Smith. 2007. "Open economy DSGE-VAR forecasting and policy analysis - head to head with the RBNZ published forecasts," *International Journal of Forecasting* 27, 512-28.

19. Sims, C. 2002. "The Role of Models and Probabilities in the Monetary Policy Process," *Brookings Papers on Economic Activity* 2002(2), 1-40.

20. Smets, F. and R. Wouters. 2003. "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area," *Journal of European Economic Association*, 1(5), 1123-1175.

21. Smets, F. and R. Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review* 97(3), 586-607.

22. Stock, J. and M. Watson. 2007. "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money Credit and Banking* 39(1), 3-33.

23. Taylor, J. and V. Wieland. 2009. "Surprising Comparative Properties of Monetary Models: Results from a New Data Base," NBER Working Paper 14849.

24. Trehan, B. 2010. "Survey Measures of Expected Inflation and the Inflation Process," FRBSF Working Paper Series No. 2009-10

25. Tulip, P. 2009."Has the Economy Become More Predictable? Changes in Greenbook Forecast Accuracy," *Journal of Money, Credit and Banking* 41(6), 1217-1231.

26. Wieland, W. and M Wolters. 2010. `The Diversity of Forecasts from Macroeconomic Models of the U.S. Economy," Working Paper, Goethe University.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.372**	0.444*	0.442	0.455	0.470	0.612**
	(0.172)	(0.250)	(0.306)	(0.278)	(0.291)	(0.262)
Constant	0.419**	0.366	0.376	0.355	0.340	0.205
	(0.202)	(0.290)	(0.341)	(0.313)	(0.314)	(0.283)
R-squared	0.080	0.073	0.057	0.062	0.065	0.074
Observations	112	112	112	112	112	112

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.0279	-0.0521	0.0919	0.0708	0.0995	0.0323
	(0.119)	(0.126)	(0.139)	(0.130)	(0.141)	(0.206)
Constant	0.796***	0.893***	0.741***	0.759***	0.721***	0.793***
	(0.142)	(0.176)	(0.186)	(0.177)	(0.193)	(0.257)
R-squared	0.001	0.003	0.010	0.006	0.010	0.001
Observations	112	112	112	112	112	112

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.245**	0.0323	0.0578	0.0133	-0.171	-0.0532
	(0.115)	(0.125)	(0.114)	(0.130)	(0.144)	(0.181)
Constant	0.624***	0.811***	0.794***	0.824***	0.966***	0.869***
	(0.116)	(0.147)	(0.124)	(0.0838)	(0.120)	(0.128)
R-square	0.062	0.001	0.003	0.000	0.030	0.003
Observations	112	112	112	112	112	112

Table1a: Absolute Forecasting Performance of Competing Models for Real GDP Growth: 1992 to 2006. Forecasts are generated using real-time data, unsmoothed population growth, and an expanding sample that starts in 1965. Forecasts are compared against the first-final release of the series in question. The figure associated with this table is Figure 1.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.558***	0.312	0.301	0.377	0.235	0.209
	(0.113)	(0.196)	(0.291)	(0.255)	(0.201)	(0.170)
Constant	0.226***	0.336***	0.337**	0.292**	0.358***	0.370***
	(0.0530)	(0.0932)	(0.134)	(0.123)	(0.106)	(0.112)
R-squared	0.198	0.050	0.037	0.052	0.018	0.012
Observations	112	112	112	112	112	112

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.456***	0.211	0.214*	0.209*	0.0880	-0.0818
	(0.0986)	(0.141)	(0.116)	(0.108)	(0.0919)	(0.126)
Constant	0.245***	0.363***	0.353***	0.348***	0.421***	0.548***
	(0.0571)	(0.0951)	(0.0771)	(0.0783)	(0.0636)	(0.105)
R-squared	0.136	0.030	0.032	0.033	0.006	0.005
Observations	112	112	112	112	112	112

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.456***	0.211	0.214*	0.209*	0.0880	-0.0818
	(0.0986)	(0.141)	(0.116)	(0.108)	(0.0919)	(0.126)
Constant	0.245***	0.363***	0.353***	0.348***	0.421***	0.548***
	(0.0571)	(0.0951)	(0.0771)	(0.0783)	(0.0636)	(0.105)
R-squared	0.174	0.024	0.046	0.099	0.042	0.012
Observations	112	112	112	112	112	112

Table1b: Absolute Forecasting Performance of Competing Models for Inflation: 1992 to 2006. Forecasts are generated using real-time data, unsmoothed population growth, and an expanding sample that starts in 1965. Forecasts are compared against the first-final release of the series in question. The figure associated with this table is Figure 1.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.511***	0.666***	0.656*	0.777***	0.981***	1.082***
	(0.187)	(0.249)	(0.334)	(0.293)	(0.297)	(0.277)
Constant	0.245	0.110	0.146	0.0320	-0.161	-0.227
	(0.227)	(0.294)	(0.368)	(0.324)	(0.306)	(0.280)
R-squared	0.116	0.123	0.094	0.113	0.154	0.148
Observations	112	112	112	112	112	112

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.0343	0.0858	0.179	0.264	0.212	0.348*
	(0.0858)	(0.125)	(0.140)	(0.164)	(0.198)	(0.194)
Constant	0.793***	0.751***	0.652***	0.546**	0.589**	0.434
	(0.103)	(0.152)	(0.191)	(0.233)	(0.273)	(0.267)
R-squared	0.002	0.008	0.034	0.063	0.034	0.073
Observations	112	112	112	112	112	112

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.245**	0.0323	0.0578	0.0133	-0.171	-0.0532
	(0.115)	(0.125)	(0.114)	(0.130)	(0.144)	(0.181)
Constant	0.624***	0.811***	0.794***	0.824***	0.966***	0.869***
	(0.116)	(0.147)	(0.124)	(0.0838)	(0.120)	(0.128)
R-squared	0.062	0.001	0.003	0.000	0.030	0.003
Observations	112	112	112	112	112	112

Table2a: Absolute Forecasting Performance of Competing Models for Real GDP Growth: 1992 to 2006. Forecasts are generated using real-time data, smoothed population growth, and an expanding sample that starts in 1965. Forecasts are compared against the first-final release of the series in question. The figure associated with this table is Figure 4.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.539***	0.258	0.204	0.184	-0.0398	0.0252
	(0.114)	(0.206)	(0.303)	(0.273)	(0.218)	(0.181)
Constant	0.235***	0.361***	0.385***	0.390***	0.506***	0.472***
	(0.0546)	(0.0992)	(0.141)	(0.135)	(0.121)	(0.120)
R-squared	0.182	0.034	0.017	0.012	0.000	0.000
Observations	112	112	112	112	112	112

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.0349	-0.00312	-0.0136	0.0463	-0.00378	-0.126
	(0.0860)	(0.0844)	(0.100)	(0.120)	(0.130)	(0.118)
Constant	0.453***	0.487***	0.498***	0.445***	0.487***	0.605***
	(0.0823)	(0.0885)	(0.108)	(0.131)	(0.140)	(0.138)
R-squared	0.003	0.000	0.000	0.005	0.000	0.028
Observations	112	112	112	112	112	112

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.432***	0.162	0.228	0.334***	0.219*	0.121
	(0.0963)	(0.148)	(0.186)	(0.120)	(0.119)	(0.160)
Constant	0.278***	0.407***	0.378***	0.328***	0.380***	0.429***
	(0.0455)	(0.0774)	(0.0829)	(0.0554)	(0.0585)	(0.0744)
R-squared	0.174	0.024	0.046	0.099	0.042	0.012
Observations	112	112	112	112	112	112

Table 2b: Absolute Forecasting Performance of Competing Models for Inflation: 1992 to 2006. Forecasts are generated using real-time data, smoothed population growth, and an expanding sample that starts in 1965. Forecasts are compared against the first-final release of the series in question. The figure associated with this table is Figure 4.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.391***	0.651***	0.588**	0.559**	0.504*	0.339
	(0.147)	(0.178)	(0.229)	(0.252)	(0.279)	(0.341)
Constant	0.413**	0.195	0.279	0.319	0.369	0.529*
	(0.177)	(0.208)	(0.249)	(0.265)	(0.290)	(0.318)
R-squared	0.092	0.151	0.093	0.067	0.048	0.018
Observations	112	112	112	112	112	112

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.0242	0.0499	0.145	0.207	-0.0119	0.310
	(0.0797)	(0.115)	(0.119)	(0.136)	(0.204)	(0.218)
Constant	0.809***	0.806***	0.748***	0.690***	0.834***	0.566**
	(0.0849)	(0.104)	(0.121)	(0.151)	(0.212)	(0.218)
R-squared	0.001	0.003	0.025	0.039	0.000	0.047
Observations	112	112	112	112	112	112

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.245**	0.0323	0.0578	0.0133	-0.171	-0.0532
	(0.115)	(0.125)	(0.114)	(0.130)	(0.144)	(0.181)
Constant	0.624***	0.811***	0.794***	0.824***	0.966***	0.869***
	(0.116)	(0.147)	(0.124)	(0.0838)	(0.120)	(0.128)
R-squared	0.062	0.001	0.003	0.000	0.030	0.003
Observations	112	112	112	112	112	112

Table3a: Absolute Forecasting Performance of Competing Models for Real GDP Growth: 1992 to 2006. Forecasts are generated using real-time data, smoothed population growth, and a rolling twenty-year sample. Forecasts are compared against the first-final release of the series in question. The figure associated with this table is Figure 6.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.498***	0.0223	-0.290*	-0.373*	-0.565**	-0.563*
	(0.145)	(0.188)	(0.163)	(0.220)	(0.233)	(0.320)
Constant	0.239***	0.472***	0.647***	0.703***	0.823***	0.836***
	(0.0737)	(0.106)	(0.101)	(0.137)	(0.148)	(0.206)
R-squared	0.124	0.000	0.029	0.046	0.101	0.090
Observations	112	112	112	112	112	112

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.0522	0.00954	-0.00336	0.0598	0.0430	-0.0487
	(0.0795)	(0.0728)	(0.0798)	(0.0968)	(0.104)	(0.0820)
Constant	0.440***	0.476***	0.489***	0.439***	0.450***	0.525***
	(0.0758)	(0.0760)	(0.0861)	(0.101)	(0.108)	(0.101)
R-squared	0.008	0.000	0.000	0.010	0.005	0.006
Observations	112	112	112	112	112	112

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.432***	0.162	0.228	0.334***	0.219*	0.121
	(0.0963)	(0.148)	(0.186)	(0.120)	(0.119)	(0.160)
Constant	0.278***	0.407***	0.378***	0.328***	0.380***	0.429***
	(0.0455)	(0.0774)	(0.0829)	(0.0554)	(0.0585)	(0.0744)
R-squared	0.174	0.024	0.046	0.099	0.042	0.012
Observations	112	112	112	112	112	112

Table3b: Absolute Forecasting Performance of Competing Models for Inflation: 1992 to 2006. Forecasts are generated using real-time data, smoothed population growth, and a rolling twenty-year sample. Forecasts are compared against the first-final release of the series in question. The figure associated with this table is Figure 6.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.514**	0.554*	0.330	0.602*	0.175	0.198
	(0.207)	(0.312)	(0.306)	(0.327)	(0.222)	(0.442)
Constant	0.248	0.243	0.480	0.208	0.623***	0.607
	(0.256)	(0.360)	(0.341)	(0.334)	(0.211)	(0.411)
R-squared	0.124	0.095	0.027	0.085	0.006	0.006
Observations	56	56	56	56	56	56

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.0606	0.0232	0.280**	0.227**	0.257***	0.243***
	(0.112)	(0.136)	(0.113)	(0.0906)	(0.0755)	(0.0710)
Constant	0.767***	0.803***	0.610***	0.641***	0.626***	0.643***
	(0.133)	(0.159)	(0.128)	(0.0959)	(0.108)	(0.0974)
R-squared	0.008	0.001	0.103	0.062	0.077	0.068
Observations	56	56	56	56	56	56

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.0965	0.305**	-0.0807	0.164	-0.0140	0.162**
	(0.154)	(0.119)	(0.129)	(0.122)	(0.164)	(0.0704)
Constant	0.738***	0.572***	0.874***	0.657***	0.796***	0.653***
	(0.152)	(0.142)	(0.133)	(0.133)	(0.130)	(0.0851)
R-squared	0.009	0.093	0.006	0.025	0.000	0.025
Observations	56	56	56	56	56	56

Table 4a: Absolute Forecasting Performance of Competing Models for Real GDP Growth: 1992 to 2006. Forecasts are generated using final-vintage data, smoothed population growth, and a twenty-year rolling window sample. Forecasts are compared against final vintage estimate of the series in question. The figure associated with this table is Figure 8.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.384**	0.164*	0.0653	0.0569	-0.0961	-0.362
	(0.148)	(0.0826)	(0.0983)	(0.141)	(0.179)	(0.252)
Constant	0.339***	0.449***	0.501***	0.509***	0.591***	0.761***
	(0.0701)	(0.0546)	(0.0836)	(0.122)	(0.153)	(0.204)
R-squared	0.112	0.025	0.004	0.003	0.007	0.071
Observations	56	56	56	56	56	56

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	-0.173	-0.203	-0.230*	-0.153	-0.257*	-0.294**
	(0.113)	(0.127)	(0.137)	(0.160)	(0.148)	(0.127)
Constant	0.662***	0.693***	0.723***	0.666***	0.755***	0.798***
	(0.105)	(0.119)	(0.139)	(0.168)	(0.153)	(0.142)
R-squared	0.067	0.087	0.105	0.046	0.139	0.174
Observations	56	56	56	56	56	56

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.555***	0.512***	0.529***	0.522***	0.358**	0.496**
	(0.118)	(0.149)	(0.180)	(0.141)	(0.152)	(0.228)
Constant	0.236***	0.263***	0.259***	0.267***	0.351***	0.288**
	(0.0604)	(0.0705)	(0.0932)	(0.0721)	(0.0668)	(0.111)
R-squared	0.291	0.241	0.243	0.232	0.108	0.183
Observations	56	56	56	56	56	56

Table 4b: Absolute Forecasting Performance of Competing Models for Inflation: 1992 to 2006. Forecasts are generated using final-vintage data, smoothed population growth, and a twenty-year rolling window sample. Forecasts are compared against final vintage estimate of the series in question. The figure associated with this table is Figure 8.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.617***	0.657***	0.628***	0.712***	0.795***	0.686***
	(0.105)	(0.135)	(0.142)	(0.158)	(0.175)	(0.176)
Constant	0.197	0.209	0.243	0.148	0.0889	0.152
	(0.155)	(0.183)	(0.193)	(0.201)	(0.218)	(0.236)
R-squared	0.307	0.255	0.173	0.165	0.156	0.092
Observations	61	61	61	61	61	61

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	-0.0292	0.219	-0.137	-0.143	-0.0297	-0.0614
	(0.136)	(0.152)	(0.155)	(0.181)	(0.0819)	(0.0940)
Constant	0.782***	0.550**	0.922***	0.924***	0.849***	0.843***
	(0.247)	(0.256)	(0.185)	(0.189)	(0.163)	(0.164)
R-squared	0.001	0.042	0.015	0.016	0.001	0.003
Observations	61	61	61	61	61	61

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.267**	0.164	0.0889	0.0418	-0.0904	-0.0902
	(0.115)	(0.113)	(0.0767)	(0.0945)	(0.0973)	(0.101)
Constant	0.557***	0.648***	0.725***	0.757***	0.887***	0.848***
	(0.172)	(0.210)	(0.190)	(0.187)	(0.201)	(0.208)
R-squared	0.072	0.028	0.008	0.002	0.009	0.009
Observations	61	61	61	61	61	61

Table 5a : Absolute Forecasting Performance of Competing Models for Real GDP Growth: 1970 to 1984. Forecasts are generated using final-vintage data, smoothed population growth, and a twenty-year rolling window sample. Forecasts are compared against final vintage estimate of the series in question. The figure associated with this table is Figure 9.

DSGE

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.859***	0.807***	0.717***	0.575*	0.339	0.163
	(0.0893)	(0.107)	(0.182)	(0.316)	(0.444)	(0.476)
Constant	0.355***	0.491***	0.645**	0.853*	1.134*	1.323**
	(0.117)	(0.158)	(0.280)	(0.444)	(0.576)	(0.609)
R-squared	0.584	0.414	0.255	0.139	0.040	0.008
Observations	61	61	61	61	61	61

BVAR

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	-0.189	-0.202	-0.240	-0.292	-0.301*	-0.297*
	(0.150)	(0.172)	(0.199)	(0.189)	(0.173)	(0.155)
Constant	1.734***	1.731***	1.748***	1.792***	1.786***	1.765***
	(0.186)	(0.210)	(0.230)	(0.220)	(0.200)	(0.188)
R-squared	0.044	0.049	0.068	0.104	0.112	0.111
Observations	61	61	61	61	61	61

Random Walk

	1Q	2Q	3Q	4Q	5Q	6Q
Slope	0.755***	0.665***	0.615***	0.573***	0.403**	0.311
	(0.0844)	(0.116)	(0.131)	(0.162)	(0.197)	(0.263)
Constant	0.380***	0.508***	0.570**	0.633**	0.887**	1.015*
	(0.124)	(0.176)	(0.225)	(0.292)	(0.387)	(0.509)
R-squared	0.567	0.419	0.338	0.291	0.138	0.079
Observations	61	61	61	61	61	61

Table 5b : Absolute Forecasting Performance of Competing Models for Inflation: 1970 to 1984. Forecasts are generated using final-vintage data, smoothed population growth, and a twenty-year rolling window sample. Forecasts are compared against final vintage estimate of the series in question. The figure associated with this table is Figure 9.
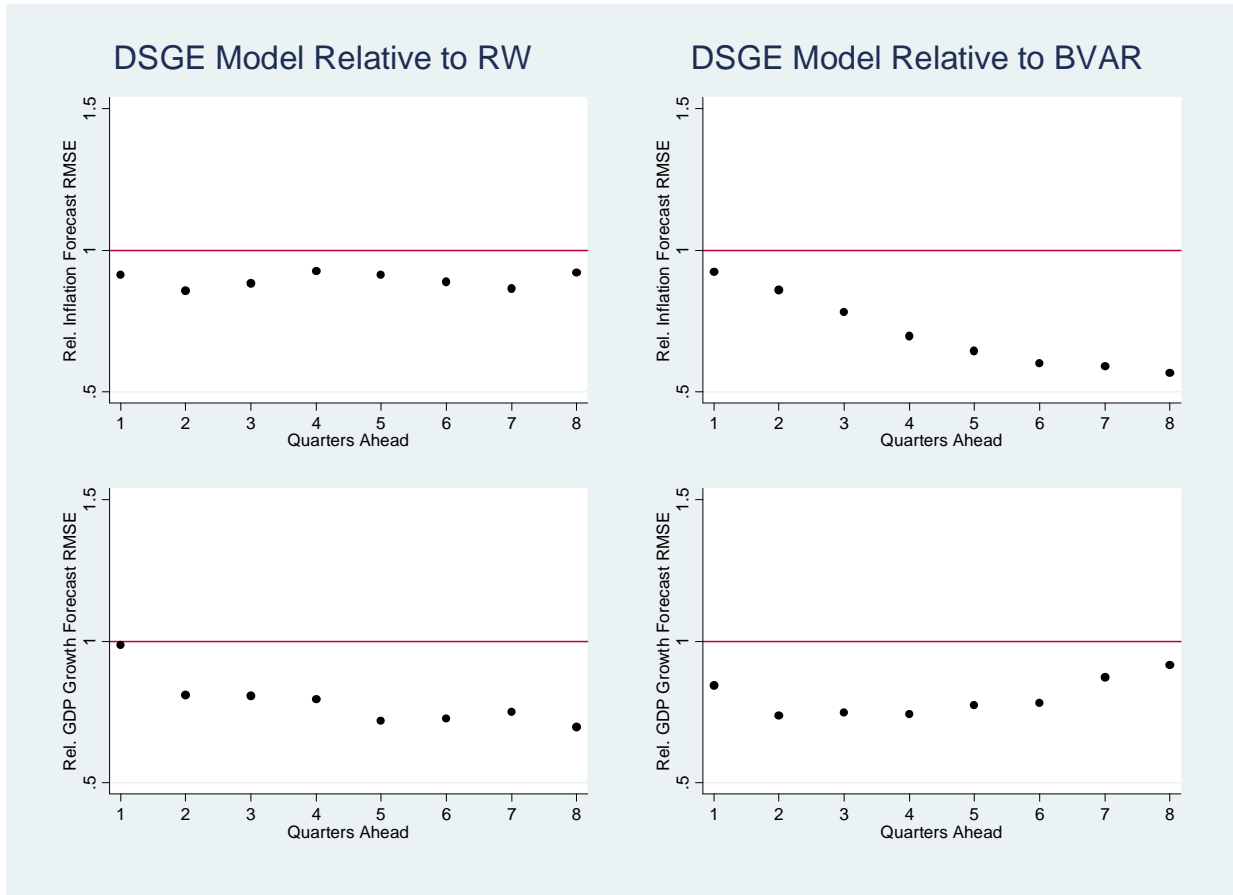
Figure 1: Relative RMSEs for Inflation and Real GDP Growth: 1992 to 2006. Forecasts are generated using real-time data, unsmoothed population growth, and an expanding sample that starts in 1965. Forecasts are compared against the first-final release of the series in question. The tables associated with this table are Tables 1a and 1b.
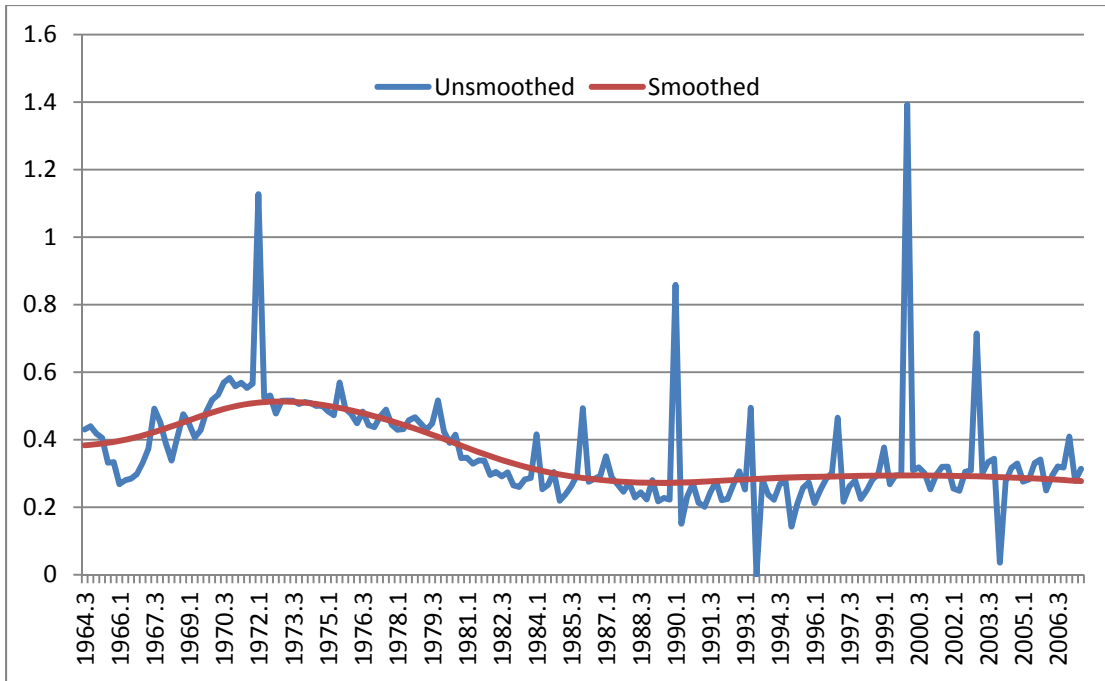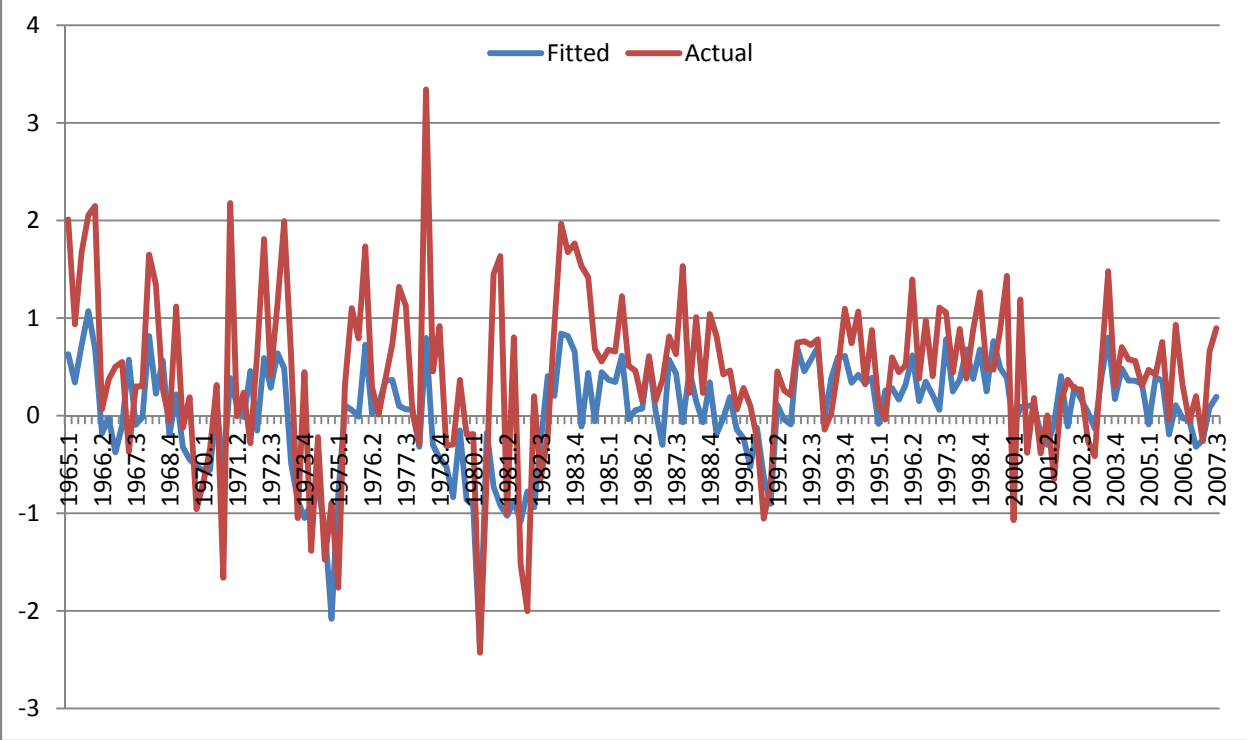
Figure 2: Unsmoothed (Smets-Wouters) and Smoothed (FRB/US) Population Growth
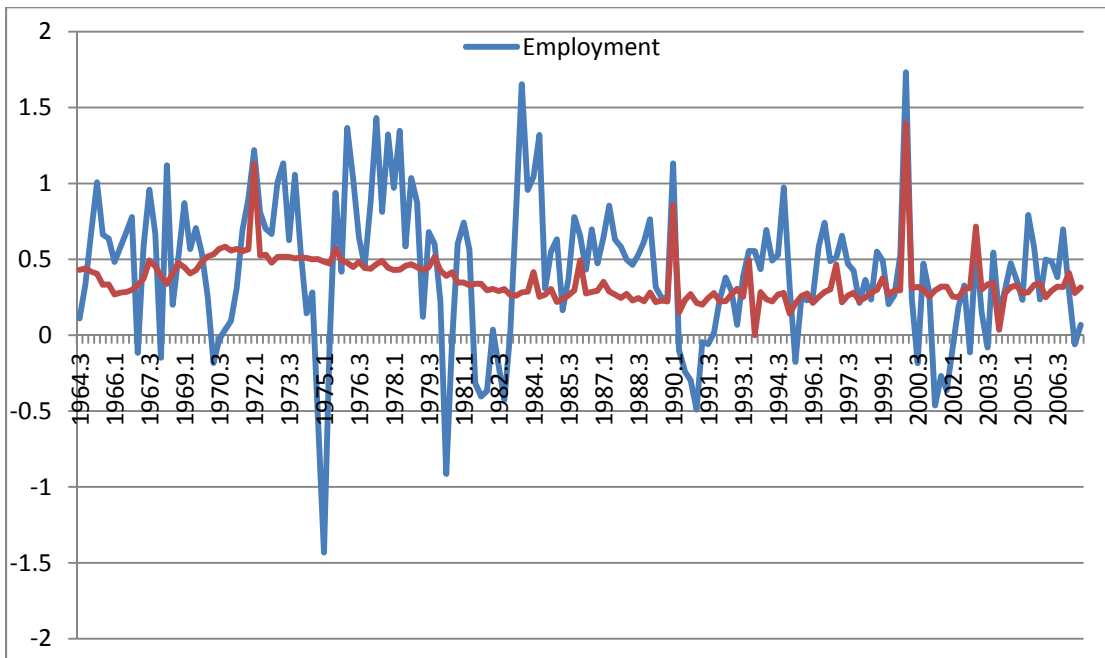
Figure 3: Employment and Population Growth
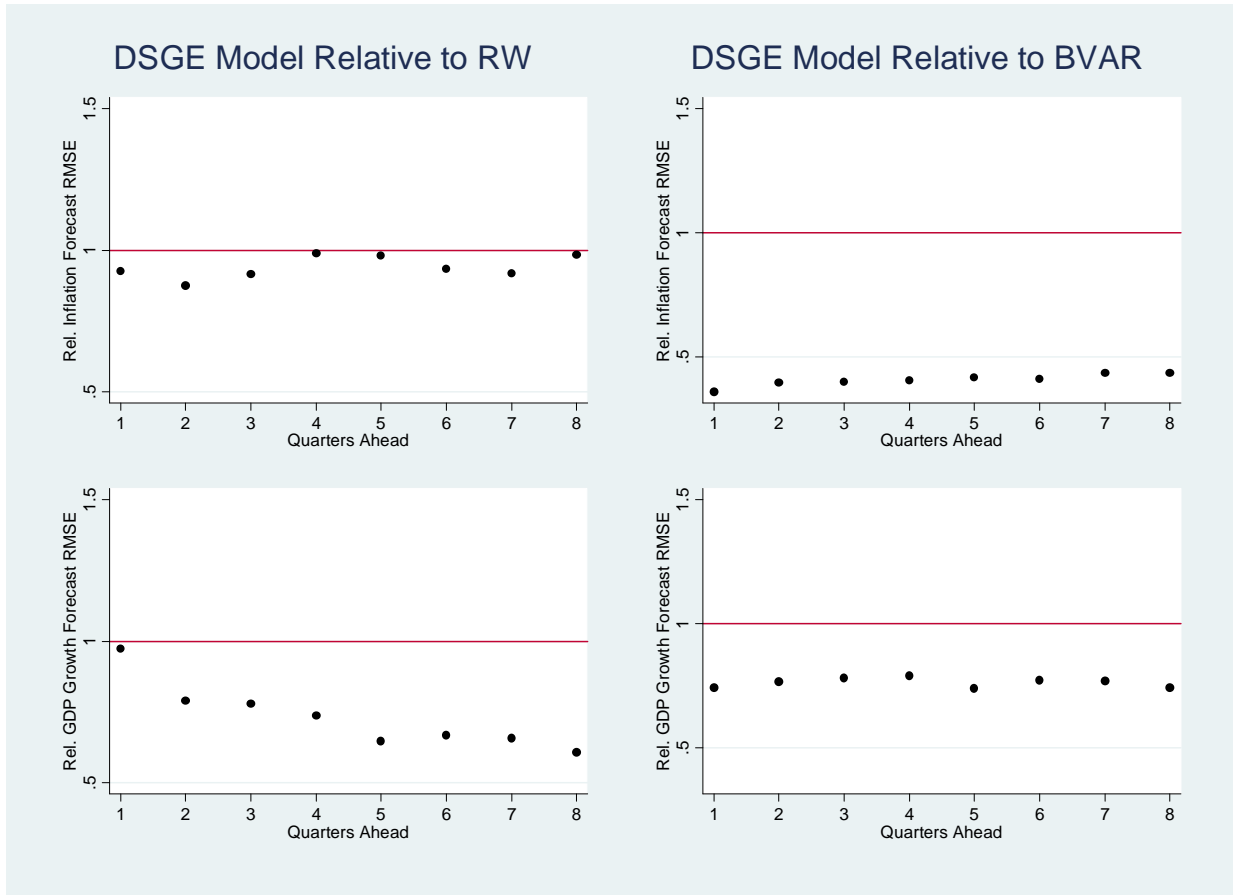
Figure 4: Relative RMSEs for Inflation and Real GDP Growth: 1992 to 2006. Forecasts are generated using real-time data, smoothed population growth, and an expanding sample that starts in 1965. Forecasts are compared against the first-final release of the series in question. The tables associated with this table are Tables 2a and 2b.
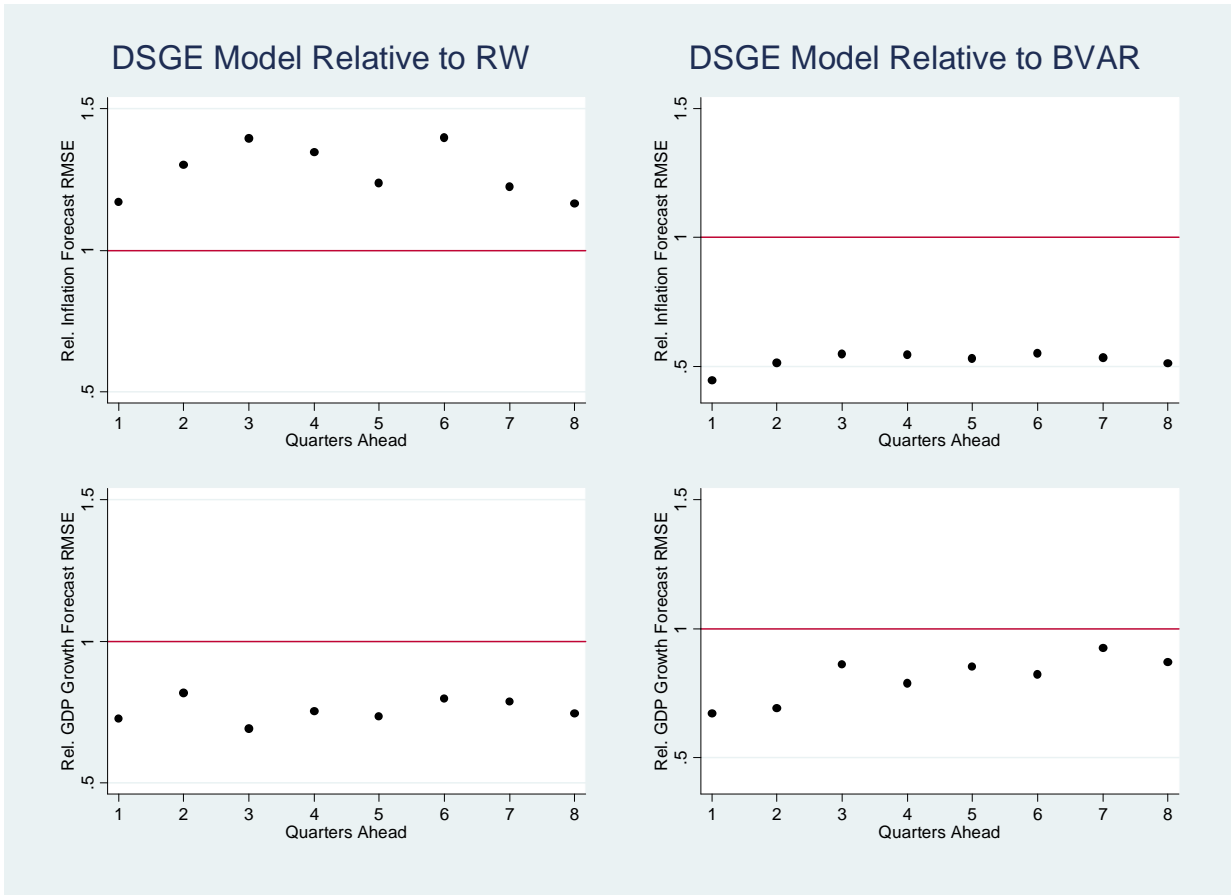
Figure 5: Fitted vs. Raw Real GDP per capita growth
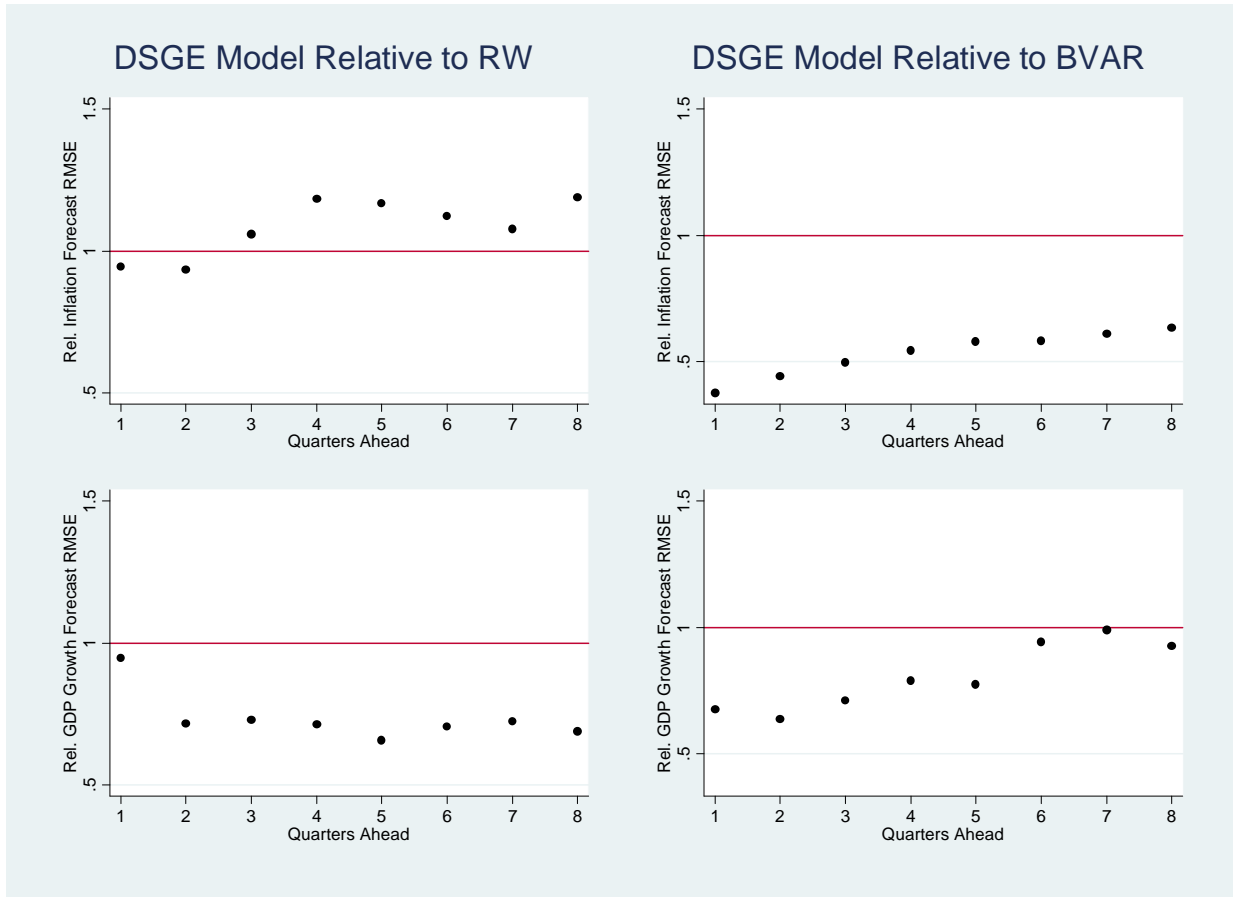
Figure 6: Relative RMSEs for Inflation and Real GDP Growth: 1992 to 2006. Forecasts are generated using real-time data, smoothed population growth, and a twenty-year rolling window sample. Forecasts are compared against the first-final release of the series in question. The tables associated with this table are Tables 3a and 3b.
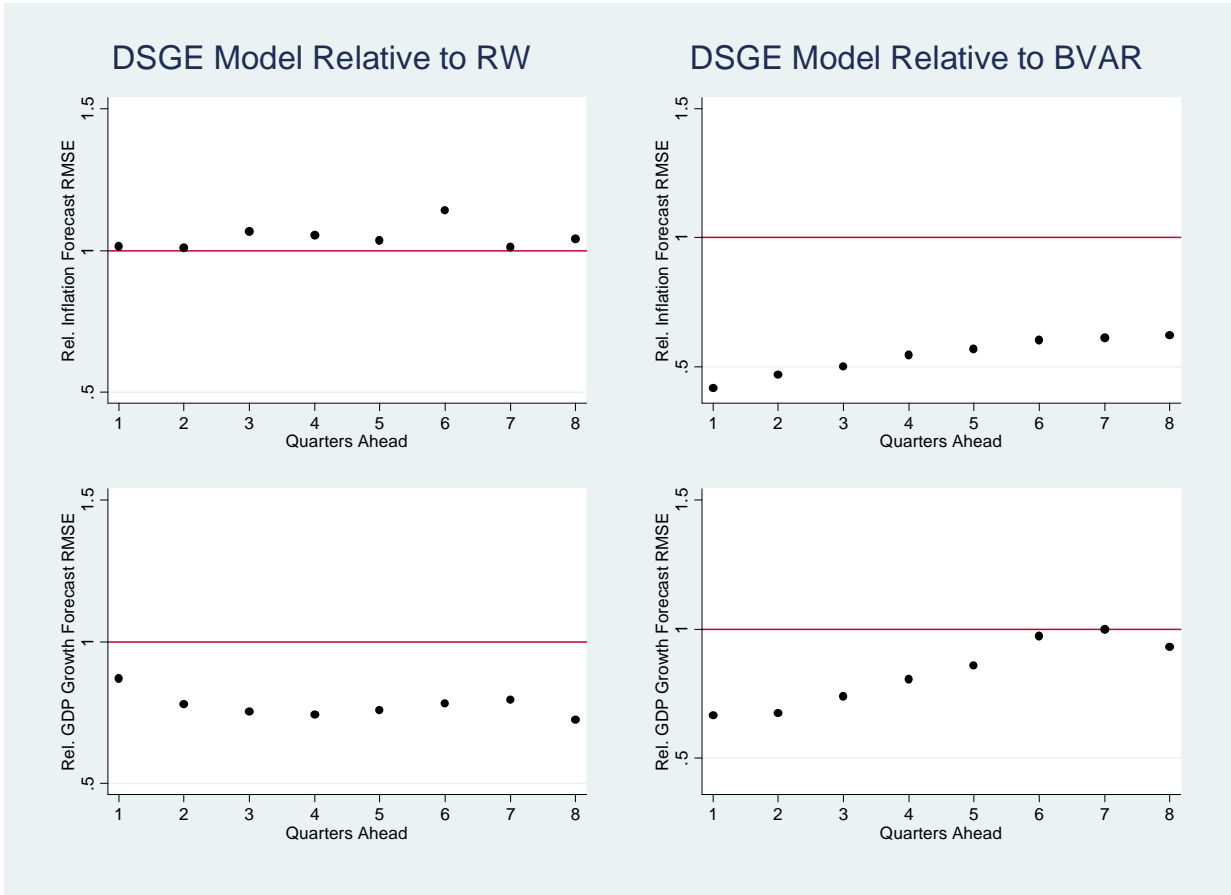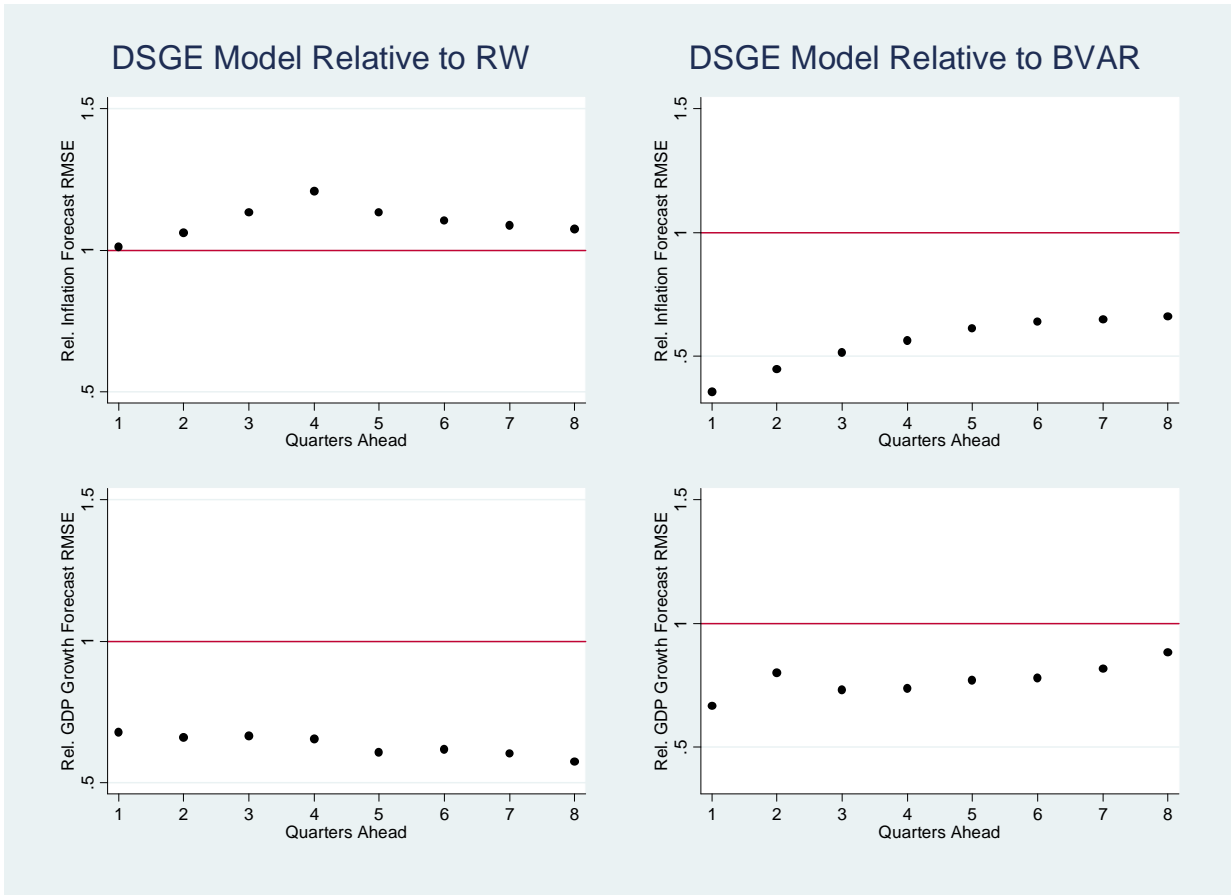
Figure 7: Relative RMSEs for Inflation and Real GDP Growth: 1992 to 2006. Forecasts are generated using real-time data, smoothed population growth, and a twenty-year rolling window sample. The random walk forecast is also generated off real-time and smoothed population data. Forecasts are compared against final vintage estimate of the series in question. The tables associated with this figure are not reported.
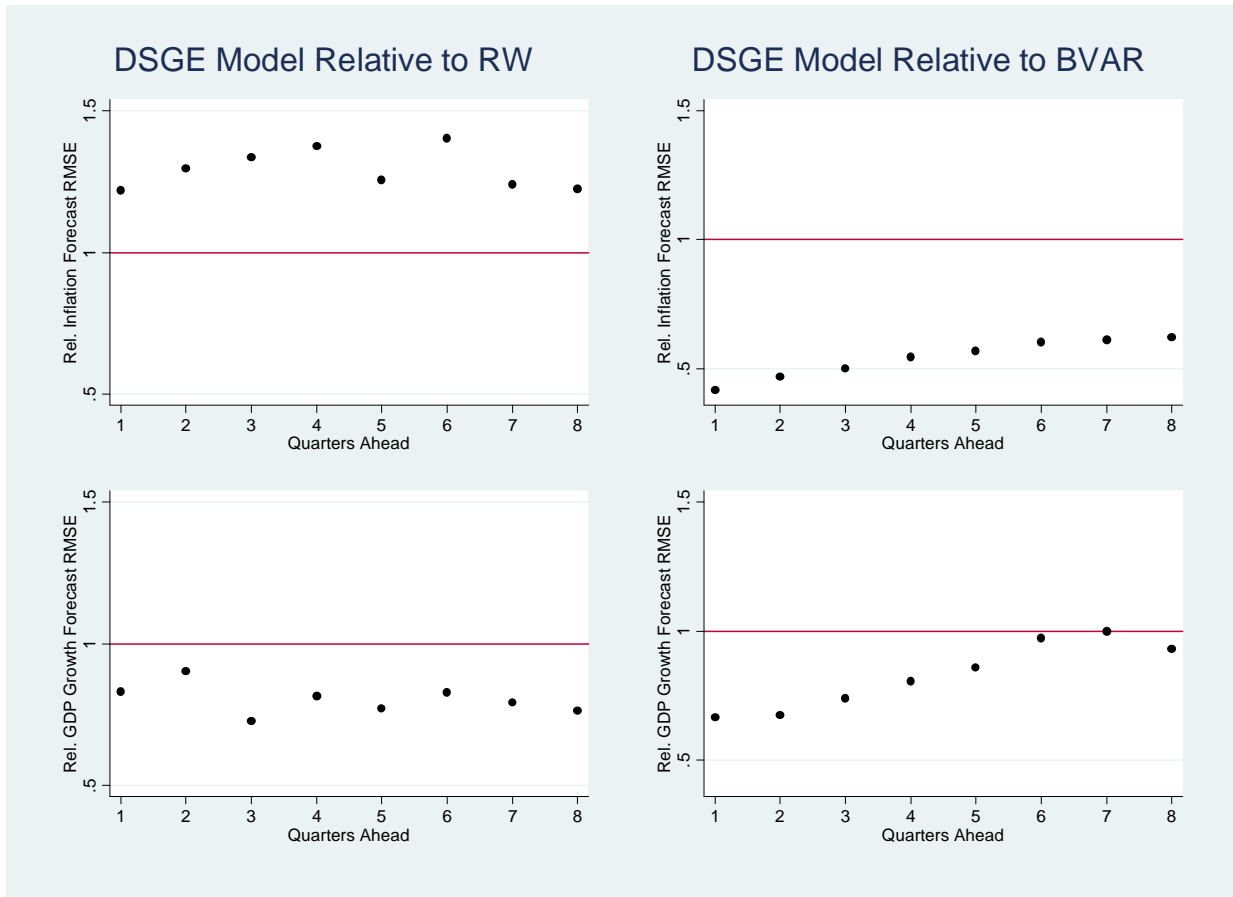
Figure 8: Relative RMSEs for Inflation and Real GDP Growth: 1992 to 2006. The DSGE and BVAR forecasts are generated using real-time data, smoothed population growth, and a twenty-year rolling window sample. The random walk forecast is generated off final vintage data (and smoothed population data). Forecasts are compared against final vintage estimate of the series in question. (The DSGE and BVAR comparison is the same as in Figure 7.) The tables associated with this figure are not reported.

Figure 9: Relative RMSEs for Inflation and Real GDP Growth: 1992 to 2006. Forecasts are generated using final-vintage data, smoothed population growth, and a twenty-year rolling window sample. Forecasts are compared against final vintage estimate of the series in question. The tables associated with this figure are Tables 4a and 4b.
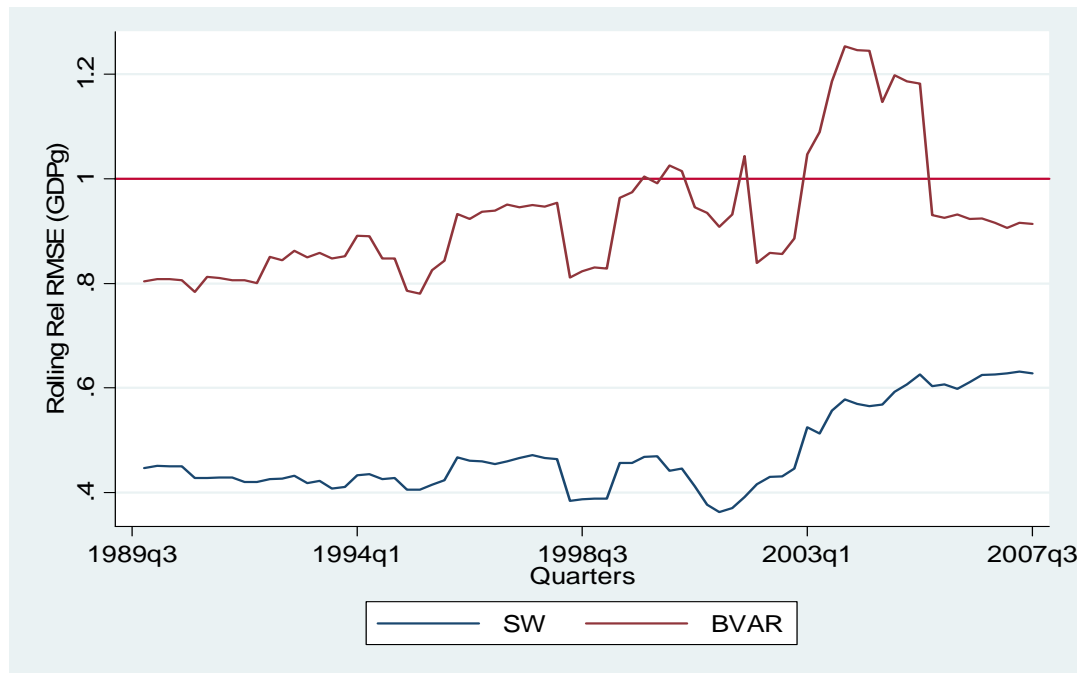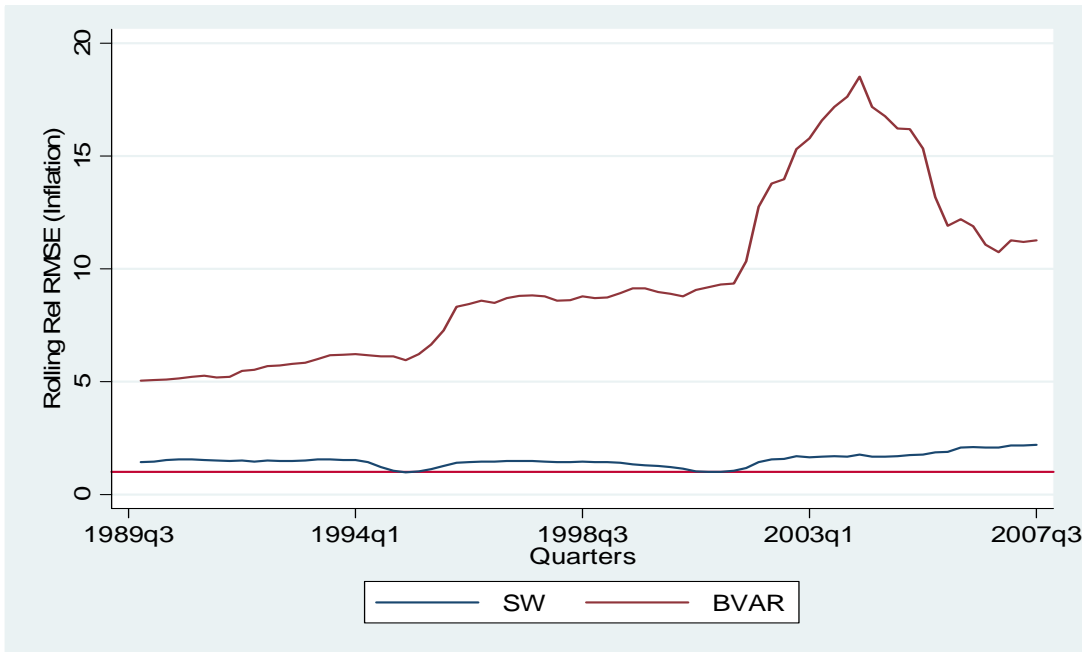
Figure 10: Relative RMSEs for Inflation and Real GDP Growth: 1970 to 1984. Forecasts are generated using final-vintage data, smoothed population growth, and a twenty-year rolling window sample. Forecasts are compared against final vintage estimate of the series in question. The tables associated with this figure are Tables 5a and 5b.
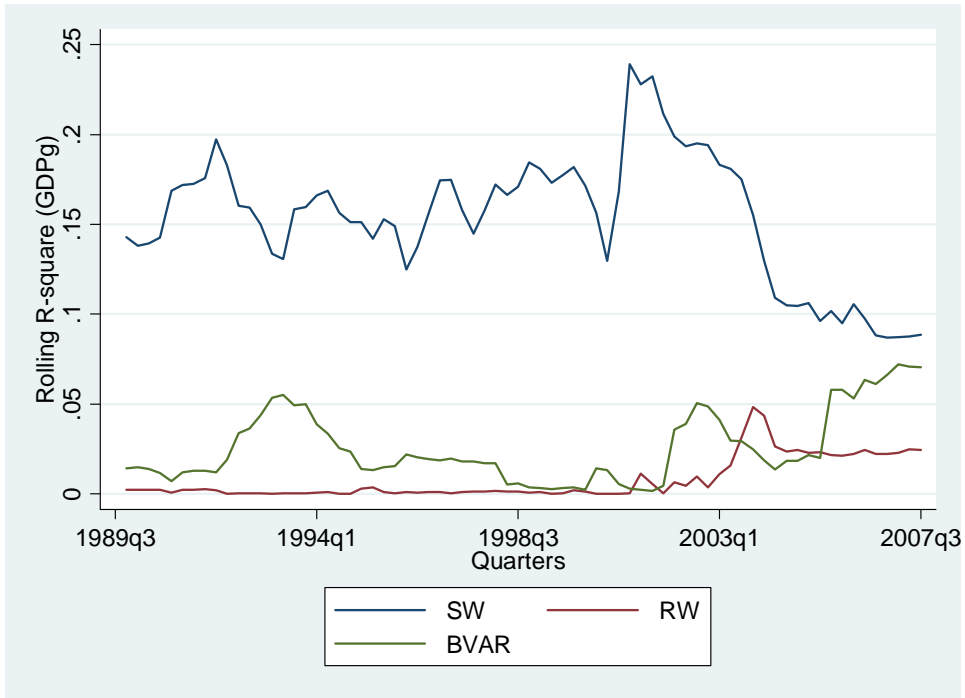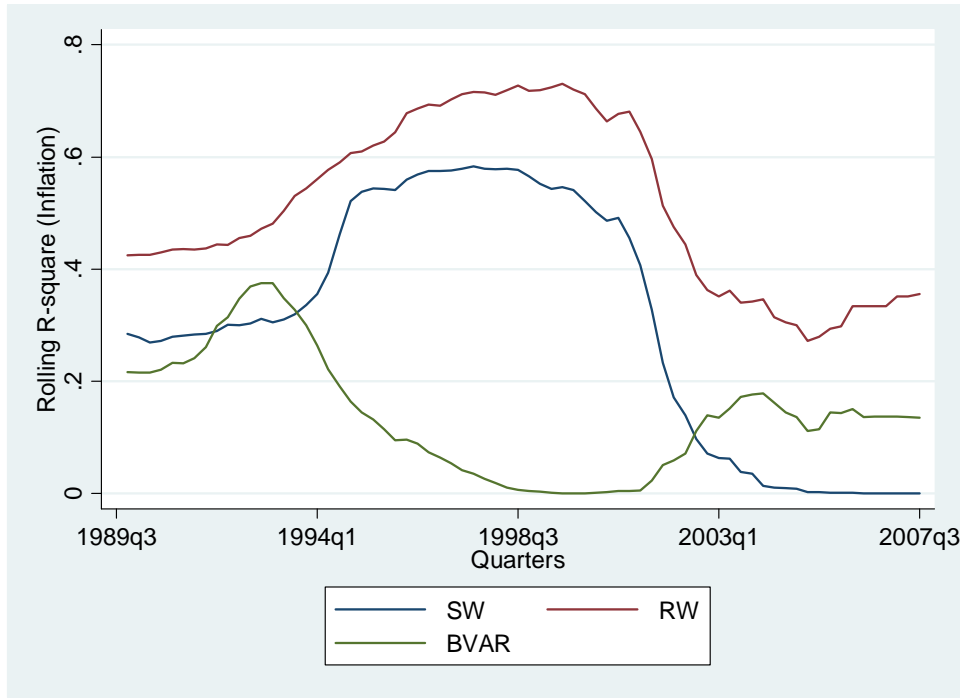
Figure 11: 20-year Rolling 4-quarter Ahead RMSEs for Inflation and Real GDP Growth using Final Vintage and Smoothed Population Growth [Will change to 15-year]

Figure 12: 20-year Rolling 4-quarter Ahead R-square for Inflation and Real GDP Growth using Final Vintage and Smoothed Population Growth [Will change to 15-year]
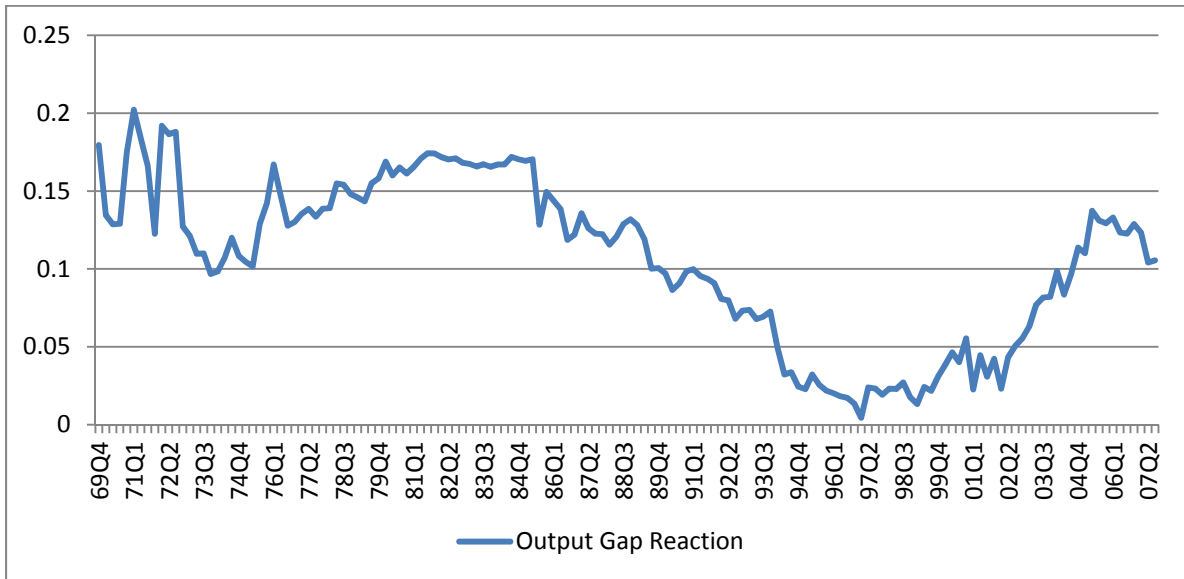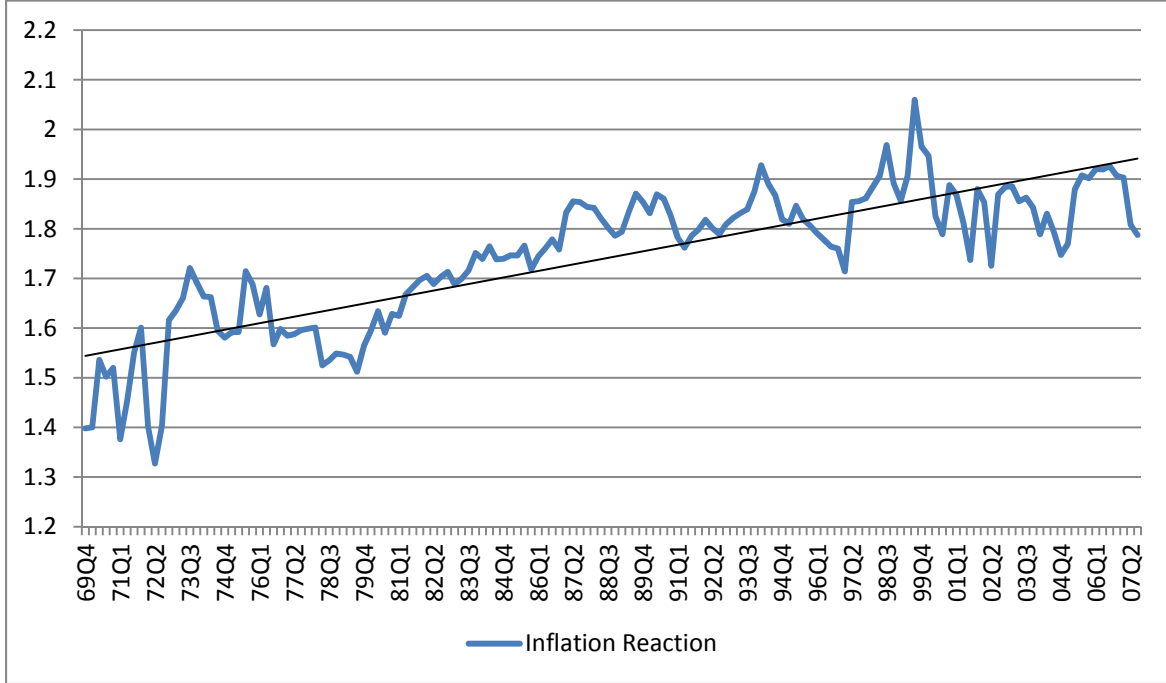
Figure 13: Estimated Rolling Taylor Rule Parameters. The date shown in the chart denotes the end of the sample period over which the model is estimate to obtain the parameter estimate.