EVALUATING DSGE MODELS FOR MONETARY AND FISCAL POLICY ANALYSIS

by

Abhishek Gupta

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

October, 2010

Abstract

This dissertation evaluates dynamic stochastic general equilibrium (DSGE) models that are widely used for policy analysis at central banks and other policy institutions. DSGE models, like all economic models, abstract in many ways from reality and are misspecified along certain known and unknown dimensions. This dissertation adapts the posterior predictive analysis, well-known in the statistics literature, to highlight the strengths and weaknesses of DSGE models as they relate to the intended task of policy analysis.

The first chapter provides a motivation and introduction to the thesis. In the second chapter we adapt the tools of prior and posterior predictive analysis to the DSGE context and illustrate their usefulness in highlighting the discrepancies between the model and the realized sample. We argue that standard criticisms of prior and posterior predictive analysis, whatever their merits in other contexts, miss the point in the DSGE context. We illustrate that posterior predictive analysis, in particular, can be useful for DSGE model evaluation and it can be viewed as a natural pragmatic Bayesian response to a murky modelling problem. In the third chapter, we apply this framework to evaluate a DSGE model for the task of monetary policy analysis. 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 (first and second moments) of the one-step ahead forecast errors is consistent with the structure observed on the realized sample. We show that in practice, in order to reconcile the U.S. marco dataset with the iconic Smets-Wouters model, we need that the observed sample must involve a highly unlikely sample correlation of structural shocks that are assumed to be uncorrelated in the model.

The fourth chapter is an empirical exercise to shed light on fiscal policy effectiveness at the zero bound for interest rates. We estimate the fiscal policy multipliers for taxes and spending in Japan both before and after the economy hit the zero lower bound in the mid nineties using a tax-code based structural VAR identification methodology. The exercise is an attempt to see if there is enough information in the data to resolve whether the fiscal policy operates differently at the zero lower bound and finds some useful results, but mainly concludes that the data are not sufficiently informative to resolve the issue using these methods.

Keywords:	Bayesian Analysis; DSGE Model Evaluation; Forecast Errors;
	Monetary Policy; Fiscal Policy; Zero Lower Bound; Structural VAR
JEL Classification:	C11, C5, E1, E5, E62, H30
Advisors :	Professor Jon W. Faust
	Professor Jonathan Wright

Acknowledgements

Prof. Jon Faust has contributed in numerous ways towards the development of this thesis. He guided me on a project about how central bank policymaking is crucially dependent on the deviation of new information from prior expectations. This research path led me to develop an entire thesis that benefitted, time and again, from his valuable insights and knowledge. Not only has his support been instrumental in the culmination of this project, but his mentoring has been influential in my transition from a graduate student to that of an economist. I could not have asked for a better adviser.

Prof. Jonathan Wright served as an ideal second adviser as he helped me develop and strengthen my econometric skills. He was always available for a quick chat in his office and his feedback was very useful in extending my thesis in a significant way. In addition, I served as a teaching assistant under Prof. Louis Maccini and learned a great deal on how to be an effective teacher. Prof. Bob Barbera provided guidance and encouragement during a difficult job market year.

In my daily work I benefitted immensely from interactions with fellow students. I enjoyed the company of my peers and friends in the economics department (Daniel Stone, Yonatan Ben-Shalom, Haomiao-Yu, Su-Hsin Chang, Sandeep Mazumder, Damiano Sandri, Silvio Contessi, Viplav Saini, Wei Xiao, Kevin Thom, Guofang Huang) over numerous lunches, parties and poker nights that made my six years more memorable and fun.

My long-time friend Raghav Virmani helped me edit multiple drafts of my job market paper. I thank Michele Julliard for answering numerous questions related to DYNARE code and Frank Smets and Rafael Wouters for providing the matlab codes and posterior draws for their DSGE model.

Besides economics, I was blessed to have wonderful house-mates in Yi-Kai Wu, Thuy-Co Hoang, Isabel Martinez, and Kiran Khaira. These guys helped me maintain my sanity by keeping a lively and humorous atmosphere at home. My Buddhist friends (Kwangsig Lee, Clint Tearnan, Guang William, Charlene Gaskins, Daniel Kunimoto) provided the necessary spiritual nourishment for developing a strong state of mind to cope with the challenges of graduate student life.

I dedicate this thesis to my parents, and my brother, Gaurav. Without their support and encouragement I could not have come so far.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

One of the main focus of the macroeconomics literature in recent decades has been to develop a dynamic stochastic general equilibrium (DSGE) model that can provide better guidance on the short-run and medium-run dynamics to aid our understanding of various economic forces at work in the economy. While there is some broad consensus on long run dynamics it can be safely said that macroeconomists widely disagree about the short to medium-run dynamics and the effects of monetary and fiscal policies over these horizons. Despite the modeling efforts of the last few decades, people have had serious doubts about the efficiency of the current set of macro-models and this was highlighted during the recent financial crisis when these models were caught off-guard as they were completely unprepared to explain the crisis in a meaningful way. Given that the current set of macro models exhibit some known and some yet unknown deficiencies, and that there is ongoing work to improve upon the existing DSGE models to aid in the policymaking process, the question that we raise and answer in this thesis is: what are the strengths and weaknesses of these models? This thesis develops tools to evaluate the adequacy of these macro models for the purpose of policymaking. This debate over the adequacy of DSGE models recently rose to the level of a Congressional hearing.

Chari (2010) raised the question:

> 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.

Colander (2010) made a recommendation about how the economics profession could do better in the future:

> ...encourage the development of a group of economists who specialize in interpreting models and applying models to the real world. The development of such a group would go a long way towards placing the necessary warning labels on models, making it less likely that fiascos, such as the recent financial crisis would happen again.

This thesis provides tools to aid in the process that Colander argues for. In particular, we provide new tools for evaluating the adequacy of DSGE models for the purpose of policymaking to check the consistency of DSGE models with regards to the realized sample. In the second chapter of this thesis, we develop tools for assessing how DSGE models can be used to inform expert judgement in the policymaking process. This chapter puts forth the applicability of prior and particularly posterior predictive analysis in assessing the strengths and weaknesses of the DSGE models. The key idea here is to evaluate the model implied posterior predictive density for any feature with the realized value for that feature and that allows us to check whether the realized sample is treated collectively as an outlier with regards to the predictive density. In doing so this chapter provides some well chosen illustrations such as the finding that these DSGE models treat all postwar recessions as essentially black-swan events that these models are incapable of generating. Lastly this chapter provides defense for the use of posterior predictive analysis against the standard criticisms and argues that these miss the point in the DSGE context.

In the third chapter we use the tool of posterior predictive analysis to evaluate the usefulness of DSGE models for the specific task of monetary policy analysis. This chapter makes the contribution of characterizing the interrelationships among the one-step ahead forecast errors as being essential to the monetary policy process. We show that in order for a DSGE model to be effective in the monetary policymaking process, it is essential that the model not only match the forecast accuracy of the observed forecast errors but more importantly also be able to generate similar cross correlations among the observed one-step ahead forecast errors. This chapter then provides a diagnostic tool for assessing the misspecifications in the model in case there is a discrepancy between the model and the realized sample on the policy relevant features of the structure of one-step ahead forecast errors. This diagnosis, we show, highlights structural inconsistencies in the model that in a simple model of AS/AD is tantamount to interpreting a demand shock as a supply shock. The main finding of this paper is that if the posterior value of a particular forecast error correlation on the realized sample is not what is typically produced by the model, then the model must need multiple shocks working in concert to get the job done, implying a non-zero cross-correlation among the smoothed structural shocks in the model: a gross violation of the model assumption.

The fourth chapter is an empirical exercise to shed light on fiscal policy effectiveness at the zero bound. We test if the effects of fiscal policy have changed in Japan after the economy hit the zero lower bound in the mid nineties using a tax-code based structural VAR identification methodology. We estimate fiscal policy multipliers for taxes and spending separately for two sample periods—one where the interest rates are not bound at zero (prior to 1995) and one where they are at the zero bound (post 1995). The exercise is an attempt to see if there is enough information in the data to resolve whether the fiscal policy operates differently at the zero lower bound and finds some useful results, but mainly concludes that the data are not sufficiently informative to resolve the issue using these methods.

Chapter 2

Posterior Predictive Analysis for Evaluating DSGE Models

2.1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have come a long way. After Kydland and Prescott (1982) demonstrated that a small DSGE model could match a few simple features of the macro dataset, there ensued a 20 year research program of adding both complexity to the model and data features to account for, and the models gradually began to approximate the richness of the macroeconomy.

A watershed event came in 2003 when Smets and Wouters (2003) demonstrated that the family of DSGE models had reached the point that it 'fit' seven key variables about as well as some conventional benchmarks. They explain their main contribution:

> [Our results] suggests that the current generation of DSGE models with sticky prices and wages is sufficiently rich to capture the time-series properties of the data, as long as a sufficient number of structural shocks is considered. These

> models can therefore provide a useful tool for monetary policy analysis in an empirically plausible setup. (2003, p.1125)

Since Smets and Wouter's demonstration, central banks around the world have rapidly been building DSGE models and pressing them into the service of monetary policymaking.

Not everyone agrees, however, that fitting a few variables as well as standard benchmarks is a sufficient condition for the models to be ready for practical policy work at central banks. For example, Sims (1980) famous critique of the models of the 1970s was more or less that the 70s models were highly problematic *despite* their impressive fit. Further, while the current DSGE models are large-scale models by standards of what is possible to solve and manipulate, they are unquestionably small relative to what would be ideal—standard models do not break out consumer durables, inventories, or housing, and have trivial financial sectors.

Perhaps surprisingly, the debate over the adequacy of DSGE models recently rose to the level of a Congressional hearing. Solow (2010) argued that the models were deeply deficient:

> The national—not to mention the world—economy is unbelievably complicated, and its nature is usually changing underneath us. So there is no chance that anyone will ever get it quite right, once and for all. Economic theory is always and inevitably too simple; that can not be helped. But it is all the more important to keep pointing out foolishness wherever it appears. Especially when it comes to matters as important as macroeconomics, a mainstream economist like me insists that every proposition must pass the smell test: does this really make sense? I do not think that the currently popular DSGE models pass the smell test. (p.1)

Chari (2010), (p.7), agreed that "We do not fully understand the sources of the various shocks that buffet the economy over the business cycle," but testified that "[Policy advice from DSGE models] is one ingredient, and a very useful ingredient, in policy making."

Of course, these views reflect a longstanding schism in macro. We seek to avoid this argument over whether the glass is essentially empty or nearly full. To push this tired metaphor to the breaking point, we argue that rather than focusing on how full the glass is, we should focus on just what liquid we are being asked to swallow.

That is, we should focus on what are the particular strengths and weaknesses of the models as they pertain to the intended task—monetary policymaking. Like Tiao and Xu (1993), (p.640), we argue for "...development of diagnostic tools with a greater emphasis on assessing the usefulness of an assumed model for specific purposes at hand rather than on whether the model is true."[1]

We believe that this perspective is particularly important when discussing models for active use in the monetary policymaking process. Policy must be made. In lieu of a formal, coherent model of important general equilibrium effects, the policymaking process employs an implicit model consisting of sector experts and general equilibrium experts hashing out the issues in (long and painful) meetings. The important issue is not some overall quality judgement about the models, but assessing in what ways the models could render the current policymaking process more reliable.

No one, we believe, is asserting DSGE models have reached the point that policy can be placed more or less on model-based autopilot. Thus, we need tools for assessing how DSGE models can best be used to inform expert judgement in the policymaking process.

This chapter supports three claims:

First, while DSGE modelling has come a long way, there remains room for im-

[1]see also, e.g., Hansen (2005).

provement in areas that are materially important for policymaking. We think this claim should be entirely uncontroversial but documenting some particulars provides a basis for the later illustration of our proposed inference tools.

Second, prior and particularly posterior predictive analysis can be valuable tools in assessing strengths and weaknesses of the DSGE models. Prior and posterior predictive analysis were popularized by Box (1980) and have been extended in many ways. Our contribution is to adapt these tools to the DSGE context and illustrate their usefulness. Most notably, perhaps, we adapt the discrepancy analysis of Gelman, Meng, and Stern (1996), which seems to have gotten little use in macro, to analyze the causal channels in DSGE models.

While prior and posterior predictive analysis has been widely used in many areas, these techniques have also been criticized as being inconsistent with coherent inference. Our third goal is to argue that standard criticisms of prior and posterior predictive analysis, whatever their merits in other contexts, miss the point in the DSGE context. In particular, posterior predictive analysis can be viewed as a natural pragmatic Bayesian response to a murky modelling problem.

These three points come in the next three sections. Along the way we discuss and illustrate various aspects of implementing prior and posterior predictive analysis using a version of the iconic Smets Wouters DSGE model. Smets and Wouters have been incredibly gracious in helping us complete this work.

2.2 Standard Approaches to Bayesian DSGE Modelling for Monetary Policymaking

2.2.1 The Macro Modelling Problem

We will take largely as given a basic knowledge of the monetary policymaking process and of DSGE modelling for use in the policymaking process. We provide the following sketch. The policymakers of the central bank meet periodically to assess changes to the values of the policy instruments under their control—the policy interest rates and other policy tools. Taking as given the view adopted at the last meeting, a major task at this meeting is to process the information that has arrived since the last meeting.

This processing is complicated by a number of features of the problem. The potentially relevant information set is high dimensional. While there is no consensus on how high-dimensional, the ongoing policymaking process closely monitors hundreds of variables.[2] Certain research supports the view that forecasts of relevant variables based on datasets of, say, 70 or more variables outperform others[3] and that the subjective process that focuses even on broader information in some cases does even better (Faust and Wright (2009)). Further, there is a rich set of dynamic feedbacks among the myriad potentially relevant variables. By general consensus, general equilibrium effects involving the expectations of a

---

[2]For example, see the Federal Reserve's policy meeting briefing materials that are now available at *http : //www.federalreserve.gov/monetarypolicy/fomc$_h$istorical.htm*

[3]e.g., Bernanke, Boivin, and Eliasz (2005); Bernanke and Boivin (2003); Faust and Wright (2009); Forni, Hallin, Lippi, and Reichlin (2005); Giannone, Reichlin, and Sala (2004); Stock and Watson (1999, 2002, 2003, 2005).

large set of heterogeneous agents may be central to the policymaking problem.[4]

The existing policy process relies on many more and less formal tools, but is ultimately heavily judgmental. The goal in DSGE modelling is to build a new model to aid in the process of interpreting incoming data, forecasting, and simulation of alternative policies.

From a modelling standpoint, forecasting does not strictly require a structural model (that is, a model with explicit causal channels). The simulation of likely outcomes under alternative counterfactual policy assumptions does require a structural model, as does the attribution of unexpected movements in incoming data to particular causes. A textbook example of the latter is attempting to sort out whether a recent unexpected rise in GDP growth is primarily due to supply or demand shocks. As these two causes may have different implications for policy, much policy analysis involves this sort of inference.

This modelling problem is made more challenging by the fact that the relevant body of theory provides only limited guidance on short-run and medium-run dynamics. While theory gives us broad guidance on overall dynamics, small adjustment costs, rule of thumb behavior, and similar effects can dominate shorter run dynamics.

Ideally, we need a model with fully articulated causal structure of a very large and complicated system where theory provides limited guidance on important aspects of the dynamics. The final complication is that we have a single historical sample for the process we are modelling, the world macroeconomy. The models will be specified and refined on this familiar dataset. New information arrives only with the passage of time. By general

[4]For example, policy is often described as expectations management. See, e.g., Bernanke, Reinhart, and Sack (2005).

consensus, the historical sample is not large enough to definitively resolve all important issues. The ongoing lack of consensus on basic questions in macro is clear testament to the idea that our only available sample is not resoundingly informative on all relevant policymaking issues.

2.2.2 Description of DSGE Modelling and the SW Model

The approach in DSGE modelling is to explicitly state the decision problems of groups of agents—e.g., households and firms—including objective functions and constraints. Together with the adding up constraints on the economy as a whole, gives a complete system. For example, households choose consumption (hence, saving) and labor supply to maximize a utility function. Firms choose investment and labor input to maximize a profit function. Ultimately the solutions these individual optimization problems, when combined with overall resource and adding up constraints, imply the dynamic behavior of the variables modelled. As noted below, given the complexity of the models, we generally end up working with some approximation to the full model-implied dynamics—most often a log-linear approximation to the deviations from some steady-state implied by the model.

A key feature of this kind of modelling is that agents are forward looking. In the simplest forms of the decision problems, forward looking agents immediately react to news about future conditions, and adjust their behavior much more quickly than is consistent with the macroeconomic data. Thus, 'frictions' are added to the decision problems in order to slow down what would otherwise be excessively jumpy behavior by agents.

Our example model is a version of the SW model. In particular, we use the model described in Smets and Wouters (2007).[5] The model is an extension of a standard closed economy DSGE model with sticky wages and sticky prices, largely based on Christiano, Eichenbaum, and Evans (2005). The model explains seven observables, with quantity variables in real, per capita terms: GDP growth, consumption growth, and investment growth, hours worked, inflation, real wage growth, and a short-term nominal interest rate.

The model introduces a rich set of frictions in the decision problems of agents—the Calvo friction for prices and wages, habit formation in consumption, and investment adjustment costs. The seven structural shocks are assumed to follow exogenous, independent persistent (autoregressive of order 1) processes and are interpreted as shocks to overall productivity, investment productivity, a risk premium, government spending, wage mark-up, price mark-up, and the monetary policy interest rate rule.

Because we will focus on consumption as an example later, it is worth going into a bit more detail on the consumption problem. Consumers are assumed to maximize the expected value of a discounted sum of period utility given by,

$$U_t = \left(\frac{1}{1-\sigma_c}(C_t - hC_{t-1})^{1-\sigma_c}\right) exp\left(\frac{\sigma_c - 1}{1+\sigma_l} L_t^{(1+\sigma_l)}\right)$$

where C_t is consumption at time t, L_t is labor hours at t and H, σ_c and σ_l are scalar parameters. The consumption portion of the utility function has habit persistence parameterized by h and risk aversion parameterized by σ_c, the coefficient of relative risk aversion (CRRA). Larger values of h tend to imply that agents will be more reluctant to change consumption.

[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.

2.2.3 Estimation

Explicitly Bayesian inference methods are the norm in this area. The methods used are, at a general level, a straightforward application of what we will call the plain vanilla Bayesian approach.[6] In a nutshell, this approach requires data, a parameterized model, and a joint prior distribution for the parameters of the model. The model implies a likelihood function for the data, and the model and prior together exhaustively characterize the state of knowledge of the researcher before the new data arrive. The plain vanilla Bayesian scheme tells us how to update the view reflected in the prior density in light of new information that arrives.

Specifically, call the data Y and the model M_θ, with parameter vector $\theta \in \Theta$. In the DSGE case, the likelihood, $L(\theta|Y)$, is implied by the specification of the economic model. Once one has specified the model and processes for the exogenous driving processes, the decision problems of the agents can be solved and this solution implies a likelihood function. Generally for computational reasons we use a log-linear approximation to the exact solution of the model where the approximation is centered on a non-stochastic steady-state of the model. This approximation gives rise to a linear, ARMA structure for the dynamics of the model.

The data tend to be taken from the standard macro data set; the sample period tends to be the longest continuous sample for which data are available and for which the structure of the economy was reasonably stable. This latter is a bit of a judgement call. SW estimate the model using US data from 1966Q1 to 2004Q4.

---

[6]Plain vanilla here describes the application of Bayesian principles to a plain vanilla context. This does not tell us what Bayesian principles imply in a richer context such as—we argue—the one described here.

Often there are multiple choices to be made for choosing model-based analogs to key quantities in the model. For example, analysts sometimes use only nondurables plus services consumption as the measure of consumption, since the model does not have durables. SW's consumption measure includes durables.

The prior, p_0 is a joint density for θ over the parameter space Θ. The conventional prior used in DSGE modelling varies, but in general terms the formal prior is often specified as a set of marginal distributions for each individual parameter. These are taken to be independent, implying the joint distribution for the prior. Generally, some natural support for each parameter is implied by economic principles, technical stability conditions for the model, and/or earlier applied work. The prior is specified to be fairly dispersed over this support. Through trial and error, the analyst may find regions of the parameter space in which the model seems ill-behaved in some way, and the support is narrowed.

While the papers often have arguments justifying the support of the prior and, perhaps, where it is centered, we have seen no argument that the joint prior implied by the independent specification of marginal priors for the parameters has any justification or has any tendency to produce results that are consistent with any subjective prior beliefs.[7]

Given the state of knowledge reflected in the model+prior and the new information in the sample, a straightforward application of Bayes' law gives us,

$$p_1(\theta) \equiv pr(\theta|Y^\tau) = \kappa p_0(\theta) L(Y^\tau|\theta)$$

where κ is the constant that makes the integral of the expression on the right (with respect to θ) integrate to one.

[7]In some cases some span of data at the beginning of the available sample is used as a 'training sample' so that the prior is tuned to reflect that dataset. We take this up below.

The prior and posterior densities for the two key consumption parameters are shown in Figure 2.1. Both posteriors are centered at about the same place as the prior, but the posteriors are considerably more peaked, indicating that that data are somewhat informative. The habit persistence parameter is fairly large, suggesting that agents are highly averse to changing consumption; the CRRA parameter is in the range that has become conventional for estimates of models like this on this sample.

2.2.4 Material Deficiencies, Omissions and Coarse Approximations

Given the size of the system being modelled and the current stage of understanding of the relevant mechanisms and modelling techniques and related algorithms, it remains the case that existing DSGE models involve coarse approximation to some economic mechanisms believed relevant for policymaking and omit other such mechanisms entirely. This is meant to be a description of the current state of development, not a criticism. To motivate the remainder of this chapter, it is useful to provide some detail on the state of modelling.

Consider omitted mechanisms and phenomena. Most standard DSGE models do not separately treat durable goods, inventories, or housing, despite conventional wisdom that these items play an important role in business cycles. Many experts believe that credit spreads have important predictive content that might be important for policymaking (e.g., Gilchrist, Yankov, and Zakrajsek (2009)), but defaultable debt is not modelled. Indeed, until the crisis, the financial sector of these models was entirely trivial. This list of omissions could obviously go much longer.

It is also true that the modelling of phenomena included in the model is often best viewed as a coarse approximation relative to the best knowledge of specialists in the particular area.

For example, important aspects of individual behavior toward risk is parameterized by the coefficient of relative risk aversion. In the best tradition of microfounded modelling, we might ask experts in individual behavior toward risk what values for this parameter might be appropriate. Unfortunately, expert opinion is overwhelmingly clear on one point: individual behavior toward risk is a rich phenomenon not well captured by this single parameter.[8] Many micro phenomena simply cannot be accounted for under this assumption. Suppose we tell the expert we are viewing this as a representative agent approximation to underlying behavior, but would still like guidance on the value. The expert should then remind us that different values will be best depending on the goals of the approximation: to 'fit' the equity premium from 1889 to 1978, CRRA > 10 (Mehra and Prescott (1985)); to 'fit' aggregate lottery revenue probably requires risk loving; to fit the reaction of consumption to changes in monetary policy, probably some value not too far from one is appropriate. As a description of individual behavior, CRRA specification is a crude approximation. The choice of parameter value should be based mainly on how best to center the approximation for a particular purpose.

Analogous issues can be raised about the treatment of habits. At the individual level there is little evidence for strong habits (Dynan (2000)), but strong habits in the model seem to be needed to 'fit' the smooth evolution of aggregate consumption data. Alternative explanations for the aggregate persistence include serially correlated measurement error (Wilcox (1992)) aggregation biases (Attanasio and Weber (1993)), and 'sticky expectations' (Carroll, Slacalek, and Sommer (forthcoming)).

[8]For a good review, see Camerer (1995).

We could do a similar analysis in these models of the labor market, investment, and the financial sector. Our goal is to provide some concrete meaning to the statement that many arguably relevant mechanisms are omitted, many included mechanisms are coarse approximations.

Finally, it also the case that the prior used in these analysis is far from the idealized case in which the model+prior fully reflects the subjective views of the relevant analysts. As Del Negro and Schorfheide (2008) note, there is no reason to suppose that taking reasonable marginal priors for the parameters and treating these as independent will lead to reasonable general equilibrium implications for the model as a whole. As we shall see below, it is not difficult to find examples where the prior is highly informative and at odds with conventional wisdom.

Rather than focusing on particulars, however, our point about that material deficiencies remain is probably best illustrated by the revealed preferences of modelers at central banks. The models remain in a state of substantial ongoing refinement and revision.

2.3 Prior and Posterior Predictive Analysis

The plain vanilla scheme described above tells us how optimally to shift our views on the relative plausibility of different parameter values $\theta \in \Theta$. But it can never cast doubt on whether the model as whole, $M_\theta, \theta \in \Theta$, is adequate. In any context, this is potentially troubling—George Box famously reminds us, all models are wrong–but it is particularly troubling in a context where we are using an *ad hoc* prior over a model with materially important aspects of approximation error and omission.

Box popularized a family of tools for checking whether an admittedly wrong model might be useful based on prior and posterior predictive analysis. Box's ideas have been elaborated in a number of ways in the statistics and economics literature.[9]

In this section, we take as given the conventional pragmatic arguments in favor of prior and posterior predictive analysis and illustrate the way it could be used to highlight strengths and weaknesses of DSGE models.[10] The basic analytics of prior and posterior predictive analysis are all well-established in the statistics literature. Our contribution is to adapt these tools in ways particularly useful in DSGE work.

2.3.1 Prior and Posterior Predictive Analysis Defined

Predictive analysis relies on simple idea: if the available sample is too freakish from the standpoint of the model+prior or model+posterior, then perhaps the model or prior should be refined.[11]

The essence of the argument can be seen in a simple example. You are attempting to do inference on the probability of observing a draw greater than 3 from some random variable. The model states that the sample is independent and identically distributed (iid) draws from a Gaussian with unknown mean and variance one. The prior for the mean is

---

[9]For example, Geweke (2005, 2007, 2010); Bernardo (1999); Gelman, Meng, and Stern (1996); Lancaster (2004).

[10]It might seem more natural to give the theoretical justification before the applications. Our theoretical arguments, however, turn on unique practical aspects of the DSGE context that are best discussed after seeing some concrete examples.

[11]It might seem most natural to change the model, but in cases like DSGE modelling where the prior is substantially arbitrary, it is not unnatural to think of deciding that the arbitrary choice of prior had put mass in 'the wrong place.'

uniform on $[0, 1]$. You obtain a sample of 50 observations and notice that the histogram of these observations is highly skewed to the right (Figure 2.2, panel (a)). This sample would be very unlikely to arise if the outcomes were indeed iid Gaussian.

The generic idea of predictive analysis is that one might want to reconsider the model+prior at this point, but two additional ideas are worth emphasizing. The right skewness may be materially important to the task at hand, since right skewness could have a large effect on the probability of observing greater than 3. Thus, one can focus on *relevant* features. Second, one sensible option would be to obtain another sample. But we are focusing on the case in large-scale, general equilibrium macroeconomics—new information will only arrive slowly.

Predictive analysis provides formal tools for judging the degree to which relevant features of a sample are freakish from the standpoint of the model+prior. By *feature*, we mean any well-behaved function of the data: $h(Y)$.[12] Following the spirit in much macroeconomics one might think of these as empirical measures corresponding to some 'stylized fact.' In the example just given, it would be natural to use the sample skewness as the data feature.[13] The sample skewness for the example sample is 1.05, whereas the population skewness of any Gaussian is zero. However, one might wonder how likely one would be to observe a *sample skewness* of 1.05 in a sample of size 50 from the model+prior at hand.

---

[12]Box called these *model checking functions*.

[13]

$$h(Y) = \sum_{t=1}^{T}(y_t - \bar{y})^3/(\sum_{t=1}^{T}(y_t - \bar{y})^2)^{3/2}$$

where y_t is the t^{th} observation and $\bar{y}$ is the sample mean.

The model+prior imply a marginal distribution for any $h(Y^{rep})$, where Y^{rep} is a sample of the size at hand drawn according to the model+prior:

$$F_h(c) \equiv \mathrm{pr}(h(Y^{rep}) \leq c) \tag{2.1}$$

Define Y^r to be the realized sample. One can plot the implied density, $f_h(x)$, along with the realized value $h(Y^r)$ on the sample get a sense of whether the realized value is freakish. Our example model+prior can indeed produce samples with large positive and negative values for the sample skewness, but would do so very rarely—the sample value of 1.05 is far in the tail of the predictive distribution.

Where large values are considered unlikely, Box suggested a prior predictive p-value defined as,

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

This is the probability of observing $h(Y)$ greater than the realized value in repeated sampling if the data were generated by the model+prior. For our example, the p-value is 0.002, or 0.2 percent.

There are, of course, dangers in summarizing a distribution with a single number such as a p-value. Such crude summaries should be used with caution, and we will largely report the entire predictive density. Still at times, p-values provide a convenient and compact summary.

We can use the posterior for the parameters of the model, p_1, instead of p_0 in (2.1) in computing the predictive density, to obtain the posterior predictive distribution and posterior predictive p-value. Once again, these predictive densities depict the likelihood of observing specified sample features in repeated sampling from the model+prior or model+posterior.

The data features we have discussed so far are a function of Y alone. In modeling causal channels we are not only interested in description, but in why events happen the way they do. To shed light on causal channels, it is also useful to consider features that are a function of the sample and θ: $h(Y, \theta)$. We'll call the former 'descriptive' features and the latter 'structural' features, to emphasize the dependence of the latter on the structural parameter.

Gelman, Meng, and Stern (1996) have written extensively on what we call structural features.[14] These seem to have received little application in macroeconometrics.

Since structural features depend on the unobserved value of θ, they are a bit more subtle to understand than descriptive features. A symptom of the difficulty is that even after observing the sample, there is no single realized value on the sample at hand. However, conditional on any fixed θ^*, we can compute $h(Y^r, \theta^*)$ and, thus,

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

where Y^{rep} is a random sample of the same size as Y^r drawn according to θ^*. Conditional on θ^*, this corresponds to the $p - value$ computed above. As always we can integrate out the dependence on the unobserved θ^* using the prior or posterior to get, $\mathrm{pr}(h(Y^{rep}, \theta^{rep}) > h(Y^r, \theta^{rep}))$, where θ^{rep} is drawn according to the prior or posterior. This is analogous to the p-value computed above.[15]

[14] Gelman et al. call $h(Y, \theta)$ a *discrepancy variable* or simply *discrepancy*. The idea is that the feature is meant to help detect a discrepancy between the model and sample.

[15] An alternative for dealing with the unobserved θ is to create some summary scalar. We could examine the value at the posterior mode, or the mean value for the feature where the mean is taken with respect to

Gelman et al. suggest the following computational approach, which may aid in understanding the above expression. Focus on the posterior version for concreteness. The model+prior imply a joint distribution for θ and Y. Thus, we can assess $\mathrm{pr}(h(Y^{rep}, \theta^{rep}) > h(Y^{r}, \theta^{rep}))$ by repeating the following steps a large number of times, where on the j^{th} step we,

1. Draw $\theta^{(j)}$ according to the posterior
2. Compute $h(Y^{r}, \theta^{(j)})$
3. Draw $Y^{(j)}$ according to $\theta^{(j)}$
4. Compute $h(Y^{(j)}, \theta^{(j)})$
5. Save the pair $h(Y^{r}, \theta^{(j)}$,$h(Y^{(j)}, \theta^{(j)})$

The marginal distribution of the $h(Y^{r}, \theta^{(j)})$s is a density corresponding to the realized value in descriptive features. The marginal distribution of the $h(Y^{r}, \theta^{(j)})$s is the posterior predictive distribution for this feature.

The scatter plot of $h(Y^{(j)}, \theta^{(j)})$ (on the vertical axis) against $h(Y^{r}, \theta^{(j)})$ (horizontal) will give a sense of the joint distribution of the two items, and the share of points falling above the 45 degree line is an estimate of the p-value described above. This p-value is the probability in repeated sampling from the model+posterior, that we observe a sample generated under a θ_0 for which the h exceeds the h implied by θ_0 on the realized sample.

Obviously, if it is small values that one wishes to detect then, the share of points under the 45 degree line constitutes a p-value. More generally, inspection of the joint

---

the prior or posterior. We discuss how our approach complements these others in the applications below.

distribution will, once again, be more informative than a simple p-value computation.

Note that the predictive p-value for descriptive statistics can be computed using a simplified version of the same algorithm exploiting the fact that h does not depend on θ. Thus, the second step above may be computed outside the loop and the realized density collapses to a point and all we have to plot is the marginal predictive density and the realized point value. These are the algorithms we use in the examples reported below.

2.3.2 Illustrations I: Descriptive Features

In this section, we illustrate how these techniques can be used to discover and highlight strengths and weaknesses of DSGE models using the SW model. This is not intended as a thorough substantive critique of this model; rather, we present examples meant to illustrate the functionality of the methods. We attempt to provide a substantive analysis in other papers (Gupta (2010), Faust and Gupta (2010b), Faust (2009)).

It is useful to keep in mind two forms of analysis that are complementary to what we are advocating: moment matching and full blown likelihood analysis. In traditional moment matching with a DSGE model, one selects (either by estimation or calibration) values for the parameter, θ, and then compares population moments implied by the model to the corresponding sample moments for the sample at hand. In a full-blown Bayesian-inspired likelihood analysis, the emphasis is on comparing models or parameter values based on the relative likelihoods, perhaps, as weighted with the prior. We seek to emphasize how posterior predictive analysis can be a complement to each.

In traditional moment matching as started in the DSGE literature by Kydland and Prescott (1982), one might focus on the some version of the variance covariance matrix of the variables—say standard deviations, correlations, and autocorrelations. In SW for example, the correlation of inflation and output growth at the posterior mode is -0.22, while the corresponding sample correlation is -0.31. It is difficult to assess whether this is a success or failure of the model, in part, because this comparison fails to represent two potential areas of uncertainty. First, summarizing the model only by the correlation at a single θ does not reflect uncertainty in the choice of parameter. Replacing the single value at the posterior mode with the posterior density for θ will bring the uncertainty in θ into the comparison (Figure 2.3, blue dashed line). To emphasize, the blue line is the posterior density for the population correlation implied by θ. Since the sample value is relatively far into the tail of the posterior density, one might once again take this as evidence against the model.

There is a second aspect of uncertainty, however: the sample correlation is not a precise estimate of the population correlation in the underlying process driving the economy. Regardless of the population correlation, the model+prior could, in principle, imply that any sample correlation might be observed in a small sample.

The posterior predictive density tells us what sort of values we would expect to see for the *sample correlation* in repeated sampling with sample size equal to the sample size at hand. It turns out (Figure 2.3, black line) that under the model+posterior, there is nothing particularly freakish about seeing *sample* correlations of -0.31 when the *population* value is -0.22 at the posterior mode.

In this case, it is important to keep the interpretation in mind: the model is consistent with the data essentially because the model implies that the sample correlation will be poorly measured—correlations like that observed in the data are likely to be observed even when the true correlation is quite different.

Posterior predictive analysis provides a way to investigate and highlight known problem areas of the model. For example, the correlation of consumption and investment growth in DSGE models has been a continuing problem. For most countries consumption and investment growth are strongly positively correlated: booms and busts tend to involve both consumption and investment. There are forces in the model, however, that tend to drive this correlation toward zero.[16] In the SW model, the posterior for this correlation is centered on low values (Figure 2.4, panel(a), black solid line), and the sample value over 0.5 is far in the tail of the posterior. In this case, the posterior predictive density is slightly more dispersed (Figure 2.4, panel(b), black solid line), but the p-value remains below 1 percent. Formally, if samples were repeatedly drawn from the model+posterior, less than 1 percent of draws would give values as extreme as that observed on the sample.

In cases where the model+posterior suggest that that the sample at hand is freakish, there are three natural diagnoses: i) strange samples happen, ii) the model needs refinement, and iii) the prior needs refinement. This last possibility, of course, arises particularly when the prior has *ad hoc* elements. To shed some light on this latter possibility we can look at the prior predictive density. If the prior predictive density strongly favors low or negative correlations of investment and consumption growth, then the posterior result could be due to the unfortunate choice of prior. In the current example, however, this is not the case (Figure 2.4, panel(b), blue dashed line). The marginal prior for this sample correlation actually favors large positive correlations.

[16]For example, a productivity shock that raises real interest rates may raise investment but reduce consumption due to the increased incentive to save.

For the SW model, the update using the data overpowered the strong prior and pushed the posterior estimate to the far side of the sample value. This illustrates the complexity of working with large dynamic systems. The θs that give large positive correlations of consumption and investment must have been downweighted by the likelihood because those θs have some other implication that is at odds with the sample. Smets and Wouters included what they call the risk premium shock in their model with the express purpose of boosting the correlation of consumption and investment growth. This shock seems to do the trick in the prior, but not the posterior. We return to this structural issue below.

These examples were intended to illustrate how posterior predictive analysis could complement or extend the sort of moment matching exercises that have been common in the literature.

Of course, defenders of full-blown Bayesian analysis have long criticized moment matching. Looking at a few marginal distributions for individual moments is no substitute for a metric on the whole system, and the likelihood itself is the natural way of summarizing the full implications of the model. Full likelihood analysis may show, for example, that posterior odds favor DSGE model A over DSGE model B. In the analysis suggested by Del Negro, Schorfheide, Smets, and Wouters (2007), one forms a Bayesian comparison of the DSGE model to a general time series model. In this case, one can learn that data shift posterior plausibility mass along a continuum from the fully articulated structural model to the general model with no causal interpretability. This sort of comparison may be very useful as an overall metric on how the model is doing.

We are considering model building for ongoing, real-time policy analysis, however. All the models have material deficiencies and are under ongoing substantial revision. Thus, echoing our second and third main points from the introduction, we argue that in addition to full-blown likelihood analysis it is important systematically to explore the particular strengths and weaknesses of the model salient to the purpose of policymaking. The fact that the posterior shifts mass from one model toward another is not very revealing of the particular strengths and weaknesses of either. We argue that by using a richer set of data features than simple moments, some important aspects of the models can be revealed.

For example, Gupta (2010) argues that an important part of policy analysis at central banks is interpreting surprising movements in the data. Policy at one meeting is set based on anticipated outcomes for the economy. At each successive meeting, policymakers assess how new information has changed the outlook and what this implies for the appropriate stance of policy. In a formal model, this amounts to interpreting the one-step (where a step is one decision making period) ahead forecast errors from the model.

The simplest substantive example of this perspective comes in the textbook aggregate supply/aggregate demand model. If prices and output come in above expectation, one deduces that an unexpected positive shock has shifted AD. If output is higher but prices lower than expected, one deduces that a favorable supply shock has shifted aggregate supply. In the textbook case, the two outcomes have different policy implications.

The key insight Gupta argues for is that policymakers need more than a model that forecasts *well* in some general sense. They need a model that properly captures the joint stochastic structure of the forecast errors. As a simple way to examine the properties of the model in this regard, we can take our descriptive feature to be the correlation of one-step forecast errors out of a benchmark time series model estimated on the sample. For example, one could use a first order vector autoregression (VAR), a Bayesian VAR, or a

VAR with lag length set by the AIC between 0 and 6. All that is required is that based on the sample alone, one can evaluate the value of the feature. For our illustration example, we use the simplest of these, the first-order VAR.

Since output growth is a major focus of policy, we focus for this example on the correlation of the output growth forecast error with the errors for the other 6 variables. The prior and posterior predictive densities along with the sample value are examined in Figure 2.5.

Several notable results can be seen. First, the prior is highly opinionated putting most mass on correlations near one. This illustrates again that although the marginal prior distributions for the parameters are fairly dispersed, the joint implications of the largely *ad hoc* prior for questions of interest in policymaking may be highly concentrated. Indeed, it may be highly concentrated in regions that do not reflect any subjective prior judgements. No expert believes that if we could just forecast output growth properly we could also nail a forecast for hours of work, but this view is reflected in the mass near one in Figure 2.5, panel(c) (blue dashed line).[17]

The realized value of the forecast error correlation is fairly generally far in the tail of the posterior predictive distribution. In particular, the relation between output and inflation (two key policy variables) is problematic. Of course, the literal meaning is that the sample is freakish from the standpoint of the model+posterior. More provocatively, in

---

[17]A complete diagnosis of this fact is beyond the scope here. However, it appears that this is due to the fact that all the shocks enter the prior with the same parameters. Despite being 'the same' in this nominal sense, a given variance shock means something different economically depending on how it enters the model. In this case the result seems to be that the prior is that demand shocks dominate.

practice on the realized sample, policymakers were systematically faced with the problem of interpreting inflation and output growth surprises of opposite signs (negative realized correlation in Figure 2.5, panel(d)). The model+posterior says that this pattern was a freak outcome and policymakers need not worry much about this problem in the future.

This simple example is only illustrative. Policymakers will in practice use a more sophisticated forecasting model. Thus, one might ideally choose a more sophisticated benchmark forecasting model. Or one could analyze the properties of optimal model-consistent forecast errors. Gupta (2010) provides a version of this more complete analysis.

2.3.3 Illustrations II: Structural Features

Ultimately, policymakers must go beyond the descriptive in order to draw inferences about the causes of economic variation and the likely causes of policy responses. Identifying causal structure in macro is very contentious, and when using a large model, the problem is multiplied by the complexity of the system. It is very difficult to look at a model and judge whether the causal structure as a whole is broadly consistent with any given view. One natural way to focus the examination is to analyze what 'causal story' the model tells of the fluctuations in the familiar sample. In doing so, we shift the large and amorphous question, 'how does the model say the world works?' to 'What light does the model shed on the sample that is the source of our current expertise and conventional wisdom?'

In the current literature it is common to present a historical decomposition of headline variables like GDP growth in terms of the underlying structural shocks. Technically, for any value of θ, we can compute our best estimate of the underlying latent structural shocks. Given the linear Gaussian structure assumed for the model, these can be computed with the Kalman smoother[18]. The standard practice seems to be to produce a historical decomposition in terms of the smoothed shocks evaluated at the posterior mode for the parameter. For example, (Figure 2.6) taken from Smets and Wouters (2007) shows that the SW model attributes much of the deep recession in 1982 to the collective effect of demand shocks in the model. Demand shocks also account for much of the recession in 2001.

These decompositions can be a very useful tool for understanding the models, and posterior predictive analysis of structural features can form a valuable complement to these historical decompositions. First, note that making judgments about the model based on decompositions like this has the same problems that arise in the simple moment matching discussed above: it ignores uncertainty in θ and if something seems amiss, it provides no systematic way to judge just how implausible or freakish the result is.

There are many ways to use posterior predictive analysis to provide more systematic results complementary to the historical decompositions. For example, for a broad overall check we can take our structural feature, our $h(Y, \theta)$, to be elements of the sample correlation matrix of the smoothed estimates of the structural shocks. Remember that the structural shocks are assumed to be mutually uncorrelated in the model.

Each θ (when combined with a sample) implies a sample correlation matrix for smoothed shocks, so this is a well-defined structural feature. Since θ is unknown, we will not have a single realized value on the sample at hand; instead we will have a posterior density for the realized value that takes into account our remaining uncertainty about θ. Following Gelman's suggestion, we can, however, consider the joint posterior density for the

[18] Harvey (1991)

feature on the realized sample and predictive samples.

For example, take as our structural feature the sample correlation of the risk premium shock and the price markup shock. We represent the posterior predictive information (Figure 2.7, panel(f)) in a scatter plot in which each point represents a joint draw of a θ and a replication sample. For each such draw we compute and plot the pair $(h(Y^r, \theta), h(Y^{rep}, \theta))$— the feature on the realized and on the predictive sample, respectively. We plot these pairs as a scatter plot with the realized value on the horizontal axis and predictive sample on the vertical axis. If a typical draw from the model+posterior implies a sample correlation like that implied for the realized sample, the points of the scatter will lie around the 45 degree line.

For the chosen correlation, the entire point cloud is well below the 45 degree line. This implies that there is no value for the structural parameter that is likely to produce a correlation of these two structural shocks as high as that implied on the realized sample.[19] (Note: in scatter plots like Figure 2.7, the number in the upper left corner of the panel is the share of points on whichever side of the 45 degree line has a smaller share of points.)

The point cloud in this case is tall and thin. The fact that the point cloud is quite narrow tells us that for the bulk of θs getting posterior mass, the sample correlation on the realized sample is a bit over 0.2: for all relevant θs, the sample correlation of these shocks was substantial and in the 0.2 range. The height of the cloud spans values from about -0.15 to +0.15 and is roughly centered on zero. This gives a sense of what we would expect to see for this sample correlation in repeated sampling. Values vary some, but are mainly clustered near the population value of zero. There is very little chance that

[19]Actually, we should say there is no θ getting nontrivial posterior mass with the stated property.

the model+posterior would generate a sample implying a sample correlation between these shocks as high at 0.2.

There is no sign of discrepancy between the data and model+posterior for the sample standard deviation of the risk premium shock (Figure 2.7, panel(a)) or its correlation with the wage markup shock (Figure 2.7, panel(g)). The correlation with the productivity (Figure 2.7, panel(b)) and government spending shocks (Figure 2.7, panel(c)), however, on the sample are far from what the model+posterior would be likely to generate.

What is the interpretation? Of course, nature may have given us a sample that just happened to imply large sample correlations among the smoothed shocks. If we set this possibility aside, we must consider misspecification of the causal channels in the model.[20] The logic of the model is that these shocks originate in behaviorally distinct sectors of the economy and the population correlation is zero in the model (for every θ). In order to accommodate the sample using the causal channels specified in the model, however, various of these causal forces had to systematically work in concert. If one strongly believes that these forces are originating in behaviorally distinct sectors of the economy (as is the standard assumption) then the model needs refining. Otherwise, some causal account of the linkages must be specified.

As noted above, the risk premium shock was included in the model to help accommodate the positive correlation of consumption and investment growth in the sample. Above, we saw that difficulties remain in this descriptive feature. This analysis provides further evidence that perhaps that aspect of the model is misspecified. Gupta (2010) argues

[20]In principle, the result could be an artifact of the prior once again, but the fact that the population correlation is zero for every θ and a check of the prior predictive distributions suggests this is unlikely.

that a more systematic look at the full set of correlations among the structural shocks can provide valuable clues to economic sources of the misspecification and guide future model refinement.

There are many ways to go beyond mere correlations of structural shocks to focus on issues important in policymaking. For example, understanding and avoiding inefficient recessions is one goal of monetary policy. Thus, it is natural to focus on how the model helps us understand the recessions in the sample.

In the U.S., it is well known that periods of at least two consecutive quarters of negative GDP growth correspond fairly closely to the NBER's definition of recessions. Thus, on any sample, we can partition the observations of the smoothed structural shocks into those occurring in an episode of at least two quarters of negative GDP growth and the others. We can examine the posterior predictive description the model provides of the recessions in the sample.

For example, take as our feature the sample standard deviation of the smoothed risk premium shock during periods of recession and boom (Figure 2.8). The analysis provides no indication of problem with the standard deviation on the full sample or during booms (Figure 2.8, panel(a) and panel(c)). The sample standard deviation during recessions on the realized sample, however, is much higher than we would expect to observe under the model+posterior (Figure 2.8, panel(b)). Indeed, we can take as our feature the difference in the sample standard deviation in recession and boom periods. This difference (Figure 2.8, panel(d)) is once again considerably larger than we would expect to see out of a sample drawn at random from the sample+posterior.

Recessions in the post-War sample, according to the model, were a collective freak occurrence of abnormally large risk premium shocks occurring systematically at business cycle frequencies. Faust and Gupta (2010b) provide a more complete analysis of this topic finding similar results for other structural shocks. Once again, we can accept that nature gave us a very strange sample in which some of the focal events for policymakers—recessions—repeatedly arose for reasons that are highly unlikely ever to be repeated, or one can consider model+prior refinement.

The main point in this section is to illustrate potential uses for posterior predictive analysis, and not to provide a substantive analysis of the SW model. Indeed, there are many thorough critiques of this model, including very incisive critique by the authors themselves, especially in joint work with Del Negro and Schorfheide (Del Negro, Schorfheide, Smets, and Wouters (2007)). We argue that the sort of posterior predictive analysis illustrated above provides a complementary tool and is especially useful for highlighting strengths and weaknesses of the models as they pertain to particular uses such as policymaking.

2.3.4 Elaborations and Abuses

We have deliberately stuck to fairly straightforward illustrations in the previous section in order to introduce these tools. There are many natural elaborations. For example, as Gelman notes, one might want to condition all the computations on certain data features. We have deliberately focused on basic applications of the ideas behind prior and posterior predictive analysis in order to introduce the ideas.

We have examined many different data features. Of course, whenever one has multiple statistics there are multiple ways they might be combined and consolidated. For example, one could take account of the full joint distribution of some group of features.

Thus, one could ask how likely the model would be to produce sample jointly showing values as extreme as the realized values. One could also combine the features into some *portmanteau feature* and only consider the marginal distribution of the overall combined feature. We have argued for the benefits of interpretability of the features and such a portmanteau would probably lose some of that.

Any discussion of the myriad features one might combine naturally leads to a discussion of how this approach might be misused and abused. On any sample, we can define features that are as 'freakish' as we like in the sense of being present in a small proportion of all samples. Indiscriminate assessment of long lists of features will, with probability one, lead one to discover that each random sample (like each child) is special in its own way. As Gelman, Meng, and Stern (1996), Hill (1996) and many others emphasize, any tool like this must be used with judgement.

In particular, we are suggesting using these tools to highlight areas of consonance and of dissonance between the model any strongly held views about the only existing sample. As we raise this topic, however, we begin to squarely face the third major point of the chapter: the standard criticisms of prior and posterior predictive analysis and what we argue is a nonstandard defense in the DSGE context.

2.4 Standard Criticisms and a Nonstandard Defense of Posterior Predictive Analysis

While we believe that posterior predictive analysis could be an extremely useful tool, uses so far in the DSGE literature have been limited. This may, in part, be because of the strong arguments often stated against the coherence of this form of inference.

In this section, we argue that standard arguments against posterior predictive analysis—whatever their merits in other cases—are moot or miss the point in the DSGE context. Indeed, posterior predictive analysis is arguably a natural pragmatic attempt to apply Bayesian principles under challenging conditions.

2.4.1 Standard Criticisms

Prior and posterior predictive analysis like all inference based on hypothetical other samples, violate the likelihood principle, which "essentially states that all evidence, which is obtained from an experiment, about an unknown quantity θ, is contained in the likelihood function of θ for the given data..." (Berger and Wolpert, 1984, p.1). Given that prior and posterior predictive analysis share with frequentist analysis an emphasis on behavior in repeated sampling, many of the objections echo the arguments in the familiar frequentist vs. Bayesian debate.

One cannot deny that troubling problems can arise whenever one attempts to cast doubt on a model when based on the fact that was observed in the sample was less likely than other samples that were not in fact observed.[21] One way of seeing the problem of casting doubt on a model due to the fact that the sample at hand is unlikely is that this approach begs the question 'unlikely compared to what?' We have no alternative model that renders the existing sample more likely. Inference without an explicit alternative is fraught a host of problems, leading some to the summary judgement (Bayarri and Berger, 1999, p.72), "The notion of testing whether or not data are compatible with a given model

[21] The literature here is immense. For a recent treatment aimed at economists, see Geweke (2010).

without specifying any alternative is indeed very attractive, but, unfortunately, it seems to be beyond reach."

Geweke (2010), (p.25), argues that, while both prior and posterior predictive analysis violate the likelihood principle, but posterior predictive analysis involves "a violation of the likelihood principle that many Bayesians regard as egregious." This is in part because in this analysis *uses the data twice* in an important sense (without taking formal account of this fact). One checks the freakishness of the sample using the posterior that was already updated using that same sample.

Berger and Wolpert (1984) offer the following judgement about the use of such techniques:

> Of course, even this use of significance testing [as proposed by Box] as an alert could be questioned, because of the matter of averaging over unobserved x. It is hard to see what else could be done with [the maintained model] alone, however, and it is sometimes argued that time constraints preclude consideration of alternatives. This may occasionally be true, but is probably fairly rare. Even cursory consideration of alternatives and a few rough likelihood ratio calculations will tend to give substantially more insight than will a significance level, and will usually not be much more difficult than sensibly choosing T [the data feature] and calculating the significance level. (p.109)

We begin our defense of posterior predictive analysis by accepting (or, at least, choosing not to contest) essentially all of Berger and Wolpert's points. In particular, until recently, Berger and Wolpert's claim that specifying explicit alternatives, perhaps in a cursory manner, is easy was nearly a tautology. Until recently, Bayesian methods were computationally infeasible except in trivial cases and constructing cursory alternatives in such cases may be easy. Constructing meaningful alternative models of the world macroeconomy with fully articulated causal channels, we argue, is not easy. And, thus, we claim this is one of the rare cases.

We also agree with Berger and Wolpert that when one comes across a rare cases where specifying alternatives is not trivial, it is difficult to imagine any systematic way to proceed other than some variant of the basic idea laid out by Box, and that is what we are proposing. But our defense of posterior predictive analysis goes considerably deeper based on a number of features that distinguish the DSGE context from that contemplated in the plain vanilla scheme.

2.4.2 Nonstandard Defense of Posterior Predictive Analysis for a Nonstandard Context

One slowly growing sample. In DSGE modeling, macroeconomists are attempting to formalize and reify our understanding of the world economy. Unfortunately, while the available sample regarding the general equilibrium process is growing, it grows sufficiently slowly that we may treat it as fixed.[22] The many rounds of refinement take place using the same data and the formal prior and posterior in all DSGE work are best viewed as two different ways of using the same sample. The 'posterior' for the current model will soon be replaced by another 'posterior' for a revised model, and this new 'posterior' will be computed on the same data as were the many previous versions.

For the remainder of this section by *posterior* in italics we simply mean what this term has come to mean in the DSGE literature: the result of the latest round of update on the familiar sample.

[22]by general consensus (confirmed over the 30 year development process) the new information in a few observations is unlikely to provide much additional information. Observations like those from the recent crisis are arguably highly informative, but generally raise more questions than answers.

Given the intertwining of expertise and the only sample it is almost inevitable that certain strongly held prior views regard the current sample. For example, at the most general level, most macroeconomists believe that systematically repeated features of the business cycle are in fact systematic features of the underlying mechanism. Posterior predictive analysis then can be seen as a pragmatic way to check the consistency of the model+*posterior* with difficult to impose prior views.

The current model by general consensus remains materially deficient. The model +*posterior* , whatever its merits, will be used to inform current decision making and as a basis for the next round of model refinement. Much of the analysis we advocate might, in principle be carried out using prior predictive analysis and Geweke (2010), (p.24), makes a strong case for doing so based on the position that we should place "specification analysis ahead of formal inference." Prior predictive analysis can be used in this specification analysis step, but any posterior analysis on what will be treated as the posterior update sample must be confined to the formal inference step.

This analysis does not conform well to the case where we have ongoing refinement of a materially deficient model that is going on in parallel with actual decision making based on the current best model. Substantive ongoing specification analysis has in the past and will for the foreseeable future accompany use of the current *posterior* . Perhaps the simplest way to avoid this debate is to consider the current *posterior* to be the prior for use in the next stage of ongoing specification analysis.

As an overarching idea, however, we think it is uncontestable that if the current model+*posterior* are thought to be materially flawed and yet will be used in policy analysis, searching for problematic predictions from this model+*posterior* must be consistent with good sense as well as Bayesian principles.

There is an alternative to the model—the current subjective policymaking machinery. In the context of DSGE modelling for policy analysis, it is important to emphasize there is an alternative: the implicit model in the current subjective policymaking process. A natural inclination might be to argue that we should render that model formal and explicit, and that this newly formalized model is the model to which we should compare the DSGE model. Of course, this misses the point entirely.

The current DSGE models built for policymaking *are* the current state of our efforts to reify the current policymaking process—to formalize the good and throw out the bad. Our defense of posterior predictive analysis is that it can render the process of enriching and refining the model more efficient by focusing attention on areas that are most troubling from the standpoint of the task at hand.

2.4.3 An Alternative: Patch up the Plain Vanilla Scheme

In our experience discussing these issues, there is a very strong and proper urge to consider the possibility that we could—perhaps in some cursory way as suggested by Berger and Wolpert—paper over the difficult aspects of the DSGE context and somehow follow some scheme that has a reasonable pragmatic interpretation as being consistent with the plain vanilla scheme.

Obviously, these suggestions involve methods to eliminate *ad hoc* aspects of the prior and to accommodate in some way the most gross deficiencies of the model. For example, Del Negro and Schorfheide (2008) what appears to be a very useful approach to training samples as a method to reducing the sort of problems with the conventional DSGE prior that they emphasize and that we have illustrated above. Geweke has recently provided a brilliant monograph on working with complete and incomplete econometric models, which involves methods for comparing explicit models based on particular features and setting aside certain types of deficiency.

All these strike us as very good ideas that should find widespread uses. We argue that that are complementary to our suggestions, however. Such steps cannot overturn the fact that for the foreseeable future, our best complete general equilibrium model will be materially deficient, in an ongoing state of refinement on a single dataset, and simultaneously in use in policymaking. We argue that under these conditions, any coherent class of analysis must allow for the examination of the flaws of the current model+*posterior* that is informing the policymaking process.

2.5 Conclusions

The modelling of causal channels in a large, dynamic system, with forward looking behavior is an incredibly daunting task. It is made even more challenging by the fact that we have a single dataset on which to both develop and test our theories. That dataset is small in relevant senses.

Perhaps it is not surprising in such a challenging context that there is little consensus in the profession on key substantive economic topics or on modelling methodology. The unprecedented Congressional hearing on DSGE modelling underscores both the unfortunate state of affairs and the fact that improving on this state of affairs is in the interest of everyone subject to the policy decisions informed by these models.

We argue for taking as a starting point that these models will actively be used in policy analysis while remaining in an ongoing state of material refinement. Seen in this light, posterior predictive analysis, we argue, can be a very useful tool for highlighting strengths and weaknesses pertinent to policy. We particularly argue for the focus on what we call structural features as a way to assess causal channels in the model. This type of analytics has received little prior use in DSGE modelling. The overarching idea is that the the type of analysis we advocate may serve to inform policymakers of limitations of the current models (which can then be judgementally allowed for) and to direct resources of the model refinement efforts to areas most relevant to policymaking.

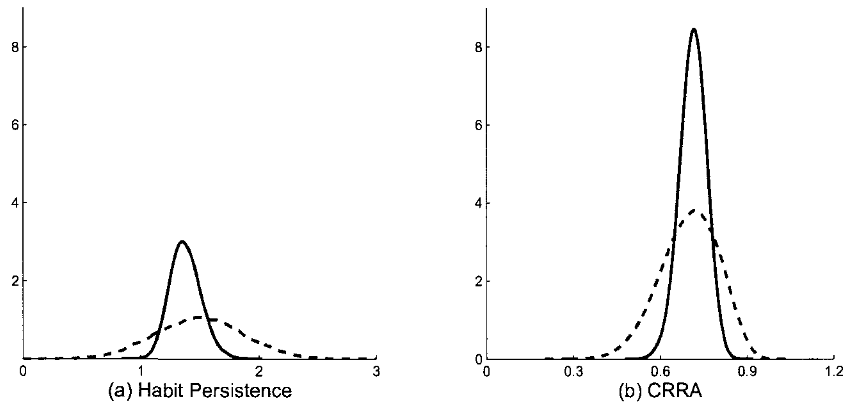


Figure 2.1: Prior and posterior densities for habit persistence and CRRA parameters. Blue dashed line is the prior; black line is the posterior. The parameters are described more fully in the text.

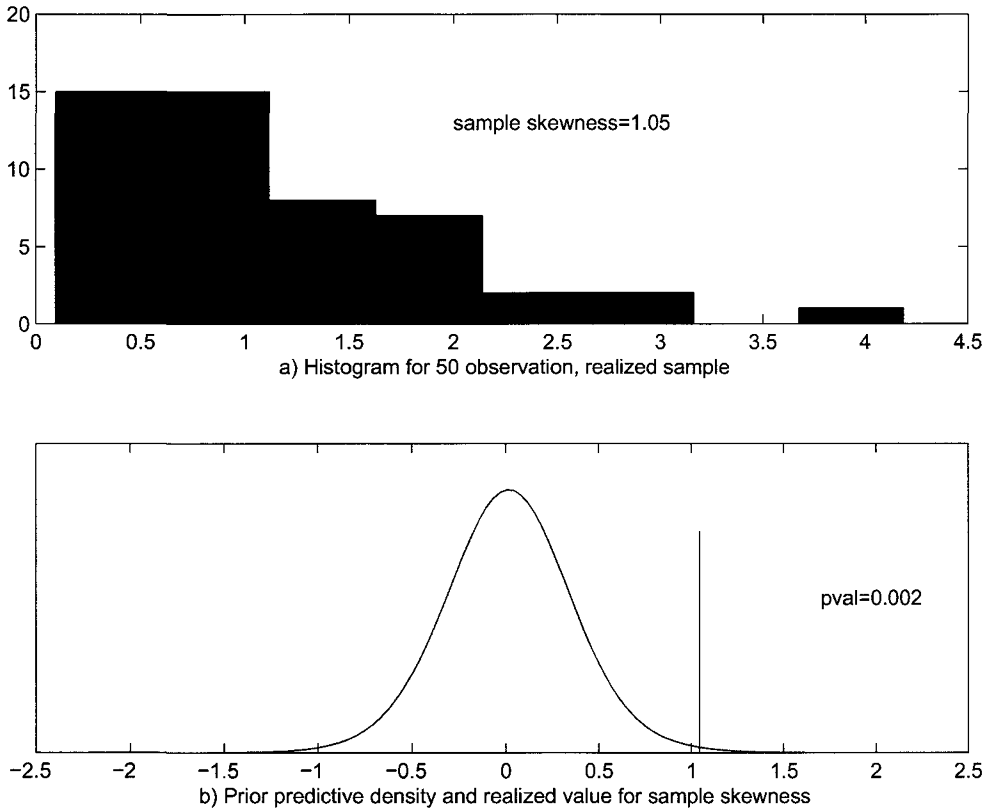


Figure 2.2: Example of prior predictive analysis. Panel a) Histogram of 50 sample observations. Panel b) The prior predictive density for sample skewness and realized value for sample. The prior is that the sample points are iid $N(\mu, 1)$ and the prior for μ is uniform on $[0, 1]$. The sample skewness of 1.05 is shown in red on the right hand panel. The stated p-value is the share of the mass of the prior predictive density exceeding the realized sample skewness.

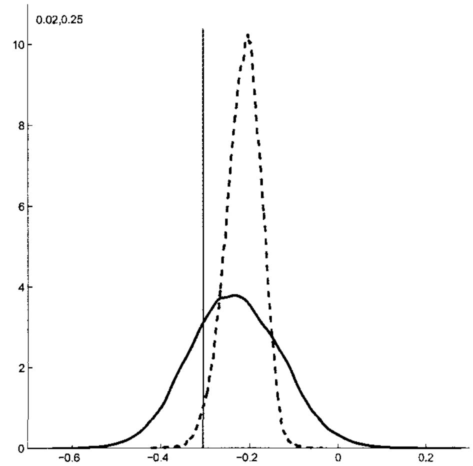
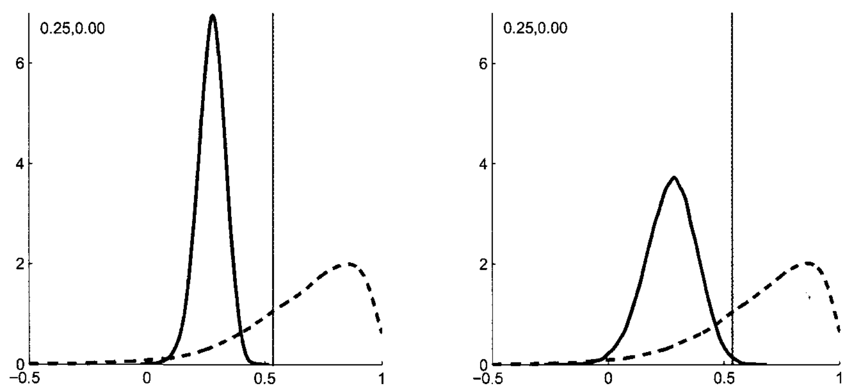


Figure 2.3: Posterior density for population correlation of output and inflation (blue dashed) and posterior predictive density for the sample correlation (black solid). Red line is the realized value on the sample. The numbers in the upper left give the proportion of mass under the posterior and posterior predictive density, respectively, that is to the left of the realized value.



(a) Prior and Posterior for Population Correlation

(b) Prior Predictive and Posterior Predictive for Sample Correlation

Figure 2.4: Correlation of consumption and investment growth. Panel a) Prior (blue dashed) and posterior (black solid) for population correlation. Panel b) Prior predictive (blue dashed) and posterior predictive (black solid) for sample correlation. The numbers in the upper left give the share of points in the smaller tail relative to the red line for the two densities on the panel, with value for prior before posterior.

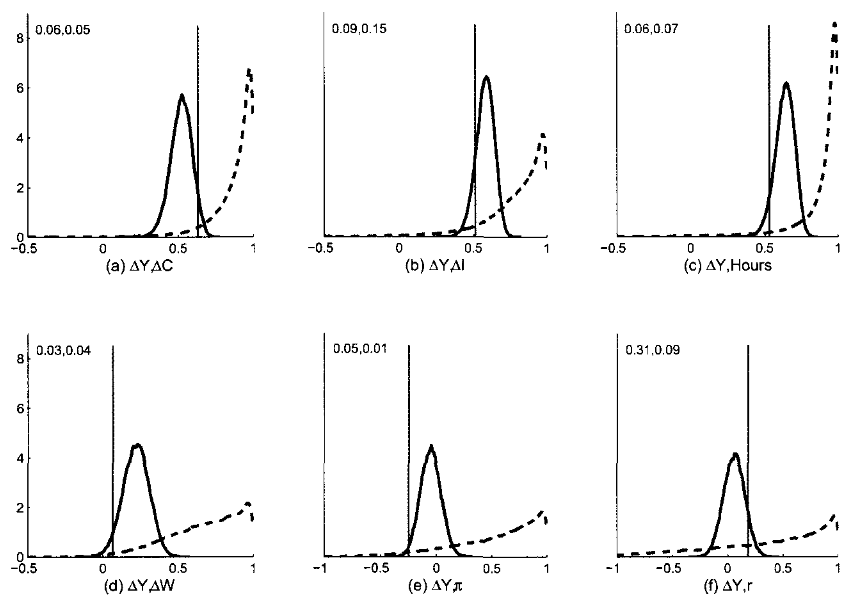


Figure 2.5: Prior predictive (dashed blue) and posterior predictive (solid black) densities for one-step forecast error correlations. Each panel is for a sample correlation of the output growth error ΔY and the error for one of the other variables in the model: ΔC, consumption growth; ΔI, investment growth; hours, Δw, wage growth; π, inflation; r, interest rate. The errors come from a VAR(1) estimated on the sample. Red line is for the VAR(1) value on the realized sample. The numbers in the upper left give the share of points in the smaller tail relative to the red line for the two densities on the panel, with value for prior before posterior.

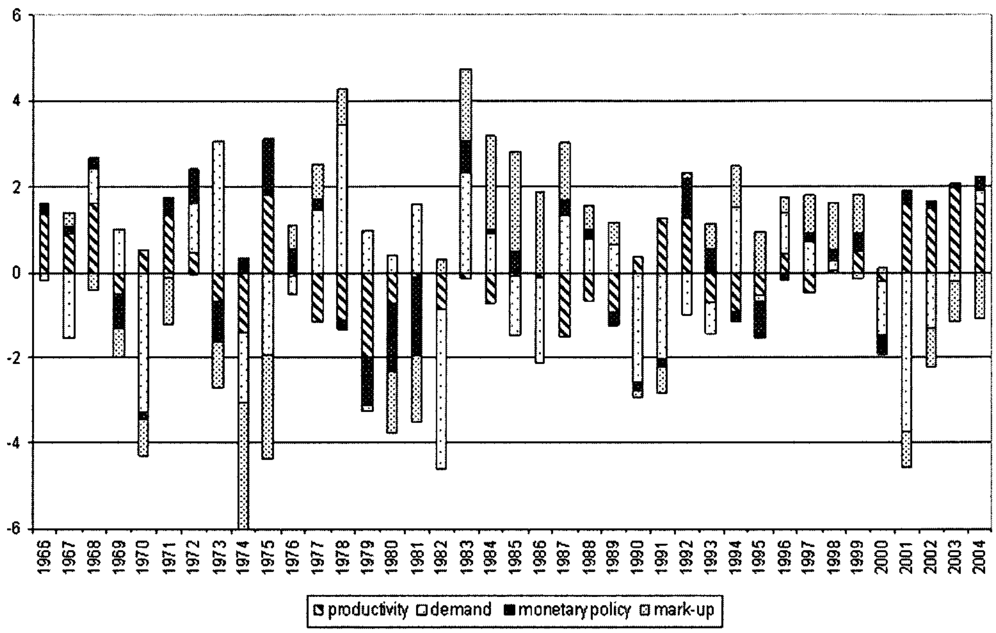


Figure 2.6: Historical decomposition of output growth in terms of the structural shocks. The 7 shocks have been averaged over calendar years and summed across broad categories. The 'demand shocks' include the risk premium, investment-specific technology, and exogenous spending shocks; the 'mark-up shocks' include the price and wage markup shocks. Source: Smets-Wouters (2007).

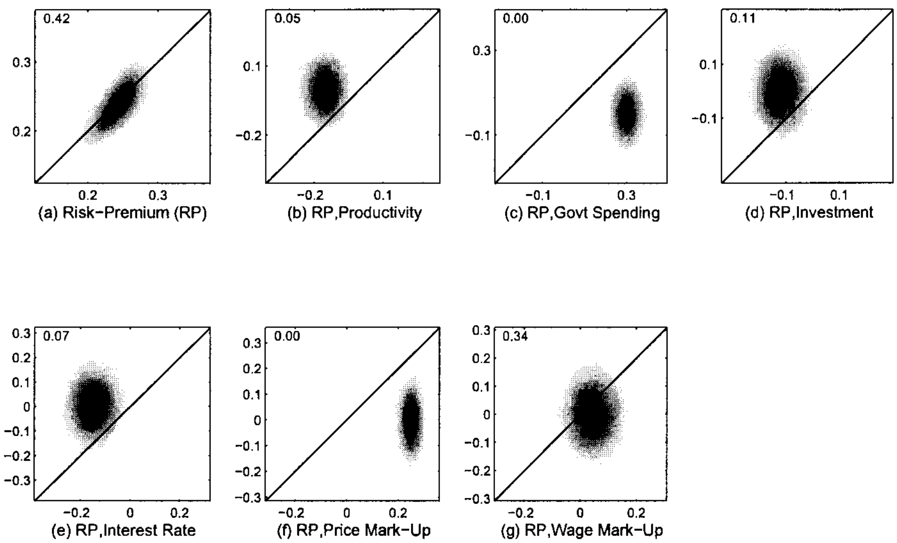


Figure 2.7: Structural feature scatter plots for the smoothed risk-premium (rp) shock. Panel a) is for the standard deviation of the rp shock; the remaining panels are for the sample correlation of the rp shock with other structural shocks in the model: productivity, investment productivity, government spending, monetary policy, price mark-up, and wage mark-up. 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.

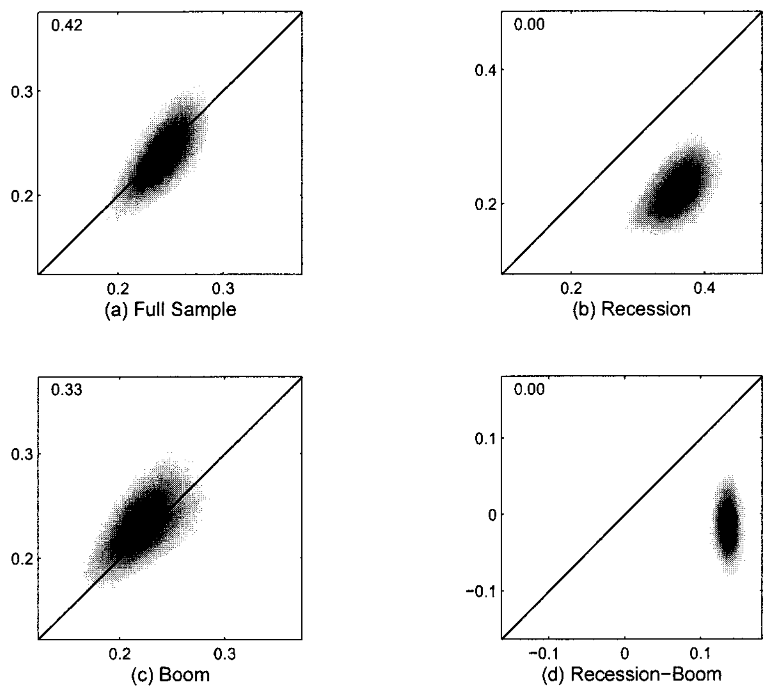


Figure 2.8: Structural feature scatter plot for the sample standard deviation of the smoothed risk-premium shock in recessions and expansions. Panel (a) is for the full sample, (b) recessions, (c) expansions. In panel (d), the feature is the difference in the standard deviation for recessions and expansions. 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.

Chapter 3

A Forecasting Metric for Evaluating DSGE Models for Policy Analysis

3.1 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 providing 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 chapter 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, Schorfheide, Smets, and Wouters (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 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 chapter, we first analyze the intended purpose of monetary policy analysis to

[1]Some of the recent papers looking at out-of-sample forecasting performance of DSGE models are Smets and Wouters (2004), Adolfson, Andersson, Linde, Villani, and Vredin (2007), Edge, Kiley, and Laforte (2009).

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 chapter 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 chapter 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 chapter 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 chapter 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 chapter. Section 3 illustrates the posterior predictive evaluation approach using the SW model and discusses the results, and Section 4 concludes.

3.2 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.

3.2.1 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

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

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 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?'

3.2.2 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:

$$\begin{aligned} \Omega &= CE(\varepsilon_t\varepsilon_t')C' \\ &= CC' \\ &= \begin{bmatrix} 10 & -4 \\ -4 & 2 \end{bmatrix} \end{aligned} \tag{3.1}$$

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:

$$\begin{aligned} \hat{\Omega} &= \tilde{C}\tilde{C}' \\ &= \begin{bmatrix} 10 & 2 \\ 2 & 2 \end{bmatrix} \end{aligned} \tag{3.2}$$

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 (3.1).

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 (3.1) and (3.2)).

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.

3.2.3 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:

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

The forecast errors, ν_t, are then given by:

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

[3]In this chapter 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.

This is an identity that must hold under the model for all versions of the shocks, so for example for the updated shocks, we will 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 did'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.

3.2.4 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, Meng, and Stern (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 chapter I consider only structural features, $h(Y, \theta)$, that depend upon θ in addition to the sample. The structural features that I consider in this chapter 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.

---

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

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 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.

3.3 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.

3.3.1 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, Eichenbaum, and Evans (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 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

[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.

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.

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

In this chapter, 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

[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

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 3.3 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 3.3, 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 volatility in these variables than what it estimates on the realized sample. Overall, we observe that the SW model does well with regards to

posterior distribution of the estimated parameter vector θ in SW.

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 3.3.

As an illustration, we focus on the FEC($Hours, \Delta W$) graphed separately in Figure 3.1 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 3.1 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 3.1 is as follows. The models needs a shock

[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.

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 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 3.2 and Table 7 provide a similar accounting analysis for the realized value of the $FEC(\Delta Y, \pi)$. Figure 3.2, 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

[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.

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 putting too much emphasis on demand shocks and in order to counteract 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 3.3 graphs the scatter plots for all the elements of Ω; Figure 3.4 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.

3.3.3 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. Therefore, due to the lack of free parameters in the model, this discrepancy shows up in the mean value of the one-step ahead forecast errors. Given these non-zero means, it becomes 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-

[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.

[10] The SPF forecasts for the next quarter are made in the middle of the current quarter. For instance,

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 3.5. 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 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

---

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.

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 3.7. Except the monetary policy shock and the wage mark-up shock, all other shocks have a non-zero realized mean value.

3.4 Conclusion

This chapter provides a way to analyze the monetary policy process using the posterior predictive analysis described in Faust and Gupta (2010a). The chapter 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 chapter 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 chapter 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 chapter show exactly how to do that. The analysis here suggests making specific structural changes to the model in order to account for the misspecifications highlighted in the model.

3.5 Appendix

3.5.1 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:

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

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:

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

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.

Table 3.1: 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 3.2: 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 3.3: 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 3.4: 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 3.5: 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 3.6: 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 3.7: 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

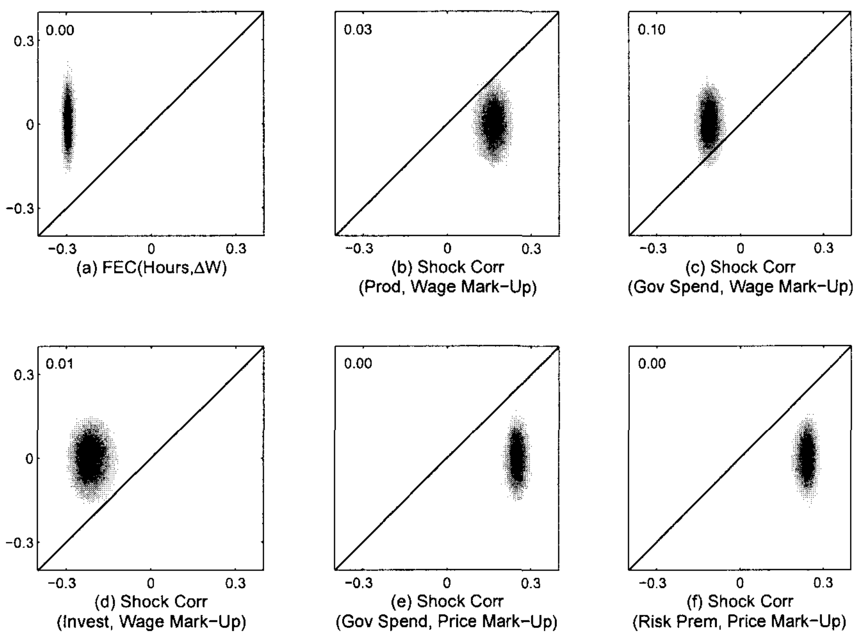


Figure 3.1: Scatter plots of correlated shocks that account for FEC($Hours, \Delta W$). Panel a) plots FEC($Hours, \Delta W$); remaining panels plot the correlated shocks that account for the negative values of this correlation on the realized sample (see Table 6). Horizontal axis plots the posterior 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.

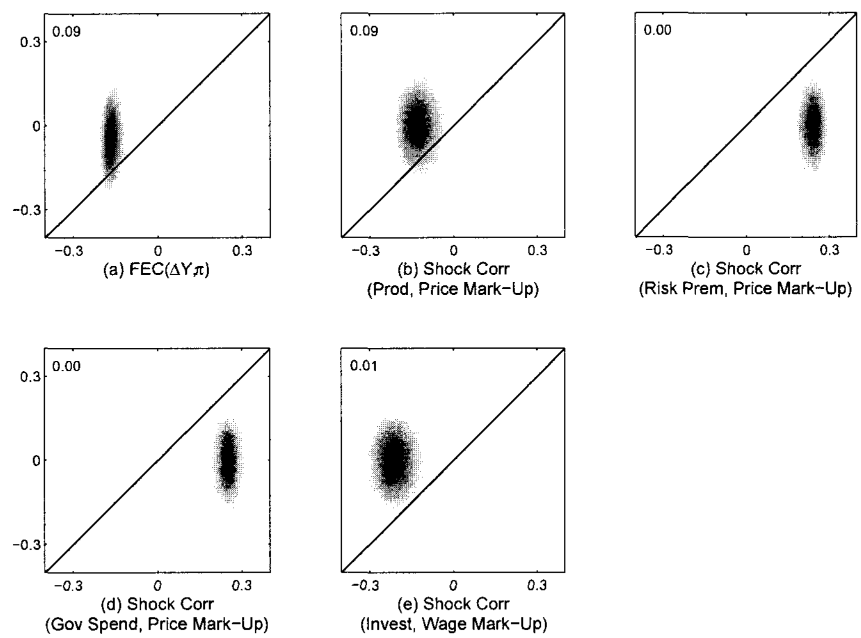


Figure 3.2: Scatter plots of correlated shocks that account for FEC($\Delta Y, \pi$). Panel a) plots FEC($\Delta Y, \pi$); remaining panels plot the correlated shocks that account for the negative values of this correlation on the realized sample (see Table 7). Horizontal axis plots the posterior density 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.

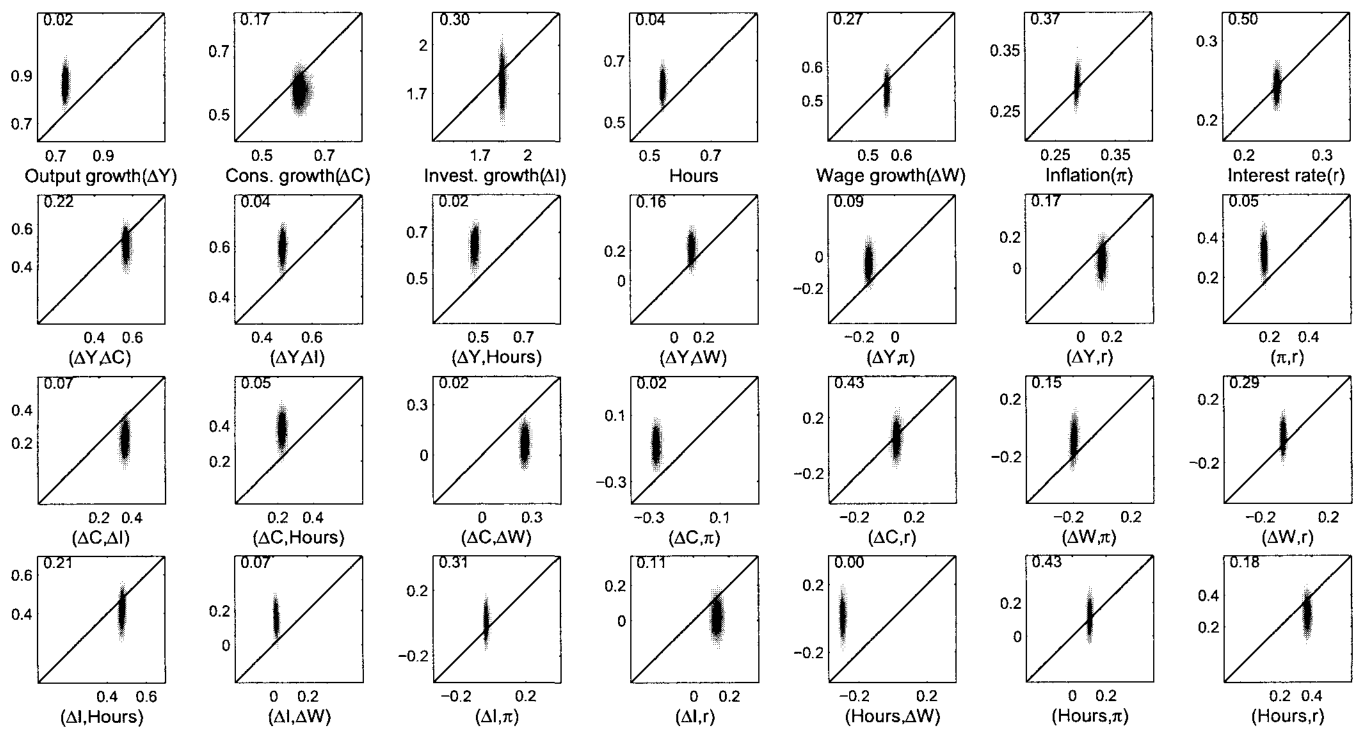


Figure 3.3: Ω matrix. Scatter plots for the standard deviations and correlations of the FE's, ν_t. First row plots the standard deviations; remaining rows plot the correlations. Horizontal axis plots the posterior values for the realized sample; vertical axis plots the posterior predictive values. The number in the upper left gives the smaller share of points on either side of the 45 degree line.

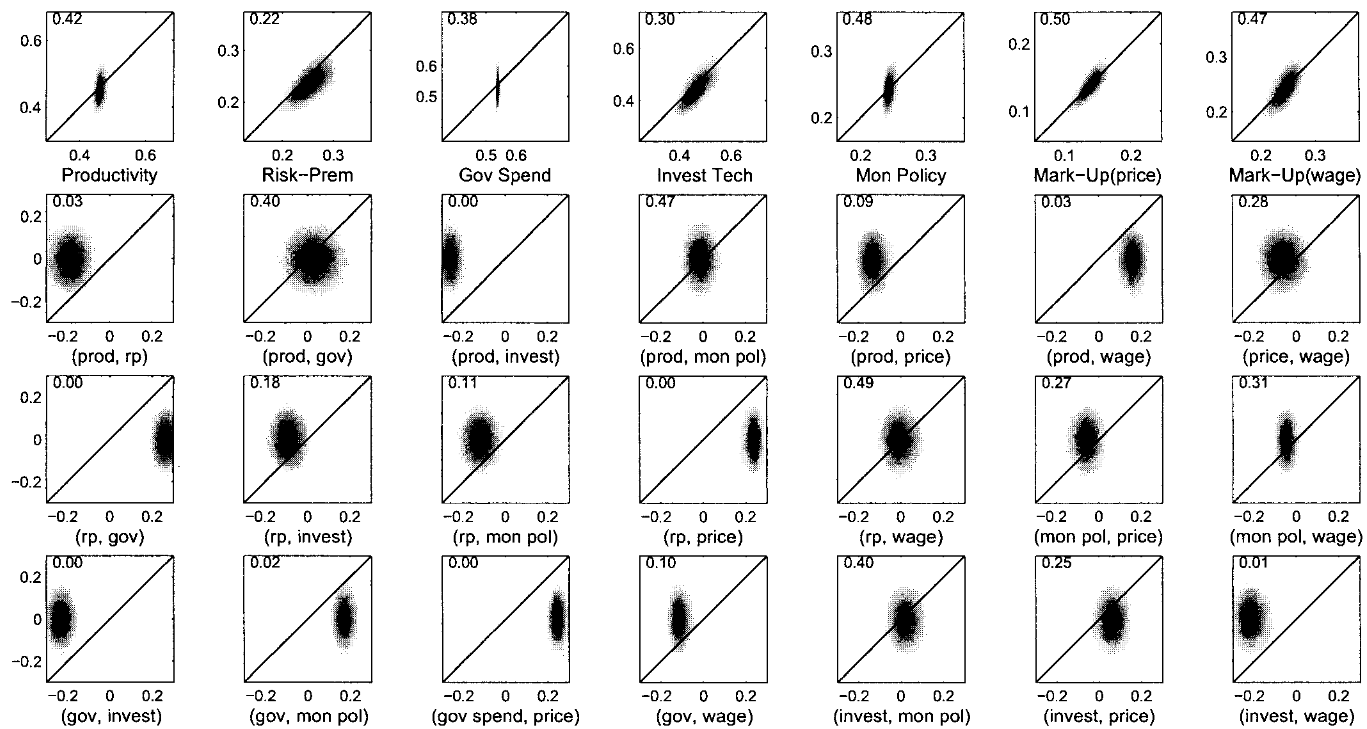


Figure 3.4: Σ matrix. Scatter plots for the standard deviations and correlations of the structural shocks, $\varepsilon_{t|t}$. First row plots the standard deviations; remaining rows plot the correlations. Horizontal axis plots the posterior values for the realized sample; vertical axis plots the posterior predictive values. The number in the upper left gives the smaller share of points on either side of the 45 degree line.

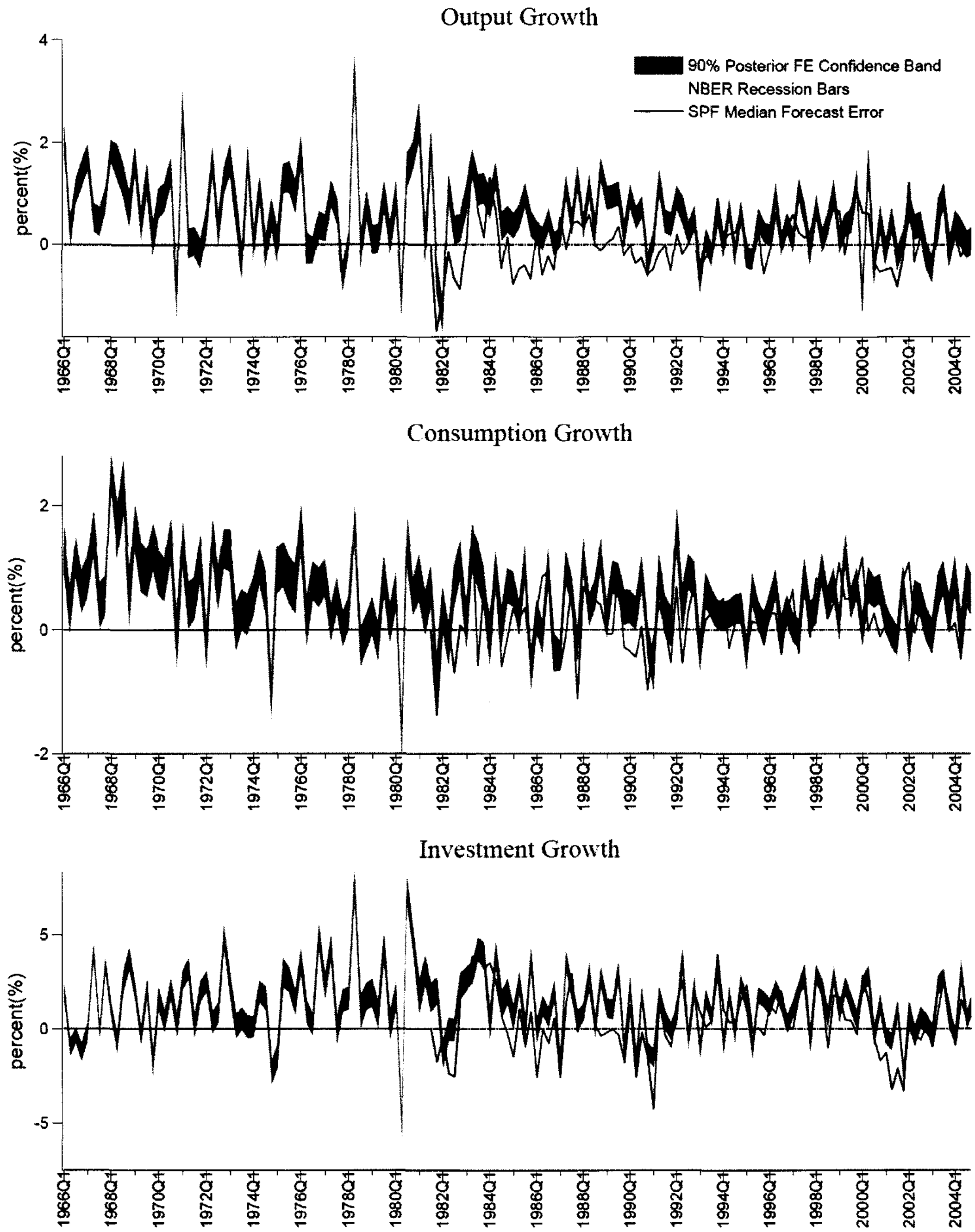


Figure 3.5: FE's in output growth, consumption growth, and investment growth.

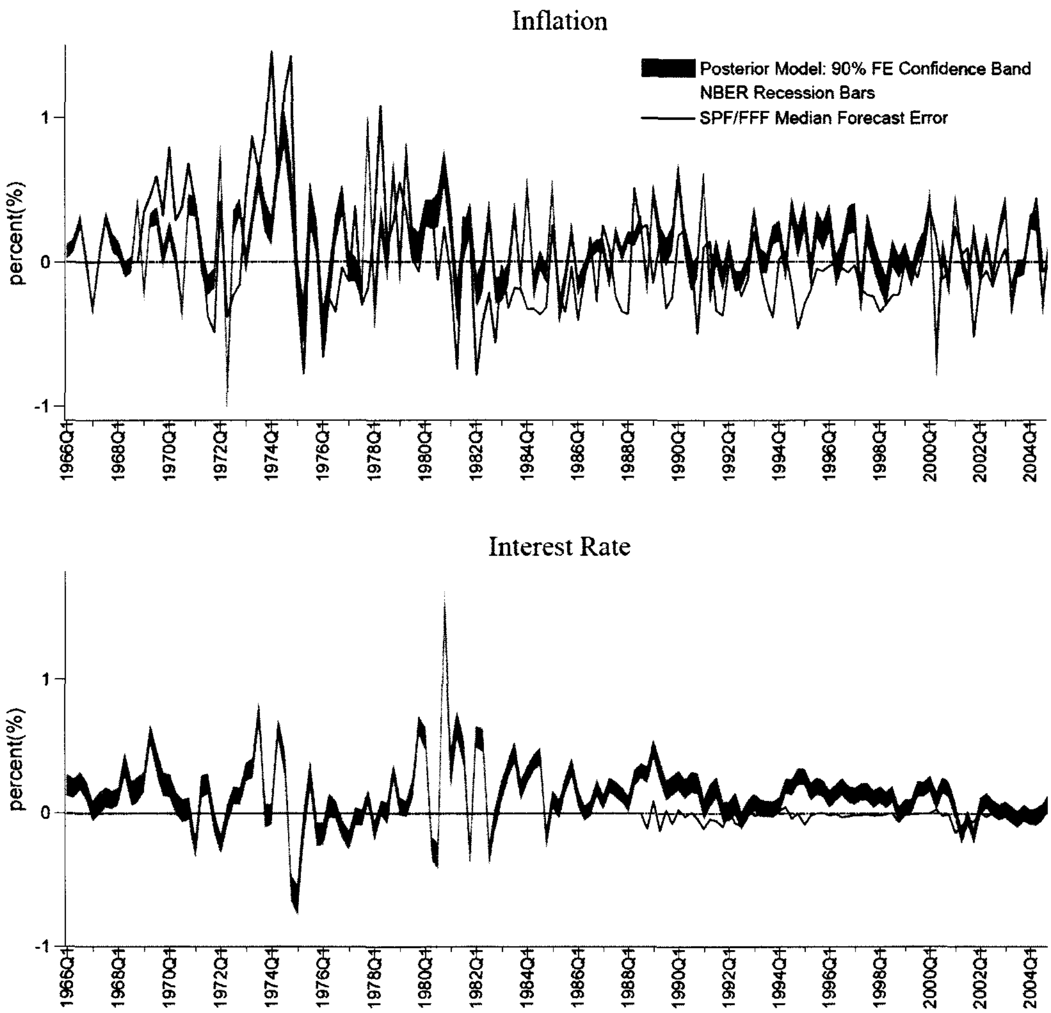


Figure 3.6: FE's in inflation and interest rate.

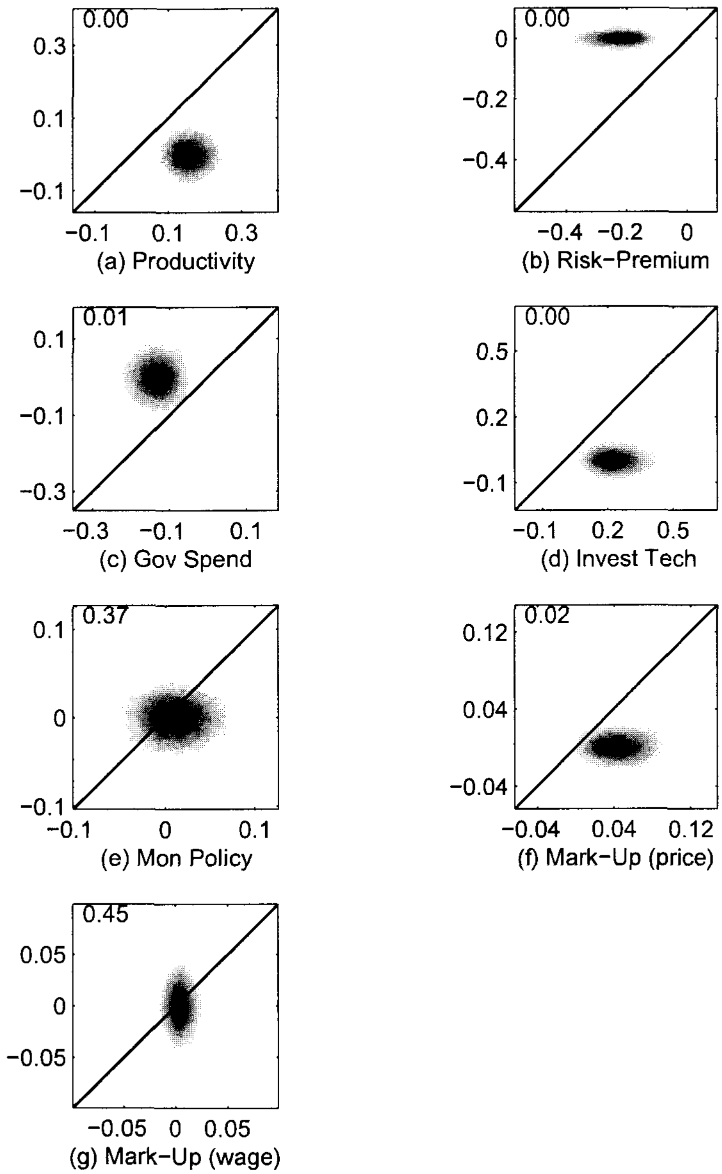


Figure 3.7: Scatter plots for the mean value of the updated structural shocks, $\varepsilon_{t|t}$. Horizontal axis plots the posterior values for the realized sample; vertical axis plots the posterior predictive values. The number in the upper left gives the smaller share of points on either side of the 45 degree line.

Chapter 4

Fiscal Policy Multipliers at the Zero Bound: The Case for Japan

4.1 Introduction

The Japanese economy has had a decade of lost growth. Since mid-nineties, Japan has been battling deflation and the overnight bank rate in Japan has been close to zero. This has been a classic textbook case of liquidity trap, in which, the nominal overnight bank interest rate hits the zero lower bound and conventional monetary policy can no longer be used to jump start the economy. In the mid-nineties this seemed like an isolated event, a Japanese problem, that was unlikely to repeat itself anywhere else in the world. However, the recent US financial crisis has caused central banks in much of the western world to bring down the bank rate close to zero. The federal funds rate in the US has been in the 0-0.25% range since December 2008.

Policy makers are turning to fiscal policy stimulus packages all over the world to enable their respective economies get back on the engine of growth. However, there is very little research available on how effective fiscal policy is at the zero lower bound. Is government spending more effective at the zero lower bound than during normal times? Is government investment spending more effective than government consumption spending at the zero lower bound? Does lowering taxes always have a stimulative effect on output and does reducing all different kinds of taxes have a similar impact on output? What fiscal policies are most effective at the zero lower bound? Recently, there have been a number of theoretical papers that have tried to answer these questions. However, there is little to no consensus and different theoretical results argue for different fiscal policies at the zero lower bound for nominal rates.

This chapter is an attempt to look at the data to find an answer to these questions. The most challenging part of this exercise is the limited number of episodes involving the zero lower bound and the short duration of the sample size within those episodes. In part, this is an attempt to see if there is enough information in the data to resolve some of these questions. Despite this caveat, we do find a few interesting and puzzling results. However, overall, we find that the Japanese data is not informative enough to say that fiscal policy operates very differently at the zero lower bound.

We estimate the fiscal policy multipliers for Japan for various macroeconomic variables and test if the short run impulse response path for these variables in response to a fiscal shock are any different for the post 1995 period. In addition to estimating these multipliers for aggregate fiscal variables of total tax revenues and total government spending, we also estimate these multipliers for disaggregated measures of these fiscal variables. We estimate the impulse response path of output, inflation and interest rates in response to a shock to government consumption and government investment; and also in response to a shock to different tax groups: personal taxes, corporation taxes, indirect taxes and net social security taxes.

The most significant result, we find is that in the short run a shock to corporate tax revenues is expansionary when the nominal rates hit the zero lower bound and contractionary when they are not bound at zero. This result is formally tested using a joint test of equality for the impulse responses up to and including 8 quarters after impact of the corporate tax shock. The Wald test rejects the null of equality of the impulse responses during the period of positive interest rates and the zero lower bound at the 1% level. We don't find support for larger government spending multipliers at the zero lower bound. Further, we find the spending multipliers for Japan to be statistically insignificant. Even the government investment spending multiplier is statistically insignificant, both during the period of positive interest rates and the zero lower bound period.

The remainder of the chapter is organized as follows: Section 2 discusses theory and the methodology used for VAR identification, Section 3 discusses data and results and introduces some checks for robustness, and Section 4 concludes.

4.2 Theory and Methodology

Recently a number of theoretical papers have been written building on the New Keynesian general equilibrium models to analyze the effects of fiscal policy when the nominal interest rate reaches the zero lower bound. Eggertsson (2009) and Christiano, Eichenbaum, and Rebelo (2009)] show that government spending multiplier should be much larger when the zero lower bound is binding than when it is not. To understand this result, one should note that at the zero bound the equilibrium in these New-Keynesian models is characterized by falling output and negative inflation unlike the equilibrium at positive nominal rates that is characterized by positive output growth and inflation. In terms of an aggregate supply-aggregate demand demand analysis, an exogenous shock that causes the nominal interest rates to hit the zero lower bound makes the aggregate demand curve upward sloping in inflation. The intuition behind this is that higher expected inflation, when the nominal interest rate is fixed at zero, lowers the real interest rate and makes spending relatively cheaper in the current period. This stimulates demand. Contrary to that a deflationary expectation at the zero lower bound raises the real interest rate and makes current spending more costly than future spending and results in a lower output.

Given this setup, a rise in government spending is more expansionary at the zero lower bound. It raises inflation as it provides increased demand for goods and services and as long as the zero lower bound is effective this reduces the real interest rate and provides additional stimulus to the economy. It is important to note that the usual crowding-out effect of government spending is not active precisely due the the interest rates being stuck at zero. Additionally, the private sector expectations about future government behavior plays an important role in explaining the larger spending multiplier in these models. The private sector agents expect output to contract and prices to fall in all future periods in which the zero lower bound is in effect. If the government can commit to increased spending until the recession is over, then that will change the private sector expectations about all future periods in which the zero lower bound in binding thereby having a much larger multiplier affect for a given amount of spending.

Eggertsson (2008) also shows that a labor tax cut has a contractionary impact on output when the zero bound on nominal rates is binding. The paper argues that the supply side effects of a tax cut are more prominent than the demand side effects during a recession. The labor supply increases and the firms are able to charge a lower price for their product relative to an earlier period in the recession state when the tax cut was not announced. This deflationary pressure on prices results in higher real interest rate in all future periods in which the zero lower bound is binding and thereby contracts output.

Erceg and Lindé (2009) also reach the same conclusion about the effect of government spending but show that this effect of the fiscal stimulus is obtained only when it is timely implemented and is financed using lump-sum taxes. They argue that in case of delays in the implementation of fiscal spending, the potential real interest rate will fall that increases the output gap and lowers inflation as long as the zero lower bound is in effect. This can cause the spending multiplier to be much lower and even negative. Taylor (2009) argues that discretionary fiscal policy is not necessarily more effective at the zero lower bound citing the Japanese example.

4.2.1 Identification

We follow the Blanchard and Perotti (2002) structural VAR approach to identify the structural fiscal policy shocks using external institutional information regarding the elasticities of fiscal variables to economic activity. Following Perotti (2004), in addition to the fiscal variables, we include inflation and interest rates in the VAR specification. The benchmark VAR specification includes a constant and 5 variables: government spending (g_t), net taxes (t_t), GDP (y_t), price level (p_t), and overnight interest rates (i_t) for the pre 1995 period when the Japanese economy was not at the zero bound and 4 variables for the zero bound period, effectively dropping the interest rate variable since that is near zero for that period. Below, we outline the identification methodology used in the chapter.

The reduced form VAR specification can be written as:

$$Y_t = A(L)Y_{t-1} + U_t \tag{4.1}$$

where $Y_t \equiv [g_t, t_t, y_t, p_t, i_t]$ is the five-variable dependent vector, $U_t = [u_t^g, u_t^t, u_t^y, u_t^p, u_t^i]$ denotes the reduced form residual vector and A(L) is a lagged polynomial allowing for 4 lags of the dependent variable. [1] After obtaining the reduced form residuals, the next step is to obtain the structural shocks. Following Blanchard and Perotti (2002), reduced form residuals for the fiscal variables, u_t^g and u_t^t, can be thought of as a linear combination of three types of shocks: automatic response of fiscal variables to changes in economic variables such as output growth, inflation and interest rates; discretionary response of policy makers to economic innovations that is assumed to be zero; and a random structural shock to the fiscal variables that is not correlated with any other shocks. We need to identify this last component of random structural shocks in order to estimate the impulse response of macro variables to an unexpected exogenous change in fiscal variables. More formally, the reduced form fiscal shocks can be expressed as a combination of automatic response of fiscal variables and random structural shocks to fiscal variables:

[1]The AIC, SIC and HQ information criterion all indicate the optimal lag length to be 2 quarters. we allow for a minimum of 4 quarters due to the annual budget cycle since the tax receipts are recorded on a cash basis, instead of an accrual basis, and would typically depend on income generated in the previous 4 quarters.

$$\begin{pmatrix} u_t^g \\ u_t^t \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_{gy} & \alpha_{gp} & \alpha_{gi} \\ \alpha_{ty} & \alpha_{tp} & \alpha_{ti} \end{pmatrix}}_{Elasticity Assumptions} \begin{pmatrix} u_t^y \\ u_t^p \\ u_t^i \end{pmatrix} + \begin{pmatrix} 1 & \beta_{gt} \\ \beta_{tg} & 1 \end{pmatrix} \begin{pmatrix} e_t^g \\ e_t^t \end{pmatrix} \tag{4.2}$$

Further, if we remove the automatic response of the fiscal variables to output, inflation and interest rate from the reduced form shocks, the resulting fiscal policy shocks can then be expressed as a linear combination of the random structural shocks to fiscal variables:

$$\begin{pmatrix} \tilde{u}_t^g \\ \tilde{u}_t^t \end{pmatrix} \equiv \begin{pmatrix} u_t^g \\ u_t^t \end{pmatrix} - \begin{pmatrix} \alpha_{gy} & \alpha_{gp} & \alpha_{gi} \\ \alpha_{ty} & \alpha_{tp} & \alpha_{ti} \end{pmatrix} \begin{pmatrix} u_t^y \\ u_t^p \\ u_t^i \end{pmatrix} = \begin{pmatrix} 1 & \beta_{gt} \\ \beta_{tg} & 1 \end{pmatrix} \begin{pmatrix} e_t^g \\ e_t^t \end{pmatrix} \tag{4.3}$$

where α_{ij} , $(i = g, t$ and $j = y, p, i)$ denote the elasticities of government spending and net taxes to output, inflation and interest rates; and e_t^t and e_t^g denote the structural fiscal shocks that are part of the structural shock vector $E_t = [e_t^g, e_t^t, e_t^y, e_t^p, e_t^i]$. To identify the two structural fiscal shocks in this 5 variable VAR, 10 restrictions are needed. Following Blanchard and Perotti (2002), we get two restrictions from normalization of the two fiscal shocks and one from the orthogonality assumption between these two shocks. Six restrictions come from spending and tax elasticity assumptions (α_{ij}'s) that are estimated using institutional information on a country's fiscal structure and are discussed below. For the last identification restriction, the literature usually assumes one of the two possible Cholesky orderings that either government spending is ordered before taxes ($\beta_{gt} = 0$) or taxes are ordered before spending ($\beta_{tg} = 0$). If we have succeeded in accounting for the systematic part of the policy responses due to the economy, then the resulting fiscal shocks are due to a shock to the revenue and spending. We argue it is more reasonable to assume that the correlation here would be small, as opposed to assuming a systematic positive or negative correlation between revenue and spending. The two fiscal elasticities (β_{gt} and β_{tg}) that determine this correlation should, therefore, be close to zero. Indeed, when we set one of these to zero, the other one comes out to be close to zero. Taking either of these Cholesky orderings doesn't affect the results and we present our baseline results for ($\beta_{gt} = 0$).

It must be admitted that this assumption of zero correlation is a substantive identifying assumption. Besides these two Cholesky orderings, there also exist infinitely many other identification possibilities. In fact, in the absence of any theory, one could assume any rotation of the two shocks. We consider some of these other identification schemes in Section 3. Allowing for these other identification schemes delivers significantly different results. We show that sufficiently 'large' values for these two fiscal elasticities can produce a statistically significant result in either direction for the government spending multiplier at positive interest rates depending on whether we set (β_{tg}) to be negative or positive. In the baseline results, government spending multiplier is not statistically significant. Assuming either of the two Cholesky orderings is, therefore, a substantive assumption that has a direct bearing on the significance of the results. We leave the complete analysis of these other identification schemes to future work.

The identification of the structural fiscal shocks is helped by two assumptions. First, the discretionary response of policymakers to innovations in macro variables is delayed due to decision and implementation lags by more than one quarter and therefore this discretionary response of policymakers at a quarterly frequency is assumed to be zero. The elasticities, therefore, only represent the automatic response. Second, we assume specific values for these elasticities that are estimated using external information on spending and tax code structure. It is not possible to compute these elasticities by simply running an OLS of u_t^g on u_t^y, u_t^p and u_t^i. This would amount to a Cholesky ordering in which the fiscal variables are ordered last and will not provide correct estimates as output, inflation and interest rates could all respond to fiscal shocks in the same quarter.

$$\underbrace{\begin{pmatrix} 1 & 0 & -g_y^* & -g_p^* & -g_i^* \\ 0 & 1 & -t_y^* & -t_p^* & -t_i^* \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} u_t^g \\ u_t^t \\ u_t^y \\ u_t^p \\ u_t^i \end{pmatrix}}_{U_t} = \underbrace{\begin{pmatrix} 1 & 0^* & 0 & 0 & 0 \\ t_g & 1 & 0 & 0 & 0 \\ y_g & y_t & 1 & 0^* & 0^* \\ p_g & p_t & p_y & 1 & 0^* \\ i_g & i_t & i_y & i_p & 1 \end{pmatrix}}_{\widehat{B}} \underbrace{\begin{pmatrix} e_t^g \\ e_t^t \\ e_t^y \\ e_t^p \\ e_t^i \end{pmatrix}}_{E_t}$$

(*) indicates identifying assumptions.

The above equation summarizes the relationship between the reduced form shocks, U_t, and the structural shocks, E_t. Here output is ordered before inflation and interest rates, and inflation is ordered before interest rates. This ordering allows us to estimate the structural shock for output by using the structural shocks for spending and taxes estimated earlier as instruments in the output equation. Once the structural shock for output is estimated we can similarly obtain the structural shocks for inflation and interest rates. The ordering of output, inflation and interest doesn't matter as long as we are only interested in studying the effect of exogenous shocks to fiscal variables. We can now write $U_t = \widehat{R}E_t$, where $\widehat{R} = A^{-1}\widehat{B}$. Writing the VAR in companion form, we get $Y_t = AY_{t-1} + U_t$ and $U_t = RE_t$. Therefore the impulse response estimates to the structural shocks at horizon i

are given by $\widehat{R}\widehat{A}^i$.

The units of the structural shocks are in percentage terms. So the impulse response estimates give us the percentage change in one variable in response to a percent change in another. For ease of interpretation and to report the multipliers in currency units, we evaluate the original impulse responses that have the dimension of elasticities at the point of means to get the yen change in one variable for a yen change in another. [2] In addition to the impulse responses for spending, taxes and output in currency units, we report the impulse response for real interest rate in percentage units estimated as the impulse response of inflation subtracted from the impulse response of nominal interest rate. For the post 1995 period, when the zero bound is in effect, the impulse response for real interest rate is estimated simply as the negative of the impulse response for inflation.

4.2.2 Elasticities

Taxes are typically divided into four categories- corporate taxes, personal income taxes, indirect taxes and social security contributions. The OECD computes the output elasticities of taxes for each of these four categories for most member countries using information on the tax code of these countries. For Japan, these elasticities are obtained from Giorno, Richardson, Roseveare, and Van den Noord (1995) for the year 1991. The output elasticity of net taxes is then computed as a weighted average of these individual elasticities, the weights given by the share of each tax category in net taxes.

---

[2] If the effect of a 1 percent increase in taxes on output is to lower it by k percent and the mean level of taxes (not log taxes) is $\overline{T}$, and the mean level of output is $\overline{Y}$, then a Yen 1 increase in taxes lowers output by Yen $k(\overline{Y}/\overline{T})$.

As noted in Perotti (2004), the price elasticity of income taxes and social security taxes, holding constant employment, output and real wage, is equal to the elasticity of the tax rate to real earnings. This is obtained by subtracting 1 from the OECD estimate of the elasticity of tax revenues to average real earnings. These are also obtained from Giorno, Richardson, Roseveare, and Van den Noord (1995). Following Perotti (2004), we assume zero price elasticity for real corporate taxes and real indirect taxes. Inflation has complex effects on corporate taxes in both directions and any attempts to quantify those effects are beyond the scope of this chapter. Nominal indirect taxes are assumed to be proportional to the price level and hence the zero price elasticity for real indirect taxes. The average price elasticity for net taxes is then computed as a weighted average of price elasticity of individual tax categories. The interest elasticity of net tax revenues is assumed to be zero. The individual income tax base includes both interest income and dividend income, while the former may positively correlated with interest rates the latter may covary negatively. Therefore, we assume zero interest elasticity of net taxes.

Output elasticity of real government spending is assumed to be zero. In most countries, like Japan, fiscal policy contains non-discretionary automatic stabilizers via transfer payments such as unemployment benefits and social security. These expenses are, however, not part of government consumption or government investment spending and real government spending is for most part unresponsive to a contemporaneous increase in output. An exception, would be the automatic indexation of government spending on some goods and services such as health care or the wage bill to the level of economic activity. However, since budgets are fixed in nominal terms, any such automatic response would be small and is, therefore, assumed to be zero. In Japan, government expenditure on wages is not indexed to inflation and therefore the price elasticity of real government spending on the wage component would equal -1. Government spending on some goods and services such as health care spending might be indexed to the price level in the current quarter thereby implying zero price elasticity. Spending on other goods and services that is not indexed would have a price elasticity of real spending equal to -1. On average price elasticity of government spending is assumed to equal -0.5 in the benchmark specification. Interest elasticity of government spending is assumed to be zero. Table 1 lists the various elasticity assumptions used in the chapter.

4.2.3 Confidence Intervals and Joint Test of Equality

We use the Kilian (1999) bootstrap after bootstrap approach to estimate the confidence intervals for the impulse response function to fiscal policy shocks. He shows that in small samples the bias-corrected bootstrap after bootstrap confidence intervals explicitly account for the bias and the skewness of the distribution of the impulse response estimator and tend to be more accurate than delta method, standard bootstrap intervals and Monte Carlo integration intervals. The steps for estimating the bias-corrected bootstrap after bootstrap impulse response path have been taken from Kilian (1999) and are provided in the appendix.

We compute 2000 bootstrap after bootstrap estimates of the impulse response path. Following the standard practice in the structural VAR literature for reporting the confidence intervals for fiscal policy multipliers, we estimate and plot 68% bootstrap after bootstrap confidence interval for the impulse responses.

4.2.4 Test Statistic and P-value

We test if the short run impulse response time path of the macro-variables in response to a fiscal shock up to and including 8 quarters is different for the zero bound period in comparison to the period with positive overnight rates. This joint test of equality is done by computing the Wald statistic for testing the null hypothesis $H_0 : \phi(i,j)^{pos} = \phi(i,j)^{zero}$ where $\phi(i,j)$ is the short run impulse response time path for variable i to shock j including the period impulse response starting from the contemporaneous quarter up to 8 quarters ahead. The Wald statistic is computed as follows:

Suppose that a is the vector of the difference in estimated impulse responses $(\phi(i,j)^{pos} - \phi(i,j)^{zero})$ in the data. Let V be the asymptotic variance-covariance matrix, where $V = Var(\phi(i,j)^{pos}) + Var(\phi(i,j)^{zero})$. And let b be the difference in estimated impulse responses in the bootstrap sample. The Wald statistic for the sample at hand is, $Wald = a' * inv(V) * a$. Next we compute the bootstrap Wald statistic for each of the 2000 bootstrap impulse response estimates as, $Wald_{boot} = (b-a)' * inv(V) * (b-a)$.

Asymptotically, the Wald statistic is distributed as a χ^2_h, where h is the number of horizons for which the equality is being tested. Since the χ^2 is a one-tailed distribution, we compute the one-tailed bootstrap p-value as the proportion of bootstrap draws for which the sample Wald statistic is less than the bootstrap Wald estimates.

4.2.5 Limitations of the VAR Approach

Perotti (2004) discusses some well noted limitations to the structural VAR approach to estimating impulse response of fiscal policy shocks. For the sake of completeness, we briefly re-visit some of these limitations here. First, it is suggested that there is often only one publicized fiscal event each year, i.e. the annual fiscal budget and therefore just one fiscal shock per year. However, as noted by Perotti, countries also tend to have meaningful mid-year budgets and together with the updates in these annual and mid-year budgets agents tend to be surprised at more like a quarterly frequency.

The second and more widely noted limitation of the VAR-based innovations is that fiscal shocks tend to be highly anticipated. Unlike decision lags that help in identification, implementation lags can be a cause for concern since most government spending and tax decisions are decided and announced much ahead of when they are implemented. Therefore, the criticism is that the estimated VAR-based innovations are valid only with respect to the information set of the econometrician and not the private sector. This criticism argues that the VAR-based innovations are not simply unexpected exogenous changes in fiscal policy but may be clouded by anticipated changes in fiscal variables. Blanchard and Perotti (2002) in their work on fiscal policy multipliers for the US, noted that explicitly allowing for implementation lag of one quarter did not change their results qualitatively and only had a negligible quantitative effect on the results. Perotti (2004) shows that the impact effect of fiscal policy shocks on output is incorrectly estimated only if the anticipated fiscal shocks or the announcement shocks that are omitted from the econometrician's estimated structural model happen to be auto-correlated.

Therefore, in the absence of any decent solution to allow for anticipation of fiscal policy in a structural VAR approach, we follow the existing literature and assume that the misestimation of the impulse responses due to anticipation effects is small. Exploring these anticipation effects and whether they would be different when the nominal interest rate is constrained at the zero lower bound is left for future work.

4.3 Data, Results and Robustness Checks

The data used in this chapter consists of quarterly observations for Japan for the period 1960Q2 to 2009Q4. The data are split into two sub-samples, pre and post 1995 representing the period of positive and near zero nominal overnight interest rate in Japan. We use the data from 1960Q2 to 1995Q1 to estimate the VAR during the positive overnight rate period for Japan and the data from 1995Q2 to 2009Q4 to estimate the VAR during the zero lower bound period for Japan. The disaggregated data for quarterly tax revenues was provided by the Tax Bureau of the Japanese Ministry of Finance. The data source for the National Accounts of Japan is SNA accounts available at web-site for the Economic and Social Research Institute of the Cabinet Office, Government of Japan. Data on overnight interest rates was taken from Bank of Japan Statistics.

Government spending, net taxes and GDP are all deflated using the GDP deflator and divided by the quarterly population to be expressed in per capita units. All variables are seasonally adjusted using the census X-12 method built in Eviews. Unit root tests (Augmented Dickey-Fuller and Phillips-Peron) were carried out to test for a deterministic and/ or stochastic trend on these seasonally adjusted real per capita variables in levels and the seasonally adjusted GDP deflator. At all levels of significance the null of a stochastic trend is not rejected for all the variables. In the unit root test, we also test for the coefficient of the time trend and the null of a deterministic trend is rejected at the 5% significance level. The unit root tests, therefore, confirm non-stationarity of the series and the presence of a stochastic trend. We take the first differences of the log levels for each variable. We follow Blanchard and Perotti (2002) in allowing for changes in the underlying drift term overtime. To do so, we subtract a changing mean[3] that is calculated as the exponential-weighted moving average of the past first differences with the decay parameter equal to 2.5% per quarter (varying this parameter makes no noticeable difference to the results). We also carried out the analysis not allowing for changes in the underlying drift term and find that it does not effect the results. Government spending variable refers to total government demand as the sum of government private consumption expenditure and government fixed capital formation. Data on government inventory is available only after 1980 and is therefore not included in government demand to keep the series consistent. The net taxes variable is the sum of income taxes, corporate taxes, indirect taxes and net social security receipts.

4.3.1 Results

Figure 4.1 plots the impulse response path of the variables to a shock to total tax revenues. The top panel plots the impulse response path for the period when the interest rates were positive (prior to 1995) and the lower panel plots the impulse response path for the zero bound period (post 1995). An increase in tax revenues lowers output at positive interest rates (Yen 1 increase in total tax revenues approximately lowers output on impact by Yen 1.2 and by Yen 0.8 in subsequent quarters) but does not have a statistically significant impact on output when the zero lower bound is in effect except for the first three quarters when it lowers output marginally by Yen 0.3 on impact. Under positive interest rates, total tax revenue shock doesn't have a statistically significant impact on real interest

[3]Certain factors such as structure of the population, human education, among others could account for the low frequency changes in the rate of growth. Explicitly accounting for these factors is beyond the scope of this chapter and these are accounted for by subtracting a changing mean from all the series.

rates. But, when the zero bound is in effect, a one percent increase in total tax revenues increases real interest rate at horizon 3 to horizon 10 by $0.02 - 0.04$ percentage points.

In Figure 4.2 we plot the impulse response path to a shock to income tax revenues. Here again we see similar qualitative results as seen in response to a shock to total tax revenues. An increase in income tax revenues puts deflationary pressure on prices and increases the real interest rate when the zero bound is in effect and doesn't have a significant impact on real interest rates under the positive interest rate period. A positive shock to income tax revenues has no significant impact on output when the nominal interest rates are constrained at zero, but it has a significant negative impact on output when positive interest rates are prevailing. These results do not seem to support the New-Keynesian theoretical argument put forward in Eggertsson (2009), namely, that a decrease in labor tax revenues at the zero bound would increase supply incentives to put deflationary pressures on the economy that would tend to increase the real interest rates and lower output.

Figure 4.3 plots the impulse response path of variables to a shock to corporation tax revenues. An increase in corporation tax revenues increases output when the nominal interest rates are restricted at the zero bound and lowers output otherwise. This happens to be one of the most statistically significant results in the chapter. Arin and Koray (2006) also find that corporation tax innovations have a positive impact on output for Canada. However their result is for the period 1960 to 1999, when Canada wasn't faced with the zero lower bound on interest rates. This result does seem to be very puzzling. In order to explain this result we assume that corporations issue higher debt in order to get a tax break and then use that debt to expand output. This explanation is rooted in the corporate finance structure. If companies want to finance a project, they can do it using debt or equity.

Debt finance allows for tax deductions, but at the same time involve a higher interest cost compared to equity finance. At the ZLB, since the interest rates are low it is prudent to use debt finance. Therefore at the ZLB, when corporate taxes are increased, companies issue even more debt to avoid higher taxes and then use this debt to expand output. On the other hand, if corporate taxes are increased during a period of positive interest rates, companies do not have a similar incentive of issuing more debt as that would be highly expensive for firms. Further, reverse causality that would render this argument is ruled out. Since taxes are proportional to income, increased output leads to increased taxes, but this has already been accounted for in the structural VAR. Any other arguments in favor of reverse causality, such as, firms having planned ahead of an output increase, should be quantitatively small and are, therefore, ignored.

Figure 4.4 plots the impulse response path of variables to a shock to indirect tax revenues. Here we find that a positive innovation in indirect taxes results in declining output at the zero lower bound, but has no significant impact on output under positive interest rates. In 1994 September, it was announced in Japan that there will be a permanent increase in indirect tax (consumption tax) starting from April 1997. Consumption tax revenues on average account for about 30% of all indirect tax revenues in Japan and this period falls in the zero lower bound sample. The negative impact on output at the zero lower bound could be explained by the anticipation effect of increased consumption tax on consumer spending following the announcement. If consumers pulled forward their spending in anticipation of an imminent permanent rise in consumption tax, on implementation of this tax it would show as a negative impact on output. An innovation in indirect tax revenues is neither contrationary nor expansionary at positive interest rates. The real interest rate falls by

$0.02 - 0.04$ percentage points in response to a 1 percent increase in indirect tax revenues during the period of positive interest rates.

Figure 4.5, Figure 4.6 and Figure 4.7 plot the impulse response path of the variables to a shock to aggregate government spending, government consumption spending and government investment spending. Aggregate government spending innovations do not have a significant impact on output either at positive interest rates or at the zero lower bound. Even government consumption spending or government investment spending innovations do not have a significant impact on output under the two different cases of positive interest rates and the zero lower bound.

Table 2 reports the Wald-statistic for the joint test of equality of the short-run impulse response time path for the two distinct time periods, up to and including 8 quarters after impact, for taxes and output to tax innovations. The number in the parenthesis is the associated bootstrap p-value for the reported Wald-statistic. We note that the impulse response path of output to aggregate tax shock is not significantly different for the positive interest rate period and the zero lower bound period. The impulse response path of taxes and output, however, in response to a corporate tax innovation and an indirect tax innovation is significantly different for the two time periods. These results therefore indicate that different tax innovations impact output differently at positive interest rates and at the zero lower bound. Using aggregate tax data may conceal the differences of tax innovations on the output during the two sub-periods considered for Japan.

4.3.2 Robustness

As shown earlier, we can express the reduced form fiscal policy shocks after accounting for the automatic response of fiscal policy to output, inflation and interest rates as a linear combination of structural fiscal shocks.

$$\begin{pmatrix} \widetilde{u}_t^g \\ \widetilde{u}_t^t \end{pmatrix} = \begin{pmatrix} 1 & \beta_{gt} \\ \beta_{gt} & 1 \end{pmatrix} \begin{pmatrix} e_t^g \\ e_t^t \end{pmatrix}$$

The benchmark results presented in the chapter assume $\beta_{gt} = 0$. As argued in the introduction, in the absence of theory, one could consider any rotation of the two shocks and these infinitely many identification schemes, in addition to the two Cholesky orderings, imply significantly different results for the fiscal policy multipliers. In this section, we want to emphasize that assuming a Cholesky ordering is a substantive assumption that is driving the results presented in the chapter. To illustrate this we plot the fiscal policy multipliers for the government spending shock for $\beta_{tg} = -5$ and $\beta_{tg} = 5$. When we set $\beta_{tg} = -5$, β_{gt} is estimated to be equal to 0.27 and the government spending multiplier for output for the positive interest rate period turns out to be significantly positive (Figure 4.8). On the other hand, if we set $\beta_{tg} = 5$, β_{gt} is estimated to be equal to -0.25 and the government spending multiplier for output for the positive interest rate period turns out to be significantly negative (Figure 4.9). Therefore, assuming one of the two Cholesky orderings in favor of the infinitely many other identification schemes is crucial to the results presented here. However, as mentioned earlier, a value for the tax revenue elasticity to a government spending shock closer to zero is a more reasonable assumption. On the other hand, it is a difficult pill to swallow that the contemporaneous effect of a government spending shock on tax revenues reduces the revenues by five fold. One should take both the qualitative and the quantitative results presented in this chapter seriously only if one believes in the identification assumption given by either of the two Cholesky orderings.

4.4 Conclusion

Up to now, there exists little empirical evidence on the effects of fiscal policy at the zero lower bound due to the lack of data. Most countries have only experienced the zero lower bound for a few quarters since the beginning of the 2008 global downturn. Japan is the only country that has been stuck at the zero lower bound for an extended period since 1995. In this chapter we assess how much information there is in the admittedly short Japanese sample that might shed some light on the various fiscal policy choices faced by policy makers today. As one might suspect, the exercise revealed that the data is inconclusive about the various issues surrounding the fiscal policy debate at the zero bound. However, we showed that behavior of overall taxes was much different from the behavior of disaggregated taxes. Even though, the impact of a cut in overall taxes was insignificant at the zero lower bound, a corporate tax cut surprisingly lowered output and an indirect tax cut raised output in Japan at zero interest rate. We also find that government spending has not been effective in Japan to raise output, not only during the zero bound period but also prior to that.

One of the contributions of this chapter is to show that the results are highly sensitive to the assumption of Cholesky ordering for the fiscal shocks that is often assumed to be innocuous in the literature. Allowing for different rotations of the fiscal shocks besides the two Cholesky orderings causes the results to be significantly different. Therefore, when looking at results from a structural VAR fiscal policy exercise one should be mindful of the specific assumptions made about the ordering of the shocks and look at the results as being contingent on a particular specification. A different specification, as we show in this chapter, can result in a statistically significant and opposite result.

Given these caveats with the structural VAR literature, alternative methodologies that bring in more information from the available data should be used in future empirical work for estimating fiscal policy multipliers. This is especially true when one has to work with a small sample size. One should, for instance, focus on Romer and Romer (2007) style tax and spending based narratives of fiscal policy changes for Japan to study the effects of fiscal policy during the zero bound period.

4.5 Appendix

4.5.1 Estimating Bootstrap After Bootstrap Impulse Response Path

Step1: I estimate a VAR(p) in companion form to get estimates of the intercept term, the coefficient matrix $\widehat{A}$ and the reduced form residuals. Step2: I generate 1000 recursive bootstrap time series. To do so I pick 4 initial lags by randomly drawing a block of 4 adjacent sample values from a uniform distribution of available sample values and then generate a time series recursively by randomly drawing residuals from a uniform distribution of the reduced form residuals estimated in Step 1. Step3: Based on the bootstrap time series, I get 1000 bootstrap estimates for the companion matrix $\widehat{A}^*$. Step4: Compute the bias as average of the bootstrap estimates minus the initial estimate. Step5: Construct the bias corrected coefficient estimate $\widetilde{A}$. $\widetilde{A}$ is set to $\widehat{A}$ if the modulus of the largest root associated with $\widehat{A}$ is ≥ 1. If not, then $\widetilde{A}$ is set to $\widehat{A}$-bias. If the bias correction puts $\widetilde{A}$ in the non-stationary region then the bias is reduced using a grid until the bias corrected estimate is stationary. Step6: I generate 2000 bootstrap time series using $\widetilde{A}$ and bootstrap residuals from step 2. Step7: On each bootstrap time series I compute the bias corrected companion matrix $\widetilde{A}^*$ using the bias estimated in step 4 and then compute the contemporaneous impulse response matrix $\widehat{R}$ using the identification strategy discussed in section 2. Step 8: For each bootstrap, I compute the impulse response path as $\widehat{R}\widetilde{A}^{*^i}(i = 0, 1, 2, ...8)$.

Table 4.1: Output, Price and Interest Elasticity of Spending and Taxes

Elasticity w.r.t.→ Real Variables↓	Output	Price	Interest Rate
Corporate Taxes	3.7	0.0	0.0
Income Taxes	1.2	1.6	0.0
Indirect Taxes	1.0	0.0	0.0
Social Security Transfers	0.6	0.0	0.0
Net Taxes	1.8	0.6	0.0
Government Consumption Spending	0.0	-0.5	0.0
Government Investment Spending	0.0	-0.5	0.0
Government Spending	0.0	-0.5	0.0

Table 4.2: Joint Test of Equality for the Impulse Response Path to a Tax Shock for Horizons (0-8): Wald Test

Type of Tax	$H_0 : \phi(T,T)^{pos} = \phi(T,T)^{zero}$ Effect of T on T	$H_0 : \phi(Y,T)^{pos} = \phi(Y,T)^{zero}$ Effect of T on Y
Income Tax	13.50	5.41
	(0.17)	(0.78)
Corporate Tax	74.43	39.76
	(0.00)	(0.00)
Indirect Tax	18.08	17.35
	(0.05)	(0.06)
Total Tax	21.25	8.06
	(0.02)	(0.51)

*The number in the parenthesis denotes the bootstrap p-value for the Wald statistic

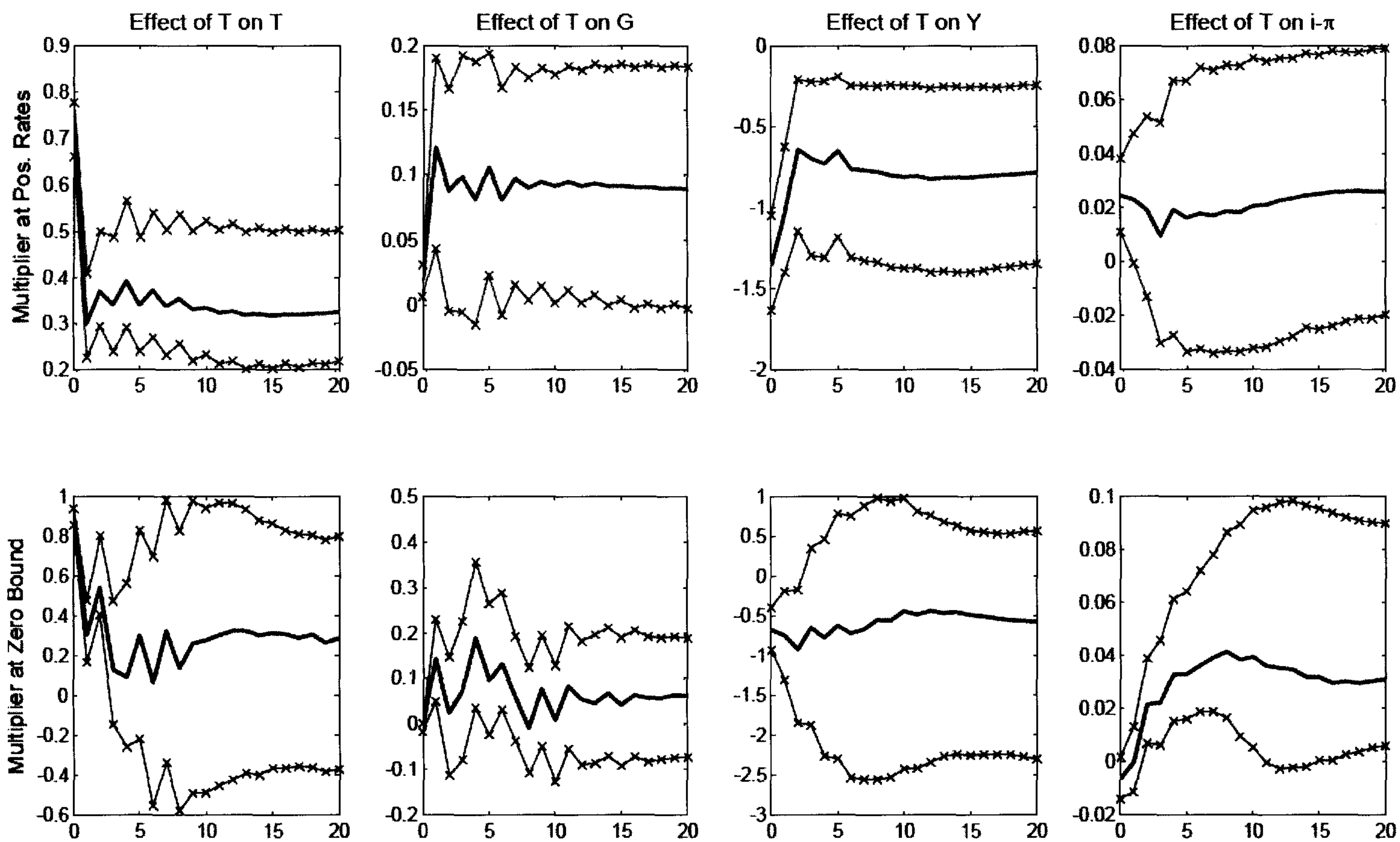


Figure 4.1: Effect of Shock to Total Tax Revenues with 68% confidence bands.

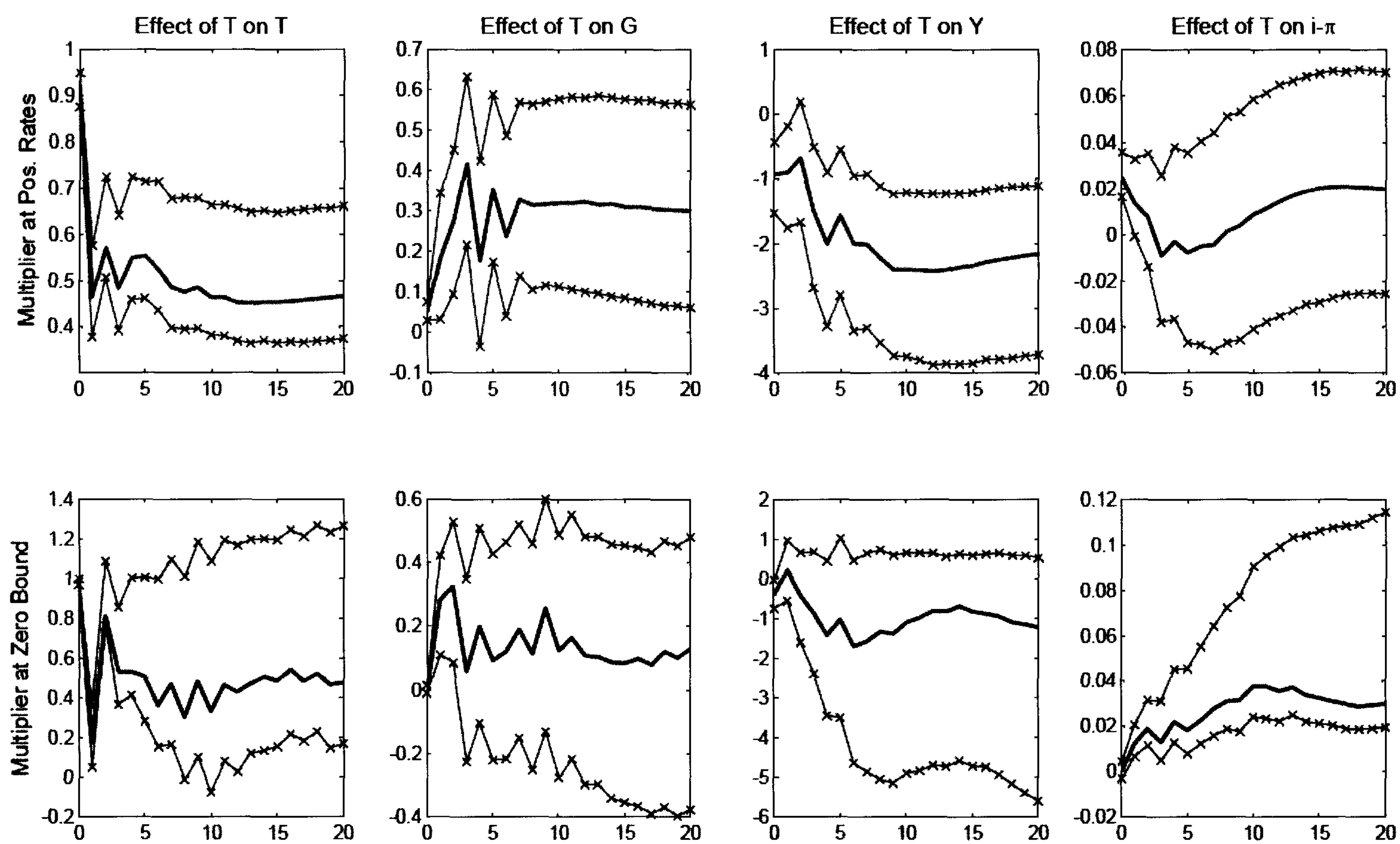


Figure 4.2: Effect of Shock to Income Tax Revenues with 68% confidence bands.

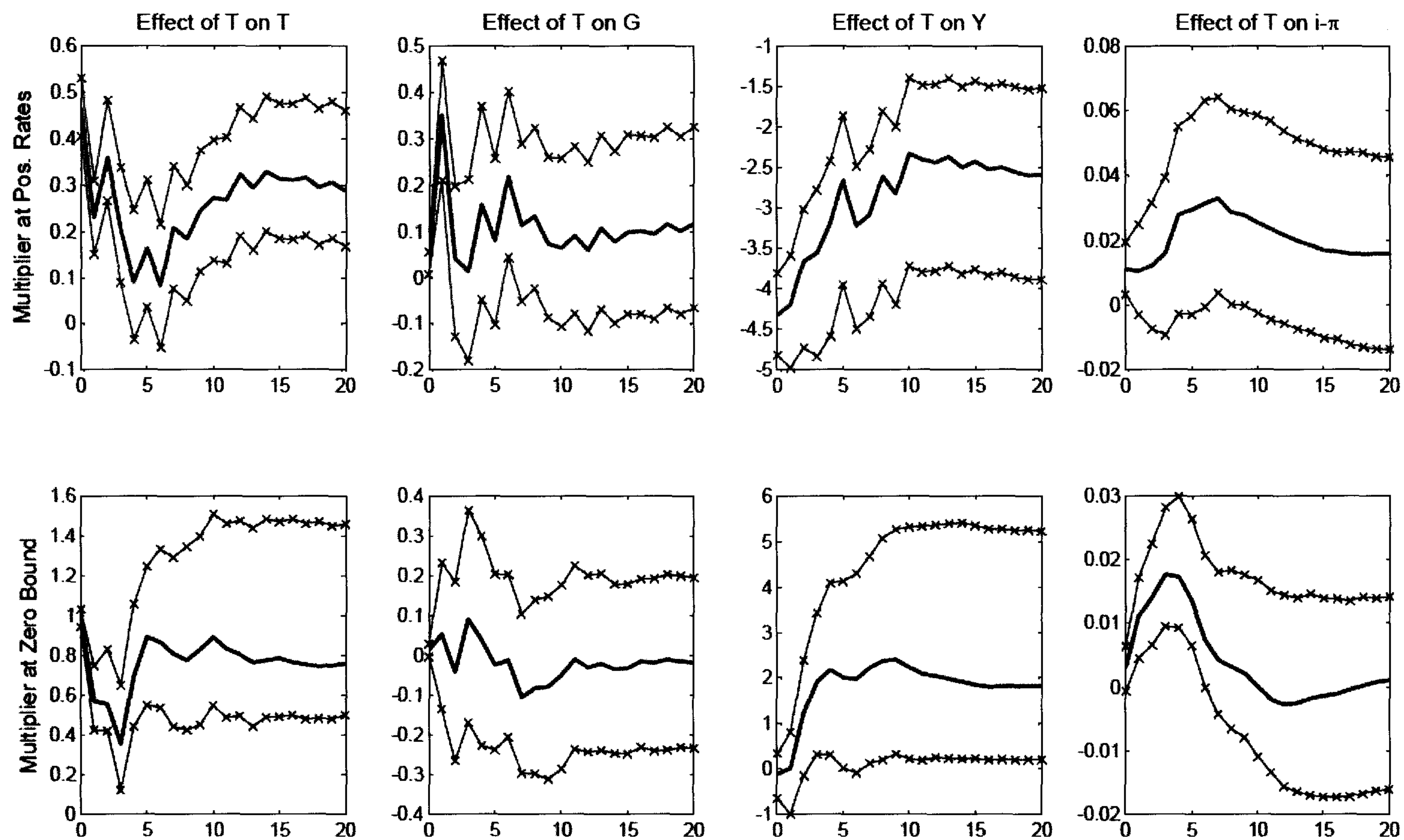


Figure 4.3: Effect of Shock to Corporation Tax Revenues with 68% confidence bands.

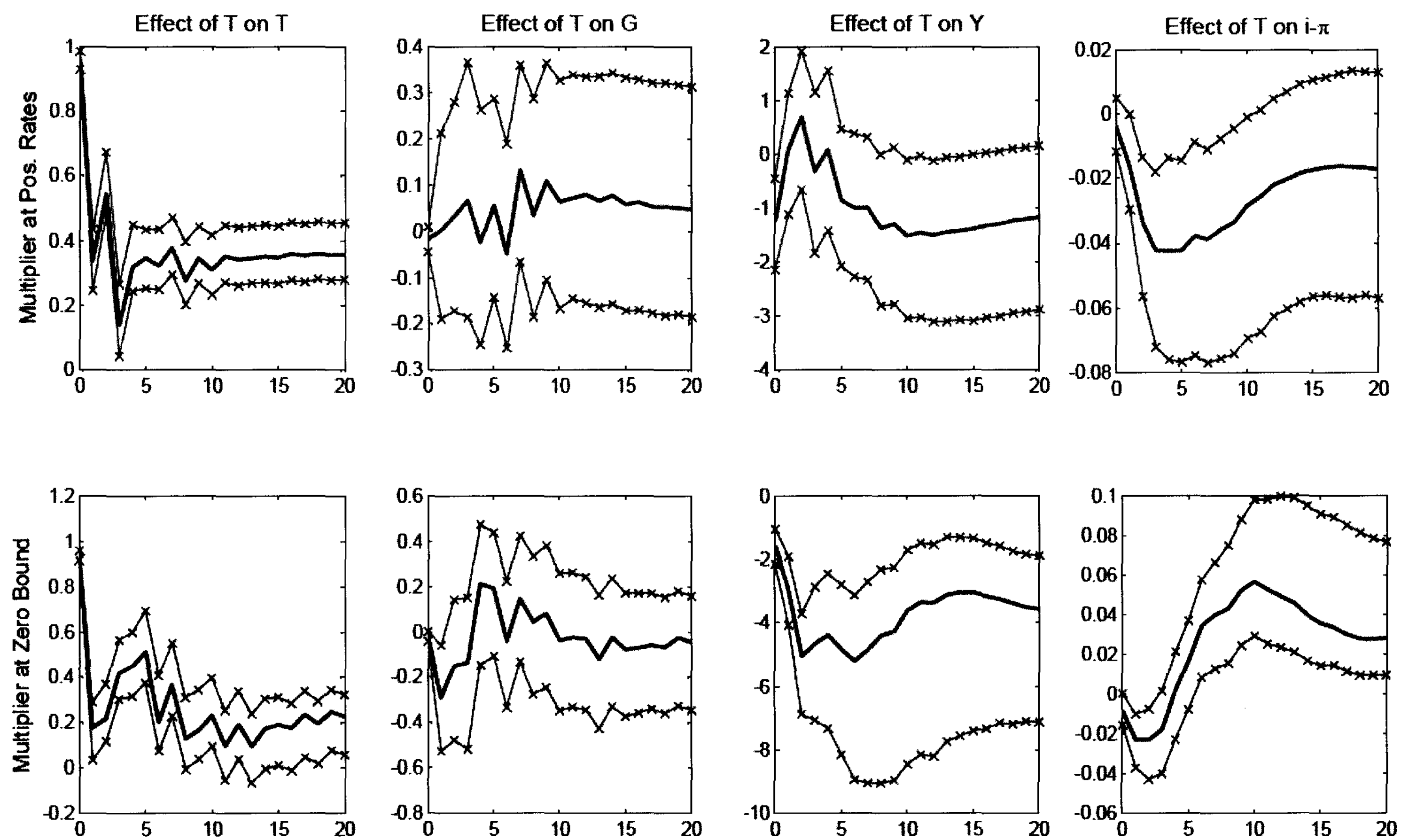


Figure 4.4: Effect of Shock to Indirect Tax Revenues with 68% confidence bands.

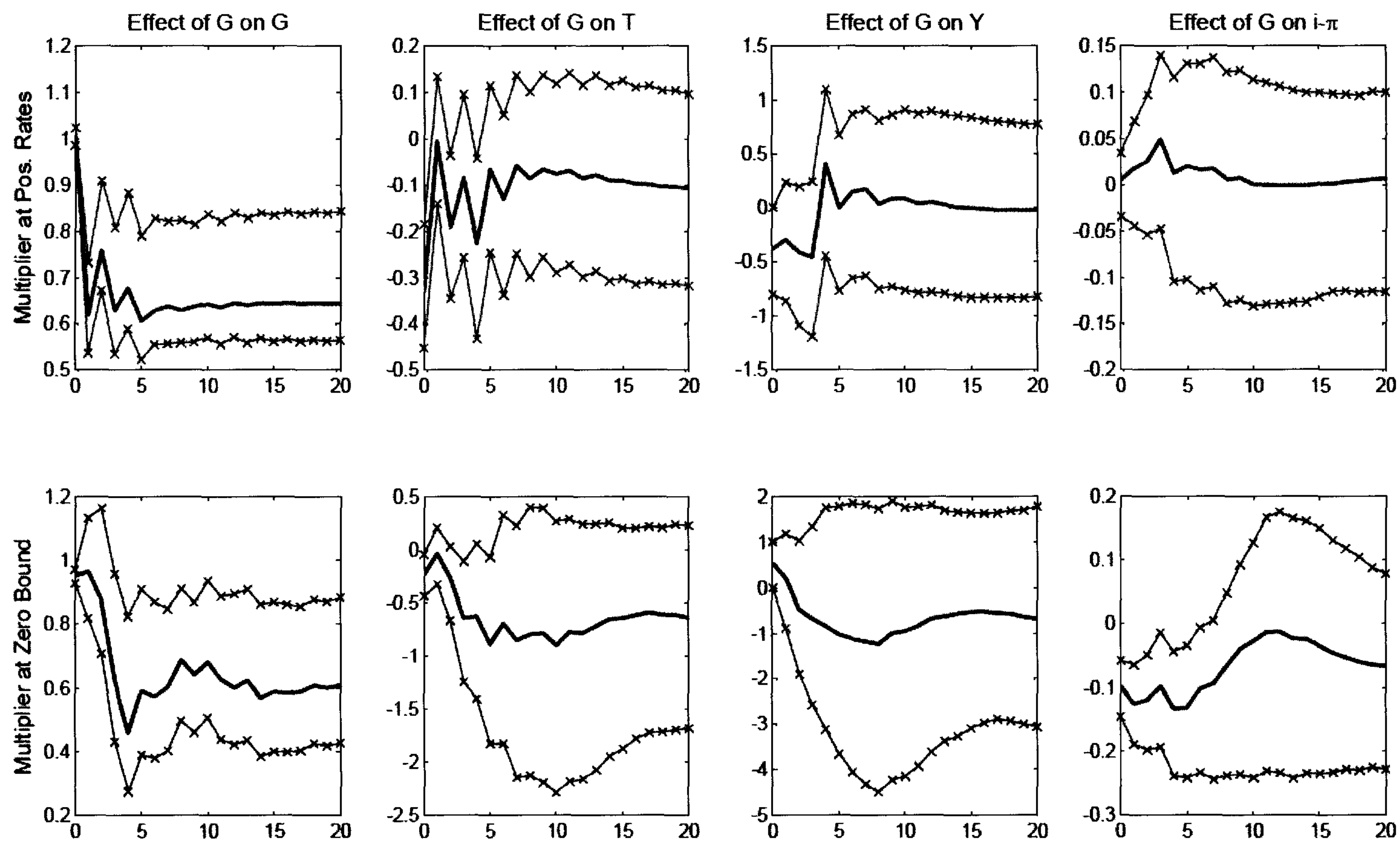


Figure 4.5: Effect of Shock to Aggregate Government Spending with 68%confidence bands.

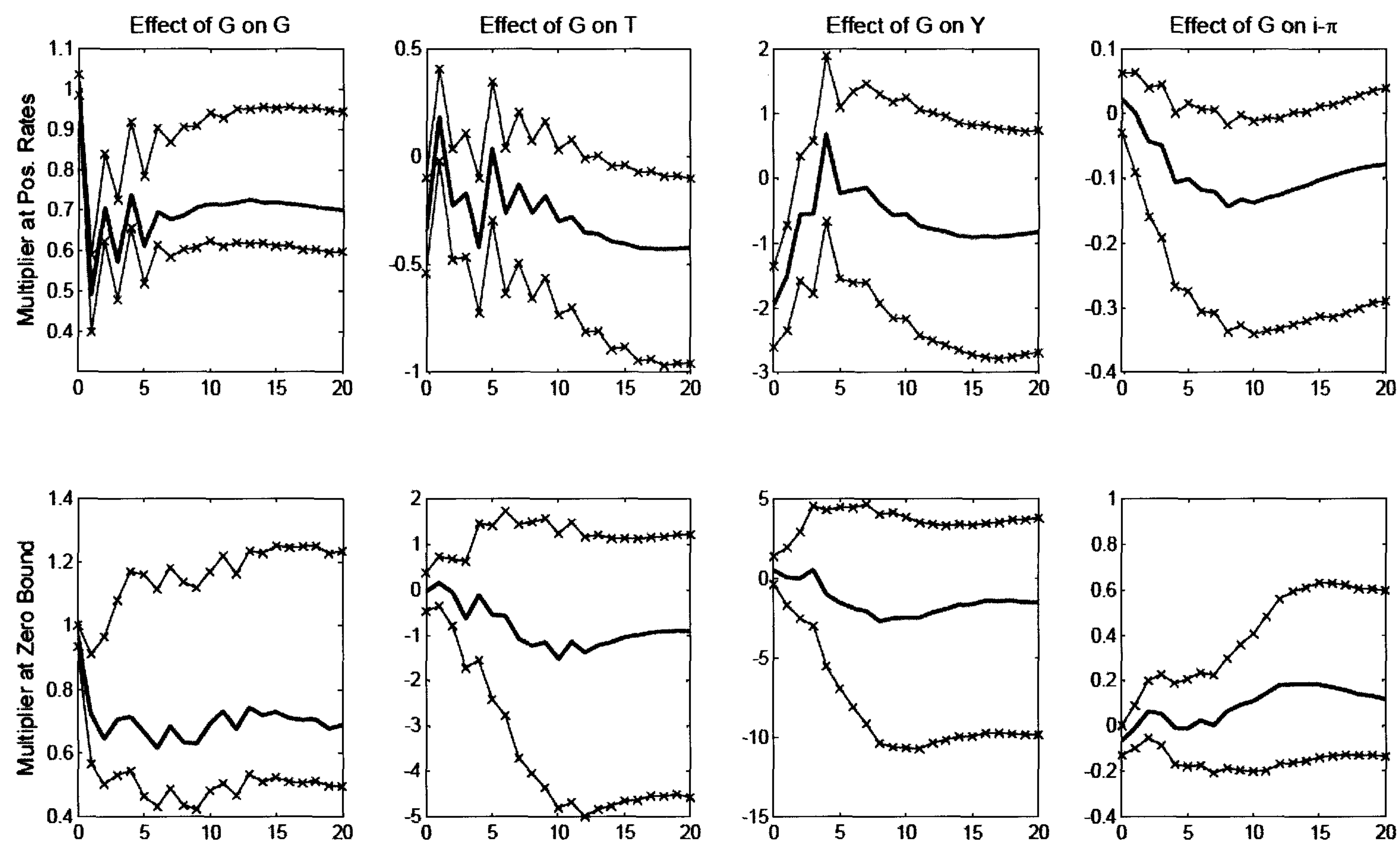


Figure 4.6: Effect of Shock to Government Consumption Spending with 68%confidence bands.

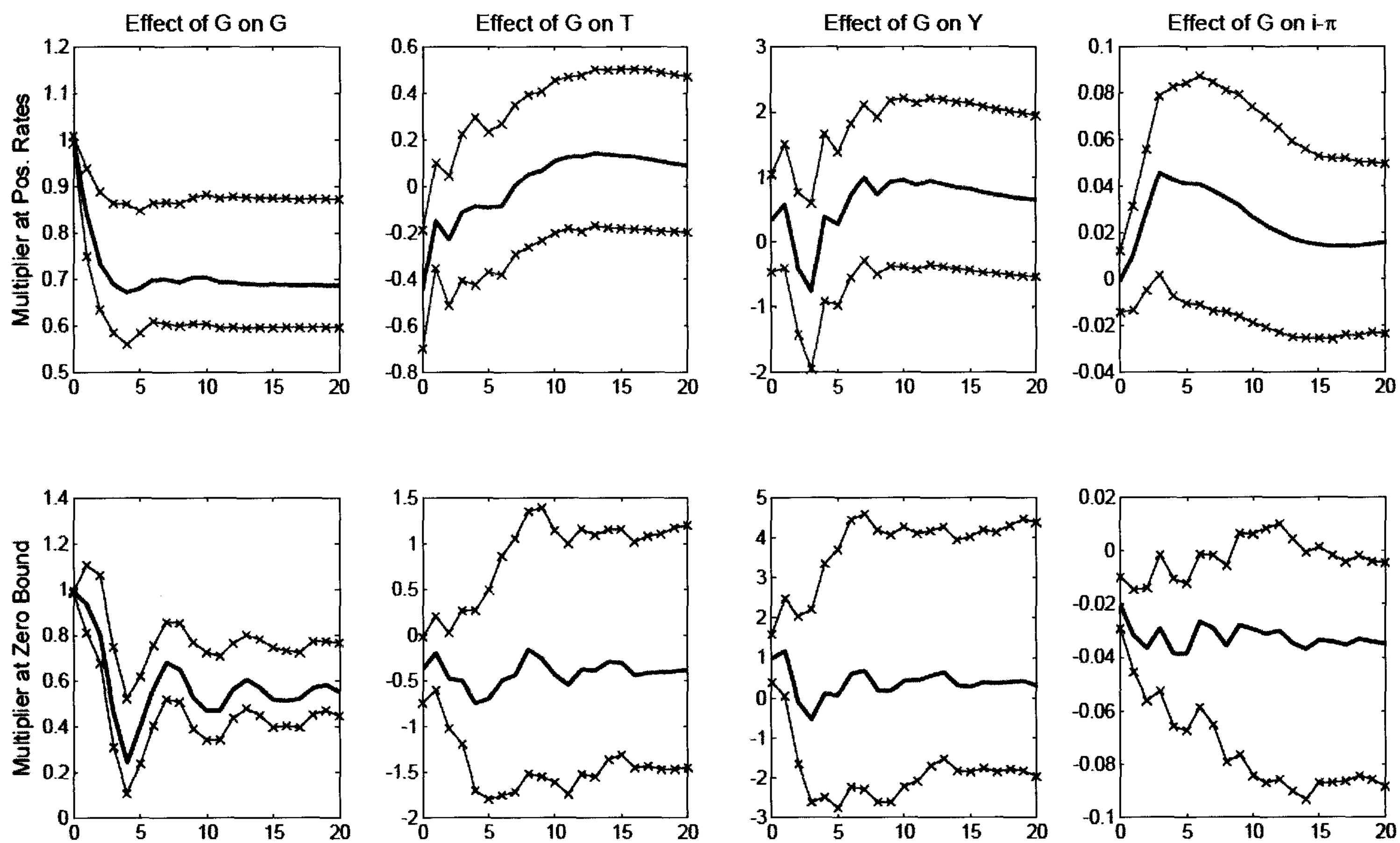


Figure 4.7: Effect of Shock to Government Investment Spending with 68% confidence bands.

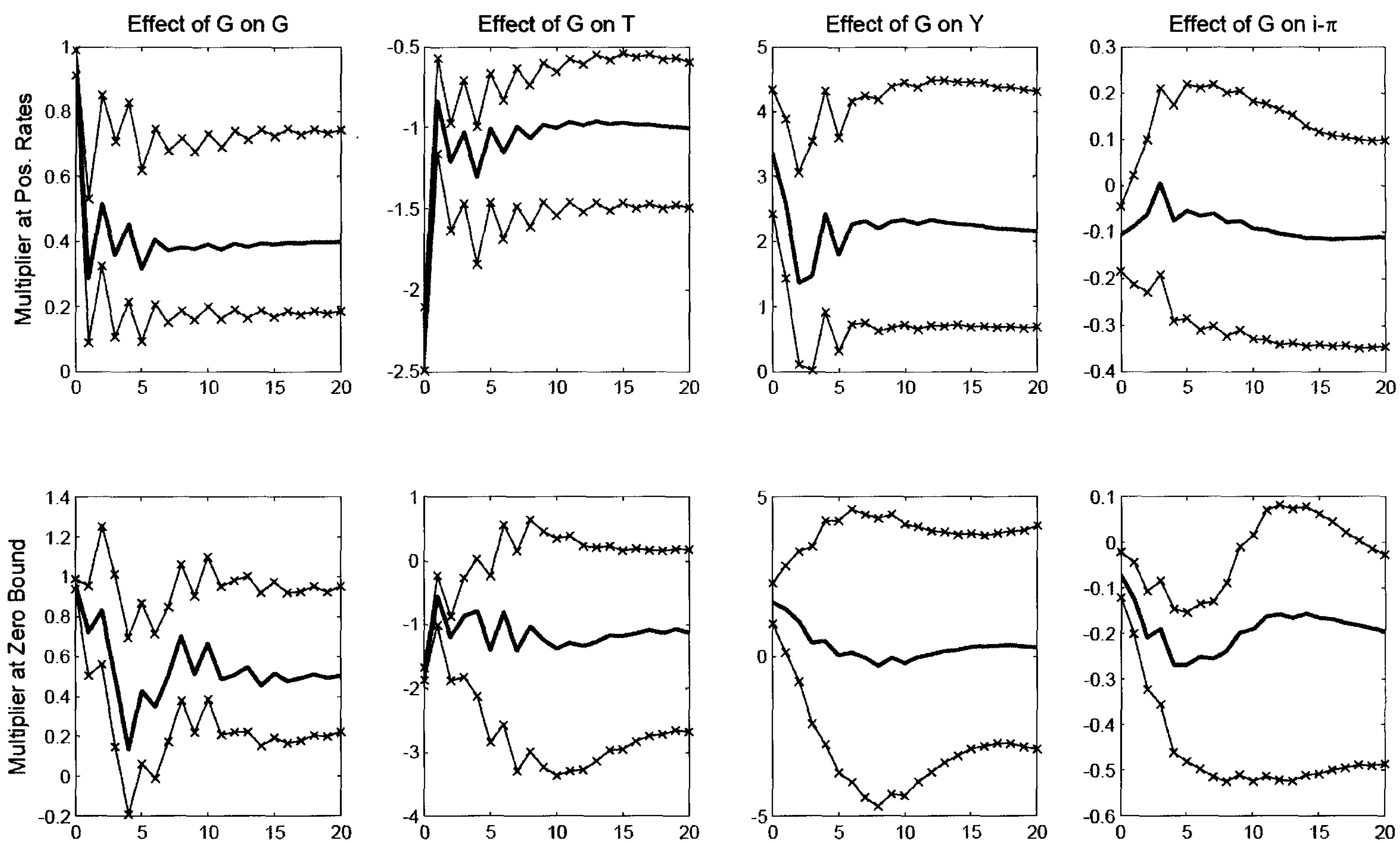


Figure 4.8: Robustness Check: Effect of Shock to Government Spending with 68%confidence bands. β_{tg}=-5; β_{gt}=0.27

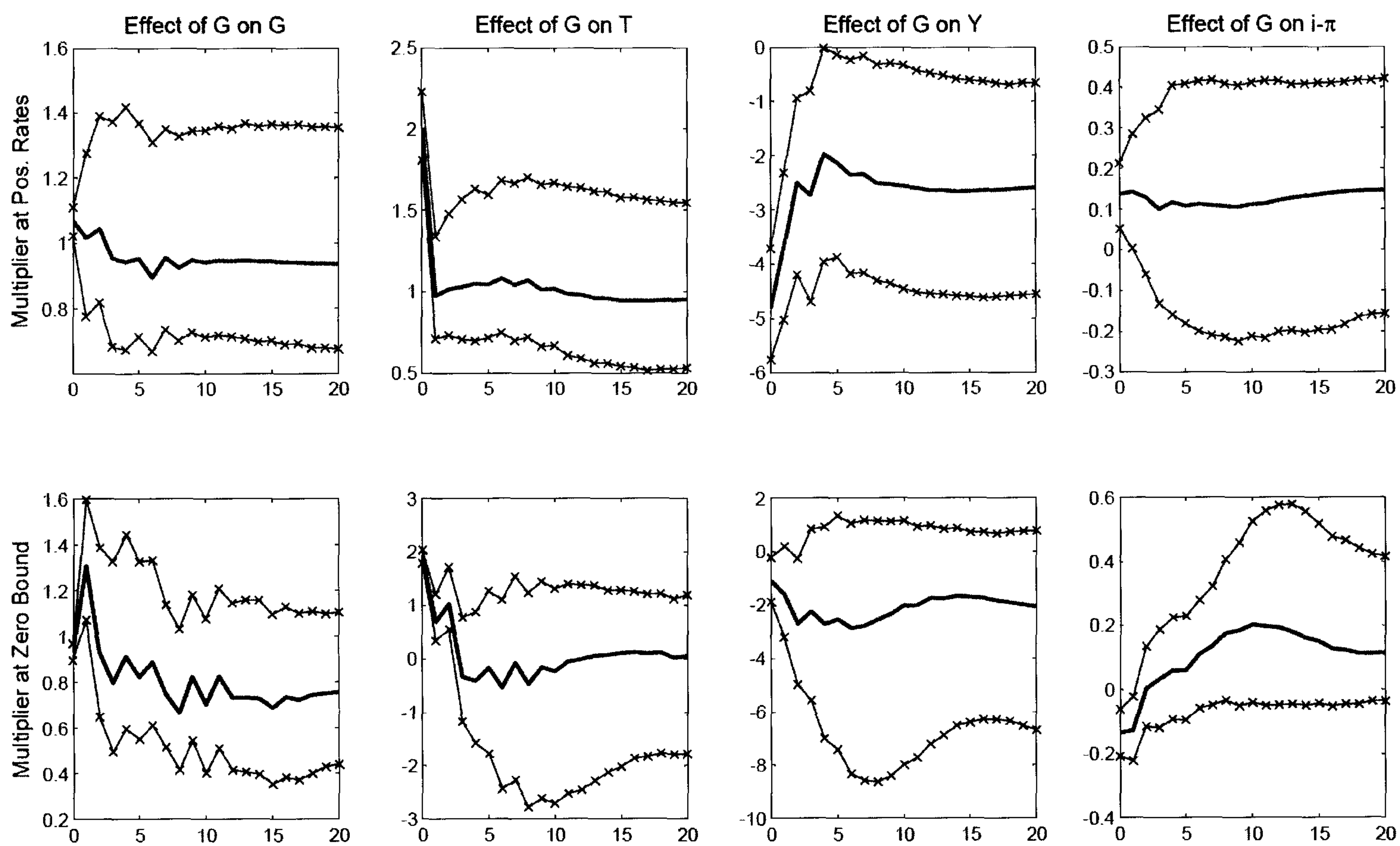


Figure 4.9: Robustness Check: Effect of Shock to Government Spending with 68% confidence bands. β_{tg}=5; β_{gt}=-0.25

Bibliography

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.*

ARIN, K., AND F. KORAY (2006): "Are some taxes different than others? An empirical investigation of the effects of tax policy in Canada," *Empirical Economics*, 31(1), 183–193.

ATTANASIO, O., AND G. WEBER (1993): "Consumption growth, the interest rate and aggregation," *The Review of Economic Studies*, 60(3), 631–649.

BAYARRI, M., AND J. BERGER (1999): "Comment on Bayarri and Berger," *Bayesian Statistics*, 6, 53–67.

BERGER, J., AND R. WOLPERT (1984): *The likelihood principle.* Institute of Mathematical Statistics.

BERNANKE, B., AND J. BOIVIN (2003): "Monetary policy in a data-rich environment* 1," *Journal of Monetary Economics*, 50(3), 525–546.

BERNANKE, B., J. BOIVIN, AND P. ELIASZ (2005): "Measuring the Effects of Monetary

Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach*," *Quarterly Journal of Economics*, 120(1), 387–422.

BERNANKE, B., V. REINHART, AND B. SACK (2005): "Monetary policy alternatives at the zero bound: An empirical assessment," *Brookings Papers on Economic Activity*, 2004(2), 1–100.

BERNARDO, J. (1999): "Quantifying Surprise in the Data and Model Verification," *Bayesian Statistics*, 6, 72–73.

BLANCHARD, O., AND R. PEROTTI (2002): "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *Quarterly Journal of Economics*, 117(4), 1329–1368.

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.

CAMERER, C. (1995): "Individual decision making," *The handbook of experimental economics*, 3, 587–704.

CARROLL, C., J. SLACALEK, AND M. SOMMER (forthcoming): "International evidence on sticky consumption growth," *The Review of Economics and Statistics.*

CHARI, V. (2010): "Testimony before the U.S. House of Representatives," *House Committee on Science and Technology, Subcommittee on Investigations and Oversight.*

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Monetary Policy Shock," *Journal of Political Economy*, 113(1), 1–45.

CHRISTIANO, L., M. EICHENBAUM, AND S. REBELO (2009): "When is the government spending multiplier large?," *National Bureau of Economic Research Cambridge, Mass., USA*.

COLANDER, D. (2010): "Written Testimony of David Colander," *House Committee on Science and Technology, Subcommittee on Investigations and Oversight.*

DEL NEGRO, M., AND F. SCHORFHEIDE (2008): "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)," *Journal of Monetary Economics*, 55(7), 1191–1208.

DEL NEGRO, M., F. SCHORFHEIDE, F. SMETS, AND R. WOUTERS (2007): "On the Fit of New Keynesian Models," *Journal of Business & Economic Statistics*, 25(2), 123–143.

DYNAN, K. (2000): "Habit formation in consumer preferences: Evidence from panel data," *American Economic Review*, pp. 391–406.

EDGE, R., M. KILEY, AND J. LAFORTE (2009): "A comparison of forecast performance between Federal Reserve staff forecasts, simple reduced-form models, and a DSGE model," *Finance and Economics Discussion Series.*

EGGERTSSON, G. (2008): "Can tax cuts deepen recessions?," *Federal Reserve Bank of New York, December.*

——— (2009): "What Fiscal Policy Is Effective at Zero Interest Rates?," .

Erceg, C., and J. Lindé (2009): "Is There a Fiscal Free Lunch in a Liquidity Trap?," .

Faust, J. (2009): "The New Macro Models: Washing Our Hands and Watching for Icebergs," *Economic Review*, pp. 45–68.

Faust, J., and A. Gupta (2010a): "Posterior Predictive Analysis for Evaluating DSGE Models," *manuscript in progress, Johns Hopkins University.*

——— (2010b): "Are All Recessions Black Swans? DSGE Models and the Post-War U.S. Business Cycle," *in progress, Johns Hopkins University.*

Faust, J., and J. Wright (2009): "Comparing Greenbook and reduced form forecasts using a large realtime dataset," *Journal of Business and Economic Statistics*, 27(4), 468–479.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005): "The generalized dynamic factor model," *Journal of the American Statistical Association*, 100(471), 830–840.

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

Geweke, J. (2005): *Contemporary Bayesian econometrics and statistics.* John Wiley.

——— (2007): "Bayesian model comparison and validation," *American Economic Review*, 97(2), 60–64.

——— (2010): *Complete and Incomplete Econometric Models.* monograph, Princeton University Press.

GIANNONE, D., L. REICHLIN, AND L. SALA (2004): "Monetary policy in real time," *NBER Macroeconomics Annual*, 19, 161–200.

GILCHRIST, S., V. YANKOV, AND E. ZAKRAJSEK (2009): "Credit market shocks and economic fluctuations: Evidence from corporate bond and stock markets," *Journal of Monetary Economics*, 56(4), 471–493.

GIORNO, C., P. RICHARDSON, D. ROSEVEARE, AND P. VAN DEN NOORD (1995): "Potential output, output gaps and structural budget balances," *OECD Economic Studies*, 24, 167–209.

GUPTA, A. (2010): "A Forecasting Metric for Evaluating DSGE Models for Policy Analysis," *in progress, Johns Hopkins University.*

HANSEN, B. (2005): "CHALLENGES FOR ECONOMETRIC MODEL SELECTION," *Econometric Theory*, 21(01), 60–68.

HARVEY, A. (1991): *Forecasting, structural time series models and the Kalman filter.* Cambridge Univ Pr.

HILL, B. (1996): "Comment on Gelman, et al.," *Statistica Sinica*, 6, 767–773.

KILIAN, L. (1999): "Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses," *Review of Economics and Statistics*, 81(4), 652–660.

KYDLAND, F., AND E. PRESCOTT (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50(6), 1345–1370.

——— (1996): "The computational experiment: an econometric tool," *The Journal of Economic Perspectives*, pp. 69–85.

LANCASTER, T. (2004): *An introduction to modern Bayesian econometrics*. Wiley-Blackwell.

MEHRA, R., AND E. PRESCOTT (1985): "The equity premium: A puzzle* 1," *Journal of monetary Economics*, 15(2), 145–161.

PEROTTI, R. (2004): "Estimating the effects of fiscal policy in OECD countries," *Working Papers*.

ROMER, C., AND D. ROMER (2007): "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *NBER Working Papers*.

SIMS, C. (1980): "Macroeconomics and reality," *Econometrica: Journal of the Econometric Society*, 48(1), 1–48.

SMETS, F., AND R. WOUTERS (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

——— (2004): "Forecasting with a Bayesian DSGE Model: An Application to the Euro Area," *Journal of Common Market Studies*, 42(4), 841–867.

——— (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 97(3), 586–606.

SOLOW, R. (2010): "Building a Science of Economics for the Real World," *House Committee on Science and Technology, Subcommittee on Investigations and Oversight*.

STOCK, J., AND M. WATSON (1999): "Forecasting in# ation," *Journal of Monetary Economics*, 44(2), 293–335.

——— (2002): "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97(460), 1167–1179.

——— (2003): "Forecasting output and inflation: the role of asset prices," *Journal of Economic Literature*, 41(3), 788–829.

——— (2005): "An empirical comparison of methods for forecasting using many predictors," *Manuscript, Princeton University.*

TAYLOR, J. (2009): "The lack of an empirical rationale for a revival of discretionary fiscal policy," *American Economic Review*, 99(2), 550–555.

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

WILCOX, D. (1992): "The construction of US consumption data: some facts and their implications for empirical work," *The American economic review*, 82(4), 922–941.

Curriculum Vitae

Abhishek Gupta was born in Delhi, India on December 16, 1981. He obtained a B.A. in Economics from University of Delhi, India in 2002, and an M.A. in Economics from the same institution in 2004. He entered the Ph.D. program in Economics at Johns Hopkins University in 2004. He is presently an Assistant Professor of Economics at Gettysburg College in Gettysburg, Pennsylvania.