

The Econometrics of DSGE Models

Jesús Fernández-Villaverde*

University of Pennsylvania, NBER, and CEPR

January 19, 2009

Abstract

In this paper, I review the literature on the formulation and estimation of dynamic stochastic general equilibrium (DSGE) models with a special emphasis on Bayesian methods. First, I discuss the evolution of DSGE models over the last couple of decades. Second, I explain why the profession has decided to estimate these models using Bayesian methods. Third, I briefly introduce some of the techniques required to compute and estimate these models. Fourth, I illustrate the techniques under consideration by estimating a benchmark DSGE model with real and nominal rigidities. I conclude by offering some pointers for future research.

Keywords: DSGE Models, Likelihood Estimation, Bayesian Methods.

JEL classification numbers: C11, C13, E30

*E-mail: jesusfv@econ.upenn.edu. Much of the research reviewed in this paper was undertaken jointly with Juan Rubio-Ramírez, the best coauthor I could have hoped for. I thank Antonio Cabrales and Pedro Mira for the invitation to deliver the lecture that led to this paper, Wen Yao for research assistance, and the NSF for financial support.

1. Introduction

This article elaborates on a basic thesis: the formal estimation of dynamic stochastic general equilibrium (DSGE) models has become one of the cornerstones of modern macroeconomics. The combination of rich structural models, novel solution algorithms, and powerful simulation techniques has allowed researchers to transform the quantitative implementation of equilibrium models from a disparate collection of ad hoc procedures to a systematic discipline where progress is fast and prospects entrancing. This captivating area of research, which for lack of a better name I call the *New Macroeconometrics*, is changing the way we think about models and about economic policy advice.

In the next pages, I will lay out my case in detail. I will start by framing the appearance of DSGE models in the context of the evolution of contemporary macroeconomics and how economists have reacted to incorporate both theoretical insights and empirical challenges. Then, I will explain why the New Macroeconometrics mainly follows a Bayesian approach. I will introduce some of the new techniques in the literature. I will illustrate these points with a benchmark application and I will conclude with a discussion of where I see the research at the frontier of macroeconometrics. Because of space limitations, I will not survey the field in exhausting detail or provide a complete description of the tools involved (indeed, I will offer the biased recommendation of many of my own papers). Instead, I will offer an entry point to the topic that, like the proverbial Wittgenstein's ladder, can eventually be discarded without undue apprehension once the reader has mastered the ideas considered here. The interested economist can also find alternative material on An and Schorfheide (2006), who focus more than I do on Bayesian techniques and less in pure macroeconomics, and in Fernández-Villaverde, Guerrón-Quintana, and Rubio-Ramírez (2008), where the work is related with general issues in Bayesian statistics, and in the recent textbooks on macroeconometrics by Canova (2007) and DeJong and Dave (2007).

2. The Main Thesis

Dynamic equilibrium theory made a quantum leap between the early 1970s and the late 1990s. In the comparatively brief space of 30 years, macroeconomists went from writing prototype models of rational expectations (think of Lucas, 1972) to handling complex constructions like the economy in Christiano, Eichenbaum, and Evans (2005). It was similar to jumping from the Wright brothers to an Airbus 380 in one generation.

A particular keystone for that development was, of course, Kydland and Prescott's 1982 paper *Time to Build and Aggregate Fluctuations*. For the first time, macroeconomists had a small and coherent dynamic model of the economy, built from first principles with optimizing agents, rational expectations, and market clearing, that could generate data that resembled observed variables to a remarkable degree. Yes, there were many dimensions along which the model failed, from the volatility of hours to the persistence of output. But the amazing feature was how well the model did despite having so little of what was traditionally thought of as the necessary ingredients of business cycle theories: money, nominal rigidities, or non-market clearing.

Except for a small but dedicated group of followers at Minnesota, Rochester, and other bastions of heresy, the initial reaction to Kydland and Prescott's assertions varied from amused incredulity to straightforward dismissal. The critics were either appalled by the whole idea that technological shocks could account for a substantial fraction of output volatility or infuriated by what they considered the superfluity of technical fireworks. After all, could we not have done the same in a model with two periods? What was so important about computing the whole equilibrium path of the economy?

It turns out that while the first objection regarding the plausibility of technological shocks is alive and haunting us (even today the most sophisticated DSGE models still require a notable role for technological shocks, which can be seen as a good or a bad thing depending on your perspective), the second complaint has aged rapidly. As Max Plank remarked somewhere, a new methodology does not triumph by convincing its opponents, but rather because critics die and a new generation grows up that is familiar with it.¹ Few occasions demonstrate the insight of Plank's witticism better than the spread of DSGE models. The new cohorts of graduate students quickly became acquainted with the new tools employed by Kydland and Prescott, such as recursive methods and computation, if only because of the comparative advantage that the mastery of technical material offers to young, ambitious minds.² And naturally, in the process, younger researchers began to appreciate the flexibility offered by the tools. Once you know how to write down a value function in a model with complete markets and fully flexible prices, introducing rigidities or other market imperfections is only one step ahead: one more state variable here or there and you have a job market paper.

¹Admittedly, Plank talked about scientific truths and not methodologies, but the original incarnation sounds too outmodedly positivist for the contemporary Foucaultian spirit.

²Galeson's (2007) insights about the two types of artistic creativity and their life cycles are bound to apply to researchers as well.

Obviously, I did not mention rigidities as a random example of contraptions that we include in our models, but to direct our attention to how surprisingly popular such additions to the main model turned out to be. Most macroeconomists, myself included, have always had a soft spot for nominal or real rigidities. A cynic will claim it is just because they are most convenient. After all, they dispense with the necessity for reflection, since there is hardly any observation of the aggregate behavior of the economy cannot be blamed on one rigidity or another.³

But just because a theory is inordinately serviceable or warrants the more serious accusation that it encourages mental laziness is certainly not proof that the theory is not true. At least since David Hume, economists have believed that they have identified a monetary transmission mechanism from increases in money to short-run fluctuations caused by some form or another of price stickiness. It takes much courage, and more aplomb, to dismiss two and a half centuries of a tradition linking Hume to Woodford and going through Marshall, Keynes, and Friedman. Even those with less of a Burkean mind than mine should feel reluctant to proceed in such a perilous manner. Moreover, after one finishes reading Friedman and Schwartz's (1971) *A Monetary History of the U.S.* or slogging through the mountain of Vector Autoregressions (VARs) estimated over 25 years, it must be admitted that those who see money as an important factor in business cycles fluctuations have an impressive empirical case to rely on. Here is not the place to evaluate all these claims (although in the interest of disclosure, I must admit that I am myself less than totally convinced of the importance of money outside the case of large inflations). Suffice it to say that the previous arguments of intellectual tradition and data were a motivation compelling enough for the large number of economists who jumped into the possibility of combining the beauty of DSGE models with the importance of money documented by empirical studies.

Researchers quickly found that we basically require three elements for that purpose. First, we need monopolistic competition. Without market power, any firm that does not immediately adjust its prices will lose all its sales. While monopolistic competition can be incorporated in different ways, the favorite route is to embody the Dixit-Stiglitz framework into a general equilibrium environment, as so beautifully done by Blanchard and Kiyotaki (1987). While not totally satisfactory (for example, the basic Dixit-Stiglitz setup implies counterfactually constant mark-ups), the framework has proven to be easy to handle and surprisingly

³A more sophisticated critic will even point out that the presence of rigidities at the micro level may wash out at an aggregate level, as in the wonderful example of Caplin and Spulber (1987).

flexible. Second, we need some role to justify the existence of money. Money in the utility function or a cash-in-advance constraint can accomplish that goal in a not particularly elegant but rather effective way.⁴ Third, we need a monetary authority inducing nominal shocks to the economy. A monetary policy rule, such as a money growth process or a Taylor rule, usually nicely stands in for such authority. There were, in addition, two extra elements that improve the fit of the model. First, to delay and extend the response of the economy to shocks, macroeconomists postulated factors such as habit persistence in consumption, adjustment cost of investment, or a changing utilization rate of capital. Finally, many extra shocks were added: to investment, to preferences, to monetary and fiscal policy, etc.⁵

The stochastic neoclassical growth model of Kydland and Prescott showed a remarkable ability to absorb all these mechanisms. After a transitional period of amalgamation during the 1990s, by 2003, the model augmented with nominal and real rigidities was sufficiently mature as to be put in a textbook by Mike Woodford and to become the basis for applied work. For the first time, DSGE models were sufficiently flexible to fit the data sufficiently well as to be competitive with VARs in terms of forecasting power (see Edge, Kiley, and Laforge, 2008, for the enchantingly good forecast record of a state-of-the-art DSGE model) and rich enough to become laboratories where realistic economic policies could be evaluated. The rest of the history is simple: DSGE models quickly became the standard tool for quantitative analysis of policies and every self-respecting central bank felt that it needed to estimate its own DSGE model.⁶ However, as surprising as the quick acceptance of DSGE models outside academic circles was, even more unexpected was the fact that models were not only formally estimated, leaving behind the rather unsatisfactory calibration approach, but they were estimated from a Bayesian perspective.

⁴Wallace (2001) has listed many reasons to suspect that these mechanisms may miss important channels through which money matters. After all, they are reduced forms of an underlying model and, as such, they may not be invariant to policy changes. Unfortunately, the profession has not developed a well-founded model of money that can be taken to the data and applied to policy analysis. Despite some recent promising progress (Lagos and Wright, 2005), money in the utility function or cash-in-advance will be with us for many years to come.

⁵Also, researchers learned that it was easy to incorporate home production (Benhabib *et al.*, 1991), an open-economy sector (Mendoza, 1991 and 1995, Backus, Kehoe, and Kydland, 1992 and 1995, and Correia, Neves, and Rebelo, 1995) or a financial sector (Bernanke, Gertler, and Gilchrist, 1999) among other extensions that I cannot discuss here.

⁶Examples include the Federal Reserve Board (Erceg, Guerrieri, and Gust, 2006), the European Central Bank (Christoffel, Coenen, and Warne, 2007), the Bank of Canada (Murchison and Rennison, 2006), the Bank of England (Harrison *et al.*, 2005), the Bank of Sweden (Adolfson *et al.*, 2005), the Bank of Finland (Kilponen and Ripatti, 2006 and Kortelainen, 2002), and the Bank of Spain (Andrés, Burriel, and Estrada, 2006).

3. The Bayesian Approach

I took my first course in Bayesian econometrics from John Geweke at the University of Minnesota in the fall of 1996. I remember how, during one of the lectures in that course, Geweke forecasted that in a few years, we would see a considerable proportion of papers in applied macro being written from a Bayesian perspective. I was rather skeptical about the prediction and dismissed Geweke's claim as an overly optimistic assessment by a committed Bayesian. Fortunately, Geweke was right and I was wrong. The last decade has indeed experienced an explosion of research using Bayesian methods; so much so that, during a recent talk, when I was presenting an estimation that for several reasons I had done using maximum likelihood, I was assailed by repeated instances of the question: why didn't you use Bayes?, a predicament rather unimaginable even a decade ago.

How did such a remarkable change come about? It would be tempting to re-enumerate, as has been done innumerable times before, the long list of theoretical advantages of Bayesian statistics and state that it was only a matter of time before economists would accept the obvious superiority of the Bayes choice. In fact, I will momentarily punish the reader with yet one more review of some of those advantages, just to be sure that we are all on board. But the simpler truth is that, suddenly, doing Bayesian econometrics was easier than doing maximum likelihood.⁷

The reason is that maximizing a complicated, highly dimensional function like the likelihood of a DSGE model is actually much harder than it is to integrate it, which is what we do in a Bayesian exercise. First, the likelihood of DSGE models is, as I have just mentioned, a highly dimensional object, with a dozen or so parameters in the simplest cases to close to a hundred in some of the richest models in the literature. Any search in a high dimensional function is fraught with peril. More pointedly, likelihoods of DSGE models are full of local maxima and minima and of nearly flat surfaces. This is due both to the sparsity of the data (quarterly data do not give us the luxury of many observations that micro panels provide) and to the flexibility of DSGE models in generating similar behavior with relatively different combination of parameter values (every time you see a sensitivity analysis claiming that the results of the paper are robust to changes in parameter values, think about flat likelihoods).

⁷This revival of Bayesian tools is by no means limited to econometrics. Bayesian methods have become extremely popular in many fields, such as genetics, cognitive science, weather analysis, and computer science. The forthcoming *Handbook of Applied Bayesian Analysis* edited by O'Hagan and West is a good survey of Bayesian statistics across many different disciplines.

Consequently, even sophisticated maximization algorithms like simulated annealing or the simplex method run into serious difficulties when maximizing the likelihoods of dynamic models. Moreover, the standard errors of the estimates are notoriously difficult to compute and their asymptotic distribution a poor approximation to the small sample one.

In comparison, Markov chain Monte Carlo (McMc) methods have a much easier time exploring the likelihood (more precisely, the likelihood times the prior) of DSGE models and offer a thorough view of our object of interest. That is why we may want to use McMc methods even when dealing with classical problems. Chernozhukov and Hong (2003) is a path-breaking paper that brought that possibility to the attention of the profession. Even more relevantly, McMc can be transported from application to application with a relatively small degree of fine tuning, an attractive property since the comparative advantage of most economists is not in numerical analysis (and, one suspects, neither their absolute advantage).

I promised before, though, that before entering into a more detailed description of techniques like McMc, I would inflict upon the reader yet another enumeration of the advantages of Bayesian thinking. But fortunately, this will be, given the circumstances of this paper, a rather short introduction. A whole textbook treatment of Bayesian statistics can be found in several excellent books in the market, among which I will recommend Robert (2001) and Bernardo and Smith (2000).

I start with a point that Chris Sims repeatedly makes in his talks: Bayesian inference is a way of thinking, not a “basket” of methods. Classical statistics searches for procedures that work well *ex ante*, i.e., procedures that applied in a repeated number of samples will deliver the right answer in a prespecified percentage of cases. This prescription is not, however, a constructive recipe. It tells us a property of the procedure we want to build and not how to do it. Consequently, we can come up with a large list of procedures that achieve the same objective without a clear metric to pick among them. The best possible illustration is the large number of tests that can be defined to evaluate the null hypothesis of cointegration of two random variables, each with its strengths and weaknesses. Furthermore, the procedures may be quite different in their philosophy and interpretation. In comparison, Bayesian inference is summarized in one simple idea: the Bayes’ theorem. Instead of spending our time proving yet one more asymptotic distribution of a novel estimator, we can go directly to the data, apply Bayes’ theorem, and learn from it. As simple as that.

Let me outline the elements that appear in the theorem. First, we have some data $y^T \equiv \{y_t\}_{t=1}^T \in \mathbb{R}^{N \times T}$. For simplicity, I will use an index t that is more natural in a time

series context like the one I will use below, but minimum work would adapt the notation to cross-sections or panels. From the Bayesian perspective, data are always given and, in most contexts, it does not make much sense to think about it as the realization of some data-generating process (except, perhaps when exploring some asymptotic properties of Bayesian methods as in Phillips and Ploberger, 1996, and Fernández-Villaverde and Rubio-Ramírez, 2004). Second, we have a model, motivated either by economic theory or some other type of reasoning. The model is indexed by i and it may be an instance of a set of possible models to consider M , i.e., we have $i \in M$. The model is composed by:

1. A parameter set, $\Theta_i \in \mathbb{R}^{k_i}$, that defines the admissible value of the parameters that index the functions in the model. Some restrictions come from statistics. For instance, variances must be positive. Others come from economic reasoning. For example, it is common to bound the discount factor in an intertemporal choice problem to ensure that total utility is well defined.
2. A likelihood function $p(y^T|\theta, i) : \mathbb{R}^{N \times T} \times \Theta_i \rightarrow \mathbb{R}^+$ that tells us the probability that the model assigns to each observation given some parameter values. This likelihood function is nothing more than the restrictions that our model imposes on the data, either coming from statistical considerations or from equilibrium conditions.
3. A prior distribution $\pi(\theta|i) : \Theta_i \rightarrow \mathbb{R}^+$ that captures pre-sample beliefs about the right value of the parameters (yes, “right” is an awfully ambiguous word; I will come back later to what I mean by it).

Bayes’ theorem tells us that the posterior distribution of the parameters is given by:

$$\pi(\theta|y^T, i) = \frac{p(y^T|\theta, i)\pi(\theta|i)}{\int p(y^T|\theta, i)\pi(\theta|i) d\theta}$$

This result, which follows from a basic application of the laws of probability, tells us how we should update our beliefs about parameter values: we combine our prior beliefs, $\pi(\theta|i)$, with the sample information embodied in the likelihood, $f(y^T|\theta, i)$, and we obtain a new set of beliefs, $\pi(\theta|y^T, i)$. In fact, Bayes’ theorem is an optimal information processing rule as defined by Zellner (1988): it uses efficiently all of the available information in the data, both in small and large samples, without adding any extraneous information.

Armed with Bayes’ theorem, a researcher does not need many more tools. For any possible model, one just writes down the likelihood, elicits the prior, and obtains the posterior. Once

we have the posterior distribution of the parameters, we can perform inference like point estimation or model comparison given a loss function that maps how much we select an incorrect parameter value or model. For sure, these tasks can be onerous in terms of implementation but, conceptually, they are straightforward. Consequently, issues such as nonstationarity do not require specific methods as needed in classical inference (see the eye-opening helicopter tour of Sims and Uhlig, 1991). If we suspect non-stationarities, we may want to change our priors to reflect that belief, but the likelihood function will still be the same and Bayes' theorem is applicable without the disconcerting discontinuities of classical procedures around the unit root.

But while coherence is certainly an attractive property, at least from an esthetic consideration, it is not enough by itself. A much more relevant point is that coherence is a consequence of the fact that Bayes' theorem can be derived from a set of axioms that decision theorists have proposed to characterize rational behavior. It is not an accident that the main solution concepts in games with incomplete information are Bayesian Nash equilibria and sequential equilibria and that Bayes' theorem plays a critical role in the construction of these solution concepts. It is ironic that we constantly see papers where the researcher specifies that the rational agents in the model follow Bayes' theorem and, then, she proceeds to estimate the model using classical procedures, undaunted by the implied logical contradiction.

Closely related to this point is the fact that the Bayesian approach satisfies by construction the Likelihood Principle (Berger and Wolpert, 1988) that states that all of the information existing in a sample is contained in the likelihood function. Once one learns about how Birnbaum (1962) derived the Likelihood Principle from more fundamental axioms, it is rather difficult not to accept it.

The advantages of Bayesian inference do not end here. First, Bayesian econometrics offers a set of answers that are relevant for users. In comparison, pre-sample probability statements are, on most occasions, rather uninteresting from a practical perspective. Few policy makers will be very excited if we inform them that in 95 of 100 possible samples, our model measures that a certain policy increases welfare but that we cannot really know if the actual data represents one of the 95 positive cases or one of the negative 5. They want to know, conditional on what we have observed in the data, what is the probability that we would be doing the right thing by, for instance, lowering the interest rate. A compelling proof of how unnatural it is to think in frequentist terms is to teach introductory statistics. Nearly all students will interpret confidence intervals at first as a probability interval. Only the

repeated insistence of the instructor will make a disappointingly small minority of students understand the difference between the two and provide the right interpretation. The rest of the students, of course, would simply memorize the answer for the test in the same way they would memorize a sentence in Aramaic if such a worthless accomplishment were useful to get a passing grade. Neither policy makers nor undergraduate students are silly (they are ignorant, but that is a very different sin); they just think in ways that are more natural to humans.⁸ Frequentist statements are beautiful but inconsequential.

Second, pre-sample information is often amazingly rich and considerably useful and not taking advantage of it is an unforgivable omission. For instance, microeconomic evidence can guide our building of priors. If we have a substantial set of studies that estimate the discount factor of individuals and they find a range of values between 0.9 and 0.99, any sensible prior should take this information into consideration.

The researcher should be careful, though, translating this micro evidence into macro priors. Parameter values do not have an existence of their own, like a Platonic entity waiting to be discovered. They are only defined within the context of a model, and changes in the theory, even if minor, may have a considerable impact on the parameter values. A paradigmatic example is labor supply. For a long time, labor economists criticized the real business cycle models because they relied on what they saw as an unreasonably high labor supply elasticity (Alesina, Glaeser, and Sacerdote, 2006, is a recent instance of such criticism). However, the evidence that fed their attacks was gathered mainly for prime age white males in the United States (or a similarly restrictive group). But representative agent models are not about prime age white males: the representative agent is instead a stand-in for everyone in the economy. It has a bit of a prime age male and a bit of old woman, a bit of a minority young and a bit of a part-timer. If much of the response of labor to changes in wages is done through the labor supply of women and young workers, it is perfectly possible to have a high aggregate elasticity of labor supply and a low labor supply elasticity of prime age males. To illustrate this point, Rogerson and Wallenius (2007) construct an overlapping generations economy where the micro and macro elasticities are virtually unrelated. But we should not push the previous example to an exaggerated degree: it is a word of caution, not a licence to concoct wild priors. If the researcher wants to depart in her prior from the micro estimates, she must have at least some plausible explanation of why she is doing so (see Browning,

⁸See also the psychological evidence that humans' cognitive processes are well described by Bayes' theorem presented by Griffiths and Tenenbaum (2006).

Hansen, and Heckman (1999) for a thorough discussion of the mapping between micro and macro estimates).

An alternative source of pre-sample information is the estimates of macro parameters from different countries. One of the main differences between economists and other social scientists is that we have a default belief that individuals are basically the same across countries and that differences in behavior can be accounted for by differences in relative prices. Therefore, if we have estimates from Germany that the discount factor in a DSGE model is around 0.98, it is perfectly reasonable to believe that the discount factor in Spain, if we estimate the same model, should be around 0.98. Admittedly, differences in demographics or financial markets may show up as slightly different discount factors, but again, the German experience is most informative. Pre-sample information is particularly convenient when we deal with emerging economies, when the data is extremely limited, or when we face a change in policy regime. A favorite example of mine concerns the creation of the European Central Bank (ECB). If we were back in 1998 or 1999 trying to estimate a model of how the ECB works, we would face such a severe limitation in the length of the data that any classical method would fail miserably. However, we could have used a Bayesian method where the prior would have been that the ECB would behave in a way very similar than the German Bundesbank. Yes, our inference would have depended heavily on the prior, but why is this situation any worse than not being able to say anything of consequence? Real life is full of situations where data are extremely sparse (or where they speak to us very softly about the difference between two models, like a unit root process and an AR(1) with coefficient 0.99) and we need to make the best of a bad situation by carefully eliciting priors.⁹

Third, Bayesian econometrics allows a direct computation of many objects of interest, such as the posterior distribution of welfare gains, values at risk, fan charts, and many other complicated functions of the underlying parameters while capturing in these computed objects all the existing uncertainty regarding parameter values. For example, instead of computing the multiplier of an increase in public consumption (per se, not a very useful number for a politician), we can find the whole posterior distribution of employment changes in the next year conditioning on what we know about the evolution of the economy plus

⁹We can push the arguments to the limit. Strictly speaking we can perform Bayesian inference without any data: our posterior is just equal to the prior! We often face this situation. Imagine that we were back in 1917 and we just heard about the Russian revolution. Since communism had never been tried, as economists we would need to endorse or reject the new economic system exclusively based on our priors about how well central planning could work. Waiting 70 years to see how well the whole experiment would work is not a reasonable course of action.

the effect of an increase in public consumption. Such an object, with its whole assessment of risks, is a much more relevant tool for policy analysis. Classical procedures have a much more difficult time jumping from point estimates to whole distributions of policy-relevant objects.

Finally, Bayesian econometrics deals in a natural way with misspecified models (Monfort, 1996). As the old saying goes, all models are false, but some are useful. Bayesians are not in the business of searching for the truth but only in coming up with good description of the data. Hence, estimation moves away from being a process of discovery of some “true” value of a parameter to being, in Rissanen’s (1986) powerful words, a selection device in the parameter space that maximizes our ability to use the model as a language in which to express the regular features of the data. Coming back to our previous discussion about “right” parameters, Rissanen is telling us to pick those parameter values that allow us to tell powerful economic histories and to exert control over outcomes of interest. These parameter values, which I will call “pseudotrue,” may be, for example, the ones that minimize the Kullback-Leibler distance between the data generating process and the model (Fernández-Villaverde and Rubio-Ramírez, 2004, offer a detailed explanation of why we care about these “pseudotrue” parameter values).

Also, by thinking about models and parameters in this way, we come to the discussion of partially identified models initiated by Manski (1999) from a different perspective. Bayesians emphasize more the “normality” of a lack of identification than the problems caused by it. Bayesians can still perform all of their work without further complications or the need of new theorems even with a flat posterior (and we can always achieve identification through non-flat priors, although such an accomplishment is slightly boring). For example, I can still perfectly evaluate the welfare consequences of one action if the posterior of my parameter values is flat in some or all of the parameter space. The answer I get may have a large degree of uncertainty, but there is nothing conceptually different about the inference process. This does not imply, of course, that identification is not a concern.¹⁰ I only mean that identification is a somehow different preoccupation for a Bayesian.

I would not be fully honest, however, if I did not discuss, if only briefly, the disadvantages of Bayesian inference. The main one, in my opinion, is that many non-parametric and semiparametric approaches sound more natural when set up in a classical framework. Think about the case of the Generalized Method of Moments (GMM). The first time you hear about

¹⁰Identification issues ought to be discussed in more detail in DSGE models, since they affect the conclusions we get from them. See Canova and Sala (2006) for examples of non-identified DSGE models and further discussion.

it in class, your brain (or at least mine!) goes “ah!, this makes perfect sense.” And it does so because GMM (and all its related cousins in the literature of empirical likelihood, Owen, 2001, and, in economics, Kitamura and Stutzer, 1997) are clear and intuitive procedures that have a transparent and direct link with first order conditions and equilibrium equations. Also, methods of moments are a good way to estimate models with multiple equilibria, since all of those equilibria need to satisfy certain first order conditions that we can exploit to come up with a set of moments.¹¹ Even if you can cook up many things in a Bayesian framework that look a lot like GMM or empirical likelihood (see, for example, Kim, 1998, Schennach, 2005, or Ragusa, 2006, among several others), I have never been particularly satisfied with any of them and none has passed the “ah!” test that GMM overcomes with such an excellent grade.

Similarly, you can implement a non-parametric Bayesian analysis (see the textbook by Ghosh and Ramamoorthi, 2003, and in economics, Chamberlain and Imbens, 2003). However, the methods are not as well developed as we would like and the shining building of Bayesian statistics gets dirty with some awful discoveries such as the potentially bad asymptotic properties of Bayesian estimators (first pointed out by Freedman, 1963) or the breakdown of the likelihood principle (Robins and Ritov, 1997). Given that the literature is rapidly evolving, Bayesian methods may end up catching up and even overcoming classical procedures for non-parametric and semiparametric problems, but this has not happened yet. In the meantime, the advantage in this sub-area seems to be in the frequentist camp.

4. The Tools

No matter how sound were the DSGE models presented by the literature or how compelling the arguments for Bayesian inference, the whole research program would not have taken off without the appearance of the right set of tools that made the practical implementation of the estimation of DSGE models feasible in a standard desktop computer. Otherwise, we would probably still be calibrating our models, which would be, in addition, much smaller and simpler. I will classify those tools in three sets. First, better and improved solution methods. Second, methods to evaluate the likelihood of the model. Third, methods to explore the likelihood of the model.

¹¹A simple way to generate multiplicity of equilibria in a DSGE model that can be very relevant empirically is to have increasing returns to scale, as in Benhabib and Farmer (1992). For a macro perspective on estimation of models with multiplicity of equilibria, see Jovanovic (1989) or Cooper (2002).

4.1. Solution Methods

DSGE models do not have, except for a very few exceptions, a “paper and pencil” solution. Hence, we are forced to resort to numerical approximations to characterize the equilibrium dynamics of the model. Numerical analysis is not part of the standard curriculum either at the undergraduate or the graduate level. Consequently, the profession had a tough time accepting that analytic results are limited (despite the fact that the limitations of close form findings happens in most other sciences where the transition to numerical approximations happened more thoroughly and with less soul searching). To make things worse, few economists were confident in dealing with the solution of stochastic difference functional equations, which are the core of the solution of a DSGE model. The first approaches were based on fitting the models to be solved into the framework of what was described in standard optimal control literature textbooks. For example, Kydland and Prescott (1982) substituted the original problem by a linear quadratic approximation to it. King, Plosser, and Rebelo (in the widely disseminated technical appendix, not published until 2002) linearized the equilibrium conditions, and Christiano (1990) applied value function iteration. Even if those approaches are still the cornerstone of much of what is done nowadays, as time passed, researchers became familiar with them, many improvements were proposed, and software circulated.

Let me use the example of linearization, since it is the solution method that I will use below.¹² Judd and Guu (1993) showed that linearization was not an ad hoc procedure but the first order term of a mainstream tool in scientific computation, perturbation. The idea of perturbation methods is to substitute the original problem, which is difficult to solve, for a simpler one that we know how to handle and use the solution of the simpler model to approximate the solution of the problem we are interested in. In the case of DSGE models, we find an approximated solution by finding a Taylor expansion of the policy function describing the dynamics of the variables of the model around the deterministic steady state. Linearization, therefore, is just the first term of this Taylor expansion. But once we understand this, it is straightforward to get higher order expansions that are both analytically informative and more accurate (as in Schmitt-Grohé and Uribe, 2004).¹³ Similarly, we can apply all of

¹²Other solution methods for DSGE models, such as projection algorithms and value function iteration, are described and compared in Aruoba, Fernández-Villaverde, and Rubio-Ramírez (2006). Judd (1998) is a comprehensive textbook.

¹³For example, a second order expansion includes a term that corrects for the standard deviation of the shocks that drive the dynamics of the economy. This term, which captures precautionary behavior, breaks the certainty equivalence of linear approximations that makes it difficult to talk about welfare and risk.

the accumulated knowledge of perturbation methods in terms of theorems or in improving the performance of the method.¹⁴ Second, once economists became more experienced with linearization, software disseminated very quickly.

My favorite example is Dynare and Dynare++, an extraordinary tool developed by Michel Juillard and a team of collaborators. Dynare (a toolbox for Matlab and Scilab) and Dynare++ (a stand-alone application) allow the researcher to write, in a concise and intuitive language, the equilibrium conditions of a DSGE model and find a perturbation solution to it, up to second order in Dynare and to an arbitrary order in Dynare++. With Dynare and Dynare++, a moderately experienced user can write code for a basic real business cycle model in an hour and compute the approximated solution in a few seconds. The computation of the model presented below (a fairly sophisticated one) requires a bit more effort, but still coding can be done in a short period of time (as short as a day or two for an experienced user) and the solution and simulation take only a few seconds. This advance in the ease of computation is nothing short of breathtaking.

4.2. Evaluating the Likelihood Function

In our previous description of Bayes' theorem, the likelihood function of the model played a key role, since it was the object that we multiplied by our prior to obtain a posterior. The challenge is how to obtain the likelihood of a DSGE model for which we do not even have an analytic solution. The most general and powerful route is to employ the tools of state space representations and filtering theory.

Once we have the solution of the DSGE model in terms of its (approximated) policy functions, we can write the laws of motion of the variables in a state space representation that consists of:

1. A transition equation, $S_t = f(S_{t-1}, W_t; \theta)$, where S_t is the vector of states that describe

¹⁴Here I can cite the idea of changing variables (Fernández-Villaverde and Rubio-Ramírez, 2006). Instead of writing a Taylor expansion in terms of a variable x :

$$f(x) \simeq f(a) + f'(a)(x - a) + H.O.T.$$

we can write it in terms of a transformed variable $Y(x)$:

$$g(y) = h(f(X(y))) = g(b) + g'(b)(Y(x) - b) + H.O.T.$$

where $b = Y(a)$ and $X(y)$ is the inverse of $Y(x)$. By picking the right change of variables, we can significantly increase the accuracy of the perturbation. A common example of change of variables (although rarely thought of in this way) is to loglinearize instead of linearizing in levels.

the situation of the model in any given moment in time, W_t is a vector of innovations, and θ is a vector with the structural parameters that describe technology, preferences, and information processes.

2. A measurement equation, $Y_t = g(S_t, V_t; \theta)$, where Y_t are the observables and V_t a set of shocks to the observables (like, but necessarily, measurement errors).

While the transition equation is unique up to an equivalent class, the measurement equation depends on what we assume we can observe, selection that may imply many degrees of freedom (and not trivial consequences for inference; see the experiments in Guerrón-Quintana, 2008).¹⁵

The state space representation lends itself to many convenient computations. To begin with, from $S_t = f(S_{t-1}, W_t; \theta)$, we can compute $p(S_t|S_{t-1}; \theta)$, from $Y_t = g(S_t, V_t; \theta)$, we can compute $p(Y_t|S_t; \theta)$, and from $S_t = f(S_{t-1}, W_t; \theta)$ and $Y_t = g(S_t, V_t; \theta)$, we have:

$$Y_t = g(f(S_{t-1}, W_t; \theta), V_t; \theta)$$

and hence we can compute $p(Y_t|S_{t-1}; \theta)$ (here I am omitting the technical details regarding the existence of these objects).

All of these conditional densities appear in the likelihood function in a slightly disguised way. If we want to evaluate the likelihood function of the observables y^T at parameter values θ , $p(y^T; \theta)$, we can start by taking advantage of the Markov structure of our state space representation to write:

$$\begin{aligned} p(y^T|\theta) &= p(y_1|\theta) \prod_{t=2}^T p(y_t|y^{t-1}; \theta) \\ &= \int p(y_1|s_1; \theta) ds_1 \prod_{t=2}^T \int p(y_t|S_t; \theta) p(S_t|y^{t-1}; \theta) dS_t \end{aligned}$$

Hence, knowledge of $\{p(S_t|y^{t-1}; \theta)\}_{t=1}^T$ and $p(S_1; \theta)$ allow the evaluation of the likelihood of the model.

Filtering theory is the branch of mathematics that is preoccupied precisely with finding the sequence of conditional distributions of states given observations, $\{p(S_t|y^{t-1}; \theta)\}_{t=1}^T$. For

¹⁵Also, I am assuming that there exists a state space representation that is Markov in some vector of states. By admitting no-payoff relevant states, like Lagrangian multipliers that encode continuation utilities (Abreu, Pearce, and Stachetti, 1990), we can fit a large class of economic models into this setup.

this task, it relies on two fundamental tools, the Chapman-Kolmogorov equation:

$$p(S_{t+1}|y^t; \theta) = \int p(S_{t+1}|S_t; \theta) p(S_t|y^t; \theta) dS_t$$

and Bayes' theorem (yes, again) :

$$p(S_t|y^t; \theta) = \frac{p(y_t|S_t; \theta) p(S_t|y^{t-1}; \theta)}{p(y_t|y^{t-1}; \theta)}$$

where

$$p(y_t|y^{t-1}; \theta) = \int p(y_t|S_t; \theta) p(S_t|y^{t-1}; \theta) dS_t.$$

is the conditional likelihood.

The Chapman-Kolmogorov equation, despite its intimidating name, tells us only that the distribution of states tomorrow given an observation until today, $p(S_{t+1}|y^t; \theta)$, is equal to the distribution today of $p(S_t|y^t; \theta)$ times the transition probabilities $p(S_{t+1}|S_t; \theta)$ integrated over all possible states. Therefore, the Chapman-Kolmogorov equation just gives us a forecasting rule for the evolution of states. Bayes' theorem updates the distribution of states $p(S_t|y^{t-1}; \theta)$ when a new observation arrives given its probability $p(y_t|S_t; \theta)$. By a recursive application of forecasting and updating, we can generate the complete sequence $\{p(S_t|y^{t-1}; \theta)\}_{t=1}^T$ we are looking for.

While the Chapman-Kolmogorov equation and Bayes' theorem are mathematically rather straightforward objects, their practical implementation is cumbersome because they involve the computation of numerous integrals. Even when the number of states is moderate, the computational cost of these integrals makes an exact (or up to floating point accuracy) evaluation of the integrals unfeasible.

4.2.1. The Kalman Filter

To fix this computational problem, we have two routes. First, if the transition and measurement equation are linear and the shocks are normally distributed, we can take advantage of the observation that all of the relevant conditional distributions are Gaussian (this just from the simple fact that the space of normal distributions is a vector space). Therefore, we only need to keep track of the mean and variance of these conditional normals. The tracking of the moments is done through the Riccati equations of the Kalman filter (for more details, see any standard textbook, such as Harvey, 1989, or Stengel, 1994).

To do so, we start by writing the first order linear approximation to the solution of the model in the state space representation we introduced above:

$$s_t = As_{t-1} + B\varepsilon_t \quad (1)$$

$$y_t = Cs_t + D\varepsilon_t \quad (2)$$

$$\varepsilon_t \sim \mathbf{N}(0, I)$$

where we use lower case letters to denote realizations of the random variable and where ε_t is the vector of innovations to the model that stacks W_t and V_t .

Let us define the linear projections $s_{t|t-1} = \mathbb{E}(s_t|Y_{t-1})$ and $s_{t|t} = \mathbb{E}(s_t|Y_t)$ where $Y_t = \{y_1, y_2, \dots, y_t\}$ and the subindex tracks the conditioning set (i.e., $t|t-1$ means a draw at moment t conditional on information until $t-1$). Also, we have matrices of variances-covariances $P_{t-1|t-1} = \mathbb{E}(s_{t-1} - s_{t-1|t-1})(s_{t-1} - s_{t-1|t-1})'$ and $P_{t|t-1} = \mathbb{E}(s_{t-1} - s_{t|t-1})(s_{t-1} - s_{t|t-1})'$. Given these linear projections and the Gaussian structure of our state space representations, the one-step-ahead forecast error, $\eta_t = y_t - Cs_{t|t-1}$, is white noise.

We forecast the evolution of states:

$$s_{t|t-1} = As_{t-1|t-1} \quad (3)$$

Since the possible presence of correlation in the innovations does not change the nature of the filter (Stengel, 1994), so it is still the case that

$$s_{t|t} = s_{t|t-1} + K\eta_t, \quad (4)$$

where K is the Kalman gain at time t . Define variance of forecast as $V_y = CP_{t|t-1}C' + DD'$.

Since η_t is white noise, the conditional loglikelihood of the period observation y_t is just:

$$\log p(y_t|\theta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log \det(V_y) - \frac{1}{2} \eta_t V_y^{-1} \eta_t$$

The last step is to update our estimates of the states. Define residuals $\xi_{t|t-1} = s_t - s_{t|t-1}$ and $\xi_{t|t} = s_t - s_{t|t}$. Subtracting equation (3) from equation (1)

$$\begin{aligned} s_t - s_{t|t-1} &= A(s_{t-1} - s_{t-1|t-1}) + Bw_t, \\ \xi_{t|t-1} &= A\xi_{t-1|t-1} + Bw_t \end{aligned} \quad (5)$$

Now subtract equation (4) from equation (1)

$$\begin{aligned} s_t - s_{t|t} &= s_t - s_{t|t-1} - K [Cs_t + Dw_t - Cs_{t|t-1}] \\ \xi_{t|t} &= \xi_{t|t-1} - K [C\xi_{t|t-1} + Dw_t]. \end{aligned} \quad (6)$$

Note $P_{t|t-1}$ can be written as

$$\begin{aligned} P_{t|t-1} &= \mathbb{E}\xi_{t|t-1}\xi'_{t|t-1}, \\ &= \mathbb{E}(A\xi_{t-1|t-1} + Bw_t)(A\xi_{t-1|t-1} + Bw_t)' \\ &= AP_{t-1|t-1}A' + BB'. \end{aligned} \quad (7)$$

and for $P_{t|t}$ we have:

$$\begin{aligned} P_{t|t} &= \mathbb{E}\xi_{t|t}\xi'_{t|t} \\ &= \mathbb{E}(\xi_{t|t-1} - K [C\xi_{t|t-1} + Dw_t])(\xi_{t|t-1} - K [C\xi_{t|t-1} + Dw_t])' \\ &= (I - KC)P_{t|t-1}(I - C'K') + KDD'K' - KDB' \\ &\quad - BD'K' + KCBD'K' + KDB'C'K'. \end{aligned} \quad (8)$$

The optimal gain K minimizes $P_{t|t}$ with the first order condition

$$\frac{\partial Tr(P_{t|t})}{\partial K} = 0$$

and solution

$$K = [P_{t|t-1}C' + BD'] [V_y + CBD' + DB'C']^{-1}$$

Consequently, the updating equations are:

$$\begin{aligned} P_{t|t} &= P_{t|t-1} - K_{opt} [DB' + CP_{t|t-1}], \\ x_{t|t} &= x_{t|t-1} + K_{opt}\eta_t \end{aligned}$$

and we close the iterations. We only need to apply the equations from $t = 1$ until T and we can compute the loglikelihood function. The whole process takes only a fraction of a second on a modern laptop computer.

4.2.2. The Particle Filter

Unfortunately, linearity and Gaussanity are quite restrictive assumptions. For example, linearization eliminates asymmetries, threshold effects, precautionary behavior, big shocks, and many other phenomena of interest in macroeconomics. Moreover, linearization induces an approximation error. Even if we were able to evaluate the likelihood implied by that solution, we would not be evaluating the likelihood of the exact solution of the model but the likelihood implied by the approximated linear solution of the model. Both objects may be quite different and some care is required when we proceed to perform inference (for further details, see Fernández-Villaverde, Rubio-Ramírez, and Santos, 2006). The effects of this are worse than you may think. First, there are theoretical arguments. Second order errors in the approximated policy function may imply first order errors in the loglikelihood function. As the sample size grows, the error in the loglikelihood function also grows and we may have inconsistent point estimates. Second, linearization complicates the identification of parameters (or makes it plainly impossible as, for example, the coefficient of risk aversion in a model with Epstein-Zin preferences as introduced by Epstein and Zin, 1989 and 1991). Finally, computational evidence suggests that those effects may be important in many applications.

Similarly Gaussanity eliminates the possibility of talking about time-varying volatility in time series, which is a fundamental issue in macroeconomics. For instance, McConnell and Pérez-Quirós (2000), Kim and Nelson (1998), Fernández-Villaverde and Rubio-Ramírez (2007), and Justiniano and Primiceri (2008) have accumulated rather compelling evidence of the importance of time-varying volatility to account for the dynamics of U.S. data. Any linear Gaussian model cannot talk about this evidence at all. Similarly, linear models cannot deal with models that display regime switching, an important feature of much recent research (see Sims and Zha, 2006, and Farmer, Waggoner, and Zha, 2006a and b).

When the state space representation is not linear or when the shocks are not normal, filtering becomes more complicated because the conditional distributions of states do not belong, in general, to any known family. How do we keep track of them? We mentioned before that analytic methods are unfeasible except in a few cases. Therefore, we need to resort to some type of simulation. An algorithm that has been used recently with much success is the particle filter, a particular example of a Sequential Monte Carlo (see the technical appendix to Fernández-Villaverde and Rubio-Ramírez (2007) for alternative approaches).

Because of space constraints, I will not discuss the filter in much detail (Fernández-Villaverde and Rubio-Ramírez, 2005 and 2007, provide all the technical background; see

Arulampalam *et al.*, 2002 for a general introduction, and Doucet, de Freitas, and Gordon, 2001, for a collection of applications). The main idea, however, is extremely simple: we replace the conditional distribution $\{p(S_t|y^{t-1}; \theta)\}_{t=1}^T$ by an empirical distribution of N draws $\left\{ \left\{ s_{t|t-1}^i \right\}_{i=1}^N \right\}_{t=1}^T$ from the sequence $\{p(S_t|y^{t-1}; \theta)\}_{t=1}^T$ generated by simulation. Then, by a trivial application of the law of large numbers:

$$p(y^T|\theta) \simeq \frac{1}{N} \sum_{i=1}^N p(y_1|s_{0|0}^i; \theta) \prod_{t=2}^T \frac{1}{N} \sum_{i=1}^N p(y_t|s_{t|t-1}^i; \theta)$$

The problem is then to draw from $\{p(S_t|y^{t-1}; \theta)\}_{t=1}^T$. But, following Rubin (1988), we can apply sequential sampling:

Proposition 1. *Let $\left\{ s_{t|t-1}^i \right\}_{i=1}^N$ be a draw from $p(S_t|y^{t-1}; \theta)$. Let the sequence $\{\tilde{s}_t^i\}_{i=1}^N$ be a draw with replacement from $\left\{ s_{t|t-1}^i \right\}_{i=1}^N$ where the resampling probability is given by*

$$\omega_t^i = \frac{p(y_t|s_{t|t-1}^i; \theta)}{\sum_{i=1}^N p(y_t|s_{t|t-1}^i; \theta)},$$

Then $\{\tilde{s}_t^i\}_{i=1}^N$ is a draw from $p(S_t|y^t; \theta)$.

Proposition 1 recursively uses a draw $\left\{ s_{t|t-1}^i \right\}_{i=1}^N$ from $p(S_t|y^{t-1}; \theta)$ to draw $\left\{ s_{t|t}^i \right\}_{i=1}^N$ from $p(S_t|y^t; \theta)$. But this is nothing more than the update of our estimate of S_t to add the information on y_t that Bayes' theorem is asking for.

The reader may be surprised by the need to resample to obtain a new conditional distribution. However, without resampling, all of the sequences would become arbitrarily far away from the true sequence of states and the sequence that is closer to the true states dominates all of the remaining ones in weight. Hence, the simulation degenerates after a few steps and we cannot effectively evaluate the likelihood function, no matter how large N is.

Once we have $\left\{ s_{t|t}^i \right\}_{i=1}^N$, we draw N vectors of exogenous shocks to the model (for example, the productivity or the preference shocks) from their corresponding distributions and apply the law of motion for states to generate $\left\{ s_{t+1|t}^i \right\}_{i=1}^N$. This step, known as forecast, puts us back at the beginning of Proposition 1, but with the difference that we have moved forward one period in our conditioning, from $t|t-1$ to $t+1|t$, implementing in that way the Chapman-Kolmogorov equation.

The following pseudo-code summarizes the description of the algorithm:

Step 0, Initialization: Set $t \rightsquigarrow 1$. Sample N values $\{s_{0|0}^i\}_{i=1}^N$ from $p(S_0; \theta)$.

Step 1, Prediction: Sample N values $\{s_{t|t-1}^i\}_{i=1}^N$ using $\{s_{t-1|t-1}^i\}_{i=1}^N$, the law of motion for states and the distribution of shocks ε_t .

Step 2, Filtering: Assign to each draw $(s_{t|t-1}^i)$ the weight ω_t^i in Proposition 1.

Step 3, Sampling: Sample N times with replacement from $\{s_{t|t-1}^i\}_{i=1}^N$ using the probabilities $\{\omega_t^i\}_{i=1}^N$. Call each draw $(s_{t|t}^i)$. If $t < T$ set $t \rightsquigarrow t + 1$ and go to step 1. Otherwise stop.

With the simulation, we just substitute into our formula

$$p(y^T | \theta) \simeq \frac{1}{N} \sum_{i=1}^N p(y_1 | s_{0|0}^i; \Psi) \prod_{t=2}^T \frac{1}{N} \sum_{i=1}^N p(y_t | s_{t|t-1}^i; \theta) \quad (9)$$

and get an estimate of the likelihood of the model given θ . Del Moral and Jacod (2002) and Künsch (2005) show weak conditions for the consistency of this estimator and for a central limit theorem to apply.

4.3. Exploring the Likelihood Function

Once we have an evaluation of the likelihood function from filtering theory, we need to explore it, either by maximization or by description. As I explained before when I motivated the Bayesian choice, maximization is particularly challenging and the results are often not very robust. Consequently, I will not get into a discussion of how we can attempt to solve this complicated optimization. The Bayesian alternative is, of course, to find the posterior:

$$\pi(\theta | y^T) = \frac{p(y^T | \theta) \pi(\theta)}{\int p(y^T | \theta) \pi(\theta) d\theta}$$

(where I have eliminated the index of the model to ease notation). With the result of the previous subsection, we can evaluate $\pi(\theta | y^T)$ for a given θ (up to a proportionality constant), but characterizing the whole posterior is nearly impossible, since we do not even have a close form solution for $p(y^T | \theta)$.

This challenge, which for a long time was the main barrier to Bayesian inference, can nowadays easily be addressed by the use of MCMC methods. A full exposition of MCMC methods would occupy an entire book (as in Robert and Casella, 2007). Luckily enough, the basic point is rather straightforward. We want to somehow produce a Markov chain whose ergodic distribution is $\pi(\theta|y^T)$. Then, we simulate from the chain and, as the Glivenko-Cantelli theorem does its magic, we approximate $\pi(\theta|y^T)$ by the empirical distribution generated by the chain. This twist of MCMC methods is pure genius. Usually, we have a theory that implies a Markov chain. For example, our DSGE model implies a Markov process for output and we want to characterize it (this is what chapters 11 to 14 of Stokey, Lucas, and Prescott, 1989, do). In MCMC, we proceed backward: we have $\pi(\theta|y^T)$ (or at least a procedure to evaluate it) and we come up with a Markov chain that generates it.

This idea would not be very practical unless we had a constructive method to specify the Markov chain. Fortunately, we have such a procedure, although, interestingly enough, only one. This procedure is known as the Metropolis-Hastings algorithm (the Gibbs sampler is a particular case of Metropolis-Hastings). In the Metropolis-Hastings algorithm, we come up with a new proposed value of the parameter and we evaluate whether it increases the posterior. If it does, we accept it with probability 1. If it does not, we accept it with some probability less than 1. In such a way, we always go toward the higher regions of the posterior but we also travel, with some probability, towards the lower regions. This procedure avoids getting trapped in local maxima. A simple pseudo-code for a plain vanilla Metropolis-Hastings algorithm is as follows:

Step 0, Initialization: Set $i \rightsquigarrow 0$ and an initial θ_i . Solve the model for θ_i and build the state space representation. Evaluate $\pi(\theta_i)$ and $p(y^T|\theta_i)$. Set $i \rightsquigarrow i + 1$.

Step 1, Proposal draw: Get a draw θ_i^* from a proposal density $q(\theta_{i-1}, \theta_i^*)$.

Step 2, Solving the Model: Solve the model for θ_i^* and build the new state space representation.

Step 3, Evaluating the proposal: Evaluate $\pi(\theta_i^*)$ and $p(y^T|\theta_i^*)$ with (9).

Step 4, Accept/Reject: Draw $\chi_i \sim U(0, 1)$. If $\chi_i \leq \frac{p(y^T|\theta_i^*)\pi(\theta_i^*)q(\theta_{i-1}, \theta_i^*)}{p(y^T|\theta_i)\pi(\theta_i)q(\theta_i^*, \theta_{i-1})}$ set $\theta_i = \theta_i^*$, otherwise $\theta_i = \theta_{i-1}$.

Step 5, Iteration: If $i < M$, set $i \rightsquigarrow i + 1$ and go to step 1. Otherwise stop.

This algorithm requires us to specify a proposal density $q(\cdot, \cdot)$. The standard practice (and the easiest) is to choose a random walk proposal, $\theta_i^* = \theta_{i-1} + \kappa_i$, $\kappa_i \sim \mathcal{N}(0, \Sigma_\kappa)$, where Σ_κ is a scaling matrix that the researcher selects to obtain the appropriate acceptance ratio of proposals (Roberts, Gelman and Gilks, 1997, provide the user with guidelines for the optimal acceptance ratios that maximize the rate of convergence of the empirical distribution toward the ergodic distribution). Of course, we can always follow more sophisticated versions of the algorithm, but for most researchers, the time and effort involved in refinements will not compensate for the improvements in efficiency.

If we are using the particle filter, we need to keep the random numbers of the simulation constant across iterations of the Metropolis-Hastings algorithm. As emphasized by McFadden (1989) and Pakes and Pollard (1989), fixing the random numbers across iterations is required to achieve stochastic equicontinuity. Thanks to it, the pointwise convergence of the likelihood (9) to the exact likelihood we stated above becomes uniform convergence. Although not strictly necessary in a Bayesian context, uniform continuity minimizes the numerical instabilities created by the “chatter” of random numbers across iterations.

Once we have run the algorithm for a sufficient number of iterations (see Mengersen, Robert, and Guihenneuc-Jouyaux, 1999, for a review of convergence tests), we can perform inference: we have an empirical approximation of the posterior of the model and finding means, standard deviations, and other objects of interest is a trivial task. In the interest of space, I omit a discussion of how to select a good initial value θ_0 .¹⁶ The values we would have for a standard calibration exercise are, in general, a good default choice.

The reader who is not familiar with the Metropolis-Hastings algorithm may feel that the previous discussion introduced many concepts. Yes, but none of them is particularly deep once one has thought about them a bit more carefully. Most important, once you get the gist of it, McMc methods are surprisingly easy to code, much more, in fact, than even simple optimization algorithms, and they can be easily be recycled for future estimations. This is why I said in section 3 that nowadays doing Bayesian econometrics is easier than doing classical inference.

¹⁶The initial states S_0 for the filter can also be thought of as parameters of the model. However, it is usually is easier to sample from the ergodic distribution of states implied by θ_0 .

5. An Application

The previous pages would look dry and abstract without an application that illustrates how we do things in practice. Hence, I am presenting a simple estimation exercise that I borrow from a recent paper I coauthored with Juan Rubio-Ramírez (Fernández-Villaverde and Rubio-Ramírez, 2008). Since in that paper we were trying to explore how stable over time were the parameters of DSGE models when we let them vary over time, we took care in estimating a model that could be easily accepted by as many readers as possible as embodying the standard New Keynesian DSGE model. Even if my objectives now are different, the same justification for the model still holds: this will be as standard a model as I know how to write.

Despite any possible vehement claims to the contrary, my choice of application implies an implicit endorsement of the New Keynesian model. If I thought that the model was completely worthless, my behavior would be slightly schizophrenic. And yes, there are good things in the model. It is built around the core of the neoclassical growth model, which is the workhorse of modern macroeconomics and which offers a reasonable account of a set of important stylized facts, both at the long run and business cycle frequencies. Keynes (1936) complained in the *General Theory* that David Hume had a foot and a half in the classical world. Modern DSGE models fully share this nature.¹⁷ In addition, the model introduces a number of real and nominal rigidities that generate the higher degree of persistence we see in the data and allow for a non-trivial role of monetary policy, which as we discussed in section 2, perhaps we also find in the data.

However, we need to remember the many shortcomings of the model. We may as well begin with its core, the neoclassical growth model. Growth theorist have accumulated many objections to the basic growth model: it does not have an endogenous mechanism for long-run growth, the existence of a balanced growth path violates some observations in the data, the model does not account for the large cross-country differences in income per capita, and so forth. Our model will suffer from all of those objections.

The second line of criticism regards the nominal rigidities, which are added in an ad hoc way through Calvo pricing. Beyond the lack of microfoundations, Calvo pricing misses many

¹⁷That is why many argue, with some plausibility, that New Keynesian models are not that Keynesian after all (see Farmer, 2007). Given the importance they give to a neoclassical core, the key role of money, and the preference they generate for low and stable inflation, we could just as well call them neomonetarist models. However, after seeing the recent uproar at the University of Chicago regarding the new *Milton Friedman Institute for Research in Economics*, it is clear that the New Keynesian brand still sells better in many quarters.

aspects of the microeconomic evidence of pricing behavior by firms documented over the last few years (Álvarez *et al.* 2005, Bils and Klenow, 2004, Bils, Klenow and Malin, 2008, Dhyne *et al.* 2006, Klenow and Kryvtsov, 2008, and Nakamura and Steinsson, 2008, among many others). Finally, the model does a terrible job of pricing financial assets, a point I will revisit in section 6.

In the interest of space, I will present only the most basic description of the model without the full background. The interested reader can get many more details at the online appendix www.econ.upenn.edu/~jesusfv/benchmark_DSGE.pdf. The basic structure of the model includes a continuum of households that work, consume, and save, a continuum of intermediate good producers that rent capital and labor to manufacture intermediate goods, a final good producer that aggregates the intermediate goods, a labor “packer” that aggregates different types of labor into a homogeneous input, and a government that implements monetary policy by fixing the short-run nominal interest rate through open market operations. Both prices and wages will be subject to rigidities that limit how often they can be changed.

5.1. The Households

The first type of agents in our model will be the households. We want to have a continuum of them because, in that way, we can generate a whole distribution of wages in the economy, with each household charging its own differentiated wage. At the same time, we do not want to have too much heterogeneity, because this will make computing the model a daunting task. The trick to combine different wages but not a lot of heterogeneity is to assume a separable utility function in consumption and labor and complete markets. Complete markets give us the basic risk-sharing result that, in equilibrium, marginal utilities are equated. If utility is separable in consumption, then perfect risk-sharing implies that all households consume the same amount of the final good and hold the same level of capital, collapsing the distribution of agents along that dimension. Finally, the requirement that we have a balanced growth path implies that we want to consider utility functions of the form:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t d_t \left\{ \log(c_{jt} - hc_{jt-1}) + v \log\left(\frac{mo_{jt}}{p_t}\right) - \varphi_t \psi \frac{l_{jt}^{1+\vartheta}}{1+\vartheta} \right\}$$

where j is the index of the household, \mathbb{E}_0 is the conditional expectation operator, c_{jt} is consumption, mo_{jt}/p_t are real money balances, p_t is the aggregate price level, and l_{jt} is hours worked. In addition, we have the discount factor, β , a degree of habit persistence, h , which

will help to induce inertia in the responses of consumption to shocks, and the Frisch labor supply elasticity, $1/\vartheta$.

The period utility function is shifted by two shocks. First, a shock to intertemporal preferences, d_t , that works as a demand shock, inducing agents to consume more or less in the current period. Second, a shock to labor supply, to capture the movements in the observed wedge in the first order condition relating consumption and labor (Hall, 1997). For simplicity, we postulate that both shocks follow an autoregressive process of order 1 in logs:

$$\begin{aligned}\log d_t &= \rho_d \log d_{t-1} + \sigma_d \varepsilon_{d,t} \text{ where } \varepsilon_{d,t} \sim \mathcal{N}(0, 1), \\ \log \varphi_t &= \rho_\varphi \log \varphi_{t-1} + \sigma_\varphi \varepsilon_{\varphi,t} \text{ where } \varepsilon_{\varphi,t} \sim \mathcal{N}(0, 1).\end{aligned}$$

The standard deviation of the shocks, σ_d and σ_φ , is constant over time, but we could easily introduce a time component to it (Fernández-Villaverde and Rubio-Ramírez, 2007). Time-varying volatility in the shocks helps to understand the changing volatility of U.S. and other Western economies over the last decades that has been named the “Great Moderation” by Stock and Watson (2003).

Households trade on the whole set of Arrow-Debreu securities, contingent on idiosyncratic and aggregate events. My notation a_{jt+1} indicates the amount of those securities that pay one unit of consumption in event $\omega_{j,t+1,t}$ purchased by household j at time t at (real) price $q_{jt+1,t}$. To save on notation, we drop the explicit dependence on the event. Households also hold an amount b_{jt} of government bonds that pay a nominal gross interest rate of R_t and invest x_t . Then, the $j - th$ household’s budget constraint is:

$$\begin{aligned}c_{jt} + x_{jt} + \frac{m_{jt}}{p_t} + \frac{b_{jt+1}}{p_t} + \int q_{jt+1,t} a_{jt+1} d\omega_{j,t+1,t} \\ = w_{jt} l_{jt} + (r_t u_{jt} - \mu_t^{-1} \Phi[u_{jt}]) k_{jt-1} + \frac{m_{jt-1}}{p_t} + R_{t-1} \frac{b_{jt}}{p_t} + a_{jt} + T_t + F_t\end{aligned}$$

where w_{jt} is the real wage paid per unit of labor, r_t the real rental price of capital, $u_{jt} > 0$ the utilization rate of capital, $\mu_t^{-1} \Phi[u_{jt}]$ is the physical cost of rate u_{jt} in resource terms (where $\Phi[u] = \Phi_1(u-1) + \frac{\Phi_2}{2}(u-1)^2$ and $\Phi_1, \Phi_2 \geq 0$), μ_t is an investment-specific technological shock that shifts the relative price of capital, T_t is a lump-sum transfer from the government, and F_t is the household share of the profits of the firms in the economy. This budget constraint is slightly different from a conventional one because households are monopolistic suppliers of their own type of work j . Therefore, the household fixes w_{jt} (subject to some rigidities to be

specified below) and supplies the amount of labor l_{jt} demanded at that wage. We can think of the household either as an independent businessman who can set its own rate or as a union that negotiates a particular wage rate. This assumption is relatively inconsequential. At the cost of some additional algebra, we could also let firms set wages and households supply the desired labor at such wages.

The law of motion for capital is given by:

$$k_{jt} = (1 - \delta) k_{jt-1} + \mu_t \left(1 - S \left[\frac{x_{jt}}{x_{jt-1}} \right] \right) x_{jt}.$$

where δ is the depreciation rate. We have a quadratic adjustment cost function

$$S \left[\frac{x_t}{x_{t-1}} \right] = \frac{\kappa}{2} \left(\frac{x_t}{x_{t-1}} - \Lambda_x \right)^2,$$

where $\kappa \geq 0$ and Λ_x is the long-run growth of investment. The specification of the adjustment cost function captures the idea that the costs are with respect to moving away from the path of investment growth that we would have in the balanced growth path. In front of investment, we have an investment-specific technological shock μ_t that also follows an autoregressive process:

$$\mu_t = \mu_{t-1} \exp(\Lambda_\mu + z_{\mu,t}) \text{ where } z_{\mu,t} = \sigma_\mu \varepsilon_{\mu,t} \text{ and } \varepsilon_{\mu,t} \sim \mathcal{N}(0, 1).$$

The investment-specific technological shock accounts for the fall in the relative price of capital observed in the U.S. economy since the Second World War and it plays a crucial role in accounting for long-run growth and in generating business cycle fluctuations (see the rather compelling evidence in Greenwood, Herkowitz, and Krusell, 1997 and 2000). The process for investment-specific technological change generates the first unit root in the model and it will be one source of growth in the economy.

The Lagrangian function that summarizes the problem of the household is given by:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\lambda_t \left\{ \begin{aligned} & d_t \left\{ \log(c_{jt} - hc_{jt-1}) + v \log\left(\frac{m_{jt}}{p_t}\right) - \varphi_t \psi \frac{l_{jt}^{1+\theta}}{1+\theta} \right\} \\ & c_{jt} + x_{jt} + \frac{m_{jt}}{p_t} + \frac{b_{jt}}{p_t} + \int q_{jt+1,t} a_{jt+1} d\omega_{j,t+1,t} \\ & -w_{jt} l_{jt} - (r_t u_{jt} - \mu_t^{-1} \Phi[u_{jt}]) k_{jt-1} - \frac{m_{jt-1}}{p_t} - R_{t-1} \frac{b_{jt-1}}{p_t} - a_{jt} - T_t - F_t \\ & -Q_t \left\{ k_{jt} - (1 - \delta) k_{jt-1} - \mu_t \left(1 - S \left[\frac{x_{jt}}{x_{jt-1}} \right] \right) x_{jt} \right\} \end{aligned} \right\} \right]$$

where the household chooses c_{jt} , b_{jt} , u_{jt} , k_{jt} , x_{jt} , w_{jt} , l_{jt} and a_{jt+1} (maximization with

respect to money holdings comes from the budget constraint), λ_t is the Lagrangian multiplier associated with the budget constraint, and Q_t the Lagrangian multiplier associated with the law of motion of capital. Since I argued before that with complete markets and separable utility, marginal utilities will be equated in all states of nature and all periods, I do not need to index the multipliers by j .

The first order conditions with respect to c_{jt} , b_{jt} , u_{jt} , k_{jt} , and x_{jt} are:

$$\begin{aligned} d_t (c_{jt} - hc_{jt-1})^{-1} - h\beta\mathbb{E}_t d_{t+1} (c_{jt+1} - hc_{jt})^{-1} &= \lambda_t \\ \lambda_t &= \beta\mathbb{E}_t \left\{ \lambda_{t+1} \frac{R_t}{\Pi_{t+1}} \right\} \\ r_t &= \mu_t^{-1} \Phi' [u_{jt}] \\ Q_t &= \beta\mathbb{E}_t \left\{ (1 - \delta) Q_{t+1} + \lambda_{t+1} (r_{t+1} u_{jt+1} - \mu_{t+1}^{-1} \Phi [u_{jt+1}]) \right\} \\ -\lambda_t + Q_t \mu_t \left(1 - S \left[\frac{x_{jt}}{x_{jt-1}} \right] - S' \left[\frac{x_{jt}}{x_{jt-1}} \right] \frac{x_{jt}}{x_{jt-1}} \right) &+ \beta\mathbb{E}_t Q_{t+1} \mu_{t+1} S' \left[\frac{x_{jt+1}}{x_{jt}} \right] \left(\frac{x_{jt+1}}{x_{jt}} \right)^2 = 0. \end{aligned}$$

I do not include the first order conditions with respect to Arrow-Debreu securities, since we do not need them to solve for the equilibrium of the economy. Nevertheless, those first order conditions will be useful below to price the securities. In particular, from the second equation, we can see that

$$\frac{\lambda_{jt+1}}{\lambda_{jt}} = \frac{d_t (c_{jt+1} - hc_{jt})^{-1} - h\beta\mathbb{E}_{t+1} d_{t+2} (c_{jt+2} - hc_{jt+1})^{-1}}{d_t (c_{jt} - hc_{jt-1})^{-1} - h\beta\mathbb{E}_t d_{t+1} (c_{jt+1} - hc_{jt})^{-1}}$$

is the pricing kernel of the economy.

If we define the (marginal) Tobin's Q as $q_t = \frac{Q_t}{\lambda_t}$ (the value of installed capital in terms of its replacement cost), we find:

$$\begin{aligned} q_t &= \beta\mathbb{E}_t \left\{ \frac{\lambda_{t+1}}{\lambda_t} \left((1 - \delta) q_{t+1} + r_{t+1} u_{jt+1} - \mu_{t+1}^{-1} \Phi [u_{jt+1}] \right) \right\} \\ 1 &= q_t \mu_t \left(1 - S \left[\frac{x_{jt}}{x_{jt-1}} \right] - S' \left[\frac{x_{jt}}{x_{jt-1}} \right] \frac{x_{jt}}{x_{jt-1}} \right) + \beta\mathbb{E}_t q_{t+1} \mu_{t+1} \frac{\lambda_{jt+1}}{\lambda_{jt}} S' \left[\frac{x_{jt+1}}{x_{jt}} \right] \left(\frac{x_{jt+1}}{x_{jt}} \right)^2. \end{aligned}$$

The first equation tells us that the relative price of capital is equal to the (expected) return we will get from it in the next period

$$\underbrace{(1 - \delta) q_{t+1}}_{\text{Sale Value}} + \underbrace{r_{t+1} u_{jt+1}}_{\text{Rental Payment}} - \underbrace{\mu_{t+1}^{-1} \Phi [u_{jt+1}]}_{\text{Compensation for Utilization Rate}}$$

times the pricing kernel. The second equation determines that if $S[\cdot] = 0$ (i.e., there are no adjustment costs), we get:

$$q_t = \frac{1}{\mu_t}$$

i.e., the marginal Tobin's Q is equal to the replacement cost of capital (the relative price of capital), which falls over time as μ_t increases. Furthermore, if $\mu_t = 1$ (as we have in the basic real business cycle model), the relative price of capital is trivially equated to 1.

The necessary conditions with respect to labor and wages are more involved. There is a labor “packer” that aggregates the differentiated labor supplied by each household into a homogeneous labor unit that intermediate good producers hire in a competitive market. The aggregation is done through the following production function:

$$l_t^d = \left(\int_0^1 l_{jt}^{\frac{\eta-1}{\eta}} dj \right)^{\frac{\eta}{\eta-1}} \quad (10)$$

where $0 \leq \eta < \infty$ is the elasticity of substitution among different types of labor and l_t^d is the aggregate labor demand.

The labor “packer” maximizes profits subject to the production function (10), taking as given all differentiated labor wages w_{jt} and the wage w_t for l_t^d . Consequently, its maximization problem is:

$$\max_{l_{jt}} w_t l_t^d - \int_0^1 w_{jt} l_{jt} dj$$

After some algebra, we get the input demand functions associated with this problem:

$$l_{jt} = \left(\frac{w_{jt}}{w_t} \right)^{-\eta} l_t^d \quad \forall j \quad (11)$$

which shows that the elasticity of substitution also controls the elasticity of demand for j -th type of labor with respect to wages. Then, by using the zero profit condition for the labor “packer”:

$$w_t = \left(\int_0^1 w_{jt}^{1-\eta} dj \right)^{\frac{1}{1-\eta}}.$$

Now we can specify the wage-setting mechanism. There are several mechanisms for introducing wage rigidities but one that is particularly clever and simple is a time-dependent rule called Calvo pricing. In each period, a fraction $1 - \theta_w$ of households can reoptimize their wages and set a nominal value $p_t w_{jt}$. All other households can only partially index their

wages by past inflation with an indexation parameter $\chi_w \in [0, 1]$. Therefore, the real wage of a household that has not been able to reoptimize it for τ periods is:

$$\prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^{\chi_w}}{\Pi_{t+s}} w_{jt}$$

The probability $1 - \theta_w$ is the reduced-form representation of a more microfounded origin of wage rigidities (quadratic adjustment costs as in the original Calvo paper, 1983, contract costs, Caplin and Leahy, 1991 and 1997, or information limitations, Mankiw and Reis, 2002, or Sims, 2002), which we do not include in the model to keep the number of equations within reasonable bounds. In section 6, I will discuss the problems of Calvo pricing in detail. Suffice it to say here that, despite many potential problems, Calvo pricing is so simple that it still constitutes the natural benchmark for price and wage rigidities. This is due to the memoryless structure of the mechanism: we do not need to keep track of when wages reoptimized the last time, since θ_w is time independent.

Relying on the separability of the utility function and the presence of complete markets, the only part of the Lagrangian that gets affected by the wage and labor supply decisions of the household is:

$$\max_{w_{jt}} \mathbb{E}_t \sum_{\tau=0}^{\infty} (\beta \theta_w)^\tau \left\{ -d_t \varphi_t \psi \frac{l_{jt+\tau}^{1+\vartheta}}{1+\vartheta} + \lambda_{t+\tau} \prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^{\chi_w}}{\Pi_{t+s}} w_{jt} l_{jt+\tau} \right\} \quad (12)$$

subject to

$$l_{jt+\tau} = \left(\prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^{\chi_w}}{\Pi_{t+s}} \frac{w_{jt}}{w_{t+\tau}} \right)^{-\eta} l_{t+\tau}^d \quad \forall j$$

Note how we have modified the discount factor to include the probability θ_w that the household has to keep the wage for one more period. Once the household can reoptimize, the continuation of the decision problem is independent from our choice of wage today, and hence, we do not need to include it in the section of the Lagrangian in equation (12). We also assume that

$$(\beta \theta_w)^\tau \lambda_{t+\tau} \prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^{\chi_w}}{\Pi_{t+s}}$$

goes to zero for the previous sum to be well defined.

Also, because of complete markets, all of the households reoptimizing wages in the current period will pick the same wage and we can drop the j th from w_{jt} . The first order condition

of this problem is then:

$$\begin{aligned} & \frac{\eta - 1}{\eta} w_t^* \mathbb{E}_t \sum_{\tau=0}^{\infty} (\beta \theta_w)^\tau \lambda_{t+\tau} \left(\prod_{s=1}^{\tau} \frac{\Pi_{t+s}^{\chi_w}}{\Pi_{t+s}} \right)^{1-\eta} \left(\frac{w_t^*}{w_{t+\tau}} \right)^{-\eta} l_{t+\tau}^d = \\ & \mathbb{E}_t \sum_{\tau=0}^{\infty} (\beta \theta_w)^\tau \left(d_{t+\tau} \varphi_{t+\tau} \psi \left(\prod_{s=1}^{\tau} \frac{\Pi_{t+s}^{\chi_w}}{\Pi_{t+s}} \frac{w_t^*}{w_{t+\tau}} \right)^{-\eta(1+\vartheta)} (l_{t+\tau}^d)^{1+\vartheta} \right) \end{aligned}$$

where w_t^* is the new optimal wage. This expression involves infinite sums that are difficult to handle computationally. It is much simpler to write the first order conditions as $f_t^1 = f_t^2$ where we have the recursive functions:

$$f_t^1 = \frac{\eta - 1}{\eta} (w_t^*)^{1-\eta} \lambda_t w_t^\eta l_t^d + \beta \theta_w \mathbb{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{1-\eta} \left(\frac{w_{t+1}^*}{w_t^*} \right)^{\eta-1} f_{t+1}^1$$

and:

$$f_t^2 = \psi d_t \varphi_t \left(\frac{w_t}{w_t^*} \right)^{\eta(1+\vartheta)} (l_t^d)^{1+\vartheta} + \beta \theta_w \mathbb{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{-\eta(1+\vartheta)} \left(\frac{w_{t+1}^*}{w_t^*} \right)^{\eta(1+\vartheta)} f_{t+1}^2.$$

Now, if we put the previous equations together and drop the j 's indexes (that are redundant), we have the first order conditions

$$\begin{aligned} & d_t (c_t - h c_{t-1})^{-1} - h \beta \mathbb{E}_t d_{t+1} (c_t - h c_t)^{-1} = \lambda_t \\ & \lambda_t = \beta \mathbb{E}_t \left\{ \lambda_{t+1} \frac{R_t}{\Pi_{t+1}} \right\} \\ & r_t = \mu_t^{-1} \Phi' [u_t] \\ & q_t = \beta \mathbb{E}_t \left\{ \frac{\lambda_{t+1}}{\lambda_t} [(1 - \delta) q_{t+1} + (r_{t+1} u_{t+1} - \mu_{t+1}^{-1} \Phi [u_{t+1}])] \right\} \\ & 1 = q_t \mu_t \left(1 - S \left[\frac{x_t}{x_{t-1}} \right] - S' \left[\frac{x_t}{x_{t-1}} \right] \frac{x_t}{x_{t-1}} \right) + \beta \mathbb{E}_t q_t \mu_{t+1} \frac{\lambda_{t+1}}{\lambda_t} S' \left[\frac{x_{t+1}}{x_t} \right] \left(\frac{x_{t+1}}{x_t} \right)^2. \end{aligned}$$

the budget constraint:

$$\begin{aligned} & c_{jt} + x_{jt} + \frac{m_{jt}}{p_t} + \frac{b_{jt+1}}{p_t} + \int q_{jt+1,t} a_{jt+1} d\omega_{j,t+1,t} \\ & = w_{jt} l_{jt} + (r_t u_{jt} - \mu_t^{-1} \Phi [u_{jt}]) k_{jt-1} + \frac{m_{jt-1}}{p_t} + R_{t-1} \frac{b_{jt}}{p_t} + a_{jt} + T_t + F_t \end{aligned}$$

and the laws of motion for f_t :

$$f_t = \frac{\eta - 1}{\eta} (w_t^*)^{1-\eta} \lambda_t w_t^\eta l_t^d + \beta \theta_w \mathbb{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{1-\eta} \left(\frac{w_{t+1}^*}{w_t^*} \right)^{\eta-1} f_{t+1}$$

and:

$$f_t = \psi d_t \varphi_t (\Pi_t^{*w})^{-\eta(1+\vartheta)} (l_t^d)^{1+\vartheta} + \beta \theta_w \mathbb{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{-\eta(1+\vartheta)} \left(\frac{w_{t+1}^*}{w_t^*} \right)^{\eta(1+\vartheta)} f_{t+1}.$$

where $\Pi_t^{*w} = w_t^*/w_t$.

The real wage evolves as a geometric average of the past real wage and the new optimal wage:

$$w_t^{1-\eta} = \theta_w \left(\frac{\Pi_{t-1}^{\chi_w}}{\Pi_t} \right)^{1-\eta} w_{t-1}^{1-\eta} + (1 - \theta_w) w_t^{*1-\eta}.$$

5.2. The Final Good Producer

There is one final good producer that aggregates intermediate goods y_{it} with the production function:

$$y_t^d = \left(\int_0^1 y_{it}^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}. \quad (13)$$

where ε is the elasticity of substitution. Similarly to the labor “packer,” the final good producer is perfectly competitive and maximizes profits subject to the production function (13), taking as given all intermediate goods prices p_{it} and the final good price p_t . Thus, the input demand functions associated with this problem are:

$$y_{it} = \left(\frac{p_{it}}{p_t} \right)^{-\varepsilon} y_t^d \quad \forall i,$$

where y_t^d is aggregate demand and the price level is:

$$p_t = \left(\int_0^1 p_{it}^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}.$$

5.3. Intermediate Good Producers

As mentioned above, there is a continuum of intermediate goods producers, each of which has access to a production function $y_{it} = A_t k_{it-1}^\alpha (l_{it}^d)^{1-\alpha} - \phi z_t$ where k_{it-1} is the capital rented by the firm, l_{it}^d is the amount of the labor input rented from the labor “packer,” and where

A_t , a neutral technology level, evolves as:

$$A_t = A_{t-1} \exp(\Lambda_A + z_{A,t}) \text{ where } z_{A,t} = \sigma_A \varepsilon_{A,t} \text{ and } \varepsilon_{A,t} \sim \mathcal{N}(0, 1)$$

This process incorporates a second unit root in the model. The fixed cost of production ϕ is indexed by the variable $z_t = A_t^{\frac{1}{1-\alpha}} \mu_t^{\frac{\alpha}{1-\alpha}}$ to make it grow with the economy (think, for example, of the fixed cost of paying some fees for keeping the factory open: it is natural to think that the fees will increase with income). Otherwise, the fixed cost would become asymptotically irrelevant. In a balanced growth path, z_t is precisely the growth factor in the economy that we want to scale for. The role of the fixed cost is to roughly eliminate profits in equilibrium and to allow us to dispense with the entry and exit of intermediate good producers.

Since $z_t = A_t^{\frac{1}{1-\alpha}} \mu_t^{\frac{\alpha}{1-\alpha}}$, we can combine the processes for A_t and μ_t to get:

$$z_t = z_{t-1} \exp(\Lambda_z + z_{z,t}) \text{ where } z_{z,t} = \frac{z_{A,t} + \alpha z_{\mu,t}}{1 - \alpha} \text{ and } \Lambda_z = \frac{\Lambda_A + \alpha \Lambda_\mu}{1 - \alpha}.$$

Many of the variables in the economy, like c_t , will be cointegrated in equilibrium with z_t . This cointegration captures the evidence of constant main ratios of the economy in a stochastic trend environment with the advantage that, with respect to the empirical literature, the cointegration vector is microfounded and implied by the optimization decision of the agents in the model and not exogenously postulated by the econometrician (for the origin of this idea, see King *et al.*, 1991).

The problem of intermediate goods producers can be chopped into two parts. First, given input prices w_t and r_t , they rent l_{it}^d and k_{it-1} to minimize real cost:

$$\min_{l_{it}^d, k_{it-1}} w_t l_{it}^d + r_t k_{it-1}$$

subject to their supply curve:

$$y_{it} = \begin{cases} A_t k_{it-1}^\alpha (l_{it}^d)^{1-\alpha} - \phi z_t & \text{if } A_t k_{it-1}^\alpha (l_{it}^d)^{1-\alpha} \geq \phi z_t \\ 0 & \text{otherwise} \end{cases}$$

The solution of this problem implies that all intermediate good firms equate their capital-labor ratio to the ratio of input prices times a constant:

$$\frac{k_{it-1}}{l_{it}^d} = \frac{\alpha}{1 - \alpha} \frac{w_t}{r_t}$$

and that the marginal cost mc_t is

$$mc_t = \left(\frac{1}{1-\alpha} \right)^{1-\alpha} \left(\frac{1}{\alpha} \right)^\alpha \frac{w_t^{1-\alpha} r_t^\alpha}{A_t}$$

A useful observation is that neither of these expressions depends on i since A_t and input prices are common for all firms.

The second part of the problem is to set a price for the intermediate good. In a similar vein to the household, the intermediate good producer is subject to Calvo pricing, where now the probability of reoptimizing prices is $1 - \theta_p$ and the indexation parameter is $\chi \in [0, 1]$.

Therefore, the problem of the firms is:

$$\max_{p_{it}} \mathbb{E}_t \sum_{\tau=0}^{\infty} (\beta\theta_p)^\tau \frac{\lambda_{t+\tau}}{\lambda_t} \left\{ \left(\prod_{s=1}^{\tau} \Pi_{t+s-1}^\chi \frac{p_{it}}{p_{t+\tau}} - mc_{t+\tau} \right) y_{it+\tau} \right\}$$

subject to

$$y_{it+\tau} = \left(\prod_{s=1}^{\tau} \Pi_{t+s-1}^\chi \frac{p_{it}}{p_{t+\tau}} \right)^{-\varepsilon} y_{t+\tau}^d,$$

where future profits are valued using the pricing kernel $\lambda_{t+\tau}/\lambda_t$.

The first order condition of this problem, after some algebra and noticing that $p_{it}^* = p_t^*$,

$$\mathbb{E}_t \sum_{\tau=0}^{\infty} (\beta\theta_p)^\tau \lambda_{t+\tau} \left\{ \left((1-\varepsilon) \left(\prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^\chi}{\Pi_{t+s}} \right)^{1-\varepsilon} \frac{p_t^*}{p_t} + \varepsilon \left(\prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^\chi}{\Pi_{t+s}} \right)^{-\varepsilon} mc_{t+\tau} \right) y_{t+\tau}^d \right\} = 0$$

This expression tells us that the price is equal to a weighted sum of future expected mark-ups.

We can express this condition recursively as:

$$\begin{aligned} \varepsilon g_t^1 &= (\varepsilon - 1) g_t^2 \\ g_t^1 &= \lambda_t mc_t y_t^d + \beta\theta_p \mathbb{E}_t \left(\frac{\Pi_t^\chi}{\Pi_{t+1}} \right)^{-\varepsilon} g_{t+1}^1 \\ g_t^2 &= \lambda_t \Pi_t^* y_t^d + \beta\theta_p \mathbb{E}_t \left(\frac{\Pi_t^\chi}{\Pi_{t+1}} \right)^{1-\varepsilon} \left(\frac{\Pi_t^*}{\Pi_{t+1}^*} \right) g_{t+1}^2 \end{aligned}$$

where:

$$\Pi_t^* = \frac{p_t^*}{p_t}.$$

Given Calvo's pricing, the price index evolves as:

$$1 = \theta_p \left(\frac{\Pi_{t-1}^X}{\Pi_t} \right)^{1-\varepsilon} + (1 - \theta_p) \Pi_t^{*1-\varepsilon}$$

5.4. The Government Problem

The last agent in the model is the government. To simplify things I forget about fiscal policy and I assume that the government follows a simple Taylor rule:

$$\frac{R_t}{R} = \left(\frac{R_{t-1}}{R} \right)^{\gamma_R} \left(\left(\frac{\Pi_t}{\Pi} \right)^{\gamma_\Pi} \left(\frac{\frac{y_t^d}{y_{t-1}^d}}{\Lambda_z} \right)^{\gamma_y} \right)^{1-\gamma_R} \exp(m_t)$$

that sets the short-term nominal interest rates as a function of past interest rates, inflation and the “growth gap”: the ratio between the growth of aggregate demand, $\frac{y_t^d}{y_{t-1}^d}$, and the average growth of the economy, Λ_z . Introducing this growth gap avoids the need to specify a measure of the output gap (always somehow arbitrary) and, more important, fits the evidence better (Orphanides, 2002). The term m_t is a random shock to monetary policy such that $m_t = \sigma_m \varepsilon_{mt}$, where $\varepsilon_{mt} \sim \mathcal{N}(0, 1)$.

The other elements in the Taylor rule are the target level of inflation, Π , and the steady state nominal gross return of capital, R . Since we are dealing with a general equilibrium model, the government can pick either Π or R but not both (R is equal to Π times the steady state real interest rate).

The nominal interest rate can be implemented either through open market operations (as has been the norm for the last several decades) or through paying interest on bank reserves (as the Fed has recently begun to do in the United States). In both cases, monetary policy generates either a surplus (or a deficit) that is eliminated through lump-sum transfers T_t to households.

5.5. Aggregation, Equilibrium, and Solution

Now, we can add all of the previous expressions to find aggregate variables and define an equilibrium. First, we have aggregate demand, $y_t^d = c_t + x_t + \mu_t^{-1} \Phi[u_t] k_{t-1}$. Second, by noticing that all the intermediate good firms will have the same capital-labor ratio, we can find aggregate supply:

$$y_t = \frac{A_t (u_t k_{t-1})^\alpha (l_t^d)^{1-\alpha} - \phi z_t}{v_t^p}$$

where

$$v_t^p = \int_0^1 \left(\frac{p_{it}}{p_t} \right)^{-\varepsilon} di$$

is an inefficiency factor created by price dispersion, and

$$l_t^d = \frac{l_t}{v_t^w}$$

is labor packed where

$$v_t^w = \int_0^1 \left(\frac{w_{jt}}{w_t} \right)^{-\eta} dj.$$

is an inefficiency factor created by wage dispersion. Furthermore, by Calvo's pricing

$$v_t^p = \theta_p \left(\frac{\Pi_{t-1}^X}{\Pi_t} \right)^{-\varepsilon} v_{t-1}^p + (1 - \theta_p) \Pi_t^{*- \varepsilon}$$

and

$$v_t^w = \theta_w \left(\frac{w_{t-1}}{w_t} \frac{\Pi_{t-1}^{X_w}}{\Pi_t} \right)^{-\eta} v_{t-1}^w + (1 - \theta_w) (\Pi_t^{w*})^{-\eta}$$

A definition of equilibrium in this economy is standard and it is characterized by first order conditions of the household, the first order conditions of the firms, the recursive definitions of g_t^1 and g_t^2 , the Taylor rule of the government, and market clearing.

Since the model has two unit roots, one in the investment-specific technological change and one in the neutral technological change, we need to rescale all the variables to avoid solving the model with non-stationary variables (a solution that is feasible, but most cumbersome). The scaling will be given by the variable z_t in such a way that, for any arbitrary variable x_t , we will have $\tilde{x}_t = x_t/z_t$. Partial exceptions are the variables, $\tilde{r}_t = r_t\mu_t$, $\tilde{q}_t = q_t\mu_t$, and $\tilde{k}_t = \frac{k_t}{z_t\mu_t}$. Once the model has been rescaled, we can find the steady state and solve the model by loglinearizing around the steady state.

Loglinearization is both a fast and efficient method for solving large-scale DSGE models. I have documented elsewhere (Aruoba, Fernández-Villaverde, and Rubio-Ramírez, 2006), that it is a nice compromise between speed and accuracy in many applications of interest. Furthermore, it can easily be extended to include higher order terms (Judd, 1998). Once I have solved the model, I use the Kalman filter to evaluate the likelihood of the model, given some parameter values. The whole process takes less than 1 second per evaluation of the likelihood.

5.6. Empirical Results

I estimate the DSGE model using five time series for the U.S. economy: 1) the relative price of investment with respect to the price of consumption, 2) real output per capita growth, 3) real wages per capita, 4) the consumer price index, and 5) the federal funds rate (the interest rate at which banks lend balances at the Federal Reserve System to each other, usually overnight). This series captures the main aspects of the dynamics of the data and model much of the information that a policy maker is interested in. The sample is 1959.Q1 - 2007.Q1.

To find the real output per capita series, I first define nominal output as nominal consumption plus nominal gross investment. Nominal consumption is the sum of personal consumption expenditures on nondurable goods and services while nominal gross investment is the sum of personal consumption expenditures on durable goods, private residential investment, and nonresidential fixed investment. Per capita nominal output is equal to the ratio between our nominal output series and the civilian noninstitutional population between 16 and 65. I transform nominal quantities into real ones using the investment deflator computed by Fisher (2006), a series that unfortunately ends early in 2000:Q4. Following Fisher's methodology, I have extended the series to 2007:Q1. Real wages are defined as compensation per hour in the nonfarm business sector divided by the CPI deflator.

My next step is to specify priors. To facilitate the task of the reader who wants to continue exploring the estimation of DSGE models, I would follow the choices of Smets and Wouters (2007) with a few trivial changes. Instead of a long (and, most likely, boring) discussion of each prior, I just point out that I am selecting mainstream priors that are centered around the median value of estimates of micro and macro data. Also, I fix some parameters that are very difficult to identify in the data. The priors are given by:

Table 1: Priors					
$100(\beta^{-1} - 1)$	h	ψ	θ_p	χ	θ_w
$Ga(0.25, 0.1)$	$Be(0.7, 0.1)$	$N(9, 3)$	$Be(0.5, 0.1)$	$Be(0.5, 0.15)$	$Be(0.5, 0.1)$
χ_w	γ_R	γ_y	γ_π	$100(\Pi - 1)$	ϑ
$Be(0.5, 0.1)$	$Be(0.75, 0.1)$	$N(0.12, 0.05)$	$N(1.5, 0.125)$	$Ga(0.95, 0.1)$	$N(1, 0.25)$
κ	α	ρ_d	ρ_φ	$\exp(\sigma_A)$	$\exp(\sigma_d)$
$N(4, 1.5)$	$N(0.3, 0.025)$	$Be(0.5, 0.2)$	$Be(0.5, 0.2)$	$IG(0.1, 2)$	$IG(0.1, 2)$
$\exp(\sigma_\varphi)$	$\exp(\sigma_\mu)$	$\exp(\sigma_e)$	$100\Lambda_\mu$	$100\Lambda_A$	
$IG(0.1, 2)$	$IG(0.1, 2)$	$IG(0.1, 2)$	$N(0.34, 0.1)$	$N(0.178, 0.075)$	

while the fixed parameters are:

δ	ε	η	ϕ	Φ_2
0.025	10	10	0	0.001

Perhaps the only two fixed parameters that are interesting to discuss are ε and η , both with a value of 10. These values imply an average mark-up of around 10 percent, in line with many estimates.

I generate 75,000 draws from the posterior using a random walk Metropolis-Hastings. While 75,000 draws is a comparatively low number, there was a substantial and long search for good initial parameter values, which means that the estimates were stable and passed all the usual tests of convergence. The posterior medians and the 5 and 95 percentile values of the 23 estimated parameters of the model are reported in table 3. Figure 1 plots the histograms of each parameter (one can think of the likelihood as the combination of all those histograms in a highly dimensional object).¹⁸

β	h	ψ	ϑ	κ	α
0.998 [0.997,0.999]	0.97 [0.95,0.98]	8.92 [4.09,13.84]	1.17 [0.74,1.61]	9.51 [7.47,11.39]	0.21 [0.17,0.26]
θ_p	χ	θ_w	χ_w	γ_R	γ_y
0.82 [0.78,0.87]	0.63 [0.46,0.79]	0.68 [0.62,0.73]	0.62 [0.44,0.79]	0.77 [0.74,0.81]	0.19 [0.13,0.27]
γ_π	Π	ρ_d	ρ_φ	σ_A	σ_d
1.29 [1.02,1.47]	1.010 [1.008,1.011]	0.12 [0.04,0.22]	0.93 [0.89,0.96]	-3.97 [-4.17,-3.78]	-1.51 [-1.82,-1.11]
σ_φ	σ_μ	σ_e	Λ_μ	Λ_A	
-2.36 [-2.76,-1.74]	-5.43 [-5.52,-5.35]	-5.85 [-5.94,-5.74]	$3.4e - 3$ [0.003,0.004]	$2.8e - 3$ [0.002,0.004]	

[FIGURE 1 HERE]

What do we learn from our estimates? First, the discount factor β is very high, 0.998. This is quite usual in DSGE models, since the likelihood wants to match a low interest rate.

¹⁸These results are also reported in Fernández-Villaverde, Guerrón-Quintana, and Rubio-Ramírez (2008).

Since we have long-run growth in the model, the log utility function generates a relatively high interest rate without the help of any discounting. Second, we have a very high degree of habit, around 0.97. This is necessary to match the slow response of the economy to shocks as documented by innumerable number of VAR exercises. Third, the Frisch elasticity of labor supply is 0.85 (1/1.17). This is a nice surprise, since it is a relatively low number, which makes it quite close to the estimates of the micro literature (in fact, some micro estimates are higher than 0.85!). Since one of the criticisms of DSGE models has traditionally been that they assumed a large Frisch elasticity, our model does not suffer from this problem.¹⁹

Investment is subject to high adjustment costs, 9.51. Again, this is because we want to match a slow response of investment to shocks. The elasticity of output to capital, 0.21, is very low but similar to the results by Smets and Wouters (2007). When we interpret this number, we need to remember that, on top of the payments to capital, we have the profits of the intermediate good producers. Since the national income and product accounts lump together both quantities as gross operating surplus, the result is consistent with the evidence on income distribution.

The estimates also reveal a fair amount of nominal rigidities. The Calvo parameter for price adjustment, θ_p , is 0.82 (an average five-quarter pricing cycle) and for wages it is 0.68 (an average three-quarter wage cycle). The indexation level for prices, χ , is 0.63, and the indexation for wages is nearly the same, 0.62. Despite the fair amount of uncertainty in the posterior (for example, χ ranges between 0.46 and 0.79), the model points out the important role of nominal rigidities. There is, of course, the counterargument that since the only way the model can account for the effects of monetary shocks is through picking up large nominal rigidities, the likelihood takes us to zones with high rigidity. In a model with other channels for monetary policy to play a role (for example, with imperfect common knowledge), the likelihood may prefer less nominal rigidities. In that sense, if the DSGE model is extremely misspecified, our inference may lead us to wrong conclusions.

The estimates for the coefficients of the Taylor rule are in line with the estimates of single equation models (Clarida, Galí, and Gertler, 2000). The coefficient on inflation, $\gamma_\pi = 1.29$, shows that the Fed respects the Taylor principle (without entering here into a discussion of whether it did in different subperiods as defended by Lubick and Schorfheide, 2004). The coefficient on output, $\gamma_y = 0.19$, signals a weak but positive response to the output growth

¹⁹On the other hand, the model, like all New Keynesian models, requires quite large preference shocks. It is not clear to me that we have made much progress by substituting a high Frisch elasticity for these large shocks.

gap. The coefficient on lagged interest rates, $\gamma_R = 0.77$, indicates a strong desire to smooth the changes on nominal interest rates over time, which has been attributed either to an avoidance of disruptions in the money market or to allow new information about the state of the economy to emerge more fully before a large change in monetary policy is fully passed on. The estimated target inflation is a quarterly 1 percent, perhaps high by today's standards but in line with the behavior of prices during the whole sample.

The growth rates of the investment-specific technological change, Λ_μ , and of the neutral technology, Λ_A , are roughly equal. This means that the estimated average growth rate of the U.S. economy in per capita terms, $(\Lambda_A + \alpha\Lambda_\mu) / (1 - \alpha)$ is 0.43 percent per quarter, or 1.7 percent annually. Finally, the estimated standard deviations of shocks show an important role for both technological shocks and for preference shocks.

6. Areas of Future Research

In the next few pages, I will outline some of the potential areas of future research for the formulation and estimation of DSGE models. I do not attempt to map out all existing problems. Beyond being rather foolish, it would take me dozens of pages just to briefly describe some of the open questions I am aware of. I will just talk about three questions I have been thinking about lately: better pricing mechanisms, asset pricing, and more robust inference.

6.1. Better Pricing Mechanisms

In our application, we assumed a simple Calvo pricing mechanism. Unfortunately, the jury is still out regarding how bad a simplification it is to assume that the probability of changing prices (or wages) is fixed and exogenous. Dotsey, King, and Wolman (1999), in an important paper, argue that state-dependent pricing (firms decide when to change prices given some costs and their states) is not only a more natural setup for thinking about rigidities but also an environment that may provide very different answers than the basic Calvo pricing.

More recently, Bils, Klenow, and Malin (2008) have presented compelling evidence that state-dependent pricing is also a better description of the data. Bils, Klenow and Malin's paper is a remarkably nice contribution because the mapping between microevidence of price and wages changes and nominal rigidity in the aggregate is particularly subtle. An interesting characteristic of our Calvo pricing mechanism is that all the wages are being changed in

every period, some because of reoptimization, some because of indexing. Therefore, strictly speaking, the average duration of wages in this model is one period. In the normal setup, we equate a period with one quarter, which indicates that any lower degree of price changes in the data implies that the model display “excess” price volatility. A researcher must then set up a smart “mousetrap.” Bills, Klenow, and Malin find their trap in the reset price inflation that they build from micro CPI data. This reset inflation clearly indicates that Calvo pricing cannot capture many of the features of the micro data and the estimated persistence of shocks. The bad news is, of course, that handling a state-dependent pricing model is rather challenging (we have to track a non-trivial distribution of prices), which limits our ability to estimate it. Being able to write, solve, and estimate DSGE models with better pricing mechanisms is, therefore, a first order of business.

6.2. Asset Pricing

So far, assets and asset pricing have only made a collateral appearance in our exposition. This is a defect common to much of macroeconomics, where quantities (consumption, investment, hours worked) play a much bigger role than prices. However, if we take seriously the implications of DSGE models for quantities, it is inconsistent not to do the same for prices, in particular asset prices. You cannot believe the result while denying the mechanism: it is through asset prices that the market signals the need to increase or decrease current consumption and, in conjunction with wages, the level of hours worked. Furthermore, one of the key questions of modern macroeconomics, the welfare cost of aggregate fluctuation, is, in a precise sense, an exercise in asset pricing. Roughly speaking, a high market price for risk will denote a high welfare cost of aggregate fluctuations and low market price for risk, a low welfare cost.

The plight with asset prices is, of course, that DSGE models do a terrible job at matching them: we cannot account for the risk-free interest rate (Weil, 1989), the equity premium (Mehra and Prescott, 1985), the excess volatility puzzle, the value premium, the slope of the yield curve, or any other of a long and ever-growing list of related observations (Campbell, 2003).

The origin of our concerns is that the stochastic discount factor (SDF) implied by the model does not covariate with observed returns in the correct way (Hansen and Jagannathan, 1991). For ease of exposition, let me set $h = 0$ (the role of habits will become clearer in a moment) and use the equilibrium condition that individual consumption is equal to aggregate

consumption. Then, the SDF m_t is:

$$m_t = \beta \frac{c_t}{c_{t+1}} = \beta \frac{z_t \tilde{c}_t}{z_{t+1} \tilde{c}_{t+1}} = \beta e^{-\Lambda_z - z_{z,t+1}} \frac{\tilde{c}_t}{\tilde{c}_{t+1}}$$

Since (detrended) consumption is rather smooth in the data:

$$\frac{\tilde{c}_t}{\tilde{c}_{t+1}} \approx 1$$

and the variance of $z_{z,t+1}$ is low, we have that $\mathbb{E}_t m_t \approx \beta e^{-\Lambda_z}$ and $\sigma_t(m_t)$ is small.

To get a sense of the importance of the first result, we can plug in some reasonable value for the parameters. The annual long-run per capita growth of the U.S. economy between 1865-2007 has been around 1.9 percent. I then set $\Lambda_z = 0.019$. For the discount factor, I pick $\beta = 0.999$, which is even higher than our point estimate in section 5 but which makes my point even stronger. Thus, the gross risk-free real interest rate, R_t is equal to:

$$R_t = (\mathbb{E}_t m_t)^{-1} \approx \beta^{-1} e^{\Lambda_z} = 1.02$$

However, in the data, we find that the risk-free real interest rate has been around 1 percent (Campbell, 2003). This is, in a nutshell, the risk-free interest rate: even in a context where agents practically do not discount the future and where the elasticity of intertemporal substitution (EIS) is 1, we create a high interest rate. By lowering β or the EIS to more reasonable numbers, we only make the puzzle stronger. The extension of the previous formula for the general constant relative risk aversion utility function is:

$$R_t = (\mathbb{E}_t m_t)^{-1} \approx \beta^{-1} e^{\frac{1}{\psi} \Lambda_z}$$

where ψ is the EIS. Even by lowering the EIS to 0.5, we would have that $e^{\frac{1}{\psi} \Lambda_z}$ would be around 1.04, which closes the door to any hope of ever matching the risk-free interest rate.

The second result, m_t fluctuates very little, implies that the market price for risk,

$$\frac{\sigma_t(m_t)}{\mathbb{E}_t m_t}$$

is also low. But this observation just runs in the completely opposite direction of the equity premium puzzle, where, given historical stock returns, we require a large market price for risk.

How can we fix the behavior of the SDF in a DSGE model? My previous argument relied on three basic components. First, that consumption fluctuates little. Second, that the marginal utility of consumption fluctuates little when consumption changes a small amount, and third, that the pricing of assets is done with the SDF. The first component seems robust. Notwithstanding recent skepticism (Barro, 2006), I find little evidence of the large fluctuations in consumption that we would need to make the model work. Even during the Great Depression, the yearly fluctuations were smaller than the total drop in consumption over the whole episode (the number that Barro uses) and they were accompanied by an increase in leisure. Exploring the third component, perhaps with incomplete markets or with bounded rationality, is to venture into a wild territory beyond my current fancy for emotions. My summary dismissals of the first and last argument force me to conclude that marginal utility must, somehow, substantially fluctuate when consumption moves just a little bit.

The standard constant relative risk aversion utility functions just cannot deliver these large fluctuations in marginal utility in general equilibrium. As we raise risk aversion, consumers respond by making their consumption decisions smoother. Indeed, for sufficiently large levels of risk aversion, consumption is so smooth that the market price for risk actually falls (Rouwenhorst, 1995).

Something we can do is to introduce habits, as I did in the model that I estimated before. Then, the SDF becomes

$$m_t = \beta \frac{d_t (c_{t+1} - hc_t)^{-1} - h\beta \mathbb{E}_{t+1} d_{t+2} (c_{t+2} - hc_{t+1})^{-1}}{d_t (c_t - hc_{t-1})^{-1} - h\beta \mathbb{E}_t d_{t+1} (c_{t+1} - hc_t)^{-1}}$$

and for a sufficiently high level of h , we can obtain large fluctuations of the SDF. The intuition is that, as $h \rightarrow 1$, we care about the ratio of the first differences in consumption and not the ratio of levels, and this ratio of first differences can be quite large. Habits are plausible (after a few trips in business class, coming back to coach is always a tremendous shock) and there may be some good biological reasons why nature has given us a utility function with habits (Becker and Rayo, 2007). At the same time, we do not know much about the right way to introduce habits in the utility function (the simple form postulated above is rather arbitrary and rejected by the data, as shown by Chen and Ludvigson, 2008) and habits generate interest rates that are too volatile.

Consequently, a second avenue is the exploration of “exotic preferences.” Standard expected utility functions, like the one used in this paper, face many theoretical limitations.

Without being exhaustive, standard expected utility functions do not capture a preference for the timing of resolution of uncertainty, they do not reflect attitudes toward ambiguity, and they cannot accommodate loss aversion. Moreover, the standard model assumes that economic agents do not fear misspecification: they are sure that the model in their heads is the same as the true description of the world. These limitations are potentially of empirical importance as they may be behind our inability to account for many patterns in the data, in particular the puzzling behavior of the prices of many assets and the risk premia (Bansal and Yaron, 2004). Over the last several years, economists have paid increasing attention to new forms of the utility function or with fear of misspecification. As a result of this interest, there is a growing excitement about the potentialities of this research area (see the survey by Backus, Routledge, and Zin, 2005, and the monograph by Hansen and Sargent, 2007, for models where the agents want to behave in a way that is robust to misspecification mistakes). However, disappointingly little work has been done in the empirical estimation of DSGE models (or even partial equilibrium models) with this type of preferences (see the review of Hansen *et al.*, 2007). A better and more realistic understanding of utility functions is bound to deliver high yields and this understanding must rely on good econometrics (for some recent attempts, see some of my own work on estimation of models with Epstein-Zin preferences: Binsbergen *et al.*, 2008).

6.3. More Robust Inference

The relative disadvantage of Bayesian methods when dealing with semiparametrics that we discussed in section 3 is unsatisfactory. DSGE models are complex structures. To make the models useful, researchers add many mechanisms that affect the dynamics of the economy: sticky prices, sticky wages, adjustment costs, etc. In addition, DSGE models require many parametric assumptions: the utility function, the production function, the adjustment costs, the distribution of shocks, etc.

Some of those parametric choices are based on restrictions that the data impose on the theory. For example, the observation that labor income share has been relative constant since the 1950s suggests that a Cobb-Douglas production function may not be a bad approximation to reality (although this assumption itself is problematic: see the evidence in Young, 2005, among others). Similarly, the observation that the average labor supplied by adults in the U.S. economy has been relatively constant over the last several decades requires a utility function with a marginal rate of substitution between leisure and consumption that is linear

in consumption.

Unfortunately, many other parametric assumptions do not have much of an empirical foundation. Instead, researchers choose parametric forms for those functions based only on convenience. For example, in the prototypical DSGE model that we presented in the previous section, the investment adjustment cost function $S(\cdot)$ plays an important role in the dynamics of the economy. However, we do not know much about this function. Even the mild restrictions that we imposed are not necessarily true in the data.²⁰ For example, there is much evidence of non-convex adjustment costs at the plant level (Cooper and Haltiwanger, 2006) and of nonlinear aggregate dynamics (Caballero and Engel, 1999). Similarly, we assume a Gaussian structure for the shocks driving the dynamics of the economy. However, there is much evidence (Geweke, 1993 and 1994, Fernández-Villaverde and Rubio-Ramírez, 2007) that shocks to the economy are better described by distributions with fat tails.

The situation is worrisome. Functional form misspecification may contaminate the whole inference exercise. Moreover, Heckman and Singer (1984) show that the estimates of dynamic models are inconsistent if auxiliary assumptions (in their case, the modelling of individual heterogeneity in duration models) are misspecified. These concerns raise the question of how we can conduct inference that is more robust to auxiliary assumptions, especially within a Bayesian framework.

Researchers need to develop new techniques that allow for the estimation of DSGE models using a Bayesian framework where we can mix tight parametric assumptions along some dimensions while keeping as much flexibility as possible in those aspects of the model that we have less confidence with. The potential benefits from these new methods are huge. Our approach shares many lines of contact with Chen and Ludvigson (2008), a paper that has pioneered the use of more general classes of functions when estimating dynamic equilibrium models within the context of methods of moments. Also, I am intrigued by the possibilities of ideas like those in Álvarez and Jermann (2004), who use data from asset pricing to estimate the welfare cost of the business cycle without the need to specify particular preferences. In a more theoretical perspective, Kimball (2002) has worked out many implications of DSGE models that do not depend on parametric assumptions. Some of these implications are potentially usable for estimation.

²⁰If we are linearizing the model or computing a second order approximation, we do not need to specify more of the function than those properties. However, if we want to compute arbitrarily high order approximations or use a projection solution method, we will need to specify a full parametric form.

7. Concluding Remarks

I claimed in the introduction that the New Macroeconometrics is a new and exciting area of research. The previous pages, even if brief, have attempted to show the reader why the field is important and how it defines the new gold standard of empirical research in macroeconomics. But there is an even better part of the deal. Much needs to be done in the field: the number of papers I can think about writing in the next decade, both theoretical and applied, is nearly unbounded (and, of course, I can only think about a very small subset of all the possible and interesting papers to write). Since my ability and the ability of other practitioners in the New Macroeconometrics are limited by the tight constraints of time, we need more eager young minds to join us. I hope that some readers will find this call intriguing.

References

- [1] Abreu, D., D. Pearce, and E. Stacchetti (1990). "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica* 58, 1041-1063.
- [2] Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2005). "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through." *Sveriges Riksbank Working Paper Series* 179.
- [3] Alesina, A., E. Glaeser, and B. Sacerdote (2006). "Work and Leisure in the U.S. and Europe: Why So Different?" *NBER Macroeconomics Annual* 2005, 1-64.
- [4] Álvarez, F. and U.J. Jermann (2004). "Using Asset Prices to Measure the Cost of the Business Cycle." *Journal of Political Economy* 112, 1223-1256.
- [5] Álvarez, L. J. , P. Burriel, and I. Hernando (2005). "Do Decreasing Hazard Functions for Price Changes Make Sense?" Working Paper No. 461, European Central Bank.
- [6] An, S. and F. Schorfheide (2006). "Bayesian Analysis of DSGE Models." *Econometric Reviews* 26, 113-172.
- [7] Andrés, J., P. Burriel, and A. Estrada (2006). "BEMOD: a DSGE Model for the Spanish Economy and the Rest of the Euro Area." *Documento de Trabajo del Banco de España* 0631.
- [8] Aruoba, S.B., J. Fernández-Villaverde, and J. Rubio-Ramírez (2006). "Comparing Solution Methods for Dynamic Equilibrium Economies." *Journal of Economic Dynamics and Control* 30, 2477-2508.
- [9] Arulampalam, A.S., S. Maskell, N. Gordon, and T. Clapp (2002). "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking." *IEEE Transactions on Signal Processing* 50, 174-188.
- [10] Backus, D. K., P. Kehoe, and F. Kydland (1992). "International Real Business Cycles." *Journal of Political Economy* 100, 745-775.
- [11] Backus, D. K., P. Kehoe, and F. Kydland (1995). "International Business Cycles: Theory and Evidence." in Tom Cooley (ed.), *Frontiers of Business Cycle Research*, Princeton University Press.
- [12] Backus, D.K., B.R. Routledge, and S.E. Zin (2005). "Exotic Preferences for Macroeconomists." *NBER Macroeconomics Annual* 2004, 319-390.
- [13] Bansal, R. and A. Yaron (2004). "Risks For The Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance* 59, 1481-1509.
- [14] Barro (2006). "On the Welfare Costs of Consumption Uncertainty." *NBER Working Paper* 12763.
- [15] Becker, G.S. and L. Rayo (2007). "Habits, Peers, and Happiness: An Evolutionary Perspective." *American Economic Review Papers and Proceedings* 97, 487-91.
- [16] Benhabib, J. and R.E. Farmer (1992). "Indeterminacy and Increasing Returns." *Journal of Economic Theory* 63, 19-41.

- [17] Benhabib, J., R. Rogerson, and R. Wright (1991). “Homework in Macroeconomics: Household Production and Aggregate Fluctuations.” *Journal of Political Economy* 99, 1166-87.
- [18] Berger, J.O. R.L. and Wolpert (1988). *The Likelihood Principle* (2nd edition). The Institute of Mathematical Statistics.
- [19] Bernanke, B. S., M. Gertler, and S. Gilchrist (1999). “The Financial Accelerator in a Quantitative Business Cycle Framework.” In J. B. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics* 1, 1341-1393. North-Holland Elsevier.
- [20] Bernardo, J.M. and A.F.M. Smith (2000). *Bayesian Theory*. Wiley Series in Probability and Statistics.
- [21] Bills, M. and P. Klenow (2004). “Some Evidence on the Importance of Sticky Prices.” *Journal of Political Economy* 112, 947-985.
- [22] Bills, M, P. J. Klenow, and B. A. Malin (2008). “Reset Price Inflation and the Impact of Monetary Policy Shocks.” *Mimeo*, University of Rochester.
- [23] Binsbergen, J.H., J. Fernández-Villaverde, R.S.J. Koijen, and J. Rubio-Ramirez (2008). “Working with Epstein-Zin Preferences: Computation and Likelihood Estimation of DSGE Models with Recursive Preferences.” *Mimeo*, University of Pennsylvania.
- [24] Birnbaum, A. (1962). “On the Foundations of Statistical Inference.” *Journal of the American Statistical Association* 57, 269–326.
- [25] Blanchard, O.J. and N. Kiyotaki (1987). “Monopolistic Competition and the Effects of Aggregate Demand,” *American Economic Review* 77, 647-66.
- [26] Browning, M., L.P. Hansen, and J.J. Heckman (1999). “Micro Data and General Equilibrium Models.” in J.B. Taylor & M. Woodford (eds.), *Handbook of Macroeconomics*, volume 1, chapter 8, 543-633. Elsevier.
- [27] Caballero, R.J. and E. M.R.A. Engel (1999). “Explaining Investment Dynamics in U.S. Manufacturing: a Generalized (S,s) Approach.” *Econometrica* 67,783-826.
- [28] Calvo, G. A. (1983). “Staggered prices in a utility-maximizing framework.” *Journal of Monetary Economics* 12, 383-398.
- [29] Campbell, J. Y. (2003). “Consumption-Based Asset Pricing.” In G.M. Constantinides, M. Harris, and R. M. Stulz (eds.), *Handbook of the Economics of Finance* 1, 803-887. North-Holland Elsevier.
- [30] Canova, F. (2007). *Methods for Applied Macroeconomics Research*. Princeton University Press.
- [31] Canova, F., and L. Sala (2006). “Back to Square One: Identification Issues in DSGE Models.” *Mimeo*, Pompeu Fabra University.
- [32] Caplin, A. and J. Leahy (1991). “State-Dependent Pricing and the Dynamics of Money and Output.” *Quarterly Journal of Economics*, 106, 683-708.

- [33] Caplin, A. and J. Leahy (1997). "Aggregation and Optimization with State-Dependent Pricing." *Econometrica* 65, 601-626.
- [34] Caplin, A. and D. Spulber (1987). "Menu Costs and the Neutrality of Money". *Quarterly Journal of Economics* 102, 703-725.
- [35] Chamberlain G. and G.W. Imbens (2003). "Nonparametric Applications of Bayesian Inference." *Journal of Business Economics and Statistics* 21, 12-18.
- [36] Chen, X. H. and S. C. Ludvigson (2008). "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models", Forthcoming in the *Journal of Applied Econometrics*
- [37] Chernozhukov, V., and H. Hong (2003). "A MCMC approach to classical estimation." *Journal of Econometrics* 115(2), p. 293–346.
- [38] Christiano, L.J. (1990). "Linear-Quadratic Approximation and Value-Function Iteration: A Comparison." *Journal of Business Economics and Statistics* 8, 99-113.
- [39] Christiano, L., M. Eichenbaum, and C.L. Evans (2005). "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113, 1-45.
- [40] Christoffel, K, G. Coenen, and A. Warne (2007). "The New Area-Wide Model of the Euro Area: Specification and First Estimation Results." *Mimeo*, European Central Bank.
- [41] Clarida, R., J. Galí, and M. Gertler (2000). "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics* 115, 147-180.
- [42] Cooper, R.W. (2002). "Estimation and Identification of Structural Parameters in the Presence of Multiple Equilibria." *Les Annales d'Economie et Statistique* 6, 1-25.
- [43] Cooper, R. W., J. C. Haltiwanger (2006). "On the Nature of Capital Adjustment Costs." *Review of Economic Studies* 73, 611-633.
- [44] Correia I., J. Neves, and S. Rebelo (1995). "Business Cycles in Small Open Economies." *European Economic Review* 39, 1089-1113.
- [45] DeJong, D.N. and C. Dave (2007). *Structural Macroeconometrics*. Princeton University Press.
- [46] Del Moral P. and J. Jacod (2002), "The Monte-Carlo Method for Filtering with Discrete Time Observations. Central Limit Theorems." in T. J. Lyons and T. S. Salisbury (eds.), *Numerical Methods and Stochastics*. The Fields Institute Communications, American Mathematical Society.
- [47] Dhyne, E., L. J. Álvarez, H. L. Bihan, G. Veronese, D. Dias, J. Hoffmann, N. Jonker, P. Lunnemann, F. Rumler, and J. Vilmunen (2006). "Price Setting in the Euro Area and the United States: Some Facts From Individual Consumer Price Data." *Journal of Economic Perspectives* 20, 171-192.

- [48] Dotsey, M., R.G. King, and A. Wolman (1999). “State Dependent Pricing and the General Equilibrium Dynamics of Money and Output.” *Quarterly Journal of Economics* 114, 655-690.
- [49] Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. Springer Verlag.
- [50] Epstein, L., and S.E. Zin (1989). “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: a Theoretical Framework.” *Econometrica* 57, 937-969.
- [51] Epstein, L., and S.E. Zin (1991). “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: an Empirical Analysis.” *Journal of Political Economy* 99, 263-286.
- [52] Erceg, C.J., L. Guerrieri, and C. Gust (2006). “SIGMA: A New Open Economy Model for Policy Analysis.” *International Journal of Central Banking* 2, 1-50.
- [53] Farmer, R.E. (2007). “Aggregate Demand and Supply.” *NBER Working Paper* 13406.
- [54] Farmer, R.E., D.F. Waggoner, and T. Zha (2006a). “Minimal State Variable Solutions to Markov-Switching Rational Expectations Models.” *Mimeo*, Federal Reserve Bank of Atlanta.
- [55] Farmer, R.E., D.F. Waggoner, and T. Zha (2006b). “Indeterminacy in a Forward Looking Regime Switching Model.” *Mimeo*, Federal Reserve Bank of Atlanta.
- [56] Fernández-Villaverde, J. and J. Rubio-Ramírez (2004). “Comparing Dynamic Equilibrium Models to Data: a Bayesian Approach.” *Journal of Econometrics*, 123, 153-187.
- [57] Fernández-Villaverde, J. and J. Rubio-Ramírez (2005). “Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood.” *Journal of Applied Econometrics*, 20, 891-910.
- [58] Fernandez-Villaverde, J. and J. Rubio-Ramirez (2006). “Solving DSGE Models with Perturbation Methods and a Change of Variables.” *Journal of Economic Dynamics and Control* 30, 2509-2531.
- [59] Fernández-Villaverde, J. and J. Rubio-Ramírez (2007). “Estimating Macroeconomic Models: A Likelihood Approach.” *Review of Economic Studies* 74, 1059-1087.
- [60] Fernandez-Villaverde, J. and J. Rubio-Ramirez (2008). “How Structural are Structural Parameters?” *NBER Macroeconomics Annual* 2007, 83-137.
- [61] Fernández-Villaverde, J., J. Rubio-Ramírez, and M.S. Santos (2006). “Convergence Properties of the Likelihood of Computed Dynamic Models.” *Econometrica* 74, 93-119.
- [62] Fernández-Villaverde, J., P. Guerrón-Quintana, and J. Rubio-Ramírez (2008). “The New Macroeconometrics: A Bayesian Approach.” in A. O’Hagan and M. West (eds.) *Handbook of Applied Bayesian Analysis*, Oxford University Press.
- [63] Fisher, J., (2006). “The Dynamic Effects of Neutral and Investment-Specific Technology Shocks.” *Journal of Political Economy* 114, 413-52.

- [64] Freedman, D. A. (1963). “On the Asymptotic Behavior of Bayes Estimates in the Discrete Case.” *Annals Mathematical Statistics* 34, 1386-1403.
- [65] Friedman, M. and A.J. Schwartz (1971). *A Monetary History of the United States, 1867-1960*. Princeton University Press.
- [66] Galeson, D. (2007). *Old Masters and Young Geniuses: The Two Life Cycles of Artistic Creativity*. Princeton University Press.
- [67] Geweke, J.F. (1993). “Bayesian Treatment of the Independent Student-t Linear Model.” *Journal of Applied Econometrics* 1993, 8, S19-S40.
- [68] Geweke, J.F. (1994). “Priors for Macroeconomic Time Series and Their Application.” *Econometric Theory* 10, 609-632.
- [69] Ghosh, J.K. and R. V. Ramamoorthi (2003). *Bayesian Nonparametrics*. Springer Verlag.
- [70] Griffiths, T. L., and J.B. Tenenbaum (2006). “Optimal Predictions in Everyday Cognition.” *Psychological Science* 17, 767-773.
- [71] Greenwood, J, Z. Hercowitz, and P. Krusell (1997). “Long-Run Implications of Investment-Specific Technological Change.” *American Economic Review* 87, 342-362.
- [72] Greenwood, J, Z. Hercowitz, and P. Krusell (2000). “The Role of Investment-Specific Technological Change in the Business Cycle.” *European Economic Review* 44, 91-115.
- [73] Guerrón-Quintana P. (2008). “What you match does matter: The Effects of Data on DSGE Estimation.” *Mimeo*, North Carolina State University.
- [74] Hall, R. (1997). “Macroeconomic Fluctuations and the Allocation of Time.” *Journal of Labor Economics* 15, S223-250.
- [75] Hansen, L. P. and R. Jagannathan (1991). “Implications of Security Market Data for Models of Dynamic Economies” *Journal of Political Economy* 99, 225-262.
- [76] Hansen, L.P., J. Heaton, J. Lee, and N. Roussanov (2007). “Intertemporal Substitution and Risk Aversion.” In J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics* 6, 3967-4056. North Holland-Elsevier.
- [77] Hansen, L.P. and T.J. Sargent (2007). *Robustness*. Princeton University Press.
- [78] Harrison, R., K. Nikolov, M. Quinn, G. Ramsay, A. Scott, and R. Thomas (2005). “The Bank of England Quarterly Model.” Bank of England.
- [79] Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- [80] Heckman, J. and B. Singer (1984). “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data.” *Econometrica* 68, 839-874.
- [81] Jovanovic, B. (1989). “Observable Implications of Models with Multiplicity of Equilibria.” *Econometrica* 57, 1431-1436.

- [82] Judd, K.L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge.
- [83] Judd, K.L. and S.M. Guu (1993). "Perturbation Solution Methods for Economic Growth Model." In Varian, H. (Ed.), *Economic and Financial Modelling in Mathematica*. Springer-Verlag.
- [84] Justiniano A. and G.E. Primiceri (2008). "The Time Varying Volatility of Macroeconomic Fluctuations." *American Economic Review* 98, 604-641.
- [85] Keynes, J.M. (1936). *The General Theory of Employment, Interest, and Money*. MacMillan.
- [86] Kilponen, J. and A. Ripatti (2006). "Introduction to AINO." *Mimeo*, Bank of Finland.
- [87] Kim, J. (1998). "Large Sample Properties of Posterior Densities, Bayesian Information Criterion and the Likelihood Principle in Nonstationary Time Series Models", *Econometrica* 66, 359-380.
- [88] Kim, C. and C.R. Nelson (1998). "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle". *Review of Economics and Statistics* 81, 608-616.
- [89] Kimball, M.S. (2002). *Real Business Cycle Theory: a Semiparametric Approach*. *Mimeo*, University of Michigan.
- [90] King, R.G., C.I. Plosser, and S.T. Rebelo (2002). "Production, Growth and Business Cycles: Technical Appendix." *Computational Economics* 20, 87-116.
- [91] King, R.G., C.I. Plosser, J.H. Stock, and M.W. Watson (1991). "Stochastic Trends and Economic Fluctuations." *American Economic Review* 81, 819-40.
- [92] Kitamura, Y. and M. Stutzer (1997). "An Information-theoretic Alternative to Generalized Method of Moment Estimation". *Econometrica* 65, 861-874.
- [93] Klenow, P. J., and O. Kryvtsov (2008). "State-Dependent or Time-Dependent Pricing: Does it Matter for Recent US Inflation?" *Quarterly Journal of Economics* 123, 863-904.
- [94] Kortelainen, M. (2002). "Edge: A Model of the Euro Area with Applications to Monetary Policy." *Bank of Finland Studies* E:23-2002, Bank of Finland.
- [95] Künsch, H.R. (2005). "Recursive Monte Carlo Filters: Algorithms and Theoretical Analysis." *Annals of Statistics* 33, 1983-2021.
- [96] Kydland, F. E. and E. C. Prescott (1982). "Time to Build and Aggregate Fluctuations." *Econometrica* 50, 1345-1370.
- [97] Lagos, R. and R. Wright (2005). "A Unified Framework for Monetary Theory and Monetary Analysis". *Journal of Political Economy*, 113, 463-484.
- [98] Lubick, T. and F. Schorfheide (2004). "Testing for Indeterminacy: An Application to U.S. Monetary Policy." *American Economic Review* 94, 190-217.
- [99] Lucas, R., (1972). "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4(2),103-124.

- [100] Mankiw, G.N. and R. Reis (2002). “Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve.” *Quarterly Journal of Economics* 117, 1295-1328.
- [101] Manski, C.F. (1999). *Identification Problems in the Social Sciences*. Harvard University Press.
- [102] McConnell, M.M. and G. Pérez-Quirós (2000). “Output Fluctuations in the United States: What Has Changed Since the Early 1980’s?” *American Economic Review* 90, 1464-1476.
- [103] McFadden, D.L. (1989). “A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration.” *Econometrica* 57, 995-1026.
- [104] Mehra, R. and E. C. Prescott, (1985). “The Equity Premium: A Puzzle.” *Journal of Monetary Economics* 15, 145-161.
- [105] Mendoza, E. (1991). “Real Business Cycles in a Small Open Economy.” *American Economic Review* 81, 797-818.
- [106] Mendoza, E. (1995). “The Terms of Trade, the Real Exchange Rate, and Economic Fluctuations.” *International Economic Review* 36, 101-37.
- [107] Mengersen, K.L., Robert, C.P. and Guihenneuc-Jouyaux, C. (1999). “MCMC Convergence Diagnostics: a ‘reviewww’.” In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.), *Bayesian Statistics 6*, Oxford Sciences Publications.
- [108] Monfort, A. (1996). “A Reappraisal of Misspecified Econometric Models.” *Econometric Theory* 12, 597-619.
- [109] Murchison, S. and A. Rennison (2006). “ToTEM: The Bank of Canada’s New Canadian Projection Model.” *Bank of Canada Technical Report*, Bank of Canada.
- [110] Nakamura, E. and J. Steinsson (2008). “Five Facts About Prices: A Reevaluation of Menu Cost Models.” *Quarterly Journal of Economics* 123, 1415-1464.
- [111] O’Hagan, A. and M. West (eds.) (2009). *Handbook of Applied Bayesian Analysis*. Oxford University Press.
- [112] Orphanides, A. (2002). “Monetary Policy Rules and the Great Inflation.” *American Economic Review* 92, 115-120.
- [113] Owen, A.R. (2001). *Empirical Likelihood*. Chapman & Hall/CRC.
- [114] Pakes, A. and D. Pollard (1989). “Simulation and the Asymptotics of Optimization Estimators”. *Econometrica* 57, 1027-1057.
- [115] Phillips, P.C.B. and W. Ploberger (1996). “An Asymptotic Theory of Bayesian Inference for Time Series”. *Econometrica* 64, 381-412.
- [116] Ragusa, G. (2006). “Bayesian Likelihoods for Moment Condition Models.” *Mimeo*, UC-Irvine.
- [117] Rissanen, J. (1986). “Stochastic Complexity and Modeling”. *The Annales of Statistics* 14, 1080-1100.

- [118] Robert, C.P. (2001), *The Bayesian Choice* (2nd edition), Springer Verlag.
- [119] Robert, C.P. and G. Casella (2005). *Monte Carlo Statistical Methods* (2nd edition). Springer Verlag.
- [120] Roberts, G., A. Gelman, and W. Gilks (1997). “Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms.” *Annals of Applied Probability* 7, 110-120.
- [121] Robins, J.M. and Y. Ritov (1997). “Toward a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-Parametric Models.” *Statistics in Medicine* 16, 285–319.
- [122] Rogerson, R. and J. Wallenius (2007). “Micro and Macro Elasticities in a Life Cycle Model with Taxes.” *NBER Working Paper* 13017.
- [123] Rouwenhust, K.G. (1995). “Asset Pricing Implications of Equilibrium Business Cycle Models.” In Cooley, T. F. (ed.), *Frontiers of Business Cycle Research*. Princeton University Press.
- [124] Rubin, D.B. (1988). “Using the SIR Algorithm to Simulate Posterior Distributions”. in J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (eds), *Bayesian Statistics 3*, 395-402, Oxford University Press.
- [125] Schennach, S.M. (2005). “Bayesian Exponentially Tilted Empirical Likelihood.” *Biometrika* 92, 31-46.
- [126] Schmitt-Grohé, S., Uribe, M. (2004). “Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function.” *Journal of Economic Dynamics and Control* 28, 755-775.
- [127] Sims, C.A. (2002). “Implications of Rational Inattention.” *Mimeo*, Princeton University.
- [128] Sims, C. A. and H. Uhlig (1991). “Understanding Unit Rooters: A Helicopter Tour.” *Econometrica* 59, 1591-1599.
- [129] Sims, C.A. and T. Zha (2006). “Were There Regime Switches in U.S. Monetary Policy?” *American Economic Review* vol 96, 54-81.
- [130] Smets, F. and R. Wouters (2007). “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *American Economic Review* 97, 586-606.
- [131] Stengel, R.F. (1994). “Optimal Control and Estimation.” Dover Publications.
- [132] Stock, J.H. and M.W. Watson (2003). “Has the Business Cycle Changed, and Why?” *NBER Macroeconomics Annual* 2002, 159-218.
- [133] Stokey, N.L., R.E. Lucas, and E.C. Prescott (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.
- [134] Wallace, N. (2001). “Whither Monetary Economics?”. *International Economic Review* 42, 847-869.
- [135] Weil, P. (1989). “The Equity Premium Puzzle and the Risk-Free Rate Puzzle.” *Journal of Monetary Economics* 24, 401-421.

- [136] Woodford, M.D. (2003). *Interest and Prices*. Princeton University Press.
- [137] Young, (2005). “One of the Things that We Know that Ain’t So: Why U.S. Labor Share is Not Relatively Stable.” *Mimeo*, University of Mississippi.
- [138] Zellner, A. (1988). “Optimal Information Processing and Bayes’ Theorem”, *American Statistician* 42, 278-284.

Figure 1: Posterior Distribution, Smets-Wouters Priors

