Université Catholique de Louvain
Faculté des sciences économiques, sociales et politiques
Département des Sciences Economiques

Essays on Formulating and Estimating DSGE Models

Giulio Nicoletti

Thèse présentée en vue de l'obtention du grade de

Docteur en Sciences Economiques et de Gestion

Louvain la Neuve
Mars 2010

Composition du jury:

Raouf Boucekkine (*Université catholique de Louvain*) (Promoteur)
Luc Bauwens (*Université catholique de Louvain*)
Michel Juillard (*CEPREMAP, Paris*)
Henri Sneessens (*Université catholique de Louvain*)
Raf Wouters (*Banque Nationale de Belgique*)

Ai miei genitori, con affetto e gratitudine.

Acknowledgements

This list is long but probably not exhaustive. I was lucky enough to get over the years more good friends and 'supporters' than I actually (probably) deserved.

A big thank goes to the Bank of Italy lunch crew, a handful of nice and brilliant people I am used to going to lunch with. Among those people, let me name the name of closest friends: Enrico (the Wise) and Claudia (the Quest for Perfection, included in this list 'ad honorem'); Juri (The big man with an even bigger heart) and Francesca (the same ... only relating to the heart), to Giordano (the Master of Diplomacy) and to Gianluca (The Companion). I could always completely rely on these friends for discussing both research and private life matters; briefly I shared with them the most important parts and moments of my recent life. Juri and Francesca even managed to take me to a gym and make me like it; Claudia brought me to the Outlet for shopping.

A thanks goes to my coauthors, Marco; Gianluca; Olivier. Gianluca was my companion as we started working together at the Bank of Italy, at the very same office. I learned from him almost all nice tricks and methods I know about good Matlab programming, some refinements of which I got lately from Claudia Miani and Fabrizio Venditti. Thanks also to Hania for her patience while me and Gianluca were working on the paper and running Matlab codes until (probably) later than due. Hania revised the English of the joint paper with Gianluca (now the third chapter of this script), thanks. Music, beer and the ciuri ciuri granitas. When shall I start Salsa lessons ?

Fabio Busetti immediately set the example of hard and effective working in both institutional and research work at the Bank. I bothered Michele Caivano with an innumerable amount of (sometimes stupid) questions about the Bayesian Econometrics of dynamic models when I was a rookie on the topic. My grati-

tude to Mike goes clearly beyond the econometrics.

From Louvain la Neuve my flatmates, Alessandro and Gabriele, and the 'big brother crew' in LLN. Besides the economics, Gabriele introduced me to the zeneise language and to the music of Fabrizio De Andre' (something I later shared and improved with Claudia and Enrico). With them and Oscar Amerighi, Augusta Badriotti, Mario Denni, Filomena Garcia, Luca Pensieroso, Andrea Silvestrini, Antonio Tesoriere, Cecilia Vergari, we enjoy friendship and still endless discussions concerning the state of the economy and the fate of the System we live in. Luca helped me in tempering some sharp and rather extreme attitude of mine while I was still green. In those early days I started my path on DSGE modelling with Pierre Goffinet, a very close friend since then. We taught DEA students together to run nicely Dynare with several different DSGE models: at that time we only had version 2. Thanks also to Marisol and best wishes to, in chronological order, little Elisa and Alexandra.

Alessandra 'the Doc' Camboni-Pensieroso helped me out a (potentially very serious) medical situation in 2008. To her goes my thanking and my admiration as the strongest woman I know (and let me say bravo Luca !).

A thank goes to Axel Gautier who, while I was a complete stranger to him, practically exempted me from teaching one course (out of five) at UCL when I had to leave for my visiting at SSE. To my 'swedish' friends, Alessandra and Silvia; Virginia; the 'great Faustinho'.

I thank all the jury members and my very kind supervisor Raouf for comments, feedbacks and guidance. Both Raf Wouters and Michel Juillard were extremely kind and always available for discussions concerning my research also when I visited them, respectively at the Bank of Belgium and CEPREMAP (an unforgettable month in Paris). This helped a lot. The visiting at the NBB was undertaken within the 'External Work Experience' program for the Central Banks in the Eurosystem, also gratefully acknowledged. For this experience my thank goes to Stefano 'the Boss' Siviero, Salvatore Rossi first and Fabio Panetta later who allowed me to go there. Stefano was also quite 'patient' for what concerns joint research projects during the time I was writing this thesis. For the EWE my thank also goes to the people at the Bank of Italy Human Resources. At the NBB I got a really perfect environment thanks also to (besides Raf Wouters)

Reuben, Engin and Maarten.

Finally, Luc Bauwens's comments helped in improving the rigour of the work.

As I left Rome my family provided me love and moral support. From the financial side I was never a burden for them: my gratitude for that goes to the Fellowship from the Ente Einaudi (now EIEF) and then to a grant from La Sapienza University. In my third year in Belgium I worked as assistant at UCL. For this very important experience I thank Raouf, Henri Sneessens and, remarkably, Bruno Van Der Linden.

For the teaching and the help with LaTeX, both typing and style, I thank Luca Missori; all remaining errors are mine.

While I was abroad and always Andrea has been like a brother to me in both good and hard times. An enlarged family was set with Cocco, Marta and Sofia. Thanks guys from the bottom of my heart.

Concerning an extended notion of family, the time spent with Alfie and Bettina taught me much and over several years, about food, wine and most importantly, life. Bettina helped me correcting the English of some of this thesis drafts, all remaining mistakes are mine.

As working papers parts of this thesis were presented in conferences; it would be long to mention all feedbacks. Let me just mention here that Stefano Neri is always the right guy to discuss with here at the Bank. Also, the 'Gandolfians', Marianna Belloc, Luca Colantoni, Giorgione Lagrasta and, particularly, Sergio Santoro. Sergione, will we succeed in writing this famous paper together ?

Besides all the people mentioned above, Claudia Mura and Nicola Napoli helped in making 2009 a much better year for me than 2008.

A last thank goes to the Belgian Health care system (in particular the St. Luc Hospital) which solved a difficult elbow problem of mine in 2003 (after more than two years of failed attempts in many other places) at practically no expense. Giordano Mion took me home from hospital by car. Thanks Gio'.

Preface

[...] We are such stuff
As dreams are made on; and our little life
Is rounded with a sleep.

William Shakespeare
The Tempest, Act 4, scene 1

It was a long journey. The quest started in the fall 2002 from Louvain la Neuve. In the year of the Master I was living in a *kot* with 4 belgian students; my place was so messy that I earned the nickname of *entropy*, still in place with my friends. After touching lands and sees, and a visiting at Stockholm School of Economics of 6 months, I landed back to Rome in late 2005. The Journey was clearly not completed yet.

My house is still messy, another proof that human knowledge is finite.

The great part of this journey was (and still is today) the company and friendship of the many young, talented and passionate people I met. They come from many different places and sometimes disciplines, maybe a proof that even if economics can be much self-referential, some economists might not be as such. The related beautiful part are those who have been friends from before and all along this journey. I name some of them in the acknowledgments but I thank them all.

As I started working at the Bank of Italy in early 2006 my main research focus became the econometric estimation of DSGE models, the motivated reader might directly jump to chapters 3 and 4. I give a rather non-technical introduction of what I do in the first chapter, the Introduction.

The journey was in the end much more demanding intellectually than from a traveling point of view. My thesis advisor, Raouf, was patient about that and I thank him (also) for that.

As I lived the recent recession at a policy institution my focus has probably shifted again to more urgent problems in economics. We shall see whether fruits are ripe for that in the years to come.

Rome 23rd February 2010

Contents

Chapter 1

Introduction

Progress, don't regress

Edward C. Prescott

From the seminal work of Kydland and Prescott (1982) a huge amount of effort and work has been devoted in order to improve our understanding of how an economy works. After the breakthrough of general equilibrium analysis, macroeconomics rapidly became a mathematical science based on the paradigm that economies can be described as intertemporal equilibria.

A Dynamic Stochastic General Equilibrium model (DSGE) describes the over time evolution of an economy as function of primitive objects, so called *deep parameters*, which relate to preference, technology and endowments of economic agents, namely firms, households and government. It is a Stochastic model as the source of economic fluctuations is provided by random shocks. Those latter can be measured and named, but their ultimate explanation is left out of the model. It is Dynamic General Equilibrium since agents take optimal intertemporal decisions (under the constraints they face) which are mutually consistent, without being necessarily optimal from a social welfare point of view.

This thesis is an attempt to add one further page to the book of knowledge of DSGE models. In particular, in each chapter I treat a different issue concerning the link between exogenous shocks and macroeconomic variables.

In the first chapter I answer the following: 'How does credit relate to the transmission mechanism of shocks into macroeconomic variables ?' We intro-

duce liquidity frictions into a standard Real Business Cycle model. While standard models assume that business projects can immediately find financing, in our set-up we introduce a lengthy search process for potential Entrepreneurs to match with loans from financial intermediaries. Every period spent on the search process entails a cost. When a match occurs the Entrepreneur can start a business and she is willing to give up part of her profit to the intermediary in exchange for not getting back into the search process. This is a kind of *liquidity premium* which is paid by active firms on top of the risk-free rate: when credit markets are tight, meaning that Entrepreneurs are abundant compared to available loans, the premium is higher. We show that when labor markets are perfect, the effects of our liquidity premium on economic dynamics are small. Instead, when also labor markets are subject to frictions, i.e. workers and firms are in another search process to find one another, credit frictions can either accelerate or reduce the impact of the stochastic shocks hitting the economy, depending on the value of deep parameters. For realistic calibrations of the model, based one US data, an accelerating effect seems to be in place.

The second chapter deals with the econometric estimation of DSGE models. Earlier contributions used simple models as small laboratory experiments which could be calibrated in order to reproduce relevant features of an economy. As the discipline evolved, models were taken to the data by the means of more formal econometric techniques. Our second question is as follows: 'Provided the generating process of shocks is unknown to researchers what is the best way to bring a model to the data ?'. This work relaxes one of the common assumption used in the literature, that is the statistical form of exogenous shocks is known to the researcher. We extend Kalman Filter Maximum Likelihood methods to cope with the problem at hand. We report better estimates of the deep parameters of the economy (i.e. more in line with those obtained by using other sources of data) and a large and statistically significant improvement in out-of-sample forecast of macroeconomic variables. We also provide empirical reasons to believe that long memory can be a relevant issue concerning DSGE models, a feature much overlooked in the literature. The ingenuity of our method is that it is an effective way of filtering out from the data the amount of persistence which would be left unexplained by the model.

In the last chapter I take a Bayesian point of view to estimating DSGE models. A bayesian researcher uses not only the set of data at hand, but also his prior knowledge. The contribution of the chapter is twofold. On the one hand we investigate some earlier problematic results, namely those of Negro and Schorfheide (2008). A critical issue in their procedure is fully exposed and some ways to solve it are outlined. Then we propose a simple way of eliciting priors, which is to use information concerning so called impulse response functions. Impulse responses are the responses of macroeconomic variables, conditional on some exogenous shock occurring in the economy. Those are the most studied object in modern macroeconomics and they constitute an ideal source of prior knowledge. Some implications of the use of impulse responses as prior knowledge are explored in this chapter.

The technical evolution of the chapters (from calibration to Bayesian Estimation) reflects the evolution of the literature on DSGE models and also the learning process of the author himself.

Chapter 2

Capital Market Frictions and the Business Cycle

2.1 Introduction

Liquidity is one of the hot topics emerging from the (ongoing) 2008 financial crisis. Liquid capital markets reduce the time it takes for capital supplied by households to be matched with available projects in the economy in a relevant way for aggregate dynamics. Earlier literature, such as Eisfeldt and Rampini (2006), have related the idea of liquidity to turnover in the market for used capital and to the ability to make old capital goods useful for different and new purposes. More recently Lagos and Rocheteau (2009) tackle the issue of liquidity in financial markets by modelling over the counter (OTC) transactions as a search process between buyers and sellers: in such a set up the average amount of time for a trade to occur provides a natural measure of how liquid is a financial system. On their thread we propose a model which features a matching process between capital supply, available from households, and firms projects: 'credit market tightness' is defined as the ratio between projects to be funded and loans made idle by households in the economy.

In particular, in our paper we study the linkage between capital markets and labor market dynamics. Following Wasmer and Weil (2004) (WW heretofore) capital and labor are modeled as complementary parts of a matching process in

order for a firm to be able to produce. In a first stage, firms need to match with capital, then they look up for workers. Finally, production takes place and firms pay back their loans in the form of an interest rate until they are destroyed. In this environment credit and labor markets interact: firms are not free to enter in the labor market until they are matched in the capital market, and in turn the profitability of households investments depends on those conditions prevailing in the labor markets. Structural parameters of both credit and labor markets play an important role in either bolster or reduce liquidity in capital markets following a positive exogenous shock. A drop in credit market tightness is associated with a stronger reaction of labor market, a kind of financial accelerator, the reverse when credit market tightness rises after a shock.

In our set-up the supply of capital from households is intermediated by a representative intermediary, generically called 'a bank', which is able to fully diversify the credit risk of different projects but it has to undertake a costly search process in order to find projects. Our set-up of credit frictions differs from the dominant model of agency costs (e.g. the financial accelerator by Bernanke, Gertler, and Gilchrist (1999)) as already applied to labor matching models where firms take up external finance (e.g. standard debt) in order to pay for vacancy posting, (see Petrosky-Nadeau (2009)). The set-up is not meant to be taken literally: real banks do not use households deposits to enter into firms capital but they rather provide debt. Still, by adopting an environment in which there is no difference between debt and equity (e.g. the Modigliani-Miller theorem holds), we can focus only on liquidity premia generated by search frictions.

Our model draws on the steady state analysis of WW, but, differently from them, we develop a dynamic general equilibrium model, where all relative prices are endogenous. To close the model, we assume there are two types of firms: small intermediate producers which are labor intensive and subject to the type of matching process described above and large final producers. The first type of firms uses one unit of labor and one unit of capital to produce, and is subject to an exogenous destruction rate every period. The large type firm buys the intermediate products and adds risk-free capital to it as according to a standard Cobb-Douglas production function. Household investments are divided into two types: one is standard capital, lent directly to large firms; the second

is risky capital which is lent to small intermediate firm through the intermediation of banks. Capital frictions apply in the relationship between banks and small firms, while labor frictions apply to small firms and workers. In the spirit of the RBC literature, we provide a developed statistical analysis, by studying the impact of technology shocks on the simulated economy.

Our environment relates to some recent research concerning matching models. First, the free entry condition for posting a vacancy is modified with respect to a standard labor market matching model. The incentive for a firm to post an additional vacancy is not constant but it is subject to conditions prevailing on the financial market: when credit market tightness drops the implicit cost of posting a vacancy also drops, amplificating thus the shock. When financial market tightness rises following a shock, the nexus is reversed and the impact of the shock is buffeted. Second, imperfections in the setting of prices in one market (e.g. pricing norms either in wages or in the rental rates of capital) tend to produce a spillover also in the other market. We discuss this point in our framework and we relate that to the literature on real wage rigidity (for example to Shimer (2009)).

The rest of the paper is organized as follows. We first present a simplified model of credit frictions modelled as search processes in section 2.2; we introduce our full model in 2.3; the determination of prices and the definition of equilibrium is provided in 2.4. The specification, adopted functional forms and the calibration of the model are discussed in 2.5. Some steady state properties of the model are discussed in 2.6 and dynamics in 2.6.1, leaving some technicalities to the appendix. In 2.7 we discuss the role of different pricing schemes in our economy. Section 2.8 concludes.

2.2 A bare-bone model of search credit frictions

In this section we outline a simplified set-up where no labor market is assumed, we add it in the next section. The aim of the section is to provide an outline of how a liquidity premium is generated by search frictions.

There are three agents in our (simplified) economy: small firms, households

and banks. Small firms post projects E_t and they must look up for financing by households by paying a per period search c. Supply of capital is intermediated by a representative agent, generically called 'a bank', which issues deposits to households and posts loans L_t with a search cost k. The so called credit market tightness is given by $\phi_t \equiv E_t/L_t$ and as usual in the matching literature the number of matches in our economy is equal to $H(L_t, E_t)$, a matching function of L_t, E_t. The probabilities for a single firm and for a loan to be matched are given by, respectively, to $p^F = \mathcal{H}(L_t, E_t)/E_t$ and $p^B = \mathcal{H}(L_t, E_t)/L_t$.

The value functions for small firms, being in the matched or unmatched state, denoted respectively by W_t^1 and W_t^2:[1]

$$W_t^1 = -c + \beta p^F(\phi_t) E_t W_{t+1}^2 + \beta(1 - p^F(\phi_t)) E_t W_{t+1}^1, \tag{2.1}$$

$$W_t^2 = d_t - \rho_t + \beta(1-s) E_t W_{t+1}^2, \tag{2.2}$$

d_t is the price of small firm products, exogenous in this simplified set-up. ρ_t is the rental price of capital, which is corresponded by active small firms to the intermediary. We assume that firms have free entry in the market for capital; this brings the value of an unmatched firm W_t^1 to zero in each period; from (2.2), this entails that the value of a matched firm is simply given by the average cost of searching:

$$E_t W_{t+1}^2 = \frac{c}{\beta p^F(\phi_t)}, \tag{2.3}$$

The number of active firms N_t increases according to the law of motion:

$$N_{t+1} = (1-s) N_t + p^B(\phi_t) L_t,$$

where s is the (exogenous) destruction rate for firms.

A bank has to choose the number of loans every period t, each entailing a search cost k; funds are collected by the means of deposits from households. In equilibrium, deposits are remunerated at a rate r_t^d given by the risk free interest

[1] In this economy there are only two states which are relevant for small firms, being matched (state i = 2) or unmatched (state i = 1). The arguments of the small firms value functions should be prices (when relevant) and the probabilities of being matched, which firms take as given in the matching framework. To ease notation we omit prices and probabilities in the arguments of all value functions. Then small firms value functions are denotes simply as W_t^i.

rate r_t in the economy, augmented for the risk of destruction of small firms s. With this respect banks are a modeling device to separate premia on default risk (s) from the liquidity premium.

The first order condition for loans equates the marginal costs and benefits of loan posting for the bank:

$$k = \beta p^B(\phi_t) E_t \frac{\partial W^B(N_{t+1})}{\partial N_{t+1}}. \tag{2.4}$$

A loan brings an expected additional value for the bank equal to $p^B(\phi_t)\frac{\partial W^B(N_{t+1})}{\partial N_{t+1}}$, where the last term is the derivative of the value function for the bank with respect to the stock of small firms. The derivative evolves as a function of the rental rate perceived from active firms minus the rate of return on deposits to be paid to households:

$$\frac{\partial W^B(N_t)}{\partial N_t} = \rho_t - (r_t + s) + \beta E_t \left[(1-s) \frac{\partial W^B(N_{t+1})}{\partial N_{t+1}} \right], \tag{2.5}$$

the marginal value in the present period is equal to the rent paid by the small firm -a bargained yield ρ_t, minus the cost of paying rates on one more deposit - plus the expected value of having one more matched unit next period, net of destruction rate s.

In the basic set-up, the rental rate ρ is computed from the Nash Bargaining solution from the two agents, the small firm and the bank:

$$\rho_t \equiv \arg\max_{\rho_t} \left(\frac{\partial W^B(N_t)}{\partial N_t} \right)^{\zeta} (W_t^2 - W_t^1)^{1-\zeta},$$

This provides a relation between $\frac{\partial W^B(N_t)}{\partial N_t}$ and W_t^2 as follows:

$$\zeta \frac{\partial W^B(N_t)}{\partial N_t} = (1-\zeta)(W_t^2 - W_t^1). \tag{2.6}$$

The bargained rental rate can be computed from (2.6) by plugging (2.2) and (2.5) respectively on its left and right hand side (after using the free entry condition),

and using (2.3) and (2.4) to substitute out the resulting expectation terms:

$$\rho_t = \zeta \left(d_t + \frac{(1-s)k}{p^B(\phi_t)} \right) + (1-\zeta) \left((r_t + s) + \frac{c(1-s)}{p^F(\phi_t)} \right),$$

The terms $\frac{c(1-s)}{p^F(\phi_t)}$ and $\frac{k(1-s)}{p^B(\phi_t)}$ evaluate the expected cost of waiting for the vacancy to be filled respectively by a bank and by a firm; this varies with credit market tightness and can be thought as a measure of the *liquidity premium*.

The main shortcoming of this simple model is that in equilibrium credit market tightness (and liquidity premia) are constant and given by:[2]

$$\phi_t \equiv \frac{E_t}{L_t} = \frac{k(1-\zeta)}{c\zeta}.$$

Kurmann and Petrosky-Nadeau (2007) adopt a model similar to the one sketched above where the only transmission mechanism is given by the fact that firms are allowed to choose labor as in a Walrasian market. As they show, this provides no acceleration mechanism; this is because they depart too little from the basic set-up highlighted above. In the rest of the paper we show how the basic set-up changes when search labor market frictions come into the picture and they are complemented with real wage rigidities, as discussed by Blanchard and Gali (2007) and Hall (2005). We then extend the discussion to pricing norms for capital rents.

2.3 Adding labor market frictions

In the complete model there are four agents: small firms, banks, large firms and households. We introduce large firms mainly to close the model. Large firms produce output and demand small firms products; by the same token they constitute one of the assets for households to invest into (we call it 'safe capital' with a little abuse of terminology in the remainder of the analysis). The

[2]Move (2.6) one period forward, take its expectation and use equations (2.4) and (2.5) to substitute the terms.

large firms maximization problem is static by nature:

$$\Pi(N_t, K_t) = \max_{N_t, K_t} \{\varepsilon_t \mathcal{F}(K_t, N_t) - d_t N_t - (r_t + \delta) K_t, \} \tag{2.7}$$

where ε_t is an aggregate productivity shock, $\mathcal{F} = K_t^{\mu} N_t^{1-\mu}$ is a Cobb-Douglas production function, N_t is the amount of intermediate goods (or equivalently the amount of small firms producing, or equivalently the amount of employed workers), K_t is the capital stock, r_t is the capital interest rate paid to the households, δ is the depreciation rate of capital. Maximizing (2.7) with respect to N_t and K_t gives the expression for the price d_t and the rental rate of 'safe' capital:

$$\varepsilon_t \frac{\partial \mathcal{F}(K_t, N_t)}{\partial N_t} = d_t, \tag{2.8}$$

$$\varepsilon_t \frac{\partial \mathcal{F}(K_t, N_t)}{\partial K_t} = r_t + \delta. \tag{2.9}$$

In the extended set-up both labor and capital are subject to matching frictions: N_t denotes both the stock of small firms and the employment level and $V_t + N_t$ is the capital level used by small firms; unemployment in our simple framework is given by $U_t = (1 - N_t)$, since total labor force is normalized to one. Every period the number of matches in the labor market are given by a matching function similar to the one previously introduced as $M_t = \mathcal{M}(V_t, 1 - N_t)$, where $\mathcal{M}$ is a matching function, increasing in its arguments and satisfying constant returns to scale. The probability q_t^F for a firm to find a worker and the probability q_t^H for a worker to find a firm are:

$$q_t^F = \frac{\mathcal{M}(V_t, 1 - N_t)}{V_t} \qquad \text{and} \qquad q_t^H = \frac{\mathcal{M}(V_t, 1 - N_t)}{1 - N_t}. \tag{2.10}$$

Concerning credit, we follow the presentation highlighted above. The probability p_t^B for a bank to find a firm and the probability p_t^F for a firm to find a bank are again:

$$p_t^B = \frac{\mathcal{H}(L_t, E_t)}{L_t} \qquad \text{and} \qquad p_t^F = \frac{\mathcal{H}(L_t, E_t)}{E_t}, \tag{2.11}$$

and labor market and credit tightness are defined as

$$\theta \equiv \frac{V}{U};\ \ \phi \equiv \frac{E}{L},$$

and, under the assumption of the matching function being homogenous of degree one, probabilities can be rewritten as a function of their market tightness by:

$$q^f = g(\theta);\ q^h = \theta g(\theta);\ p^f = f(\phi);\ p^B = \phi f(\phi),$$

with the following standard properties: $\partial g(\theta)/\partial\theta < 0$; $\partial\theta g(\theta)/\partial\theta > 0$ and symmetric for the credit matching function. Some more technical assumptions are introduced in section 2.5.

It is interesting to note that in the extended set-up both the number of vacancies and small firms are state variables; we assume that capital and labor matches are respectively destroyed with a common exogenous probability s,[3] the dynamics between the different states is therefore given by:[4]

$$V_{t+1} = (1-s)V_t - \mathcal{M}(V_t, 1-N_t) + \mathcal{H}(L_t, E_t), \tag{2.12}$$

$$N_{t+1} = (1-s)N_t + \mathcal{M}(V_t, 1-N_t), \tag{2.13}$$

A small firm can be in three different states rather than two as before:

- *state 1*: firms searching for a bank able to provide one unit of capital (E_t projects)
- *state 2*: firms with the unit of capital searching for one worker (V_t vacancies)
- *state 3*: firms with one unit of capital and one worker, producing one intermediate good (priced d_t) and paying back a capital rent to the bank and

[3] Shimer (2005) shows on US data that the separation probability is nearly acyclical, particularly during the last decades. Hall (2005) also emphasizes that for the past 50 years in the US, the separation rate is nearly constant while the job-finding rate shows high volatility at business cycle.

[4] The assumption of a common destruction rate of jobs and firms, might be seen as restrictive, but it is more consistent with the framework one firm, one job; we provide some more details about that in our calibration section (2.5).

a labor rent (wage) to the household (N_t firms),

The asset values for firms being in each of three states are respectively given by:

$$W^{F,1} = -c + \tilde{\beta}_t \mathrm{E}_t \left[(1 - p_t^F) W_{t+1}^{F,1} + p_t^F W_{t+1}^{F,2} \right], \tag{2.14}$$

$$W^{F,2} = -\gamma + \tilde{\beta}_t \mathrm{E}_t \left[s W_{t+1}^{F,1} + (1 - q_t^F - s) W_{t+1}^{F,2} + q_t^F W_{t+1}^{F,3} \right], \tag{2.15}$$

$$W_t^{F,3} = -\rho_t - w_t + d_t + \tilde{\beta}_t \mathrm{E}_t \left[s W_{t+1}^{F,1} + (1 - s) W_{t+1}^{F,3} \right], \tag{2.16}$$

where now γ is the firm cost of searching for a worker (labor demand cost), w_t is the wage paid to the worker, and $\tilde{\beta}_t$ is the rate at which future profits are discounted. Since the small firms are held by the households, the rate at which future profits are discounted is provided by the usual consumption kernel (see equation (2.30)).

As before, the free entry condition states that the value of a firm which is searching for capital is equal to zero:

$$W_t^{F,1} = 0. \tag{2.17}$$

The representative household has an income that can be consumed (C_t), directly invested (I_t) into large firms or indirectly (through banks that behave as intermediaries and take risk) invested (H_t) into small firms. Its welfare satisfies the Bellman equation:

$$W^H(V_t, N_t, K_t) = \max_{I_t, H_t} \left\{ \mathcal{U}(C_t) + \beta \mathrm{E}_t \left[W^H(V_{t+1}, N_{t+1}, K_{t+1}) \right] \right\}, \tag{2.18}$$

under the budget constraint:

$$C_t + I_t + H_t = N_t w_t + (r_t + \delta) K_t + (r_t^b)(V_t + N_t) + \Pi_t^F + \Pi_t^B, \tag{2.19}$$

and the investment definitions:

$$I_t = K_{t+1} - (1-\delta)K_t \tag{2.20}$$

$$H_t = (V_{t+1} + N_{t+1}) - (1-s)(V_t + N_t). \tag{2.21}$$

$\mathcal{U}$ is an increasing and concave utility function and β is the household discount factor. Income includes wages w_t (paid by the small firms), interest r_t on capital in large firms and interest $r_t^b = (r_t + s)$ on capital in small firms (paid by the banks). We assume that households hold small firms and banks, and therefore receive their whole profits, respectively Π_t^F from firms and Π_t^B from banks. Maximizing (2.18) with respect to I_t and H_t, under the constraints (2.19), (2.20) and (2.21) gives the arbitrage conditions for deposits, vs capital from large firms:

$$\frac{\partial \mathcal{U}(C_t)}{\partial C_t} = \beta \mathrm{E}_t \left[(1 + r_{t+1}) \frac{\partial \mathcal{U}(C_{t+1})}{\partial C_{t+1}} \right], \tag{2.22}$$

$$\frac{\partial \mathcal{U}(C_t)}{\partial C_t} = \beta \mathrm{E}_t \left[(1 + r_{t+1}^b) \frac{\partial \mathcal{U}(C_{t+1})}{\partial C_{t+1}} \right]. \tag{2.23}$$

The final and intermediate goods markets and the large firms capital market are assumed to be perfectly competitive.

2.3.1 Representative bank

The problem for a bank changes only slightly from 2.2. The bank decides the level of loans L_t it wants to supply and collects the resulting number of new capital matches $H(L_t, E_t)$. The bank receives income ρ_t from the producing firms (in number N_t) but have to pay interests to households on the whole capital stock ($V_t + N_t$ units of capital). The representative bank's asset value satisfies the Bellman equation:

$$W^B(N_t, V_t) = \max_{L_t} \left\{ \rho_t N_t - k L_t - (r_t^b)(V_t + N_t) + \tilde{\beta}_t \mathrm{E}_t \left[W^B(N_{t+1}, V_{t+1}) \right] \right\}, \tag{2.24}$$

under the flow constraints:

$$V_{t+1} = (1 - s - q^F(\theta_t))V_t + p^B(\phi_t)L_t, \tag{2.25}$$

$$N_{t+1} = (1 - s)N_t + q^F(\theta_t)V_t, \tag{2.26}$$

where k is the bank cost of searching for a firm (capital supply cost). Maximizing (2.24) with respect to L_t and under the constraints (2.25) and (2.26) gives:

$$k = \tilde{\beta}_t p^B(\phi_t) \mathrm{E}_t \left[\frac{\partial W^B(N_{t+1}, V_{t+1})}{\partial V_{t+1}} \right]. \tag{2.27}$$

By the envelope theorem, we also have:

$$\begin{aligned} \frac{\partial W_t^B(N_t, V_t)}{\partial V_t} &= -(r_t^b + s) + \tilde{\beta}_t(1 - s - q^F(\theta_t))\mathrm{E}_t \left[\frac{\partial W_{t+1}^B(N_{t+1}, V_{t+1})}{\partial V_{t+1}} \right] \\ &+ \tilde{\beta}_t q^F(\theta_t)\mathrm{E}_t \left[\frac{\partial W^B(N_{t+1}, V_{t+1})}{\partial N_{t+1}} \right] \end{aligned} \tag{2.28}$$

$$\frac{\partial W_t^B(N_t, V_t))}{\partial N_t} = -(r_t^b + s) + \rho_t + \tilde{\beta}_t(1 - s)\mathrm{E}_t \left[\frac{\partial W^B(N_{t+1}, V_{t+1})}{\partial N_{t+1}} \right]. \tag{2.29}$$

Since the banks are held by the households, the rate at which future profits are discounted is:

$$\tilde{\beta}_t = \beta \mathrm{E}_t \left[\frac{\partial \mathcal{U}(C_{t+1})}{\partial C_{t+1}} \frac{\partial C_t}{\partial \mathcal{U}(C_t)} \right]. \tag{2.30}$$

2.4 Price determination and equilibrium

In a search matching framework prices do not have an allocative role and they are determined after demand and supply have already met. Prices need to lie within a set of values, so called bargaining set, insuring that for both sides of the bargain there is some non zero gain of being in the match: usually the chosen point corresponds to the outcome of the Nash axiomatic solution for the bargain. In our framework this would apply to both capital rents and wages. Nevertheless, Shimer (2004) have shown that wage dynamics have a sort of persistence which might be hard to replicate by the means of a Nash bargaining

solution. Empirical estimates show that interest rates can also have a staggered dynamics (e.g. with respect to monetary policy rates) but with a much faster adjustment than wages (Kok and Werner (2006)). Following Hall (2005) and for analytical convenience we first assume that real wages are fixed at their steady state level; section 2.7 presents results for the case when prices are determined by simple rules which describe some convex combination between previous period prices and their nash bargaining solution outcome.

We recap our main assumptions concerning prices, ranging the benchmark case to the full flexible case and we present the relevant equations.

a Wages are fixed at some steady state level (see section 2.5 for a discussion), while interest rates are determined by the Nash Bargaining solution between active firms and banks. We have then:[5]

$$w_t = \bar{w}, \forall t \tag{2.31}$$

and

$$\rho \equiv \arg\max_{\rho_t} \left(W_t^{F,3} - W_t^{F,1}\right)^{\zeta} \left(\frac{\partial W_t^B(N_t, V_t)}{\partial N_t}\right)^{1-\zeta}, \tag{2.32}$$

where $0 < \zeta < 1$ is firm's bargaining power.

b In the fully flexible case wages are also determined by solving a Nash bargaining, between households and firms:

$$w_t \equiv \arg\max_{w_t} \left(W_t^{F,3} - W_t^{F,1}\right)^{\xi} \left(\frac{\partial W^H(N_t, K_t)}{\partial N_t} \frac{\partial C_t}{\partial \mathcal{U}(C_t)}\right)^{1-\xi}, \tag{2.33}$$

where $0 < \xi < 1$ is firm's bargaining power. $W^{F,3}$ is the value of a matched firm and $\frac{\partial W^H(N_t,K_t)}{\partial N_t}$ provides the incentive of an household of having one more active small firm.

c We use simple rules as in Shimer (2009) to provide an intermediate case between (a) and (b), we describe them below.

[5]Under certain circumstances (too high wages combined with strongly negative shocks), firms and banks could want to destroy the job. It is easy to check *ex post* that it never happens (see appendix 3).

For the Nash bargaining solution a simple algebra can be provided as follows. Concerning rental rates, maximizing (2.32) with respect to ρ_t under the usual constraint defining the surplus to be shared between the parties (e.g. $S_t = W^{F,3} + \frac{\partial W_t^B(N_t,V_t)}{\partial N_t}$, gives:

$$(1-\zeta)\left(\frac{\partial W_t^B(N_t,V_t))}{\partial N_t} + W_t^{F,3} - W_t^{F,1}\right) = \frac{\partial W_t^B(N_t,V_t)}{\partial N_t}, \tag{2.34}$$

while for wages we have[6]:

$$(1-\xi)\left(\frac{\partial W_t^H(N_t,K_t)}{\partial N_t} + \frac{\partial \mathcal{U}(C_t)}{\partial C_t}(W_t^{F,3} - W_t^{F,1})\right) = \frac{\partial W_t^H(N_t,K_t)}{\partial N_t}. \tag{2.35}$$

In the case of wage norms the pricing formulas needs to be changed slightly: first assume that the actual wage w_t is formed as according to a simple rule which is a convex linear combination between past wages and the nash bargaining solution:

$$w_t = \tau w_t^{nash} + (1-\tau)w_{t-1}.$$

The ongoing wage w_t gives only some weight τ to the Nash bargaining solution. This latter can then be consistently computed by following the derivations from Shimer (2009) (pages 117–119). By inspecting equation (2.16) it is clear that the value function of a matched small which pays some wage rate $\tilde{w}$, different from the ongoing wage –call it ($\tilde{W^{F,3}}$)– will be given by the value function at w_t plus the difference $(\tilde{w}_t - w_t)$, i.e.

$$\tilde{W_t^{F,3}} = W_t^{F,3} + (\tilde{w}_t - w_t),$$

the same can be done for all value functions having the wage among their arguments (e.g. $\partial W^H(N)/\partial N$). The nash bargained wage will be the solution of a bargaining problem as the one outlined above where each value function x is replaced by $\tilde{x}$ as computed above. The nash bargaining equation consistent

[6]Alas in the complete set up closed form expressions for wage and rental rate of capital are in general not easy to interpret

with the pricing norm is then given by :

$$(1-\zeta)\left(\frac{\partial W_t^B(N_t,V_t))}{\partial N_t}+W_t^{F,3}-W_t^{F,1}\right)=\frac{\partial W_t^B(N_t,V_t)}{\partial N_t}+(w_t^{rule}-w_t^{nash}), \quad (2.36)$$

The same can be done for the rental rate of capital. Section (2.7) assesses the relevance of pricing norms.

The equilibrium definition for our economy it is given by the following description. Given initial conditions on V_t, N_t and K_t, an equilibrium of the economy is given by a sequence of prices $\{\mathcal{P}_t\}_{t=0}^{\infty} = \{r_t, r_t^b, d_t, w_t, \rho_t\}_{t=0}^{\infty}$ and a sequence of quantities $\{\mathcal{Q}_t\}_{t=0}^{\infty} = \{C_t, L_t, E_t, V_{t+1}, N_{t+1}, K_{t+1}\}_{t=0}^{\infty}$ such that:

- given a sequence of prices $\{\mathcal{P}_t\}_{t=0}^{\infty}$, $\{C_t\}_{t=0}^{\infty}$ is solution to the household problem (2.22)
- given a sequence of prices $\{\mathcal{P}_t\}_{t=0}^{\infty}$, $\{L_t\}_{t=0}^{\infty}$ is solution to the bank problem (2.27)
- given a sequence of prices $\{\mathcal{P}_t\}_{t=0}^{\infty}$, $\{E_t, V_{t+1}\}_{t=0}^{\infty}$ are solutions to the small firm problem (2.17) and the accumulation equation (2.12)
- given a sequence of prices $\{\mathcal{P}_t\}_{t=0}^{\infty}$, $\{N_{t+1}, K_{t+1}\}_{t=0}^{\infty}$ are solutions to accumulation equation (2.13) and the large firm problem (2.9)
- given a sequence of quantities $\{\mathcal{Q}_t\}_{t=0}^{\infty}$, $\{\mathcal{P}_t\}_{t=0}^{\infty}$ clears the final goods market:

  $$Y_t = \gamma V_t + kL_t + C_t + I_t + H_t,$$

  the intermediate goods market (2.8) and satisfies the parity condition (2.23)
- wages and small firms capital prices are set according to the determination mechanisms described above as a, b or c in this section.

A resume of the model equations and of the market clearing conditions is given in appendix 1.

2.5 Calibration

We calibrate our model on quarterly data to reproduce some stylized facts for the US economy. We choose a logarithmic utility function for the household and we set β equal to 0.99, implying a steady state real interest rate of roughly 4% per year. Large firms production technology is Cobb Douglas and the capital share in production (μ) is set to 0.33. The depreciation rate of capital δ is set to 0.025 and implies a steady state capital-production ratio of 9. For what regards the productivity shock we set an autocorrelation $\eta = 0.9$ and a standard deviation $\sigma_u = 0.0075$, as in standard real business cycle models (King and Rebelo (1999)).

We follow den Haan, Ramey, and Watson (2000) and define the matching functions as:

$$\mathcal{H}(L, E) = \frac{L\ E}{(L^h + E^h)^{\frac{1}{h}}}, \tag{2.37}$$

$$\mathcal{M}(V, 1-N) = \frac{V\ (1-N)}{(V^m + (1-N)^m)^{\frac{1}{m}}}, \tag{2.38}$$

where (i) the parameter $0 < h < \infty$ represents the matching efficiency (or the inverse of the friction level): if $h \to 0$ we have $H = 0$ (infinite frictions), if $h \to \infty$ we have $H = \min(L, E)$ (no frictions) and (ii) differently from the standard Cobb-Douglas formulation, whatever the friction level, we always have probabilities $0 < p^B, p^F < 1$[7].

Using the definitions of labor market tightness $\theta = V/U$ and of credit market tightness $\phi = E/L$, we define the following functions:

$$g(\theta) \equiv \frac{1}{(1+\theta^m)^{\frac{1}{m}}} \qquad \text{and} \qquad f(\phi) = \frac{1}{(1+\phi^h)^{\frac{1}{h}}}\ , \tag{2.39}$$

this allows us to simply rewrite the probabilities:

$$p^B = \phi f(\phi) \qquad \text{and} \qquad p^F = f(\phi), \tag{2.40}$$

$$q^F = g(\theta) \qquad \text{and} \qquad q^H = \theta g(\theta). \tag{2.41}$$

[7]We of course have symmetric properties for the matching function $\mathcal{M}$ and the probabilities $0 < q^F, q^H < 1$.

We get $g(0) = 1$, $g(\infty) = 0$ and $\partial g(\theta)/\partial\theta < 0$. Moreover $\lim_{m\to 0} g(\theta) = 0$, $\lim_{m\to\infty} g(\theta) = \min(1, 1/\theta)$ and $\partial g(\theta)/\partial m > 0$.

We get $\lim_{\theta\to 0} \theta g(\theta) = 0$, $\lim_{\theta\to\infty} \theta g(\theta) = 1$ and $\partial\theta g(\theta)/\partial\theta > 0$. Moreover $\lim_{m\to 0} \theta g(\theta) = 0$, $\lim_{m\to\infty} \theta g(\theta) = \min(1, \theta)$ and $\partial\theta g(\theta)/\partial m > 0$. We of course have symmetric properties for function f.

We assume that the bargaining power ζ between the firm and the bank is 0.5, hence the total surplus is equally distributed. We choose the steady state wage ($\bar{w}$) to obtain an unemployment rate of 5% ($1 - N^* = 0.05$), which is approximately the average US unemployment rate over the last 30 years (see for instance OECD data). From monthly estimations by Hall (2005), we take an average (over the last 55 years) US job destruction rate of 4%. Drawing on OECD (2004) survival rates of new firms entering the market, we compute an implied average destruction rate for firms entering the US market around 3 to 4 percent per quarter: therefore setting an equal destruction rate for jobs and firms not yet producing does not appear too restrictive.[8]. Matching efficiencies m and h are fixed to 2.53: a figure which implies a (steady state) matching elasticity of 0.5.

2.6 Steady state and cyclical properties

The steady state of our model can be reduced to two conditions describing the behaviour of banks and firms in the steady state as a function of tightness in the market for labor and credit. The system of equations describing the capital and the labor market is given by:[9]

$$\frac{k(s+g(\theta))}{\phi f(\phi)} + (s+r) = (1-\zeta)\frac{g(\theta)}{s}\left((d-w)-(s+r)\right), \tag{2.42}$$

$$\frac{c(s+g(\theta))}{f(\phi)} + \gamma = \zeta\frac{g(\theta)}{s}\left((d-w)-(s+r)\right). \tag{2.43}$$

[8]We checked that most of the results are robust to using $s = 0.01$ which is used in models of financial acceleration.

[9]See Appendix 5 for derivations

$g(\theta)$ is the probability for a firm to find a worker ($\theta g(\theta)$ is the job finding rate for a worker) and $f(\phi)$ is the probability for a bank to find a worker.

Equation (2.42) is the (θ, ϕ) equilibrium for the bank (*bank equation*). The left hand side is the overall cost to keep a loan open and the right hand side is its expected income. This curve is upward sloping: if θ is low (loose labor market), the probability for a firm to match a worker is high and the expected income for the bank is therefore also high, leading to high supply of capital by the bank (small ϕ) (see proposition 1, appendix 5).

Equation (2.43) represents the (θ, ϕ) equilibrium for the firm (*firm equation*). The left hand side is the expected cost for a firm to keep a vacancy open and the right hand side is the expected income. This curve is downward sloping: if ϕ is high, there is only a few banks and the average time (cost) for a firm to find a bank is also high, which implies that a firm needs a loose labor market (low searching costs) to compensate, that is a low θ (shown in proposition 2, appendix 5). The two curves describe a unique steady state state equilibrium (proposition 3, appendix 5).

The systen in (2.42–2.43) is useful to characterize the behaviour of the system following a permanent technology shock, which is reflected in a change of d_{ss}. The price d_{ss} measures the productivity of labor; a permanent productivity shock induces the K_{ss}/N_{ss} ratio to raise (capital deepening) in order to match the equality $1 + r = 1/\beta$. This in turn increases d_{ss}. We show the effect of a permanent productivity shock in figure 2.1:

A permanent shock on the price d makes labor hiring more profitable for firms and for an equilibrium to exist this must be compensated by lower probabilities of finding workers, that is, more congestion on the labor market. In our graph a positive shock corresponds to outwards shift of both the bank curve and the firm curve (proposition 4 in appendix 5). As according to economic intuition the labor market tightness θ unambiguously increases, while capital market tightness can either increase or decrease, as we discuss in section 2.6.1. On the basis of the benchmark calibration outlined above credit market tightness drops, as shown in Figure 2.1.

To show the cyclical properties of our model we simulate three different

models: (i) a basic RBC model *à la* Hansen (1985) where all markets are perfectly competitive, (ii) a RBC model *à la* Merz (1995) with frictions on the labor market (model that nests labor market frictions *à la* Pissarides (2000) into the basic RBC model) and (iii) our model with frictions on both the financial and the labor markets. We use comparable calibrations for all models and compare results to business cycle characteristics of US data: sources and methodology are reported in appendix 4.

The simulation results for the first experiment as well as the US stylized facts are summarized in table 2.1. The basic RBC model displays some shortcomings when compared to the real data: (i) most variables are not persistent enough unless we introduce a highly autocorrelated productivity shock, and (ii) all the variables are perfectly contemporaneous, while consumption leads and employment lags output in real data. Introducing frictions on the labor market only partly solves the problems: persistence in simulated data increases but still not enough; and although unemployment is now lagging, consumption is still coincident with output.

Our model further improves the simulated cyclical properties: consumption

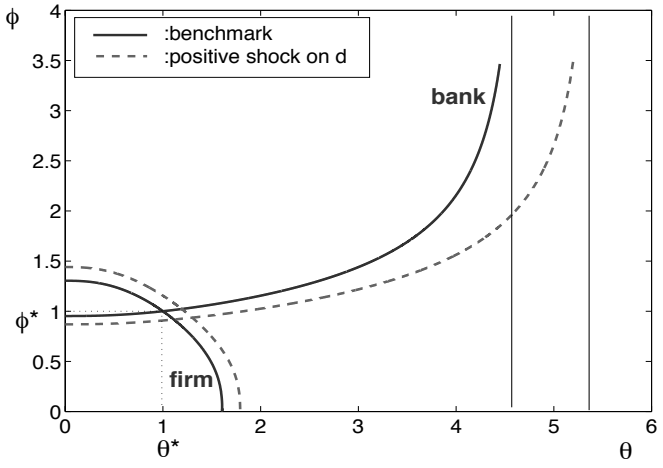


Figure 2.1: Equilibrium labor and financial markets: a permanent productivity shock

US economy (1970-2004)

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.29	0.88	1.00		
consumption	0.80		0.10	0.69	0.85	0.89	0.49
investment	2.84		0.36	0.88	0.95	0.85	0.28
unemployment	7.44		-0.45	-0.90	-0.88	-0.75	-0.13

model without frictions

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.07	0.69	1.00		
consumption	0.23		0.45	0.74	0.77	0.77	0.77
investment	3.58		-0.01	0.65	0.99	0.99	0.99
unemployment	15.28		0.05	-0.63	-0.98	-0.98	-0.98

model with labor frictions

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.14	0.86	1.00		
consumption	0.28		0.50	0.78	0.83	0.78	0.77
investment	3.50		0.04	0.83	0.99	0.94	0.92
unemployment	10.8		-0.27	-0.97	-0.90	-0.73	-0.69

model with labor and financial frictions

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.15	0.86	1.00		
consumption	0.27		0.45	0.72	0.77	0.79	0.74
investment	3.34		0.07	0.84	0.99	0.59	0.43
unemployment	11.6		-0.32	-0.99	-0.86	-0.59	-0.43

HP filtered quarterly data, with 1600 weight. US stylized facts: sources and methodology in appendix 3. Simulated data: GDP$\equiv C_t + I_t + H_t$, consumption$\equiv C_t$, investment$\equiv I_t + H_t$, unemployment$\equiv 1 - N_t$.

Table 2.1: Business cycle statistics

leads output, and unemployment and output are more persistent. Creating employment now first necessitates to obtain capital from a bank. The employment creation process is therefore much slower and that explains why employment/output has a stronger persistence.[10] The model with a Nash bargained wage and its statistical properties are displayed in appendix 2.

In our model a positive technology shock stimulates both labor demand and capital supply by banks. This latter generates a positive externality for the firms (higher probability to get matched) that increases further their labor demand, i.e. some kind of financial accelerator is at play.

Some further intuition concerning the role of both labor and capital market frictions can be gauged by comparing the response of the simulated economy to the same technology shock when either labor market or capital market frictions are increased. increasing illustrated in figure 2.1. We represent the impulse response functions to an aggregate productivity shock in our benchmark economy, when there are more frictions on the labor market (more precisely, we decrease the parameter m in equation (2.38) from 2.53 to 2) and in an economy with more frictions on the financial market (more precisely, we decrease the parameter h in equation (2.37) from 2.53 to 2). In steady state, this (rather strong) enhancement of frictions induces an increase in tightness of each of the markets in which frictions are raised: in the case of higher credit market frictions by roughly one percent and by almost ten percent in the case of higher labor market frictions. The overall reaction is very different in the two cases: when credit market frictions raise, the labor market tightness drops by a large amount, since firms are no longer able to finance vacancies; the reductions in steady state GDP is pretty large (GDP shrinks by almost 90 %), when the same experiment is performed on labor market frictions the drop of the steady state output is roughly equal to ten percent. Figure 2.2 shows that from a dynamic perspective, increasing the imperfections on capital markets amplifies and propagates the macroeconomic shock more than in the case of higher labor market frictions. (see also Wasmer and Weil (2004) for a similar evidence in a static model).

[10]To have a fair comparison we imposed an exogenous wage also on the model with only labor market frictions.

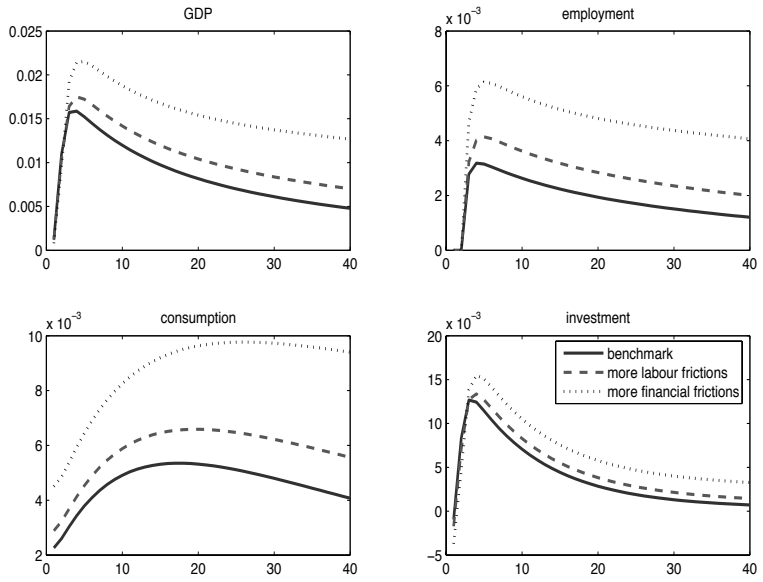


Figure 2.2: Impulse responses to a positive aggregate productivity shock

2.6.1 Credit market tightness

The way in which credit relates to labor market tightness is quite clear cut: since labor and credit markets act in complementary way we expect that a drop or rise of financial market tightness will be associated to a respectively stronger or weaker reaction of labor market tightness. Nevertheless the effect of the shock on ϕ, the credit market tightness, is ambiguous: credit market tightness can drop or rise depending on whether a technology shock makes economic conditions more convenient for new projects E_t or for loans L_t; when the relative economic convenience favors banks, the bank equation will shift more strongly downwards and vice versa in the case of small firms. When credit market tightness drops, existing projects will face an higher probability to be matched, providing an accelerating device to the system, in spite of the fact that, only on impact, the number of new projects can be comparatively reduced.

Which effect will predominate depends on many parameters and it is not straightforward to grasp, still, when firms and banks have equal bargaining

power ($\zeta = 0.5$), there is a simple condition such that credit market tightness decreases/increases after the shock can be given as follows:

$$\gamma \gtrless (s + r), \tag{2.44}$$

credit market tightness $\phi \equiv E/L$ increases (decreases) depending on the cost an open idle vacancy for the bank $s + r$ with respect to that of a small firm γ. To show the condition above we focus on a system composed by the *bank equation* and the *hybrid equation*; this latter is the locus of all the intersection points between the firm and the bank equation. When $\zeta = 0.5$, it is computed by simply equating the left hand sides of (2.42) and (2.43) and the curve does not depend on d.

$$\frac{k(s + g(\theta))}{\phi f(\phi)} + (s + r) = \frac{g(\theta)}{2\ s} \left((d - w) - (s + r) \right), \tag{2.45}$$

$$\frac{c(s + g(\theta))}{f(\phi)} + \gamma = \frac{k(s + g(\theta))}{\phi f(\phi)} + (s + r) \tag{2.46}$$

Equation 2.45 is the bank equation; a productivity shock shifts it outwards. It does not depend on the parameters (γ, c) since they only pertain to the firms.

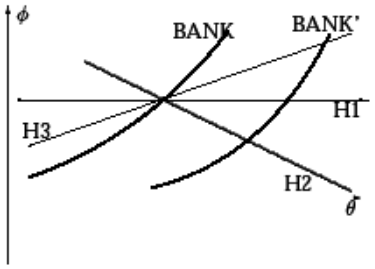
Equation 2.46 is the hybrid equation; when $\zeta = 0.5$ this is not influenced by the productivity shocks. The hybrid curve rotates and shifts with (γ, c). The hybrid curve will be downwards, flat or upwards sloping as according to 2.44, as shown in proposition 6, appendix 5.

In our exercise the hybrid curves rotates around the initial steady state (prop 7 in appendix 5):[11]

We display the effects of a permanent technology shock to market tightness for different calibrations of the costs of search of small firms (c, γ), keeping constant all the other parameters and the initial steady state. We have three possible cases:

H1 $\gamma = (s + r)$: Credit market tightness is not influenced by a productivity

[11] Only for simplicity the hybrid equation in the figure below is drawn as linear



shock: we call this case *financial neutrality*. c^* is the reference cost to search for a loan by a small firm in this case.

H2 $\gamma < (s+r)$ High labor market costs are compensated by low financial costs. ($c < c^*$). Credit market tightness *drops* after a technology shock, making it harder for small firms to find a bank. This entails a weaker effect of productivity shocks on labor market tightness. We call this case *financial deceleration*.

H3 $\gamma > (s + r)$ Low labor market costs are compensated by high financial costs ($c > c^*$). In this case credit market tightness *rises* after a technology shock, making it easier for small firms to find a bank; congestion on the labor market will be stronger than in the first two cases. We call this case *financial acceleration*.

The bank equation shifts outwards with d since higher profit opportunities have to be compensated by an increase in labor market tightness (it is less likely for a firm to find a worker) and a decrease in credit market tightness (it less likely for a bank to find a small firm). The hybrid curve does not move with the shock by construction but depending on the cost structure (c, γ) born by small firms it displays a particular rotation, H1, H2, H3. Economies with the same initial steady state labor and credit market tightness have very different responses depending of the structure of the costs of searching: when financial search costs are high, the impact of technology shocks on labor market tightness and on output are also high; in other words developed financial markets play a buffer role in the economy.

A temporary productivity shock for different calibration of the (c, γ) menu, corresponding to the H1,H2,H3 is shown in figure 2.3:

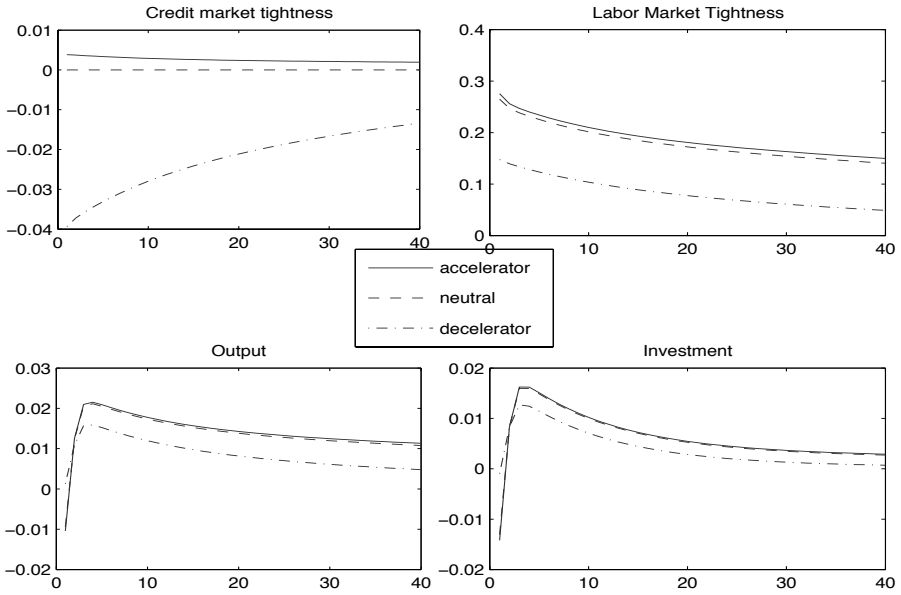


Figure 2.3: Impulse responses of a technology shock

The impact of technology shocks on labor market tightness and output decreases as long as the slope of the hybrid equation decreases, i.e. when we go from the accelerative environment $\gamma < (s+r)$ to the decelerating $\gamma > (s+r)$. The effect on output is also consistent with the steady state analysis: the strongest effect on labor market tightness also corresponds with the strongest effect on output[12]. This is not surprising since by using the production function we have:

$$Y_t = \left[\frac{K_t}{N_t}\right]^{\alpha} N_t,$$

where the term in square brackets depends only on relative prices which are weakly related to the reaction of market tightness and the reaction of the level of output is instead directly related to labor market tightness.

[12]In the figure we follow a plausible calibration on the US data to calibrate the deceleration environment only. As the reader can see, we use the deceleration with respect to the neutral case is much stronger than the acceleration. $0.6849 = c^a > c^* = 0.6667 > c^d = 0.367$

2.7 Different pricing rules

In this section we show results with pricing norms: ongoing prices p, p being either wages or the rental price of capital, are defined as a weighted average between the nash bargained solution and past prices:

$$p_t = \tau p_t^{nash} + (1-\tau)p_{t-1},$$

and the p_t^{nash} is consistently computed as shown in section 2.4. Results from a wage norm are shown in table 2.2:

Benchmark model with labor and financial frictions

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.15	0.86	1.00		
consumption	0.27		0.45	0.72	0.77	0.79	0.74
investment	3.34		0.07	0.84	0.99	0.59	0.43
unemployment	11.6		-0.32	-0.99	-0.86	-0.59	-0.43

Model with wage norms, $\tau = 0.1$

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.19	0.87	1.00		
consumption	0.38		0.26	0.65	0.79	0.93	0.92
investment	3.1		0.15	0.86	0.98	0.76	0.55
unemployment	11.5		-0.32	-0.99	-0.88	-0.75	-0.48

Model with wage norms, $\tau = 0.5$

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.19	0.84	1.00		
consumption	0.47		0.44	0.78	0.97	0.96	0.96
investment	2.6		0.13	0.85	0.99	0.99	0.96
unemployment	3.7		-0.54	-0.85	-0.63	-0.53	-0.47

Simulated data using wage norms with a different smoothing parameter. HP filtered quarterly data, with 1600 weight. GDP$\equiv C_t + I_t + H_t$, consumption$\equiv C_t$, investment$\equiv I_t + H_t$, unemployment$\equiv 1 - N_t$.

Table 2.2: Business cycle statistics

We obtain two results. First the model strongly needs wage rigidities in order to fit the data, second the role of pricing norms for interest rates turns out to be rather limited and, to save space we do not show results. For calibrations

of the τ which appear to be more plausible than the extreme case of a fully rigid wage the model falls short in providing enough volatility of the unemployment rate.

2.8 Conclusion

In this paper, we extend the Merz (1995) and Andolfatto (1996) RBC models with frictions on the labor market, by adding imperfections on the capital market. To do so, rather than to rely on asymmetric information and moral hazard problems as in Bernanke, Gertler, and Gilchrist (1999), we rely on localization problems as in Wasmer and Weil (2004) and model the frictions on the capital market in a fully symmetric way to the frictions on the labor market.

Our main results show that (i) imperfect capital market help explaining the data cyclical properties (ii) besides imperfect capital markets, rigid wages helps too. (iii) economies that feature lower credit market search costs can also entail attenuated effect of technology shocks on the labor market.

Comparing our model with a set-up that only entails labor market frictions, we are able to reproduce, as in an adjustment cost model, the empirical finding that consumption leads GDP over the cycle[13]. Second, higher credit frictions, modeled as higher search costs paid by small firms amplify the cycle. In a standard model a technology shock raises total factor productivity of the large firms and it raises the value of small firm products: this increases vacancy openings and it stimulates capital accumulation. An added twist of our model is that credit market tightness drops or raises depending on which side of the market has more convenience to add one vacancy after the shock. Since credit and labor are complementary a raise in credit market tightness is associated with a raise in labor market tightness. This implies that for a given steady state higher credit search costs imply a stronger reaction of macroeconomic variables after a technology shock.

Our model generalizes the results of Wasmer and Weil (2004) where credit market tightness does not depend on technology. Moreover since there are no

[13]We thank Marcello Savioz for this observation.

household deposits in their model the deposit cost for a bank to keep is implicitly set equal to zero and any positive vacancy opening cost for small firms does the job; implicitly it is as if we always set $\gamma > (s + r)$.

To keep this simple framework, this model is very stylized and could then be developed along several dimensions:

- Neither the exogenous wage (no flexibility at all) nor the Nash bargained wage (too much flexibility) lead to a fully convincing modelization. The model does not support wage norms when wage rigidities are low.
- Our credit view is not related to liquidity since money is not introduced in the model. It should rather be interpreted as propensity to lend. This model could be extended by adding a monetary, liquidity and by marching credit empirical facts found for example by the work of Dell'Ariccia and Garibaldi (2005).

We leave these extensions for future research.

2.9 Appendix

2.9.1 Appendix 1: model equations

Matching functions:

$$\mathcal{H}(L_t, E_t) = \frac{L_t E_t}{\left(L_t^h + E_t^h\right)^{\frac{1}{h}}}, \tag{2.47}$$

$$\mathcal{M}(V_t, 1 - N_t) = \frac{V_t(1 - N_t)}{\left(V_t^m + (1 - N_t)^m\right)^{\frac{1}{m}}}, \tag{2.48}$$

$$Y_t = \varepsilon_t K_t^{\mu} N_t^{1-\mu}, \tag{2.49}$$

$$p_t^B = \frac{\mathcal{H}(L_t, E_t)}{L_t} \qquad \text{and} \qquad p_t^F = \frac{\mathcal{H}(L_t, E_t)}{E_t}, \tag{2.50}$$

$$q_t^F = \frac{\mathcal{M}(V_t, 1 - N_t)}{V_t} \qquad \text{and} \qquad q_t^H = \frac{\mathcal{M}(V_t, 1 - N_t)}{1 - N_t}, \tag{2.51}$$

$$
\begin{aligned}
V_{t+1} &= (1-s-q^F(\theta_t))V_t + p^B(\phi_t)L_t, && (2.52)\\
N_{t+1} &= (1-s)N_t + q^F(\theta_t)V_t, && (2.53)\\
I_t &= K_{t+1} - (1-\delta)K_t, && (2.54)\\
H_t &= (V_{t+1}+N_{t+1}) - (1-s)(V_t+N_t), && (2.55)
\end{aligned}
$$

$$
\tilde{\beta}_t = \beta \mathrm{E}_t\left[\frac{C_t}{C_{t+1}}\right], \tag{2.56}
$$

$$
\begin{aligned}
\varepsilon_t(1-\mu)(K_t^{\mu}N_t^{-\mu}) &= d_t, && (2.57)\\
\varepsilon_t\mu(K_t^{\mu-1}N_t^{1-\mu}) &= r_t+\delta. && (2.58)
\end{aligned}
$$

Banks:

$$
\begin{aligned}
\frac{\partial W^B(V_t,N_t)}{\partial V_t} &= -(r_t^b+s) + \tilde{\beta}_t(1-s-q^F(\theta_t))\mathrm{E}_t\left[\frac{\partial W^B(N_{t+1},V_{t+1})}{\partial V_{t+1}}\right] && (2.59)\\
&+ \tilde{\beta}_t q^F(\theta_t)\mathrm{E}_t\left[\frac{\partial W^B(N_{t+1},V_{t+1})}{\partial N_{t+1}}\right]\\
\frac{\partial W^B(N_t,V_t)}{\partial N_t} &= -(r_t^b+s) + \rho_t + \tilde{\beta}_t(1-s)\mathrm{E}_t\left[\frac{\partial W^B(N_{t+1},V_{t+1})}{\partial N_{t+1}}\right], && (2.60)\\
k &= \tilde{\beta}_t p^B(\phi_t)\mathrm{E}_t\left[\frac{\partial W^B(N_{t+1},V_{t+1})}{\partial V_{t+1}}\right]. && (2.61)
\end{aligned}
$$

Small firms:

$$
\begin{aligned}
c &= \tilde{\beta}_t \mathrm{E}_t\left[p^F(\theta_t)W_{t+1}^{F,2}\right], && (2.62)\\
W_t^{F,2} &= -\gamma + \tilde{\beta}_t \mathrm{E}_t\left[(1-q^F(\theta_t)-s)W_{t+1}^{F,2} + q^F(\theta_t)W_{t+1}^{F,3}\right], && (2.63)\\
W_t^{F,3} &= -\rho_t - w_t + d_t + \tilde{\beta}_t \mathrm{E}_t\left[(1-s)W_{t+1}^{F,3}\right]. && (2.64)
\end{aligned}
$$

Interest rate bargaining and wages:

$$(1-\zeta)W_t^{F,3} = \zeta\frac{\partial W^B(N_t,V_t)}{\partial N_t}, \tag{2.65}$$

$$w_t = \bar{w}. \tag{2.66}$$

An aggregate productivity shock is modeled as:

$$\varepsilon_t = \varepsilon^{1-\eta}\varepsilon_{t-1}^{\eta}e^{u_t}, \tag{2.67}$$

where η is the autoregressive parameter and $u_t \sim N(0,\sigma_u^2)$ and the goods market clear:

$$Y_t = \gamma V_t + kL_t + C_t + I_t + H_t.$$

Appendix 2: Nash bargained wage

Table 2.3 compares these simulations to our benchmark model with exogenous wage and to the empirical facts. We see that adding the bargained wage deteriorates our results, by drastically reducing the unemployment volatility and correlations.

2.9.2 Appendix 3: Ex post check

A match (job) involves 3 agents: the worker, the firm and the bank. In *state 2*, a small firm and a bank are matched and the firm is trying to find a worker. As long as $W_t^{F,2} > W_t^{F,1} = 0$ and $\partial W_t^B/\partial V_t > 0$, both the firm and the bank have no incentive to voluntary destroy the match.

In *state 3*, a small firm and a bank are matched through a financial market and a small firm and a worker are matched through a labor market. As long as $W_t^{F,3} - W_t^{F,1} + \partial W_t^B/\partial N_t > 0$, both the firm and the bank have no incentive to voluntary destroy the match because the surplus to share between them is positive. The

US economy (1970-2004)

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.29	0.88	1.00		
consumption	0.80		0.10	0.69	0.85	0.89	0.49
investment	2.84		0.36	0.88	0.95	0.85	0.28
unemployment	7.44		-0.45	-0.90	-0.88	-0.75	-0.13

model with labor and financial frictions, exogenous wage

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.15	0.86	1.00		
consumption	0.27		0.45	0.72	0.77	0.79	0.74
investment	3.34		0.07	0.84	0.99	0.59	0.43
unemployment	11.6		-0.32	-0.99	-0.86	-0.59	-0.43

model with labor and financial frictions, Nash bargained wage

	st.dv.		cross corr. with GDP(t+k)				
variables	rel/GDP	k	-4	-1	0	1	4
GDP	1.00		0.10	0.72	1.00		
consumption	0.29		0.43	0.77	0.86	0.85	0.85
investment	3.21		0.00	0.67	0.99	0.99	0.99
unemployment	0.54		-0.56	-0.69	-0.44	-0.42	-0.42

HP filtered quarterly data, with 1600 weight. US stylized facts: sources and methodology in appendix 3. Simulated data: GDP$\equiv C_t + I_t + H_t$, consumption$\equiv C_t$, investment$\equiv I_t + H_t$, unemployment$\equiv 1 - N_t$.

Table 2.3: Business cycle statistics

worker has no incentive to destroy the match as long as $\partial W_t^H / \partial N_t > 0$.

To avoid these endogenous job destruction problems, we check *ex post* that these 4 inequalities are never violated during our simulations.

2.9.3 Appendix 4: Description of the data source and methodology

US data used in the paper include GDP, Consumption, Investment and Unemployment rate. The first three series come from the Quarterly National Accounts database of the OECD, from 1970q1 to 2004q4. All data are transformed into their logarithm and HP filtered with a 1600 weight.

The last series comes from the OECD Main Economic Indicator database, from 1970m1 to 2004m12. This is a monthly series and we take the three months average value for the quarter of interest. Unemployment rate data are HP filtered with a 1600 weight.

2.9.4 Appendix 5: analytical properties and derivation of the steady state

Bank equation

$$\frac{k(s+g(\theta))}{\phi f(\phi)} + (s+r) = (1-\zeta)\frac{g(\theta)}{s}\left((d-w)-(s+r)\right),$$

Plugging equations (2.29) and (2.16) into (2.34), we have

$$\rho = (1-\zeta)(d-w) + \zeta(r+s).$$

By injecting (2.27) and (2.29) into (2.29), and using the definitions of the probabilities (expressions (2.40) and (2.41)) and ρ, we get equation (2.42).

Proposition 1

ϕ is increasing on $[0, \theta^B]$ and takes values between $[\phi^B, \infty]$.

Proof

The increasing part can be shown by the means of the implicit function theorem on equation (2.42). From the definitions (2.39), we know that $\partial\phi f(\phi)/\partial\phi > 0$ and $\partial g(\theta)/\partial\theta < 0$. From (2.42), we can also show that $\partial\phi f(\phi)/\partial g(\theta) < 0$. From the last 3 inequalities, we can immediately deduce $\partial\phi/\partial\theta > 0$. From equation (2.42) we deduce ϕ^B: if θ=0 then $g(\theta) = 1$. If $\phi \to \infty$ then $\phi f(\phi) = 1$ and we can directly deduce the vertical asymptote θ^B.

QED

Firm equation

$$\frac{c(s + g(\theta))}{f(\phi)} + \gamma = \zeta \frac{g(\theta)}{s} \left((d - w) - (s + r)\right).$$

By injecting (2.14), (2.16) into (2.15), and using the definitions of the probabilities (expressions (2.40) and (2.41)) and ρ, we get equation (2.43).

Proposition 2

ϕ is decreasing on $[0, \theta^F]$ and takes values between $[0, \phi^B]$.

Proof

The decreasing part is shown by using the implicit function theorem. From the definitions (2.39), we know that $\partial f(\phi)/\partial\phi < 0$ and $\partial g(\theta)/\partial\theta < 0$. From (2.43), we can also show that $\partial f(\phi)/\partial g(\theta) < 0$. From the last 3 inequalities, we can immediately deduce $\partial\phi/\partial\theta < 0$. The remaining part is proved as follows. We look at equation (2.43). If θ=0 then $g(\theta) = 1$ and we can directly deduce ϕ^F. If $\phi = 0$ then $f(\phi) = 1$ and we can directly deduce θ^F.

QED

Existence and uniqueness of the steady state equilibrium

Proposition 3

If $0 < \phi^B < \phi^F$ then we have the existence of a unique and well-defined solution.

Proof

From the two previous propositions.

QED

Friction levels and shock effects

Proposition 4

A positive productivity shock (increase in d) increases the labor market tightness (increase in θ).

Proof

From the f properties, we have $\partial \phi f(\phi)/\partial \phi > 0$. Using equation (2.42) with θ constant, we have $\partial \phi f(\phi)/\partial d < 0$, that is, given the previous result: $\partial \phi/\partial d < 0$. In other words, a positive productivity shock moves the bank curve to the right. From the f properties, we also have $\partial f(\phi)/\partial \phi < 0$. Using equation (2.43) with θ constant, we have $\partial f(\phi)/\partial d < 0$, that is, given the previous result: $\partial \phi/\partial d > 0$. In other words, a positive productivity shock moves the firm curve to the right. If the two curves are moving to the right, this will automatically give a higher θ.

QED

Proposition 5

If banks and firms have an equal bargaining power $\zeta = 1 - \zeta = 0.5$, then $\frac{d\phi^{ss}}{dd} < 0$ if the following condition hold:

$$\gamma > (s + r). \tag{2.68}$$

Proof

When $\zeta = 1 - \zeta$, the left hand sides of the bank (2.9.4) and firm (2.9.4) equations can be equalized. After some rearrangements we have that:

$$k(s + g(\theta)) = -\phi f(\phi)(s + r) + [c(s + g(\theta)) + f(\phi)\gamma]\phi, \tag{2.69}$$

Now consider the system of equations (2.9.4) and (2.69). The bank equation (2.9.4) is upward sloping (see proposition above) and it shifts to the right with a positive technology shock. Equation (2.69) is independent on d. So it is sufficient (and necessary) to check that equation (2.69) is downward sloping in the space (ϕ, θ) to ensure that ϕ^{ss} will be negatively affected by a productivity shock. Applying the implicit function theorem to equation (2.69) we get:

$$\frac{d\phi}{d\theta} = -\frac{g'(\theta)(k - c\phi)}{(-\gamma + s + r)(f(\phi) + \phi f'(\phi)) - c(s + g(\theta)).} \tag{2.70}$$

Numerator and denominator must have opposite signs: since $g', f' < 0$, the condition on the numerator is as follows:

$$kL^{ss} - cE^{ss} > 0 \tag{2.71}$$

The denominator is a sum of two parts and it will always be negative provided that

$$\gamma > (s + r).$$

To complete the proof it is sufficient to rearrange equation 2.69 as

follows:

$$k(s+g(\theta)) - c(s+g(\theta))\phi = \phi f(\phi)[\gamma - (s+r)],$$

and to notice that 2.71 is implied by the condition $\gamma - (s+r) > 0$.

QED

2.9.5 Controlling for the rotation of the hybrid curve

In this section we develop the algebra to show how changes in γ are controlled by changes in the c term.

Looking at the equation (63), the hybrid equation, we see that only the term in square brackets contains the c and γ parameters:

$$c(s+g(\theta)) + f(\phi)\gamma,$$

for the solution (ϕ, θ) of the bank-hybrid system to be left unchanged by a parameter change this term has to be constant. This implies that a change $d\gamma$ should be offset by a dc equal to:

$$dc = \frac{-f(\phi)d\gamma}{s+g(\theta)}. \tag{2.72}$$

By looking at the steady state solution of the model (see file cal_{ww}) it can be shown that this is what is happening and that small discrepancies are only due to the approximation $\beta = 1$.

In fact from cal_{ww} we get that:

$$c = \frac{\beta p^f(-\gamma + \beta q^f W^{F,3})}{1-\beta(1-q^f-s)} \simeq \frac{-f(\phi)\gamma + q^f W^{F,3}}{g(\theta)+s}, \tag{2.73}$$

where $W^{F,3}(\rho, r, s)$ is a function of ρ, which does not change by moving γ, c since w, d are kept constant. By differentiating 2.73 we then obtain 2.72.

Chapter 3

Estimating DSGE models with unknown data persistence

3.1 Introduction

Over the last few years, DSGE models have become the workhorse of modern macroeconomic modeling and both academics and practitioners often use them to produce macroeconomic forecasts. The reduced form of these models, obtained once they are solved for expectations, is usually cast into a state space form, which describe the data as a (dynamic) linear combination of fundamental exogenous shocks and endogenous states. Under some general conditions this implies a (restricted) finite order VARMA representation for the data generating process (DGP).[1]

The appealing feature of DSGE models is that they are deeply grounded on economic theory. They link macroeconomic variables to the microeconomic behavior of agents, imposing meaningful restrictions on the data generating process, so called 'cross equation restrictions'. An important drawback of DSGE models is that their reduced form representation might be too stylized to possibly capture the persistent dynamics of the data, as highlighted by Cogley and Nason (1992) and, back in the early days, already by Kydland and Prescott

[1]The order of the VARMA depends upon the type of model at hand. In particular it is related to the order of the AR in the exogenous states as well as the number endogenous states which are treated as non observed variables, see Ravenna (2007) for more details.

(1982). An important consequence is that, when models are taken to the data, the value of estimated deep parameters fail to square well with available evidence either from national accounting (as for labor share, capital depreciation) or, more in general, from microeconometric evidence. As highlighted by Prescott and McGrattan (2007) (appendix), this issue (plus other problems) seriously weakens the use of 'microfounded models'.

Earlier attempts to reconcile the theoretical framework and the statistical properties of the data have mostly focused on modifying cross equation restrictions in order to add persistence to the simulated economy and have a better match with the data. This led to models which included a large number of 'frictions', endogenous state variables and exogenous state processes. While still at the forefront, Chari, Kehoe, and McGrattan (2008), criticized this line of research as 'lacking discipline', e.g. weak microfoundations and economic reasoning. A second line of research has focused on finding the "right" data transformation to resolve the mismatch between the data and the model implied dynamics. This has been done by either filtering the data before estimation, for example with either an HP or a Band Pass filter, or by explicitly including one or more unit roots in the model. The implications of these procedures are not always crystal clear: Chang, Doh, and Schorfheide (2007) show that we can not statistically discriminate *whether the 'true' DGP i*s better approximated by a model that includes unit roots in the exogenous states or by one that has more 'frictions'. At the same time, Canova (2009) and Canova and Ferroni (2009), show that different data filtering can influence the estimation results in a non-trivial way.

In this paper we take a different route from most of previous literature. We propose a method that does not impose any strong assumption on the statistical nature of model exogenous dynamics and makes the structure of DSGE model consistent with the persistence of the data at hand. This is done by adding a non-parametric step to the standard Kalman filter to adapt the VAR(MA) order of the DSGE model to the data persistence without modifying cross equation restrictions. Our procedure relates to Kalman filter in a similar way as the Generalized Least Squares refer to OLS, we name it 'Generalized' Kalman Filter ('GKF' heretofore). We show that our technique is equivalent to filtering the

data in a way that optimally removes the persistence which is left unexplained by the model structure.

The theoretical justification for our approach is the following. The order of the VAR(MA) representation of the data implied by a DSGE model depends crucially on the autoregressive order of the exogenous state processes. Highly persistent data might require a large number of lags to approximate accurately the data generating process. However, on the economic side there is an important drawback of using 'ad hoc' high order exogenous processes. By formulating a specific type of process for exogenous states, the researcher also casts assumptions on how expectations about future variables are formed; if a researcher modifies the order of the exogenous state processes to get a better fit of the data, the economic interpretation associated with the model might change dramatically every time a new dataset is considered. Probably this is one of the main reasons why exogenous state processes tend to be cast in few conventional ways, such as simple AR(1), or unit roots, with few exceptions in the literature.

For these reasons our aim is instead to make the structure of DSGE models consistent with the degree of persistence of the data at hand, in particular with the possibility of a long memory data generating process, a feature stressed by recent empirical literature (Mayoral (2005), Abadir, Caggiano, and Talmain (2006)). As we show below, this implies that we do not violate the cross-equation restrictions, which are applied to a set of filtered, rather than raw data. The method is also parsimonious since we use a non-parametric approach which enables us to approximate the dynamics of persistent processes (e.g. many lags) by estimating a few parameters. In particular, our filter spans from the case of strictly stationary to the most extreme case of long memory processes. Those have the appealing feature of having non-negligible spectral mass at low frequencies, 'the typical spectral shape of economic time series' (Granger (1966)), but their approximation with short order ARMA processes might be very poor, in particular when standard statistical lag-selection criteria are employed (Granger and Joyeux (1980)).

Our results can be summarized as follows. First we show that a strong degree of persistence in the data can substantially bias the structural parameters estimates of a DSGE model when exogenous state dynamics is misspecified

with respect to the true one. Following the thread of McGrattan (2006) and Ruge-Murcia (2007) we recur to simulated data. We evaluate the ability of the Maximum Likelihood (ML) methods to estimate the DSGE structural parameters and find a relevant amount of bias in the estimates: the stronger the persistence of the data the larger the bias. We show in the simulation that our approach dramatically reduces the bias in estimated parameters, which turns out to be negligible even when data are highly persistent. Finally, we turn to real data, and in line with Ireland (2004), we estimate a standard RBC model on U.S. data using both the Standard Kalman filter and the GKF. We report different and more plausible parameter estimates with the GKF than with the standard filter. Further evidence of the ability of the generalized Kalman filter to better capture the dynamics of the data, comes from more accurate out of sample forecasts compared to the standard filter. Finally, as a further check, we show that our method outperforms the case where a unit root is imposed on the technology shock and the model is estimated in levels.

The plan of the work is the following. In section 3.2 we present references needed to apply long memory methods to DSGE models, appendix 3.9 contains further details on general long memory processes. In 3.3 we describe the Generalized Kalman filter and the related estimation technique. Since the method can be applied also outside the field of DSGE models, we make this section as self-contained as possible, in order to be used as a cookbook recipe even by non-DSGE researchers. In section 3.4 we run some simulation to evaluate the effects of dynamic misspecification of exogenous states for the estimation of unobserved endogenous states (e.g. the capital stock) and deep parameter estimates. In this section we explain how the cross equation restrictions of the DSGE model come into the picture and how in our approach measurements are filtered in way which is consistent with the model. In section 3.5 we estimate a RBC model using real data for the U.S. economy and we undertake a forecasting exercise as in Ireland (2004) in section 3.6. In section 3.7 we discuss how our analysis relates to the case of pure unit roots in technology shocks. Some conclusions follow.

3.2 Background and objectives

In this section we provide some background relating long memory processes and DSGE models, some more details on long memory processes are in appendix (3.9).

Long memory processes have been extensively studied in time series analysis and a good review of their properties can be found in Robinson (2003) and Baillie (1996). Their main characteristic is an autocorrelation function that decays slowly to zero: compared to ARMA processes, it decays hyperbolically rather than exponentially. A well known class of long memory processes, introduced by Granger (1980), is the Autoregressive Fractional Moving Average Process of order d (ARFIMA(p,d,m)), which is defined as

$$(1-L)^d \phi(L) y_t = B(L) e_t, \tag{3.1}$$

where e_t is a white noise process with variance σ^2, $\phi(L)$ is a lag polynomial of finite order p, $B(L)$ is a lag polynomial of order m and $d \in [0,1]$. The basic building block of fractionally integrated processes is called the 'fractional white noise': it is a special case of (3.1), defined by setting $p, m = 0$ in (3.1). Since d can take any value in the interval 0 and 1, ARFIMA processes fill the gap between a strictly stationary process $(I(0))$, when $d = 0$, and a unit root process, when $d = 1$. For $0 < d < 0.5$ the process is "stationary"with hyperbolic rather than exponential decay of the autocorrelation function. When $0.5 < d < 1$ the process is non stationary (the squared sums of its autocorrelations do not converge), but, differently from unit root processes, it is still mean reverting. When $d < 0$ similar properties hold, with the difference that the autocorrelation function oscillates around zero in a way which shows an 'overdifferencing' pattern of the time series.

Several empirical studies argue that many macroeconomic variables are better described by long memory rather than AR(I)MA models: Diebold and Rudebusch (1989) were the first to report evidence of long memory in many macroeconomic time series; their results were recently confirmed by Mayoral (2005) with an updated version of the dataset. Abadir, Caggiano, and Talmain (2006)

show that long memory processes fit better the dynamics of the Nelson and Plosser dataset compared to ARIMA processes. Gadea and Mayoral (2006) provide robust empirical evidence supporting the hypothesis fractional integration for the inflation rates of the OECD countries while Altissimo, Mojon, and Zaffaroni (2009) reach the same conclusion for the euro area inflation rates. Evidence of long memory dynamics for hours worked has been found by Gil-Alana and Moreno (2006).

The main theoretical reason for long memory behaviour in macroeconomic data is the aggregation of heterogenous dynamic processes: fractional integrated process can be generated by *linearly* aggregating heterogenous ARMA models (Granger (1980)). More recently, Haubrich and Lo (2001) and Abadir and Talmain (2002) show that long memory processes can be produced in microfounded macroeconomic models with different sectors. In particular, Abadir and Talmain show that introducing heterogenous sectors in a monopolistic competition DSGE model leads to a *non-linear* long memory process for the aggregate output whose autocorrelation function reproduces the empirical shape found on the US GDP.

3.2.1 Background: DSGE and long memory

This section discusses our approach to incorporate potentially persistent data into the state space representation of a DSGE model. Consider a scalar state space model:

$$\theta_{t+1} = \Phi\theta_t + \varepsilon_{t+1} \tag{3.2}$$

$$y_t = H\theta_t, \tag{3.3}$$

where θ_t is an (unobserved) state variable, modeled by an AR(1) process; y_t is the measurement variable and the innovation ε_t is a normal i.i.d. process with standard deviation σ_ε. The dynamic properties of the data y_t closely depend on the specification of the transition equation for θ_t, in particular to its autoregressive order. For the case at hand, it is straightforward to show that y_t has an

autoregressive representation of order one:

$$y_t(1 - \Phi L) = H\varepsilon_t.$$

The reduced form of DSGE models can be represented in a state space form; for instance, 3.2 - 3.3 is consistent with the case of some DSGE model having only one unobserved exogenous state θ_t and one observable variable where the matrices $\Phi \equiv \Phi(\psi)$, $H \equiv H(\psi)$ are derived from the solution of the model as functions of some deep parameters ψ. In the more standard case (e.g. an RBC model) of one unobserved endogenous state variable (e.g. capital) and one AR(1) exogenous state (e.g. productivity), the measurement y_t (e.g. the GDP) would follow a (restricted) ARMA(2,1) process (see Ravenna (2007)).

A way to account for strong persistence in the y_t is by allowing the innovations ε_t to be a long memory process. Assume for example the following fractional noise representation:

$$(1 - L)^d \varepsilon_t = e_t. \tag{3.4}$$

with $e_t \sim IID$, then by construction the exogenous variable θ_t is a fractional autoregressive process,

$$(1 - \Phi L)(1 - L)^d \theta_{t+1} = e_{t+1}, \tag{3.5}$$

and the measurements have the following representation:

$$y_t(1 - \Phi L)(1 - L)^d = He_t, \tag{3.6}$$

which is a ARFIMA(1,d,0) process.

As we show in the paper, this way of introducing long memory in a state space allows us to account for the persistence of the data at hand without any implication for the structural cross-equation restrictions of the DSGE model. Broadly viewed, our approach connects to well-established methods in DSGE modelling based on the idea of data-filtering. In fact, as discussed in details in

section 3.4, the gist of our approach is to introduce a new variable $\tilde{y}_t$ defined as

$$\tilde{y}_t = \sum_{j=0}^{\infty} l_j y_{t-j},$$

where the weights $\{l_j\}_{j=0}^{\infty}$ are function of the persistence of the original data which is not captured by the model structure[2] and to estimate the model using this new variable:

$$\tilde{y}_t(1 - \Phi L) = He_t. \tag{3.7}$$

3.7 embeds the same structure (Φ, H) of the original model but the potential 'excess' persistence of the data is removed.[3]

An alternative and more straightforward strategy would be to directly increase the order of the exogenous state process and use conventional state space techniques to estimate the enlarged set of parameters. However, adding (too many) ad hoc lags would change the model structure, in the specific, the way agents form expectations about future variables, every time a new data set is considered. This would make the model exposed to some type of Lucas' critique. Some further discussion of important technical estimation issues of this alternative approach are relegated to a later section, (3.3.1).

3.3 Generalized Kalman Filter

In this section we introduce a modification of the Kalman Filter that can be used when the data persistence is unknown and possibly very strong. We start presenting the equations of the generalized filter and assume that the autocorrelation structure of the data is known. We remove this assumption in section 3.3.1 where we describe our estimation strategy .

[2]The fractional noise representation used in this section is merely for expositional convenience and it does not play any role in our estimation strategy. Our method, and in particular the estimation of the weights l_j, can be applied in both the case of errors being fractionally integrated or following even more general types of long memory processes. For the case of fractionally integrated processes, the appendix (3.9) shows how the weights l_j relate to the parameter d.

[3]This is implemented with an iterative estimation procedure.

Consider the following state-space model:

$$\theta_{t+1} = \Phi\theta_t + \varepsilon_{t+1} \tag{3.8}$$

$$\mathbf{y}_t = H\theta_t + v_t \tag{3.9}$$

$$E\left(\varepsilon_t\varepsilon'_{t+k}\right) = \left\{ \begin{array}{ll} Q & k=0 \\ 0 & otherwise \end{array} \right\} \tag{3.10}$$

$$E\left(v_t v'_{t+k}\right) = \left\{ \begin{array}{ll} R & k=0 \\ 0 & otherwise \end{array} \right\} \tag{3.11}$$

The $n-$dimensional state θ_t in the transition equation is modeled as an autoregressive process with innovation ε_t, Φ is its $n \times n$ transition matrix and $\mathbf{y}_t$ a $l-$dimensional vector of measurement variables, related to the states via an $n \times l$ matrix H. v_t is a vector of measurement errors, also of dimension l. Q is the variance covariance matrix of ε_t while R is an $l \times l$ diagonal matrix for the variance covariance matrix of the independent measurement errors.

Although we assume, for expositional simplicity, to have only one state variable in the θ_t vector, in the remaining part of this section we explain our methodology by referring to a generic state space model of type (3.8-3.11), our approach can be easily extended to more general case. To provide a short reference here, the simple state space (3.8-3.11) refers to a DSGE where θ_t corresponds to exogenous states (e.g. total factor productivity), commonly modelled as autoregressive processes of order one. In section 3.4 we examine how to apply our method to DGSEs and to extend the state space in order to encompass the case of one endogenous state variable (e.g. capital).

In (3.8-3.11) the state variable θ_t is assumed to be not observed. When the transition and measurement equations are linear and the shocks ε_t and v_t are normally distributed, the Kalman filter provides the best estimate of the state, conditionally on the assumed autoregressive process of the transition equation. In the following discussion we assume that both linearity of the model and normality of the shocks are good approximations of the data generating process.[4]

[4]This is also consistent with what is typically done in the DSGE literature, where models are typically linearized around their steady state and normal shocks are considered.

Define then:

$$\begin{aligned}\theta_{t|t-1} &= E\left(\theta_t|\Im_{t-1}\right), \\ P_{t|t-1} &= E\left[\left(\theta_t-\theta_{t|t-1}\right)\left(\theta_t-\theta_{t|t-1}\right)',\right]\end{aligned}$$

respectively the prediction of the state θ_t based on all information up to $t-1$ and its dispersion matrix, then the Kalman filter equations are given by:

$$\theta_{t|t-1} = \Phi\theta_{t-1|t-1}, \tag{3.12}$$

$$P_{t|t-1} = \Phi P_{t-1|t-1}\Phi' + Q, \tag{3.13}$$

$$\nu_t \equiv y_t - H\theta_{t|t-1}, \tag{3.14}$$

$$F_t = E\left[\left(y_t-H\theta_{t|t-1}\right)\left(y_t-H\theta_{t|t-1}\right)'\right] = HP_{t|t-1}H' + R, \tag{3.15}$$

$$K_t = P_{t|t-1}H_t'\left(HP_{t|t-1}H' + R\right)^{-1}, \tag{3.16}$$

$$\theta_{t|t} = \theta_{t|t-1} + K_t\left(y_t - H\theta_{t|t-1}\right), \tag{3.17}$$

$$P_{t|t} = \left(I - K_tH\right)P_{t|t-1}, \tag{3.18}$$

Equation (3.12) and (3.13) are respectively the state prediction and its variance, given the information set at $t-1$. Equation (3.14) defines the one-step-ahead prediction error of the measurements and (3.15) its dispersion matrix, given the information available at time $t-1$. This group of equations is called the *projection step*, in which states are predicted forward in time, leaving the information set of the observer constant. The last block of equations gives the so called *information updating step* in which the projection of the states is revised due to the arrival of time t information. The Kalman gain, (3.16), is a weighting matrix that minimizes the variance of the state forecast error, given available information:

$$K_t \equiv \underset{K_t}{\arg\min}\, E\left[\left(\theta_t-\theta_{t|t-1}-K_t\left(y_t-H\theta_{t|t-1}\right)\right)\left(\theta_t-\theta_{t|t-1}-K_t\left(y_t-H\theta_{t|t-1}\right)\right)'\right]$$

and it is used to extract information from time t prediction errors ν_t. By construction, past predictions $\theta_{t|t-1}$ and the constructed forecast errors ν_t are assumed to be orthogonal, an assumption which holds true when the transition

equation is not misspecified. Finally, $P_{t|t}$ in (3.18) is the variance covariance matrix of the state conditional on information at time t.

Equations (3.12) to (3.18) represent a system whose parameters can be estimated by maximum likelihood. In fact, if we assume that the innovations $\{e_t\}$ and $\{v_t\}$ are multivariate Gaussian, then the distribution of y_t conditional on the state θ_t and the information set at time $t-1$, is given by

$$f_{y_t|\theta_t,\Im_{t-1}}\left(y_t|\theta_t,\Im t-1\right)=(2\pi)^{-\frac{1}{2}}\left|HP_{t|t-1}H^{'}+R\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(y_t-H\theta_{t|t-1}\right)\left(HP_{t|t-1}H'+R\right)^{-1}\left(y_t-H\theta_{t|t-1}\right)'\right\} \tag{3.19}$$

which can be maximized with respect to the unknown parameters.

When the dynamics of the state process are misspecified, (e.g. the order of the autoregressive process is too short to approximate the DGP), the state space model described above no longer can be taken as a good approximation of the data generating process: the stronger the persistence in the original DGP, the more severe this problem is likely to be. As a result, both the projection step and the updating step, as designed in the standard Kalman filter will fail to be optimal and the parameter estimates will be biased. We cope with this problem by setting up a state space model where we allow ε_t to be serially correlated,

$$\theta_{t+1} = \Phi\theta_t+\varepsilon_{t+1} \tag{3.20}$$

$$\mathbf{y_t} = H\theta_t+v_t \tag{3.21}$$

$$E\left(\varepsilon_t\varepsilon'_{t+k}\right) = \begin{cases}\sigma^2\rho_\varepsilon(k) & \forall k <= m\\ 0 & otherwise\end{cases} \tag{3.22}$$

$$\Omega^m = \sigma^2\begin{bmatrix}1 & \rho_\varepsilon(1) & \cdots & \rho_\varepsilon(m)\\ \rho_\varepsilon(1) & 1 & \cdots & \rho_\varepsilon(m-1)\\ \vdots & \vdots & \ddots & \vdots\\ \rho_\varepsilon(m) & \rho_\varepsilon(m-1) & \cdots & 1\end{bmatrix} \tag{3.23}$$

$$E\left(v_tv'_{t+k}\right) = \begin{cases}R & k=0\\ 0 & otherwise\end{cases} \tag{3.24}$$

where σ^2 is the variance of ε_t and Ω^m is the matrix containing the autocorrelation structure of ε_t, $\{\rho_\varepsilon(k)\}_{k=1}^m$, up to lag m.

This set up can be handled in such a way to reconcile it with the standard approach, by constructing a 'bridge variable' which is an AR(1) with i.i.d. innovations, as follows:

1. Consider the Cholesky decomposition of Ω^m, namely

$$\Omega^m = \sigma^2 \Gamma \Gamma'$$

where Γ is a lower triangular matrix with elements $\{\gamma_{i,j}\}_{i,j=1}^{m+1}$.

2. Invert the matrix Γ and define $L \equiv \Gamma^{-1}$. Construct using the original states $\theta_{\mathbf{t}}$ (all vectors are denoted with bold typeset in the remaining part of this section) a vector of transformed variables $\mathbf{z}_t$ defined as

$$\mathbf{z}_t = \begin{bmatrix} z_{t-m} \\ \vdots \\ z_{t-1} \\ z_t \end{bmatrix} = \begin{bmatrix} l_{0,1} & 0 & \cdots & 0 \\ l_{1,1} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ l_{m,1} & l_{m,2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \theta_{t-m} \\ \vdots \\ \theta_{t-1} \\ \theta_t \end{bmatrix} \equiv L\theta_{\mathbf{t}} \tag{3.25}$$

3. By construction the transformed variable $z_t \equiv \theta_t + \sum_{j=0}^{m-1} l_{m,m-j}\theta_{t-j-1}$ is an AR(1) process, i.e.

$$z_{t+1} = \Phi z_t + e_{t+1}, \text{such that holds: } E(e_t \mid z_t, z_{t-1}, \ldots, z_{t-m}) = 0 \tag{3.26}$$

this step removes any autocorrelation in the residuals in equation (3.20), as in a GLS procedure; in the following text we refer to that as the 'GLS transformation'.

Coefficients of the L matrix, denoted as $\{l_{m,m-j}\}_{j=0}^{m-1}$, can be thought of as optimal weights such that $E(e_t|\mathbf{z_t}) = 0$ is satisfied when regressing z_t on its lagged value. This is the moment condition that we implicitly impose in the next section when estimating the parameters of the model. It can be immediately seen that if the ε_t are serially uncorrelated, then the coefficient $\{l_{m,m-j}\}_{j=0}^{m-1}$

are all equal to zero [5] and θ_t is equal to z_t which leads us back to the standard case. We are now able to define what we call henceforth the Generalized State Space model.

Definition 1 *Let's consider the variables defined in the state space model (3.20-3.24), where v_t is a (vector) of independent white noise with (diagonal) variance matrix R. The generalized state-space model is then defined as*

$$\begin{bmatrix} z_{t-m+1} \\ \vdots \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{(m)} & I_{(m)} \\ \mathbf{0}'_{(m)} & \Phi \end{bmatrix} \begin{bmatrix} z_{t-m} \\ \vdots \\ z_t \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{(m)} \\ \\ e_{t+1} \end{bmatrix} \tag{3.27}$$

or

$$\begin{aligned} \mathbf{z}_{t+1} &= \Psi \mathbf{z}_t + \mathbf{e}_{t+1} \\ \theta_t &= z_t + \sum_{j=0}^{m-1} \gamma_{m,m-j} z_{t-j-1} = \begin{bmatrix} \mathbf{0}_{m-1} & 1 \end{bmatrix}' \Gamma \mathbf{z_t} = D_m' \Gamma z_t \quad (3.28) \\ \mathbf{y_t} &= H\theta_t + v_t, \quad (3.29) \\ E\left(\mathbf{e}_t \mathbf{e}_t'\right) &= \left\{ \sigma^2 D_m D_m' = \tilde{Q} \right. \quad (3.30) \\ E\left(v_t v_t'\right) &= \{R \quad (3.31) \end{aligned}$$

where vectors are denoted with a bold and scalars with normal typeset: consistently with previous notation $\mathbf{y_t}$ is an $1 \times l$ vector, while $\mathbf{z}_t$ is an $m \times 1$ vector whose last element is z_t. $\mathbf{0}_{(m)}$ is a zero (column) vector of m-elements and $\gamma_{m,m-j}$ correspond to the m,m-j elements of the matrix Γ, D_m is a vector selecting the last row in the Γ matrix.

The main difference with the standard state-space models is given by equation (3.28) which maps the vector $\mathbf{z_t} \equiv [z_t, \ldots, z_{t-m}]$ into the scalar θ_t.[6] This can be considered as a "bridge equation" that embodies all the information on the autocorrelation function of shocks ε_t by transforming them into the i.i.d. innovations e_t. Once again, if the ε_t are serially uncorrelated, then the generalized state space model reduces to the standard state space model.

[5] since then $\Gamma\Gamma' = I$

[6] The filter by construction requires that we drop the first m observation in order to avoid any effect of the intimal condition at the beginning of the sample.

We are now ready to write the Kalman filter equations for the state-space model defined in (3.27-3.31). Following previous notation and using the generalized state space model, in particular equation (3.28), it is easy to construct prediction errors for the state θ_t as function of prediction errors of the GLS-transformed variable z_t and the weighting matrix Γ, as follows:

$$\theta_t - \theta_{t|t-1} = \left(z_t - z_{t|t-1}\right) + \sum_{j=0}^{m-1} \gamma_{m,m-j} \left(z_{t-j|t-j} - z_{t-j|t-j-1}\right) = D_m' \Gamma \left(\mathbf{z_t} - \mathbf{z_{t|t-1}}\right), \tag{3.32}$$

where the vector of filtered states is defined as $\mathbf{z_{t|t-1}} \equiv [z_{t|t-1}, \ldots z_{t-m+1|t-m}]$ so that no smoothing process of the past states is involved. From the state prediction errors all moments which are relevant for the Kalman filter can easily be derived. For example the prediction error variance for θ is written as:

$$P^{\theta}_{t|t-1} = P^{z}_{t|t-1} + \sum_{j=0}^{m-1} \gamma^2_{m,m-j} P^{z}_{t-j-1|t-j-2} = D_m \Gamma' P^{\mathbf{z}}_{t|t-1} \Gamma D_m', \tag{3.33}$$

which depends on matrix Γ and it is generally larger than the matrix $P_{t|t-1}$ defined in the standard Kalman filter (3.13).

The complete Kalman filtering equations can then be written as follows:

$$\begin{aligned}
\mathbf{z_{t|t-1}} &= \Psi \mathbf{z_{t-1|t-1}} && (3.34)\\
P^{\mathbf{z}}_{t|t-1} &= \Psi P^{\mathbf{z}}_{t-1|t-1}\Psi^{'} + \tilde{Q} && (3.35)\\
\theta_{t|t-1} &= D^{'}_{m}\Gamma \mathbf{z_{t|t-1}} && (3.36)\\
P^{\theta}_{t|t-1} &= D_m\Gamma^{'} P^{\mathbf{z}}_{t|t-1}\Gamma D^{'}_{m} && (3.37)\\
F_t &= E\left[\left(\mathbf{y_t} - H\theta_{t|t-1}\right)\left(\mathbf{y_t} - H\theta_{t|t-1}\right)^{'}\right] = HP^{\theta}_{t|t-1}H^{'} + R \\
&= HD_m\Gamma^{'} P^{\mathbf{z}}_{t|t-1}\Gamma D^{'}_{m}H^{'} + R && (3.38)\\
K^G_t &= P^{\mathbf{z}}_{t|t-1}H^{'}_t D^{'}_{m}\Gamma\left(HD_m\Gamma^{'} P^{\mathbf{z}}_{t|t-1}\Gamma D^{'}_{m}H^{'} + R\right)^{-1} && (3.39)\\
\mathbf{z_{t|t}} &= \mathbf{z_{t|t-1}} + K_t\left(\mathbf{y_t} - HD^{'}_{m}\Gamma \mathbf{z_{t|t-1}}\right) && (3.40)\\
\theta_{t|t} - \theta_{t|t-1} &= = D^{'}_{m}\Gamma\left(\mathbf{z_{t|t}} - \mathbf{z_{t|t-1}}\right) && (3.41)\\
P^{\mathbf{z}}_{t|t} &= \left(\mathbf{I_m} - K_tH\right)P^{\mathbf{z}}_{t|t-1} && (3.42)
\end{aligned}$$

The first two equations (3.34-3.35) correspond to the projection step applied to the transformed variable vector $\mathbf{z}$ rather than the original θ. The whole vector $\mathbf{z}$ is projected forward by using the matrix Ψ, large but very sparse in nature. Since the GLS transformation is a linear filter, coefficients in the Ψ matrix are the same as for the original untransformed state variable θ (e.g. the Φ), after taking into account lagged states as in our generalized state space model. The matrix $P^{\mathbf{z}}_{t|t-1}$ denotes the variance covariance matrix for the whole vector $\mathbf{z_{t|t-1}}$, while $P^{z}_{t|t-1}$ is its submatrix related only to $z_{t|t-1}$. If the $\{\varepsilon_t\}^T_{t=1}$ are uncorrelated then $\Gamma = I_m$ and $\theta_t = z_t$; we are back to the standard Kalman filter. The remaining main difference with the standard filter lies in how the Kalman gain is defined. As equation (3.39) displays, our filter solves the following problem:

$$\begin{aligned}
K^G_t \equiv \arg\min_{K^G_t} E\Bigg[&\left(z_t - z_{t|t-1} - K^G_t\left(y_t - HD^{'}_{m}\Gamma \mathbf{z_{t|t-1}}\right)\right) \\
&\left(z_t - z_{t|t-1} - K^G_t\left(y_t - HD^{'}_{m}\Gamma \mathbf{z_{t|t-1}}\right)\right)\Bigg] \quad (3.43)
\end{aligned}$$

We use it to update the GLS transformed state, $z_{t|t-1}$, by using prediction

errors defined on the original data vector $\mathbf{y}$. The Kalman gain K_t^G defined in (3.39) will be larger than its equivalent in the standard Kalman filter due to the larger $P^{\theta}_{t|t-1}$. This is because the generalized filter embodies all the information contained in the autocorrelation function of the state θ_t and consequently gives more weight (compared to the standard filter) to the observed variable when estimating the state variable at time t given the information at time $t-1$. On the other hand, the standard Kalman filter, by imposing a specified AR formulation to the state dynamics, would discard all the residual persistence and regard it as a noise component (e.g. measurement error) of the observed variable, as we discuss in section 3.4.

A problem of our generalized state space model is that in order to approximate persistent dynamics, a large number of lags m should be introduced in estimating the Ω^m matrix . A one-step estimation of such a parametric model by maximum likelihood is known to be likely to be troublesome with this respect, the interested reader may want to check the summary of the maximum likelihood procedure (Sowell (1992)) given in Baillie (1996). In the next section we present an approach, borrowed from Abadir and Talmain (2005), which allows to estimate parsimoniously (e.g. few parameter estimates are needed) the autocorrelation structure of the residuals even assuming very long processes. This allows us to apply the Generalized State Space Model described in this section.

3.3.1 Estimation

In general the autocorrelation matrix Ω^m is unknown and it can be derived either from assumptions on the nature of the shock, for example if we assume it is a fractionally integrated process, we can estimate the parameter d, and infer from that the autocorrelation structure, or by approximating the autocorrelation function of the innovations by using a large number of parameters. The former way is not very convenient in a state space model,[7] while the latter is unfeasible in the case of highly persistent shocks since we would need a large number of

[7]Furthermore since d is not a structural parameter, its estimation does not add any additional information from an economic point of view.

parameters to be estimated by Maximum Likelihood.

Following Abadir and Talmain (2005), we take a different path and estimate $\rho_{\varepsilon}(k)$ by fitting a functional form to the ACFs of the residual ε_t:

$$\rho_{\varepsilon}(k) \simeq \frac{1}{(1+a_1 k^{a_2})^{a_3}} \tag{3.44}$$

where a_1, a_2 and a_3 are parameters to be estimated. This functional form[8] was derived in Abadir and Talmain (2002) and corresponds to the decay rate of the ACF of a long memory process which includes the fractional integrated processes as a *special case.*

Using 3.44 and assuming that the innovations $\mathbf{e}_{t+1}$ and v_t are normally distributed, we can estimate the parameters of the generalized filter by maximizing the following likelihood function

$$\begin{aligned} f_{y_t|\theta_t,\Im_{t-1}}(\mathbf{y_t}|\theta_t,\Im_{t-1}) &= (2\pi)^{-\frac{1}{2}}\left|HP^{\theta}_{t|t-1}H^{'}+R\right|^{-\frac{1}{2}} \\ &\exp\Big\{-\frac{1}{2}\left(\mathbf{y_t}-H\theta_{t|t-1}\right)\left(HP^{\theta}_{t|t-1}H^{'}+R\right)^{-1}\left(\mathbf{y_t}-H\theta_{t|t-1}\right)^{'}\Big\} \\ &= (2\pi)^{-\frac{1}{2}}\left|HD\Gamma P^{\mathbf{z}}_{t|t-1}\Gamma^{'}D^{\prime}H^{'}+R\right|^{-\frac{1}{2}} \\ &\exp\Big\{-\frac{1}{2}\left(\mathbf{y_t}-HD^{'}\Gamma^{'}\mathbf{z_{t|t-1}}\right)\left(HD\Gamma P^{\mathbf{z}}_{t|t-1}\Gamma^{'}D^{\prime}H^{'}+R\right)^{-1} \\ &\left(\mathbf{y_t}-HD^{'}\Gamma^{'}\mathbf{z_{t|t-1}}\right)^{'}\Big\} \end{aligned} \tag{3.45}$$

[8]This functional form has been used in a number of papers. Abadir et. al. showed that it can capture very closely (and better than ARMA processes) the dynamic properties of many economic variables; Abadir and Talmain (2005) used it to construct a GLS approach, similar to the one proposed here, and to effectively deal with the uncovered interest rate puzzle; Moretti (2007) used it to develop a test for long memory co-movements between two macroeconomic time series.

where

$$\begin{aligned} \Gamma\Gamma' &= \sigma^2\Omega^m; \\ \Omega^m &= \{\omega_{i,j} : \omega_{i,j} = \rho_\varepsilon(k),\ k = |i-j|,\ i,j = 1,...,m\} \\ \rho_\varepsilon(k) &= \frac{1}{(1+a_1 k^{a_2})^{a_3}} \end{aligned}$$

with respect to the parameters of the filter matrices and the parameters of the functional form, implicitly requiring the moment condition $E(e_t|z_t) = 0$ to be satisfied.

This approach can be seen as a generalized method of moments estimation with a "GLS type" correction where Γ' is an optimal weighting matrix such that e_t and z_t are uncorrelated. In the next section, we show through simulation the ability of our approach to estimate the structural parameters of a simple DSGE model under the hypothesis that the DGP of the state is a fractional AR process.

3.4 The artificial DSGE economy

In this section we evaluate the accuracy of our approach in estimating a DSGE model in which the Kalman filter reconstructs both the path of an exogenous state (e.g. the capital stock) and that of an endogenous state (e.g. the capital stock). We show that when the persistence of the data at hand is unknown and stronger than what assumed by the researcher, the performance of the standard Kalman filter in reconstructing of the unobserved endogenous states (e.g. the capital stock) deteriorates and parameter estimates can be very biased. Conversely, that our generalized filter is able to effectively cope with the issue at hand.

Our experiment is based on simulated data in line with the recent literature (see for an example Ruge-Murcia (2007)). The framework is as simple as possible, we choose the Ramsey model which is the core of many, if not all, DSGE/RBC models. We introduce the misspecification by simulating a Ramsey model where the technology is an autoregressive process of order one with an innovation that follows a fractional noise (the process is described in the appen-

dix). A fractional noise allows us to easily control the degree of persistence by changing a single parameter, the order d of fractional integration; furthermore, it is also a good way to check that our method is able to approximate processes that, by nature, do not have a finite order state space representation.

A simple Ramsey model is sufficient to highlight one of the main points of the paper: unaccounted persistence in *exogenous* state processes distorts the path of the unobserved *endogenous* state variables and leads to substantial bias in the parameter estimates. The more persistent is the data generating process the larger is the root mean square error of the estimated capital stock. At the same time, the estimated deep parameters, which determine the reduced form parameters in the transition equation for capital accumulation, are more and more biased with respect to the true ones.

For each sample we first estimate the unobserved states and the deep parameters by the means of the usual Kalman filter, documenting the effects of dynamic misspecification. Then we apply our Generalized Kalman Filter to the same set of data and show that the estimation bias decreases significantly compared to the standard case. A further robustness check is conducted in subsection 3.7.1 where we investigate the extent to which long memory dynamics can be tackled by introducing a unit root in technology.[9]

Our set-up features a productivity shock (a_t) as unobserved exogenous state, capital (k_t) as unobserved endogenous state, consumption (c_t), output (y_t), the real interest rate (r_t) as measurements.[10] We simulate the Ramsey model to generate 1000 artificial data samples each one of 170 observations for consumption, output and the real interest rate. We repeat this procedure for each integer degree of fractional order d ranging from 0.1 to 0.9.

In the following we provide first a short summary of the model we use and we explain how our method is applied to it. We show that our method is equivalent to building a filter on the data which removes the 'excess persistence',

[9] In this case, in order to simulate the data, we use the same model used for real data estimation. We postpone the discussion accordingly. This also makes the presentation of the unit-root case more self contained.

[10] As there is a single structural shock it would also be possible to estimate the model using only one series, e.g. the GDP. Here we stick to the treatment of Ireland (2004), where one shock, technology, is used as the source of comovement of variables

which is left unexplained by the model. Simulation results are then provided in subsection 3.4.1.

The model can be summarized as a standard (decentralized) market problem as follows. Households choose consumption C and investments I such as to maximize their objective function:

$$U_0 = E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \log \left(C_t \right) \right\}, \tag{3.46}$$

subject to a budget constraint:

$$C_t + I_t \leq W_t + R_t K_{t-1}, \tag{3.47}$$

where W_t is the real wage rate and R_t is the real interest rate on previous period capital stock K_{t-1}; since household face no leisure choice, we normalized labor to one.

Capital is set by the households with the standard law of motion, involving the capital depreciation parameter δ:

$$K_t = (1 - \delta) K_{t-1} + I_t,$$

where no growth trend in productivity is assumed. Firms use capital according to the production function:

$$Y_t = A_t K_{t-1}^{\alpha}, \tag{3.48}$$

where technology evolves as:

$$\log(A_t) = \rho \log(A_{t-1}) + \varepsilon_t, \tag{3.49}$$

where ε_t is considered as i.i.d. by the households with a constant variance $var(\varepsilon_t) = \sigma_{\varepsilon}^2$.

Using equation (3.47), the absence of profits, capital accumulation can be rewritten as:

$$K_t = (1 - \delta) K_{t-1} - C_t + W_t + R_t K_{t-1}, \tag{3.50}$$

The first order necessary condition associated with the maximization of the objective (3.46) subject to (3.50) is given by the standard Euler equation:

$$C_t^{-1} = E_t \left[\beta(1 + R_{t+1} - \delta)C_{t+1}^{-1}\right] \tag{3.51}$$

Firms rent capital K_t by paying a rental price R_t and maximize their profits by choosing capital such that the real interest rate equals the marginal productivity of capital minus the depreciation rate δ:

$$R_t + \delta = \alpha A_t K_{t-1}^{\alpha-1} \tag{3.52}$$

Finally we have the goods market clearing condition

$$Y_t = C_t + I_t.$$

The competitive equilibrium for the economy is the sequence of prices $\{R_t, W_t\}_0^\infty$ and quantities $\{Y_t, K_t, C_t\}_0^\infty$ such that firms maximize profits, agents maximize utility and all markets clear. This amounts to estimating the system of equations given by the log-linear approximation of (3.48-3.52). When calibrating the model for the simulations deep parameters are set in a way which is consistent with the previous literature. The capital share α is set equal to 0.33; the preference term β is equal to 0.99 which corresponds to a real interest rate of 0.04 on annual basis; the depreciation rate δ is set equal to 0.025; the autoregressive term is set to 0.9. To gain on clarity σ_ε^2 is set equal to 1.[11] In order to remove any singularity in the system of equations we add two i.i.d. normal measurement errors with zero mean and standard deviation equal to 0.01 to consumption and the real interest rate. To keep the estimation process as clean as possible we do not introduce measurement errors as autocorrelated processes in order to ensure technology shock is the only dynamic factor in the model. The model is then log-linearized around its steady state [12] and solved under the assumption that innovations ε_t are i.i.d.

By denoting log-deviations from the steady state with lower case variables,

[11] We checked that results are unchanged by setting more realistic values.

[12] The steady state values of the variables are denoted with ^{ss}

we write the model solution in the following form:

$$\begin{aligned} a_{t+1} &= \rho a_t + \varepsilon_{t+1}, && (3.53)\\ k_{t+1} &= p_{kk} k_t + q_{ka} a_t, && (3.54)\\ \mathbf{y_t} &= m_{yk} k_t + n_{ya} a_t, && (3.55) \end{aligned}$$

For clarity we separate here exogenous and endogenous state variables, respectively

$$a_t \equiv \ln(A_t/A^{ss}), k_t \equiv \ln(K_t/K^{ss}),$$

and the vector $\mathbf{y_t}$ collects measurements:

$$\mathbf{y}_t = [\ln(C_t/C^{ss}),\, \ln(Y_t/Y^{ss}),\, \ln(R_t/R^{ss})].$$

The $\{p, q, m, n\}$ matrices describe the (linear) rational expectation solution of the model, the so called policy functions of the model. p and q are in this case scalars which denote how endogenous state variables respond respectively to lagged endogenous states and to exogenous states. Matrices m, n (3×1) denote how measurements react to the same set of variables.

It is straightforward to cast the model in a state space representation: define a vector of state variables, $\theta_{\mathbf{t}} \equiv [a_t, k_t]$, stack the equations (3.53-3.54) to form the transition equations of the state space. Add a set of measurement errors to (3.55) to specify the measurement equation. Finally, we can use the Kalman Filter innovations to estimate parameters.

When the unobserved innovations ε_t are serially correlated, equation (3.53) is not a good approximation of the exogenous state dynamics. Our Generalized State Space framework deals with this problem as follows. Append an equation to the state space model describing the residuals' autocorrelation:

$$\begin{aligned} a_{t+1} &= \rho a_t + \varepsilon_{t+1}, && (3.56)\\ k_{t+1} &= p_{kk} k_t + q_{ka} a_t, && (3.57)\\ \mathbf{y}_t &= m_{yk} k_t + n_{ya} a_t, && (3.58)\\ E[\varepsilon_t \varepsilon_{t+k}] &= \rho_\varepsilon(k),\ k = 1, \ldots, \infty && (3.59) \end{aligned}$$

where the autocorrelations ρ_ε, up to lag m, are estimated using the functional form in 3.44. Then, as in section 3.3, invert the cholesky of the variance matrix of ε_t and get a set of coefficients $\{l_{m,j}\}_{j=1}^{m-1}$. Define a new exogenous variable:

$$\tilde{a}_t = a_t - \sum_{j=1}^{m-1} l_{m,j} a_{t-j}, \tag{3.60}$$

where $\tilde{a}_t$ corresponds to the 'bridge variable' defined in section (3.3) and $\{l_{m,j}\}_{j=1}^{m-1}$ to the last row of the matrix L in the generalized state space model (3.25).[13] Once the bridge variable is described, one can form the generalized state space (3.27–3.31), use the formulas in (3.34–3.42) in section 3.3.1, to respectively build the filter and undertake the estimation. The economic interpretation of this approach is now given. First, notice that the 'GLS-transformed' variable $\tilde{a}_t$ is now by construction an AR(1) process. Then since the GLS transformation is a linear filter, it can be thought as applying to all variables through the exogenous shock $\tilde{a}_t$ resulting in the following model:

$$\begin{aligned}
\tilde{a}_{t+1} &= \rho\tilde{a}_t + \eta_{t+1}, && (3.61)\\
\tilde{k}_{t+1} &= p_{kk}\tilde{k}_t + q_{ka}\tilde{a}_t, && (3.62)\\
\tilde{\mathbf{y}}_t &= m_{ka}\tilde{k}_t + n_{ya}\tilde{a}_t, && (3.63)
\end{aligned}$$

Where $\eta_t \sim IID$ and each GLS-filtered variable x is denoted by $\tilde{x}$. The filtered capital stock $\tilde{k}_t$ and measurements $\tilde{\mathbf{y}}_t$ are defined as:[14]

$$\begin{aligned}
\tilde{k}_t &= k_t - \sum_{j=1}^{m-1} l_{m,j} k_{t-j}, && (3.64)\\
\tilde{\mathbf{y}}_t &= \mathbf{y}_t - \sum_{j=1}^{m-1} l_{m,j} \mathbf{y}_{t-j}. && (3.65)
\end{aligned}$$

[13] Here $\tilde{a}_t$ and a_t play respectively the role of z_t and θ_t in the equation $z_t = \theta_t + \sum_{j=0}^{m-1} l_{m,m-j}\theta_{t-j-1}$.

[14] There is a little abuse of notation since $\mathbf{y_t}$ is a vector; in this case it is sufficient to stack the same filter in different columns, one for each variable of vector $\mathbf{y_t}$.

Since the GLS-transformation is a linear filter, our approach is equivalent to applying the cross equation restrictions, defined in the matrices (p, q, m, n), to the set filtered variables. In other words, our method can be seen as a way of filtering the data and applying model restrictions to filtered, rather than to raw, data. With this respect, the filter endogenously account for the part of the data not explained by the model and corrects variables consistently to that information.[15] The same type of logic, albeit within a different technical framework, is followed by Cogley and Sbordone (2008) where the cross equation restrictions (e.g. the New Phillips curve in their application) hold for an inflation gap obtained using a time varying trend inflation. Our framework is also related to the recent work of Canova and Ferroni (2009) where they show that different data filtering lead to different behaviour of the structural shocks of the model (e.g. in some cases there will be too much persistence left, in others some signs of overdifferencing). For these reasons, they propose the "naive" approach to combine in the measurement equation several filtered versions of the same variable (e.g. Beveridge Nelson filtered, HP and others) aggregated with estimated weights in order for residuals of the Kalman filter to be closer to i.i.d. Finally a way of dealing with persistence in DSGE models which shares some common points with ours can be found in Gorodnichenko and Ng (2009), who propose to treat the reduced form of the DSGE model with the filter defined by the quasi difference of exogenous states, e.g. $1 - \rho L$, where ρ is the persistence parameter of AR(1) process of the exogenous state and L is the lag operator.

3.4.1 Simulation results

We simulate 1000 artificial data for DSGE model described in the previous section. For each sample, we estimate the parameters and the unobserved state variables first with standard likelihood methods (see 3.19) and then with the generalized state space approach in sections 3.3-3.3.1. The functional form (3.44) is used to recover the autocorrelation matrix Ω^m of the residuals of technology, as explained in 3.3.1, then its Cholesky factorization to recover the weights

[15] As this procedure improves the fit one might want to evaluate a model by how much of the data it is able to explain versus how much it is left to the filter. We leave this question to further research.

$\{l_{m,j}\}_{j=1}^{m-1}$ contained in the matrix L defined above. We set the number of lags m equal to 30 since it seemed sufficient to successfully approximate fractional integration behaviour.

Table one reports simulation results for the standard filter. The second column shows the true model parameters, while in the following ones we report the sample means of each estimated parameter for different degree of persistence (e.g. an increasing d). The last two rows report our assessment, measured by the Root Mean Square Error, of how well the two filters are able to reconstruct the true unobserved state variables (the rows are denoted with $a_{t|t}$ and $k_{t|t}$, the RMSE refers to the error $a_t - a_{t|t}$ and $k_t - k_{t|t}$).[16] states example annual also reliable the Results show that, when the persistence of exogenous states is higher than what was assumed by the researcher, due to long memory, the reconstruction of the unobserved states will be poor. In the estimation process, deep parameters which determine the persistence of the transition equation will tend to be biased in order for the reduced form of the model to try to match the persistence in the data.[17]

[16] For each simulation sample we estimate the structural parameters of the model

[17] The actual pattern of deep parameter estimates is hard to tell in advance since it will depend on which measurements are used. If the researcher uses several measurements which are deeply correlated with the unobserved state one might expect a better estimate of the deep parameters in spite of the misspecification. We leave this point, to further research.

		Fractional Integration Parameter d								
	True	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
α	0.330	0.3478	0.3556	0.3645	0.3704	0.3702	0.3768	0.4109	0.4605	0.5483
δ	0.025	0.0266	0.0272	0.0279	0.0279	0.028	0.0285	0.0309	0.0343	0.0407
β	0.99	0.9875	0.988	0.9881	0.991	0.9901	0.9903	0.9902	0.9923	0.9954
ϕ	0.9	0.8903	0.8967	0.9004	0.9066	0.9067	0.9078	0.9126	0.9226	0.9388
σ_e	1	1.0404	1.0404	1.1358	1.2554	1.4744	1.703	1.8521	1.9403	1.9852
σ_{v_1}	0.01	0.1122	0.1384	0.0898	0.0698	0.0344	0.0258	0.0211	0.0211	0.0218
σ_{v_2}	0.01	0.1064	0.1314	0.1228	0.1289	0.1237	0.1359	0.2217	0.3609	0.6956
RMSE										
$a_t - a_{t\|t}$		1.0439	1.0967	1.1964	1.3862	1.7154	2.4849	5.6623	11.1167	19.6298
$k_t - k_{t\|t}$		0.3827	0.4714	0.5357	0.6491	0.8617	1.5441	5.3139	10.8163	19.0695

Table 3.1: Results for degree of fractional integration d: Standard Kalman filter

Our experiment delivers the following results. First, the bias in parameter estimates is relevant and it tends to increase as the persistence in the simulated data becomes stronger. For d larger than 0.5, the bias becomes quite substantial, especially for the capital share α, the standard deviation of the shock σ_e, e.g. the innovation on productivity, and the standard deviation of measurement errors. In particular, for large values of d the estimates of the capital share are similar to what found by Ireland (2004) on real data. The upward bias in the capital share is also related to the distortion in the capital dynamics; in fact, a value of α above its true value increases the return on capital and consequently it makes the capital more persistent in the reduced form. In our simulations this effect is partially compensated by an increase in the capital depreciation δ (which lowers the persistence of capital) but the resulting errors in the reconstruction of capital dynamics are still considerable. Further experiments showed that the increase in δ is mostly due to the inclusion of the real interest rate among the measurements; this variable is difficult to grasp in real data and it is almost never included in the dataset of a DSGE model.[18] Second, we find that the ac-

[18]When we estimate a similar model on U.S. data and we do not include a real rate measure among the observables, the estimate of the depreciation rate is much lower than what one should expect, see section 3.5.2 for discussion. We leave a more complete assessment of which variables should be selected in the measurement equation of a DSGE model to future research.

curacy of the standard filter in reconstructing both the exogenous $a_{t|t}$ and the endogenous state $k_{t|t}$ is poor (RMSE shown in the last two rows of Table one) and it gets much worse as the parameter d increases. As both components are typically unobserved, this check is only possible by using simulated data, but it is important as, in general, the results of many empirical papers may depend on how unobserved states are reconstructed.

Our third result is that,as far as d increases, the variance of measurement errors also tends to increase: persistent dynamics is discarded as measurement error. One caveat holds for this result: a further investigation showed us that the pattern of measurement errors estimates depends by few very large estimates that bias upwards the mean of the simulation samples.[19] not depend on If we remove those samples, the estimates of all other parameters are almost unchanged, but we get an almost accurate estimate for σ_{v1} and a clearer pattern for σ_{v2} (e.g. its bias rises progressively as d increases):

True σ_{v2}	0.01
$d = 0.1$	0.0101
$d = 0.2$	0.0128
$d = 0.3$	0.0199
$d = 0.4$	0.0213
$d = 0.5$	0.0254
$d = 0.6$	0.0311
$d = 0.7$	0.1129
$d = 0.8$	0.2637
$d = 0.9$	0.5756

Table 3.2: Estimates for σ_{v2} after removing problematic samples

In Table 3.3 we report the results of the same exercise for the Generalized Filter.[20]

[19]The estimates were just large, they did not touch any of the boundaries imposed for the parameter estimation

[20]We choose a value of m, number of lags in productivity, equal to 30. This seems to be a reasonable compromise; while implementing a rather effective correction it does not exclude too many observations.

		Fractional Integration Parameter d								
	True	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
α	0.330	0.3302	0.3301	0.3301	0.3299	0.3299	0.3299	0.3309	0.3342	0.3411
δ	0.025	0.025	0.025	0.025	0.0249	0.025	0.02499	0.0251	0.0253	0.0258
β	0.99	0.99	0.99	0.99	0.99	0.99	0.9900	0.99	0.9901	0.9903
ϕ	0.9	0.90	0.90	0.90	0.90	0.9	0.899	0.9002	0.9007	0.902
σ_e	1	0.9964	1.0017	1.0106	1.01	1.0295	1.0437	1.0598	1.0836	1.1301
σ_{v_1}	0.01	0.0099	0.0099	0.0099	0.01	0.0099	0.0099	0.0099	0.0099	0.0099
σ_{v_2}	0.01	0.0099	0.0099	0.0099	0.01	0.0099	0.0100	0.0119	0.0249	0.0524
RMSE										
$a_t - a_{t\|t}$		1.0154	1.0172	1.0342	1.0654	1.1241	1.2155	1.3512	2.6863	5.3401
$k_t - k_{t\|t}$		0.2073	0.2542	0.3135	0.4014	0.5165	0.6899	0.9249	2.5774	5.1981

Table 3.3: Results for degree of fractional integration d: Generalized Kf

As it can be readily seen, the amount of bias provided by the generalized filter is almost negligible. This is true for all the degree of fractional integration, even when the state variables become non-stationary $(d > 0.5)$. The generalized filter also provides much more accurate prediction of the endogenous and exogenous state variable compared to the standard filter. In fact, the RMSE errors of the state predictions are up to 4 times smaller than those obtained with the standard filter.

This shows the ability of the modified filter to capture the dynamic properties of the data. In the last section we show how this accuracy in predicting the data dynamics also holds when undertaking an out-of-sample forecasting exercise.

3.5 Real data estimation

3.5.1 Model

In this section we take our model to the real data and repeat the Maximum Likelihood estimation for a RBC model as in Ireland (2004); the same type of model is also used by Ruge-Murcia (2007) in order to compare different estimation techniques for a different case of models misspecification. Households

choose consumption and labor/leisure and save by investing in stocks of capital. Furthermore, they maximize the following utility function:

$$U_0 = E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \log(c_t) + \gamma(1 - n_t) \right\}, \tag{3.66}$$

subject to the budget constraint:

$$c_t + i_t \leq w_t n_t + r_t k_{t-1},$$

there is no population growth and the total amount of labor is normalized to one and leisure is given by $1 - n_t$. Capital accumulates with the standard law of motion:

$$\eta k_t = (1 - \delta) k_{t-1} + i_t, \tag{3.67}$$

expressed in efficiency units in order to take into account a log-linear trend in technology η. This gives rise to the (standard) first order conditions:

$$c_t^{-1} = E_t \left[\beta(\eta + r_{t+1} - \delta) \right] c_{t+1}^{-1}, \tag{3.68}$$

$$w_t = \gamma c_t \tag{3.69}$$

The production function is given by the Cobb-Douglas:

$$y_t = a_t (\eta^t n_t)^{1-\alpha} k_{t-1}^{\alpha}, \tag{3.70}$$

where technology evolves as:

$$\log(a_t) = (1 - \rho) \log(a^{ss}) + \rho \log(a_{t-1}) + \varepsilon_t. \tag{3.71}$$

In equilibrium the real interest rate equals the marginal productivity of capital minus the depreciation rate δ:

$$r_t + \delta = \alpha a_t n_t^{1-\alpha} k_{t-1}^{\alpha-1}, \tag{3.72}$$

and the real wage rate equals the marginal productivity of labor:

$$w_t = (1-\alpha)a_t n_t^{\alpha} k_{t-1}^{\alpha}. \tag{3.73}$$

The model is closed by the market clearing condition:

$$y_t = c_t + i_t. \tag{3.74}$$

The competitive equilibrium for the economy is the sequence of prices $\{r_t, w_t\}_0^{\infty}$ and quantities $\{y_t, k_t, c_t, n_t, i_t\}_0^{\infty}$ such that firms maximize profits, agents maximize utility and all markets clear. The model is linearized by using the Taylor expansion of the system of equations ((3.67)-(3.74) around the deterministic steady state of the model. Then we solve the model following Klein (2000) and rewrite the reduced form solution into the (generalized) state space representation.

As before, we define with $\mathbf{y_t}$ the vector of observable variables of the model: we use output, consumption and hours worked. $\theta_\mathbf{t}$ is the vector of unobservable states, both the capital stock and productivity. Following Ireland (2004), we introduce in the state space model three mutually independent autocorrelated measurement errors, $\eta_t = \left[\eta_t^y,\ \eta_t^c, \eta_t^h\right]^{'}$ that are assumed to evolve as AR(1) processes

$$\begin{bmatrix} \eta_t^y \\ \eta_t^c \\ \eta_t^h \end{bmatrix} = \begin{bmatrix} \rho_{\eta^y} & 0 & 0 \\ 0 & \rho_{\eta^c} & 0 \\ 0 & 0 & \rho_{\eta^h} \end{bmatrix} \begin{bmatrix} \eta_{t-1}^y \\ \eta_{t-1}^c \\ \eta_{t-1}^h \end{bmatrix} + \begin{bmatrix} \zeta_t^y \\ \zeta_t^c \\ \zeta_t^h \end{bmatrix}$$

with

$$E\left[\zeta_t^{'}\zeta_t\right] = \begin{bmatrix} \sigma_{\zeta^y}^2 & 0 & 0 \\ 0 & \sigma_{\zeta^c}^2 & 0 \\ 0 & 0 & \sigma_{\zeta^h}^2 \end{bmatrix}.$$

We have $\mathbf{y}_t = [y_t, c_t, h_t]$ and $\theta_t = [k_t, a_t, \eta_t^y, \eta_t^c, \eta_t^h]$.

The following state space representation is then used to estimate the model:

$$\theta_{t+1} = \Phi\theta_t + \varepsilon_{t+1}$$
$$\mathbf{y}_t = H\theta_t,$$

where

$$\Phi = \begin{bmatrix} p_{kk} & [q_{ka} & 0 & 0 & 0] \\ 0 & [& F & &] \end{bmatrix} \; ; \; F = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & \rho_{\eta^y} & 0 & 0 \\ 0 & 0 & \rho_{\eta^c} & 0 \\ 0 & 0 & 0 & \rho_{\eta^h} \end{bmatrix}$$

$$H = \begin{bmatrix} m_{ck} & n_{ca} & 1 & 0 & 0 \\ m_{yk} & n_{ya} & 0 & 1 & 0 \\ m_{rk} & n_{ra} & 0 & 0 & 1 \end{bmatrix}$$

where the elements of the matrices Φ and H are obtained from the solution of the model, for example:

$$k_t = p_{kk}k_{t-1} + q_{ka}a_t,$$

$$c_t = m_{ck}k_{t-1} + n_{ca}a_t.$$

where the matrices (p, q, m, n) denote as before elements of the decision rules and in general x_{ij} denote the effect of changes of variable j into the dynamics of variable i for $x \in (p, q, m, n)$.

As it is clear from the state space model, we focus only on one special treatment of measurement errors: three autocorrelated but mutually uncorrelated measurement errors. This is consistent with Ireland (2004) which shows that such specification, compared with one having correlated measurement errors, have the best out-of sample forecasting properties in spite of less plausible values of the deep parameters. As we show below, the trade off between forecasting accuracy and realistic parameter estimation vanishes once we allow for persistent dynamics in the data.

The measurements are hours worked, consumption and GDP, taken respec-

tively from BLS data (Current Employment Statistics) and US NIPA national accounting: data run from 1948:1 to 2002:2.[21] We estimate the following set of structural parameters: $\psi \equiv [\alpha,\ \eta,\ \gamma,\ \delta,\ \rho,\ \rho_{\eta^y},\ \rho_{\eta^c},\ \rho_{\eta^h},\ \sigma_{\varepsilon},\ \sigma_{\zeta^y},\ \sigma_{\zeta^c},\ \sigma_{\zeta^h}]$ plus the level of technology a^{ss} which enters the steady state expressions of the variables. The estimated parameters are the capital share, the log-linear trend of technology, the parameter which pins down the amount of hours worked in steady state, the depreciation rate of capital, the persistence parameter of the technology shock and the measurement errors together with their standard deviations. The discount preference term β is calibrated at the value of 0.99, as standard in the literature. Differently from Ireland, we also estimate the depreciation term δ it has important implication for the dynamics of the endogenous state k_t.

3.5.2 Parameter estimates

In this section we compare full-sample estimates of the deep parameters obtained with the Generalized and the Standard Kalman Filter. Table (3.4) reports the estimation results:

Parameters	α	δ	ρ_a	η	γ	σ_{ε}
Gen. Kalman Filter	0.2917	0.0246	0.9686	0.0055	0.0041	0.0048
Kalman Filter	0.5189	0.0031	0.9999	0.0057	0.0036	0.0092
Parameters	$\rho_{\eta_t^y}$	$\rho_{\eta_t^c}$	$\rho_{\eta_t^h}$	σ_{ζ^y}	σ_{ζ^c}	σ_{ζ^h}
Gen. Kalman Filter	0.998	-0.8041	0.9995	0.0039	0.0004	0.0062
Kalman Filter	0.9999	0.9994	0.9990	0.0036	0.0001	0.0060

Table 3.4: Generalized and standard Kalman filter parameter estimates for the U.S. economy.

Since the 'true' deep parameters are unknown we can not directly asses the goodness of the estimates obtained with the standard and the generalized

[21] We use the same dataset as in Ireland (2004) which is based on the 1996 chained data

Kalman Filter. However, since some parameters have a clear structural interpretation it is possible to compare them with the evidence coming from either national accounts or previous microeconometric studies (see discussion in Prescott and McGrattan (2007). We start discussing the capital share α and the depreciation rate of capital δ: these estimates should be consistent with figures derived from national accounts data.[22] Table (3.4) shows that both the capital share and the depreciation rate estimated by the generalized filter are close to those obtained using national accounts data, while the standard filter falls short in providing good estimates. This latter finding is consistent with the results in Ireland (2004) and with the fact that in most of the literature those parameters are calibrated rather than estimated since it is believed that DSGE models fail in reproducing reasonable estimates for them. This failure can be attributed to the role of the capital share and the depreciation rate of capital in determining the persistence of the capital stock in the transition equation: larger values of the capital share and smaller depreciation rates imply higher returns on capital and lower depreciation in capital accumulation, this produces a more persistent capital dynamics. Since the capital persistence is transmitted to output and consumption through the other equations of the RBC model, such unrealistic parameter estimates might be seen as the consequence of the Kalman Filter trying to replicate the persistence of the data. As the second row of Table (3.4) shows, once we control for persistent data the RBC model is able to convey estimates which are in line with national accounts data.

With regards to the other parameters, estimation results differ substantially between the two filters.[23] There is a substantial difference in the values of both the persistence and variance parameters of the technology shock. The productivity trend η is also slightly larger than in the standard case. The measurement error on consumption is negatively autocorrelated, a result found also by Ireland (2004) when allowing measurement errors to be cross-correlated.[24] An

[22]The capital share can be derived from the labor share, computed as the average share of value added which is paid to labor (around 0.3 for US data), while the depreciation rate should square with the average ratio between investments and the capital stock (around 0.025 assuming quarterly data)

[23] We found similar results when we calibrate $\delta = 0.025$.

[24] Overall, it seems that the generalized filter produces estimates which are in line with es-

important remark concerns the high estimated value of the persistence of exogenous states (measurement errors and the technology shock) delivered by the Generalized Filter which might sound at odds with the fact that our device controls for persistence. As a robustness check, we re-estimated the model using i.i.d. measurement errors with both the Standard and the Generalized Filter: the persistence of the technology shock estimated by the Standard filter is $\rho = 0.98$ while the Generalized Filter delivers a much lower $\rho = 0.85$. This suggests that the high values of persistence shown in table 3.1 are mostly an artifact of the introduction of the autocorrelated measurement errors, rather than a problem of our method not being able to capture persistence. of parameter is filter.

To complete the section, we provide a warning against inferring too quickly (e.g. without looking carefully also at what innovation residuals have to say) that a unit root process is the correct description of exogenous states when a Standard Kalman Filter delivers an high value of persistence for the exogenous states, further discussion of the unit-root case is in section 3.7. In particular, consider the estimate delivered by the Standard Filter ($\rho = 0.9999$), shown in the second row of table 3.1, this value makes the process of technology as follows

$$a_t = 0.9999 a_{t-1} + \varepsilon_t,$$

which is undistinguishable from a unit root process, since $a_t - 0.9999 a_{t-1} \simeq \Delta a_t$. Now, if a unit root process describes correctly the dynamics of the technology, then we should expect the innovations $\{\varepsilon_{t=1}^T\}$ to be i.i.d.. However, if we estimate with a Local Whittle Estimator (Robinson(1995)) the fractional integration order d of the innovations $\{\varepsilon_{t=1}^T\}$, as reconstructed by the Kalman Filter, we get a significant negative value around -0.3, which hints to series being over-differenced.[25]

timates produced by models with less restrictions on the variance-covariance matrices of the exogenous shocks, such as the one with cross-correlated shocks in Ireland (2004).

[25] For the innovations from our generalized filter we could not reject the hypothesis that d is equal to zero.

Reconstructed residuals $\{\varepsilon_{t=1}^{T}\}$	Estimated d
Exact LW estimator	-0.30

We can get to the same conclusion by looking at the dynamics of the log detrended GDP. If we assume that in our RBC model the output y_t has the same statistical nature as the technology a_t (this is consistent with Christiano and Vigfusson (2003)), we can gauge the order of fractional integration of a_t directly from that of (log-detrended) GDP. The Local Whittle Estimator for this series is equal to 0.7:

Detrended Log-GDP	Estimated d
LW estimator	0.69
Exact LW estimator	0.71
Feasible ELW estimator	0.71
Feasible ELW estimator with detrending	0.71
2-step feasible ELW estimator	0.71
Feasible ELW estimator w/o detrending	0.71

Straightforward computations show that if $a_t \sim I(d = 0.7)$, meaning that $a_t(1-L)^{0.7} = e_t$, where $e_t \ \ i.i.d.$, it follows that:

$$a_t - 0.9999a_{t-1} \simeq \Delta a_t = (1-L)(1-L)^{-0.7}e_t,$$

which, consistently with what already shown, tells us that the first difference of productivity is overdifferenced by an order 0.3:

$$\Delta a_t(1-L)^{-0.3} = e_t, \text{that is:} \Delta a_t \sim I(d = -0.3).$$

Following Baillie (1996), long memory processes can be successfully employed to describe time series which appear as non stationary in their levels but they look over-differenced when their first difference is considered.

3.6 Forecasting

In order to assess the ability of our approach to capture the dynamic properties of the data we compare the out of sample forecast of the Generalized filter with those of the standard Kalman filter. The rationale behind this exercise follows the results in Granger and Joyeux (1980) who showed that while a short order AR representation can adequately fit long memory dynamics in sample, the forecasts produced by such AR models will not be very accurate.

The exercise is implemented as follows. We estimate the RBC model described in the previous section with both the standard and our approach for a subsample of data, precisely from 1948:1 until 1987:4. We generate out-of-sample forecasts one through four quarters ahead for each variable and compare the root-mean-squared forecast errors from the modified model to those from the standard Kalman filter. We then extend the subsample by one period and repeat the estimation and forecasting. We continue this way until all the sample is covered in 2002.

Table 4 reports the RMSE together for the forecasts generated by the standard Kalman filter (KF) and those generated by the generalized filter (GKF). In order to asses whether any difference of the two RMSE is significant we also report the statistic proposed by Diebold and Mariano (1995).[26]

The results indicate that forecasts from the modified filter significantly outperform those from the standard Kalman filter. In particular, for output, the RMSE of the generalized filter are up to 60% smaller than those of the standard filter. This shows the better performance of the Generalized Filter with respect to the normal one[27]. On one hand, the forecast from the generalized filter are more accurate than those from standard filter for all the steps ahead; on the other, the forecast accuracy of the modified filter improves, relatively to the standard filter, as the forecast horizon arises. This result indicates the presence of very persistent dynamics in the data that can not be fully captured by the autoregressive structure of the standard filter. On the other hand, it also shows

[26]The critical values for the Diebold and Mariano test have been obtained using a bootstrap procedure.

[27] Similar, albeit less striking, results hold for the case when we calibrate the δ.

Steps ahead	1	2	3	4
Output				
RMSE: GKF	0.6492	1.2294	1.8137	2.3585
RMSE: KF	0.984	1.8935	2.881	3.7871
D-M test	-6.8582***	-5.9845***	-4.8295***	-3.7659***
Consumption				
RMSE: GKF	0.4588	0.7022	1.0155	1.3689
RMSE: KF	0.5443	0.961	1.4172	1.9213
D-M test	-3.6916***	-3.7701***	-3.0075***	-2.1948**
Hours				
RMSE: GKF	0.6352	1.1373	1.6295	2.0847
RMSE: KF	0.8135	1.4583	2.0558	2.6461
D-M test	-4.8812***	-3.4673***	-2.2781**	-1.8388*

Table 3.5: Root mean square error of the forecasts from one to four quarters ahead for the standard and Generalized Kalman filter. *** is significant at 1%, ** is significant at 5% and * is significant at 10%

that by accounting for such persistence it is possible to considerably improve the dynamic properties of the model.

3.7 Comparison with unit root case

So far, we implicitly ruled out of our comparison the case where data exhibit a unit root behaviour since in that case the level of technology would have permanent memory of each innovation and our method could not be readily applied. In general terms this is not a drawback of the method we propose; if the researcher believes that the growth rates of macroeconomic variables rather than their level are persistent, then our methodology could be readily applied to the growth rates of technology rather than their levels. Nevertheless since we believe this can be of some interest, in this section we provide some evidence concerning whether a long memory consistent specification of exogenous processes can also improve the performance of DSGE models with respect to the case where the growth rate of technology is an AR(1) process (unit root case).

In the case of a unit root and considering log-variables as before, the state

space is written as:

$$\begin{aligned}
\hat{k} &= k_t - a_t && (3.75)\\
\hat{y}_t &= y_t - a_t && (3.76)\\
\hat{c}_t &= c_t - a_t && (3.77)\\
\hat{h}_t &= h_t && (3.78)\\
\Delta a_{t+1} &= \rho \Delta a_t + \varepsilon_{t+1}, && (3.79)\\
\hat{k}_{t+1} &= p_{kk}\hat{k}_t + q_{ka}\Delta a_t, && (3.80)\\
\hat{\mathbf{y}}_t &= m_{yk}\hat{k}_t + n_{ya}\Delta a_t, && (3.81)\\
a_t &= \mu + a_{t-1} + \varepsilon_{t,} && (3.82)
\end{aligned}$$

where the vector $\hat{\mathbf{y}}_t$ is now defined in detrended terms in order ensure stationarity: $\hat{\mathbf{y}}_t \equiv [\hat{y}_t, \hat{c}_t, \hat{h}_t]$. An important remark is that in this set up hours worked are a fully $I(0)$ stationary variable, in the Generalized state space approach we filtered them using the same filter as for the remaining variables.
The state space model (3.79-3.81) has now a non-stationary state variable. As common in the literature, we cope with this problem by implementing the Exact Kalman Filter as in Durbin and Koopman (2001).[28]

While we checked that our routines are able to recover deep parameter estimates for simulated data our application on true data was less impressive. First, a model with three autocorrelated measurement errors such as our benchmark and Ireland's seems to be not identifiable with unit roots, so that our results concern only a case with i.i.d. measurements which displayed no identification problem.[29] Second, parameter estimates show a reverse pattern with respect to

[28] In particular we use their 'univariate version' in which the updating step is repeated by adding one measurement at the time and the information set is expanded accordingly. This set-up more easily allows to mix measurements which are cointegrated with those which are not (hours). In the case of cointegration, in our case between output and consumption, the D-K filter uses the first observation of the first ordered cointegrated variable (output) to initialize the level of the technology state; this value is then used in the other cointegrated variables (consumption).

[29] This might not be seen as problematic from the point of view of the out-of-sample forecast horse race. When assessing whether a model with unit roots is able to beat one which allows for long memory one would want to compare the best specifications for the two categories: with this respect there is no logical inconsistency in comparing a long memory version having auto-

what we obtained with standard procedure, we report them below for our full sample:

Parameters	α	δ	ρ_a	η	γ	σ_ε
Gen. Kalman Filter	0.2917	0.0246	0.9686	0.0055	0.0041	0.0048
Unit Root Filter	0.1871	0.1073	0.7317	0.0050	0.0047	0.0050
Parameters	$\rho_{\eta_t^y}$	$\rho_{\eta_t^c}$	$\rho_{\eta_t^h}$	σ_{ζ^y}	σ_{ζ^c}	σ_{ζ^h}
Gen. Kalman Filter	0.998	-0.8041	0.9995	0.0039	0.0004	0.0062
Unit Root Filter	-	-	-	0.0197	0.0001	0.0540

Table 3.6: Full Sample parameter estimates: Generalized Kalman Filter against Unit Root.

The capital share is in this case lower than consensus estimates and the depreciation rate is far higher; as we argued before these can be interpreted as signs of having 'too much persistence' in the data generating process.

Concerning out-of-sample forecasts they were acceptable for cointegrated variables, e.g. consumption and output, even if they did not outperform any of previous procedures (from 2 to 4 percent is the RMSE of output, around 1 per cent for consumption). Concerning hours worked the root mean square errors raised by an order of ten with respect to previous procedures. Consistently with our parameter estimates hours worked appear to be mainly 'driven' by measurement errors.

Steps ahead	1	2	3	4
Output				
RMSE: Unit Root Model	2.4087	3.1181	3.6301	4.0673
Consumption				
RMSE: Unit Root Model	1.0172	1.0201	1.0406	1.3298
Hours				
RMSE: Unit Root Model	9.6450	10.0057	10.1787	10.2798

Table 3.7: Root mean square error of the forecasts from one to four quarters ahead for the unit root model.

correlated measurement errors and the unit root having i.i.d. measurements. Anyway, as we performed out-of sample forecasts with the generalized filter and no autocorrelated measurement errors, we saw that results were not overturned.

The inability of tracking hours worked with a unit root technology shock is not surprising, given previous literature. Many studies using technology as permanent component found that it has a limited explanatory power on variables which are modelled as stationary, e.g. hours worked in our case. Early findings are King, Plosser, Stock, and Watson (1991) which show that the ability of unit root technology shock to explain the dynamics of variables which are not cointegrated with GDP is limited. The same result is found in more recent studies (for example the New Area Wide Model used at the European Central Bank) where the permanent component of technology usually plays a limited role in the variance decomposition of non cointegrated variables and in particular of hours worked. This is probably related to the strong restriction a unit root technology shock imposes on the data generating process, since the level of technology provides the common component of cointegrated variables, while its growth rate explains the stationary components (such as hours worked). When technology is the only shock present in the model the unit root assumption implies a very tight relation between the variance of the cycle and that of the trend which might be seen as too restrictive. A further remark concerns the statistical nature of hours worked, which is assumed to be stationary in the conventional approach both when technology is a unit root and when it is modelled as stationary. Empirical studies show that hours worked have non-clear cut statistical properties, (e.g. they can be integrated of order one, zero or long memory, see Gil-Alana and Moreno (2006)). In our Generalized Kalman Filter procedure, we account for the possibility of long memory behaviour, as we filter hours in the same way as the other variables, that is we allow for a 'common long range component' between GDP and hours worked.

Overall, results in this section suggest that a long memory specification, which is in between full stationarity and pure unit root, can be preferred over the two extremes. We are well aware that in a general equilibrium setting our conclusions are likely to depend on assumptions concerning the structure of the DSGE model at hand, as already found by Chang, Doh, and Schorfheide (2007), which show that the inference one can make concerning whether shocks are I(0) or I(1), depends upon the structure of the model at hand, therefore we do not claim generality with this remark.

3.7.1 Comparison with unit roots: simulated data

In this section we provide a further experiment using simulated data. We simulate data from the stationary model described by Ireland (2004), the same as in section 3.5 and we add a long memory component as described in section 3.4. We then use the simulated dataset to estimate a model with a unit root in technology such as (3.75-3.82). This complements the analysis of section 3.4. We do not use the same simple Ramsey model, as we got serious identification problems with it.[30] To run this experiment it is also crucial that the signal-to-noise ratio (e.g. the standard deviation of the technology shock relative to those of measurement errors) is set at values which are plausible with respect to those estimated on real data, when the signal-to-noise ratio is too high the optimizer simply fails (e.g. lack of identification) to estimate the unit-root model. To sum-up, the model used to simulate data is the one described by Ireland (2004), adding three i.i.d. measurement errors and the model used for estimation is its unit root version described by (3.75-3.82). The calibration of parameters to be estimated is as follows:[31]

$$\alpha = 0.33,\ \rho_e = 0.99,\ \delta = 0.025,\ \sigma_e = 0.002,\ \sigma_i = 0.001,$$

where α is the capital share in production ($1-\alpha$ is the labor share), δ is the depreciation rate, ρ_e, σ_e denote the persistence and the standard deviation of the growth rate of technology, while $\sigma_i, i \in \{y, c, h\}$ is the standard deviation of the three i.i.d. measurement errors, on output, consumption and hours worked. Results are means from 200 samples;[32] to save on space we only report the cases $d = 0, 0.1, 0.4, 0.6, 0.9$. The $d = 0$ case is relevant as in this case the estimated model is somehow misspecified with respect to the true Data Generating Process since we impose a unit root on stationary data.

[30] Results were cross-checked by using Dynare 4.0.2

[31] No substantially different results were found when the β was included in the estimated parameters. Also, in this simple example data are considered in deviation from their steady state, then we do not estimate a drift η nor the parameter which pins down the steady state level of hours worked γ.

[32] We eliminate those very few samples in which the optimizer got stuck at some bound; the sampling distribution looks quite symmetric and no important difference can be found from using median or means

Parameter	d = 0	d =0.1	d=0.4	d=0.6	d=0.9
α	0.4021	0.4012	0.4220	0.4719	0.5941
ρ_e	0.0425	-0.0198	0.1527	0.5439	0.8444
δ	0.0589	0.0351	0.0359	0.0332	0.0295
σ_c	0.0002	0.0002	0.0004	0.0012	0.0094
σ_y	0.0029	0.0027	0.0022	0.0013	0.0011
σ_h	0.0027	0.0029	0.0060	0.0159	0.0419
σ_e	0.0016	0.0020	0.0035	0.0053	0.0090

Table 3.8: Deep parameter estimates : data generating process is stationary (plus long memory), estimating model embeds a unit root in technology

The first worthy remark is that even when no long memory is introduced in the DGP, the unit root model produces substantial bias in the estimates.[33] This appears to be rather counterintuitive, as it occurs even if the persistence of the exogenous parameter is set at a very high value, close to a unit root; dramatic results, not shown to save space, appear also for a lower $\rho = 0.95$. After some trials we found that, as highlighted in the previous section, one serious problem seems to be the presence of a stationary time series among the measurements, hours worked. The true DGP delivers a relatively high amount of persistence in hours due to the presence of the level of technology, $a_t = 0.99a_{t-1} + \varepsilon_t$, in their reduced form equation; in the unit root model hours worked are instead a function of the first difference of technology, whose law of motion is given by $\Delta a_t = \rho_e \Delta a_{t-1} + \epsilon_t^u$. Given the true exogenous state process, a correct estimate should be $\rho_e = 0.99 - 1 = -0.01$ but such a low value would make persistence in hours worked very low in the unit-root model. When a long memory component is added to the innovations of the DGP this makes hours even more persistent and, as d goes high and it tends to exacerbate the problem. It is not surprising then that estimates of ρ_e are upward biased and so are those for the other parameters, especially the capital share, making capital to be more per-

[33] A potential critique is that the direction of bias shown here is not entirely consistent with what we have found in the previous section on real data. While this should deserve further enquiry in the future, we do not claim that a simple stationary RBC with a fractional noise in the innovations is the True Data Generating Process for US data, so we do not see this potential criticism as really critical

sistent. When removing hours worked from the set of measurements, estimates improve (not shown) but they still suffer from some upward bias. A plausible, though non exhaustive, explanation for this pattern is that the unit root model delivers deviations of cointegrated variables from the technology level (their trend) which are more limited and short-lived than the DGP. To see that, using the reduced form of the unit-root model we can write the deviation of the level of capital from its trend as:

$$k_t - a_t(1 - p^u_{kk}L) = q^u_{aa}\Delta a_t,$$

where p^u, q^u are the solution of the unit-root model and L is the lag operator. Under the true DGP Δa_t should almost be a white noise, as ρ_e is very low. Plugging the true values of the remaining deep parameters, we see that the resulting p^u, q^u give a more short lived variation of the term $k_t - a_t$ with respect to the DGP (this is also consistent with the findings of Rotenberg and Woodford, see below).

Overall, our results from this and the previous section suggest that even when the data generating process has an exogenous process which is very close a unit root ($\rho = 0.99$) a model which has unit-root in technology behaves fundamentally in a different way from a stationary model; furthermore it is impaired in reproducing the behaviour of the stationary time series. This is consistent with most of the finding in previous literature, coming from disparate sources: Rotemberg and Woodford (1996) find that an RBC with unit roots cannot explain the variability of the predictable component of the business cycle. Other researchers such as Canova (2009) found that the variance decomposition of a shocks changes dramatically when either a unit root in technology is introduced or data are linearly detrended. This is related to our results, in the sense that when the unit-root is introduced it will be less able to capture cyclical fluctuations and it will make the other shocks look more important in the variance decomposition.

3.8 Conclusion

In this paper we have shown a simple method to bring general equilibrium dynamic models to persistent data. This can be done by allowing one (or more) common factor to have a long memory behaviour. By doing this one can improve the plausibility of estimates and the out-of-sample forecast performance with respect to both the pure stationary and the pure unit root case: we report an average (significant) reduction in the forecast error of about 30%.

Our method does not alter the cross equations restrictions implied by the DSGE model, but it rather applies them to observations which have been detrended. Contrary to other studies, the detrending procedure is not assumed a priori but it is directly derived from the data in a way which is consistent with the structure of the model, e.g. by exploiting information which the model have already deemed as not useful and put into innovation residuals. With this respect our method also differs from other methods proposed for example by Canova (2009), where an hybrid model (DSGE plus filter) is estimated.

In simulation experiments we have shown that persistent data can have important effects in the way the Kalman filter reconstructs unobserved endogenous states. We also provide evidence that the introduction of a stochastic trend as a unit-root imposes is a constraint the data generating process which can come at some cost and should not be seen as a general way to cope with persistence. We leave the full development of this line of research to further work.

Our contribution is not limited to DSGE models but it can be used any time long memory data and latent component are present; moreover, while here we apply it to the case of a simple univariate unobserved shock, it can be readily applied to a multivariate framework. Simulation results are provided for the case of ARFIMA shocks, the method is also able to cope with general long memory processes, since it does not cast any strong assumption concerning the nature of the long memory dynamics, whether it is a fractionally integrated process or a different long memory process.

As already mentioned in the introduction, there is a widespread evidence about long memory behaviour of many macroeconomic time series, for example inflation; this would be a call for monetary and New Keynesian Models. Since

these are mostly estimated by Bayesian methods we leave it to further research to extend our framework to that case.

3.9 Appendix: Long memory processes

Long memory processes have been extensively studied in time series analysis and a good review of their properties can be found in Robinson (2003) and Baillie (1996). Long-memory processes are defined by autocorrelations decaying slowly to zero or spectral density displaying a pole at zero frequency. For instance, compared to ARMA processes, they are have a autocorrelation function that decays hyperbolically rather than exponentially. In this respect, long memory processes represent an alternative representation to the knife-edge distinction between I(0) and I(1) processes. A well known class of long memory processes, introduced by Granger (1980), is the Autoregressive Fractional Moving Average Process of order d (ARFIMA(p,d,m)), which is defined as

$$(1-L)^d \phi(L) y_t = B(L)\varepsilon_t, \tag{3.83}$$

where ε_t is a white noise process with variance σ^2, $\phi(L)$ is a lag polynomial of order p, $B(L)$ is a lag polynomial of order m and $d \in [0,1]$ indicates the fractional differencing parameter. As the parameter d can take any value in the interval between 0 and 1, ARFIMA processes fill the gap between a strictly stationary process, when $d = 0$, and a unit root process, when $d = 1$.

A special case of (3.83), which is the basic building block of fractionally integrated processes is the so called 'fractional white noise', defined by setting $p, m = 0$ in (3.83). For $0 < d < 0.5$ the process is "stationary" with hyperbolic rather than exponential decay of the autocorrelation function (ACF heretofore); when $0.5 < d < 1$ the process is non stationary, i.e. the squared sums of its autocorrelations do not converge. However, differently from unit root processes, it is still mean reverting. Finally, when d is negative the ACF presents an oscillatory behaviour around zero, which is the typical pattern we can observe in 'overdifferenced' time series. To further understand the difference between ARMA and ARFIMA processes, we plot the ACF of two autoregressive processes with roots

respectively equal to 0.9 and 0.99, and the ACF of a fractional noise with $d = 0.4$. As the number of lags increases, the latter decays at very slow hyperbolic rates implying very different statical properties compared to the AR models.

Going more into the technical details, the term $(1-L)^d$ in 3.83 is called the fractional differencing polynomial: the term 'fractionally integrated processes' denotes the family of processes generated by the application of a fractional differencing polynomial. For $d < 0.5$ the fractional differencing polynomial $(1-L)^d$ can be represented as in Granger (1980):

$$\begin{aligned}(1-L)^d &= \sum_{j=0}^{\infty} \frac{\Gamma(j-d)(-1)^j}{\Gamma(j+1)\Gamma(-d)} L^j \simeq \sum_{j=0}^{\infty} j^{-(d+1)} \frac{(-1)^j}{\Gamma(-d)} L^j \\ &\equiv \sum_{j=0}^{\infty} \pi_j L^j\end{aligned}$$

where $\Gamma(\cdot)$ is the Gamma function and the approximating term derives from a first order asymptotic expansion. Therefore, any stationary fractional noise y_t can be rewritten in an infinite order autoregressive representation (AR(∞)),

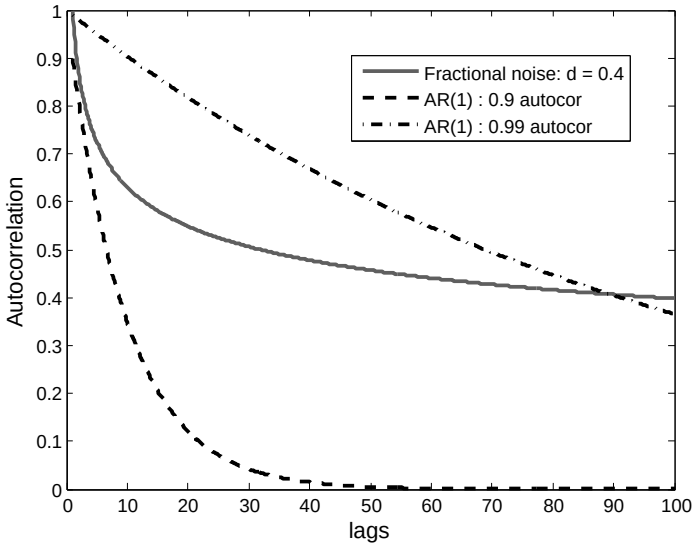
$$(1-L)^d y_t = \varepsilon_t \tag{3.84}$$

$$\sum_{j=0}^{\infty} \pi_j L^j y_t = \varepsilon_t \tag{3.85}$$

$$y_t = \sum_{j=1}^{\infty} \pi_j y_{t-j} + \varepsilon_t, \tag{3.86}$$

As Granger (2001) and Granger and Joyeux (1980) point out, even if it possible by using standard identification criteria, to fit an ARIMA model on long memory generated data, this would lead to a poor approximation of the true DGP and inaccurate out-of-sample forecasts. In particular, Granger and Joyeux (1980) show that short order autoregressive processes are not flexible enough to successfully approximate long memory behaviour and that statistical criteria have a very hard time in selecting processes with enough lags.

Figure 3.1: ACF of two ARs compared to a stationary fractional noise $d = 0.4$ process



Chapter 4

Bayesian prior elicitation in DSGE models: macro vs micro priors

4.1 Introduction

Since the seminal contributions by Sims and Zha (1999), Schorfheide (2000) and Smets and Wouters (2003), Bayesian estimation methods have gained ground as a very attractive alternative over classical methods in the field of dynamic stochastic general equilibrium models (DSGE), among both academicians and practitioners. Unlike in the frequentist approach, a bayesian researcher uses both information from the available data and prior knowledge, in order to provide so called posterior estimates. An important point in the bayesian methodology relates to how the researcher elicits prior information and the way the latter influences posterior estimates.

For DSGE models, prior knowledge is normally formulated as independent distributions, one for each of the structural parameters. Often those are informed by using previous microeconometric studies. Hence, the resulting joint prior distribution has independent components, at least as long as microeconometric studies are undertaken independently for each deep parameter. In the following, we will refer to this approach as 'microprior'.

The microprior selection scheme is very simple and ready to implement, but suffers from several shortcomings. First, a priori independence of the deep pa-

rameters entails implicit assumptions for the a priori beliefs on data moments and the responses of macroeconomic variables after some shock (e.g. impulse response functions, IRFs heretofore) which may indeed be at odds with prior knowledge which draws from macroeconometric studies. Second, for some parameters, microeconometric information can be abundant (e.g. the frequency at which firms adjust their prices, a direct measure of nominal rigidities in the economy), while for others it can be scant, (e.g. concerning the average size of some shock or its persistence). At the same time there might be abundant evidence concerning the behaviour of macroeconomic variables, such as its response after some shock occurs (IRFs), that researchers might want to exploit.

Based on such considerations, Negro and Schorfheide (2008)– DS heretofore – pioneered a different approach in which prior distributions are formed on the basis of a priori beliefs on macroeconomic indicators (for example simulated moments or IRFs) rather than on the deep parameters; we will label this approach as 'macroprior'. A macroprior implies that the a priori belief on deep parameters is described by a joint distribution whose dependence structure depends on the mapping between the deep parameters and the macroeconomic indicator at hand. More specifically, DS propose an hybrid approach in which a macroprior is used for the subset of parameters concerning the exogenous states, e.g. the persistence of exogenous shocks and their standard deviations and a microprior for the other parameters such as those related to the amount of nominal rigidities in the economy.[1] In the context of a Smets-Wouters type of model, DS(2008) find that posterior estimates, in particular those related to the nominal rigidities in the economy, which are *not* informed with the macroprior, change with respect to the case when only micro priors are used.

The contribution of this paper is twofold. First, we want to clarify if and how the dependence structure of macropriors shape posterior estimates, as hinted by the results in DS (2008) and in particular how estimates of nominal rigidities are affected by the prior. Second, we find it convenient to do that by introducing a new type of macroprior, elicited in a non parametric way from impulse response functions (IRF-prior heretofore). Concerning our IRF-prior, we build a

[1]For simplicity we omit the description of what DS (2008) introduce as priors on parameters which determine the steady state of the model since those are not essential for our discussion.

controlled experiment, where we constrain the prior to have the same location (e.g. the mean/mode) as in the microprior. By doing this, any difference in posterior estimates can then be attributed to the second order properties of the IRF-prior, e.g. the dependence structure of parameters and its spread. The latter is controlled by the researcher on a joint set of parameters but its information content can vary across parameters, depending on how much each single parameter is important in determining the shape of the impulse response functions; as a (limit) example consider a case in which a change in one parameter does affect impulse responses (lack of identification); in this case our prior would not be informative on this parameter.[2]

We compare the way our prior's dependence structure shapes posterior estimates with what is obtained with both a microprior and the DS (2008) prior. A further advantage of the IRF-prior, which however we do not fully exploit in this paper, is that, differently from the deep parameters and, albeit to a lesser extent, from the sample moments used in DS (2008), researchers and policy makers often have a clear prior view of what should be the behaviour of impulse response function without explicitly recurring to a data-based presample, but directly drawing from abundant previous (VAR) studies.

A preview of our results is as follows. When an IRF-prior is applied to exogenous state parameters we find that posterior estimates can differ with respect to the case of the benchmark independent case by a little amount. This happens when the joint prior is set to be rather tight, while the effects are negligible when the overall variance of the prior is larger. This differs from the findings of DS (2008) which report large differences in the posterior estimates: we apply their method to our case and we are able to explain how this difference is produced. The second finding is that when applying the IRF-prior to a different set of deep parameters posterior estimates change more widely. As a third result, by recurring to an exercise on simulated data, we highlight the fact that this last feature is mainly due to model misspecification.

The paper is structured as follows, in section (4.2) we describe the general

[2]Due to the non linear transformation involved in the solution of the DSGE model, imposing the same spread (for example the variance) on two priors with different shapes would not strictly guarantee the same degree of informativeness, therefore we do not try to control for that, but we rather check ex-post, from the posterior distribution, if and how a prior was informative

formulation of our IRF-prior. In section (2.3) we introduce the DSGE model used for our experiments. Section (4.3.1) details the estimation strategy we follow and (4.4) shows posterior estimates with our benchmark prior. In section (4.5) the IRF-prior is applied to the block of parameters which pertain to exogenous states. In subsection (4.5.2) we contrast our results with those obtained with the DS prior. In section (4.6) we present changes in posterior estimates obtained when a IRF-prior is applied to a different block of deep parameter and we discuss whether this might be due to misspecification vs weak identification. Some conclusions follow. A discussion of more technical parts, such as the 'intractable normalizing constants', the DSGE model specification and a presentation of DS(2008) method are relegated to the appendix.

4.2 Impulse Response Priors

The reduced form of a (linearized) DSGE model can be written in state space form (see An and Schorfheide (2007)):

$$X_t = F(\psi)X_{t-1} + G(\psi)\epsilon_t, \tag{4.1}$$

$$\epsilon_t \sim N(0, \Sigma_\epsilon), \tag{4.2}$$

$$Y_t = HX_t, \tag{4.3}$$

the transition equation corresponds to the solution (e.g. the reduced form) of the model. The n-dimensional vector of states X_t collects all variables which are known to economic agents (households, firms and government) at time t (see Sims (2002)). X_t includes so called exogenous states (e.g. total factor productivity), which follow, each one, a scalar autoregressive process. ϵ_t denotes the l-dimension vector of i.i.d. normally distributed (and uncorrelated) structural shocks, they appear either as innovations of exogenous state processes and/or directly in one or more other equations. Their standard deviation is a diagonal matrix Σ_ϵ, or, as vector, σ_ϵ. Both the $(n \times n)$ F matrix and the $(n \times l)$ G matrix in the transition equation are functions of ψ, the vector of so called deep (or structural) parameters (e.g. parameters in the utility functions of the agents

or parameters related to technology). The vector ξ gathers both σ_ϵ and ψ:

$$\xi \equiv [\psi, \ \sigma_\epsilon].$$

For simplicity we will refer to ξ as the vector of 'deep parameters' with a slight inconsistency with previous terminology.

The $l \times n$ selection matrix H links states X_t to those variables (Y_t) which are observed by the econometrician. As common in the literature, we also assume that the number of observed variables is equal to the number of shocks. Y_t is then a vector of dimension l.[3]

Referring to variable Y, its time t impulse response function, j periods after a shock (ϵ_t) occurs, is defined as the difference between the j-steps ahead projection of Y_t, conditional on both the $t-1$ information (I_{t-1}) and the knowledge of ϵ_t, and the same j-steps-ahead projection without the knowledge of ϵ_t (Koop, Pesaran, and Potter (1996)):

$$\gamma_{t,j} \equiv E[Y_{t+j} \mid \epsilon_t = \sigma_\epsilon, I_{t-1}] - E[Y_{t+j} \mid I_{t-1}], j = 0, \ldots, m \tag{4.4}$$

where, as convenient, we conditioned the shocks on the their standard deviations. In our linear state space model (4.1-4.3), following definition (4.4), the impulse $\gamma_j(\xi)$ at horizon $j = 0, \ldots m$ reduces to:

$$\gamma_j(\xi) = HF^j(\psi)G(\psi)\Sigma_\epsilon, \ j = 0, 1, \ldots m, \tag{4.5}$$

$\gamma_j(\xi)$ is a $l \times l$ matrix. Stacking all horizons we denote:

$$\gamma(\xi) \equiv [\gamma_0(\xi) \ldots \gamma_j(\xi) \ldots \gamma_m(\xi)], \ 0 < \forall j < m,$$

which is a $l \times l \times (m+1)$ matrix. We follow this notation in non-ambiguous cases.

Our non parametric prior is constructed as follows. For a given ξ, compute a distance function between the impulse $\gamma(\xi)$ and some target impulse γ^*. The

[3]This assumption is not important to define impulse responses while it plays a relevant role in the estimation process, as it is not possible to run the Kalman Filter on a stochastically singular system.

distance is denoted with $d(\gamma(\xi), \gamma^*)$ and it maps the $l \times l \times (m+1)$ space (measurement, shock and impulse horizon) into the positive real line. Our prior kernel w is given by a logistic transformation of the distance: [4]

$$w(\xi \mid \gamma^*, K) = \frac{exp(-d/K)}{K(1 + exp(-d/K))^2}, \tag{4.6}$$

where K is the variance of the logistic and it controls for our degree of belief in the prior. A proper prior probability distribution can be obtained by normalizing the kernel:

$$\bar{w}(\xi \mid \gamma^*, K) = \frac{w(\xi \mid \gamma^*, K)}{\int w(\xi \mid \gamma^*, K) d\xi}, \tag{4.7}$$

but the knowledge of the normalization constant (e.g. the denominator of 4.7) is not required since the model is estimated by MCMC techniques. As the prior kernel is directly used in the estimation process in the rest of the presentation we omit the proper prior $\bar{w}$ and we directly express formulas in terms of the kernel w.

Using a penalty function draws upon the idea of a distance between impulses, already in use for estimating DSGE models (see Christiano, Eichenbaum, and Evans (2005)). The distance we use is a quadratic function of the type used in impulse response matching:[5]

$$d(\xi \mid \gamma^*) = vec(\gamma(\xi) - \gamma^*)' \, W \, vec((\gamma(\xi) - \gamma^*)), \tag{4.8}$$

where W is an $(l^2 * (m+1)) \times (l^2 * (m+1))$ identity matrix and vec denotes the vec matrix operator.[6] A branch of research which is closely connected to

[4] Different functions can also be used; for example, we experimented the inverse of the distance, with broadly similar results. However, the inverse function is not continuous on a measure zero set, e.g. when the distance is exactly zero. In practice we found it to be not a problem unless the parameter space of the prior is of dimension one. Yet, in this case the sampler can get stuck at the point of discontinuity.

[5] We experiment also different functions: a sup distance; and the function described by Uhlig (2005) in the context of sign restrictions for VAR models. Since we have not found strikingly different results using different distance measures, we present the quadratic function case, except when stated differently.

[6] In order to avoid the criticism of Canova and Sala (2009) that the surface described by the distance between impulse responses can be very ill-behaved due to the non linearity of IRFs, we set $m = 1$, that is we use only the impact of the impulse response and the first lead; results

our approach is the literature known as Generalized Likelihood Uncertainty Estimation, within the framework of global sensitivity analysis, where some measure of distance is also used to inform distribution measures (see Saltelli, Tarantola, Campolongo, and Ratto (2004) page 182 and following).

Formula (4.6) defines a joint prior kernel over the whole vector of parameters ξ, but, as explained in the introduction, the best use of prior knowledge of researchers might entail using a macroeconomic prior only on a subset of parameters, ξ_{macro} . Partitioning the vector of parameters ξ into two parts,

$$\xi \equiv [\xi_{macro},\, \xi_b],$$

only ξ_{macro} is informed by IRF-priors and ξ_b is the subvector informed by independent distributions $p(\xi_b \mid \xi_b^*, \Sigma_b^*)$. Hyperparameters (location and spread) are now denoted ξ_b^*, Σ_b^*; were the benchmark prior applied also to the $macro$ parameters, its hyperparameters would be denoted as $\xi_{macro}^*, \Sigma_{macro}^*$. The 'hybrid' prior over the whole vector of parameters, $p(\xi|\xi_b^*, \Sigma_b^*, \gamma^*, K)$ can now be defined as (proportional to) the product of the kernel of the IRF-prior conditional to ξ_b, times an independent prior on ξ_b:

$$p(\xi|\xi_b^*, \Sigma_b^*, \gamma^*, K) \propto w(\xi_{macro} \mid \xi_b, \gamma^*, K) p(\xi_b|\xi_b^*, \Sigma_b^*), \tag{4.9}$$

where the conditional distribution is computed from the ratio between joint and marginal IRF prior:

$$w(\xi_{macro} \mid \xi_b, \gamma^*, K) = \frac{w(\xi_{macro}, \xi_b \mid \gamma^*, K)}{w(\xi_b \mid \gamma^*, K)}. \tag{4.10}$$

Expression (4.9) is of little use for estimation: while the numerator in (4.10) is computable, the denominator of (4.10) cannot be analytically computed and, by the same token, it varies over draws of ξ_b, making it unsuitable for MCMC estimation. This is a (well) known issue in the MCMC literature as the 'intractable normalizing constant problem', see Moeller, Pettitt, Berthelsen, and Reeves (2006) and our appendix for a more detailed presentation. DS(2008)

are in general robust to using $m < 5$.

propose to approximate the conditional distribution at hand by the joint IRF-distribution (i.e. the numerator of 4.10, which is known), where the conditioning parameters ξ_b are kept fixed at some value ($\xi_b^?$):

$$w(\xi_{macro} \mid \xi_b, \gamma^*, K) \approx w(\xi_{macro}, \xi_b = \xi_b^? \mid \gamma^*, K). \tag{4.11}$$

Introducing the calibrated set of parameters $\xi_b^?$, equation (4.9) is replaced by the following one:

$$p(\xi_{macro}, \xi_b = \xi_b^? \mid \xi_b^*, \Sigma_b^*, \gamma^*, K) \propto w(\xi_{macro}, \xi_b = \xi_b^? \mid \gamma^*, K) p(\xi_b \mid \xi_b^*, \Sigma_b^*), \tag{4.12}$$

where now the full expression for the kernel is written as:

$$w(\xi_{macro}, \xi_b = \xi_b^? \mid \gamma^*, K) = \frac{exp(-d(\xi_{macro}, \xi_b = \xi_b^? \mid \gamma^*, K)/K)}{K(1 + exp(-d(\xi_{macro}, \xi_b = \xi_b^? \mid \gamma^*, K)/K))^2}. \tag{4.13}$$

Equations (4.12) and (4.13) together with the definition of the distance (4.8), as modified to take into account the partitioning of the set of deep parameters, define the IRF-prior we will be using in the remainder of our analysis. The subvector ξ_b plays now a double role:

1. ξ_b is informed with an independent prior (microprior) $p(\xi_b \mid \xi_b^*, \Sigma_b^*)$.
2. It is calibrated in the IRF-prior (and in the DS prior) at some value $\xi_b^?$.

In the analysis in DS(2008), based on sample moments rather than IRFs, the same type of calibrated parameters appear.[7] From their analysis we have little guidance on how much point 1 vs. 2 above drive posterior estimates. Also it is not clear how much having a non independent prior matters for results. To shed light on those issues, in the following we run a set of experiments with our prior and we cross-check some of our results by using the DS(2008) methodology. In the first set of experiments we apply the IRF-prior on the exogenous state parameters (section 4.5) as in DS(2008), in the second one (section 4.6) we apply it to a different set of deep parameters more related to the structure of the economy, such as the amount of nominal rigidities; more details of the selected

[7] More details about DS are provided in the appendix

parameters are offered to the reader at the end of section 2.3. In order to control for the effects of point 2 in both experiments we use a target impulse γ^* derived from the DSGE model itself

$$\gamma^* = \gamma(\xi^*_{IRF}, \xi_b = \xi^?_b),$$

as ξ^*_{IRF} is the same location hyperparameter as in the microprior, we also focus on the role of second order properties (spread and correlation structure) of the macroprior in shaping the posterior distribution. As we check, once calibrated parameters are controlled for, posterior estimates do not vary much as calibrated parameters are considered at different $\xi^?_b$.[8] In section 4.5.2 we show how the DS prior is sensitive to calibrated $\xi^?_b$. In particular, when comparing results for the benchmark and the DS prior, the two posteriors come closer when also $\xi^?_b$ is closer to the posterior mode obtained under the benchmark prior. Our results show that the introduction of calibrated parameters is not at all innocuous and it is the main determinant of the large impact of prior knowledge on posterior estimates as found by DS(2008).[9] We now turn to describe the DSGE model.

4.3 The DSGE Model

We use a New Keynesian model which features Calvo pricing with indexation, no capital, no habits in consumption and no wage rigidities; the model follows the description in Rabanal and Rubio-Ramirez (2005) – RR heretofore. Some further details concerning the model and the motivation for choosing a small scale model are provided in appendix 4.8. The set of its structural parameters is given by:

[8] In a more general case, when γ^* is taken from VAR studies or when pre-sample information is considered, as it is the case of the DS prior, this is not likely to be the case. Further research is in progress by the authors on this topic.

[9] Our prior could be computed for a given set of points, integrated with respect to the subvector ξ_b and then an interpolation scheme might be used to fill missing points during the estimation procedure. Due to interpolation this procedure turned out to be very unstable from a numerical point of view and we abandoned it.

$$\xi \equiv (\beta, \epsilon, \delta, \sigma, \theta_p, \gamma, \omega, \rho_r, \gamma_y, \gamma_\pi, \rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_p),$$

β is the discount factor, ϵ is the elasticity of substitution among varieties in the bundle of commodities, δ is the share of capital in the Cobb Douglas production function, σ is the elasticity of intertemporal substitution and θ_p is the (Calvo) probability of price adjustment, which lies in a bounded space (0,1). γ is the elasticity of labor supply, ω is the degree of backward looking indexation in the Phillips curve, $\rho_r, \gamma_y, \gamma_\pi$ are the parameters concerning the Taylor rule (interest rate inertia, output term and inflation term). The rest of parameters pertain to exogenous state variables, technology (a) and preference shocks (g) which are autocorrelated (ρ_a, ρ_g), while the monetary shock (m) and the mark-up shocks (p) are modeled as i.i.d.

Some parameters which are known to be difficult to pin down, i.e. either because not separately identifiable from other model parameters or are hard to estimate with detrended data, are calibrated at values which are standard in the literature (see R-R 2005). The calibrated terms are given by $\beta = 0.99, \delta = 0.36, \epsilon = 6$. Since both the elasticity of substitution σ and the Calvo probability θ_p lie in a bounded space (i.e. between zero and one), we transform them in order for our algorithms to search in a larger space. We define:[10]

$$\sigma^{-1} \equiv 1/\sigma \ ; \ \Theta_p \equiv \frac{1}{1-\theta_p}.$$

The transformation at hand is also convenient from an economic point of view. Transformed variables are, respectively, the risk aversion of agents and the average amount of time that it takes until a price is revised. As it is common practice in literature we formalize our priors as functions of the transformed parameters.[11]

The vector of estimated parameters ξ is then reduced to:

[10] This transformation greatly improves the performance of the algorithm, by lowering the condition number of many of the matrices involved in the Kalman filter.

[11] About the average duration of prices, there is also microeconometric evidence readily available, therefore forming a prior on the transformed variables is also sound from an economic point of view.

$$\xi \equiv (\sigma^{-1}, \Theta_p, \gamma, \omega, \rho_r, \gamma_y, \gamma_\pi, \rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_p),$$

and exogenous state parameters are the last six terms.

In the next section we sketch the general estimation procedure. As we run two different experiments with IRF-priors, in each one we inform a different block of parameters by our macroprior. These blocks are respectively given by: [12]

1. The exogenous states parameters block, as in Negro and Schorfheide (2008):

$$\xi_{macro} \equiv [\rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_p].$$

   For this case the application of the DS prior is also discussed.

2. Parameters non related to the exogenous shocks, with the exception of the intertemporal elasticity of substitution which we found hard to identify when a joint prior is used:[13]

$$\xi_{macro} \equiv [\Theta_p, \gamma, \omega, \rho_r].$$

4.3.1 Estimation

From the model solution we write a state space model as $(4.1 - 4.3)$, by using the reduced form of the model as transition equation. The state vector X_t is given by real wages (rw_t), inflation (π_t), GDP (y_t), the FED funds rates (r_t); two autoregressive processes of order one: a preference shock process (g_t) and a technology shock process (a_t). The vector of i.i.d. shocks ϵ_t is given by the innovations of the two the AR(1) processes (technology and preference shock) plus monetary and mark up shock (see the appendix for more details). The

[12] We run a third experiment concerning only parameters in the Taylor rule, but we omit it since it yields no different results with respect to the benchmark prior.

[13] In linear models this parameter can be difficult to identify, see the paper by An and Schorfheide (2007)

measurement equation is given by the following:[14]

$$\text{Real Wage} = 100\, rw_t,$$

$$\text{Real Output Growth}\% = 100(y_t - y_{t-1}),$$

$$\text{Annualized inflation rate} = 400\pi_t \equiv 400(\ln P_t - \ln P_{t-1}),$$

$$\text{Annualized interest rate} = 400\, r_t,$$

where measurements are on the left hand side, their model counterparts on the right hand side. To form the likelihood of the model we use the Kalman Filter innovations: its implementation follows the description in Durbin and Koopman (2001). Since an analytical form of the posterior distribution cannot be obtained, we use Monte Carlo Markov Chain methods to draw from it,[15] in the specific, a Random Walk Metropolis Hastings algorithm.

A general description of the estimation procedure with the different priors is summarized below; we denote differently only those steps which differ across different priors (b, IRF, DS are respectively for benchmark priors, IRF prior, DS prior):

1b Set benchmark priors: $p(\xi \mid \xi^*, \Sigma^*)$. Combine with the Likelihood function to form the posterior kernel.

1IRF Set benchmark $p(\xi_b \mid \xi_b^*, \Sigma_b^*)$ and compute a target $\gamma^* = \gamma(\xi_{macro}^*, \xi_b^? = \xi_b^*)$ (compute posterior kernel).

1DS Set $p(\xi_b \mid \xi_b^*, \Sigma_b^*)$ and compute sample moments (Γ^*) on a pre-sample (compute posterior kernel).

[14] The sample is 1960:01-2001:04. Series on output, prices and wages come from the Bureau of Labor Statistics, output for the non farm business sector and its deflator and hourly compensation for the nonfarm business sector. Fed Funds rates from the FRED database of the FED of St Louis. Data have been made stationary prior to estimation. We use inflation and output growth rates and we demean them, a linear trend is subtracted from the real wage.

[15] An alternative would be to use importance sampling, drawing from some distribution f and using the weights P/f, where P is the posterior kernel, to adapt the distribution. This method has been explored in An and Schorfheide (2007), and it has been shown to be not superior to MCMC

2 Use a numerical optimizer to find the mode of the posterior defined above for benchmark priors (1b), call the mode $\bar{\xi}$.[16] To make the different estimations comparable this step is run only once with benchmark priors and the estimates are used to run estimation for all different priors.[17]

3 Compute the inverted hessian of the posterior H at $\bar{\xi}$ and decompose via eigenvalue-eigenvector decomposition to form a matrix D: $H = D'\Lambda D$.

4 Draw a candidate vector of parameters (ξ^c) as according to a random walk with normally distributed innovations $\eta \sim N(0,1)$:

$$\xi^c = \xi_i + \mu\Lambda^{-\frac{1}{2}} D\eta,$$

where $\xi^c = [\xi^c_{macro}, \xi^c_b]$ and μ is a general scale parameter, which we use to tune the acceptance rate of the chain. To make estimation with different priors numerically comparable, each chain is started at the same ξ_0, numerically found in step one by benchmark priors.

5b Combine Likelihood and benchmark prior to get a posterior kernel for draw ξ^c.

5IRF Compute the impulse response function $\gamma(\xi^c_{macro}, \xi^?_b = \xi^*_b,)$ and form the prior kernel $w(\xi^c_{macro}, \xi^?_b = \xi^*_b \mid \gamma^*, K)$, combine with $p(\xi^c_b \mid \xi^*_b, \Sigma^*_b)$ and the Likelihood to get a posterior kernel for draw ξ^c.

5DS Compute the simulated data moments implied by draw $\xi^c_{macro}, \xi^?_b$ where $\xi^?_b$ is fixed (we discuss this in section 4.5.2) and get the DS prior kernel (described in 4.23) combine with the benchmark $p(\xi^c_b \mid \xi^*_b, \Sigma^*_b)$ and with the Likelihood to get the posterior kernel of ξ^c.

6 Accept the draw ξ^c with probability depending on posterior odds ratio:

$$\xi_{i+1} = \xi^c \text{ if ratio} \equiv \frac{P(\xi^c)}{P(\xi_i)} > \text{rand(U(0,1))}, \tag{4.14}$$

$$\xi_{i+1} = \xi_i, \text{otherwise}, \tag{4.15}$$

[16] We used the csminwell.m program by Chris Sims.

[17] It is clearly possible to form a posterior, to be used for numerical maximization, applying points 1IRF and !DS; MCMC posterior results are robust to doing that.

where P is the posterior kernel (likelihood times prior) and rand is a uniformly distributed random variable. The MH always accepts a draw which provides a higher posterior kernel with respect to the previous one, but it can accept a draw that do not improve upon the previous one with a probability lower than one.[18]

7 The algorithm is run in one chain for 200.000 draws, the μ is calibrated in order to keep the acceptance rate between 20 and 40 % as in common literature concerning Bayesian DSGE estimation.

We assessed convergence by inspecting CUMSUM statistics which suggested us to select the last 50.000 draws of the chains to draw inference upon the parameters (see the appendix).

4.4 Benchmark prior and posterior results

In our benchmark prior all parameters are independent random variables which follow from the description in Rabanal and Rubio-Ramirez (2005).[19] The prior mode of the elasticity of intertemporal substitution is set to one, the logarithmic utility case when substitution and income effect compensate, while the Calvo parameter is set in such a way that prices adjust on average every two quarters; both distributions are modeled as gamma. The elasticity of labour is also set equal to one with a normal distribution. We chose a uniform distribution between zero and one for the degree of price indexation in the economy. The coefficients on the Taylor rule are informed on the basis of previous studies, the $\gamma_\pi = 1.5$ being the reference value since the original work by Taylor (1993). For the exogenous parameters block we use uniform distributions.[20] We use a slightly larger support for what concerns the cost push shock, which is known

[18]Numerically this is implemented by drawing a number from a (0,1) uniform distribution (denoted rand(U(0,1))) and comparing the drawn number with posterior odds

[19]Results do not change when more or less diffuse priors are adopted, provided they are independent distributions.

[20]Switching to an inverted gamma prior for the variances of shocks does not alter results significantly.

Parameters	Mean	Distribution
σ^{-1}	2	Gamma(2, 1)
Θ_p	2	Gamma(2, 1)
γ	1	Normal(1, 0.25)
ω	0.5	Uniform(0, 1)
ρ_r	0.5	Uniform(0, 1)
γ_y	0.25	Normal(0.25, 0.125)
γ_π	1.5	Normal(1.5, 0.25)
ρ_a	0.5	Uniform(0, 1)
ρ_g	0.5	Uniform(0, 1)
σ_a	0.5	Uniform(0, 1)
σ_m	0.5	Uniform(0, 1)
σ_g	0.5	Uniform(0, 1)
σ_p	0.75	Uniform(0, 1.5)

Table 4.1: Sufficient statistics for the benchmark prior (microprior)

to contribute to inflation dynamics with quite large movements. The chosen shapes and parameters of the prior distributions, are reported in table 4.1.

Adopting the benchmark prior we get the posterior distributions for the parameters shown in figure 4.1.

Posterior estimates are quite close to the numerical posterior mode. The elasticity of labor supply turns out quite low compared to its prior mean value; this makes marginal costs less sensitive to variations in output and therefore it can be seen as a substitute mechanism to wage rigidity. The degree of inflation indexation turns out to be quite high; $\omega = 0.77$, which might also be a figure overstated by the lack of rigid wages in the model. Estimated price rigidities (Θ_p) also turn out to be quite high compared to microeconometric evidence implied by our prior, a common feature in this class of models. Parameter estimates are similar to those obtained by Rabanal and Rubio-Ramirez (2005) with a similar dataset; we also found results to be robust to using 600.000 draws. An assessment of the convergence of the markov chain, which tends to be an issue in the estimation of a small DSGE model with few observables is provided in the appendix by the CUMSUM statistics.

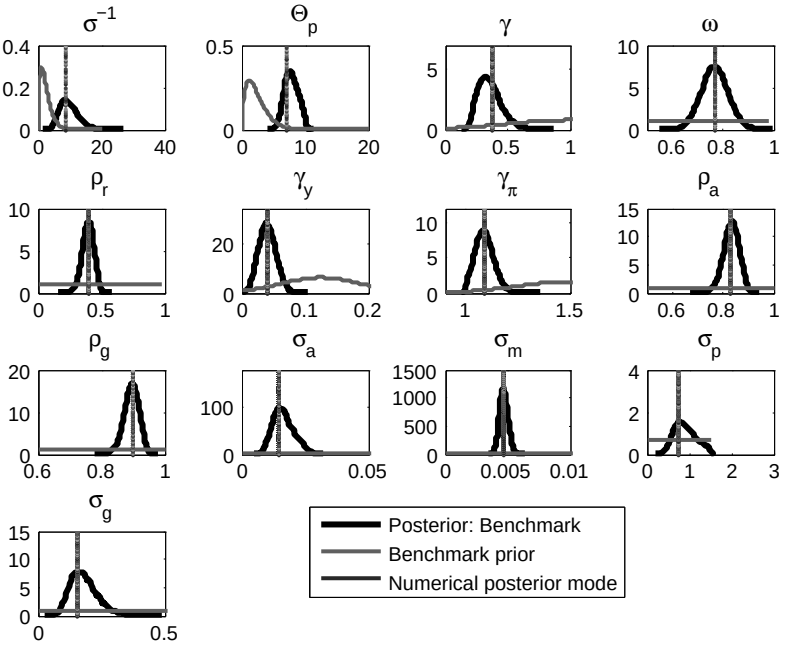


Figure 4.1: Benchmark prior, posterior distribution, numerical posterior mode

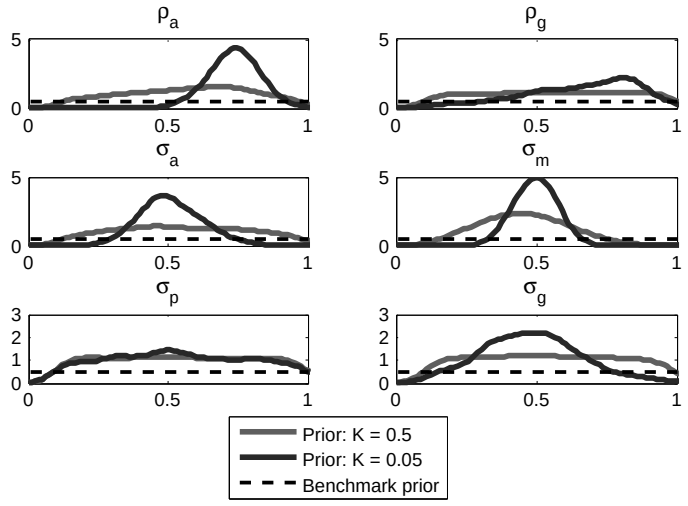
4.5 Priors on exogenous state parameter

4.5.1 IRF priors: first block

In this section we apply our IRF-prior to the exogenous states block of parameters, so that we will denote:

$$\xi_{macro} \equiv [\rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_p].$$

Kernel density plots (figure 4.5.1) show both the case $K = 0.5$ and the more informative prior $K = 0.05$, together with the benchmark prior;



,

Figure 4.2: I block of parameters: univariate kernel density

We report means and quantiles of the IRF-prior below, for the case $K = 0.5$:[21]

[21]Due to the nonlinear mapping between the structural and the reduced-form parameter spaces, analytical expressions for IRF-priors cannot be obtained, therefore we draw the para-

Parameters	0.01 perc.	Mean	0.99 perc.
ρ_a	0.06	0.64	1.00
ρ_g	0.06	0.55	1.01
σ_a	0.07	0.49	1.00
σ_m	0.09	0.46	0.96
σ_p	0.06	0.53	1.01
σ_g	0.06	0.54	1.01

The amount of information conveyed by the prior differs across parameters; it is quite high for the standard deviation of technology and monetary policy shocks, while it is lower for the remaining parameters. Compared to benchmark priors the amount of information they convey seems to be larger, both at $K = 0.5$ and $K = 0.05$. The difference in informativeness between the case $K = 0.05$ and $K = 0.5$ seems to be quite large in the plots, this were also reflected in the posteriors.

By using importance sampling we also compute the correlation structure implicit in our prior. Parameters turn out to be highly correlated.

Parameters	ρ_a	ρ_g	σ_a	σ_m	σ_p	σ_g
ρ_a	1.00	0.97	0.61	0.96	0.69	0.85
ρ_g	0.97	1.00	0.67	0.92	0.72	0.92
σ_a	0.61	0.67	1.00	0.65	0.98	0.87
σ_m	0.96	0.92	0.65	1.00	0.74	0.82
σ_p	0.69	0.72	0.98	0.74	1.00	0.88
σ_g	0.85	0.92	0.87	0.82	0.88	1.00

From the MCMC posterior estimates we found that there is little difference with respect to the benchmark prior case. Figure 4.3 shows the case $K = 0.05$, as for $K = 0.5$ the difference turned out to be almost negligible. Regarding σ_p, in spite of the fact that the IRF-prior appears to be almost as informative as the benchmark in figure 4.5.1, this is the only posterior for which there was some noticeable shrinkage using $K = 0.05$.

meters we are interested in from a uniform distribution and we compute our normalized prior. This is then used to re-weight the original draws as in an importance sampling scheme. This approach is used to compute (weighted) kernel densities, quantiles and the linear correlation measures in the tables.

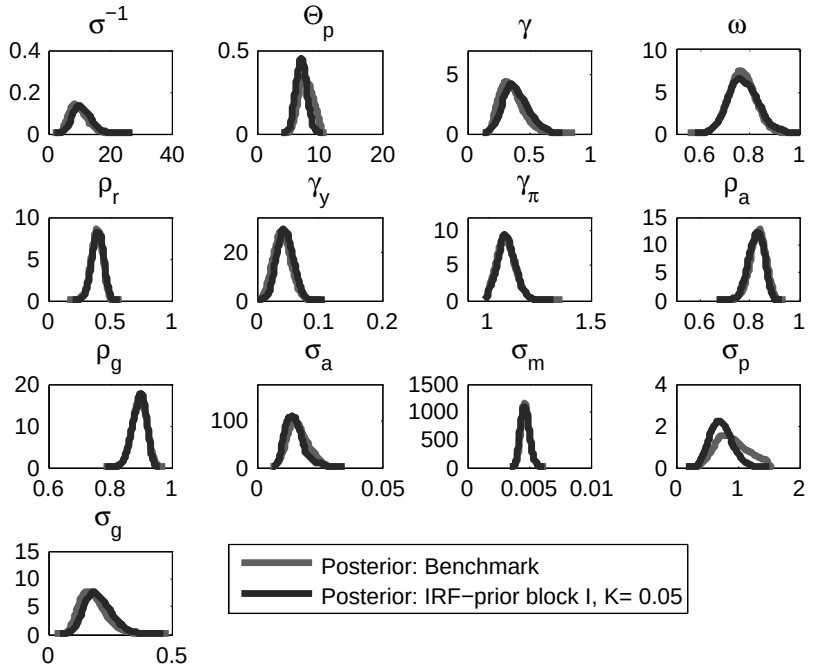


Figure 4.3: Posteriors: univariate kernel density, using benchmark prior and IRF-prior block I.

We checked that setting $\xi_b^?$ at values which differ from ξ_b^* we could not see relevant differences in the posterior estimates, as long as the change was made both regarding the target impulse (as at step 1 IRF in previous section) and in the impulse computed with draw ξ^c (as at step 5IRF in previous section). As this differs from the main finding of DS(2008), we investigate the issue in the next section.

4.5.2 DS (2008) reassessed

In this section we compare benchmark estimates with posterior results obtained by applying a DS-prior to the set of exogenous state parameters. Some description and technical details concerning the DS-prior are in appendix; estimation follows the steps outlined in section (4.3.1). We use a pre-sample of around forty observations in which we compute moments, leaving the remaining for estimation[22]. Following DS (2008), the moments computed for the observable variables are restricted to the variance covariance matrix and the order one (cross) autocorrelation. We fix the scale parameter of the strength of the prior, T^*, at 6, also following DS (2008). We checked that for T^* in the range between 4 and 8 results do not change.

For what concerns the calibrated parameters, experiments show that two parameters matter the most in shaping results: the parameter for risk aversion and the average amount of time after which a firm is able to change its price. We present two experiments, respectively DS1 and DS2 below, which differ only for the calibration of those two parameters. The following points summarize the experiments:

- Calibrated parameters are:

$$\xi_b^? = [\sigma^{?-1}, \Theta_p^?, \gamma^?, \omega^?, \rho_r^?, \gamma_y^?, \gamma_\pi^?],$$

  all but two are set at the benchmark prior mean.

DS1 ($\sigma^{?-1}$, $\Theta_p^?$) are both set equal to 2, the benchmark prior mean.

[22]Results are not very sensitive to this choice, we also tried different splits without seeing much qualitative difference, albeit a little quantitative one

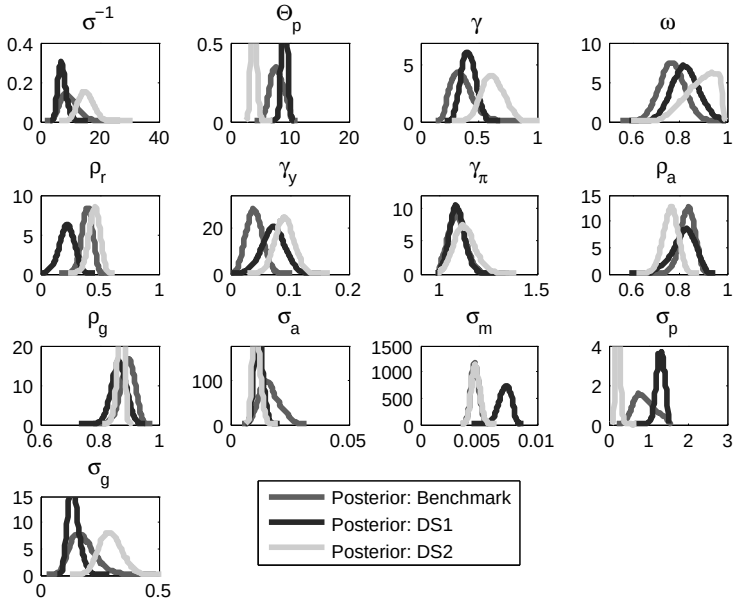


Figure 4.4: MCMC posterior distribution using the benchmark prior and DS prior. Two different settings of the calibrated rigidities, DS1, 'high' values of calibrated parameters, DS2, 'low' values of calibrated parameters.

DS2 $(\sigma^{?-1}, \Theta_p^?)$ are both set equal to 5, closer to the posterior mean found with benchmark priors.

In both experiments, differently from DS(2008), the benchmark prior applied to subvector ξ_b is the same, as described in table 4.1.

Such an (apparently) small difference in calibration between the two experiments is sufficient to produce dramatic differences in posterior estimates, as shown in figure 4.4. In particular we find that posterior nominal rigidities (Θ_p) are very low when $\Theta_p^? = 2$, which is in line with the findings of DS (2008). MCMC posterior estimates for DS1 and DS2 with respect to benchmark are shown in figure 4.4.

When the two parameters are set closer to the posterior mode obtained un-

der benchmark prior (DS1 experiment, reported in blue), estimates for nominal rigidities become much closer to those obtained by using the benchmark prior (red line). The DS prior seems to be quite informative, posterior distributions being narrower with respect to the benchmark prior case. The DS2 experiment $(\sigma^{?-1}, \Theta_p^? = 2)$ is reported in green; in this case the DS method produces posterior estimates for nominal rigidities which are far away and much lower than the benchmark results. With this respect we give a clue concerning how results in DS (2008) could have been produced. They found that, in general, deep parameters estimates can be sensitive to priors. Indeed not only do priors matter, but (mostly) matters how parameters which need to be calibrated are fixed by the researcher; luckily it seems to be possible to find calibrations for which benchmark and macroprior results do coincide, while maintaining an high degree of informativeness by using the DS prior.

4.6 Priors Results from IRF-prior: II block

4.6.1 IRF-priors: Second block

In our second experiment IRF-priors are used for a different block of parameters, namely the average length of price adjustment, the elasticity of labor supply, the degree of inflation indexation and the smoothing parameter in the Taylor rule: $\xi_{macro} \equiv [\Theta_p, \gamma, \omega, \rho_r]$.

For the IRF prior we impose (roughly) the same prior means as in the benchmark: means and quantiles are reported below.

Parameters	0.01 perc.	Prior Mean	0.99 perc.
Θ_p	1.01	2.11	7.03
γ	0.01	1.06	2.03
ω	0.06	0.49	0.92
ρ_r	0.01	0.43	0.78

with the kernel density plots shown in figure 4.5:[23]

[23]To show how overall prior tightness operates we show both $K = 0.5$ and $K = 0.05$

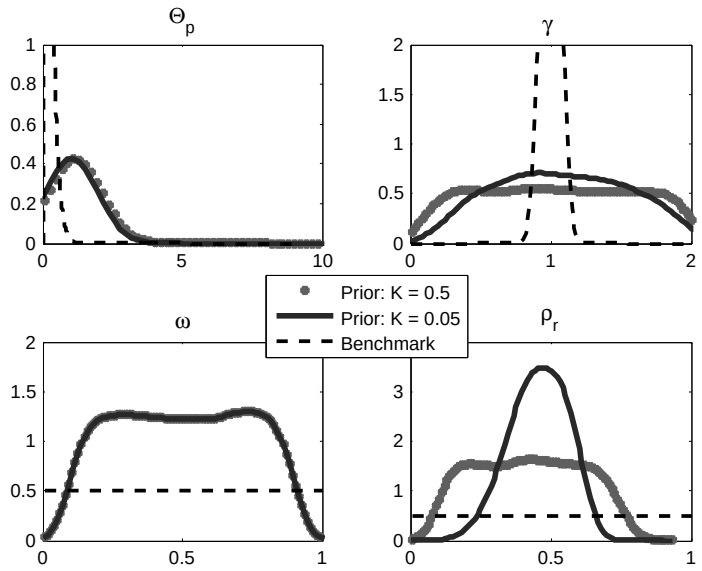


Figure 4.5: Block II. Univariate kernel density (Note that Θ_p is cut at one, kernel estimates do not take censoring into account)

Also in this case the strenght of the prior differs across parameters; differently from the previous block, the difference between setting $K = 0.5$ or a tighter $K = 0.05$ appears to be rather limited. The correlation structure of this prior turns out to be as follows:

Parameters	Θ_p	γ	ω	ρ_r
Θ_p	1.00	0.78	0.91	0.73
γ	0.78	1.00	0.96	0.98
ω	0.91	0.96	1.00	0.94
ρ_r	0.72	0.98	0.94	1.00

For what concerns the estimation, we follow the steps described in section 4.3.1; calibrated parameters were set according to $\xi_b^? = \xi_b^*$, which is according to the hyperparameters of the benchmark prior; no relevant difference was found by changing calibrated parameters both in the target and IRF implied by the metropolis draws.

From our second experiment we report changes in posterior estimates with respect to the benchmark prior case, for some parameters regarding the standard deviation of the exogenous states: technology and more dramatically, the mark-up shock.[24] Among the structural parameter the most relevant changes concerned the amount of nominal rigidities Θ_p and the elasticity of labor supply γ. For this block no sizeable difference was found between the case $K = 0.5$ and $K = 0.05$

The change in the posterior estimates produced by a change in the priors seems to follow some meaningful pattern: higher nominal rigidities (Θ_p) are met by a lower elasticity of labor supply γ and by larger mark-up shocks σ_p. This can be better understood by looking at how the Phillips curve and the marginal costs are specified in the model. The reader is referred to the appendix (4.8). After some trivial algebraic manipulation we see that,

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \kappa_{pp}(\theta_p)\left(mc_t + e_p\right), \tag{4.16}$$

$$mc_t = \left(\frac{1}{\sigma} - 1 + \frac{\gamma+1}{1-\delta}\right) y_t - \frac{\gamma+1}{1-\delta} a_t - g_t \tag{4.17}$$

[24] As a check we reran the benchmark estimation enlarging the prior of the mark-up shock, but previous result did not change substantially.

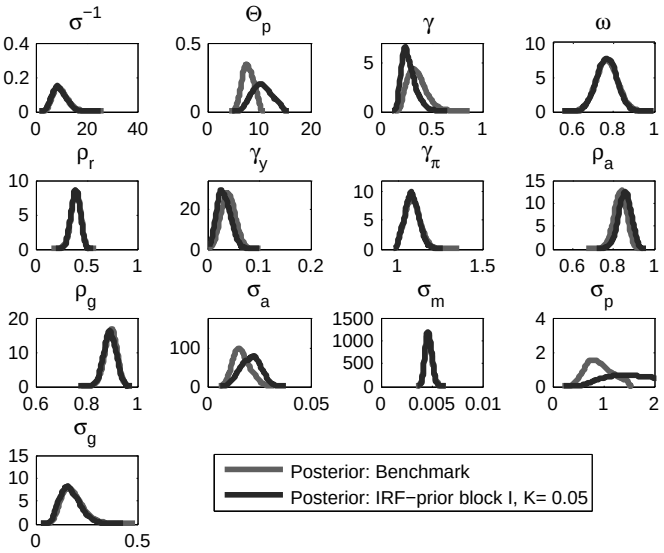


Figure 4.6: IRF prior on block II. Posteriors: univariate kernel density

the impact of the cost push shock e_p is loaded into inflation π_t by a term κ_{pp} which is a negative function of the expected duration of price rigidity (Θ_p), (see appendix). For given expectations of future inflation, more rigid prices will need a larger variance of e_p in order to match the same variance of inflation present in the data. The elasticity of labor supply enters instead in the definition of marginal costs; a lower value of the elasticity γ helps in keeping marginal costs lower after shocks to output. A lower estimate of γ would then imply an higher variance of technology shocks (σ_a), since technology a_t is multiplied by the parameter γ in equation (4.17). The specific changes in posterior estimates under the two priors might suggest a problem of identification: deep parameters are only jointly identified in the model but individually they are not well identified.

A possible interpretation of the difference in posterior estimates would then be that a correlated prior would effectively twist estimates by imposing a stronger correlation structure among parameters, when those are not well individually identified. A different interpretation would be that the model is misspecified and the structure of dependence in the prior exacerbates the problem by imposing the cross equation restrictions of the DSGE model as prior information. In order to get some insights about which of the two explanations above appears more plausible, we simulated data from the model and we estimated it by recurring to both our IRF-prior and the benchmark independent prior. Data is simulated by setting parameters consistently with the numerical posterior mode on real data.

We contrast results in figure 4.7: once misspecification is controlled for, the difference between the two posterior estimates is greatly reduced with respect to what was apparent from figure 4.6.1, and appears to affect almost only the parameter σ_p.[25] This can suggest that misspecification is to blame for the instability of the estimates. As a check for misspecification, we ran a Kalman smoother in order to recover the supposedly i.i.d. innovations of the shocks and to compute their contemporaneous correlation structure, shown in the table below.

[25]Some caution should be used since we did not undertake a proper Monte Carlo experiment. With MCMC estimation this can take more than 300 hours of computing time.

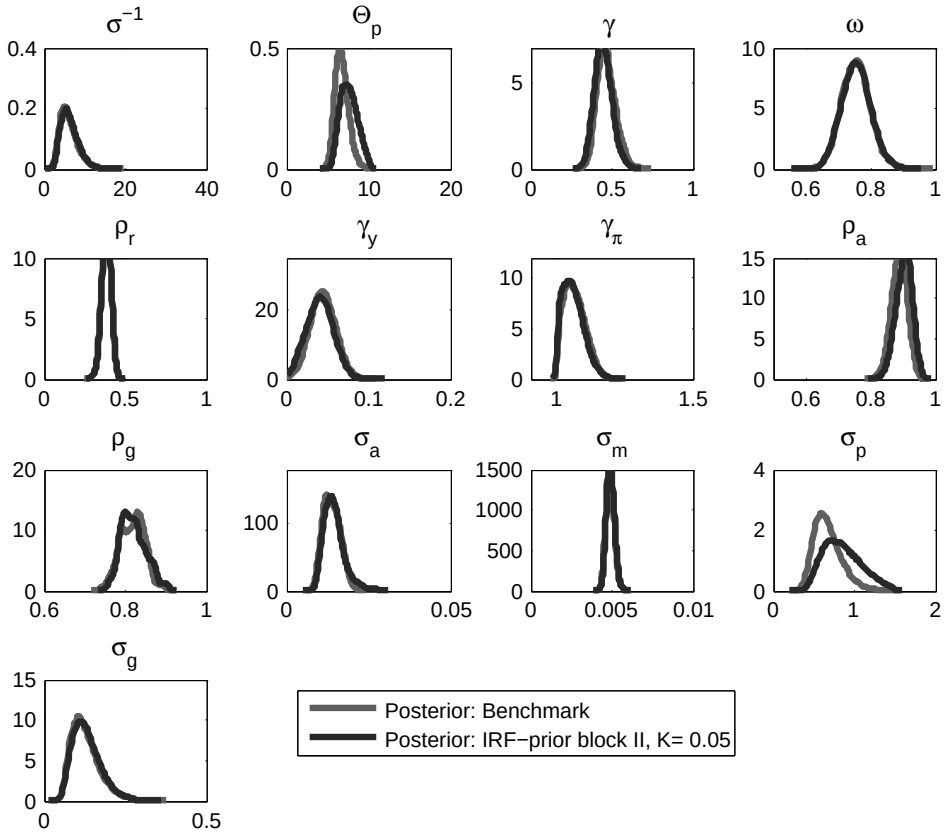


Figure 4.7: II block and independent prior benchmark: simulated data

Shocks	Technology	Monetary	Cost Push	Preference
Technology	1.00	-0.55	0.31	0.01
Monetary	-0.55	1.00	-0.57	0.1551
Cost Push	0.31	-0.57	1.00	0.0138
Preference	0.01	0.16	0.01	1.00

It is easy to see that all shocks are highly cross-correlated with the only exception of the preference shock, which is a sign of misspecification.[26]

[26]We assessed confidence bands for the correlation matrix by undertaking a little Monte Carlo

4.7 Conclusions

In this paper we studied the effect of joint priors on posterior estimates for DSGE models. We introduced a new macro prior based on impulse response functions ('IRF-prior') and we compared posterior estimates obtained using ours against a benchmark microprior.

The paper has two main implications concerning the use of macropriors. First, when a macroprior is adopted, posterior distributions depend upon parameters which need to be calibrated. The way researchers calibrate those parameters is not neutral for results. We show this to be the key to understand the Negro and Schorfheide (2008) result that posterior estimates change when a joint prior on exogenous states parameters is adopted. Once we offset the role of calibrated parameters, a joint prior on exogenous state parameters seems to have a weak or negligible influence on the posterior estimates of the parameters which regulate the amount of nominal rigidities in the economy. We confirm this finding by using both our IRF-prior and the DS prior.

Second, we verify that applying our IRF-prior on a key set of parameters can twist the posterior estimates even after calibrated parameters are taken into account. The change in the posterior estimates mostly concerns the Calvo probability of price adjustment and the variance of mark-up shocks. By the means of a simulated dataset we check that this result is due to misspecification of the DSGE model.

4.8 Appendix

The problem of intractable normalizing constants

Consistently with our definitions of the IRF-prior (see section 4.2) the Metropolis-Hastings Ratio which defines the probability of accepting a candidate draw ξ^c over the previous ξ^i will be given as follows:

experiment with 1000 simulated datasets. This provided us with the insight that in this model a reasonable confidence interval for correlations is (-0.2–0.2) with a mode of zero. Results are available upon request.

$$H(\xi^c \mid \xi^i) = \frac{L(X \mid \xi^c) g(\xi^i \mid \xi^c) w(\xi_b^c, \xi_{irf}^c)}{L(X \mid \xi^i) g(\xi^c \mid \xi^i) w(\xi_b^i, \xi_{irf}^i)} \frac{w(\xi_b^i)}{w(\xi_b^c)}, \tag{4.18}$$

where L denotes the likelihood function from the DSGE model (X are the data) and g is the proposal or candidate distribution in the Metropolis; as typically it is used a random walk as proposal this term cancels out in the ratio. The kernel w rather than its proper version $\bar{w}$ is adopted, as its normalizing constant would be eliminated in the ratio, as standard in an MCMC approach. To simplify notation we omit in the expression the term related to the benchmark prior. The last term in the expression is the ratio between 'normalizing constants'; it is not analytically available and it varies over the draws c and i so the ratio cannot be computed.

Moeller, Pettitt, Berthelsen, and Reeves (2006) show that this type of problem can be tackled by introducing an auxiliary variable x which is defined over the same space as ξ_{macro}, distributed with a conditional density $f(x \mid \xi_{macro})$ chosen by the researcher. Introduce then a proposal density $\pi(x \mid \xi_b)$ (indepedent to g) to be used in the Metropolis sampler, as follows:

$$\pi(x \mid \xi_b) = \frac{w(x, \xi_b)}{w(\xi_b)}, \tag{4.19}$$

where $w(x, \xi_b)$ is in our case the joint distribution of x, ξ_b as induced by our IRF-prior. The marginal density $w(\xi_b)$ was defined before. Now consider estimating the joint set ξ, x by a Metropolis sampler. Its ratio would be given by:

$$H(\xi^c, x^c \mid \xi^i, x^i) = \frac{L(X \mid \xi^c) g(\xi^i \mid \xi^c) w(\xi_b^c, \xi_{irf}^c)}{L(X \mid \xi^i) g(\xi^c \mid \xi^i) w(\xi_b^i, \xi_{irf}^i)} \frac{w(\xi_b^i)}{w(\xi_b^c)} \frac{f(x^c \mid \xi_b^c) \pi(x^i \mid \xi_b^i)}{f(x^i \mid \xi_b^i) \pi(x^c \mid \xi_b^c)}, \tag{4.20}$$

now plug (4.19) in (4.20) to remove the normalizing constants. Assuming also a random walk for the g we obtain:

$$H(\xi^c, x^c \mid \xi^i, x^i) = \frac{L(X \mid \xi^c) w(\xi_b^c, \xi_{irf}^c)}{L(X \mid \xi^i) w(\xi_b^i, \xi_{irf}^i)} \frac{f(x^c \mid \xi_b^c)}{f(x^i \mid \xi_b^i)} \frac{w(x^i, \xi_b^i)}{w(x^c, \xi_b^c)}, \tag{4.21}$$

the solution comes as one assumes that the distribution followed by the auxiliary variable x is given by:

$$f(x^c \mid \xi_b) = \frac{w(x^c, \xi_b = \xi_b^?)}{w(\xi_b^?)},$$

where, as in the main text, $w(x^c, \xi_b = \xi_b^?)$ denotes the IRF joint prior kernel when ξ_b is at some fixed value $\xi_b^?$. The final expression for the metropolis ratio is given by:

$$H(\xi^c, x^c \mid \xi^i, x^i) = \frac{L(X \mid \xi^c) w(\xi_b^c, \xi_{irf}^c)}{L(X \mid \xi^i) w(\xi_b^i, \xi_{irf}^i)} \frac{w(x^c \mid \xi_b = \xi_b^?)}{w(x^i \mid \xi_b = \xi_b^?)} \frac{w(x^i, \xi_b^i)}{w(x^c, \xi_b^c)}, \tag{4.22}$$

(4.22) reduces to the shortcut outlined by DS(2008) only if $x^i = \xi_{irf}^i$ for all i.

Model application

We use the same model as one of those estimated in Rabanal and Rubio-Ramirez (2005). The reason for choosing a small scale model rather than a larger one is twofold. On the one hand DS (2008) already document the problem at hand by using a medium scale model à la Smets and Wouters. Taking that latter model or even a larger one would therefore not add value to our analysis, while it would make it more complicated and less transparent, as already seen in DS (2008). On the other hand, we already know from Negro and Schorfheide (2008) that even larger models are misspecified, albeit probably less than smaller ones, but on top of that, they have more parameters and this can make identification more difficult. Overall we believe that the model we chose represents a good trade-off between transparency and realism for our analysis.

The (linearized) model is described by the following set of equations:

$$\begin{aligned}
\frac{1}{\sigma} y_t &= \frac{1}{\sigma} E_t y_{t+1} - (r_t - E_t \pi_{t+1} + E_t g_{t+1} - g_t); \\
\pi_t &= \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \kappa_{pp}(mc_t + e_p); \\
r_t &= \rho_r r_{t-1} + (1 - \rho_r)(\gamma_\pi \pi_t + \gamma_y y_t) + e_m;
\end{aligned}$$

$$y_t = a_t + (1-\delta)n_t;$$

$$mc_t = rw_t + n_t - y_t;$$

$$rw_t = y_t\frac{1}{\sigma} + \gamma n_t - g_t;$$

The complete set of variables is given by: $\{y_t, mc_t, rw_t, \pi_t, r_t, n_t\}$, respectively the GDP, marginal costs, real wages, inflation rates, interest rates, the amount of hours worked. There are four shocks in this economy, technology a, intertemporal preferences g, monetary e_m and mark up shocks e_p; the first two autoregressive processes, while the latter are i.i.d.

$$a_t = \rho_a a_{t-1} + e_a,$$

$$g_t = \rho_g g_{t-1} + e_g.$$

The first three equations are the Euler equation, the (backward looking) Phillips curve and a standard Taylor rule. The backward looking component in the Phillips curve is derived from the assumption that non updating producers partially index their prices as a function of past inflation. The remaining equations are standard: the production function, the definition of marginal costs and the supply for labour. As common in the literature, some constants in the equations above are given by a non linear combinations of deep parameters, as follows:

$$\gamma_b = \omega/(1+\omega\beta)$$

$$\gamma_f = (\beta/(1+\omega\beta));$$

$$\kappa_{pp} = \frac{(1-\delta)(1-\theta_p\beta)(1-\theta_p)}{(\theta_p(1+\delta(\epsilon-1)))(1+\omega\beta)},$$

ω is the degree of backward looking indexation in the Phillips curve, β is the discount factor (calibrated at 0.99), θ_p is the probability of price adjustment in the Calvo model, δ is the share of capital in the Cobb Douglas production function (0.36) and ϵ is the elasticity of substitution among varieties in the bundle of commodities (6).

Estimation results: tables

Table 4.2: Numerical mode results

Parameters	Numerical posterior mode
σ^{-1}	8.5361
Θ_p	7.2722
γ	0.3735
ω	0.7689
ρ_r	0.3943
γ_y	0.0415
γ_π	1.0901
ρ_a	0.8332
ρ_g	0.9001
σ_a	0.0148
σ_m	0.0047
σ_g	0.1563
σ_p	0.7501

Table 4.3: Posterior: benchmark priors

Parameters	0.01 perc.	0.99 perc.	Posterior mean
σ^{-1}	4.6972	17.7151	9.8850
Θ_p	5.5936	10.1753	7.9071
γ	0.2077	0.6842	0.3630
ω	0.6456	0.9082	0.7701
ρ_r	0.2777	0.4926	0.3950
γ_y	0.0115	0.0724	0.0408
γ_π	1.0029	1.2097	1.0890
ρ_a	0.7543	0.9036	0.8329
ρ_g	0.8409	0.9424	0.8961
σ_a	0.0085	0.0258	0.0165
σ_m	0.0040	0.0057	0.0047
σ_g	0.0874	0.3300	0.1818
σ_p	0.4454	1.4693	0.9167

Table 4.4: Posterior: IRF-prior (First block)

Parameters	0.01 perc.	0.99 perc.	Posterior mean
σ^-1	4.8299	18.6660	10.3330
Θ_p	5.5119	10.0485	7.6757
γ	0.1758	0.6198	0.3650
ω	0.6466	0.8961	0.7657
ρ_r	0.2814	0.4959	0.3948
γ_y	0.0142	0.0735	0.0425
γ_π	1.0036	1.2060	1.0918
ρ_a	0.7504	0.8955	0.8314
ρ_g	0.8161	0.9416	0.8921
σ_a	0.0092	0.0315	0.0165
σ_m	0.0040	0.0056	0.0047
σ_p	0.4285	1.4680	0.8665
σ_g	0.0909	0.3332	0.1885

Table 4.5: Posterior: IRF-prior (Second block)

Parameters	0.01 perc.	0.99 perc.	Posterior mean
$\sigma^{-}1$	4.4469	17.3490	9.5278
Θ_p	6.8017	14.3987	10.5628
γ	0.1664	0.5168	0.2802
ω	0.6540	0.8880	0.7669
ρ_r	0.2892	0.4901	0.3932
γ_y	0.0063	0.0678	0.0322
γ_π	1.0045	1.1920	1.0844
ρ_a	0.7717	0.9229	0.8550
ρ_g	0.8319	0.9491	0.8921
σ_a	0.0108	0.0323	0.0211
σ_m	0.0040	0.0055	0.0047
σ_p	0.6504	2.9236	1.6528
σ_g	0.0831	0.3150	0.1729

Some further details concerning the DS prior

The DS prior uses information about sample moments, elicited from a pre-sample of data. IRFs and moments are closely related concepts: they coincide when impulse responses from all shocks are considered. Differently from simulated moments, the advantage of IRFs, not fully exploited in this paper, is that they allow the researcher to draw prior information only from the shocks whose behaviour is most known to researchers, even if they contribute to the overall data moments by little. A good example of that is the monetary policy shock. In general, IRFs are the most studied objects in macroeconomics and potentially allow to better draw information from prior knowledge.

The kernel of the Del Negro and Schorfheide's prior corresponds to the likelihood of the VAR which (approximately) embeds the restrictions implied by the DSGE model.[27] The VAR approximation of the DSGE is denoted as follows:

$$Y_t = \Phi_1(\xi)Y_{t-1} + \Phi_2(\xi)Y_{t-2} + \ldots + \Phi_j(\xi)Y_{t-j} + \epsilon_t,$$

[27] DSGEs do not always have a finite VAR representation; a simple RBC model estimated with capital as unobserved variable has a finite VARMA structure, but no exact VAR one. To check the quality of the approximation we simulated datasets from an RBC model and we estimated both the RBC and the approximated VAR by likelihood methods. The difference in the sampling distributions seems to be rather small. Results are available upon request.

or in matrix notation

$$Y_t = \Phi(\xi)X_t + \epsilon_t,$$

with $\Phi(\xi) \equiv \Phi_1(\xi) \dots \Phi_j(\xi)$ and $X_t \equiv [Y_{t-1} \ \dots \ , Y_{t-j}]$.

The Φs are numerically constructed using the OLS regression coefficients formulas from DSGE-simulated data moments and they depend upon the whole vector of structural parameters ξ. The likelihood of the VAR approximation of the DSGE is easily expressed as a function of few data moments, conditional on the values of deep parameters:

$$L(Y, X \mid \xi) = \mid \Sigma(\xi) \mid^{-(T^*+n-1)/2} \exp\left(\frac{T^*}{2} tr[\Sigma(\xi)]^{-1}(\Gamma_{YY} - 2\Phi(\xi)\Gamma_{XY} + \Phi(\xi)'\Gamma_{XX}\Phi(\xi)'\right), \tag{4.23}$$

where $\Sigma(\xi)$ is the theoretical variance covariance matrix of the data implied by the DSGE, T^* is a scale factor for the strength of the prior. Γs are sample data moments (a pre-sample is used) with Y as the dependent variables (in our case wages, output growth, inflation and interest rates) and X are lagged data. Expression (4.23) provides the kernel of the DS-prior concerning parameters which are informed by the prior. When not the whole vector of deep parameters is informed by the DS prior, but just some ξ_{ds}, the kernel above should be multiplied with a microprior on ξ_b. The same issue of calibrated parameters as in IRF-prior also arises when computing (4.23). The rationale underlying the DS prior stems from a dummy prior approach, see Sims (2008), where the researcher constructs an augmented sample which combines the original data with some 'dummy' observations which have been simulated from the model. By approximating the DSGE by a VAR, DS are able to tune their dummy prior directly on moments of the data without the need of calibrating directly deep parameters.

The table below reports some quantiles and the mean of the DS prior as applied to our pre-sample. We sampled parameters from a uniform distribution and computed (4.23) at each draw: those are then weights for a weighted kernel density estimation. The prior is computed for two different values of calibrated parameters.

It is interesting to note that variances change widely when different values of the calibrated parameters are selected, moreover the persistence of the technol-

Table 4.6: DS Prior statistics: $\sigma^{?-1}, \Theta_p^?$ at 2

Parameters	0.01 perc.	Prior mean	0.99 perc.
ρ_a	0.6464	0.8284	0.9856
ρ_g	0.6051	0.7954	0.9691
σ_a	0	0.0515	0.2176
σ_m	0	0.1411	0.2944
σ_g	0	0.0504	0.2633
σ_p	0.0408	0.2326	0.3860

Table 4.7: DS Prior statistics: $\sigma^{?-1}, \Theta_p^?$ at 5

Parameters	0.01 perc.	Prior Mean	0.99 perc.
ρ_a	0.5640	0.7503	0.9223
ρ_g	0.5927	0.7933	0.9367
σ_a	0	0.0606	0.2030
σ_m	0	0.0606	0.2172
σ_g	0.0321	0.2172	0.3739
σ_p	0	0.1197	0.2554

ogy shock changes, while the autoregressive parameter of the preference shock is almost not affected. This is due to the fact that our set of measurements do not provide enough information to well identify the technology shock, which turns out to be very sensitive to the assumed parameters. While the preference shock is not sensitive to changes in its persistence parameter we found that the smoothed technology shock turns out to be heavily influenced by changes in both its own parameter and changes in the persistence parameter of the other shock.

Results with the DS prior

Table 4.8: Posterior Estimates with DS: ($\sigma^{?-1}$, $\Theta_p^?$ at 2)

Parameters	0.01 perc.	0.99 perc.	Posterior Mean
σ^{-1}	10.3426	23.6727	15.7359
Θ_p	3.1205	4.9284	3.9258
γ	0.4376	0.8156	0.6270
ω	0.7361	0.9786	0.8935
ρ_r	0.3380	0.5534	0.4539
γ_y	0.0594	0.1315	0.0930
γ_π	1.0176	1.2788	1.1297
ρ_a	0.6856	0.8270	0.7601
ρ_g	0.8435	0.8928	0.8724
σ_a	0.0080	0.0149	0.0109
σ_m	0.0040	0.0057	0.0048
σ_g	0.2015	0.4277	0.2993
σ_p	0.1438	0.3621	0.2282

Table 4.9: Posterior Estimates with DS: ($\sigma^{?-1}$, $\Theta_p^?$ at 5)

Parameters	0.01 perc.	0.99 perc.	Posterior Mean
σ^{-1}	4.7907	11.6364	7.3624
Θ_p	8.1756	10.1361	9.1578
γ	0.2786	0.5678	0.4165
ω	0.6984	0.9504	0.8188
ρ_r	0.0862	0.3692	0.2263
γ_y	0.0264	0.1202	0.0740
γ_π	1.0019	1.1973	1.0908
ρ_a	0.6520	0.9066	0.8142
ρ_g	0.8087	0.9217	0.8646
σ_a	0.0085	0.0173	0.0119
σ_m	0.0060	0.0082	0.0073
σ_g	0.0962	0.2134	0.1401
σ_p	1.0877	1.4887	1.3153

4.8.1 Simulated data

Table 4.10: Benchmark priors: simulated data

Parameters	0.01 perc.	0.99 perc.	Posterior mean
$\sigma^{-}1$	2.6439	12.9031	6.1677
Θ_p	5.1776	9.1008	6.7599
γ	0.3489	0.6236	0.4664
ω	0.6568	0.8689	0.7554
ρ_r	0.3141	0.4448	0.3836
γ_y	0.0092	0.0814	0.0444
γ_π	0.9997	1.1653	1.0627
ρ_a	0.8281	0.9420	0.8888
ρ_g	0.7483	0.8920	0.8176
σ_a	0.0088	0.0199	0.0138
σ_m	0.0044	0.0055	0.0049
σ_p	0.3809	1.2081	0.6743
σ_g	0.0590	0.2625	0.1284

Table 4.11: IRF-prior on block II: K = 0.05 (simulated data)

Parameters	0.01 perc.	0.99 perc.	Posterior mean
$\sigma^{-}1$	2.6833	13.0484	6.3809
Θ_p	5.6056	9.9593	7.5941
γ	0.3333	0.5923	0.4474
ω	0.6575	0.8691	0.7562
ρ_r	0.3169	0.4470	0.3844
γ_y	0.0029	0.0798	0.0405
γ_π	0.9989	1.1708	1.0583
ρ_a	0.8407	0.9547	0.9030
ρ_g	0.7601	0.9014	0.8206
σ_a	0.0092	0.0258	0.0146
σ_m	0.0044	0.0055	0.0049
σ_p	0.4497	1.4290	0.8557
σ_g	0.0591	0.2635	0.1320

Cumsum plots

We show cumsum (CS) plots of the posterior estimates, where cumsum is the difference between a rolling mean of posterior draws and the overall mean, scaled by the standard deviation of the chain:

$$CS_t = \left(\frac{1}{t} \sum_{n=1}^{t} \theta^n - \mu_\theta \right) / \sigma_\theta,$$

in order to have the percentage oscillation of the CS statistic around the mean value. In order to compute posterior estimates we retain draws after a t such that the CS statistic oscillates by no more than 5%. In general this is achieved by selecting the last 50.000 draws of our chains, which is what we do. Below we report in the graphs for the last 100.000 draws of the chains.

Figure 4.8: Cumsum plots : benchmark priors

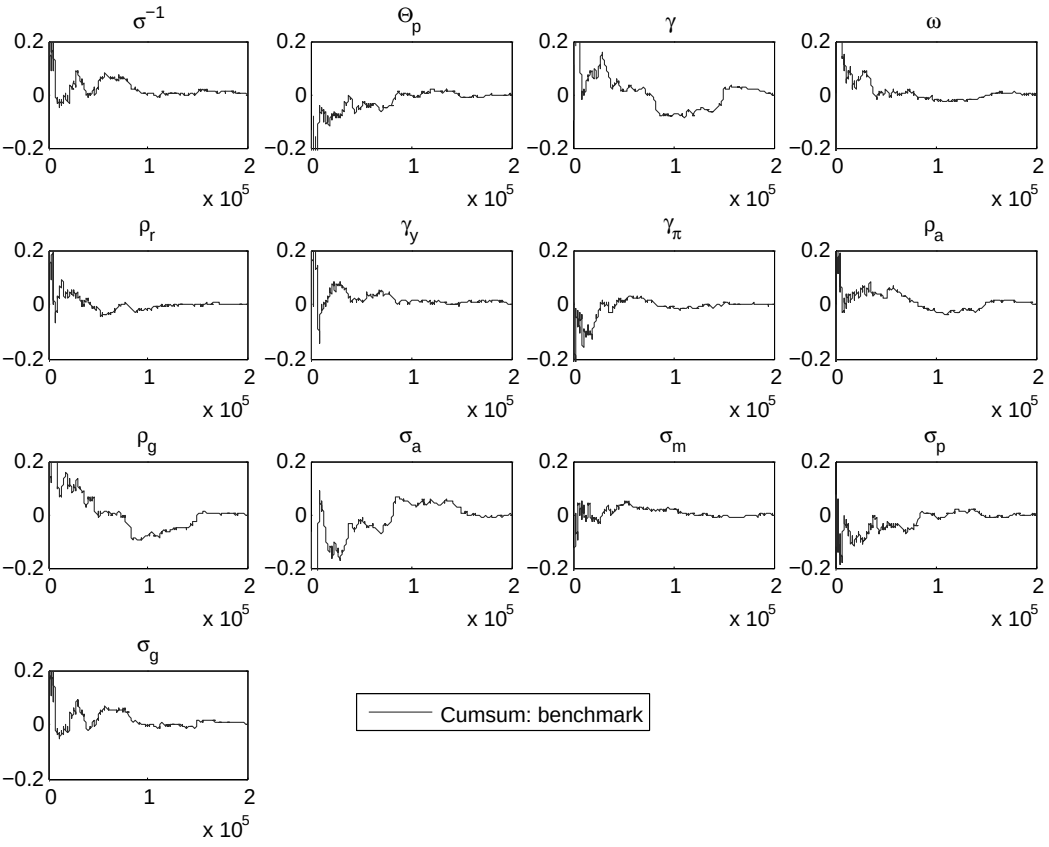


Figure 4.9: Cumsum plots : DS prior, DS1 experiment

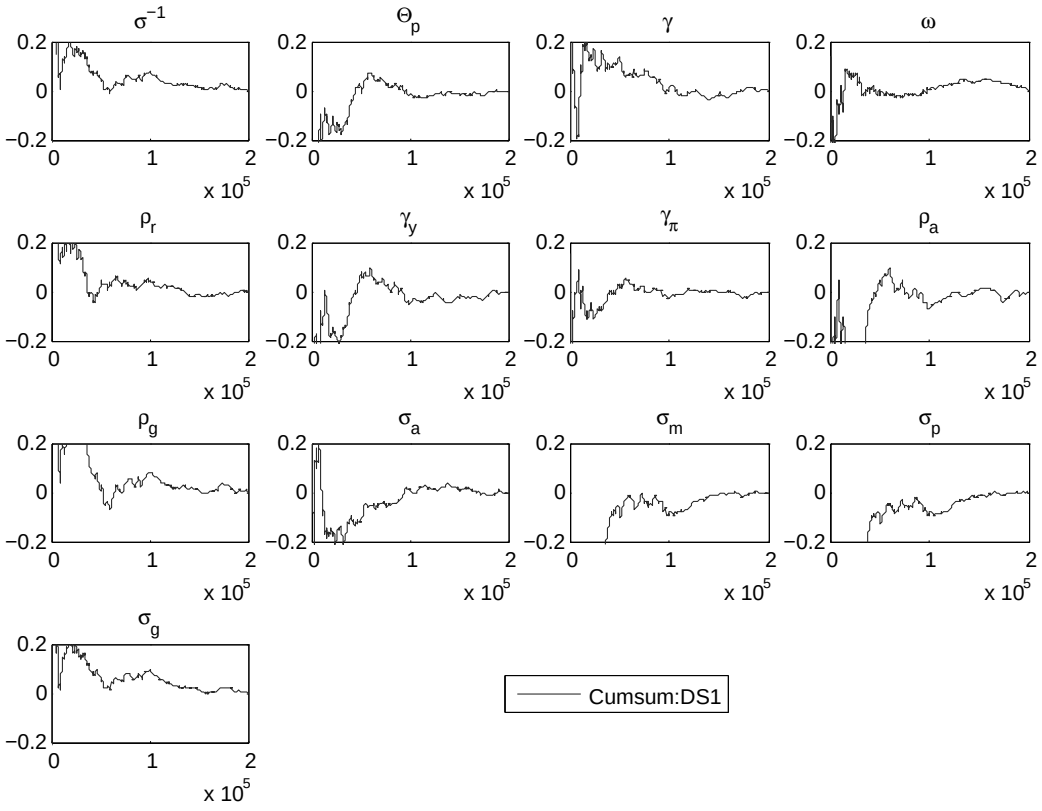


Figure 4.10: Cumsum plots : DS prior, DS2 experiment

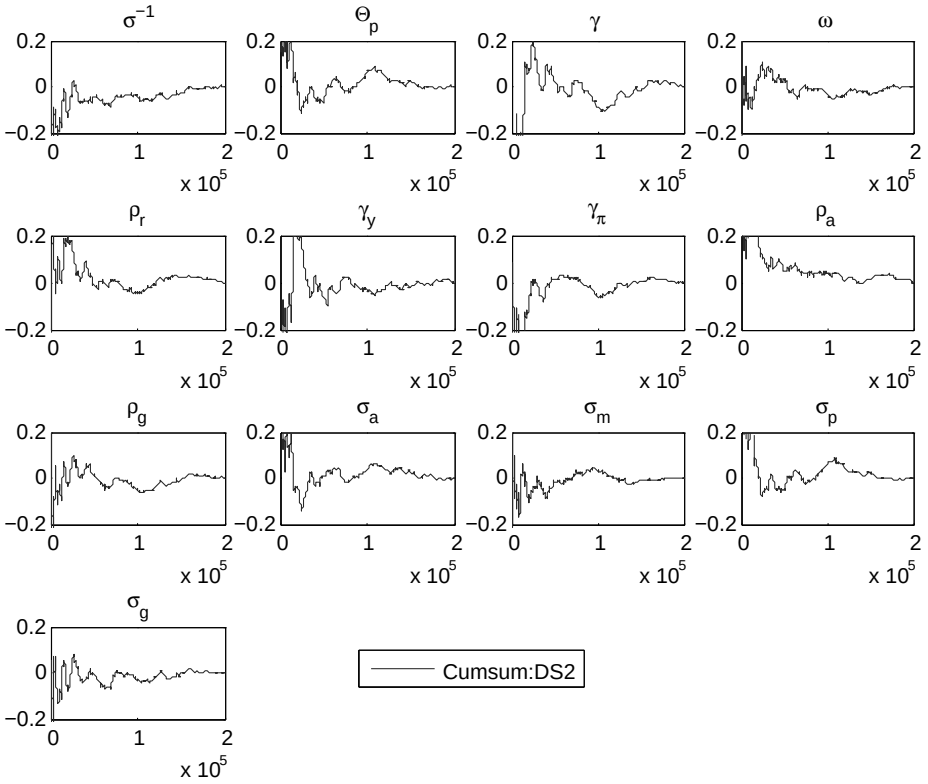


Figure 4.11: Cumsum plots: IRF-priors, block 1

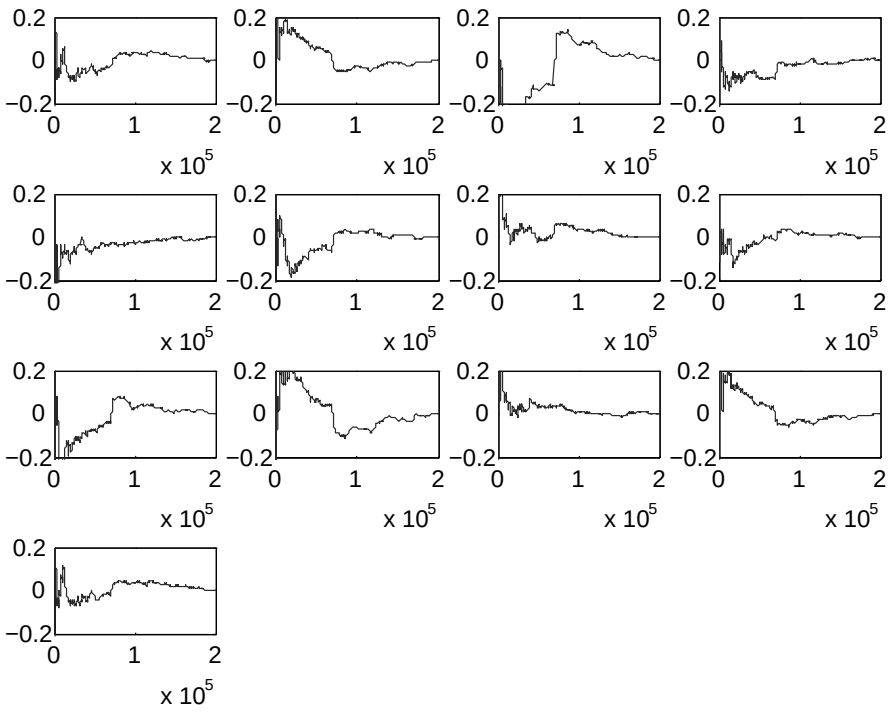
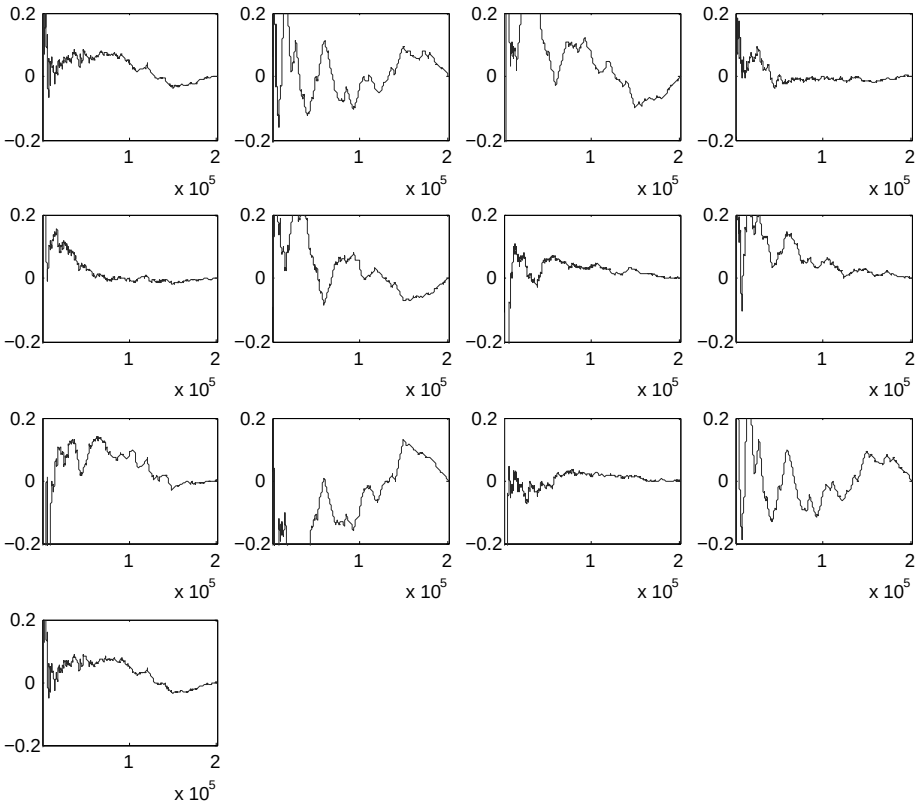


Figure 4.12: Cumsum plots: IRF-priors, block 2



Bibliography

ABADIR, K. M., G. CAGGIANO, AND G. TALMAIN (2006): "Nelson and Plosser revised: the ACF approach," University of Glasgow Working Papers 7.

ABADIR, K. M., AND G. TALMAIN (2002): "Aggregation, persistence and volatility in a macromodel," *Review of Economic Studies*, 69, 749–779.

——— (2005): "Distilling co-movements from persistence macro and financial series," ECB Working Papers 525.

ALTISSIMO, F., B. MOJON, AND P. ZAFFARONI (2009): "Fast micro and slow macro: can aggregation explain the persistence of inflation?," *Journal of Monetary Economics*, 56(2), 231–241.

AN, S., AND F. SCHORFHEIDE (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, pp. 187–192.

ANDOLFATTO, D. (1996): "Business Cycles and Labor-Market Search," *American Economic Review*, 86(1), 112–132.

BAILLIE, R. (1996): "Long Memory Processes and fractional integration in econometrics," *Journal of Econometrics*, 73, 5–59.

BERNANKE, B., M. GERTLER, AND S. GILCHRIST (1999): "The Financial Accelerator in a Quantitative Business Cycle Framework," in *Handbook of Macroeconomics*, ed. by J. Taylor, and M. Woodford, vol. 1C, chap. 21, pp. 1341–1393. Amsterdam: North-Holland.

BLANCHARD, O., AND J. GALI (2007): "Real Wage Rigidities and the New Keynesian Model," *Journal of Money Credit and Banking*, pp. 35–66.

CANOVA, F. (2009): "Bridging Cyclical DSGE Models and the Raw Data," Unpublished.

CANOVA, F., AND F. FERRONI (2009): "Multiple filtering devices for the estimation of cyclical DSGE models," Economics Working Papers 1135, Universitat Pompeu Fabra.

CANOVA, F., AND L. SALA (2009): "Back to square one: Identification issues in DSGE models," *Journal of Monetary Economics*, 56(4), 431–449.

CHANG, Y., T. DOH, AND F. SCHORFHEIDE (2007): "Non-stationary Hours in a DSGE Model," *Journal of Money, Credit and Banking*, 39(6), 1357–1373.

CHARI, V. V., P. KEHOE, AND E. MCGRATTAN (2008): "New Keynesian Models: Not Yet Useful for Policy Analysis," Staff Report 409 Minneapolis FED.

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

CHRISTIANO, L. J., AND R. J. VIGFUSSON (2003): "Maximum likelihood in the frequency domain: the importance of time-to-plan," *Journal of Monetary Economics*, 50(4), 789–815.

COGLEY, T., AND J. NASON (1992): "Output Dynamics in Real Business Cycle Models," *American Economic Review*, JUNE.

COGLEY, T., AND A. SBORDONE (2008): "Trend Inflation, Indexation and Inflation Persistence in the New Keynesian Phillips Curve," *American Economic Review*, 98(6), 2101–2126.

DELL'ARICCIA, AND GARIBALDI (2005): "Gross Credit Flows," *Review of Economic Studies*, (72), 665–685.

DEN HAAN, W., G. RAMEY, AND J. WATSON (2000): "Job Destruction and Propagation of Shocks," *American Economic Review*, 90, 482–498.

DIEBOLD, F. X., AND R. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–263.

DIEBOLD, F. X., AND G. D. RUDEBUSCH (1989): "Long Memory and Persistence in Aggregate Output," *Journal of Monetary Economics*, 24(2), 189–209.

DURBIN, J., AND S. J. KOOPMAN (2001): *Time series analysis by state space methods*. Oxford University Press.

EISFELDT, A., AND A. RAMPINI (2006): "Capital Reallocation and Liquidity," *Journal of Monetary Economics*, 53(3), 369–399.

GADEA, M., AND L. MAYORAL (2006): "The Persistence of Inflation in OECD Countries: A Fractionally Integrated Approach," *International Journal of Central Banking*, 2(1).

GIL-ALANA, L. A., AND A. MORENO (2006): "Technology Shocks and Hours Worked: A Fractional Integration Perspective," Working papers Universitad de Navarra 03/06.

GORODNICHENKO, Y., AND S. NG (2009): "Estimaton of DSGE Models When the Data are Persistent," Columbia University.

GRANGER, C. W. J. (1966): "The typical spectral shape of an economic variable," *Econometrica*, 34(1), 150–161.

——— (1980): "Long memory relationship and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238.

——— (2001): "On Model Approximation for Long-Memory Processes: A Cautionary Result," *Annals of Economics and Finance*, 2, 97–100.

GRANGER, C. W. J., AND R. JOYEUX (1980): "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis 1*, pp. 15–30.

HALL, R. (2005): "Job Loss, Job Finding, and Unemployment in the U.S. Economy over the Past Fifty Years," Paper prepared for the NBER Macro Annual Conference, April 2005.

HANSEN, G. (1985): "Indivisible Labor and the Business Cycle," *Journal of monetary Economics*, 16, 309–327.

HAUBRICH, J. G., AND A. W. LO (2001): "The sources and nature of long-term memory in aggregate output," *Economic Review, Federal Reserve Bank of Cleveland*, Q(II), 15–30.

IRELAND, P. (2004): "A method for taking models to the data," *Journal of Economic Dynamics and Control*, 28(6), 1205–1226.

KING, R., AND S. REBELO (1999): "Ressuscitating Real Business Cycles," in *Handbook of Macroeconomics*, ed. by J. Taylor, and M. Woodford, vol. I B, chap. 14. Amsterdam: North-Holland.

KING, R. G., C. I. PLOSSER, J. STOCK, AND M. WATSON (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81(4), 819–40.

KLEIN, P. (2000): "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Journal of Economic Dynamics and Control*, 24(10), 1405–1423.

KOK, SORENSEN, C., AND T. WERNER (2006): "Bank interest rate pass-through in the euro area: a cross country comparison," ECB Working Paper 580.

KOOP, G., M. H. PESARAN, AND S. M. POTTER (1996): "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 114(74), 119–147.

KURMANN, A., AND N. PETROSKY-NADEAU (2007): "Search Frictions in Physical Capital Markets as a Propagation Mechanism," mimeo.

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

LAGOS, A., AND G. ROCHETEAU (2009): "Liquidity in asset markets with search frictions," *Econometrica*, 77, 403–426.

MAYORAL, L. (2005): "Further evidence on the statistical properties of Real GNP," Economics Working Papers 955, Universitat Pompeu Fabra.

MCGRATTAN, E. (2006): "Measurement with Minimal Theory," Meeting Papers 338, Society for Economic Dynamics.

MERZ, M. (1995): "Search in the Labor Market and the Real Business Cycle," *Journal of Monetary Economics*, 36, 269–300.

MOELLER, J., A. N. PETTITT, K. K. BERTHELSEN, AND R. W. REEVES (2006): "An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants," *Biometrika*, 93, 451–458.

MORETTI, G. (2007): "Detecting long memory co-movements in macroeconomic time series," Bank of Italy working paper 642.

NEGRO, M. D., AND F. SCHORFHEIDE (2008): "Forming Priors for DSGE Models and How it Matters for Nominal Rigidities," *Journal of Monetary Economics*, 55(7), 1191–1208.

OECD (2004): "Understanding Economic Growth," OECD.

PETROSKY-NADEAU, N. (2009): "Credit, Vacancies and Unemployment Fluctuations," mimeo.

PISSARIDES, C. (2000): *Equilibrium Unemployment Theory*. The MIT Press: Cambridge.

PRESCOTT, E. C., AND E. R. MCGRATTAN (2007): "Technical Appendix: Unmeasured Investment and the Puzzling U.S. Boom in the 1990s," Staff Report 395 Minneapolis FED.

RABANAL, P., AND J. F. RUBIO-RAMIREZ (2005): "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics*, 52, 1151–1166.

RAVENNA, F. (2007): "Vector Autoregressions and Reduced Form Representations of DSGE models," *Journal of monetary economics*, 54(7).

ROBINSON, P. M. (2003): *Time Series with long memory*. Oxford University Press.

ROTEMBERG, J. J., AND M. WOODFORD (1996): "Real-Business-Cycle Models and the Forecastable Movements in Output, Hours and Consumption," *The American Economic Review*, 86(1), 71–89.

RUGE-MURCIA, F. J. (2007): "Methods to estimate dynamic stochastic general equilibrium models," *Journal of Economic Dynamics and Control*, 31(8), 2599–2636.

SALTELLI, A., S. TARANTOLA, F. CAMPOLONGO, AND M. RATTO (2004): *Sensitivity Analysis in Practice*. Wiley-VCH.

SCHORFHEIDE, F. (2000): "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15(6), 645–70.

SHIMER, R. (2005): "Reassessing the Ins and Outs of Unemployment," mimeo.

——— (2009): *Labor Markets and Business Cycles*. Available on website.

SIMS, C. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20(1-2).

SIMS, C., AND T. ZHA (1999): "Error Bands for Impulse Responses," *Econometrica*, 67(5), 1113–1155.

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

SOWELL, F. (1992): "Modeling long-run behavior with the fractional ARIMA model," *Journal of Monetary Economics*, 29, 277–302.

TAYLOR, J. (1993): "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

WASMER, E., AND P. WEIL (2004): "The Macroeconomics of Labor and Credit Market Imperfections," *American Economic Review*, 94(4), 944–963.