A Forecasting Metric for Evaluating DSGE Models for Policy Analysis

Gupta, Abhishek

Johns Hopkins University

30 October 2010

A Forecasting Metric for Evaluating DSGE Models for Policy Analysis

Abhishek Gupta*
Gettysburg College

October 30, 2010

Abstract

This paper evaluates the strengths and weaknesses of dynamic stochastic general equilibrium (DSGE) models from the standpoint of their usefulness in doing monetary policy analysis. The paper isolates features most relevant for monetary policymaking and uses the diagnostic tools of posterior predictive analysis to evaluate these features. The paper provides a diagnosis of the observed flaws in the model with regards to these features that helps in identifying the structural flaws in the model. The paper finds that model misspecification causes certain pairs of structural shocks in the model to be correlated in order to fit the observed data.

JEL Classification: C11, C52, E1, E5
Keywords: Posterior predictive analysis; DSGE; Monetary Policy; Forecast Errors; Model Evaluation.

---

*Correspondence: Abhishek Gupta- 300 North Washington Street, Department of Economics, Gettysburg College, Gettysburg, PA, 17325; e-mail: agupta@gettysburg.edu. I thank Jon Faust for supporting this project. I also thank Jonathan Wright, Michele Julliard, Frank Smets and Rafael Wouters for useful comments and clarifications.

1 Introduction

The last decade has seen a rapid progress in dynamic stochastic general equilibrium (DSGE) models. The pioneering work of Smets and Wouters (2003) caught the attention of academics and policy makers as they showed for the first time that these models can forecast as well as standard atheoretical benchmarks. In recent years, DSGE models have become an important tool at policy institutions for forecasting and policy analysis and a lot of research, both at central banks and universities, has been directed towards developing better and larger models. However at the same time, there remain important questions regarding the strengths and weaknesses of these models to check if these new models being developed are getting better at the intended purpose of monetary policymaking.

The current state of macro models, and in particular, the DSGE models by design were essentially silent about the financial crisis as they have no meaningful financial sectors. It was certainly unusual, but maybe not unwarranted, that this debate about the adequacy of DSGE models rose to the level of a Congressional hearing. Chari (2010) testified:

> The recent crisis has raised, correctly, the question of how best to improve modern macroeconomic theory. I have argued we need more of it. After all, when the AIDS crisis hit, we did not turn over medical research to acupuncturists. In the wake of the oil spill in the Gulf of Mexico, should we stop using mathematical models of oil pressure? Rather than pursuing elusive chimera dreamt up in remote corners of the profession, the best way of using the power in the modelling style of modern macroeconomics is to devote more resources to it. (p.9)

Colander (2010) agreed with Chari but emphasized the need for devoting more resources to interpret rather than develop models:

> ...increase the number of researchers explicitly trained in interpreting and relating models to the real world. This can be done by explicitly provid-

> ing research grants to interpret, rather than develop, models. In a sense, what I am suggesting is an applied science division of the National Science Foundations social science component. This division would fund work on the appropriateness of models being developed for the real world. (p.7)

This paper provides tools to aid in the process that Colander argues for. In particular, we provide new diagnostic tools for evaluating the adequacy of DSGE models for the intended purpose of monetary policymaking.

Much work in the area of evaluating DSGE models has focussed solely on evaluating the overall fit of these models. The most notable among these is the analysis suggested by Del Negro et al. (2007), in which they form a Bayesian comparison of the DSGE model to a general time series model. They show that the degree to which the data shifts the posterior plausibility mass along a continuum from the fully articulated structural model to the general model with no causal interpretability reflects the degree of misspecification in the structural model. They claim,"...the degree of misspecification in this large scale DSGE model is no longer so large as to prevent its use in day-to-day policy analysis, yet is not small enough to be ignored..." Besides this, since the iconic work of Smets and Wouters (2003), a number of papers have evaluated DSGE models in terms of their out-of-sample forecasting performance and have noted that richly specified DSGE models now belong in the forecasting toolbox of central banks.[1] The current set of evaluation tools, in our opinion, are highly insightful in informing us about the overall likelihood of different competing models, but offer little guidance on the evaluation of a particular structural model for its usefulness in day-to-day policy analysis.

We take the view that current DSGE models are misspecified in some known and some unknown dimensions and yet may still offer valuable insights for the policy process. We argue that evaluating a flawed model using an overall fit metric is uninformative about the specific nature of misspecification. Tiao and Xu (1993), Hansen (2005) and

[1] Some of the recent papers looking at out-of-sample forecasting performance of DSGE models are Smets and Wouters (2004), Adolfson et al. (2007), Edge et al. (2009).

Kydland and Prescott (1996) have argued that model evaluation should be based on a question asked of the model and not on a global measure of fit, and that models should be designed and evaluated for a specific question. Therefore, instead of evaluating these models for their overall fit, we evaluate these models for their usefulness in the task of monetary policymaking.

In this paper, we first analyze the intended purpose of monetary policy analysis to highlight particular model features that should be of primary interest. Then we show how to assess the DSGE model with regards to these selected features and develop a diagnostic tool to link the discrepancies between the model and the data with regards to these features to specific structural misspecifications in the model. In particular, we argue that policymaking at central banks can be characterized as interpreting the structural sources of unexpected outcomes in the observed data and accordingly acting upon it.

In the DSGE context this amounts to checking whether the model implied structure of the one-step (where a step is one decision making period) ahead forecast errors is consistent with the observed data on two counts: a) forecast accuracy as reflected by the standard deviations of the one-step ahead forecast errors (FEs) and b) the cross-correlations among the FEs that are crucial to understanding the correct source of the structural shocks causing the economy to deviate from its efficient path. These two requirements of the DSGE models are akin to speed and skill. In any activity if you focus only on developing one of these and not the other, the end result is most likely going to be a disaster. For instance, in the sport of endurance horse racing, if the jockey focusses only on developing speed and not skill, then he might not be prepared to jump over natural obstacles such as creeks and ditches and very likely might fall into one. We, therefore, want to emphasize that this paper is not about a horse race for only achieving a better overall fit and that it differs significantly from the existing literature that focuses on the forecast accuracy of DSGE models.

We illustrate our approach using the Smets and Wouters (2007) DSGE model, henceforth SW. We find, for example, that the one-step ahead FE correlation between output growth and inflation for the realized sample implied by the model has a larger magnitude of negative correlation than what can be accounted for by the DSGE model. We similarly find other significant discrepancies between the structure of FEs estimated on the observed data and that predicted by the model. We trace these discrepancies to certain structural shocks, suggesting the source of the model misspecification. In particular, we find that the consistency of the model with the data requires a non-zero cross-correlation among the smoothed structural shocks of the model: a gross violation of the model assumption. This can be viewed as one of the following two statements: One, the realized sample is collectively treated as an outlier from the standpoint of the DSGE model and it is unlikely for the DSGE model to produce a sample that is similar to the realized sample. Two, the DSGE model is misspecified and the overidentifying restrictions of the model are not consistent with the data.

In this paper we analyze the SW model using the posterior predictive tools described in Faust and Gupta (2010a) and uncover strengths and weaknesses of that model from the standpoint of monetary policymaking. Faust and Gupta (2010b) have used the same tools of posterior predictive analysis to show that DSGE models are highly unlikely to produce recessions similar to the ones observed in the post-War US sample, implying that conditional on these DSGE models, the only available historical dataset can be viewed as an abnormality. It is these kinds of evaluation tools, we argue, that are needed to put the development of these DSGE models back on track. We believe that analysis like that illustrated in this paper can be highly informative for policymakers, who—in lieu of an immediate fix—can judgementally allow for these models in policymaking, and also for model developers, who can use this information to highlight specific misspecifications in the model and, thus, focus their attention on improving those portions of the models that are showing stress.

The rest of the paper is organized as follows: Section 2 characterizes the monetary policy process as assessing the interrelationships among the FEs, provides a structural diagnosis of the FEs, and describes the posterior predictive analysis used in the paper. Section 3 illustrates the posterior predictive evaluation approach using the SW model and discusses the results, and Section 4 concludes.

2 Model Evaluation Purpose and Tools

In this section, we provide a characterization of model-based monetary policy analysis that suggests model diagnostics that are particularly illuminating in highlighting certain policy relevant deficiencies in these models. In the following sections we also show how these suggested diagnostic tools can help us improve upon these models in their ongoing development to aid in the monetary policymaking process.

2.1 Policy Analysis

Policy makers meeting at time t do the following things: they observe new data since the last meeting at $t-1$ (thus, t is measured as the index of meetings), set the policy rate for the current meeting, $i_{t|t}$ and make a forecast for the policy rate for the next meeting, $i_{t+1|t}$.[2] Thus, on an ongoing basis policymakers come into the meeting at time t with the anticipated policy decided at the previous meeting, $i_{t|t-1}$, and at the meeting they decide how to update that view of optimal policy in light of information that has arrived since the last meeting. Under time consistency at least, policy makers will deviate from their expected policy path, $i_{t|t-1}$ only if they have observed new data that is not consistent with their expectations.

Practical policymaking at central banks is thus characterized as interpreting the structural sources of the news in the observed data and accordingly acting upon it. In the DSGE context, the news is entirely reflected in the one-step ahead forecast errors for the observable variables, Z_t:

$$\nu_t = Z_t - Z_{t|t-1}$$

Let us suppose that policy is given by a simple Taylor rule (any linear policy rule will

[2]In a forward looking model, forming the expectations about the future path of policy is an inherent part of setting policy today.

do here),

$$i_{t|t} = a + b\pi_{t|t} + cy_{t|t}$$

where $\pi_{t|t}$ is the assessment of inflation at t given time t information and $y_{t|t}$ is view of the output gap at t given information at t. As written, value of these two variables at t is not perfectly observed at t. Although, inflation is measured pretty well later, the gap between actual and efficient output remains imprecisely measured indefinitely.

The update in policy rule is written as:

$$i_{t|t} - i_{t|t-1} = a + b(\pi_{t|t} - \pi_{t|t-1}) + c(y_{t|t} - y_{t|t-1})$$

The crucial idea is that the update on these two latent variables under the linear and Gaussian structure of a DSGE model is given by the Kalman filter as a linear function of the news:

$$\begin{bmatrix} \pi_{t|t} - \pi_{t|t-1} \\ y_{t|t} - y_{t|t-1} \end{bmatrix} = \Gamma\nu_t = \Gamma(Z_t - Z_{t|t-1})$$

Thus, policy analysis, in this simple structure, is a matter of computing the news or the surprises in observables. The structural interpretation of this news is then given by the interrelationship among the structural shocks in the model and the Γs reflect the implications of this interrelationship for the latent variables.

To see a simple version of this, consider a simple textbook aggregate demand/ aggregate supply (AS/AD) framework where the observables are output and some indicator of inflation. The basic idea is that if output and prices come in higher than expected, then we might infer that a positive AD shock has shifted the AD curve outward, which would raise both output and inflation, and warrant a higher interest rate. If on the other hand, inflation comes in higher than expected but the output indicator is lower than expected, then we might infer a negative supply shock has shifted the AS curve inward, which would reduce output and raise inflation. The optimal policy response is this case might be to leave rates approximately unchanged if, say, the fall in output is the efficient response to the adverse supply shock.

This stylized account of policy suggests that we analyze the structure of 'news' according to the model. In particular, the model will imply a correlation matrix for one-step ahead forecast errors. Given a sample data, we can ask whether the realized forecast errors implied by the model appear to have the correlation structure that is implied by the model. This is a different question from pure forecast accuracy. We are not asking 'are the errors small?,' we are asking 'do the errors have the right interrelationships?'

2.2 Diagnosis of the One-Step Ahead Forecast Errors: A Simple Example

Consider a simple model in which the data are generated by two supply shocks that both push output growth and inflation in the opposite direction:

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = A \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \end{bmatrix} + C\varepsilon_t$$

where $\varepsilon_t \sim \text{iid}N(0, I)$, and the first row of the shock impact matrix, C, is negative and the second row is positive, e.g.,

$$C = \begin{bmatrix} -1 & -3 \\ 1 & 1 \end{bmatrix}$$

Setting aside small sample issues, the FEs are given by:

$$\nu_t = C\varepsilon_t$$

We denote the variance-covariance matrix of FEs, ν_t, by Ω and that is given by:

$$\begin{aligned}
\Omega &= CE(\varepsilon_t \varepsilon_t')C' \\
&= CC' \\
&= \begin{bmatrix} 10 & -4 \\ -4 & 2 \end{bmatrix}
\end{aligned} \tag{2.1}$$

We see that the FE covariance for output growth and inflation generated by this simple model will be negative: both shocks move output and prices in different directions.

Suppose the true model driving the data is as stated above except that one shock moves both output and inflation in the same direction and the other shock moves them in opposite direction as before. This can be captured in the above model if we replace C by:

$$\tilde{C} = \begin{bmatrix} -1 & 3 \\ 1 & 1 \end{bmatrix}$$

so the analyst is using a misspecified model that has two different supply shocks but in reality the data are generated by a process with one supply and one demand shock. Since both models have the same A, which we assume for simplicity is known, the optimal forecast of the two models are identical. We have chosen C and $\tilde{C}$ so that the variance of the two FEs is the same. This is to emphasize the difference between the diagonal elements of the variance-covariance matrix of the FEs that tell us about the accuracy of forecast errors, and the off-diagonal elements that tell us about the interrelationship among the FEs.

Suppose we observe a large sample of data generated according to the true model driven by A, $\tilde{C}$ and iid $N(0, I)$ shocks. The FEs estimated on this large realized sample will then have the variance-covariance matrix approximately equal to $\tilde{C}\tilde{C}'$, and in particular the FE covariance between the two variables will be positive:

$$\begin{aligned} \hat{\Omega} &= \tilde{C}\tilde{C}' \\ &= \begin{bmatrix} 10 & 2 \\ 2 & 2 \end{bmatrix} \end{aligned} \tag{2.2}$$

Note that the diagonal elements of the variance-covariance matrix of FEs given by the true model in the above equation are equal to those given by the misspecified model in (2.1).

If one is working with the misspecified model, then all errors would be interpreted as supply shocks, and policy would be chosen to be the optimal response to the observed mix of supply shocks. The only misspecification one would observe is that the covariance of forecast errors estimated on the realized sample would be different from that predicted by the model (the off-diagonal elements in (2.1) and (2.2)).

We can diagnose the above symptom to provide a structural analysis of the misspecified model that would help us in figuring out the true model. In particular, any estimate of the FEs, $\hat{\nu}_t$, will imply an estimate of the structural shocks, $\hat{\varepsilon}_t$. Under the model, we know that:

$$\hat{\nu}_t = C\hat{\varepsilon}_t$$

The estimate of the variance-covariance matrix of FEs implied by the misspecified model on the realized sample is then given by:

$$\begin{aligned} \hat{\Omega} &= CE(\hat{\varepsilon}_t\hat{\varepsilon}_t')C' \\ &= C\hat{\Sigma}C' \\ \Rightarrow \hat{\Sigma} &= C^{-1}\hat{\Omega}C'^{-1} \end{aligned}$$

where $\hat{\Sigma}$ is the sample variance-covariance matrix of the estimated structural shocks.

For our example values:

$$\hat{\Sigma} = \begin{bmatrix} 10 & -6 \\ -6 & 4 \end{bmatrix}$$

Thus, the structure of this misspecified model reflected in C, along with the realized variance-covariance matrix of the FEs, $\hat{\Omega}$, implies a realized value for the variance-covariance matrix of the structural shocks that does not obey the assumptions of the model. In other words, the estimated structural shocks on the realized sample turn out to be correlated to accommodate the misspecification in the model.

The symptom of misspecification we observe is that our estimate of the realized supply shocks on the observed sample, that is the $\hat{\varepsilon}'s$, have a negative sample correlation. The intuition: in the misspecified model both shocks move output and inflation in different directions, but in the realized sample output and inflation tended to move in the same direction. In order to reconcile the misspecified AS/AD model with the realized sample, we need that the two supply shocks work together in just the right way. That is we need just the right mix of negative correlation between the two structural shocks. When one shock is positive and tends to raise output and prices, the other is negative and tends to lower output and prices, and the positive effect dominates. In this very simple case, when observing that the model 'needed' the two supply shocks to be negatively correlated to explain the sample, we quickly deduce that what we need is a shock that moves output and inflation in the same direction—a demand shock. In the next section we show how to apply this analysis in the case of a more complex DSGE model.

2.3 Diagnosis of the One-Step Ahead Forecast Errors in the DSGE Context

The analysis works the same in a larger and more complex DSGE model, but given higher dimensions, the diagnosis is a bit more subtle. The diagnosis of the one-step ahead forecast errors in the DSGE context focusses on three additional subtleties we abstracted from in the simple case: a) sampling fluctuation in estimated parameters and sample variance-covariance matrices, b) the model may be misspecified in terms of the conditional mean, which was not the case above, and c) the DSGE models generally imply a vector autoregression-moving average (VARMA) structure.

We will take account of a), sampling fluctuations, using posterior predictive analysis as discussed in the next section. Issue b), misspecified conditional mean, adds no real problems and it simply adds to the list of problems we may detect. Issue c) requires a bit more discussion.

Regarding the third point, standard DSGE models imply a VARMA process for the data instead of a pure vector autoregression as assumed in the simple example above. In the VAR case, conditional on initial conditions we observe the ε's, but this is not the case with MA components. Thus, we must work with our best estimate of the ε's, which will be implied by the Kalman filter and/or Kalman smoother. Therefore, rather than working with ε_t, we can work with the updated structural shocks, $\hat{\varepsilon}_{t|t}$ that reflect only the information up to period t, or the smoothed structural shocks, $\hat{\varepsilon}_{t|T}$.[3] To see the effects of the MA terms, write the MA representation of the observables as,

$$Z_t = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}$$

where C_i denotes the coefficients for the MA parts and C_0 is the lag zero impact matrix equivalent to C in the example above. Taking expectations conditional on information available at time $t-1$, we get:

$$Z_{t|t-1} = \sum_{i=1}^{\infty} C_i \hat{\varepsilon}_{t-i|t-1}$$

[3] In this paper I only report the results using $\varepsilon_{t|t}$ to be consistent with the one-step ahead decision making problem of the policymakers. However, one might want to look at the smoothed shocks to reflect on how the model relates to the full information case. The results are not noticeably different for the two cases.

The forecast errors, ν_t, are then given by:

$$\begin{aligned} \nu_t &= C_0\varepsilon_t + \sum_{i=1}^{\infty} C_i(\varepsilon_{t-i} - \hat{\varepsilon}_{t-i|t-1}) \\ &= C_0\varepsilon_t + err \end{aligned}$$

This is an identity that must hold under the model for all versions of the shocks. For example, for the updated shocks we analyze:

$$\nu_{t|t} = C_0\varepsilon_{t|t} + err$$

where the err includes the revision to the ε's due to the new information made available this period relative to the previous period. In the simple VAR example we didn't have this additional err term. It turns out that this term in practice is small and so we only focus on explaining the relation between ν_t and ε_t using the lag zero impact matrix C.

2.4 Describing Posterior Predictive Analysis

The literature on prior and posterior predictive analysis was popularized by Box (1980) and has since then been extended by many others including Gelman et al. (1996), Bernardo (1999), Geweke (2007). A complete description of this prior and posterior predictive analysis as applied to the DSGE context is provided in Faust and Gupta (2010a). We provide a brief sketch over here for the sake of completeness.

Posterior predictive analysis relies on a simple idea: if the available sample is collectively an outlier from the standpoint of the model+posterior, then perhaps the model or prior should be refined. It provides formal tools for judging the degree to which relevant features of a sample are freakish from the standpoint of the model+posterior. If the realized value is too surprising, then that calls into question the practical validity of the model in further exercises.

In a standard Bayesian estimation approach, we have an economic model (the DSGE model) that describes the full joint distribution of observed variables, Y, in terms of unobservable parameters, θ. We define a descriptive feature, $h(Y)$, as one that can be described as a function of Y alone. The model+posterior will imply a marginal distribution for this feature that is known as the posterior predictive distribution.[4] This distribution is given by:

$$F_h(c) \equiv pr(h(Y^{rep}) \leq c)$$

where Y^{rep} is a random sample drawn according to the model+posterior of the same size as the realized sample. The implied posterior predictive density of this descriptive feature is then denoted by $f_h(x)$ and this allows us to check how freakish the realized sample, Y^r, is by comparing the observed value, $h(Y^r)$, against this density. We define the posterior predictive p-value of a one-tail test in the upper tail as the proportion of points in the upper tail of this density, $f_h(x)$, relative to $h(Y^r)$.

$$1 - F_h(h(Y^r))$$

In this paper I consider only structural features, $h(Y, \theta)$, that depend upon θ in addition to the sample. The structural features that I consider are the mean value and the correlations of the optimal one-step ahead model consistent forecast errors and the first and second moments of the updated structural shocks, $\varepsilon_{t|t}$.

Any feature when evaluated on the realized sample, Y^r, is referred to as the realized value of the feature. When talking about realized features, an important difference arises between a descriptive feature and a structural feature. While the former is defined completely by the realized sample at hand, Y^r, the latter is not, because of the dependence on the unknown θ.

Due to the dependence of the realized value of the feature on θ, computing the p-value is slightly more complex for the structural features. However, conditional on a

[4] I only consider posterior predictive distribution for different features in this paper. One could similarly look at the model+prior and that will imply the prior predictive distribution for the corresponding feature.

fixed θ^*, one can compute the realized value of the structural feature, $h(Y^r, \theta^*)$, and therefore the probability that the value for this feature in repeated sampling will be greater than the realized value for a fixed θ^* is given by:

$$pr(h(Y^{rep}, \theta^*) > h(Y^r, \theta^*))$$

In order to compute the posterior predictive p-value, one can now integrate out the dependence on θ using the posterior distribution for the parameters to get the p-value as follows:

$$pr(h(Y^{rep}, \theta^{rep}) > h(Y^r, \theta^{rep}))$$

where (Y^{rep}, θ^{rep}) are drawn according to the model+posterior and Y^{rep} is of the same sample size as Y^r.

In practical terms, computing the pair $h(Y^{rep}, \theta^{rep})$, $h(Y^r, \theta^{rep})$ for enough values of (Y^{rep}, θ^{rep}) drawn from the model+posterior will allow us to characterize the posterior predictive distribution for the structural feature and the posterior distribution for the realized sample value. To analyze these two distributions jointly we can look at a scatter plot with $h(Y^r, \theta^{rep})$ on the horizontal axis and $h(Y^{rep}, \theta^{rep})$ on the vertical axis. The p-value described above is then simply the proportion of points above the 45 degree line for a one-tail (upper tail) test of the posterior predictive density. For example, if the upper tail p-value is 0.05, we will see only 5% of the scatter plot above the 45 degree line. Summarizing a distribution with a single number such as a p-value can hide a lot of information. Such crude summaries should, therefore, be used with caution, and we will largely report the entire predictive density. Still at times, p-values provide a convenient and compact summary.

If the realized structural feature is not surprising from the standpoint of model +posterior then one should expect to see most of the scatter cloud to lie around the 45 degree line. On the other hand, if the entire scatter cloud lies either mostly above or mostly below the 45 degree line, then it says that for essentially no value of the posterior parameter is the model able to produce a value similar to that observed on the realized sample. This implies that either the realized sample is freakish from the standpoint of the DSGE model and we will almost never observe a sample like that again, or that the DSGE model is misspecified with regards to that feature.

3 Application

In this section, we evaluate the iconic SW DSGE model for the task of monetary policy analysis using the diagnostic tools of posterior predictive analysis as described in the previous section. We chose this model over many other competing models because this was the first model that was shown to forecast as well as certain atheoretical benchmarks like Bayesian VARs. It introduces a rich set of frictions and as many structural shocks as observed variables, most of which have meaningful economic interpretations. In addition, this is one of the best known medium-scale DSGE models available for policy analysis. In the rest of this section I first briefly discuss the model, and then discuss the results of the posterior predictive evaluation of this model and their implications for model assessment.

3.1 DSGE Model: Smets and Wouters (2007)

SW is an extension of the standard DSGE model with sticky wages and sticky prices, largely based on Christiano et al. (2005)[5]. This model allows for sticky nominal wage and price settings with backward inflation indexation. Other features include habit formation in consumption, investment adjustment costs, variable capacity utilization, and fixed costs in production. The model introduces seven orthogonal structural shocks that include productivity, investment, risk premium, government spending, wage and price mark-up, and monetary policy shocks.

Households maximize a non-separable utility function in consumption and labor. Consumption depends on the previous period's consumption and the degree of habit

[5]The log-linearized equations of the model are provided in appendix A. Readers are referred to Smets and Wouters (2007) for a thorough explanation of the model equations and frictions.

formation is given exogenously. Labor is differentiated, so households have some market power over wages. Due to wage rigidity à la Calvo (1983), households set their optimized wages only periodically, and the households that do not optimize, partially index the wages to the previous period's inflation. Households own the capital stock and rent it out to firms. They decide how much to invest given the investment adjustment costs and also determine the rate of capital utilization in order to minimize costs. Labor aggregator firms purchase the differentiated labor input from the households and transform it into aggregate labor using the Kimball aggregator. A continuum of intermediate firms purchase this aggregated labor and rent capital from households and produce differentiated goods that are sold to the final producers. Similar to households, intermediate firms face nominal rigidities and set prices à la Calvo (1983). Prices that are not optimized are partially indexed to the previous period's inflation. The final goods firm then takes the prices of these intermediate goods as given and transforms them into a composite good sold to consumers, investors, and the government. The model is closed with a Taylor type monetary policy reaction function, where the interest rate is adjusted gradually in response to the output gap and inflation.

This model has been estimated with Bayesian techniques using quarterly U.S. data for seven key macro-economic variables from 1966 to 2004: real GDP growth, real consumption growth, real investment growth, inflation, real wage growth, hours worked, and the nominal interest rate. GDP, personal consumption expenditure and private fixed investment are all deflated using GDP price deflator and divided by a population index, thus making them real per capita variables. Hours worked is computed by multiplying civilian employment with the average weekly hours worked by all persons in the non-farm business sector. This is divided by the population index to make the series per capita. Real wage is computed by deflating the hourly compensation of all persons in the non-farm business sector by the GDP deflator. Inflation is defined as the log difference in the GDP deflator, and the nominal interest rate used is the quarterly effective federal funds rate. All growth rates are computed using quarter-to-quarter log differences.

3.2 Variance-Covariance Matrix of One-Step Ahead Forecast Errors

In this paper, we compare the realized value of the elements of the variance-covariance matrix of the FEs, Ω, implied by the model+posterior with the corresponding posterior predictive distribution.[6] The matrix Ω can be broken down into the diagonal elements, the standard deviations of the one-step ahead FEs (FESTDs), and the normalized off-diagonal elements, the one-step ahead FE correlations (FECs). We demonstrate that it is informative to evaluate these off-diagonal elements of Ω. To begin with, we provide a comparison of the point estimates at the posterior mode for the realized value and the population value. However, these are only meant to provide a benchmark reference point and we later look at the full posterior predictive distribution.

Table 1 compares the posterior mode value for the elements of Ω, $\Omega(\theta^*)$, to the realized value for these features computed at the posterior mode, $\Omega(Y^r, \theta^*)$. For the FESTDs, the posterior mode values implied by SW are "close" to the realized values implied by the model at the posterior mode. This closeness in point estimates is confirmed by the posterior distributions around these point estimates. The first row in Figure 3 graphs the scatter plots of the the FESTDs, the diagonal elements of Ω. These scatter plots provide a natural way to compare the posterior predictive density (vertical axis) to the posterior density for the realized sample (horizontal axis) for these structural features. For instance, for the FESTD of interest rates (Figure 3, row 1, column 7), the scatter cloud centers on the 45 degree line. This says that a typical sample drawn according to the model+posterior will have its interest rate FESTD similar to what is observed in the sample at hand. Except for output growth and hours (for which the scatter cloud lies mainly over the 45 degree line), this is true for all the other observed variables. The p-values reported on the upper left corner for the panels for the FESTD of output growth and hours indicates that the model+posterior produces much higher

[6] All computations are done using the software DYNARE. The posterior predictive distributions and the posterior distributions for the realized structural features are based on 60000 random draws from the posterior distribution of the estimated parameter vector θ in SW.

volatility in these variables than what it estimates on the realized sample. Overall, we observe that the SW model does well with regards to matching the forecast error accuracy as measured by the diagonal elements of the Ω matrix.

The second half of Table 1 compares the posterior mode value of the off-diagonal elements of Ω, the FECs, to the realized value for these features computed at the posterior mode. SW performs poorly with regards to some key FECs. The scatter plots comparing the posterior predictive values to the posterior values for the realized sample for these FECs are graphed in rows 2,3 and 4 in Figure 3.

As an illustration, we focus on the FEC($Hours, \Delta W$) graphed separately in Figure 1 in panel (a).[7] In this case, the entire scatter cloud lies above the 45 degree line implying that the FEC($Hours, \Delta W$) is much higher in predictive samples than what is observed in the realized sample. The posterior predictive values (on the vertical axis) are centered around zero, whereas the posterior values for the realized sample (on the horizontal axis) show a negative correlation.

The diagnosis of this discrepancy guides us to the structural misspecification in this model using the structural accounting provided in section 3.3. The remaining panels of Figure 1 graph the scatter plots for the correlations among the updated structural shocks that account for why the posterior values for the realized sample of the FEC($Hours, \Delta W$) differ substantially from its posterior predictive values. Table 6 provides a quantitative accounting at the posterior mode for how much of the negative realized value of $FEC(Hours, \Delta W)$ is accounted for by these correlated shocks.

The economic rationale behind why the model needs correlated shocks to generate the negative $FEC(Hours, \Delta W)$ shown in Figure 1 is as follows. The models needs a shock that raises wages and lowers hours at the same time to produce the negative correlation between the forecast errors for hours and wage growth. Productivity and wage-mark up shocks are the only potential candidates in a model with uncorrelated shocks. First consider the productivity shock. A positive productivity shock in this

[7]The point estimate for this correlation, given in Table 1, is -0.01 for the posterior mode value relative to -0.30 for the realized value.

model results in an increase in a firm's mark-up as sticky prices prevent the firm's prices from rising and sticky wages stall the rise in wages until the next period of optimization. Because this wedge between the marginal productivity of labor and real wage is expected to decrease over time, real wages are expected to rise in the future. This generates an intertemporal substitution effect that causes households to reduce their labor supply contemporaneously. Also, due to predetermined prices, real balances and, thus, aggregate demand remain unchanged and the same output can be produced using fewer hours. In the presence of sticky wages, the model+posterior is unable to produce a big enough increase in wage growth. Therefore, in response to a productivity shock alone, the model+posterior is quantitatively unable to account for the negative one-step ahead forecast error correlation between hours and wage growth realized in the observed sample. The wage growth shock, on the other hand, bypasses the wage stickiness and increases wages on impact but has a limited negative impact on hours worked as the negative wealth effect is countered by the positive substitution effect.

Overall, whatever little negative correlation is generated by the productivity and wage mark-up shocks is countered by positive correlation generated between hours and wage growth by other shocks. The model, therefore, requires certain pairs of shocks estimated on the realized sample to happen together in order to produce the negative value of realized $FEC(Hours, \Delta W)$. For instance, a positive correlation between the productivity and wage mark-up shocks causes wages to rise and hours to fall on impact thereby explaining some of the additional negative realized value of $FEC(Hours, \Delta W)$. The negative realized structural shock correlation between government spending and wage mark-up also accounts for the negative realized value of $FEC(Hours, \Delta W)$. As discussed earlier, a positive wage mark-up shock that raises wages but does not have a significantly negative impact on hours, when accompanied by a negative spending shock causes hours worked to fall substantially producing a negative correlation in the FEs for hours and wage growth. Similarly, the other correlated shock pairs can be shown to account for the observed $FEC(Hours, \Delta W)$ using the lag zero shock impact matrix,

C, provided in Table 3.[8]

This diagnosis tells us that the SW DSGE model+posterior is highly unlikely to produce samples with a negative value for $FEC(Hours, \Delta W)$. This symptom of misspecification then tells us that the model needs a shock that raises wages and lowers hours worked at the same time and does so in a quantitatively significant way. A leisure preference shock, that amounts to people being voluntarily unemployed, could do the trick. However, this is not a satisfactory explanation and we argue that other potential channels such as the role of labor market frictions and efficiency wages should be explored within the framework of these DSGE models to account for this observed discrepancy.

Figure 2 and Table 7 provide a similar accounting analysis for the realized value of the $FEC(\Delta Y, \pi)$. Figure 2, panel (a), graphs the scatter plot for $FEC(\Delta Y, \pi)$ and the remaining panels plot the correlated pairs of structural shocks that account for the realized value of this FEC. We see that even though this correlation tends to be negative for the posterior values for the realized sample, the posterior predictive values can produce such a low negative correlation only about nine percent of the times. This might be considered as a crucial issue if the model is to be taken seriously for use in policy analysis as output growth and inflation are the two key policy variables. As was shown in our simple example in section 3.2, it is important for the model to get right not only the standard deviation of the FEs in output growth and inflation, but more importantly how these two forecast errors are correlated. The diagnosis of this FEC depends, once again, on the nature of the realized estimate of structural shock correlations and the shock impact matrix, C, that shows that productivity and mark up-shocks are the only candidates to generate a negative correlation between the FEs for output growth and inflation. However, these shocks are not large enough to get the desired negative correlation on their own and the model requires certain shocks to happen together. The problem here seems to be that the model+posterior is

[8] It is important to note that matrix C reports the impact effect of a one standard deviation shock and not a one unit shock.

putting too much emphasis on demand shocks and in order to counter-act this effect, the model requires a combination of positive and negative correlations among the various structural shocks as shown in Table 7. Faust and Gupta (2010a) highlight that the model+prior in the SW model heavily leans towards a bigger role for the demand shocks. In light of that result, to fix this issue at hand, one could either change the prior or think of a new model specification that has a larger role for supply shocks in these models. However, since we are looking at a general equilibrium model, it is not certain how this would affect the overall likelihood of the model. This analysis, nevertheless, provides a starting point for future model refinement as it highlights a crucial weakness of the model.

Figure 3 graphs the scatter plots for all the elements of Ω; Figure 4 graphs the scatter plots for all the elements of Σ. We could repeat the above analysis for all the elements of Ω by relating them to the elements of Σ using the impact matrix, C.

3.3 Mean of One-Step Ahead Forecast Errors

It might seem surprising that the mean value of the structural shocks happens to be non-zero in the SW DSGE model. This is, however, because the mean is not freely estimated in the model. Thus, as in a regression with no intercept term, the residuals need not be mean zero.[9] Output and investment growth, are both being under-predicted by the model (evaluated at posterior mode) by approximately 0.5% per quarter and investment growth by as much as 1.27% per quarter. On the other hand, inflation and interest rate variables have independent parameters estimating the trend value for these variables and these turn out to be under-predicted (at the posterior mode) by approximately 0.1% per quarter which could be ascribed to sampling or model uncertainty. Edge et al. (2008) disaggregate US private demand into four categories of private expenditures and confirm that these expenditure categories tend to have different average trend growth

[9] In particular, the model parameterizes the same deterministic trend value for all the real growth variables namely, output growth, consumption growth, investment growth and real wage growth. It is precisely these variables that tend to have a much higher mean value of the FEs.

rates. Therefore, if the different expenditure categories are constrained to a common trend growth rate as in the analyzed SW model for the US economy then the lack of free parameters shows up in the mean value of the FEs. Given these mean FEs are non-zero, it is nevertheless interesting to ask whether the observed in-sample bias is consistent with errors obtained by real-world forecasters on this sample.

In this section, we compare the posterior distribution for the realized value of the mean of the in-sample FEs with some simple model-free benchmarks. These are: (i) the median Survey of Professional Forecasters (SPF) FEs for the observed variables—output growth, consumption growth, investment growth, and inflation; and (ii) the federal funds futures (FFF) FEs.[10] The one-quarter ahead SPF and federal funds futures are both real-time forecasts and, as a result, are disadvantaged in comparison with the model forecasts on two counts: first, they are out-of-sample forecasts; second, the current quarter information for most variables is not available for real-time forecasts and they, instead, have to be conditioned on the nowcast—the forecast for the current quarter for the variable concerned. The in-sample DSGE model forecast, on the other hand, assumes full knowledge of current period information, including all information for the current quarter that is made available in the future. Having noted the differences in the timing conventions, these real-time forecasts are meant to provide only a benchmark for the mean value of the FEs over the sample period.

As an illustration, we discuss the mean value of the FE's for investment growth graphed in the bottom row in Figure 5. The grey band shows the 90% posterior distribution for the realized value of this forecast error. This band consistently lies over the 0 line for most part of the sample and the mean realized value of the forecast error in investment growth for the full sample from 1966 to 2004, implied by the model+posterior is 1.27% on a quarterly basis. This value is given in Table 4. The median Survey of Professional Forecasters mean FE for investment growth for the sample from 1981

[10] The SPF forecasts for the next quarter are made in the middle of the current quarter. For instance, the second-quarter SPF forecasts are made in the middle of February. The FFF forecasts are made at the end of the current quarter. Here, we use the average FFF forecast for the month of January, February, and March made at the last day of December.

to 2004 is 0.39%. The realized value for the corresponding sample, implied by the model+posterior, is 1.12% per quarter. One can use the similar diagnostic analysis as used for the variance-covariance matrix of FEs, to diagnose the realized mean value of the one-step ahead investment growth FE using the impact matrix C and the realized mean value of the updated structural shocks given in Table 5. The main non-zero mean values for the structural shocks that contribute to the non-zero FE for investment growth are the productivity shock, the investment technology shock, the risk-premium shock and the government spending shock.

The posterior distributions for the realized value of the mean structural shocks are given in Figure 7. Except the monetary policy shock and the wage mark-up shock, all other shocks have a non-zero realized mean value.

4 Conclusion

This paper provides a way to analyze the monetary policy process using the posterior predictive analysis described in Faust and Gupta (2010a). The paper characterizes monetary policy analysis as being divided into two steps: first, estimating the first and second moments of the FE's and, second, filtering the structural implications of these forecast errors using the shock impact matrix, C, and the realized value of the estimated structural shocks. The paper illustrates the application of these tools to the SW model, highlighting the model's strengths and weaknesses. The model+posterior does reasonably well on the FESTDs but performs poorly with regards to certain key FECs. In addition, the mean FE's for the observables implied by the model+posterior are nonzero. The paper also highlights specific misspecifications in the model with regards to the structural shocks. We show that the model is highly over-identified and the only way it can accommodate the observed data is by assigning nonzero cross-correlations and nonzero means to the realized value of the one-step ahead forecast errors and this is not consistent with a structural interpretation of the shocks.

The ultimate goal of any model evaluation tool should be to highlight specific flaws in the structure of the model and identify possible areas of improvement for future model building. The evaluation tools discussed in this paper diagnose the problem areas in these models at a structural level and highlight what pairs of shocks in these models are the trouble areas. We strongly encourage the use thee tools as it can help the DSGE modelling experts to concentrate their efforts in model refinement in areas that are particularly troublesome from the standpoint of policymaking.

5 Appendix

5.1 Log-linearized Model equations

The consumption Euler equation is given by:

$$\begin{aligned}\hat{c}_t &= \frac{h/\gamma}{1+h/\gamma}\hat{c}_{t-1} + \frac{1}{1+h/\gamma}E_t\hat{c}_{t+1} + \frac{(\sigma_c-1)(W_*^h L_*/C_*)}{\sigma_c(1+h/\gamma)}(\hat{l}_t - E_t\hat{l}_{t+1}) \\ &\quad - \frac{1-h/\gamma}{\sigma_c(1+h/\gamma)}(\hat{r}_t - E_t\hat{\pi}_{t+1} + \hat{\epsilon}_t^b)\end{aligned}$$

Current consumption depends on a weighted average of past and expected future consumption, the ex-ante real interest rate($\hat{r}_t - E_t\hat{\pi}_{t+1}$), expected employment growth ($\hat{l}_t - E_t\hat{l}_{t+1}$) and a risk-premium shock, $\hat{\epsilon}_t^b$. h represents the habit formation coefficient, σ_c is the inverse of the intertemporal elasticity of substitution and γ represents the labor-augmenting deterministic growth rate in the economy. The investment Euler equation is given by:

$$\hat{i}_t = \frac{1}{1+\beta\gamma^{1-\sigma_c}}\hat{i}_{t-1} + \frac{\beta\gamma^{1-\sigma_c}}{1+\beta\gamma^{1-\sigma_c}}E_t\hat{i}_{t+1} + \frac{1}{(1+\beta\gamma^{1-\sigma_c})\gamma^2\varphi}\hat{q}_t + \hat{\epsilon}_t^i$$

Current investment, $\hat{i}_t$, depends on past and expected future investment, the value of the existing capital stock, $\hat{q}_t$, and an investment-specific technology shock, $\hat{\epsilon}_t^i$. β is the rate of time preference and φ is the steady-state elasticity of the investment adjustment cost function. The value of capital is given by:

$$\hat{q}_t = \beta\gamma^{-\sigma_c}(1-\delta)E_t\hat{q}_{t+1} + (1-\beta\gamma^{-\sigma_c}(1-\delta))E_t\hat{r}_{t+1}^k - (\hat{r}_t - E_t\hat{\pi}_{t+1} + \hat{\epsilon}_t^b)$$

The value of the capital stock depends positively on its expected future value and the expected rental rate of capital, $E_t\hat{r}_{t+1}^k$, and negatively on the ex-ante real interest rate and the preference shock. The current capital used in production, $\hat{k}_t^s$ is given by:

$$\hat{k}_t^s = \hat{k}_{t-1} + \hat{z}_t$$

Current capital used in production is equal to the capital installed in the previous period, $\hat{k}_t$, plus the degree of capital utilization, $\hat{z}_t$. The degree of capital utilization as determined by the rental rate of capital is given by:

$$\hat{z}_t = \frac{(1-\psi)}{\psi}\hat{r}_t^k$$

where ψ reflects the capital utilization adjustment costs. The rental rate of capital is given by:

$$\hat{r}_t^k = -(\hat{k}_t - \hat{l}_t) + \hat{w}_t$$

The capital accumulation equation is given by:

$$\hat{k}_t = \frac{(1-\delta)}{\gamma}\hat{k}_{t-1} + (1 - \frac{(1-\delta)}{\gamma})\hat{i}_t + (1 - \frac{(1-\delta)}{\gamma})(1+\beta\gamma^{1-\sigma_c})\gamma^2\varphi\hat{\epsilon}_t^i$$

where δ is the rate at which capital depreciates. On the supply side, aggregate production is given by:

$$\hat{y}_t = \phi_p(\alpha\hat{k}_t^s + (1-\alpha)\hat{l}_t + \hat{\epsilon}_t^a)$$

Output is produced using capital and labor where α denotes the share of capital in production. $\hat{\epsilon}_t^a$ represents the productivity shock, while ϕ_p represents the fixed costs in production. On the demand side, the aggregate resource constraint is given by:

$$\hat{y}_t = (1 - g_y - i_y)\hat{c}_t + (\gamma - 1 + \delta)k_y\hat{i}_t + R_*^k k_y\hat{z}_t + \hat{\epsilon}_t^g$$

The demand for output comes from consumption, investment, capital utilization costs, and government spending. The coefficients on consumption and investment represent the steady-state consumption-output ratio and the steady-state investment-output ratio respectively. k_y is the steady-state capital-output ratio and R_*^k is the steady-state rental rate of capital. The price mark-up defined as the negative of the real marginal cost is given by:

$$\hat{\mu}_t^p = \widehat{mpl}_t - \hat{w}_t = \alpha(\hat{k}_t - \hat{l}_t) + \hat{\epsilon}_t^a - \hat{w}_t$$

Price mark-up, $\hat{\mu}_t^p$, is equal to the difference between the marginal product of labor, $\hat{mpl}_t$, and the real wage rate, $\hat{w}_t$. The price equation illustrating the profit maximization behavior of the firms is given by:

$$\hat{\pi}_t = \frac{\iota_p}{1+\beta\gamma^{1-\sigma_c}\iota_p}\hat{\pi}_{t-1} + \frac{\beta\gamma^{1-\sigma_c}}{1+\beta\gamma^{1-\sigma_c}\iota_p}E_t\hat{\pi}_{t+1} - \frac{1}{1+\beta\gamma^{1-\sigma_c}\iota_p}\frac{(1-\beta\gamma^{1-\sigma_c}\xi_p)(1-\xi_p)}{((\phi_p-1)\epsilon_p+1)\xi_p}\hat{\mu}_t^p + \hat{\epsilon}_t^p$$

Current inflation depends on past and expected future inflation, the marginal cost, $-\hat{\mu}_t^p$, and a price mark-up shock, $\hat{\epsilon}_t^p$. ι_p is the degree of indexation to past inflation, $(1-\xi_p)$ is the Calvo probability of being allowed to re-optimize prices, and ϵ_p represents the curvature of the Kimball goods market aggregator. A higher ϵ_p increases the complementarity with other prices, and, therefore, slows the speed of adjustment to the desired mark-up. The wage mark-up defined as the difference between real wage and the marginal rate of substitution is given by:

$$\hat{\mu}_t^w = \hat{w}_t - m\hat{r}s_t = \hat{w}_t - (\sigma_l\hat{l}_t + \frac{1}{1-h/\gamma(\hat{c}_t - h/\gamma\hat{c}_{t-1})})$$

The wage equation is given by:

$$\begin{aligned}\hat{w}_t \quad = \quad & \frac{1}{1+\beta\gamma^{1-\sigma_c}}\hat{w}_{t-1} + \frac{\beta\gamma^{1-\sigma_c}}{1+\beta\gamma^{1-\sigma_c}}(E_t\hat{w}_{t+1} + E_t\hat{\pi}_{t+1}) - \frac{1+\beta\gamma^{1-\sigma_c}\iota_w}{1+\beta\gamma^{1-\sigma_c}}\hat{\pi}_t \\ & + \frac{\iota_w}{1+\beta\gamma^{1-\sigma_c}}\hat{\pi}_{t-1} - \frac{1}{1+\beta\gamma^{1-\sigma_c}}\frac{(1-\beta\gamma^{1-\sigma_c}\xi_w)(1-\xi_w)}{((\phi_w-1)\epsilon_w+1)\xi_w}\hat{\mu}_t^w + \hat{\epsilon}_t^w\end{aligned}$$

Current real wage depends on past and expected real wages, past, current, and expected inflation, the wage mark-up, $\hat{\mu}_t^w$, and a wage mark-up shock, $\hat{\epsilon}_t^w$. ι_w is the degree of wage indexation to past inflation, $(1-\xi_w)$ is the Calvo probability of being allowed to re-optimize wages, ϵ_w represents the curvature of the Kimball labor market aggregator, and $(\phi_w - 1)$ is the steady-state labor market mark-up. The model is closed using a Taylor type monetary policy reaction function given by:

$$\hat{r}_t = \rho\hat{r}_{t-1} + (1-\rho)[r_\pi\hat{\pi}_t + r_y(\hat{y}_t - \hat{y}_t^p)] + r_{\Delta y}[(\hat{y}_t - \hat{y}_t^p) - (\hat{y}_{t-1} - \hat{y}_{t-1}^p)] + \hat{\epsilon}_t^r$$

The monetary authorities gradually adjust the interest rate in response to output gap and inflation. Output gap is defined as the difference between actual and potential output, $\hat{y}_t^p$. Potential output is defined as the efficient level of output that would prevail in the absence of price and wage rigidity, i.e. under flexible prices and wages. r_π, r_y, and $r_{\Delta y}$ are coefficients of the monetary policy reaction function, and ρ represents the interest rate smoothing in the policy function.

References

Adolfson, M., Andersson, M., Linde, J., Villani, M., Vredin, A., 2007. Modern forecasting models in action: Improving macroeconomic analyses at central banks. International Journal of Central Banking.

Bernardo, J., 1999. Quantifying surprise in the data and model verification. Bayesian Statistics 6, 72–73.

Box, G., 1980. Sampling and Bayes' inference in scientific modelling and robustness. Journal of the Royal Statistical Society. Series A (General) 143 (4), 383–430.

Calvo, G., 1983. Staggered prices in a utility-maximizing framework. Journal of monetary Economics 12, 383–398.

Chari, V., July 2010. Testimony before the U.S. House of Representatives. House Committee on Science and Technology, Subcommittee on Investigations and Oversight.

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

Colander, D., July 2010. Written testimony of David Colander. House Committee on Science and Technology, Subcommittee on Investigations and Oversight.

Del Negro, M., Schorfheide, F., Smets, F., Wouters, R., 2007. On the fit of New Keynesian models. Journal of Business & Economic Statistics 25 (2), 123–143.

Edge, R., Kiley, M., Laforte, J., 2008. Natural rate measures in an estimated DSGE model of the US economy. Journal of Economic Dynamics and Control 32 (8), 2512–2535.

Edge, R., Kiley, M., Laforte, J., 2009. A comparison of forecast performance between Federal Reserve staff forecasts, simple reduced-form models, and a DSGE model. Finance and Economics Discussion Series.

Faust, J., Gupta, A., 2010a. Posterior predictive analysis for evaluating DSGE models. manuscript in progress, Johns Hopkins University.

Faust, J., Gupta, A., 2010b. Are all recessions black swans? DSGE models and the post-war U.S. business cycle. in progress, Johns Hopkins University.

Gelman, A., Meng, X., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica 6, 733–759.

Geweke, J., 2007. Bayesian model comparison and validation. American Economic Review 97 (2), 60–64.

Hansen, B., 2005. Challenges for econometric model selection. Econometric Theory 21 (01), 60–68.

Kydland, F., Prescott, E., 1996. The computational experiment: an econometric tool. The Journal of Economic Perspectives, 69–85.

Smets, F., Wouters, R., 2003. An estimated dynamic stochastic general equilibrium model of the Euro area. Journal of the European Economic Association 1 (5), 1123–1175.

Smets, F., Wouters, R., 2004. Forecasting with a Bayesian DSGE model: An application to the Euro area. Journal of Common Market Studies 42 (4), 841–867.

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

Tiao, G., Xu, D., 1993. Robustness of maximum likelihood estimates for multi-step predictions: the exponential smoothing case. Biometrika 80 (3), 623–641.

Table 1: Ω, Variance-Covariance Matrix of FE's, ν_t, at Posterior Mode.

Variables	Standard Deviations	
	Posterior	Realized Value
ΔGDP	0.85	0.74
ΔC	0.56	0.62
ΔI	1.75	1.84
hours	0.61	0.54
ΔW	0.52	0.55
inflation	0.29	0.28
interest rate	0.24	0.24
	Correlations	
ΔGDP, ΔC	0.50	0.56
ΔGDP, ΔI	0.59	0.48
ΔGDP, hours	0.63	0.49
ΔGDP, ΔW	0.19	0.12
ΔGDP, inflation	-0.05	-0.16
ΔGDP, interest rate	0.04	0.14
ΔC, ΔI	0.22	0.36
ΔC, hours	0.38	0.22
ΔC, ΔW	0.04	0.26
ΔC, inflation	-0.08	-0.28
ΔC, interest rate	0.05	0.07
ΔI, hours	0.41	0.48
ΔI, ΔW	0.14	0.02
ΔI, inflation	0.01	-0.04
ΔI, interest rate	0.03	0.14
Hours, ΔW	-0.01	-0.30
Hours, inflation	0.12	0.11
Hours, interest rate	0.29	0.38
ΔW, inflation	-0.09	-0.20
ΔW, interest rate	-0.03	-0.09
inflation, interest rate	0.31	0.17

Table 2: Σ, Variance-Covariance Matrix of Structural Shocks, $\varepsilon_{t|t}$, at Posterior Mode.

Shocks	Standard Deviations	
	Posterior	Realized Value
productivity	0.45	0.46
risk premium	0.24	0.26
govt spending	0.52	0.54
investment	0.45	0.47
monetary policy	0.24	0.24
price mark-up	0.14	0.14
wage mark-up	0.24	0.25
	Correlations (Posterior Mode Value=0)	
productivity, risk premium		-0.19
productivity, govt spending		0.03
productivity, investment		-0.27
productivity, monetary policy		-0.01
productivity, price mark-up		-0.13
productivity, wage mark-up		0.17
risk premium, govt spending		0.27
risk premium, investment		-0.09
risk premium, monetary policy		-0.12
risk premium, price mark-up		0.25
risk premium, wage mark-up		-0.03
govt spending, investment		-0.22
govt spending, monetary policy		0.18
govt spending, price mark-up		0.25
govt spending, wage mark-up		-0.13
investment, price mark-up		0.02
investment, inflation		0.06
investment, wage mark-up		-0.22
monetary policy, price mark-up		-0.06
monetary policy, wage mark-up		-0.05
price mark-up, wage mark-up		-0.06

Table 3: Lag Zero Shock Impact Matrix, C at Posterior Mode

Variables→ Shocks↓	ΔY	ΔC	ΔI	hours	ΔW	π	interest rate
productivity	0.33	0.07	0.32	-0.28	0.07	-0.05	-0.07
risk premium	-0.42	-0.51	-0.35	-0.29	-0.04	-0.02	-0.11
govt spending	0.50	-0.06	-0.12	0.34	0.01	0.01	0.03
investment	0.37	-0.04	1.64	0.25	0.03	0.04	0.04
monetary policy	-0.19	-0.19	-0.28	-0.13	-0.03	-0.04	0.18
price mark-up	-0.12	-0.05	-0.23	-0.05	-0.28	0.24	0.06
wage mark-up	-0.03	-0.08	-0.07	-0.06	0.43	0.13	0.04

Table 4: SPF vs DSGE Model: Mean Value of FE's, $E(\nu_t)$.

Variables	Period	SPF	DSGE	DSGE(full sample)
ΔGDP	81Q4 — 04Q4	0.03	0.40	0.56
ΔC	81Q4 — 04Q4	0.16	0.38	0.50
ΔI	81Q4 — 04Q4	0.39	1.12	1.27
inflation	69Q1 — 04Q4	0.02	0.10	0.10
interest rate*	89Q1 — 04Q4	-0.02	0.11	0.13

*Interest rate forecast errors are based on federal funds futures.

Table 5: Mean Value of Structural Shocks, $E(\varepsilon_{t|t})$, at Posterior Mode.

	Realized Value
productivity	0.16
risk premium	-0.25
govt spending	-0.12
investment	0.27
monetary policy	0.01
price mark-up	0.05
wage mark-up	0.00

*Posterior mode value=0

Table 6: Accounting the Realized Value of FEC($Hours, \Delta W$), at Posterior Mode

Main Correlated Shock Pairs	Shock Correlation	Contribution to FEC($Hours, \Delta W$)
(productivity, wage mark-up)	0.17	-0.07
(govt spending, wage mark-up)	-0.13	-0.06
(investment, wage mark-up)	-0.22	-0.08
(govt spending, price mark-up)	0.25	-0.08
(risk premium, price mark-up)	0.25	0.07

FEC($Hours, \Delta W$) with uncorrelated shocks= -0.01

FEC($Hours, \Delta W$) with correlated shocks= -0.30

Table 7: Accounting for Realized Value of FEC($\Delta Y, \pi$), at Posterior Mode

Main Correlated Shock Pairs	Shock Correlation	Contribution to FEC($\Delta Y, \pi$)
(productivity, price mark-up)	-0.13	-0.05
(risk premium, price mark-up)	0.25	-0.13
(govt spending, price mark-up)	0.25	0.15
(investment, wage mark-up)	-0.22	-0.05

FEC($\Delta Y, \pi$) with uncorrelated shocks= -0.05

FEC($\Delta Y, \pi$) with correlated shocks= -0.16

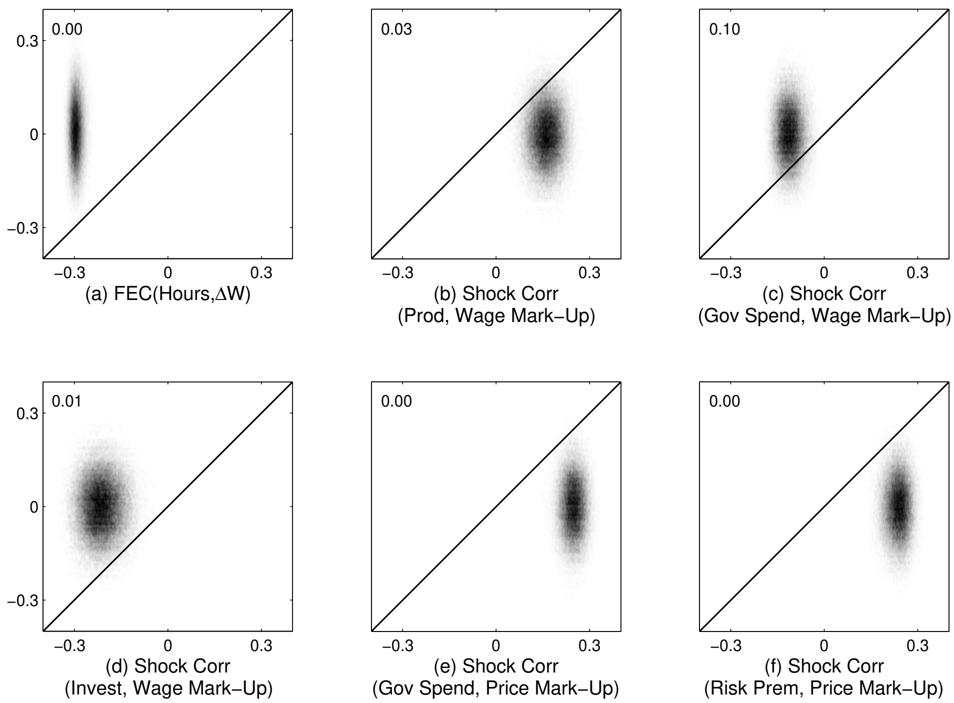

Figure 1: Scatter plots of correlated shocks that account for FEC($Hours, \Delta W$). Panel a) plots FEC($Hours, \Delta W$); remaining panels plot the correlated shocks that account for the negative values of this correlation on the realized sample (see Table 6). Horizontal axis plots the posterior density for the realized sample; vertical axis plots the posterior predictive values. The number in the upper left gives the smaller share of points on either side of the 45 degree line.

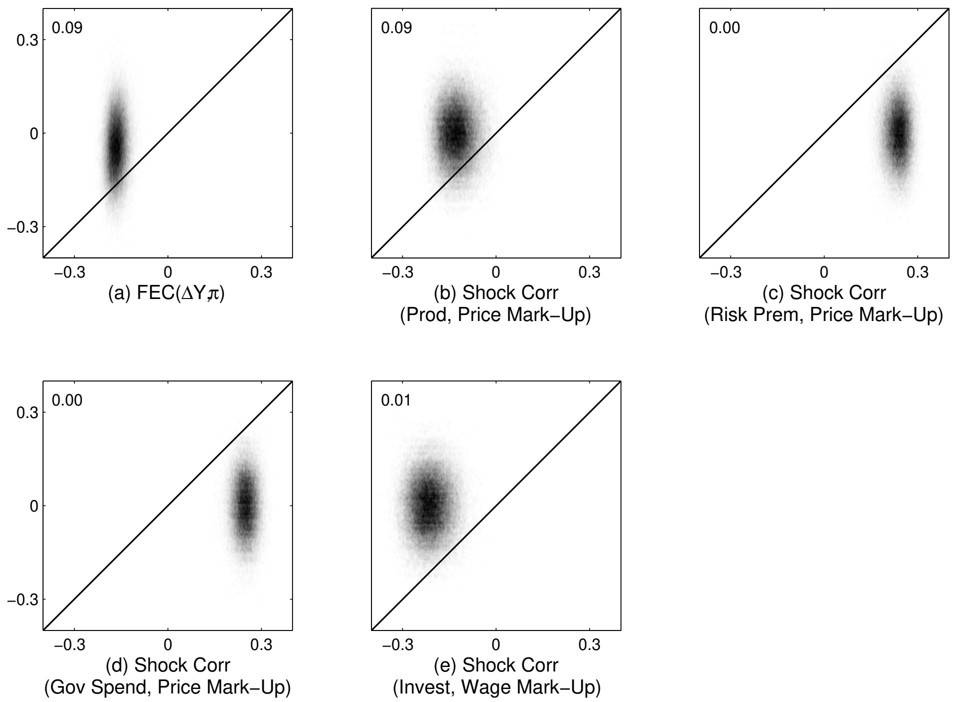

Figure 2: Scatter plots of correlated shocks that account for FEC($\Delta Y, \pi$). Panel a) plots FEC($\Delta Y, \pi$); remaining panels plot the correlated shocks that account for the negative values of this correlation on the realized sample (see Table 7). Horizontal axis plots the posterior density for the realized sample; vertical axis plots the posterior predictive values. The number in the upper left gives the smaller share of points on either side of the 45 degree line.

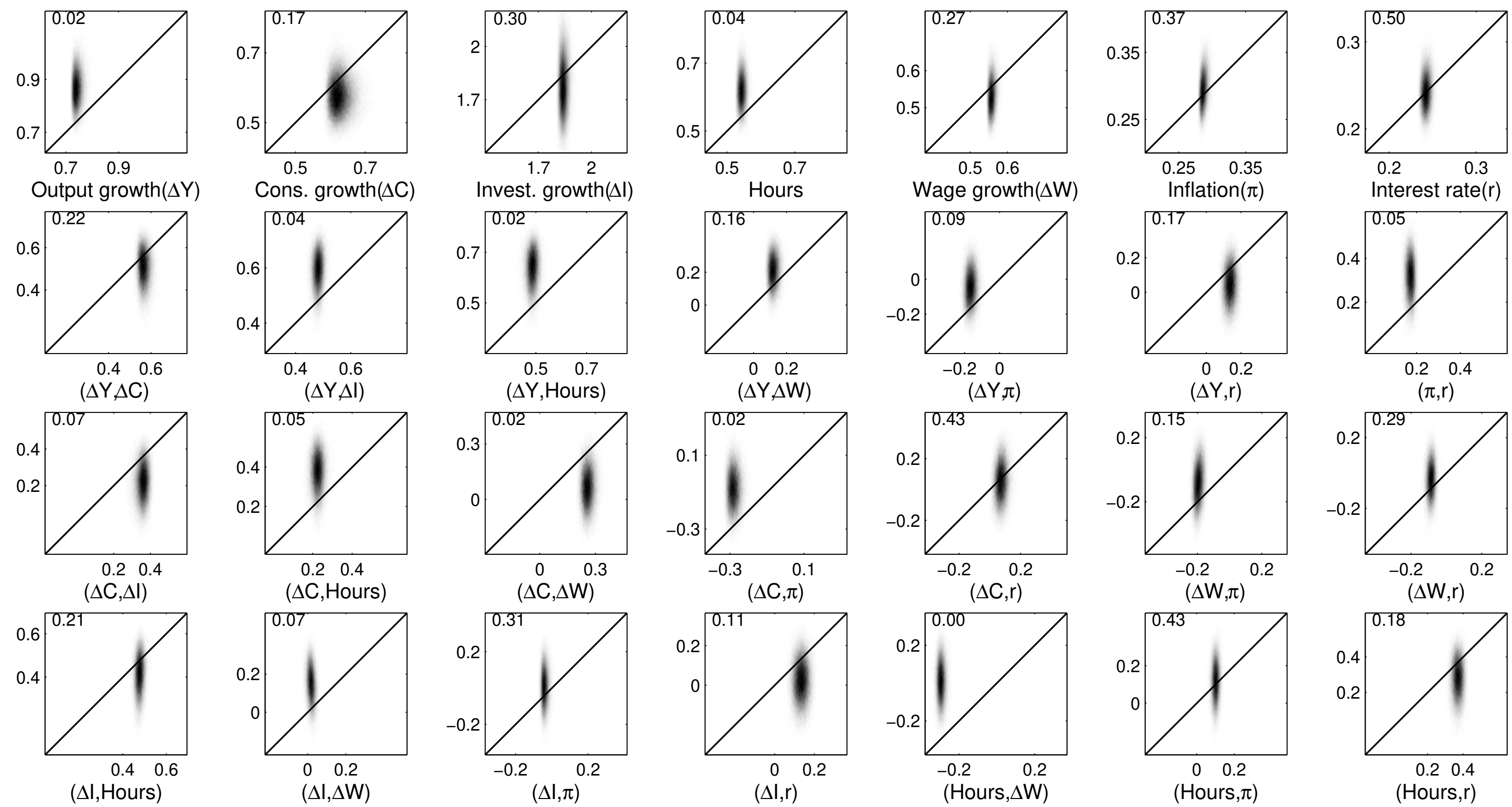

Figure 3: Ω matrix. Scatter plots for the standard deviations and correlations of the FE's, ν_t. First row plots the standard deviations; remaining rows plot the correlations. Horizontal axis plots the posterior values for the realized sample; vertical axis plots the posterior predictive values. The number in the upper left gives the smaller share of points on either side of the 45 degree line.

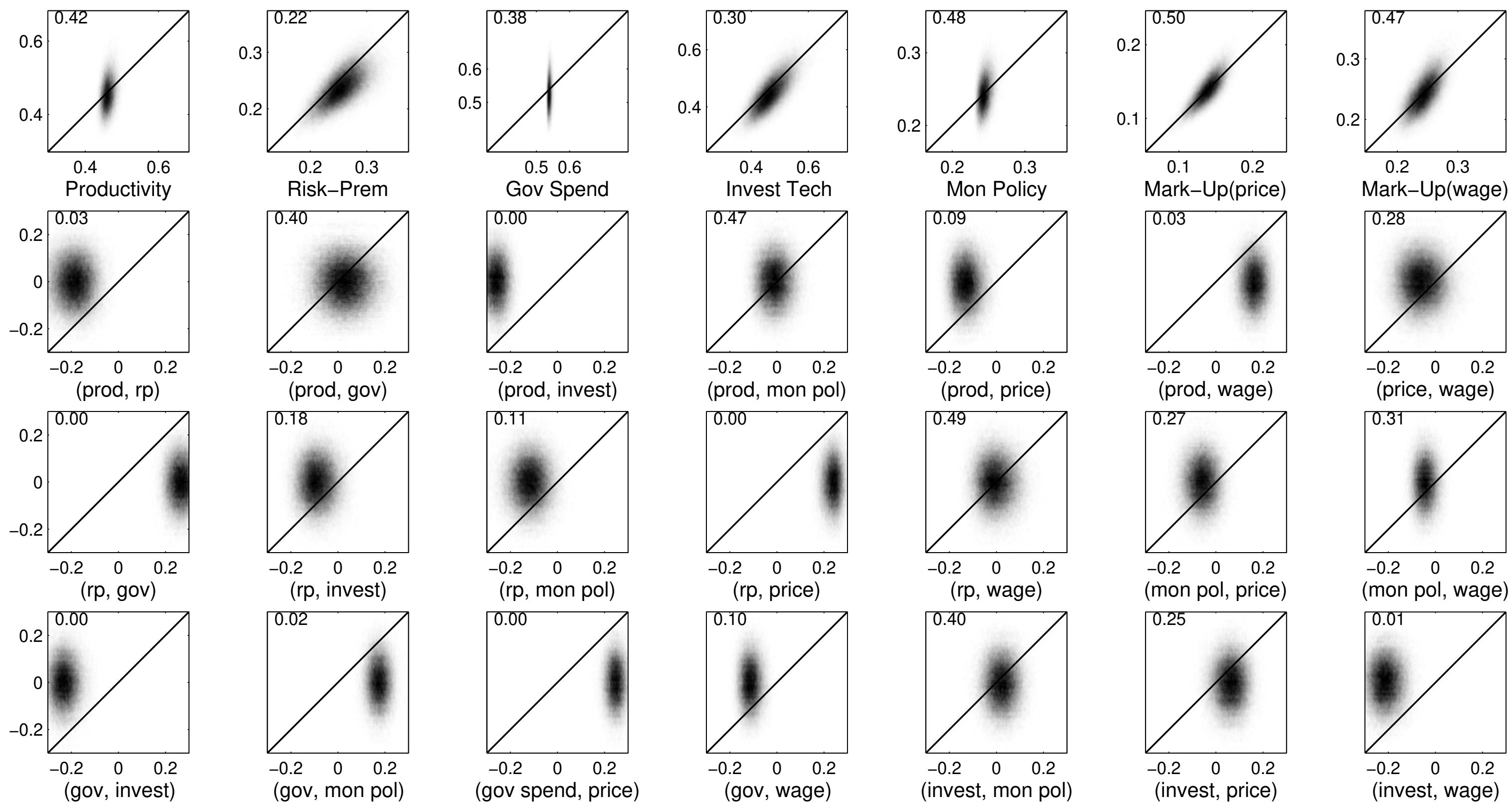

Figure 4: Σ matrix. Scatter plots for the standard deviations and correlations of the structural shocks, $\varepsilon_{t|t}$. First row plots the standard deviations; remaining rows plot the correlations. Horizontal axis plots the posterior values for the realized sample; vertical axis plots the posterior predictive values. The number in the upper left gives the smaller share of points on either side of the 45 degree line.

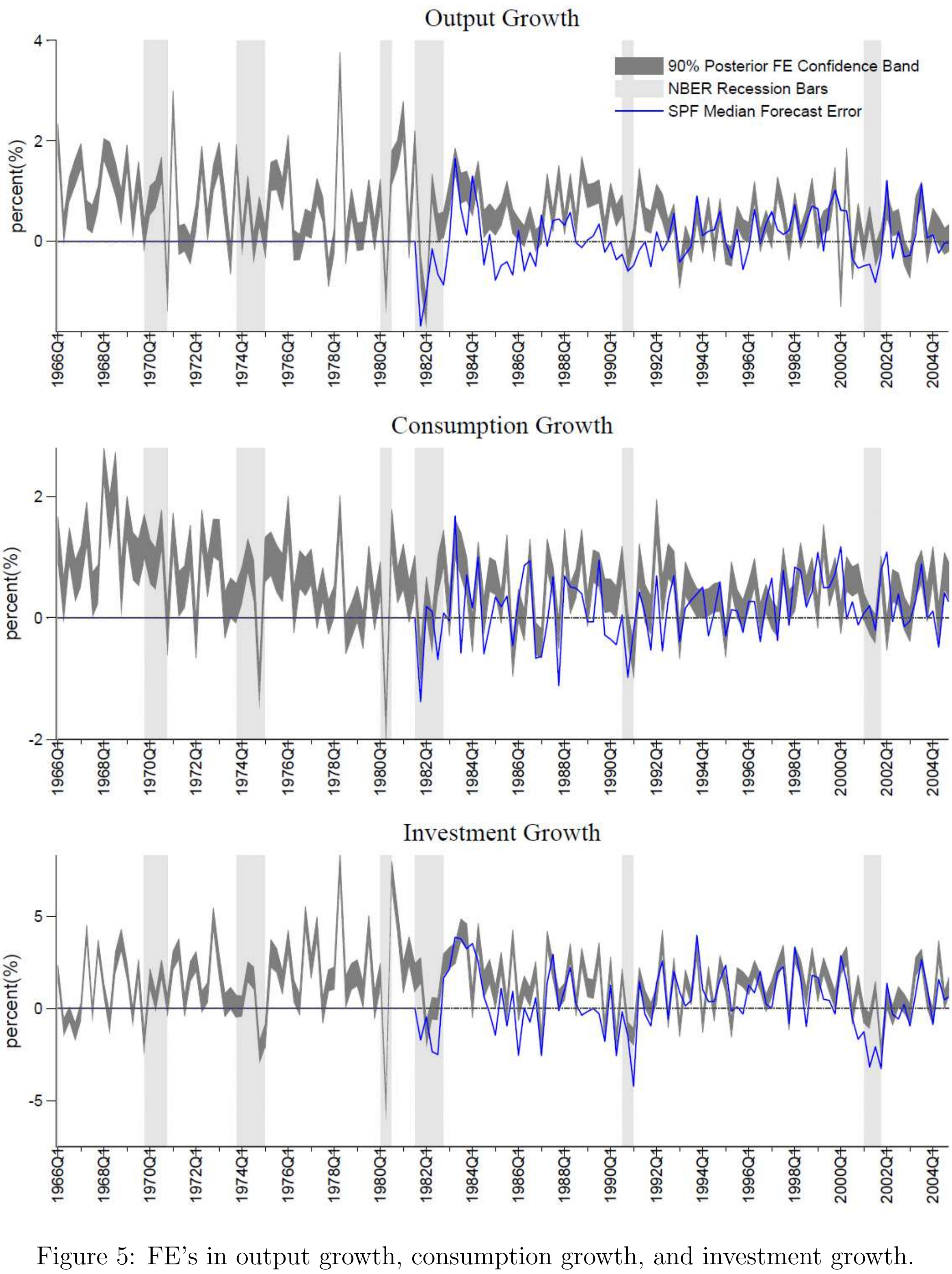

Figure 5: FE's in output growth, consumption growth, and investment growth.

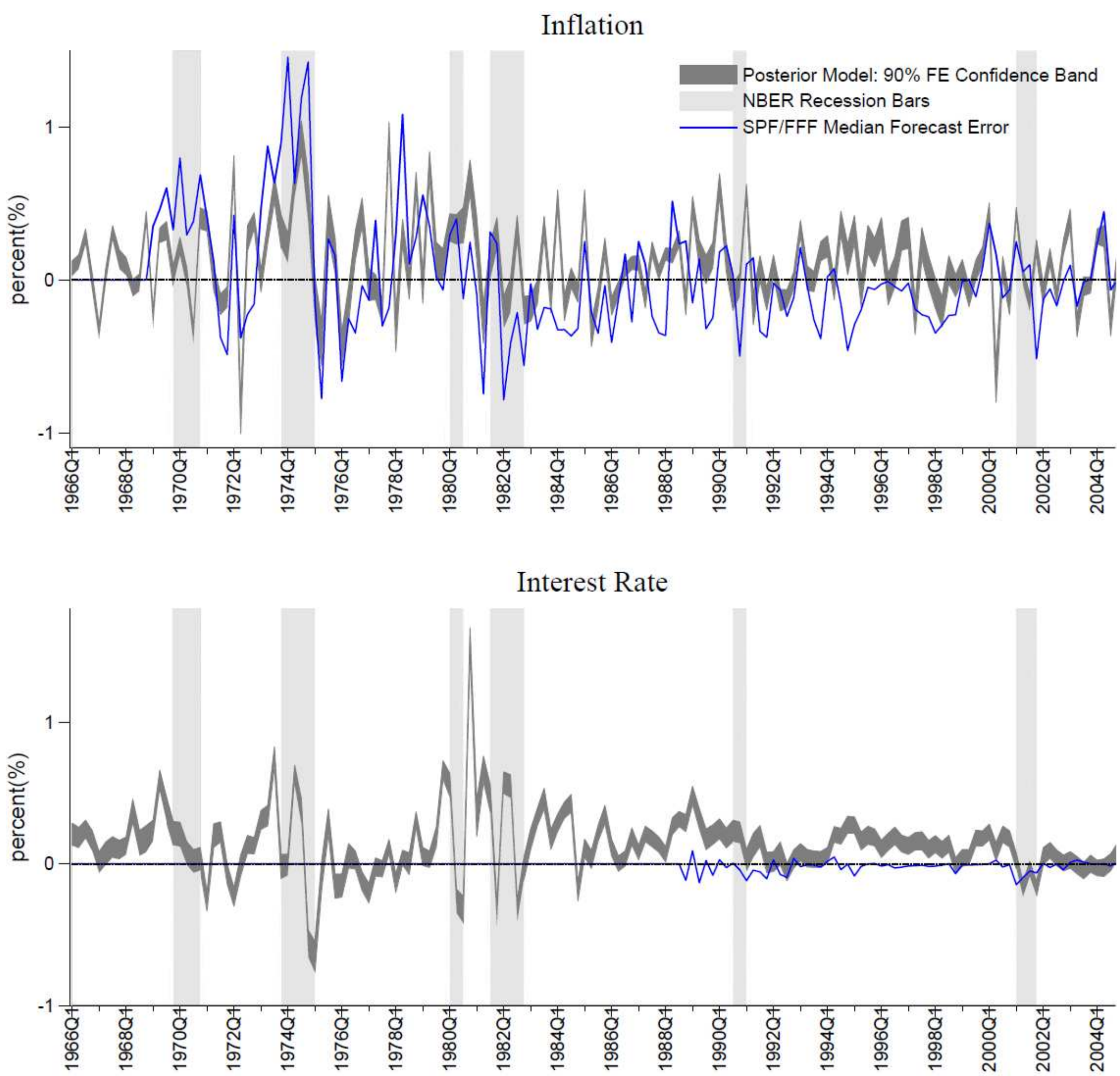

Figure 6: FE's in inflation and interest rate.

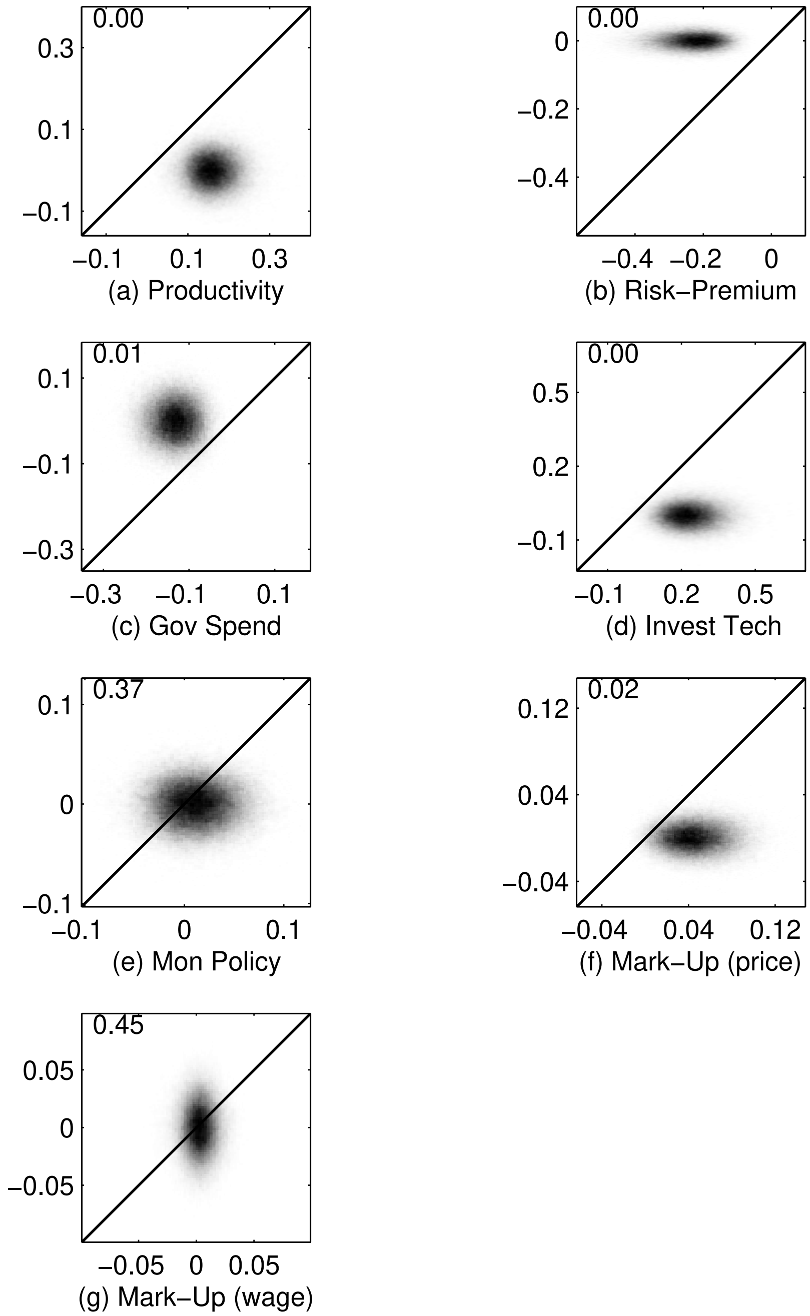

Figure 7: Scatter plots for the mean value of the updated structural shocks, $\varepsilon_{t|t}$. Horizontal axis plots the posterior values for the realized sample; vertical axis plots the posterior predictive values. The number in the upper left gives the smaller share of points on either side of the 45 degree line.