Notes on Business Cycle Theory from a Dynamic Stochastic General Equilibrium Perspective

Solomon, Bernard Daniel

Universite de Montreal

16 February 2015

Notes on Business Cycle Theory from a Dynamic Stochastic General Equilibrium Perspective

Bernard Daniel Solomon

March 13, 2016

Contents

Notes on Business Cycle Theory from a Dynamic Stochastic General Equilibrium Perspective

Daniel Solomon[1]

Abstract

These notes go over some basic aspects of the analysis of business cycles and aggregate fluctuations from a dynamic stochastic general equilibrium (DSGE) perspective. I build a cannonical DSGE model with a small number of representative agents and a large set of distortionnary wedges standing for various frictions as an organising framework. I use this model to discuss fundamental properties of business cycle dynamics. I start with some of the basic assumptions common to most applied DSGE models, and the modeling of household and firm behaviour. Then I discuss general equilibrium and the response of the economy to various shocks with flexible prices and wages, as well as ways of applying DSGE models with actual data. Finally I add nominal price rigidities to get the standard New Keynesian model, and discuss some open economy issues, fiscal policy and unconventional monetary policy.

[1] bernardsolomon79@gmail.com

Preliminary and incomplete. Please do not distribute without permission.

1 Introduction

"Nothing is less real than realism ... It is only be selection, by elimination, by emphasis that we get at the real meaning of things."

Georgia O'Keeffe

These notes survey business cycle theory in a dynamic stochastic general equilibrium (DSGE) framework. The DSGE approach presents a formal representation of an economy with households and firms doing the best they can to attain certain objectives subject to different constraints and interacting in competitive markets. I revise certain key points, highlight others that are often implicit or omitted in the standard analysis and propose a framework that tries to synthesize a vast literature and the experience of many analysts who have worked with DSGE models.

The main objective is to use the discipline and logic of the DSGE approach as a general framework that tries to cut through the complexity of the real world and bring out some of the core factors driving the economy, eventually also allowing a rough quantitative assesment of these factors and their effects. To do this, I will construct a simplified artificial economy. This artificial economy, or model, allows us to more precisely analyse some of the key mechanisms behind the functioning of the real economy, because despite its falseness it is much easier to control and manipulate either on the computer or mentally than the real economy.

I develop a a cannonical DSGE model that encompasses many existing models as special cases. In comparison to some of the large scale macroeconomic models in use, the model I develop has a small number of variables and equations. Yet in comparison to other expositions of the DSGE approach my model is more complex, with a relatively large number of features and shock processes. The added complexity is required to provide a synthesis of some of the key issues arising in applied macroeconomic analysis, issues that can be ignored in a purely academic analysis.

I try to focus on general results that hold across a large range of plausible models and model calibrations without examining any specific calibrated or estimated model's conclusions. Because the DSGE approach is quantitative, in more specific applications we can assess its level of "realism" along important dimensions and improve its fit to the real world through standard econometric techniques. Most of the analysis should hopefully be self contained. But inevitably, since this is more a set of personal notes than a formal textbook, much of this document is better seen as a complement and an extension to more fundamental textbook readings on dynamic general equilibrium theory (for example Ljunqvist and Sargent (2004) [140]ch. 1-4, 8 and 10-13, Gali (2007) [91] (ch. 2-5) and Romer (2001) [161] ch 4,7,8 and 11).

I investigate business cycles in the general definition of the term as "economy-wide fluctuations in production or economic activity over several months or years"... "that occur around a long-term growth trend, and typically involve shifts over time between periods of relatively rapid economic growth (an expansion or boom), and periods of relative stagnation or decline (a contraction or recession)." (http://en.wikipedia.org/wiki/Business_cycles). But some of the mechanisms that I highlight are also important for explaining long run income differences across countries, or more prolonged economic depressions and medium-run fluctuations such as the 1930's in most of the world, or the lost decades in Latin America in the 1980's or in Japan in the 1990's. This ability to use a common framework to study both short run and long run issues is one of the key strengths of the DSGE approach.

I generalise the typical textbook DSGE model by introducing a relatively large number of stochastic wedges (otherwise known as shocks) to the decision margins of agents in the economy. A wedge is any factor that causes a deviation of economic agents' behaviour from the benchmark decision rules of the neoclassical growth model without any taxes or distortions. It can be seen as a generalisation of a distortionnary (non lump-sum) tax. To use a common analogy, the undistorted neoclassical growth model represents an economy in which all basic "plumbing" systems necessary for a well functioning society work well. The wedges represent various factors that interfere with the "plumbing" of the economy, making it run less efficiently and smoothly.

The DSGE with stochastic wedges framework allows us to analyse both business cycles and long run income differences between countries in a unified way (for examples of the use of this framework to study long run income differences see Parente and Prescott (1999) [156], Restuccia and Rogerson (2007) [160], Buera and Shin (2012) [36], Jones (2011) [117]). It also also provides an intuitive description of the mechanisms behind different business cycle events (often called "shocks") as changes in the "taxes" on various activities.

Chari et al. (2007) [50] demonstrate that many if not all models with realistic frictions such as credit constraints, imperfect competition or price rigidities can be represented using a benchmark RBC economy with wedges that cause deviations of agents' decisions from the equilibrium conditions of the undistorted benchmark economy. Buera and Moll (2012) [35], Zha et al (2009), Curdia and Woodford (2009) , Challe and Ragot (2010) [48] show how incomplete financial markets and credit frictions can be represented in representative agent economies as generating wedges that distort saving, investment and consumption decisions. Labour market distortions can be thought of as taxes on either labour supply or labour demand (Shimer 2009)[165]. Inefficient distribution of production inputs across firms can be treated as lower total factor productivity (TFP).

As a result, we can analyse business cycles with similar tools and concepts as those used in public finance. The cost of the distortionnary wedges approach is that we can no longer treat the representative household's objective function as the relevant measure for welfare analysis. When wedges are exogenous and the baseline model has flexible prices and wages, we can compute the competitive equilibrium of the economy using a social planner maximisation problem, but we should nevertheless remember that this does not imply a Pareto optimal allocation from the perspective of households in the economy. Since we focus on positive analysis, the limited normative validity of the model is not a big concern. [2]

I start by developing the real business cycle (RBC) model, in which only real (inflation adjusted) variables are determined and only relative prices matter. I then add nominal pricing rigidities that link the DSGE framework with the traditional sticky price Keynesian analysis. Sticky price models can be seen as a formalisation of the popular notion of business cycle fluctuations being caused by disequilibrium between supply and demand due to inefficient prices. Flexible price RBC models instead try to explain business cycles as equilibrium responses to factors that shift output despite relatively efficient price setting. Both frameworks can fit the data well along many dimensions but with radically different prescriptions for economic policy.

Many business cycle analysts emphasize the primary role of sticky prices in explaining most of the short term dynamics of the economy. They would argue that the RBC model is at best just a prelude to the actual business cycle analysis using a Keynesian sticky price model, or at worst a distraction. But there are several reasons for dedicating significant time and effort to the RBC model before extending it with nominal price and wage rigidities.

First, RBC analysis highlights the role of other frictions and market imperfections besides nominal rigidities, showing that they can cause large and inefficient economic fluctuations even if prices and wages were quite flexible. The original RBC model (e.g King and Rebelo [128]) assumed all business cycle fluctuations were efficient, in the Pareto optimality sense: the government could not improve economic outcomes without hurting at least some agents. Modern RBC theories allow for all sorts of market inefficiencies that can be potentially reduced through government policies. The common denominator of RBC models is the absence of monetary frictions or nominal price and wage rigidity (e.g cash in advance constraints, staggered nominal price settig), not the lack of any

[2] Chari et al (2008) argue that the basis of formal welfare analysis using agents' objective functions in the most recent generation of monetary DSGE models is tenouous at best. They also suggest that many of the shocks in those models should be interpreted as wedges as in BCA. The key distinction from this perspective between BCA and the current generation of monetary policy DSGE models is that the former allow for more pervasive static and dynamic correlations between the shocks, while the latter usually assume the shock processes are independent of each other. But once we accept the shocks are not structural, there is no reason to assume they are not correlated with each other. While BCA has been performed so far using RBC models, it could also be implemented using an economy with nominal price rigidities as the prototype/baseline model augmented with wedges.

sources of frictions and inefficiences (e.g incomplete financial markets, market power and real wage rigidity).

Second, even with nominal price rigidity the response of the economy to shocks comes closer to the dynamics of the RBC model over time. Factors highlighted by RBC analysis are likely to play a key role in the profit optimisation of firms that are readjusting their prices in a sticky price economy. And targetting the natural short-term interest rate from the RBC equilibrium is in many cases an optimal monetary policy. Therefore, central banks spend a lot of effort analysing the "natural" levels of output and interest rates that would emerge under flexible prices.

Before continuing, it is worth emphasizing that the goal is not to present a completely realistic model of the economy, even ignoring the difficulty of defining criteria for realism. The complexity of modern society with a large number of agents interacting almost simultaneously makes it extremely unlikely that any mental or mathematical model of the economy would ever be complete or very realistic anyways. At best, we can aim for for a model or a set of models that will allow us to keep track of the main variables in the economy, explore the logical consequences of certain assumptions, approximately mimic and study some of the main features of the economy and through its relative simplicity improve our grasp of reality.

2 Sorting out some core assumptions

The essence of the DSGE approach is the requirement for the economist to carefully specify all demand and supply functions, or their imperfect competition equivalents, and all budget and other accounting constraints of the economy that are necessary to conduct the analysis. To accomplish this, we need to say something about the main agents, or actors, in our model.

The model economy consists of many almost identical households that can be summarized by one or 2 representative households that own all the firms, and many identical firms that can be summarized by a representative firm (formally, there is a continuum of households and firms). Households and firms interact through competitive markets for goods, labour, capital and financial assets. Their decisions are affected by several stochastic wedges or taxes. These represent various financial markets' imperfections, search frictions, imperfect information and deviations from rational expectations.

The DSGE framework here can be seen as a generalisation of the Business Cycle Accounting (BCA) framework of Chari et al (2007) [50]. I extend BCA to contain a larger number of wedges, allowing a more detailed identification of various frictions distorting the economy, for example through data on wages and asset prices. BCA usually assumes that the wedges are exogenous and follow a general reduced form stochastic process such as a VARMA(p,q) after the appropriate stationarity inducing transformations (such as linear detrending or first differencing). They can be correlated among themselves both dynamically (for example through a VAR(p) component of their joint stochastic process) and statically (e.g through correlated shock innovations). The wedges can also be linked to the model's endogenous variables to take into account interactions between the full state vector of the economy and the wedges. For example, the feedback between financial frictions and the level of economic activity (frequently called the financial accelerator) is captured up to a linear approximation by a VAR(p) that includes both the wedges and the endogenous states. [3]

2.1 Rational choice

2.1.1 Definition and motivating assumptions

The supply and demand functions in the DSGE approach are usually derived from the rational choice assumption: agents do the best they can to attain certain goals summarized by a value function, subject to the perceived and actual constraints they face. Formally they pick a vector of decision variables x to solve

$$\max_{x \in C(x,z)} V(x,z),$$

for some value function $V(.)$ and constrained set $C(.)$, where z is a vector of exogeneous variables. This principle for deriving decision rules can be made almost tautological by allowing for enough flexibility in agents' perceptions. There are theorems guaranteeing that any behaviour can be represented as the solution to a dynamic optimisation problem for some pattern of beliefs and preferences (Rust,2006 [162]). This makes the rational choice assumption in its most general form more of a framework for structuring our thinking rather than a falsifiable hypothesis: a method for deriving the dynamics of different variables by always trying to "rationalise" decisions, asking (mathematically) "What were they thinking ?".

The rational choice approach tries to capture on average at least the complexity of human and organisational decision making. It provides a convenient recipe for discovering potentially key effects of the agents' economic environment on their decisions without actually having access to what is in their mind when making those decisions, when surveys are likely to be highly unreliable and of limited external validity.

[3] Note that in this case the competitive equilibrium is no longer equivalent to a social planner optimisation problem.

It provides an abstract, idealised, description of the response of economic agents to key factors such as uncertainty, diminishing marginal utility of consumption (more generally diminishing marginal sensitivity to different stimuli),search costs in labour or product markets, quantity adjustment costs. Classic examples include the buffer stock/permanent income consumption-saving model (Carroll (2001) [46], Attanasio et al (2010)[14]), the search frictions model of labour markets (e.g Ljunqvist and Sargent (2004) ch. 26 [140], Hall (2014) [107]) and the neoclassical capital investment model with fixed adjustment costs (Khan and Thomas (2008) [124], Bloom et al (2012) [34]). In formulating these abstract descriptions it forces the analyst to think more carefully about the motivations of people and organisations in the economy and the constraints they face in trying to achieve certain goals.

In practice, DSGE models usually simplify the analysis by equating agents' perceptions of the process governing the evolution of their value functions and constraints with the analyst's perceptions of the actual process along most dimensions. Obviously, this strict perfect optimisation version of the rational choice approach is highly unrealistic. Relative to the benchmark that it establishes, actual decisions must be subject to significant optimisation errors.

The strict form of the rational choice approach should really be seen as an idealised abstraction of agents that take reasonably good decisions. It is a useful vehicles for quantification and improving our intuition as long as agents in the economy on average approximately solve the optimisation problems they face correctly: as long as there are some fundamental reasons explaining and justifying most of agents' decisions, at least as a rough approximation. In this case, the focus of formal analysis should first be on the fundamental motivations and constraints that shape behaviour. Adjustments for misperceptions and decision making biases can be added afterwards using reduced form statistical models to make the analysis more realistic.

Things are more complicated because any model of the economic environment (whether built formally by an economist or informally by economic agents) is likely to be missing a lot of factors, and therefore can lead to highly suboptimal behaviour. But the rational choice assumption only requires that agents' actions are approximately optimal in the context of the simplified approximate reality of the model. Agents' actions may be far from the optimal decisions because our understanding of the economy is highly incomplete and full of errors. But this type of decision error due to the complexity of the real world is also one made by the expert economist or financial advisor. It is fundamentally different from the decision errors due to psychological decision biases or inefficient heuristics that are the subject of cognitive psychology and behavioural economics.

The claim of the strict rational choice assumption is that the economists or decision making experts and the agents inside the model are on average (across many economists or many agents) roughly on an equal footing in terms of their ignorance of the truly optimal decisions. Under the strict rational choice assumption, the expert analyst can improve decisions of the agents in the model only by a modest amount. The gains from expertise cannot be too large. And if our simplified model of the economy does end up capturing the main features of the economy well, the rational choice approach assumes agents will on average make approximately optimal choices.

A weaker interpretation of rational choice theory is that it describes average behaviour of agents in the economy under the assumption that decision errors across agents are random and hard to predict systematically. In this case, ir represents a focal point, around which agents' decisions are likely to be distributed. Formally, this is captured by the Random Utility Model (RUM) (Mas Collel et al (1995) [146], chapter 1, Webb (2013), [179]). the RUM adds stochastic preference shifters or perception errors into the exogeneous variables z. Once we recognise how misspecified economic models can be, the rational choice assumption takes on a more modest role. It reduces to saying that the part of economic behaviour that can be captured most reliably by the endogenous, structural, part of a model is the rational part. Trying to endogenise irrational behaviour is subject to all sorts of modeling noise and biases, so irrationality is best captured by reduced form exogenous processes such as expectations or preference shocks capturing various decision errors.

A purely IID error implies that the analyst gives up on an attempt to systematically predict agents' optimisation errors. More systematic deviations from strict rational choice, that do not wash out in aggregate, can be captured by assuming agents optimise subject to stochastic wedges that distort decisions. For example these stochastic wedges can make capital investment look more attractive than it actually is given rational expectations on future profits, or they can increase the desired level of household consumption relative to the optimal consumption under rational expectations about future income. From an econometric perspective, it can be difficult to disentangle the component of any stochastic wedge that is due to optimisation error, from the component that is due to other factors that are not observed by the outside analyst- factors that would make the agents' decisions look optimal.

Recent research in Neuroscience finds evidence of a physical, neuron interaction based, counterpart to the value function used to rank decisions at the core of rational choice theory. At the same time it suggests that choice must be viewed as fundamentally stochastic, and is likely to be systematically biased for example due to constraints on decision making time (Fehr and Rangel , 2011)[83]. Webb (2013) [179] shows how the RUM approach can be derived as a reduced form of the standard neural decision making model. He shows how neurological studies of decisions can improve the specification of the error terms in RUM, for example suggesting patterns of correlation between different types of errors or implying heteroskedastic errors.

Research by neuroscientists may one day lead to a much more accurate model of decision making based on a realistic understanding of the brain. But for now, we still do not have a general neuroscience based model of decision making in complicated economic situations such as the consumption/saving choice or labour supply decision. As a result, the standard rational choice approach is likely to be a mainstay of decsion making models for many years to come.

2.1.2 Alternatives to rational choice

The main alternative to a rational choice model tries to posit directly relations between variables and functional forms without referring to any optimisation subject to constraints. This is often called reduced form modeling. In general, the reduced form model does not require the analyst to think as hard about the environment and objectives of economic agents. This makes it easier to develop new model specifications, and increases the speed of analysis. At the same time, this approach is more likely to neglect important effects that are not a priori obvious to an analyst that hasn't faced the same decisions under the same incentives as the agents being studied. For example, in typical reduced form approaches it is hard to know to what degree psychological features such as risk aversion, or various features of agents' constraints such different types of credit constraints or the persistence of income shocks affect agent behaviour.

In more sophisticated versions, analysts appeal to so called rules of thumb or heuristics. The main idea is that the complexity of the economic environment and the costs of thinking drive agents to adopt simple, routine, rules of behaviour that are adequate in satisfying agents' objective despite being suboptimal. These rules are often chosen rather arbitrarily, without sufficient justification. At least at this stage, the rules of thumb and heuristics postulated in economic models without the rational choice assumption tend to underestimate the sophistication of human learning and approximate optimisation abilities (Levine 2012, ch.3). For small shocks, these rules of thumb may be quasi-optimal. For the large shocks and changes that are most interesting to economists and the public it is more likely that agents pay more attention to their environment and try harder to optimise their decisions. In these cases, the simple rules of thumb may no longer apply and the optimal rules derived in DSGE models may be a better approximation to reality.

Furthermore, the optimal decision rules in DSGE models are almost always derived as a state space model of the form

$$\begin{aligned} x_t &= h(s_t, v_t), \\ s_t &= g(s_{t-1}, e_t), \end{aligned}$$

in which the decisions depend on a small number of state variables s_t. Conditional on following the rule, agents only have to optimise once- when picking the rule. Afterwards, they just need to keep on following the rule in order to attain their objectives. But then, what prevents us from describing the optimal decision rule itself as a simple rule of thumb? The problem is that the non optimisation based approaches have not adequately defined the meaning of simplicity in decision rules, nor have they clearly defined the threshold for agents to be satisfied with a suboptimal decision rule.

There has been some work on developing models in which agents optimise subject to information processing constraints (e.g Gabaix (2014) [87], Sims (2010)). Gabaix (2014, 2015 [87][88]) develops a model of optimisation subject to constraints on the complexity of the agent's mental representation of reality. These constraints lead the agent to optimise subject to a simplified perception of its environment, taking into account only a small number of factors and underreacting to others. Formally, the agent first chooses a vector of attention levels m, where $m_j = 1$ implies full attention to a variable z_j (the rational choice benchmark), while $m_j = 0$ means that the agent ignores changes in z_j, The agent chooses m by solving

$$\max_m \tilde{V}(x(z,m), z) - c(m),$$

where $c(m)$ is the cost of attention m, and $\tilde{V}$ is an approximation of the optimal value function V. Gabaix assumes that $\tilde{V}$ is a Taylor approximation around a default value of the exogeneous variables $\bar{z}$, such the sample average or last period value of z. This captures the idea that the agent already knows how to optimise quite well in a simplified default environment, such as one with constant interest rates or an environment that is the same as last period. After choosing the optimal m^*, the agent chooses x^* to solve

$$\max_{x \in C(x,z)} V(x, z, m^*).$$

Gabaix calls the resulting 2 step optimisation "sparse maximisation", since the agent has a preference for sparsity in its understanding of the environment in which it is operating.

This framework delivers many of the biases documented by behavioural economics. For example the default value creates a status quo bias, in which decisions are tilted in part towards past actions. It can also lead agents to underestimate differences in the prior probability of events (the base rate fallacy/representativeness heuristic), if agents do not devote full attention to those differences and rely instead on a default of equal prior probabilities. Or it can cause agents to underestimate the amount of mean reversion in shocks to their environment, leading to excessive extrapolation from the past. It also leads agents to put greater weight on more easily accessible and immediately important variables, such as current income or prices , reducing the weight of future income or prices in decisions.

An alternative, more complicated framework is the rational inattention model (Sims (2010)). In this model agents optimise within a standard bayesian inference model but subject to imperfect signals about reality and constraints on their ability to improve the precision of these signals. The signal processing constraints force agents to allocate their limited attention to more important sources of uncertainty, while barely responding to other less important shocks. So far, these models have only been applied in a few simplified settings and are often computationally harder to work with. In the future, one can envision a model combining the insistence of the DSGE approach on forward looking behaviour and explicit incorporation of the effects of budget constraints on preferences behaviour with some sort of optimisation under bounded rationality constraints. In the meantime, the rational choice model is its current form may be the only operational, widely used framework for explicitly modeling purposeful actions by intelligent agents.

At the same time, the rational choice framework often ends up significantly exaggerating the intelligence and access to information of agents in the economy. So there may be a role for complementing these models with other models that assume simpler rules of thumb behaviour, afterwards using some form of model combination. Or one could use hybrid models mixing rational choice based decision rules with rules of thumb, capturing an intermediate reality in which both the implications of optimal choice and rules of thumb are captured to some degree. The frequently used 2 household type DSGE model with unconstrained saving households and rule of thumb households that always consume all their income can be seen as an example of that (e.g Gali and Lopez Salido (2007) [92], GIMF (2013)[7]). A common model of financially constrained entrepreneurs assumes that except for a small fixed consumption or dividend payment rate, they invest as much as possible up to their borrowing constraint (Bernanke et al (1999)[28], Christiano et al (2012) [56]). Other examples include DSGE models in which a subset of agents have adaptive expectations (e.g Shleifer et al, 2013).

2.1.3 Hybrid models as a way of combining rational choice and

bounded rationality

Hybrid models that mix DSGE dynamics with reduced form models such as VAR's can also be seen as a way of mixing rational choice theory with simpler bounded rationality rules. A simple bounded rationality rule is to set a vector of variables x_t to some habitual level x^h up to some random decision shifter that could reflect various environmental signals that cause the agent to deviate from her habitual decision. At the same time, agents are partially rational and have some awareness of the optimal actions identified for example in a DSGE model.

One could easily imagine that these extra signals are persistent, leading for example to a VAR(1) This leads to a model for the simple rule of thumb x_t^{rot}, mixing with the DSGE dynamics x_t^{dsge}

$$\begin{aligned} x_t &= \omega x_t^{dsge} + (1-\omega) x_t^{rot}, \\ x_t^{rot} &= x^h + e_t, \\ e_t &= \rho e_{t-1} + v_t, \\ 0 &< \omega < 1, \\ &\Rightarrow \\ x_t^{rot} &= x^h + \rho(x_{t-1}^{rot} - x^h) + v_t, \\ x_t &= x^0 + a x_t^{dsge} + b(x_{t-1}^{rot} - x^h) + u_t, \end{aligned}$$

for appropriate definitions of $x^0, a.b, u_t$. The hybrid model can also be reformulated as a version of the popular partial adjustment model

$$x_t = x_t^{rot} + \omega(x_t^{dsge} - x_t^{rot}).$$

In this interpretation actual behaviour is governed by the rule of a thumb with some adjustment each period towards the rational decision.

2.2 Representative agents and the importance of income and wealth distributions

Our core model will have 1 or 2 representative household types, and a single representative type of firm. In its most extreme form the single representative household and firm model assumes that the distribution of resources across agents is irrelevant when studying aggregate fluctuations, except through its impact on the wedges. Any reduction in spending by a borrower due to higher borrowing costs is offset by an equivalent increase in spending by lenders. Income losses by lenders due to an increase in bankruptcies are offset by equivalent gains to borrowers. Because of the irrelevance of the wealth and income distributions for aggregate variables, we can study a symmetric distribution of resources with identical infinitely lived agents subject to wedges without loss of generality. The model also assumes that the demographic structure of the population has only a small effect on business cycles, so we can consider agents with an arbitrarily long (infinite) decision horizon.

In an extended model with a small number of representative agents (e.g borrowers and savers), the approach assumes that the wealth and income distributions only matters for macroeconomic aggregates through a small number of agent types. This assumption does not mean that more realistic wealth and income distributions do not affect social welfare. But the working hypothesis is that the representative agent model can capture with reasonable accuracy the changes in macroeconomic aggregates, which can then be used as inputs into welfare analysis taking into account more realistic heterogeneity between households. [4]

The representative agent model with time varying wedges makes it easier to consider the joint impact of many different frictions by simplifying computations and economic intuition. It also makes global nonlinear solution methods easier to apply. These methods are particularly valuable in studying large shocks such as financial crises. In contrast, with current computing power nonlinear solution methods may not be feasible for sufficiently detailed heterogeneous agent models. Economies with a larger number of agent types or a more structural model of any given friction are harder to solve and analyse while taking into account the joint impact of several frictions.

For some issues explicit modeling of the effect of the distribution of resources on growth and fluctuations is obviously desirable. For example, one cannot directly study the impact of the flow of debt among different households and firms without introducing wealth heterogeneity. Demographic transition issues such as the pension funding crisis also require models with overlapping generations of

[4] See Maliar and Maliar [142] for conditions under which this hypothesis is completely justified for households. In particular, we need complete financial markets. Krusell and Smith (2006) [134] discuss conditions for approximate aggregation into a few representative agents to hold even with incomplete financial markets. For firms, we can get an aggregate production function with aggregate total factor productivity as a function of the cross-section distribution of productivities, at least in the Cobb-Douglas case (see for example Moll(2009), Hsieh and Klenow (2009)). The business cycle fluctuations in the cross-section distribution of firms can be partially captured by the total factor productivity wedge. Perhaps the key point from these papers and others on aggregation such as Mulligan (2011) [151] is that one should not take the representative agent too literally as an actual agent in the economy in the same way that the average household with a non integer number of kids does not actually exist. Therefore, it does not make a lot of sense to ask if there is any agent in the economy who actually behaves like the representative agent. The representative agent represents average behaviour over many different agents. Under some conditions, studying the behaviour of such an average over many agents is sufficient to (approximately) capture the aggregate dynamics of the economy.

finite life households. Though within the infinite horizon representative household, the essence of the demographic shock from an aging population can be studied as a shift towards a higher weight on leisure in the utility function.

I view the detailed examination of specific frictions in more focused,more microfounded, heterogeneous agent models (frequently looking at only 1 or 2 frictions at a time) as complementary with the business cycle accounting approach, or the currently popular medium scale DSGE policy models with a large number of shocks and relatively loose microfoundations. One can use the more microfounded models to discipline the process for wedges in the economy studied here.

I will also examine an extension of the baseline model with 2 types of representative households, in which one type is always credit constrained. This extension nests two tractable models that have become almost standard in fiscal policy analysis (e.g the GIMF model (2013) [7] Iacoviello (2005) [111]): the PIH/ROT (Rule of Thumb) model in which a subset of households simply consume their income each period, and the borrower/saver model in which a subset households always borrow the maximum allowed by their borrowing constraints. It is a convenient way of capturing most of the insights from more realistic heterogeneous agent models with credit constraints and precautionary household saving such (e.g Carroll (2000) [45]).

2.3 Rational expectations

2.3.1 Definition and motivating assumptions

Our analysis "officially" assumes rational expectations (RE). though we will sometimes interpret the shocks hitting the economy as caused by agent misperceptions about the future. In its weakest form RE simply means that agents in the economy form expecations optimally given the available information and their information processing constraints. Using this definition, RE is simply an extension of the rational choice framework to the formation of expectations.

In its stronger form RE mean that the model builder and all agents in the model share the same conditional expectations function induced by the model. This is equivalent to saying that from the perspective of the model agents' forecasts are unbiased, i.e. there are no systematic forecast errors. It is also equivalent to saying that the model builder has no information advantage over the agents in the model. This strong form of the RE hypothesis is also frequently called "Model Consistent Expectations". Clearly these are unrealistic assumptions, given the cost of acquiring and processing information in the real world. It is highly unlikely that any agent in the economy will ever know the true data generating process of the economy. As a result, systematic forecasting biases in at least some directions are unavoidable.

The presence of strong heterogeneity in information and views on economic issues and the resulting debates obviously contradict the strict RE hypothesis that all agents share the same common knowledge of the economy. However, we can be flexible and extend our interpretation of RE to allow for differences in opinions and perceptions of the economy. The RE assumption is compatible with an economy in which every single agent's decisions deviate from the best decision under rational expectations, but in aggregate these errors in decision making approximately cancel out. This requires sufficiently small decision making units relative to the size of the economy, and low correlation between expectational errors to allow application of the law of large numbers. In applied work, we can easily extend this perspective to the case in which the average of individual errors is an IID residual (this can be treated as measurement error in a state space estimation framework).

Except for linear decision rules, this outcome is not what necessarily emerges from an environment in which agents' beliefs themselves randomly deviate from rational expectations. However, it may still be a reasonable approximation (see McCallum (1980) for a similar argument). It is almost certainly a better approximation to random expectations errors around RE than most adaptive expectations schemes. The perspective of RE as describing the average forecast error over many agents also reconciles RE equilibrium analysis with the traditional role of the economist as advisor to either the government or the private sector. If RE literally held for each agent, there would be little role for economic experts to come in and improve decision making through their advice. But if RE only holds on average, expert advice can help individual agent by reducing the bias in their own personal forecasts and decisions.

At a more fundamental level, RE enforce a certain humility of the modeler: they embody the belief that a priori the average forecasts over many different agents in the economy at any one time are usually as good as those of an individual economist building a model. This hypothesis receives some support from the typical finding of the litterature on economic forecasting that consensus or average survey forecasts usually dominate any individual economic forecast in accuracy. When building a model, the economist knows the true data generating process by default. In this case, acknowledging that the dominance of the average forecast across agents requires that on average agents have RE.

RE emerge as a focal point of the analysis when even though we know there must be some common bias in agents' beliefs (since they cannot know the correct model), we cannot a priori be certain in

which direction it will go. In this case, the RE assumption examines the outcome under the mean of a symmetric prior distribution of the direction of the bias. Ideally, one would augment this analysis by a sensitivity analysis exploring the effects of different biases in important dimensions. One could even build probability regions taking into account uncertainty on the direction of expectation biases. But due to computational and time constraints, analysts almost always jump this step. So far, the litteratures on decision making biases has suggested that agents can both overreact and underreact to information. However, it has not provided a clear framework explaining when we should observe under or over reaction. Assuming both types of biases are a priori equally common and of equal size leads to RE distorted by random expectations errors as a natural benchmark.

2.3.2 RE, adaptive expectations, learning and misperceptions

The usual alternative to RE (e.g the Fair model (Fair (2004), or the Moody's Analytics macromodel (2011)) is to assume that agents form expectations on the basis of the lags of some time series that happen to be easily observable by the economist. The adaptive expectations approach appears to better reflect the limited information and information processing capacities of some economic agents, but it is also sure to miss many factors that enter the expectations of economic agents. These are probably best modelled in terms of a relation between observable aggregates such as GDP and unobserved states as in the DSGE approach. The typical implementation of adaptive expectations also does not directly incorporate the differences in opinions and perceptions of the economy that we have highlighted as the main deviation from a strict intepretation of RE. It just imposes a model of the average bias of economic agents' expectations, often without giving it sufficient economic or psychological motivation.

The growing research on learning in DSGE models (e.g Eusepi and Preston (2011) [78]) is about how economic agents use the past to learn about the evolution of the factors driving the economy (the states). Models with learning are harder to solve than RE models. But computational constraints are likely to drop significantly over the next decade, and learning models have already shown promise in generating more persistent and hump shaped dynamics in response to shocks.

Even if we think the models with learning dynamics are more realistic, the RE benchmark still serves as a valuable point of comparison. Examining RE equilibria allows for a clear distinction between the conjenctured effects of an economic policy, regime change or shock that are due to the fundamental structure of the economy and the effects that are due to misperceptions about the economy. Essentially, RE allows us to isolate the economic equivalent of the placebo effect.

If an economic theory does not satisfy the requirement for existence of an RE equilibrium, then it cannot be that 1) people on average believe the theory and 2) the theory is actually true. Otherwise the theory would imply an RE equilibrium. In that case, the theory can only be true if many people are not aware of it or do not actually believe it, that is it cannot be a common "conventional" opinion. Placebo effects are usually less powerful or less reliable than more fundamental effects that are compatible with RE.

Finally, by clearly separating the role of expectations in in decision making, RE automatically capture the 2 direction feedback between agent expectations and expected outcomes. Agents' decisions are influenced by their expectations about the future. But at the same time if many agents change their decisions in a similar direction based on shifts in expectations, this will affect the future. Formally, an RE equilibrium has to solve a fixed point problem:

$$\begin{aligned} x_t &= F(\{E_t x_{t+j}\}_j, z_t) \text{ and } E_t x_{t+j} = g(x_t, z_t) \\ &\Rightarrow x_t = F(x_t, z_t). \end{aligned}$$

Adaptive expectations capture the feedback from agent expectations to expected outcomes, but they assume any feedback in the other direction happens with a potentially important lag or does not happen at all. The RE assumption emphasizes that agents' current actions and new shocks hitting the economy should affect agents' expectations of the future, further affecting their actions today.

In the business cycle accounting framework, some shocks can be interpreted as persistent deviations of agents' decisions from RE. For example excessive pessimism can be mapped into a higher investment wedge, or a higher household discount factor. I will highlight these links between deviations from RE and shocks at several points in the text. There is one key difference between these approaches to "irrational" expectations and the traditional adaptive expectationas approach. Adaptive expectations imply predictable and easy to exploit (for policy makers) mistakes by economic agents. While shock based deviations from of RE allow us to explore agent misperceptions, the way they are modelled as random processes, with the ensuing difficulties in identifying and quantifying them, emphasizes the difficulty in systematically predicting and exploiting "irrationality" for policy purposes or for private gain.

A challenge to strong reliance on easily predictable irrational beliefs is that if 1) these expectation errors are the main cause of business cycles, 2) it's easy to measure irrational beliefs shocks in the private sector, and 3) most people in private sector agree with this theory of business cycles, then the government should be able to eliminate recessions simply by informing private sector agents they're too pessimistic. But in reality it's hard for the government or a smart economist to credibly tell people to be less pessimistic in their decisions, and by how much they should adjust their decisions to align them with RE. This suggests some limits on non RE shifts in expectations as an explanation for the business cycle. At the very least, correctly measuring the amount of deviations from RE is not an easy task.

2.4 Competitive markets

The baseline model without price rigidity assumes perfect competition. However, one of the key points of business cycle accounting is that many models of imperfect competition can be represented as economies with perfect competition with appropriate adjustment in the wedges that distort agents' supply and demand decisions relative to an environment with actual perfect competition (Chari et al (2007) [50]). Agents with market power behave as if they take prices as given subject to an extra tax or lower production efficiency that induce lower supply and higher prices in equilibrium. From that perspective, assuming perfect competition is without loss of generality as long we adjust the wedges correctly. We will still have standard supply and demand factors playing a key role in shaping market outcomes despite the imperfect competition.

Perfect competition is the limit of many market structures as transaction, search and information costs decline and as the number of participants increases (e.g Gale (1986) [89], Burdett and Judd (1983) [37], Moen (1997) [150]), Kircher (2009) and Mas Collel et al (1995) [146], ch 10. The general idea is that agents taking prices as given becomes a good approximation as it becomes easier for agents in the market to meet competing offers for products of a given quality or characteristics within a short enough amount of time. In the case of constant marginal costs and homogeneous products, perfectly competitive outcome can be established with just 2 competing firms (see the Bertrand model in Mas Collel et al (1995)[146], ch. 12).

With increasing marginal costs or capacity constraints, perfect competition is attained for homogeneous goods in the limit where search frictions decrease and the number of firms increases, as long as buyers are allowed to simultaneously compare multiple price offers or recall previous price offers and as long as firms are allowed to ration buyers. Even without simultaneous search, the ability of buyers to recall previous price offers allows them to search first for cheaper goods without worrying that they will end up without any purchase. As a result, as the number of competitors increases it becomes easier for buyers to find goods at the market clearing price. A firm offering goods at a price significantly above the market clearing price would not be able to find buyers as the number of competitors increases. The ability to ration buyers allows firms to offer slightly lower prices than the competition without worrying that they will have to make too many sales at excessively high marginal costs. A very low price would lead to a potentially large number of customers that would have to be served at a marginal revenue below the marginal cost of production. As a result firms will not offer prices significantly below the market clearing price.

Research on more realistic levels of search frictions has developed the most for labour markets(see Ljunqvist and Sargent (2004) [140],ch 26 for an introduction). But we can apply most of the results from labour markets search models to product markets if we think of firms as selling jobs worth a value X at a price equal to X minus the wage. Firms offering higher wages (or a combination of higher wages and amenities) sell at a lower price, but with a lower probability for the job seeker to get the job. The value functions and functional forms will of course differ in product markets.[5] I depart from the perfectly competitive framework in discussing the New Keynesian sticky price extension of the baseline model. The baseline framework for studying economies with sticky prices is monopolistic competition: a market structure where each firm has some market power coming from product differentiation or imperfect substitution among goods, and conditional on an index of market power (e.g the level of substitutability among competing goods) profit margins are independent of the number of competitors. In the flexible price model, monopolistic competition can be easily mapped into a perfectly competitive market with a distortionary wedge.

2.5 Shocks

DSGE models generate business cycles through shocks to the stochastic wedges. Shocks can represent a mixture of unexpected shifts in fundamentals and shifts in expectations that could reflect rational sunspot/sentiment shocks or random deviations from RE. The DSGE frameworkI makes no attempt to generate endogeneous deterministic fluctuations, as done by some authors for example in the Post Keynesian litterature. Instead, it suggests different sources of randomeness that cannot be fully explained or predicted in a deterministic model. These shocks are combined with several mechanisms that amplify them, make their effects more persistent and generate hump shaped/ boom-bust (or recession-recovery) dynamics that are typically observed in the data. The focus on random shocks as the cause of

[5]Old Keynesian macroeconomic models would frequently assume partially exogenous processes for prices in the short run, with agents deciding on desired supply and demand taking prices as given, and actual sales equal to the minimum of desired supply and demand. From a modern perspective, the problem is that price taking is not rational given the disequilibrium between desired supply and demand: agents have incentives to bargain or make alternative price offers to try to avoid the rationing in the old Keynesian model (see the Barro critique, Barro (1977) [20]). In contrast, modern theories of search frictions derive the number of unmatched buyer or seller vacancies and prices as endogenous outcomes of the economy.

economic fluctuations is in line with the difficulty of finding any deterministic cycle in the economy.

Policy makers and the business press often classify shocks to the wedges as demand and supply shocks. Roughly speaking, demand shocks are those shocks that shift desired spending on components of GDP such as consumption holding all general equilibrium and model based endogenous channels fixed. Supply shocks are shocks that shift the desired output of firms and desired production factor supplies (such as labour supply) once again holding all general equilibrium and other endogenous model channels fixed. For example an increase in the external finance premium on household borrowing would count as a demand shock, while a change in business hiring regulation will count as a supply shock. In some cases the classification is more ambiguous, and a shock is both a supply and demand shock. An example of this ambiguity is a change in the consumption tax, or in the level of product market competition. The demand and supply classification is often discounted by DSGE researchers (e.g Plosser (1989)), because most shocks have simultaneous effects on supply and demand factors. However, it is deeply ingrained in policy makers and in non-economists' minds. Therefore, it is useful to be able to translate analysis of DSGE models into those terms, while acknowledging that may interesting scenarios map into a combination of supply and demand shocks.

An alternative shock classification is into intertemporal and intratemporal shocks. Intertemporal shocks such as an external finance premium or an investment wedge shock directly affect agents' intertemporal optimisation conditions. In contrast intratemporal shocks' main direct effect is on budget constraints and intratemporal optimisation conditions. Once again, the classification is not mutually exclusive: in the earlier example the consumption tax shock is both intertemporal and intratemporal.

Another feature that will be important in analysing shocks is the pattern of their dynamics. Front loaded shock processes have a bigger magnitude in initial periods: $|z_{t+j+1}| < |z_{t+j}|$ for $j > 0$. They hit hardest on impact and die out gradually over time if they are stationnary. Back loaded shocks have a bigger magnitude with some delay during initial periods: $|z_{t+j+1}| > |z_{t+j}|$ for $0 < j < J$. They build up gradually over a few periods before hitting a peak and then eventually dying out if they are stationnary. These two patterns of shocks can lead to significant differences in the response of the economy. Front loaded shocks have been most common until recently, usually in the form of an AR(1) process. More recently, growth rate shocks and news shocks which are back loaded have become the focus of attention due to their ability to capture much of the common intuition about expectations driven business cycles, and the gradual diffusion of productivity improvements or new business practices. Back loaded shocks can also capture anticipated tax or government spending shocks. **Of course transitory back loaded shocks must eventually have an anticipated front loaded pattern as the shock dies out. The response of the economy to the front loaded final phase of the shock will have some similarities to the initial response to unanticipated front loaded shocks, except one has to take into account the shift in capital and other endogenous state variables due to the initial back loaded part of the shock.**

In the basic real business cycle model with upward sloping production factor supplies and downward sloping production factor demands, it is difficult for demand shocks to generate business cycles. Therefore, an important departure from the standard real business cycle model in our framework is the addition of a distribution/marketing sector that reflects search frictions in product markets as in Rios Rull et al (2012) [17]. Fluctuations in the efficiency of this sector in response to e.g credit market shocks can capture the common notion that a recession is a time when it is harder to find customers for one's products. It also allows us to formalise the popular notion of demand driven business cycles effects even with flexible prices and wages. Finally it introduces well defined notions of production capacity and capacity utilisation.

2.6 Approximations

We will use several common approximations to simplify the analysis at various stages. A frequent approximation for production and utility functions is the constant elasticity of substitution (CES) function

$$u(\{x_i\}) \simeq a\left(\Sigma_i b_i x_i^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} .$$

For this function the elasticity of substitution (the percentage change in the ratio of two inputs $\frac{x_j}{x_k}$ required to achieve a given level of u in response to a 1% change in the ratio of marginal effects of x_j and x_k on u) is approximately constant:

$$\sigma \simeq \left| \frac{\Delta(\frac{x_j}{x_k})/(\frac{x_j}{x_k})}{\Delta\left(\frac{u_{x_j}}{u_{x_k}}\right)/\left(\frac{u_{x_j}}{u_{x_k}}\right)} \right| .$$

[6] **A higher σ means that different production inputs or utility providing goods x_i are easier to substitute in response to relative price**

[6] The IES is exactly constant at the limit

$$\sigma = \left| \frac{\partial(x_j/x_k}{\partial(u_{xj}/u_{xk})} \frac{\left(\frac{u_{x_j}}{u_{x_k}}\right)}{(\frac{x_j}{x_k})} \right| .$$

changes.For example, as $\sigma \to \infty$ u becomes a linear function and all inputs are perfect substitutes in producing u. In some cases we will take a loglinear approximation for utility functions and production functions,

$$u(\{x_i\}) \simeq a\Pi_i x_i^{\alpha_i}.$$

This Cobb-Douglas specification often emerges as the only functional form compatible with certain restrictions such as a constant elasticity of substitution between different production factors or utility function variables, and the existence of a long run balanced growth path (BGP) for the economy. Formally, the Cobb Douglas function is the limit of the CES function as $\sigma \longrightarrow 1$. Of course the 2 restrictions we have just mentionned are themselves convenient approximations, roughly supported by data on long run trends. In any case, loglinear approximation often provides a tractable middle ground between the frequently (over)used linear approximation and allowing for stronger nonlinear effects. Using a 1st order Taylor approximation, it implies the intuitive restriction that a 1% change in x_i changes u by approximately α%.

Another common approximation will be the assumption of perfect foresight or its extension in the form of certainty equivalence. Certainty equivalence means that we distribute the expectations operator through the frequently non linear function, that is

$$Ef(x) \simeq f(E(x)).$$

This approximation violates Jensen's inequality, stating that for strictly concave functions

$$f(E(x)) > Ef(x).$$

, with the reverse inequality holding for the strictly convex function $g(.) = -f(.)$. Typical functions in economic models can be decomposed into a finite number of strictly concave, strictly convex or linear regions. Certainty equivalence is only exact if $f(.)$ is linear. Nevertheless for small levels of uncertainty, the certainty equivalence approximation can be reasonably accurate and provide useful intuition.

For cases with high levels of uncertainty, we can frequently get a useful decomposition of the total effects of a shock or a change in parameters into the certainty equivalence effect and a correction for the nonlinear uncertainty effects. For example, a typical no arbitrage condition for an asset j with return $R_{j,t+k}$ states that

$$1 = E_t M_{t,t+k} R_{j,t+k}, k \geq 1$$

where $M_{t,t+k}$ is a stochastic discount factor at period t for payoffs in period $t+k$**. Using the covariance equation, we can rewrite this as**

$$\begin{aligned} E_t M_{t,t+k} E_t R_{j,t+k} &= E_t M_{t,t+k} R_{j,t+k} - cov(M_{t,t+k}, R_{j,t+k}) \Longleftrightarrow \\ E_t R_{j,t+k} &= \frac{1 - cov(M_{t,,t+k}, R_{j,t+k})}{E_t M_{t,t+k}}. \end{aligned}$$

Note that for a risk free k period zero-coupon bond, the return

$$R_{t+k} = \frac{1}{E_t M_{t,t+k}},$$

so that the risk premium can be written as

$$\left(\frac{E_t R_{j,t+k}}{R_{t+k}}\right)^{1/k} = \left[1 - cov(M_{t,t+k}, R_{t+k})\right]^{1/k}.$$

The risk premium is positive whenever the rate of return on asset j covaries negatively with the stochastic discount factor, i.e the rate of return on j is lower just when the investor values a high return the most. This sort of relation is at the heart of much of asset pricing research (see Cochrane, 2011 [59]), as well as nonlinear models with financial frictions (e.g Mendoza (2010) [148]). In some cases, the effects that are shut down by default with certainty equivalence are actually 1st order. I will point out several situations where taking into account Jensen's inequality reveals important new economic mechanisms for generating and propagating busines cycles.

We will often imagine that firms and households face continuous choices, with the optimal choice given by the the calculus first order condition that the derivative of the objective function with respect to the decision variable be set to zero. Agents in our model keep on hiring, investing and accepting jobs as long as the marginal benefit exceeds the marginal cost and stop hiring, investing or accepting jobs when the marginal benefit is below the marginal cost. By the intermediate value theorem, if the net benefit function is continuous then there exists a point at which the marginal benefit equals the marginal cost. This is the optimisation first order condition. In reality, most choices are discrete so that there isn't a choice exactly setting marginal cost equal to marginal benefit. Nevertheless, the first order conditions can be a useful approximation as along as individual agents face a sufficiently large number of discrete choices or as long as there is a sufficiently large number of agents in the economy.

With these preliminary remarks out of the way, it is time to introduce the two main types of agents in the model, households and firms.[7]

[7] The word "agents" is the conventional way of labelling the most basic decision making units inside an economic model. From a marketing perspective, perhaps the term "economic actors" would have been more appropriate, while conveying the same meaning. In some sense

3 Households

I start with the classic DSGE model assumption of a single representative household type, before moving on to my preferred two household types (borrowers and savers) model There is a large number of measure 1 of identical households with an infinite planning horizon. Households try to maximise a discounted sum of each period's payoff function, usually called the utility function, subject to a sequence of budget constraints for each period.Their utility function for each period $u(c, l_t)$ is defined over consumption c_t and leisure l_t.

We follow the standard expected utility assumption to deal with aggregate uncertainty so that given any set of uncertain outcomes $\{u_i\}$ with probabilities π_i households rank decisions based on

$$v = \Sigma_i \pi_i u_i = Eu.$$

We can see this as a simplification in which *1) the contribution to v of π_i or u_i are independent of those of u_k or π_k for $k \neq i$, 2) the contribution of u_i to v is zero if $\pi_i = 0$ (event i never occurs) and u_i if $\pi_i = 1$ (event i occurs with certainty).* Yet another way of thinking about this is that each outcome contributes to the value function a utility of u_i adjusted by some weight ω_i that is linear in the probability of the outcome π_i. Furthermore the ratio of weights is proportional to the ratio of probabilities, $\frac{\omega_i}{\omega_k} = \frac{\pi_i}{\pi_k}$.

These simplifications are fully justified under certain reasonable but sometimes empirically violated assumptions (see Mas Collel et al (1995), ch. 6 [146]). The expected utility theorem says that agents' observed choices y^* are equivalent to maximising the expected utility v above for some utility function $u(y)$ as long as

1) choices satisfy the standard rationality assumptions of completeness and transitivity:the decisions maker can always rank any two choices, either by expressing a strict preference or saying that he/she is indifferent between the options. Also, if x_A is preferred to x_B, and x_B is preferred to x_C, then x_A is preferred to x_C as long as preferences do not change.

2) the independence axiom is satisfied: if households prefer outcome x_A to an outcome x_B then, given any third outcome x_C , they must prefer the outcome in which x_A occurs with probability π and x_C occurs with probability $1-\pi$ to the one where x_B occurs with probabilty π and x_C with probability $1-\pi$,

3) the continuity axiom is satisfied: if x_A is preferred to x_B which is preferred to x_C, there exists some probability $\pi > 0$ such that the outcome in which x_A occurs with probability π and x_C with probability $1-\pi$ is preferred to x_B.

a microfounded/structural macroeconomic model is at its core a mathematical story or play. The mathematics makes the evolution of the story more concise and more logically coherent (this is almost tautological since mathematics itself can be seen as a branch of logic). At the same time, there is no doubt that it makes economic models less appealing to those with a more literary inclination.

The independence axiom is essentially a requirement that the ranking between x_A and x_B remains the same when we mix in the possibility of a 3rd outcome x_C. Continuity excludes lexicographic preferences in which certain options are always dominated because there is a small probability of an extremely undesirable outcome x_C. The independence axiom is the most controversial in terms of empirical violations, and has led to development of several models relaxing it (perhaps the main alternative being prospect theory (Barberis,2012)).[8]

Households care about consumption and work effort over many periods into the future, but usually they are impatient. They care more about payoffs closer to the present and discount payoffs further into the future more heavily. Formally, they apply a discount factor $\beta_t < 1$ to each period's utility. We make the common simplifying assumption that the payoff in each period depends on payoffs in other periods only through budget constraints: there is no direct link between the payoffs in each period. This implies the total value of of intratemporal payoffs is linearly additive (see Mas Collel et al (ch. 20) [146] for further discussion of this assumption).

Households value any plan for choices of consumption, leisure and saving using the objective function

$$V_0^h = E_0 \Sigma_{t=0}^{\infty} \beta_t u(c_t, l_t)$$

, where without loss of generality we consider the decision at $t = 0$, and we assume rational expectations ($E_0(.)$ is taken under the same probability measure as that used by the analyst building the model).

The benchmark model assumes a time separable utility, so that past consumption does not affect current utility through preferences. There is no habit formation, a feature which is commonly used to explain sluggish adjustment of consumption to some shocks. However, note that we can always rewrite a utility function without habits as

$$u(c_t, l_t) = u(\frac{c_t}{s_t}, l_t)$$

with $s_t = s$ as a constant habit stock that does not affect affect optimal decisions. In this sense, for any time separable utility function the household values consumption relative to a highly persistent, slowly evoving, habit stock. The

[8] Formally our arguments only establish that the value function must be of the form

$$v = a + b(\Sigma_i \pi_i u_i),$$

but we can set $a = 0$ and $b = 1$ without loss of generality (other choices of a or b are equivalent to redefining u_i with $a = 0, b = 1$.).

key feature of preferences with habit formation is that the habit stock changes endogenously in response to business cycle conditions. To the degree that we think habits are more slowly evolving and do not respond much to business cycles, we may prefer the simpler utility function with a constant habit stock. Incidentally, micro level studies have failed to find significant support for habit formation, in the sense of a habit stock that responds to higher frequency shocks.

3.1 The household's choice problem

The representative household picks consumption c_t, risk-free bond holdings b_{t+1} and labour supply n_t^s sequences to solve

$$
\begin{aligned}
& & V_0^h = \max_{\{c_t, n_t^s, b_{t+1}\}} E_0 \Sigma_{t=0}^{\infty} \beta_t u(c_t, l_t) \\
& & \text{subject to a sequence of constraints} \\
c_t(1+\tau_{c,t}) + b_{t+1} & \leq & b_t R_t efp_{h,t} + w_t(1-\tau_{nst}) n_{s,t} + d_t + T_t,\ (\lambda_t), \\
& & \text{and a no Ponzi game (NPG) condition,} \\
\lim_{T\to\infty} E_t \frac{u_{c,t+T} b_{t+1+T}}{\Pi_{j=1}^T efp_{h,t+j+1} R_{t+1+j}} & \geq & 0. \\
n_{s,t} & = & 1 - l_t, \\
u_c & > & 0, u_{cc} < 0, u_l > 0, u_{ll} < 0, u_{c,l} \leq 0.
\end{aligned}
$$

, where d_t are dividends from firms, w_t is the wage rate, R_t is the risk free gross interest and b_t are risk free assets. $efp_{h,t+1}$ is the household external finance premium that taxes borrowing and subsidizes lending. We will show latter that it can emerge as a reduced form for a richer model in which a subset of households are borrowing constrained. Like R_{t+1}, $efp_{h,t+1}$ is predetermined with respect to $t+1$.

The taxes here can be interpreted as a mixture of actual government taxes, legislative or regulatory barriers and financial frictions. For example changes in the cost of consumer financing can be partially mapped into the consumption tax $\tau_{c,t}$ and in the borrowing external finance premium $efp_{h,t}$. Changes in the level of competition in consumer product markets or in the cost of consumer search and shopping can be mapped into $\tau_{c,t}$, with higher costs raising $\tau_{c,t}$. The labour supply wedge $\tau_{ns,t}$ can represent income taxes and welfare payments that discourage work effort. It can also represent time varying search frictions that make harder for households to find the jobs to which they are a best match in a recession (e.g Beaudry and Portier (2013)[24]).

T_t are lump-sum rebates for all the taxes and wedges in the economy. Lump sum taxes are a simple device to isolate the effect of wedges at the margin and avoid the effect through aggregate resources' loss, which many people would argue is important for long run growth but is secondary for the purpose of studying business cycles. [9]

The discount factor of the representative agent is more than a measure agents' actual psychological impatience. For example, even in a long run balanced growth path without uncertainty, precautionary saving can generate what looks like higher patience from the perspective of a representative household (e.g Aiyagari (1994) [1]). Therefore, in the balanced growth path where $\beta_t = \tilde{\beta}^t, \tilde{\beta}$ can be very close to 1 or even above 1 despite the fact that a purely psychological impatience discount factor must satisfy $\tilde{\beta} < 1$ (people generally prefer getting something today rather than tomorrow). This flexibility can be important in allowing the model to match low average real interest rates and high equity premia in the long run. [10]

The optimisation problem above is known as the sequential representation of the household's problem, since it represents the household as choosing sequences of state-contingent decisions to maximise a weighted sum of lifetime payoffs. A more intuitive representation of this problem is to make choices each period to maximise the sum of current payoffs and the current value of future payoffs, assuming the continuation of optimal decisions in the future. This equivalence of choosing infinite sequences of decisions and of making the best choices now condional on making the best choices in the future is known as the Bellman optimality principle (Stokey et al (1989) ch. 3-4 [171], Rust (2006) [162]). The principle of optimality allows us to reformulate the optimisation problem as

$$V^h(b_{t,}, \tau_t, S_t) = \max_{c_t, n_t^s, b_{t+1}} u(c_t, l_t) + E_t \frac{\beta_{t+1}}{\beta_t} V^h(b_{t+1}, \tau_{t+1}, S_{t+1})$$

$$\text{subject to}$$

$$c_t(1+\tau_{c,t}) + b_{t+1} \quad \leq \quad b_t R_t efp_{h,t} + w_t(1-\tau_{nst})n_{s,t} + d_t + T_t, \; (\lambda_t).$$

Here τ_t is a vector of wedges, and S_t is a vector of other aggregate state variables that are required for the household to form its optimal plans. This "recursive" representation of the optimisation problem is often more intuitive and useful in numerical solution methods. It shows that households do not explicitly need to form plans while contemplating a large number of periods into the future. It is sufficient that they have (on average at least) a correct

[9] Assuming the wedges are rebated to the representative household in this way is also equivalent to modeling the wedges as binding quantity constraints on the variables to which the wedges apply. For example the wedge on bond holdings could come from a borrowing constraint.

[10] $\tilde{\beta} > 1$ is compatible with a finite objective function for the representative households as long as $\tilde{\beta} G_y^{1-\sigma} < 1$, where G_y is the long run growth rate of consumption and $\sigma = \frac{-u_{xx}}{u_x}$ is the curvature of the utility function $u(f(c,l))$ with respect to $x = f(c,l)$. $\tilde{\beta} G_y^{1-\sigma} < 1$ requires $G_y > 1$ and $\sigma > 1$, both of which are empirically plausible.

assessment of the value of their future payoffs, as summarised in the value function $V^h(.)$.

Optimisation errors or misperceptions can be included in the discount factor $\frac{\beta_{t+1}}{\beta_t}$ and the wedges τ_t. More generally, we can use the recursive representation of of the optimisation problem as a starting point for various bounded rationality models in which agents replace the value of future outcomes $V^h(b_{t+1}, \tau_{t+1}, S_{t+1})$ by various approximations.

Optimal household choices imply the following conditions for each state and period:

$$
\begin{aligned}
c_t &: \lambda_t(1 + \tau_{c,t}) = u_{c,t} \\
n_t^s &: u_{l,t} = \lambda_t w_t(1 - \tau_{ns,t}) \\
b_{t+1} &: \lambda_t = E_t \frac{\beta_{t+1}}{\beta_t} efp_{h,t+1} R_{t+1} \lambda_{t+1}.
\end{aligned}
$$

In addition, optimality requires that the NPG condition hold with equality (this is usually called the transversality condition), otherwise the household would overaccumulate wealth.

3.2 Consumption and saving decisions

The first order condition for bonds, known as an Euler equation is at the centre of the modern theory of consumption and saving behaviour. At the optimal choice, households equalise the marginal utility of consuming this period to the appropriately discounted expected marginal utility of consumption next period. In the special case of a separable utility function with constant intertemporal elasticity of substitution

$$u(c, l) = \frac{c^{1-\sigma}}{1-\sigma} + v(l),$$

and assuming constant discount factors and interest rates, and certainty equivalences,we get

$$\mathbf{E}_t \frac{c_{t+1}}{c_t} = \left(\tilde{\beta} Refp_h\right)^{1/\sigma}.$$

The equation above captures the main insight of the simplest form of the modern theory of consumption: households try to keep consumption relatively constant across periods. This is often referred to as consumption smoothing (Romer (2001) [161], ch. 7). More generally, households try to smooth the marginal utility of consumption across periods, but for plausible utility function this still leads to a significant desire to stick to a smooth consumption path.

From a microeconomic perspective, the bond Euler equation is a demand curve stating that the demand for consumption today relative to consumption tomorrow is negatively related to the relative price of consuming more today than tomorrow, $R_{t+1}efp_{h,t+1}$. Perhaps more intuitively for non economists, the Euler equation says agents shift consumption from the present to the future when the total interest payment on a given loan becomes more expensive relative to a bundle of consumption goods. The many debates about the fit of aggregate Euler equations essentially revolve around what other control variables belong in this demand function besides the interest rate, and about the aggregation of the many individual household level demand curves into an aggregate relative consumption demand function.

This aggregate relation between interest rates and consumption is a reasonable approximation to an environment with more realistic fixed portfolio transaction or optimisation costs in adjusting to interest rates and asset prices at the individual level, as long as there is sufficient dispersion in idiosyncratic shocks or dispersion in adjustment costs (so absent such adjustment costs not everyone would respond the same way). In this case, the aggregate consumption response to interest rates can be quite flexible despite micro level rigidities. Similar arguments can be made for the relation between aggregate investment and interest rates (See Caplin and Leahy (2002) [42], Khan and Thomas (2008) [124]and Golosov and Lucas (2007)[95] for examples of this aggregate flexibility despite micro level fixed adjustment costs).

In some cases non convex adjustment effects will have significant effects, that may only be partially be captured by the structure of our reduced form wedges. This is especially the case for highly illiquid assets, such as housing (see for example Iacoviello and Pavan (2013) [113]). But an approximation assuming very sluggish adjustment in the aggregate is likely to be even more misspecified.

The consumption Euler equation, like all dynamic first order conditions in our model, can also be interpreted as an asset pricing equation. Here, we get the price of a 1 period, inflation adjusted, risk free bond promising 1 unit of consumption (the numeraire in our model), as a function of its payoff and the stochastic discount factor:

$$p_{bond,t} = \frac{1}{R_{t+1}} = E_t \frac{\beta_{t+1}}{\beta_t} \frac{\lambda_{t+1}}{\lambda_t} efp_{h,t+1}.$$

This links macroeconomics to the vast litterature on general equilibrium financial asset pricing (see Cochrane (2011) [59] for a survey).

3.2.1 Discount factor shocks as proxies for time varying precautionary saving and sentiment shocks

Discount factor shocks are important in modern business cycle theory, for example in Keynesian models (e.g Smets and Wouters (2002) [167], Christiano et al (2010)) and in asset pricing models (Cochrane (2011) [59]). More generally discount factor shocks are important in analysing changes in aggregate consumption and saving across the business cycle.

The stochastic component of the discount rate can be linked under certain circumstances to changes in perceived income risk. Higher uninsured idiosyncratic income risk, or lower cash on hand (the sum of income and liquid financial assets) in the presence of uninsured idiosyncratic income risk imply a lower ability to smooth consumption in case of negative income shocks. This leads to higher uncertainty, lower consumption and higher saving. The increase in saving due to higher income perceived income risk is called the preautionary saving motive (Carroll (2001) [46]). From the perspective of the representative agent, higher precautionary saving looks like a higher discount factor.

Carroll (2001) [46] argues that the implications of higher income uncertainty for consumption uncertainty growth are essentially the same as those of certain types of credit constraints. This can be easily seen in our framework by noting that consumption growth in the Euler equation is a function of the product

$$\frac{\beta_{t+1}}{\beta_t} efp_{h,t+1}.$$

Without further information than consumption growth, we cannot separate external finance premia from precautionnary saving.

The link between precautionnary saving and discount factor shocks can be seen more easily in the case of a separable utility function

$$u(c,l) = \frac{c^{1-\sigma}}{1-\sigma} + v(l), \sigma > 0.$$

Consider an economy with many households indexed by i. Assume any credit constraint does not bind and $efp_{h,t+1} = 1$. With uncertainty we have the Euler equation

$$\frac{1}{\tilde{\beta} R_{t+1}} = E_t \left(\frac{c_{i,t+1}}{c_{i,t}} \right)^{-\sigma}.$$

Using the convexity of $f(x) = x^{-\sigma}$, for $x = \frac{c_{i,t+1}}{c_{i,t}}$, Jensen's inequality implies

$$\begin{aligned} \left(E_t \frac{c_{i,t+1}}{c_{i,t}} \right)^{-\sigma} \quad &< \quad E_t \left(\frac{c_{i,t+1}}{c_{i,t}} \right)^{-\sigma} = \frac{1}{\tilde{\beta} R_{t+1}}. \\ &\iff \quad E_t \frac{c_{i,t+1}}{c_{it}} > \left(\tilde{\beta} R_{t+1} \right)^{1/\sigma} = \frac{c_{i,t+1}^{PF}}{c_{i,t}^{PF}} \iff \\ E_t \frac{c_{i,t+1}}{c_{i,t}} \quad &= \quad v_{i,t+1}^b \left(\tilde{\beta} R_{t+1} \right)^{1/\sigma}, v_{t+1}^b > 1, \end{aligned}$$

where x^{PF} is the value of x under perfect foresight. Roughly speaking moving from a low uncertainty environment to one with high uncertainty decreases current consumption and increases saving. The average $v^b_{i,t+1}$ above is an overestimate of the effect of higher uncertainty on aggregate consumption growth (as opposed to average consumption growth), since households for which the effect of uncertainty is stronger will typically be poorer and have less weight in aggregate consumption. Nevertheless, in general higher uncertainty is expected to reduce consumption and raise saving (see Challe and Ragot (2010) [48] for a more formal proof in a simplified model of uninsured idiosyncratic income risk).

In representative agent models like ours, a convenient proxy for changes in countercyclical income risk may be the unemployment rate. Much of the countercyclical income risk comes from a higher risk of job loss in a recession, and unemployment is much more likely to be observed than other measures of income risk such as the variance of earnings across different households. Therefore, one shortcut to incorporate the effect of countercyclical risk in simpler representative agent models is to add the unemployment rate as part of the discount factor in the Euler equation.

We will formalise this shortcut using a simplified model of precautionnary saving from Carrol and Toche (2009) [47]. In their model each employed household faces a probability p^u_{t+1} of job loss next period. Assume for simplicity perfect foresight on all other variables including the job loss probability.The Euler equation for an employed household is

$$\frac{c^e_{t+1}}{c^e_t} = (\tilde{\beta} R_{t+1})^{1/\sigma} \left(1 + p_{u,t+1} \left[\left(\frac{c^u_{t+1}}{c^e_{t+1}} \right)^{-\sigma} - 1 \right] \right)^{1/\sigma}.$$

Also assume for simplicity that there is an unemployment system that provides a constant replacement rate of what the newly unemployed worker would spend if he kept on working, so that

$$\frac{c^u_t}{c^e_t} = \tilde{c}_u < 1.$$

We can see immediately from this equation that an increase in the expected job loss probability increases the expected growth rate of consumption conditional on staying employed and boosts saving by the employed.

A first order Taylor approximation of the new unemployment risk factor around $\tilde{c}_u = 1$ leads to

$$\frac{c^e_{t+1}}{c^e_t} \simeq (\tilde{\beta} R_{t+1})^{1/\sigma} \left(1 + p_{u,t+1}\sigma(1 - \tilde{c}_u)\right)^{1/\sigma}.$$

Unemployment risk behaves like an increase in the return to saving. The extra return is approximately the product of the unemployment risk, the percentage consumption loss in unemployment (the level of social insurance) and risk aversion.

Caroll and Toche's model does not lend itself conveniently to aggregation via a representative household, therefore in general equilibrium it still requires the use of more computationally intensive solution methods for heterogenous agent models (e.g the Krusell and Smith (1998) [133] algorithm). However we can capture most of the aggregate implications of the original model if we assume that ex post households get complete consumption insurance against unemployment so that

$$c_t^e = c_t^u = c_t.$$

Ex-ante, we assume that employed househods do not take into account this insurance, deviating from rational expectations. Instead they think they will get $\tilde{c}_u$ of their former income if they become unemployed. As a result the representative household's consumption path is determined by the Euler equation

$$\frac{c_{t+1}}{c_t} = (\tilde{\beta} R_{t+1})^{1/\sigma} \left(1 + p_{u,t+1}\left[\tilde{c}_u^{-\sigma} - 1\right]\right)^{1/\sigma} = (\tilde{\beta}_{t,t+1} R_{t+1})^{1/\sigma},$$

, where we have redefined the stochastic discount factor to take into account unemployment risk. In models without an explicit distinction between unemployment and non-employment, and assuming small fluctuations in the number of hours per job, unemployment is is roughly proxied by $1 - n_t$.

The discount factor shock can also be mapped into overoptimism or pessimism about the future. Excessive pessimism makes the representative household more patient and less eager to consume now relative to the future. The increase in "patience" shows up as an increase in

$$\tilde{\beta}_{t,t+j} = \beta_{t+j}/\beta_t$$

for some $j > 0$ in the first order condition for bonds. Conversely, higher optimism maps into a a lower $\tilde{\beta}_{t,t+j}$. More formally, an excessively pessimistic household's Euler equation dictates lower consumption today given the expectation of lower future consumption and a constant discount factor. But under the true rational expectations operator and a constant discount factor current consumption would be too low:

$$\left(E_t \tilde{\beta}_{t,t+1} R_{t+1} efp_{h,t+1}\right)^{1/\sigma} \simeq \frac{c_{t+1}}{c_t} > \left(\tilde{\beta} E_t R_{t+1} efp_{h,t+1}\right)^{1/\sigma},$$

where we have made a certainty equivalence approximation.

This interpretation of the discount factor shock can be useful for decomposing the impact of another shock (for example a change in government spending or taxes) into the rational expectations component that is justified when households on average understand the effect of the shock, and into a an expectations shock component caused by households' misperceptions of the effect of the shock.

Cochrane (2011) [59] discusses a similar equivalence between deviations from RE and the stochastic discount factor used to price financial assets. He emphasizes the difficulty of separating shifts in the discount factor due to time varying uncertainty or risk aversion from shifts in the discount factor due to time varying deviations from RE based on typical macroeconomic and financial time series data. Shefrin (2007) [164] surveys in detail the behavioural interpretation of discount factor shocks, under the assumption that deviations from RE are the only source of the discrepancy from the single representative agent rational asset pricing model. He calls the shift in the discount factor of the representative household due to deviations from RE "market sentiment". A positive (negative) market sentiment indicates excessive optimism (pessimism) when aggregating expectations over many households. RE hold in the aggregate (zero market sentiment) if expectations errors cancel on average in the cross section, and if these errors are uncorrelated with household wealth.

3.2.2 The consumption function

We can use the household optimal bond holding conditions (the Euler equations) and household budget constraints to obtain an aggregate consumption function. By incorporating other variables such as debt to income ratios or measures of idiosyncratic household risk and the wealth inequality into efp_h, t and β_t, we get a consumption function that combines the insights of the classic Permanent Income Hypothesis (PIH, Romer ch. 7 [161])and some of the insights of the vast literature on credit constraints and precautionary saving.

The PIH emphases wealth as an important determinant of consumption spending, both in the form of various assets and in the form of future incomes . Both higher current income and higher wealth increase consumption, in contrast to traditional Keynesian analysis which emphasizes current income.

A key insight of the PIH is that much of that wealth is in the form of future earnings. By making consumption a function of expected future income, the PIH tries to partially endogenise the residual or consumption "shock" in traditional old Keynesian consumption functions, often interpreted as a shock to consumer confidence. Higher expectations of future income make consumers more optimistic, increasing their consumption. In reverse, lower future income expectations make consumers more pessimistic and reduce consumption.

The original PIH model has been heavily criticised for exaggerating the importance of wealth for consumption and underestimating the importance of credit conditions and current income (see Romer (2001) ch. 7 [161] for a basic exposition, and Attanasio and Weber (2010) [14] for a modern survey of the PIH and the main empirical and theoretical issues surrounding it). In its representative agent version, it has been criticised for exaggerating the amount of insurance available to households against idiosyncratic risks. The modern theory of the consumption function (e.g Carroll (1997) [44],Carroll (2001) [46], Kaplan and Violante (2014)[122]) usually predicts a higher average sensitivity of consumption to current income and changes in credit constraints than to wealth.

While future income is less important for large segments of the population that are credit constrained or close to their borrowing limit, it is more important for the consumption decisions of middle income and rich households, most of whom are further from their borrowing limits, and who constitute the bulk of consumption spending. Surveys such as the SCF, and modern models of consumption that take into account uninsured income risk, life cycle effects and the partial illiquidity of wealth find that around 30-42% of households are credit constrained or close to being credit constrained (Kaplan and Violante (2014)[122]). But other households in the economy do have significant net liquid wealth relative to their maximum borrowing limit, and these tend to be the richer households whose contribution to aggregate consumption is significantly above their proportion of the population.

In the US in recent years, the top 20% of income earners have accounted for up to 60% of aggregate consumption. Most of the households with low net liquid wealth have a lot of partially illiquid net wealth such as stock portfolios or housing that can be adjusted in response to bigger shocks. [11] For example a big enough negative shock may induce the liquidation of some pension assets or moving to a smaller apartment or house.

Furthermore, while uninsured idiosyncratic household risks are clearly important, recent empirical evidence suggests significant levels of insurance and much larger foresight by households regarding income changes than estimated by simple econometric methods (e.g Kaplan and Violante (2010)[121], Guvenen and Smith (2014) [105]. Guvenen and Smith (2014) [105] estimate that 45% of idiosyncratic shock innovations are ensured by a combination of formal and informal mechanisms. These results explain why empirically the marginal propensity to consume (MPC) out of transitory income shocks such as tax rebates is around 0.25 rather than the higher numbers suggested by older Keynesian consumption functions (Kaplan and Violante (2014)[122]). Note that

[11] Net wealth is assets minus liabilities: that is, Kaplan and Violante (2012) subtract credit card debt from liquid assets and mortgage debt from illiquid assets to get their wealth measures.

these MPC's are for nondurable consumption. Higher MPC's can be found if we include durable goods as part of consumption, but standard microeconomic theory argues durables are more like a form of saving than consumption, since durables are essentially a type of capital. Therefore, it is quite natural that there is a higher MPC for durable goods.

Finally, comparison of extended versions of the PIH model to micro data on the level of net worth of older households suggests that around 75-85% of households in the US succeed in saving enough for retirement relative to the optimal saving benchmark. In contrast simpler models, such as the old Keynesian consumption function fail to explain household wealth accumulation (Scholz et al ,2009 [90]). Retirement is perhaps the biggest predictable income shock that any household is likely to suffer. The ability of most households to successfully smooth their marginal utility of consumption in this case suggests that the PIH provides a reasonable approximation to the consumption and saving behaviour of most households. Unfortunately, the more sophisticated models with uninsured income risk, forward looking behaviour and partial insurance are harder to solve and aggregate for the purposes of DSGE analysis. We try to approximate the richer insights from those models by starting with the original PIH model and adjusting it to take into account credit constraints and incomplete financial markets.

The consumption function: a closed-form solution **Solving for the**

consumption function of our representative household usually requires numerical methods. We will examine a special case that delivers some insights that are likely to hold in more general solutions. We focus on a perfect foresight equilibrium with separable log utility

$$u = \ln c + v(l).$$

We also simplify the budget constraint by dropping all wedges except $efp_{h,t}$ and putting all income sources together into one variable y_t.**Define** $R_t efp_{h,t} = R_{h,t}$. **Iterating forward on the budget constraint and using the transversality condition,**

$$\Sigma_{k=0}^{\infty} \frac{c_{t+k}}{\Pi_{j=1}^{k} R_{h,t+k}} = b_t R_{h,t} + \Sigma_{k=0}^{\infty} \frac{y_{t+k}}{\Pi_{j=1}^{k} R_{h,t+j}},$$

where for any x_{t+j}, $\Pi_{j=1}^{0} x_{t+j} = 1$. This is the standard intertemporal budget constraint that is at the heart of the PIH. Note that the intertemporal budget constraint holds for any sequence of shocks, and regardless of the borrowing constraints, as long as the no Ponzi game condition holds.

Now write

$$\frac{c_{t+k}}{c_t} = \Pi_{j=1}^{k} \frac{c_{t+j}}{c_{t+j-1}}.$$

Using this equation together with the Euler equation

$$\frac{c_{t+j}}{c_{t+j-1}} = \tilde{\beta}_{t+j} R_{h,t+j},$$

where $\tilde{\beta}_{t+j} = \frac{\beta_{t+j}}{\beta_{t+j-1}}$**, we can solve out for** c_t **from the intertemporal budget constraint,**

$$\begin{aligned} c_t &= \frac{1}{\Sigma_{k=0}^{\infty} \Pi_{j=1}^{k} \tilde{\beta}_{t+j}} \left(b_t R_{h,t} + \Sigma_{k=0}^{\infty} \frac{y_{t+k}}{\Pi_{j=1}^{k} R_{h,t+j}} \right), \\ \text{where } \beta_0 &= 1 \text{ for } k = 0. \end{aligned}$$

This consumption function is the result of combining the intertemporal budget constraint with the desire to smooth consumption as captured by the Euler equation.

$\frac{1}{\Sigma_{k=0}^{\infty} \Pi_{j=1}^{k} \tilde{\beta}_{t+j}}$ **is often called the Marginal Propensity to Consume (MPC) out of wealth**

$$b_t R_{h,t} + \Sigma_{k=0}^{\infty} \frac{y_{t+k}}{\Pi_{j=1}^{k} R_{h,t+j}},$$

, i.e the change in consumption in response to a marginal/small change in wealth. However, this assumes that the $\tilde{\beta}_{t+j}$ **are independent of wealth. In the special case of** $\tilde{\beta}_{t+j} = \tilde{\beta}$ **the MPC is** $1 - \tilde{\beta}$**. It is typical of the PIH in that consumption depends on the present value of all future income and on current financial wealth.**

The main modifications of our consumption function relative to the standard PIH is that we allow for financial distortions in $R_{h,t+j}$ **and for uninsured idiosyncratic risk effects or deviations from RE through time varying discount factors** $\tilde{\beta}_{t+j}$. In this case, $\frac{1}{\Sigma_{k=0}^{\infty} \Pi_{j=1}^{k} \tilde{\beta}_{t+j}}$ is no longer the consumption out of a marginal change in wealth, though it may still provide a good approximation to the true MPC if uncertainty levels are low.

- The key prediction of the PIH is that more persistent changes in income y_t **with a significant impact over many periods will have bigger effects on current consumption** c_t **than purely transitory shocks. A permanent shock will have the largest impact, with consumption almost responding by the full amount of the change in income.** This prediction is at the origin of the name of the PIH. The idea is that consumption is most strongly related to the component of income composed of permanent shocks.

- Changes in income in a later period y_{t+k} **have a smaller effect on** c_t **than a change in current income** y_t **due to the higher discount**

rate attached to future income streams. This may seem inconsistent with the earlier statement that more persistent income shocks have a bigger effect on consumption. The reconciliation comes from the fact that a more persistent shock has bigger effects on income in later periods. So even though each future income y_{t+k} individually has a smaller impact on consumption than current income y_t , the cumulative effect is larger.

- In the extended PIH model, an increase in income uncertainty or a strengthening of the precautionary saving motive due to lower income or wealth would show up as a higher sequence of $\left\{\tilde{\beta}_{t+j}\right\}$ lowering consumption c_t. An increase in financing frictions for borrowers would have a similar effect through higher $\{R_{h,t+j}\}$.

- Modern consumption/saving theories put a lot of emphasis on the difference in response to changes in cash on hand,

$$x_t = y_t + b_t R_{h,t},$$

versus the present value of future income,

$$PV_{t+1} = \Sigma_{k=1}^{\infty} \frac{y_{t+k}}{\Pi_{j=1}^{k} R_{h,t+j}}.$$

The consumption function can be written as

$$c_t = \frac{1}{\Sigma_{k=0}^{\infty} \Pi_{j=1}^{k} \tilde{\beta}_{t+j}} \left(x_t + PV_{t+1}\right).$$

- The precautionary saving motive is usually decreasing in cash on hand. Therefore, the marginal propensity to consume (MPC) out of cash on hand is higher than that out of the present value of future incomes. More generally,

$$\frac{\partial \tilde{\beta}_{t+j+1}}{\partial x_{t+j}} < \mathbf{0}$$

implies a higher MPC out of income over a short term horizon (e.g 1-4 quarters) than out of medium or long term income (e.g more than 4 quarters ahead). See Carroll (1997) [44],Carroll (2001) [46] for more analysis of these points using the standard model of precautionary savings with uninsured idiosyncratic income risk. The effect of precautionary saving in reducing the MPC out of future income can also be captured through a higher $\{R_{h,t+j}\}$ beyond the effect of explicit external finance premia (see for example the FRB/US model documentation[12]).

[12] http://www.federalreserve.gov/econresdata/notes/feds-notes/2014/a-tool-for-macroeconomic-policy-analysis.html

Furthermore, the reduction in precautionary saving as cash on hand increases is stronger when starting from a lower cash on hand:

$$\frac{\partial^2 \tilde{\beta}_{t+j+1}}{\partial^2 x_{t+j}} > \mathbf{0}.$$

As a result, the consumption function is concave in cash on hand,

$$\mathbf{c}_x > \mathbf{0}, \mathbf{c}_{xx} < \mathbf{0}.$$

This nonlinearity becomes important in the presence of uninsured idiosyncratic risk. Using Jensen's inequality, concavity implies that

$$\begin{aligned} c(\Sigma \omega_i x_i) &> \Sigma_i \omega_i c(x_i), \text{ for any set of population weights } \{\omega_i\}_i, \\ 0 &\leq \omega_i \leq 1, \ \Sigma_i \omega_i = 1. \end{aligned}$$

Any increase in the dispersion of cash on hand $\{x_i\}_i$ across the population reduces consumption. For example, higher inequality in a recession can amplify the negative effect on consumption. In a representative agent model, this effect can be incorporated by making $\tilde{\beta}_{t+j}$ a positive function of income inequality (similar to the earlier incorporation of the unemployment rate into $\tilde{\beta}_{t+j}$).

- An increase in income uncertainty or a reduction in household wealth, by raising $\left\{\tilde{\beta}_{t+j}\right\}$ also reduces the sensitivity of consumption to interest rates $\{R_{h,t+j}\}$. Since income uncertainty and household wealth typically deline in a recession, household consumption should be less sensitive to interest rates in general, and to monetary policy more particularly, in an economic downturn.

While our derivation has used log utility, the perfect foresight consumption function can be easily generalised to the case of other utility functions that imply an Euler equation of the form

$$\frac{c_{t+j}}{c_{t+j-1}} = g_c(x_{t+j}, \tilde{\beta}_{t+j}, R_{h,t+j}) R_{h,t+j}.$$

x **is a vector of variables excluding consumption. The same steps as for the log utility case lead to the consumption function**

$$c_t = \frac{1}{\Sigma_{k=0}^{\infty} \Pi_{j=1}^{k} g_c(x_{t+j}, \tilde{\beta}_{t+j}, R_{h,t+j})} \left(b_t R_{h,t} + \Sigma_{k=0}^{\infty} \frac{y_{t+k}}{\Pi_{j=1}^{k} R_{h,t+j}} \right).$$

For example, this Euler equation holds for the frequently used Cobb-Douglas function

$$u(c,l) = \frac{(c^{\theta} l^{1-\theta})^{1-\sigma}}{1-\sigma}$$

In this case, x_{t+j+1} includes the relative leisure time $\frac{l_{i,t+j+1}}{l_{t+j}}$. More generally we could allow for demographic variables to reflect the age composition of the population. For the common case of

$$\boldsymbol{\sigma > 1} \text{ we have } \frac{\partial g_c(.)}{\partial R_h} \boldsymbol{< 0}$$

so that higher future interest rates increase the marginal propensity to consume out of wealth. The intuition is that in this case the income effect of higher interest rates dominates the substitution effect and households want to consume more of their wealth when interest rates increase.

At the same time, higher interest rates directly reduce wealth for borrowers via $b_t R_{h,t}$ and for all households via the discounted value of future income $\Sigma_{k=0}^{\infty} \frac{y_{t+k}}{\Pi_{j=1}^{k} R_{h,t+j}}$. As a result, current consumption c_t is usually still decreasing in interest rates when $\sigma > 1$.

Note that in equilibrium all the variables on the right hand side of this equation will usually follow a Markov process, so the the consumption function is also a Markov process in the model's state variables. This makes predicting and simulating the behaviour of consumption relatively straightforward once we have solved the household optimisation problem. Also, it simplifies comparison the consumption function from our optimisation based model to simpler statistical models such as VAR's that are also Markov processes.

3.2.3 The external finance premium and borrowing constrained households

Here we show how the external finance premium for the representative household $efp_{h,t+1}$ can emerge in an environment with heterogeneity, where a subset of households are borrowing constrained. The model of household credit constraints in Iacoviello (2005) [111] and the savers-spenders model of Mankiw (2000) [143] are standard examples. Imagine that a fixed proportion θ_{bo} of households always borrow at a rate $R_{bo,t} > R_t$. The remaining $\theta_{sa} = 1 - \theta_{bo}$ of households are savers who always lend at $R_{sa,t} = R_t$. Assume they both have a utility function of the form

$$\frac{c_i^{1-\sigma}}{1-\sigma} + v(l_i), i = bo, sa.$$

Define the percentage deviation from the steady state of a variable x_t to be

$$\hat{x}_t = \frac{x_t - x}{x}.$$

Then a first order approximation in percentage deviations from the steady-state for each households' Euler equation is

$$\hat{c}_{i,t} = E_t \hat{c}_{i,t+1} - \frac{1}{\sigma} \hat{R}_{i,t+1} + v_{i,t+1}^{b},$$

where $v^b_{i,t+1}$ is a household specific discount factor shock. Aggregate consumption can be written in percentage deviations from the steady state as

$$\hat{c}_t = \frac{\theta_{bo}c_{bo}}{c}\hat{c}_{bo,t} + \frac{\theta_{sa}c_{sa}}{c}\hat{c}_{sa,t}.$$

Using this with the Euler equations implies

$$\hat{c}_t = E_t\hat{c}_{t+1} - \frac{1}{\sigma}\left(\frac{\theta_{bi}c_{bo}}{c}\hat{R}_{bo,t+1} + \frac{\theta_{sa}c_{sa}}{c}\hat{R}_{t+1}\right) + (\frac{\theta_{bo}c_{bo}}{c}v^b_{bo,t+1} + \frac{\theta_{sa}c_{sa}}{c}v^b_{sa,t+1}).$$

Now, $\frac{\theta_{bo}c_{bo}}{c}\hat{R}_{bo,t+1} + \frac{\theta_{sa}c_{sa}}{c}\hat{R}_{t+1}$ can be seen as a first order approximation for some $R_{t+1}efp_{h,t+1}$. We can derive a similar result for the nonlinear Euler equation under perfect foresight, though there the weight of the different $R_{i,t+1}$ in the aggregate $R_{t+1}efp_{h,t+1}$ will no longer be constant, and heterogeneity in the discount factor shocks will also affect the definition of $efp_{h,t+1}$. To see this, start with the individual nonlinear Euler equations for $i = sa, bo$

$$c_{i,t+1} = \left(\tilde{\beta}_{i,t}R_{i,t+1}\right)^{1/\sigma} c_{i,t}.$$

Using

$$c_t = \Sigma_i\theta_i c_{i,t},$$

and defining

$$\tilde{\beta}_t = \Sigma_i\theta_i\tilde{\beta}_{i,t}$$

we can rewrite the Euler equation as

$$\begin{aligned}
c_{t+1} &= \Sigma_i(\tilde{\beta}_{i,t+1}R_{i,t+1})^{1/\sigma}\theta_i c_{i,t} \\
&= c_t\tilde{\beta}^{1/\sigma}_{t+1}\Sigma_i\left(\frac{\tilde{\beta}_{i,t+1}}{\tilde{\beta}_{t+1}}R_{i,t+1}\right)^{1/\sigma}\frac{\theta_i c_{i,t}}{c_t} \\
&= c_t\left(\tilde{\beta}_{t+1}\left[\Sigma_i\left(\frac{\tilde{\beta}_{i,t+1}}{\tilde{\beta}_{t+1}}R_{i,t+1}\right)^{1/\sigma}\frac{\theta_i c_{i,t}}{c_t}\right]^{\sigma}\right)^{1/\sigma} \\
&\equiv c_t\left(\tilde{\beta}_{t+1}R_{t+1}efp_{h,t+1}\right)^{1/\sigma}
\end{aligned}$$

with the appropriate definitions.

Note that the household external finance premium can reflect more general distortions in household credit markets than just credit spreads between risk free assets and household loans. In particular it also captures quantity constraints on borrowing that show up in first order conditions as Lagrange multipliers. A tighter credit constraint due to lower income or collateral increases Lagrange multiplier and raises the expected growth rate of consumption. Therefore, the tighter credit constraint can be mapped into a higher $efp_{h,t+1}$.

The most tractable model capturing credit constraints is one where a subset of households simply consume all their labour income each period. These "Rule of Thumb" (ROT) households can be seen as highly impatient but without any collateral for borrowing. Therefore they cannot borrow, and are constrained to simply consuming their income.

Krusell and Smith (2006) [134] suggest that this 2 household model can provide a good approximation to the aggregate consumption behaviour of richer heterogeneous household models with uninsured idiosycratic income risk and differences in discount factor across households. We will examine a slight generalisation of this model, where a subset of households are constrained by an exogenous borrowing limit $\bar{b}_{t+1}$. Ultimately, once we add durable goods, this 2 household setup will be part of our preferred cannonical DSGE model.

Borrowers are more impatient than savers: that is their discount factor on the BGP is $\tilde{\beta}_{bo} < \tilde{\beta}_{sa}$. [13]**As a result, borrowers borrow up to the maximum credit limit on the BGP.** We assume fluctuations are close enough to the BGP, and ROT households are impatient enough such that the credit constraint remains binding outside the BGP. Alternatively, we can imagine that big positive shocks that would normally cause households to save some of their income gains also induce significant reductions in uninsured idiosyncratic risk or significanly higher optimism about future income prospects. As a result, borrowers end up borrowing as much as possible despite the higher income. The borrowing limit is stochastic, allowing us to capture in reduced form a tightening of credit constraints. We also assume that borrowing households do not own shares in firms. This captures the fact that poorer or more highly indebted households typically own little stock market wealth or private equity.

The representative borrower household solves

$$
\begin{aligned}
V_0^{bo} & \quad \max_{\{c_{bo,t}, n_{bo,t}, b_{bo,t+1}\}} E_0 \Sigma_{t=0}^{\infty} \beta_{bo,t} u(c_{bo,t}, l_{bo,t}) \\
& \quad \text{subject to a sequence of constraints} \\
c_{bo,t}(1+\tau_{c,t}) + b_{bo,t+1} & \leq b_{bo,t} R_t efp_{bo.h,t} + w_t(1-\tau_{nst}) n_{bo,t} + T_t,\ (\lambda_t), \\
& \quad -b_{bo,t+1} \leq \bar{b}_{t+1},\ (\mu_{t+1}),\ \bar{b}_{t+1|} \geq 0. \\
n_{bo,t} & = 1 - l_{bo,t}, \\
u_c & > 0, u_{cc} < 0, u_l > 0, u_{ll} < 0, u_{c,l} \leq 0.
\end{aligned}
$$

The optimality condition for bonds changes to

$$
b_{bo,t+1} : \lambda_{bo,t} = E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} efp_{bo,h,t+1} R_{t+1} \lambda_{bo,t+1} + \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \mu_{t+1}.
$$

[13] Higher impatience can be seen as a proxy for higher expected income growth of younger consumers, which given their relatively low current income and the consumption smoothing motive encourages them to borrow more.

The other optimality conditions remain the same.

The first thing to note is that as promised, through appropriate redefinitions a binding borrowing constraint with $\mu_{t+1} > 0$ maps into a higher external finance premium $\frac{R_{bo,t+1}}{R_{t+1}}$, so we can analyse a tightening of the borrowing constraint as an increase in the household external finance premium. Second, in this model with an always binding borrowing constraint the consumption function of borrowers is trivial:

$$c_{bo,t}(1+\tau_{c,t}) = w_t(1-\tau_{nst})n_{bo,t} + T_t - \bar{b}_t(R_t efp_{bo,h,t} - 1) + (\bar{b}_{t+1} - \bar{b}_t).$$

The household consumes its wage income net of taxes and debt repayments. This leads to a traditional Keynesian multiplier effect on aggregate consumption: shocks that reduce output lead to lower wage income, which reduces the consumption of income constrained households. Under some conditions (e.g sticky prices, or spillovers from lower aggregate demand to TFP), this leads to a further decline in output.

Note that in the model in which the borrowing constraint always binds, the borrower's behaviour is similar to a PIH household that perceives all income shocks to be permanent. In both cases, the MPC out of income is close to 1. This highlights a potential difficulty in identifying borrowing constraints from the response of households to income shocks. High MPC's may be observed either due to borrowing constraints, or because households think that the shock will be quite persistent (e.g due to imperfect attention to mean reversion).

Debt repayments have been decomposed into interest payments $\bar{b}_t(R_t efp_{bo,h,t} - 1)$ and principal repayments $-(\bar{b}_{t+1} - \bar{b}_t)$. The typical model with rule of thumb households assumes $\bar{b}_{t+1} = 0$ for all t, that is households cannot borrow and simply consume their after tax income. Finally, in this model market clearing conditions change to

$$\begin{aligned}
\Sigma_i \theta_i c_{it} &= c_t, \\
\Sigma_i \theta_i n_{i,t} &= n_{st}, \\
\Sigma_i \theta_i b_{i,t} &= 0.
\end{aligned}$$

Aggregate consumption c_t is now a function of future income levels, interest rates and current income with the coefficient on current income depending on the proportion of income earned by rule of thumb households.

Estimates of the proportion of credit constrained households θ_{bo} range between 0.15 and 0.5, with numbers in the lower range being more common in medium size policy models (e.g the IMF's GIMF model (2013) [7], Bank of

England COMPASS model (2013) [39],Gali et al et al (2007) [92]). Hurst (2006) conducts tests of excess sensitivity to predictable income shocks at the level of individual US households (using the PSID). He finds a group of around 20-30% of households that have a very high MPC, together with significant drops in consumption income in retirement that are hard to justify using various control variables. He links this group of households explicitly to questions in the PSID about lack of financial planning, arguing that 20-30% of households can be seen as ROT households that do not try to smooth their consumption. Note that the actual consumption share of these households can be significantly lower, because usually their average consumption $Ec_{bo,t} < Ec_{sa,t}$.

Finally, note that the aggrecate consumption function can also be reinterpreted as a model of optimisation error by the representative household, with θ_{bo} reflecting a tendency of households to systematically over respond to changes in their current income. For example this optimisation error could arise if current income is more salient in people's minds when judging how much to save.

3.3 Labour supply

The labour supply first order condition equates the marginal cost of labour effort in terms of leisure to the marginal benefit. Here the marginal benefit is the wage adjusted for the shadow cost (Lagrange multiplier) of the constraint - i.e. the value of the extra marginal labour effort in relaxing the budget constraint. Combining first order conditions, **we can rewrite the labour supply condition as**

$$u_{l,t} = u_{c,t} \frac{(1 - \tau_{ns,t})}{1 + \tau_{c,t}} w_t.$$

If we interpret our model literally, the representative household makes a continuous choice of working hours. In other words, labour supply is divisible. In reality there are significant indivisibilities in labour supply choices at the household level, e.g one can work at a part-time or full time job with limited adjustments in terms of overtime or hours reductions. But aggregate labour supply is continuous. The divisible labour supply model can always be trivially reinterpreted as the decisions of a large family that measures family welfare according to $u(c, 1 - n_s)$ on what proportion of its members n_s to send to work. The family sends more members to work until the marginal disutility of sending an extra member to work is equal to the wage evaluated at the family's marginal utility of wealth.

The large family assumption is clearly extreme. However, even if hours of work are indivisible the assumption of divisible labour supply of the representative household can be fully justified in the presence of highly competitive labour markets and complete financial markets (Mulligan (2001)[151]). The divisible labour representative agent's decisions are equivalent under these conditions to the aggregate of the decisions of many agents with partially divisible labour supply. So a reduction in the labour supply of the representative household in response to lower wages can be mapped into a lower rate of job acceptance by the unemployed facing lower wage offers, as well as higher separations due to firms being more reluctant to pay salaries above the reservation wage for some workers.

Following Mulligan (2001) [151], we can interpret the labour supply first order condition in a model with complete financial markets as representing the tradeoff between the cost of working u_l and the benefit of working $u_{c,t}\frac{(1-\tau_{ns,t})}{1+\tau_{c,t}}w_t$ of the marginal worker who is (approximately) indifferent between working and being out of the labour force. Each worker j has to choose between working a fixed amount of hours for a benefit of $u_{c,t}\frac{(1-\tau_{ns,t})}{1+\tau_{c,t}}w_{j,t}\varepsilon_{j,t}$ and not working with a benefit of $u_{l,t}$. Agents for which the opportunity cost of working is lower so that

$$u_{l,t} < u_{c,t}\frac{(1-\tau_{ns,t})}{1+\tau_{c,t}}w_{j,t}\varepsilon_{j,t}$$

end up working. Agents for which working is costly relative to the benefits such that

$$u_{l,t} > u_{c,t}\frac{(1-\tau_{ns,t})}{1+\tau_{c,t}}w_t\varepsilon_{j,t}$$

are unemployed or out of the labour force. For the marginal worker, $\varepsilon_{j,t} = 1$ and the labour supply condition holds with equality.

I use a utility function $u(c,l)$ that is compatible with a a balanced growth path to capture the lack of a a strong trend in hours of work over the second half of the 20th century (there have been some regime changes for example in Europe reducing hours of work, but they do not seem to follow a deterministic trend). Within the class of constant elasticity of substitution functions, the long run invariance of hours worked to trend output growth requires Cobb-Douglas utility function of the form

$$u(c,l) = \frac{(c^{\theta}l^{1-\theta})^{1-\sigma}}{1-\sigma}.$$

I allow for nonseparable utility due to the strong evidence that consumption and leisure are substitutes. Some models of the extensive labour supply margin also predict non-separable utility functions for the representative household (Basu and Kimball (2002) [21],Euseppi and Preston (2009)[77]). Complementarity of work and consumption can explain some of the comovement between current income and

consumption, as well as part of the difference in consumption levels between workers and the unemployed. To see this consider the consumption Euler equation with a non separable Cobb-Douglas utility function under perfect foresight:

$$\left(\frac{c_{t+1}}{c_t}\right)^{1-\theta(1-\sigma)} = \tilde{\beta}_{t+1} R_{t+1} efp_{h,t+1} \left(\frac{l_{t+1}}{l_t}\right)^{(1-\theta)(1-\sigma)}.$$

For $\sigma = 1$ we are back in the case of a separable utility function, and consumption growth is only a function of discount factor shocks and interest rates and external financing constraints. If $\sigma > 1$, consumption and work effort are complements. In this case an increase in working effort at $t+1$ relative to t (lower $\frac{l_{t+1}}{l_t}$) leads to an increase in the expected growth rate of consumption $\frac{c_{t+1}}{c_t}$.

The expected work effort term $\frac{l_{t+1}}{l_t}$ acts as a shift in the discount factor, with higher work effor now relative to the future making the household more "impatient". For moderate non separability the conclusions with separable utility can still provide a reasonable approximation. But the complementarity between work and consumption improves the empirical fit of the consumption function by explaining some of the observed excess sensitivity of consumption to current income relative to the separable utility function benchmark. Nonseparable preferences also weaken the impact of income effects on labour supply by strengthening the link between work hours and consumption.

A meta analysis of many studies by Chetty (2006) [52] suggests that the degree of risk aversion and non-separability that is compatible with the empirical evidence on labour supply and changes in consumption in response to unemployment or disability cannot be too high. In particular the relative risk aversion coefficient σ should not exceed 2, with values closer to 1 being more plausible. Havranek et al (2015) provide a comprehensive meta-analysis of empirical estimates of the intertemporal elasticity of substitution $\frac{1}{\sigma}$. Their analysis suggests an average estimate of $\sigma \in [1, 2]$, depending on the country or the weighting of different estimats in the same study.

In the general case with uninsured idiosyncratic risk, individual labour supply with indivisibilities does not aggregate exactly to that of a representative agent (Chang et al (2011)[49]). Nevertheless the representative household model can provide a useful approximation in cases when full modeling of indivisibility with heterogenous agents is computationally too costly (see Chang et al (2011) [49] for some shortcuts that can be used to improve the match of the representative agent model to results from a more realistic heterogeneous agents model).

Early models of indivisible labour were used to justify a very high macro elasticity of labour supply, helping to explain large employment fluctuations as movements along the labour supply curve (see Kydland (1995). [62]). Mulligan

(2001) [151] , Erosa et al (2014)[75] and Mustre del Rio (2012) show that when incorporating a realistic level of heterogeneity in workers' reservation wages the aggregate Frisch (constant marginal utility of consumption) labour supply elasticity with indivisible labour is closer to that of reasonably calibrated divisible labour models at around 0.7 to 1.75. The lower number takes into account only the extensive margin, while the higher number also takes into account the response of hours per job (the intensive margin). [14] It is common to dismiss changes in the intensive margin as insignificant. Yet in the US hours per job adjustment typically accounts for one third of the volatility of total hours. In the recent financial crisis of 2008, hours per job changes accounted for most of the total hours adjustment in some European countries such as Germany (Ohanian and Raffo 2009).

The divisible labour representative household model also becomes a better approximation once we take into account that many of the marginal workers whose employment fluctuates the most across the business cycle usually have a choice between looking for part time of full time employment, and to the degree that the typical household contains at least 2 working age members. Mulligan and Rubinstein (2004) analyse household labour supply aggregating over all household members, and find that in contrast to individual worker employment where the extensive employment status dominates, the intensive hours margin dominates labour supply from the household perspective. They argue that labour supply of couples is much smoother than that of individual workers. In part this is because it is much less likely for both members of a couple to be unemployed than for a single member. Finally, in studying tax policies one has to take into account differences in labour supply elasticity between different income or wealth and education groups. From the representative household perspective, we need to use the labour supply elasticity of the group that is actually most affected by a tax policy in studying its macroeconomic impact, adjusted by its weight in the active population.

The standard labour supply equation is at the heart of the difficulties of the basic neoclassical business cycle model in generating comovement between

[14] With a cobb-douglas utility function, the representative's household labour supply elasticity is approximately $\frac{l}{n^s}$ close to the BGP. This makes it hard to reconcile typical estimates of the proportion of hours worked in the population (e.g 1/4 of the hours in a year or less spent working, and an employment rate around 60%-70%) with such a low aggregate labour supply elasticity if we interpret leisure l to include all non working hours. However, a reconciliation is possible if we interpret the utility function of the representative household to be over the amount of non work hours relative to a minimal fixed amount of non work time, dedicated to sleeping and other core home production activities. From that perspective we can think of the utility function as

$$\begin{aligned} & u(c, 1 + l_{\min} - n^s - l_{\min}) \\ = \quad & u(c, 1 - n^s), \end{aligned}$$

where $l_{\min}$ is a fixed minimum amount of time that has to be spent outside of work, and $l = 1 - n^s$ as before.

consumption and investment in response to demand shocks or to intertemporal shocks. Consider a separable utility function. An increase in consumption reduces the marginal utility of consumption and therefore the marginal utility of leisure for a given wage and taxes. By the concavity of utility in leisure, leisure must increase and labour supply must decline. The fall in labour supply makes it harder for an increase in consumption to be accompanied by an increase in output and investment. Bilbiie (2006) [30] shows this logic also holds when consumption and leisure are substitutes as long the utility function is concave. There are several mechanisms that have been shown to partially solve this comovement problem.

An indivisible labour model with nonseparable utility in consumption and leisure and both extensive and intensive margin can generate positive comovement between aggregate consumption and labour supply(Euseppi and Preston (2009) [77]). The idea is that even though at the individual worker level higher consumption is associated with lower labour supply, because consumption and leisure are substitutes an increase in wealth can cause consumption and aggregate labour supply to both fall because of a composition effect: suppose a shock causes an increase in individual household consumption and a reduction in labour supply. At the individual household we still have negative consumption-labour supply comovement. However, because part of the reduction in labour supply comes through lower labour force participation and the employed consume more, aggregate consumption declines.

Consider news of a future increase in productivity. This typically generates a positive wealth effect, reducing labour supply. The reduction in labour supply generates a recession in response to a positive news shock. Euseppi and Preston show that if we add investment adjustment costs to the model, the substitution effect of higher future productivity dominates and the positive news lead to higher labour supply and output. In this case at the individual level consumption falls and labour supply goes up. Aggregate consumption nevertheless increases because of a higher labour force participation rate. However, the realism of a reduction in employed households' consumption in response to positive news is doubtful, so this probably cannot replace other mechanisms generating expectations driven business cycles through shifts in labour demand. [15]

Habit formation as in Christiano et al (2007) [55] is another factor that can contribute to comovement for demand or news shocks. With habit formation, while higher current consumption lowers the marginal utility of wealth and reduces labour supply, higher future consumption increases the marginal utility of wealth and raises labour supply. Habit formation or other consumption adjustment costs can be captured in our setup as part of the consumption wedge $\tau_{c,t}$. In this case $\tau_{c,t}$ is an increasing function of the deviation of consumption from the stock of habits $S_t^c = f(c_{t-1}, .., c_{t-k})$. Incorporating durable goods can

[15]Note that their mechanism works through a non-concave utility function of the representative household, so it is compatible with the analysis in Bilbiie (2006).

similarly alleviate the comovement problem under some circumstances. We will discuss several mechanisms that can enhance comovement of consumption and investment over the business cycle, including effects of demand or news shocks on TFP, labour adjustment costs and the presence of consumer durables investment.

Finally, we examine more formally the effect of worker heterogeneity of the borrower/saver form. In general, differences between workers will generate movements in the effective labour tax rates faced by the representative household. Consider the simplest case of

$$u(c,l) = \frac{(c^b l^{1-b})^{1-\sigma}}{1-\sigma}.$$

The first order condition for type i household is

$$\frac{1-b_i}{l_{i,t}} = \frac{b_i w_t}{c_{i,t}} \frac{1-\tau_{ns,i,t}}{1+\tau_{c,i,t}}.$$

Manipulating this equation and aggregating over all households,

$$\begin{aligned} l_t &= \Sigma_i \theta_i l_{i,t} = \frac{c_t}{w_t} \Sigma_i \frac{1-b_i}{b_i} \frac{1+\tau_{c,i,t}}{1-\tau_{ns,i,t}} \frac{\theta_i c_{it}}{c_t}, \\ c_t &= \Sigma_i \theta_i c_{i,t}. \end{aligned}$$

Clearly if all wedges affecting labour supply $\tau_{c_i}, \tau_{ns,i}$ and leisure preference parameters are identical across households, we get perfect aggregation of individual labour supplies into aggregate labour supply. In the more realistic case in which different groups of households have different preferences for leisure and face different labour supply or consumption wedges we will not get perfect aggregation. Significant shifts in the composition of aggregate consumption across different groups or in the wedges facing different groups will affect aggregate labour supply. We will have more to say on these effects when discussing shocks to the labour wedge.

4 Firms

4.1 The firm's decision problem

There is a measure 1 continuum of firms. The representative firm picks sequences of n_t, m_t to maximise

$$\begin{aligned} d_t &= g(A_t m_t, e_{m,t}) z_t f(A_t n_t) - w_t(1+\tau_{nd,t})(n_t + m_t), \\ g_m &> 0,\ g_{mm} < 0,\ g_{e_m} > 0,\ f_n > 0,\ f_{nn} < 0, \\ g(am,.)f(an) &\leq a g(m,.) f(n) \text{ for } a > 0.\ A_t = G_y A_{t-1}, G_y \geq 1. \end{aligned}$$

A_t **is the labour augmenting component of total factor productivity, whose growth rate** $G_y - 1$ **determines the long run growth, absent any shocks, of output.**

Following the analysis of the benchmark New-Keynesian model we have abstracted from capital. More generally we have assumed $f_{nn} < 0$, a fixed capital stock and no feedback from the price of capital into any credit constraints. The fixed capital stock allows us to ignore investment.

z_t **is a total factor productivity (TFP) shock, also known as the efficiency wedge using the terminology of business cycle accounting. In addition to actual changes in the technology, it can also reflect more generally any factor that changes the efficiency of the economy for a given level of aggregate capital and labour input.**

An increase in the proportion of financially distressed firms can reduce the utilisation rate of capital and labour if a firm's normal functioning is hindered by liquidation procedures or by harsher debt covenant restrictions. Procyclical competition in product markets can generate procyclical TFP fluctuations from the perspective of the aggregate production function (Jaimovich and Floetotto (2008) [114]).

Misallocation of production inputs across firms due to differences in costs of financing or in the degree of imperfect competition faced by different producers also reduces the output of the economy for any given level of aggregate labour or capital inputs. For example if higher financial frictions disproportionately affect smaller firms, the dispersion in marginal products of inputs across firms increases. With diminishing marginal productivity this higher dispersion reduces aggregate TFP (see Khan and Thomas (2013) [125], and (Buera and Moll (2012) [35]for examples).

Higher costs of intermediate production inputs also reduce TFP from the perspective of a value added production function in terms of labour and capital (see Chari et al 2007 [50], Jones (2011,2011b) [117], [116] and the section on the input output structure in business cycle models). In all these cases, there is a decline in z_t.Chari et al call z_t the efficiency wedge and find that it can explain most of the output decline during the US Great Depression in the 1930's. The section on real business cycle models with variable capital discusses these endogenous TFP channels in more detail. [16]

$g(.)$ **represents distribution, matching marketing and commercialisation efforts due to e.g. search frictions as in Rios-Rull et al. (2012)**

[16] Note that these mechanisms for endogenising TFP allow us to reduce the scope of the typical critique of these shocks that we do not seem to observe technological regress in recessions (see King and Rebelo (2000) for a typical statement of this critique).

[17]. We can think of $g(.)$ as the utilisation rate of production capacity $z_t f(A_t n_t)$. Here we take it as a function of the amount of labour dedicated to marketing and distribution, and vector of other factors $e_{m,t}$. $e_{m,t}$ can reflect changes in regulation of shopping, time varying entry and exit rates of new stores, changes in consumer shopping efforts or frictions in consumer financing that affect the efficiency of the distribution and marketing sector. In some models (e.g Rios Rull et al (2012) [17]) shopping effort is endogeous. The impact of a shock to $e_{m,t}$ on business cycles is qualitatively similar to a TFP shock. We will examine several extensions that try to endogenise $e_{m,t}$.

The general idea captured in models with a non trivial $g(.)$ function is that the distribution, delivery and matching of goods and services to buyers are key components of their value as measured by GDP. In fact distribution costs can represent more than 40% of the retail price of a product (Burstein et al. (2001)). A reduction in the efficiency of the distribution and marketing process as measured by $g(,)$ reduces the efficiency of the economy just like a reduction in the productivity of its factories or a worse allocation of production inputs across firms. Therefore, from the perspective of a value added production function in capital and labour, any decline in the efficiency of the matching and distribution process $g(.)$ will act like a negative TFP shock.

$g(.)$ shocks also capture more general shifts in the efficiency of the matching process between buyers and sellers. For example changes in consumer tastes across sectors or negative sector specific technology shocks usually require adjustments in the composition of output across sectors. But in an economy with frictions, it may take some time for the matching process in new sectors or in sectors with higher TFP to adjust to the inflow of new buyers and sellers into these sectors, especially if the new entrants are slow to adapt their marketing and search strategies to the new market. While this adjustment occurs, matching in the economy is less efficient (see for example Black (1986) [32]). This effect may be especially relevant in economies that have gone through strong business or real estate investment booms, when the boom turns to a bust. In that case, difficulties in shifting the structure of the economy from an excessive reliance on investment can amplify a recession. Note that this effect is assymetric: the same intersectoral adjustment costs would dampen the response of the economy to positive shocks requiring reallocation across sectors.

Using the terminology of Rios Rull et al (2012) [17], these are all demand shocks that act like productivity shocks because they change the aggregate TFP of the economy. In an economy with randomeness in the matching process (e.g Angeletos and Lao (2012) [11]) a financial shock can destroy existing longer term trade relations, increasing the idiosyncratic matching risk facing firms. The increased uncertainty in matching raises the costs of finding buyers for one's products and reduces the efficiency of the matching process. This mechanism would also map into a negative shock to $g(.)$.

As for the other firm level wedges, once again we interpret taxes more broadly here to include various private sector distortions as well as government taxation. For example, we will show that $\tau_{nd,t}$ can include financing frictions in the funding of the firms' wage payments, countercyclical markups in an environment with imperfect competition or labour market search frictions that raise the cost of hiring. Note that the TFP wedge is equivalent to a revenue tax that is not rebated to the households in T_t. We could also have included a revenue tax rebated to the household in T_t, but this would be equivalent to assuming a symmetric tax on all production inputs. Therefore, we can ignore the revenue tax without loss of generality.

Optimal firm decisions imply

$$\begin{aligned} n_t &: \quad g_t z_t f_{n,t} = w_t(1+\tau_{nd,t}), \\ m_t &: \quad g_{m,t} z_t f_t = w_t(1+\tau_{nd,t}). \end{aligned}$$

These optimality conditions have the classic interpretation that the firm hires more labour as long as the extra benefit exceeds the extra cost $w(1+\tau_{nd})$.**A higher wage or higher labour tax means that at the margin the firm requires a higher return to hiring someone or to raising the number of hours. With diminishing marginal productivity this causes the firm to reduce employment. An increase in productivity makes the firm increase employment for activities that were not profitable before.**

Formally, since the higher productivity raises the marginal product of labour, and at the optimum the firm hires until the marginal product equals the marginal cost of employment, the firm keeps the marginal product equal to the marginal cost by hiring more. Note that in this formulation where the only cost of distribution is in terms of wages, the ratio $\frac{m_t}{n_t}$ is independent of $w_t(1+\tau_{nd,t})$.[17]

4.2 The neoclassical labour market model, unions, long term contracts and wage rigidity

The description of labour supply and labour demand above constitutes the neoclassical labour market model, a benchmark for much of the analysis both in labour economics and in macroeconomics. The key condition characterising labour markets in our model is

$$\frac{u_{l,t}}{u_{c,t}} \frac{1+\tau_{c,t}}{(1-\tau_{ns,t})} = \frac{g_t z_t f_{n,t}}{(1+\tau_{nd,t})}.$$

The left hand side of this equation can be seen as the reservation wage of the representative worker. At any wage above the reservation wage, the representative household has an incentive to further

[17] If we had separate wedges on the wage bill of the two types of workers, $\frac{m_t}{n_t}$ could be affected.

increase labour supply and vice versa if the wage is below the reservation wage. The right hand side of the equation is marginal value of an extra unit of labour to the representative firm. At any wage above this marginal value the firm has an incentive to reduce employment and vice versa for any wage below the marginal value.

We have already discussed how these aggregate conditions reflect aggregation of labour supply over heterogenous workers who can only adjust their hours in discrete amounts (partial labour indivisibility). In those circumstances, the left hand side of the labour market equilibrium condition represents the marginal worker's reservation wage (the worker who is just indifferent between working and not working). Labour market equilibrium equalises the representative worker reservation wage and the representative firm's marginal return to hiring. This represents an outcome in which all mutually beneficial trades between employers and workers have been realised.

The neoclassical labour market equilibrium condition seems at first hard to relate to labour markets with significant union power and centralised bargaining such as those in some European countries. But in fact, if unions and employers bargain efficiently the equilibrium equation above provides a good approximation. If the union demands wages above the reservation wage for the marginal workers (those whose jobs are at risk or who haven't been hired yet), labour supply would exceed labour demand. Therefore, there would be union members who would increase their welfare by working at the given wage and would accept a lower wage to work, but who would be rationed out of the labour market. Furthermore, firms would be willing to hire these workers at the lower wage. In reverse if employer representatives demand wages below the marginal return to hiring, some employers would benefit from proposing lower wages in order to hire more workers at a profit and workers would be willing to work at those higher wages. Efficient bargaining between firms and workers would not allow these mutually beneficial transactions to go unfulfilled.

More formally, as the number of workers and firms represented in labour market negotiations increases, then under plausible conditions the set of bargaining outcomes compatible with no deviations by various coalitions of firms and workers shrinks towards the competitive equilibrium condition above (for a proof, see the analysis of convergence of the core of an economy towards competitive equilibrium in Mas Collel et al (1995)[146], ch 18)). Of course, to achieve this outcome requires significant flexibility in terms of the ability to have different wages for different worker experience levels (possibly proxied by tenure) and skills.

Real world imperfect information about effort and worker abilities can limit contract flexibility. We can capture inefficiency in the labour market bargaining process, e.g due to excessive wage rigidity in union demands, in reduced form as part of the labour supply wedge τ_{ns}. [18] Similarly, it is easy to extend this analysis to allow for some market power by unions (imperfect substitutability between unionised and non unionised workers) leading to monopoly rents through the labour supply wedge τ_{ns} (with higher τ_{ns} representing higher monopoly power) . But conditional on this market power, efficient bargaining between unions and employers should still lead to something close to the neoclassical labour market condition above.

The neoclassical labour market model embodied in the representative household and firm maximisation problems above also seems to clash with the overwhelming empirical evidence that wages in ongoing employment relations are quite stable in response to most shocks (Babecky et al (2009a) [15], Basu et al (2010) [19]). However, it is compatible with a more realistic model in which firms enter long term contracts that provide their workers partial insurance against shocks while the wage of the marginal workers (new hires or the first workers that would be fired at a constant wage) is set competitively (see Barro (1977) [20] and Pissarides (2009)[157] for related arguments about the relevance of wage rigidity in determining aggregate labour hours).

This poses a measurement problem in matching w_t in our model directly to average wage series or to wages reported in microeconomic panel data sets that do not distinguish between ongoing jobs and new jobs. w_t in the model is more directly comparable to the wages of newly hired workers, though even this measure does not fully capture the actual cost of the marginal worker to the firm in a world with long term labour contracts. At any rate wages of new hires are usually found to be quite responsive to aggregate conditions (Carneiro et al (2009) [43],Bellou and Kaymak (2011) [25] ,Pissarides (2009)[157]), even in some European countries known for their rigid labour legislation such as Portugal.

Babecky et al (2009b)[16] examine more generally compensation practices in a survey of European firms. They find that rigidity in wage setting is compensated at least to some degree by adjustments to other margins such as overtime pay, bonuses, delayed promotions or hiring new workers with lower wages. Pissarides (2009)[157] summarizes evidence from studies that distinguish wages on new jobs from those of job stayers. He argues that labour market models with flexible wage adjustment match the empirical cyclicality of the wages of new hires in the US and Europe quite well. Kudlyak (2010) [135] tries to measure the actual cost of the marginal worker using more sophisticated methods in the US. Unlike simply looking at the wages of new hires, Kudlyak takes into account the fact that in long term contracts wage reductions could take the form of lower future wages relative to the market wage, rather than a cut in the entry wage of a new worker. She finds the cost of labour is up to 3 times as procyclical as the average wage. While her method has not yet been applied in other countries, the cumulative evidence just cited suggests that the most relevant measure of wages for labour demand can be quite flexible.

[18] See the section on nominal wage rigidity for a formalisation of this insight.

Here we provide a basic illustration of the relation between the neoclassical labour market model, long term labour contracts and wage rigidity, as long the marginal workers' wages are flexible. Suppose the representative firm uses 2 kinds of workers, permanent workers n_t^p and marginal workers n_t^m. The firm commits to keep a fixed amount of permanent worker hours $\bar{n}^p + \bar{m}^p$ at a fixed wage rate w^p. Meanwhile marginal workers can be be adjusted flexible at at a competitive wage w_t. We assume that both types of workers enter the production function as perfect substitutes. Define

$$n_t = n_t^p + n_t^m, m_t = m_t^p + m_t^m.$$

Also we start with the case in which shocks are small enough that the firm would never want to have less worker hours than the number of permanent hours. The firm's cash flow is now

$$d_t = g(A_t m_t, e_{m,t}) z_t f(A_t n_t) - w_t(1+\tau_{nd,t})(n_t^m + m_t^m) - w^p(1+\tau_{nd,t})(\bar{n}^p + \bar{m}^p).$$

The first order conditions are exactly the same as in the original model as long as the firms' labour demand satisfies $n_t > \bar{n}^p$ and $m_t > \bar{m}^p$ (which we have assumed to hold).

Now go back to the representative household's optimisation problem. We have to think a bit more carefully about the household choice between permanent and marginal labour effort. One possibility is to allow the household to choose both types of hours subject to a constraint that

$$n_t^p \leq \bar{n}^p + \bar{m}^p.$$

assume that $w_p > w_t$, which is in line with typical case in which more permanent staff with longer tenure have higher wages. In this case the constraint above always binds, and the representative household is constrained in its choice of permanent workers. In this case only the choice of marginal workers is unconstrained. A simpler but more ad hoc assumption is to force the constraint above to hold with equality for some exogenous unspecified reason. The work hours of permanent and marginal workers are perfect substitutes in the utility function. Defining

$$n_t^{s,m} = n_t^m + m_t^m,$$

we get the same first order conditions as for the problem without permanent workers as long as

$$n_t^s > \bar{n}^p + \bar{m}^p.$$

But this second inequality will hold in any competitive equilibrium in which the earlier inequality conditions on labour demand also hold.

Under these conditions, combining the household budget constraint and the firms' profits and equating supply and demand for marginal labour hours, we get the same system of equations characterising equilibrium in the model with long term contracting as in the model with only static flexible labour contracts.

Therefore we get the same solution for quantities. The relevant wage for hiring or firing decision is that of the marginal workers w_t. Average wages in the economy will typically be more rigid. Looking a a histogram of wage changes in such an economy, we will see a spike at 0 (here we have assumed real wage rigidity for permanent workers, but a similar argument could be made for nominal rigidity). One question that arises is what happens when firms labour demand declines strongly enough that they want to cut back on their permanent employees. In general, the outcome will no longer be the same as with static contracts. The one exception is if a sufficient number of permanent workers can be converted into marginal workers, in which case the constraints $n_t > \bar{n}_t^p$ and $m_t > \bar{m}_t^p$ never bind (where we now allow $\bar{n}_t^p, \bar{m}_t^p$ to occasionally adjust in response to economic conditions). Essentially, this amounts to assuming bargaining between workers and employers is sufficiently flexible so as to avoid any inefficient separations, as in Barro (1977) [20].[19]

5 Competitive Equilibrium

We use the

1) market clearing conditions,

2) budget constraints and

3) optimisation conditions (e.g first order conditions for differentiable problems or Value functions for non differentiable optimisation problems) to form

4) a system of equations defining the solution of the model.

This system of equations is called the competitive equilibrium.

5.1 State Space Representation of the Competitive Equilibrium

We partition the variables in the system into state variables and non-state variables. The state variable vector is sufficient to describe the dynamics of the system, in the sense that once we know the path of the states it is easy to predict the path of non state variables from that of the states.

There are several ways of defining the state vector. At the very least the states at time t must include all predetermined variables: i.e. variables whose forecast error at t is exogenous (Klein (2000) [130]). Typically this includes, exogenous stochastic shock processes z_t, capital stocks and the lags of non state

[19] The equivalence with our basic model will also fail if the two types of workers are imperfect substitutes in production on the utility function, but in that case we should still get equivalence with a neoclassical labour market model featuring the same substitutability patterns.

variables (k_t, y_{t-1}). Using just the predetermined variables as states gives us the Minimum State Vector (MSV) solution. We also have the fundamental source of uncertainty in the system, a vector of IID variables (often known as innovations or error terms) e_t.

Defining the state vector s_t and the non state vector y_t we seek a solution to the system of equations

$$\begin{aligned} E_t f(s_t, s_{t+1}, y_t, y_{t+1}, e_{t+1}) &= 0, \\ E e_{t+1} &= E_t e_{t+1} = 0. \end{aligned}$$

[20] **We are usually looking for a first order Markov solution**

$$\begin{aligned} s_{t+1} &= g(s_t, e_{t+1}), \\ y_t &= h(s_t). \end{aligned}$$

The equations above are called a state space system. Compared to typical systems of equations in econometrics or statistics, the key innovation in the state space equations is the recognition that the underlying factors that drive the system can be unobserved. In contrast, a subset of the non state variables y_t is usually observed. In applied work, it is common to allow for an extra residual or measurement error ε_t in the equation for the non state variables, leading to

$$y_t = h(s_t) + \varepsilon_t.$$

Once we have obtained the state space representation, we can use a combination of explicit formulas and simulation (applying the law large numbers) to forecast,do scenario analysis ,find probability bands or event probabilities for different variables,evaluate the ability of the model to match various features of the data or link historical events to their more fundamental unobserved causes (the states) through variance decompositions. Fernandez Villaverde (2010)[84] and the references he cites are an excellent introduction to solution methods and Bayesian inference for DSGE models. For generalisations that handle a large number of heterogenous agents (which are also applicable to the more standard models with a small number of agents as a special case) see the survey by Algan et al (2009) [5].

Note that the expression of the equilibrium system $E_t f(.)$ and the solution $g(.)$ and $h(.)$ using only t and $t+1$ is without loss of generality: we can always handle further lead and lags by expanding the state vector. For example, if we need y_{t-1} in our state vector, we can define the new variable $s_{-1,t} = y_{t-1}$. For y_{t-2}, we would define $s_{-2,t+1} = s_{-1}, t$ and so on for higher order lags.

[20] Inequality constraints can be transformed into equations by introducing Lagrange multipliers and forming the corresponding complementary slackness conditions.

Finally, note that in the case of exogeneous shock processes z_t the solution for the endogenous decision variables of agents in y_t can be done independently from the actual shock processes used in model simulation. This gives us some flexibility in modeling deviations from rational expectations due to misperceptions on the exogenous shocks. For example, we can solve for decision rules

$$y_t = h(k_t, z_t) + \varepsilon_{1,t},$$

assuming that agents perceive z_t as following

$$z_t = f^1(z_{t-1}) + \varepsilon_{1,t},$$

, while the true process that we use in simulating the model is

$$z_t = f^2(z_{t-1}) + \varepsilon_{2,t}.$$

This allows us to capture situations where agents may overestimate the persistence of shocks leading them to overreact (see for example Laibson et al 2011), or their decision rules may misperceive the risk that they face leading them to errors in the amount of precautionnary behaviour.We could also use this procedure to model situations where agents overestimate the importance of certain shocks, or more generally situations in which agents have misperceptions about the causes and economic mechanisms behind business cycles.

5.2 Filtering and Forecasting with the State Space Representation

From the state space representation above, it is easy to get $p(y_t|s_t)$ either explicitly or through simulation. The main econometric challenge is to estimate the probability distribution of the partially unobserved states s_t conditional on the the observed variables. Define $y^t = (y_{0,\ldots,}y_t)$. In typical macroeconomic forecasting applications, we are interested in the forecast distribution

$$p(s_{t+1}|y^t)$$

, the filtered state distribution

$$p(s_t|y^t)$$

and the smoothed state distribution

$$p(s_t|y^T).$$

In general we need to go through the following steps to forecast:

1) Use the previous forecast distribution $p(s_t|y^{t-1})$ to get the conditional likelihood function $p(y_t|y^{t-1})$,

$$p(y_t|y^{t-1}) = \int p(y_t|s_t)p(s_t|y^{t-1})ds_t.$$

2) Use the conditional likelihood function to get the filtered state distribution through

$$p(s_t|y^t) = \frac{p(y_t|s_t)p(s_t|y^{t-1})}{p(y_t|y^{t-1})}.$$

3) Use the filtered state distribution to get the forecast distribution through

$$p(s_{t+1}|y^t) = \int p(s_{t+1}|s_t)p(s_t|y^t)ds_t.$$

In the special case of a linear state-space model, these steps can be accomplished using the Kalman filter. In the general nonlinear case, there is a wide variety of options going from adaptations of the Kalman filter to take into account higher order terms in the state equations (e.g the unscented Kalman filter) to methods that use simulation to approximate the integrals above, such as the particle filter. The general idea of the particle filter is to use importance sampling:

Suppose we know how to draw from a distribution $g(x)$, and we want draws from a target distribution $\tau(x)$. Then for any desired function $h(x)$,

$$\int h(x)\tau(x)dx = \int h(x)\frac{\tau(x)}{g(x)}g(x)dx.$$

By the law of large numbers the sample average using n draws from $g(x)$,

$$\frac{1}{n}\Sigma_{j=1}^n h(x_j)\frac{\tau(x_j)}{g(x_j)} \rightarrow_{a.s} \int h(x)\frac{\tau(x)}{g(x)}g(x)dx = \int h(x)\tau(x)dx.$$

See Fernandez Villaverde et al (2010)) [84] for a more detailed survey.

5.3 Solving For the Competitive Equilibrium

In most cases we can find a unique state-space form solution. The most straightforward way to do this is by Taylor approximations of the functions $h(.)$ and $g(.)$. The 1st order approximation

$$\begin{aligned} AE_t(x_{t+1} - x) &= B(x_t - x), \\ x_t &= (s_t; y_t) \end{aligned}$$

involves taking a (log)linear approximation of the equations in $E_t f(.)$ around x (usually the long run balanced growth path of the economy)

and using matrix decompositions of A and B such as the QZ or the SVD decompositions to solve the linearised system of equations (e.g Klein (2000) [130]) for a set of equations

$$s_{t+1} - s = g_1(s_t - s) + e_{t+1}$$

$$y_t - y = h_1(s_t - s).$$

This linear approximation to the state space system is similar to another popular class of time series models, the dynamic factor model (DFM) (Stock and Watson (2010)[170]). The state variables are essentially like the factors in DFM's, except that they have a clearer economic interpretation than factors extracted through purely statistical methods. The parameter restrictions from DSGE models can be used to improve (conditional) forecasts from factor models (see for example Baurle (2008), Consolo et al (2009)) and to provide economic interpretations to the factors. In the other direction, the factors extracted in reduced form statistical factor models may be useful in improving the identification of wedge processes in business cycle accounting DSGE models.

A unique rational expectations solution for the linearised state space system above requires the number of Eigenvalues of (A, B) with an absolute value below 1 to equal the number of predetermined variables. Under this condition, the QZ decomposition algorithm partitions the initial system into 2 separate blocks: the 1st block of equations allows solving for the predetermined variables, while the 2nd one solves the non predetermined variables as a function of the predetermined variables that have been solved in the 1st block (Klein (2000)[130]).

To get a sense of the manipulations involved in one of the most important procedures in DSGE model analysis and forecasting, we will go through the solution method of the 1st order approximation as outlined in Gomme and Klein (2010) [96]:

1) Use the following QZ decomposition:

Given any square matrice (A, B), if the matrix $A - zB$ is invertible for some complex number z, then we can find matrices (Q, Z, T, S) such that

- $Q^T Q = QQ^T = I$, $Z^T Z = ZZ^T = I$,
- T and S are upper triangular (i.e with all entries below the main diagonal equal to zero),
- $QA = SZ^T, QB = TZ^T$,
- either $s_{ii} \neq 0$ or $t_{ii} \neq 0$.

2) Order the pairs (s_{ii}, t_{ii}) such that those with $|s_{ii}| > |t_{ii}|$ come first. These will be the stable Eigenvalues, representing variables that have a stationnary solution solving backwards over $t, .., t - j, ...$

3) Premultiply the equations by Q to get

$$\begin{array}{rcl} QAE_t(x_{t+1}-x) & = & QB(x_t-x). \\ & \Longleftrightarrow & \\ SZ^T E_t(x_{t+1}-x) & = & TZ^T(x_t-x) \\ & \Longleftrightarrow & \\ SE_t \left[\begin{array}{c} \tilde{s}_{t+1} \\ u_{t+1} \end{array} \right] & = & T \left[\begin{array}{c} \tilde{s}_t \\ u_t \end{array} \right], \\ \text{where } \left[\begin{array}{c} \tilde{s}_t \\ u_t \end{array} \right] & = & Z^T(x_t-x). \end{array}$$

$\tilde{s}_t$ corresponds to the stable variables, that have a stationnary solution solving backwards over time (e.g shock processes or endogenous state variables). u_t corresponds to the unstable variables, that have a stationnary solution solving forward over time.

4) Since S and T are upper triangular, we can partition the matrices into blocks such that

$$\left[\begin{array}{cc} S_{11} & S_{12} \\ 0 & S_{22} \end{array} \right] E_t \left[\begin{array}{c} \tilde{s}_{t+1} \\ u_{t+1} \end{array} \right] = \left[\begin{array}{cc} T_{11} & T_{12} \\ 0 & T_{22} \end{array} \right] \left[\begin{array}{c} \tilde{s}_t \\ u_t \end{array} \right].$$

The last set of equations in this system is

$$S_{22}u_{t+1} = T_{22}u_t$$

If the eigenvalues of (S_{22}, T_{22}) satisfy $|s_{ii}| \leq |t_{ii}|$, then, these equations can be solved forward for a stationnary solution $u_t = 0$.

5) Using $u_t = 0$,

$$S_{11}E_t\tilde{s}_{t+1} = T_{11}\tilde{s}_t$$

If the eigenvalues of (S_{11}, T_{11}) satisfy $|s_{ii}| > |t_{ii}|$, then we can write

$$E_t\tilde{s}_{t+1} = S_{11}^{-1}T_{11}\tilde{s}_t$$

.

6) Using $ZZ^T = I$, we have

$$x_t - x = \left[\begin{array}{c} s_t - s \\ y_t - y \end{array} \right] = Z \left[\begin{array}{c} \tilde{s}_t \\ u_t \end{array} \right]. = \left[\begin{array}{cc} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{array} \right] \left[\begin{array}{c} \tilde{s}_t \\ u_t \end{array} \right],$$

where Z has been partitioned such that Z_{11} and Z_{12} have the same number of rows as $s_t - s$. Since $u_t = 0$,

$$\begin{array}{rcl} s_t - s & = & Z_{11}\tilde{s}_t, \\ y_t - y & = & Z_{21}\tilde{s}_t. \end{array}$$

7) If Z_{11} is invertible, we can now solve for $\tilde{s}_t$ and for the non-predetermined variables in y_t,

$$\begin{aligned}
\tilde{s}_t &= Z_{11}^{-1}(s_t - s), \\
y_t - y &= Z_{21}Z_{11}^{-1}(s_t - s) = h_1(s_t - s).
\end{aligned}$$

8) Using our solution for $\tilde{s}_t$ we can solve solve for the state variables s_t,

$$\begin{aligned}
E_t Z_{11}^{-1}(s_{t+1} - s) &= S_{11}^{-1}T_{11}Z_{11}^{-1}(s_t - s) \\
&\Leftrightarrow \\
E_t(s_{t+1} - s) &= Z_{11}S_{11}^{-1}T_{11}Z_{11}^{-1}(s_t - s) \\
&\Rightarrow \\
s_{t+1} - s &= g_1(s_t - s) + e_{t+1}, \\
g_1 &= Z_{11}S_{11}^{-1}T_{11}Z_{11}^{-1}.
\end{aligned}$$

Higher order Taylor approximations of order K are of the form

$$\begin{aligned}
s_{t+1} - s &= \Sigma_{k=1}^{K}\frac{1}{k!}g_k\begin{bmatrix} s_t - s \\ \sigma \end{bmatrix}^{\otimes[k]} + e_{t+1}, \\
y_t - y &= \Sigma_{k=1}^{K}\frac{1}{k!}h_k\begin{bmatrix} s_t - s \\ \sigma \end{bmatrix}^{\otimes[k]}
\end{aligned}$$

, where $x^{\otimes[k]}$ is the kth order Kronecker product of x. They can be obtained with the following procedure:

1. Substitute the general form of the DSGE solution into the system of equilibrium conditions:

$$\mathbf{E}_t\mathbf{f}(\mathbf{s}_t, \mathbf{g}(\mathbf{s}_t, \mathbf{e}_{t+1}), \mathbf{h}(\mathbf{s}_t), \mathbf{h}(\mathbf{g}(\mathbf{s}_t, \mathbf{e}_{t+1})), \mathbf{e}_{t+1}, \boldsymbol{\sigma}) = \mathbf{0}.$$

$\sigma \in [0, 1]$ is a parameter controlling the level of uncertainty in the

system:$\sigma = 0$ gives the deterministic system with no uncertainty, while $\sigma = 1$ represents the original system with the actual standard deviations of e_t. We will define the perturbation in (s_t, σ).

2. Solve for the linear approximation using the previously outlined QZ decomposition algorithm. This provides the 1st coefficient matrices required for the solution, g_1 and h_1.

3. For $k = 2..K$,

Take the kth total derivatives of the equilibrium conditions from 1. and evaluate them at the non-stochastic steady state $(s_t, \sigma) = (s, 0)$:

$$\begin{aligned}
\frac{d^k}{ds^k}E_t f(s, g(s), h(s), h(g(s)), 0) &= 0, \\
\frac{d^k}{d\sigma^k}E_t f(s, g(s), h(s), h(g(s)), 0) &= 0
\end{aligned}$$

4. Given knowledge of $\{g_k, h_k\}_{j=1...k-1}$ from the previous steps, we can find g_k, h_k by solving a linear system of equations.

Going over step 4 recursively for each k, we find the set of all solution matrices

$$\{g_k, h_k\}_{k=1...K}.$$

(see for example Judd (1998), ch. 14 [119] and Schmitt-Grohe and Uribe (2004) [102]). Lan and Meyer-Gohde (2012) [136] prove that the solution $\{g_k, h_k\}_{k=1...K}$ always exists and is unique. Modern software packages for solving and simulating macroeconomic models (e.g dynare) have made these calculations almost routine. With current computing resources, analysts usually restrict the approximation order $K \leq 3$, with 1st and 2nd order approximations being more common. 2nd and 3rd order Taylor approximations capture some nonlinearity, though the quality of approximation can deteriorate away from the long run deterministic growth path of the economy.

More severe nonlinearity or assymetry in dynamics requires the use of more accurate but slower global approximation methods (Judd (1998), ch. 11 [119]). Alternatively they can be captured to some degree by adjusting the approximation point x. For example, higher de facto financing frictions for a large negative shock can be captured by using an approximation around x with heavier steay-state financial distortions for the IRF.

We can also increase the accuracy of perturbation methods by using exact equations such as identities to solve out for some variables whenever it is possible. If we have m equations of the form $f(y_{1t}, y_{2,t}, s_t) = 0$, then we can use the perturbation solutions of s_t and $y_{2,t}$ to solve out precisely for $y_{1,t}$ (see Maliar and Maliar 2011 for further development of this hybrid perturbation approach).

In some cases a unique solution in the form above does not exist . In this case we need to add new shocks to the system to account for this indeterminacy. In addition, the functional form linking these "sunspot" or nonfundamental shocks v_t to the other variables is itself indeterminate. To implement this, we redeclare some of our jump variables to be predetermined and add new exogenous forecast errors to them (see Farmer and Khramov (2013)[81] for a formal analysis and examples of how to implement this in standard software packages such as Dynare). There are several economic criteria for resolving this indeterminacy (e.g Benhabib and Wen (2004) [27]). **These lead to a solution of the form**

$$\begin{aligned} s_{t+1} &= g(s_t, e_{t+1}, v_{t+1}), \\ y_t &= h(s_t, v_t). \end{aligned}$$

6 Business cycle dynamics in the flexible prices and wages economy with a fixed capital stock

In the baseline model introduced in this section life is actually much simpler. The equilibrium conditions are

$$\begin{aligned} b_{t+1} &= 0, \\ n_t^s &= n_t + m_t, \\ y_t &= c_t. \end{aligned}$$

We solve the model in several steps:
1. Solve the firm's first order conditions for m for $m(n, w, ...)$.
2. Use 1. and the first order conditions for n to solve for $n(w, ...), m(w, ...)$. The result of 2. can usually be written as $w(n)$.
3. Combining the above with the first order condition for n^s with the market equilibrium conditions we get a static equation for n_t which is independent of efp_{ht} or $efp_{h,t+1}$.

After solving the model, we are now ready to explore the response of this simple economy to various shocks.

6.1 Intertemporal shocks

A fundamentals based solution requires a unique solution in each of these steps. In that case, intertemporal wedges such as $efp_{h,t}$, or β_t do not affect output if they are uncorrelated with intratemporal wedges. like $\tau_{c,t}$ or $\tau_{nd,t}$. They only affect interest rates. A more indirect proof of this result is also instructive: if the intertemporal and intratemporal wedges are uncorrelated, a shift in the intertemporal wedge that reduces desired consumption and increases desired saving does not affect labour demand for a given wage. For consumption c_t and output $y_t = c_t$ to fall in response to this negative demand shock requires labour supply to fall for any given wage. This causes wages to rise, decreasing labour demand. But with the standard income effects on labour supply, a decline in consumption encourages an increase in labour supply at any given wage and therefore higher output. Since $y_t = c_t$ this means consumption has increased, leading to a contradiction. Instead, consumption does not change and the higher desired savings lead to an decline in interest rates to clear credit markets.

Consider the implications for financial shocks, e.g a rise in $efp_{h,t+1}$. The above result says that to have an effect on output, a financial shock must have some immediate effect on the efficiency of utilisation rate of productive resources z_t or $e_{m,t}$, or on intratemporal household wedges $\tau_{c,t}, \tau_{ns,t}$. Note that these results do not rely on perfectly competitive markets, since $\tau_{nd,t}$ and z_t can be in part related to imperfect competition. With imperfect competition, these intertemporal household level shocks cannot generate business cycles as long firms' pricing markups are exogenous. Without sunspot equilibria, only fundamental intratemporal shocks such as z_t can move output in this economy. [21]

6.2 Total factor productivity (TFP) shocks

Suppose there is a positive TFP shock in the current period that is orthogonal to the other wedges.

1) The increase in TFP causes higher output both directly as productivity increases but also indirectly by raising labour demand, since the marginal product of labour is higher.

2) To attract more workers, wages have to increase. The wage increase has two effects. First there is a substitution effect: a higher return to working encourages higher labour supply. Second, there is an income effect. Since leisure is a normal good, higher income encourages leisure and discourages labour supply.

3) With preference that are compatible with a long run balanced growth path these two effects offset each other perfectly in the economy with fixed capital stock. As a result, output and consumption only increase due to the direct effect of higher productivity. Unless, the TFP shock also leads to a shift in the labour wedge, equilibrium labour supply does not change.

$$
\begin{aligned}
\Delta z_{t+j} &> 0,\ \Delta g_{t+j} \Rightarrow \Delta n^d_{t+j} > 0 \Rightarrow \Delta w_{t+j} > 0 \\
&\Rightarrow \Delta n^s_{t+j} = 0 \Rightarrow \Delta n_{t+j} = 0 \\
&\Rightarrow \Delta c_{t+j} > 0, \Delta y_{t+j} > 0.
\end{aligned}
$$

This is easiest to see with a Cobb Douglas production function

$$y_t = z_t n_t^{\alpha}$$

and a constant elasticity of substitution utility function

$$u(c,l) = \frac{\left(c^{\theta} l^{1-\theta}\right)^{1-\sigma}}{1-\sigma}.$$

[21] With multiple solutions, output can shift due to sunspot fluctuations that look. One has to be careful though with sunspot models not to generate other counterfactual predictions such as negative responses of aggregate labour demand to cuts in payroll taxes. The controversial responses of these models to shocks occur because they frequently require an aggregate labour demand curve that is upward sloping and steeper than aggregate labour supply (see e.g. the aggregate externality RBC model in Benhabib and Wen (2004) [27] and Aiyagari's (1995) [2] critique of these models' implications for other shocks and tax policies).

, where for simplicity we set $g_t = 1$. The household's labour supply optimality condition is

$$\frac{1-\theta}{\theta}\frac{1}{l_t} = \frac{w_t(1-\tau_{ns,t})}{c_t(1+\tau_{c,t})}.$$

Labour demand satisfies

$$\begin{aligned} \alpha\frac{y_t}{n_t} &= w_t(1+\tau_{nd,t}) \\ &\Longrightarrow \quad y_t = \frac{w_t(1+\tau_{nd,t})n_t}{\alpha} \end{aligned}$$

, which using $y_t = c_t$ implies

$$\begin{aligned} \frac{1-\theta}{\theta}\frac{1}{1-n_t} &= \alpha\frac{w_t(1-\tau_{ns,t})}{w_t(1+\tau_{nd,t})(1+\tau_{c,t})n_t}, \text{ i.e.} \\ \frac{1-\theta}{\theta}\frac{n_t}{1-n_t} &= \alpha\frac{(1-\tau_{ns,t})}{(1+\tau_{nd,t})(1+\tau_{c,t})}. \end{aligned}$$

, so that n_t is independent of w_t and therefore of z_t shocks after conditioning on $\tau_{c,t}$ and $\tau_{ns,t}$. Note that equilibrium labour supply is decreasing in $\tau_{ns,t}$ and in $\tau_{nd,t}$, so a TFP shock could reduce labour supply if for example it increases the labour supply wedge $\tau_{ns,t}$. Allowing g_t to depend on m_t, we can easily derive the same result for example with

$$y_t = z_t m_t^{\gamma} n_t^{\alpha}.$$

4) Meanwhile, the interest rate follows the pattern of the change in productivity closely. Assume first that productivity is higher in the current period than in the next period, that is the TFP shock is front loaded. Then consumption smoothing households would like to reallocate resources across periods through saving. Since this is not possible in aggregate with $b_{t+1} = 0$, interest rates fall to clear credit markets.

Now assume that productivity is higher next period than now. Then households would like to borrow funds to start consuming more today. Since this is again impossible in aggregate, interest rates increase to clear credit markets (Goodfriend (2004) [98]). The countercyclicality of interest rates for front loaded TFP shocks is reduced in the presence of credit constrained households.

- For credit constrained households the improvement in the economy leads to a lower external finance premium. In this case, a reduction in $efp_{h,t+1}$ tends to push up interest rates, going in the opposite direction to the effect of consumption smoothing by non credit constrained households.

6.3 Labour wedge shocks

An increase in the labour demand wedge $\tau_{nd,t}$ could be caused by higher financing costs for wages (e.g Jermann and Quadrini (2010), Arellano et al. (2012)), higher recruitment costs per worker (Pissarides (2008) [157]), higher markups due to lower competition or an insufficient response of prices to declines in demand (See Bils et al (2012) for a state of the art empirical analysis of countercyclical markups and their interpretation, and the section on mapping price rigidity into markups in this document), higher uncertainty about the returns to hiring , greater pessimism by employers about the returns to hiring, or higher search frictions in labour markets due to the need to reallocate workers between sectors.

1) The rise in $\tau_{nd,t}$ decreases labour demand and wages.

2) Lower wages reduce labour supply through a substitution effect, but at the same time they increase labour supply through a negative income effect. However, in contrast to the TFP shock, the income effect of the lower wage is reduced by the increase in transfers T_t. As a result, the substitution effect dominates and equilibrium labour supply declines.

3) Since TFP is fixed by assumption, output in the economy also declines.

$$\begin{aligned} \Delta\tau_{nd,t+j} &> 0 \Rightarrow \Delta n^d_{t+j} < 0 \Rightarrow \Delta w_{t+j} < 0 \\ &\Rightarrow \Delta n^s_{t+j} < 0 \Rightarrow \Delta n_{t+j} < 0 \\ &\Rightarrow \Delta c_{t+j} < 0,\ \Delta y_{t+j} < 0. \end{aligned}$$

Note that the results would be the same if instead of writing out the labour demand wedge as a tax on firms, we specified it as a binding quantity constraint on labour demand. For example consider the constraint

$$w_t n_{d,t} \leq \bar{a}_t,\ (\mu_{nd,t}).$$

Now the labour demand optimality condition is

$$\alpha\frac{y_t}{n_t} = w_t(1+\mu_{nd,t}).$$

Writing $\mu_{nd,t} = \tau_{nd,t}$, we get the exact same conclusions as before. The alternative intepretation clarifies why exactly the substitution effect dominates the income effect on labour supply: the key factor is that the higher perceived cost of hiring workers increases the share of non

labour income in GDP. This could occur from a mixture of higher profits share in GDP, or if unemployment benefits spending increases proportionately to the higher labour demand wedge.A higher labour supply wedge $\tau_{n,st}$ would have a similar effect for quantities, but with wages increasing. [22]

$\tau_{nd,t}$ **and** $\tau_{ns,t}$ **are frequently combined together to define an overall index of labour market distortions. Using households' first order condition for labour supply with firms' first order condition for labour demand we get**

$$\frac{g_t z_t f_{n,t}}{(u_{l,t}/\lambda_t)} = \frac{1+\tau_{nd,t}}{1-\tau_{ns,t}}.$$

$\frac{1+\tau_{nd,t}}{1-\tau_{ns,t}}$ **is often known as the labour wedge. It captures the deviation of employment from the benchmark efficient outcome of the undistorted neoclassical growth model, with a higher wedge indicating bigger deviations. In general the labour wedge is countercyclical, indicating significantly higher procyclical volatility in hours of work compared to a benchmark labour market model without distortions.** Chari et al (2007) [50] document an important role for the labour wedge in business cycles. Mulligan (2010) shows that a real business cycle model with an increase in the labour wedge can account for much of the US great recession of 2008-2010, when large output declines were accompanied by acyclical or rising average labour productivity. Because of the procyclicality of wages (especially after adjustment for composition bias and long term labour contracts), we start by examining factors that could explain a procyclical labour demand wedge.

6.3.1 Some models of business cycle fluctuations in the labour demand wedge

This section presents some models of a higher labour wedge in recessions, due to precautionnary cutbacks in labour demand caused by the combination of a significant increase in idiosyncratic business risk, incomplete financial markets and labour adjustment costs.

Arellano et al (2012)[12] build a model where a procyclical labour wedge emerges from the interaction between hiring decision and the costs of default on the firm's debt:

1) Firms must make their employment decisions before fully knowing their productivity or the demand for their products during the period. That is, in contrast to the usual assumption that labour demand can be easily adjusted as a function of the firm's current productivity, here labour demand is less flexible.More generally there are some labour adjustment costs making labour demand depend on expected future conditions as in Bloom et al (2012) [34].

[22] The labour supply wedge is isomorphic to a preference shock to the disutility of work.

2) Because of the mismatch between the receipt of revenues and spending on wages and other production inputs, firms require short term loans (either from financial institutions, financial markets or from other firms in the form of trade credit).

3) Financial markets are incomplete: firm loan repayments cannot be made fully state contingent and default has real costs, unlike in the complete markets model. A higher wage bill increases the probability of the firm not having sufficient funds to fully repay its debt, forcing it to default on its loans. Default with incomplete markets leads to the liquidation of the firm: the firm disappears and its previous owners can only restart production by paying a reentry cost. More generally, there are significant costs of default.

4) An increase in the probability of default for a given labour input (e.g due to higher idiosyncratic risk) leads to a precautionnary reduction in labour demand as the firm tries to reduce the probability and costs of default.

Arellano et al (2012)[12] show that in a reasonably calibrated model this mechanism can match the declines in output and employment in the US during the great recession of 2008-2010 quite well while replicating the initial increase in labour productivity. Arellano et al (2012) [12] assume the firms' owners are fully diversified with respect to idiosyncratic risk. The precautionnary reduction in labour demand due to higher uncertainty is likely to be even stronger if firms are owned by undiversified entrepreneurs or large undiverisifed shareholders as in Angeletos and Calvet (2005) [9], or Angeletos (2007) [8]. With labour adjustment costs or partial irreversibility labour demand choices become like an investment. In that case the increased riskiness of labour encourages risk-averse entrepreneurs to reduce hiring and shift towards safer investments such as risk free bonds.

These labour demand effects are unlikely to be significant for most large US corporation that have positive net financial asset balances. But the distribution of corporate net financial assets is highly unequal. While in recent years the average corporate sector net financial assets balance is positive, the median balance is negative, with 57% of corporations having negative balances. Furthermore the non corporate firm sector has a negative net financial asset balance (see BEA tables, as well as Armenter and Hnatkovska (2012)). Computing the current ratio of US corporations as the ratio of liquid assets to short term debt, we see that even in aggregate short term debt is significantly larger than liquid assets with a current ratio around 35% over the last 30 years. It jumped to over 50% in the early part of the great recession (2008-2010), but that was in large part because firms cut down their expenditures and accumulated cash in reaction to the financial crisis. [23] These facts suggest that the relation between financial distress fears and hiring may be quite relevant at least for

[23] http://www.marketminder.com/c/fisher-investments-how-strong-is-corporate-americas-balance-sheet/b14e50c8-8ae3-42cb-b046-ee62a28051bb
http://conversableeconomist.blogspot.com/2013/07/us-firms-holding-18-trillion-in-liquid.html

small and medium size firms. A lot of the corporate cash balances are held for precautionary motives in the fear of future binding financial constraints. From this perspective, evidence of rising corporate net financial assets in the 2008 financial crisis are perfectly compatible with Arellano et al's (2012)[12] precautionnary fall in labour demand story: firms in part built up their cash reserves by reducing hiring.

Nonconvex labour adjustment costs can also generate similar cutbacks in employment in a recession. Bloom et al (2012) [34] build a model in which hiring and firing decisions are subject to fixed labour adjustment costs, and proportional hiring and firing costs. These nonconvex adjustment costs generate inaction bands, where firms do not adjust their labour demand if the change in their desired workforce is below certain thresholds. An increase in idiosyncratic uncertainty increases the firm's inaction bands, leading to less hiring of new workers. The higher uncertainty is modelled as a 2nd moment shock that also increases the likelihood of positive idiosyncratic shocks. This leads to a reduction in firing of workers as well, but the decline in endogenous firing is dominated by the much larger exogenous separation rate of workers (e.g due to natural attrition, or idiosyncratic changes in the efficiency of certain production processes or product lines inside the firm). Consequently, the decline in hiring dominates and labour demand falls.

This result is reinforced if we take into acount the growing evidence that the increase in idiosyncratic risk in recessions is more skewed towards negative outcomes (Guvenen et al (2012)). In this case, an increase in uncertainty would depress labour demand even more, this time also significantly increasing firing as well as reducing hiring. Again, Bloom et al (2012)[34] assume that firm owners are fully diversified with respect to idiosyncratic production risk. But similarly to the case of costly default in Arellano et al (2012) [12] we can imagine an extension of their model in which firms are owned by entrepreneurs as in Angeletos (2007) [8]. More generally firm owners or managers are not well diversified with respect to idiosyncratic risk. In this case, higher idiosyncratic uncertainty in the presence of labour adjustment costs will make having a large workforce more risky for the owners, amplifying the reduction in labour demand.

Arellano et al (2012)[12] and Bloom et al (2012) [34] generate a recession from an increase in idiosyncratic risk. An interesting question is where does the countercyclicality of idiosyncratic business risk come from. One interpretation of countercyclical risk is that there really is an increase in the dispersion of productivity or demand facing different firms. Another interpretation is that the volatility of idiosyncratic shocks in their model comes from the residual risk left after the purchase of financial insurance contracts or the short-selling of financial contracts that are positively correlated with the shocks to the firm. This intuition is formalised in Angeletos and Calvet's (2005) [9] model of idiosyncratic production risk, in which entrepreneurs can diversify their firm-specific risk by trading other financial assets. The optimal portfolio of entrepreneurs in their model minimise the variance of entrepreneur wealth for a given expected return. By the projection theorem, after purchase of other financial assets at $t-1$ the original idiosyncratic risk $z_{j,t}$ can be decomposed into

$$z_{j,t} = \bar{z}_{j,t} + \varepsilon_{j,t}$$

,where $\varepsilon_{j,t}$ is uncorrelated with $\bar{z}_{j,t}$. $\bar{z}_{j,t}$ is predetermined with respect to t and $\varepsilon_{j,t}$ is the residual undiversified risk. Higher business risk can be linked to a decline in the availability of financial contracts that provide insurance or to stricter limits on shortselling, e.g due to lower trading of various derivatives and more sophisticated loans, or due to tighter credit line limits. A reduction in the amount of insurance provided by the financial system translates into higher residual risk $\varepsilon_{j,t}$ for firms. More generally, we can envision other mechanisms raising the shadow cost of borrowing to finance production inputs. For example, collateral constraints on short term debt become tighter, more restrictive debt covenant clauses are triggered etc...

Finally, the mechanisms we have discussed can also apply to purchases of other production factors such as marketing and distribution expenses, or intermediate inputs that represent approximately 50% of the value of production. In that case higher idiosyncratic uncertainty would discourage purchase of intermediate inputs as long as some of those purchases are predetermined with respect to idiosyncratic shocks. From the perspective of a value added production function a decline in the use of intermediate inputs or of marketing and distribution would map into a TFP reduction. We will explore this mechanism in greater detail in separate section on the business cycle effects of the input-output structure of the economy.

6.3.2 Aggregation/composition effects on the labour wedge, and factors increasing the sensitivity of labour supply to shocks

Here, I look at some explanations for the labour wedge coming from the labour supply side. These labour supply channels help reconcile large movements in employment with low cyclicality of wages. I focus on factors that are independent of institutional restrictions preventing wage adjustments, i.e factors that matter even if wages are completely flexible. I will examine the consequences of wage rigidity due to contractual limitations once we get to Keynesian economics.

A higher labour wedge in a recession could reflect composition effects in labour demand and supply. What matters to the representative firm in a competitive labour market is the number of efficiency units, that is the number of hours adjusted for differences in individual worker productivity. Labour demand shifts towards more qualified and skilled workers in a downturn. The standard model of labour demand assumes workers of different qualities are perfect substitutes in production, therefore it cannot explain the tendency to fire less productive workers first. But a more realistic model in which the more productive workers are better complements with the production of other workers could easily explain why a firm keeps its most productive workers in a recession, and fires the least productive first. The change in the composition of the workforce shows up as a higher labour demand wedge. In parallel, low productivity workers are less likely to get job offers above their reservation wages in a downturn.

The combination of incomplete financial markets and labour indivisibility can also generate a labour supply wedge. In a model with incomplete markets as in Chang et al. (2011) [49], the lower wage reduces labour supply by low productivity workers via a substitution effect. But conditional on deciding not to work, the lower wage does not produce an offsetting negative income effect that would boost labour supply. So with indivisible labour and incomplete markets, the labour supply of these households declines. From the perspective of the representative household, this looks like a shock increasing the disutility of working in a recession, or a higher labour supply tax. [24]

If the labour wedge is measured using aggregate hours without adjusting for differences in worker specific productivity, then we overestimate the decline in effective labour demand in a (adjusted for differences in worker efficiency) in a downturn. Kydland (in Cooley et al (1995) [62] ch. 5) argues using micro panel data evidence that incorrectly using total hours without adjusting for differences in efficiency across workers overstates the variability of labour demand by around 40%. For a given estimate of the TFP process, this leads to a significant overestimate of labour wedge fluctuations. To see how mismeasurement of the labour input leads to the appearance of labour wedges, we can extend our model to include differences in worker efficiency per hour. Let h_{jt} be the physical actual hours of work chosen by household j. Total hours in the economy are

$$h_t = \int h_{j,t} dj.$$

Meanwhile the actual labour input relevant for firms is

$$n_t = \int \varepsilon_{j,t} h_{j,t} dj$$

[24] This highlights that some divisibility in labour supply is key for the income effect to matter: for example the income effect would be operational if the worker has a choice between part time versus full time jobs or if , as argued in Mulligan and Rubinstein (2004) ,the household can adjust the number of its members looking for work. Other sources of divisibility are the number of shifts and the effort per hour of work. Finally, if we have indivisible labour but complete markets, the lower wages in a recession boost labour supply via a negative income effect for all workers, because the lower wages translate into lower intra-household insurance payments to low productivity workers (see Chang et al (2007), Chang et al (2011)[49] for more on these points).

, where $\varepsilon_{j,t}$ is the household specific labour productivity, and $E\varepsilon_{j,t} = 1$. This measure of n_t may not be the same measure of labour effort that's relevant to the representative household. However, Maliar and Maliar (2003) [142] show that in some cases such as Cobb douglas preferences and complete financial markets, the representative household utility function can be written as $u(c_t, 1-n_t)$ with $n_t = \int \varepsilon_{j,t} h_{j,t}$. We will focus on this special case, while noting that the result will still be valid as long as the representative household's true labour effort is less procyclical than actual hours worked.

We have argued that in general n_t will be less procyclical than h_t due to a higher employment rate of less productive workers in booms than in recessions. Now suppose an economist uses h_t instead of n_t as the measure of the labour input. One effect of this measurement error is that for a given level of output, the economist will underestimate the procyclicality of TFP ($z^h > z$ in a recession, and vice versa in a boom). This may lead to a significant underestimate of the role of TFP shocks in the business cycle. We now show that the analyst using mismeasured labour input data will also overestimate the procyclicality of the labour wedge. The true labour demand equation can be written as

$$\begin{aligned} zf_n &= z^h f_h \frac{zf_n}{z^h f_h} = w, \Longleftrightarrow \\ z^h f_h &= \frac{z^h f_h}{zf_n} w = (1+\tau_{nd})w. \end{aligned}$$

In a recession a bigger fall in h than in n implies an increase in

$$\frac{z^h f_h}{zf_n} = 1 + \tau_{nd}$$

by the concavity of the production function. Therefore, using the wrong measure of the labour input the researcher finds a higher labour demand wedge in a recession, and vice versa in a boom. For labour supply

$$\begin{aligned} u_l &= u_{1-n} = u_{1-h} \frac{u_{1-n}}{u_{1-h}} = w\lambda, \Longleftrightarrow \\ u_{1-h} &= \frac{u_{1-h}}{u_{1-n}} w\lambda = (1-\tau_{ns})w\lambda. \end{aligned}$$

In a recession, leisure measured as $1-h$ increases by more than $1-n$. By the concavity of utility in leisure,

$$\frac{u_{1-h}}{u_{1-n}} = 1 - \tau_{ns}$$

declines. As a result, the wrong measure of the labour input leads the researcher to conclude that the labour supply wedge has worsened. Chang and Kim (2007) find this bias is important in an incomplete markets model with indivisible labour (see also Chang and Schorfheide (2011) for a more detailed analysis of the results of estimating and forecasting with representative agent economies when the true economy has non trivial heterogeneity due to incomplete financial markets). [25] From this perspective actual labour input fluctuations are significantly less volatile when seen from the employers' perspective. But, the welfare costs of these fluctuations is bigger than suggested by the utility function of a representative household, because it falls more heavily on the number of jobs, particularly those of lower skilled, younger and poorer workers.

Another way to increases the procyclicality of employments while reducing that of wages is to make the marginal utility of wealth λ_t at which households evaluate the benefits of raising labour supply more procyclical. One realistic feature that helps in this direction is the incorporation of durable goods investment, especially if durable and non durable consumption goods are complements. In that case, if we measure the labour wedge using a λ_t that ignores these effects, we will tend to exaggerate labour wedge fluctuations (see the section on durable goods for more discussion of this). Also, if there are increases in the costs of consumer financing (both higher interest rates or higher credit rationing), these will often map into a higher consumption wedge $\tau_{c,t}$. An increase in $\tau_{c,t}$ will reduce the purchasing power of any given wage, making work less desirable and reducing labour supply through a substitution effect.

Note that while the higher consumption wedge $\tau_{c,t}$ also has a negative income effect which tends to stimulate labour supply, the substitution effect dominates in this case because the rise in the wedge is compensated through higher transfers T_t. This is realistic if the rise in the consumption wedge represents frictions that do not have a direct resource cost, such as tighter quantity constraints on consumer credit, or increases in credit spreads on consumer credit that increase the compensation of lenders for any given loan size.

Wen (2006)[180] shows that if the shock to the consumption wedge $\tau_{c,t}$ is persistent enough, it can generate a business cycle with procyclical consumption, investment and output. [26] But if this is the single source of business cycles, real wages would be countercyclical, since the lower labour supply in a recession would reduce wages. So $\tau_{c,t}$ shocks are best seen as complements to TFP or labour demand shocks that can produe business cycles with procyclical wages. The consumption wedge shocks helps in reducing the procyclicality of wages, making them appear more "rigid". Models ignoring the consumption wedge will once again tend to overestimate the volatility of the labour wedge. Finally certain models of search frictions or intersectoral labour adjustment costs can microfound procyclical movements in the labour supply wedge $\tau_{ns,t}$ (e.g Beaudry

[25] A related point is that wage flexibility is underestimated because of composition bias, with the pool of workers shifting towards more qualified and highly paid employees in a recession and vice versa in a boom (see Romer (2001) [161] ch. 5) .

[26] Wen (2006) examines preference shocks to consumption that shift the marginal rate of substitution between consumption and leisure. But these shocks are isomorphic to our $\tau_{c,t}$ shock.

and Portier (2013) [24]). The main idea in these models is that in a recession it becomes harder for unemployed workers to match with good job opportunities at any given wage if macroeconomic adjustments require a shift to another sector of production and it is costly for workers to retrain for the new sectors. As a result, labour supply declines and workers become more expensive.

7 Aggregate demand in real business cycle models and matching frictions in product markets

While real business cycle theory shows that supply side factors can be an important driver of aggregate economic fluctuations, popular and business press analysis of the economy still seems to think in terms of aggregate demand as the main factor in business cycles. There is a sense in which in a flexible price model, any recession in which wages and other production factor prices go down in real terms must be caused by negative shifts in the supply of goods. Low relative production factor prices are equivalent to high relative final goods prices. In this case, simple supply and demand logic in markets for final goods would argue that there has been a negative supply shock. This suggests that explaining business cycles through demand shocks could be a challenge for market clearing models, unless we can find channels transforming demand shocks into supply shocks.

Consider an economy with fixed capital and labour where the only endogenous variation in output comes from changes in utilisation rates of production capacity (due to e.g. time varying search frictions). The results would be similar if m_j was labour and p_m was a wage, as long as 1) the aggregate labour demand curve $m_j(w,z)$ is downward sloping, 2) labour supply is upward sloping and 3) labour supply is independent of income or subject to standard income effects. [27] The output of firm j is

$$\begin{aligned} y_j &= z_j u_j, \\ u_j &= f(m_j, y, e_m), \\ f_m &> 0,\ f_y > 0. \end{aligned}$$

m_j is the effort dedicated to marketing,which costs $m_j p_m$, where for simplicity p_m is exogenous. $f_y > 0$ captures the aggregate demand externality. Each firm solves

$$\max_{m_j} y_j - p_m m_j.$$

, which gives a first order condition,

$$z_j u_{m_j}(...) = p_m.$$

[27] If aggregate labour demand is upward sloping the discussion in the previous section applies.

In the simplest concept of general equilibrium, distribution effects across firms don't matter. This allows us to look at a symmetric equilibrium, where

$$y = zu(m, y).$$

Solving this equation for

$$y = f(m, z)$$

and plugging the result into the first order condition, we get

$$zu_m(m, z, f(m, z), e_m) = p_m.$$

A unique solution of this equation would give $m^* = g(z, p_m, e_m)$ **independently of** y **: search and marketing effort is independent of the level of economic activity once we have taken into account the effect of other fundamental shocks. In a sense this result is obvious:** y **is endogenous, so if there exists a unique solution for it, that solution must be in terms of the exogenous fundamentals** z, p_m, e_m. **This does not mean that aggregate demand channels captured in** $u(.)$ **are irrelevant. The aggregate demand externality in** $u(.)$ **can amplify the economy's response to fundamental shocks to** z **for example, even though it has no direct effect.** This is similar to flexible price economies with countercyclical markups and unique equilibria such as Jaimovich et al (2008) [114], or to economies with financial accelerator effects amplifying productivity shocks such as Bernanke et al (1999) [28].

One way to get direct aggregate demand effects on actual output in this setup is to find multiple solutions to either

$$\begin{aligned} y &= zu(m, y, e_m), \text{ or} \\ zu_m(m, z, f(m, z), e_m) &= p_m. \end{aligned}$$

With multiple solutions we get sunspot shock equilibria which could be associated with the popular notion of aggregate demand. Otherwise, within a fundamentals based model with unique equilibria, to get aggregate demand shocks to directly cause recessions we must find a more direct relation between e.g. consumer credit conditions or consumer uncertainty/precautionary saving and the efficiency of the search and matching process that determines $u(m, y, ...)$, **linking** $e_{m,t}$ **with** $efp_{h,t+1}$ **or** $\tau_{c,t}$ **or** β_t. For example this link can occur if consumer credit availability affects shopping effort beyond its indirect effect on desired consumption spending. (see the discussion in Rios Rull et al (2012) [17]).

To formalise the role of consumer search, assume that to purchase c_t **units of the consumption good requires the representative household to expand shopping effort** S_t^c, **satisfying**

$$c_t \leq \phi_t^c S_t^c, \ (\mu_t^S).$$

The new consumption goods finding rate ϕ_t^c is taken as given by households when choosing other variables. There is a matching function

$$c_t \leq f^M(S_t^c, m_t),$$

which is increasing in both shopping effort and firms' sales effoct m_t, and has constant returns to scale. This allows us to write $\phi_t^c = \frac{c_t}{S_t^c} = f(\frac{m_t}{S_t^c})$.There are several common market mechanisms for determining ϕ_t^c. In the directed search framework (e.g Rios Rull et al (2012) [17]) sellers post prices, and buyers direct their shopping effort to the best combination of prices and probabilities of finding the desired goods ϕ_t^c. This focuses the analysis on inefficiencies in other wedges (e.g the labour wedge), since with directed search the matching process is constrained Pareto optimal. A common alternative in labour market model is to assume Nash bargaining in which the surplus from the trade is split using a constant bargaining weight, but the bargaining assumption is less appealing for goods markets where prices are usually posted. Alternatively, we can follow Farmer (2011) [80] and use self fulfilling expectations about ϕ_t^c to determine the equilibrium. Shopping requires an effort $v(S_t^c)$, where for lack of clear evidence we assume the marginal cost of shopping is independent of consumption and work hours. **The household's utility function is**

$$\begin{aligned} & u(c_t, l_t) - v(S_t^c), \\ v_{S^c} &> 0, v_{S^c S^c} \geq 0. \end{aligned}$$

The first order conditions are modified to

$$\begin{aligned} c_t &: \lambda_t(1 + \tau_{c,t}) = u_{c,t} - \mu_t^c \\ S_t^c &: v_{S_t^c} = \phi_t^c \mu_t^c \\ n_t^s &: u_{l,t} = \lambda_t w_t (1 - \tau_{ns,t}) \\ b_{t+1} &: \lambda_t = E_t \frac{\beta_{t+1}}{\beta_t} efp_{h,t+1} R_{t+1} \lambda_{t+1}. \end{aligned}$$

The shopping effort constraint (μ_t^c) is always binding. Therefore we can combine first order conditions to get

$$\lambda_t(1 + \tau_{c,t}) = u_{c,t} - \frac{v_{S_t^c}}{\phi_t^c}.$$

Note that we can rewrite this first order condition as

$$\begin{aligned} \lambda_t(1 + \hat{\tau}_{c,t}) &= u_{c,t}, \text{ where} \\ \hat{\tau}_{c,t} &= \tau_{c,t} + \frac{v_{S_t^c}}{\phi_t^c \lambda_t}. \end{aligned}$$

This version of the first order condition highlights that search frictions here act very similarly to another component of the consumption wedge. Consider an increase in $\tau_{c,t}$ or in $\frac{\beta_{t+1}}{\beta_t} efp_{h,t+1}$. Ceteris paribus this increases $\lambda_t(1+\tau_{c,t})$ and therefore reduces $v_{S_t^c}$. Since $v_{S^cS^c} > 0$, shopping effort declines. As a result, the number of matches in product markets declines and firms find new buyers at a lower rate (g_t declines).

I will sketch informally a theory of how intertemporal consumption wedges and the efficiency of matching in product markets may be linked beyond any impact through desired spending y. Suppose consumers must decide their shopping effort early during the period before they fully understand the costs of financing or the level of employment risk they will face during the period by the time they can actually make a purchase. More formally, imagine a model with search frictions in product markets and two subperiods inside each period t: to find a good match on the product markets, shopping must start in the 1st subperiod before consumers fully observe the state of the economy during the period).

A credit shock, besides affecting intertemporal consumption wedges, can raise the dispersion of financing costs across consumers and firms. This increases uncertainty about desired consumption spending for consumers both directly and by raising the dispersion in the ability of firms to pay wages and maintain their current workforce. Both these channels therefore lead to higher income risk and higher uncertainty about the desirability of a high shopping effort for consumers. This higher uncertainty can reduce shopping effort independently of the direct effect of changes of the realised y. But the decline in shopping by consumers reduces the efficiency of matching in product markets (think of a standard matching function using shopping effort and marketing and distribution expenses as inputs). As a result, output declines.

For plausible estimates of shopping time and effort, the weight of household shopping in the matching function is around 0.1 (e.g Rios-Rull et al (2012)[17], Gourio and Rudanko (2013)[99]). A more important role for shopping effort is plausilbe once we take into account the distribution (both retail and wholesale), marketing and new product development sectors. These can be seen as intermediate sectors that perform much of the shopping and matching effort indirectly for households. We can then formulate our model in terms of retail stores,purchasing agents and marketing/advertising departments that must decide on the amount of resources dedicated to marketing and distribution under imperfect information, before knowing the full state of the economy.

A recession increases uncertainty about the buying power of customers. The higher uncertainty reduces the incentives of retailers and marketers to search for new products and customers, reducing

the efficiency of the search process. Similar mechanisms would work if aggregate uncertainty increases, or if we have sentiment shocks about the level of uncertainty as in Angeletos and Lao (2012) [11]. In an economy with flexible capital, similar mechanisms would affect the market for capital due to higher firm level idiosyncratic risk in a recession. Therefore, by linking higher uncertainty directly with shopping effort we can generate a causal effect of lower "aggregate demand" on output without resorting to indeterminacy or sticky prices.

8 Adding capital,durable goods, intermediate production inputs and search frictions

The section discusses several extensions of the previous model that allow for non zero aggregate saving and intertemporal choices that affect the production capacity of the economy. We examine economies with production functions that allow for variable capital and intermediate inputs, and a distinction between durable and nondurable consumption. We also analyse the intertemporal dimension of labour demand by introducing search frictions in labour markets. We discuss intermediate inputs in this section because in the long run they are very similar to a form of capital with 100% depreciation, and because they can be an important mechanism for amplifying shocks that is often neglected. The objective is to generalise the model from the previous section, until we reach an RBC framework that encompasses most existing models as special cases.

8.1 The real business cycle model with capital

So far our model of the economy has ignored capital. Implicitly, it has assumed that labour and intermediate goods were the only inputs, or that capital was fixed (no depreciation and infinite capital adjustment costs). We now add an adjustable capital stock to the analysis.

Capital in our economy is owned by many identical firms producing the final good of the economy. The representative final good firm's problem now becomes dynamic, so there is a role for intertemporal distortions through changes in the firm's discount factor $\beta_{f,t}$ to affect its decisions. Movements in the firm's discount factor can be caused by changes in the discount factor of its shareholders (the representative household) as in standard interpretations of the complete markets model. In fact in a standard complete markets economy $\beta_{f,t} = \beta_t \lambda_t$ of the households. We allow for differences in discount factors between households and firms ($\beta_{f,t} \neq \beta_t \lambda_t$) in order to incorporate in reduced form distortions in the firms' discount factor coming from other sources. These could be justified in a world with entrepreneurs or private equity, or in a world with dominant shareholders, or in an environment where managers and shareholders have opposing interests. In these situations we get firm level credit frictions and uninsured business risk that can generate difference between household and firm discount factors.

As for households, the firm discount factor is stochastic, reflecting for example business cycle shifts in the level of precautionary saving or the external finance premium faced by firms, or shifts in business owners' expectations that are not justified by rational expectations (e.g excessive optimism or pessimism). We include a tax or wedge on investment as a way to capture regulatory distortions or financing frictions specifically affecting investment (See Chari et al (2007)[50], Buera and Moll (2012) [35]).

We follow the benchmark assumption in the literature that investment is subject to a 1 period time to build delay. This may seem to be completely at odds with the pervasiveness of multiperiod projects that are often started more than a year prior to completion. In fact, the assumption can be a good approximation in world with multiperiod projects as long as the scale of such investments can be adjusted flexibly enough to changes in business conditions over the duration of the project.

There is conflicting evidence on the importance of time to build frictions for investment dynamics. New Keynesian models that try to match the SVAR evidence on monetary policy shocks such as Smets and Wouters (2007) [168] typically require high time to build adjustment costs (modelled to a first order approximation through convex investment adjustment costs). Jaimovich and Rebelo (2009) [115] rely on significant investment adjustment costs to explain why investment increases in advance of positive news about future TFP. In contrast Liu et al (2011) [139] find that these costs are minimal in the estimation of a RBC model with credit constraints, while Sims (2011) finds that matching SVAR evidence on productivity shocks requires sticky price models to have low investment adjustment costs. We do allow for convex adjustment costs on the level of capital (e.g Romer (2001) [161] ch. 8), though matching the time series dynamics of investment requires that these capital adjustment costs be small or moderate.

We start with the representative final goods firm. It solves

$$V_0^f = \max_{\{k_{t+1}, m_t, n_t\}} E_0 \Sigma_{t=0}^{\infty} \beta_{f,t} d_t,$$

$$\begin{aligned} d_t &= g(A_t m_t, e_{m,t}) z_t f(k_t, A_t n_t) - w_t(1+\tau_{nd,t})(n_t + m_t) - p_t^k(1+\tau_{k,t})[k_{t+1} - (1-\delta_t)k_t], \\ g_m &> 0,\ g_{mm} < 0,\ g_{e_m} > 0,\ f_n > 0,\ f_{nn} < 0, f_k > 0, f_{kk} < 0 \\ ,\ g(am, .)f(ak, an) &\leq ag(m, .)f(k, n) \text{ for all } a \geq 0.\ A_t = G_y A_{t-1}, G_y \geq 1. \end{aligned}$$

We also impose the transversality condition

$$\lim_{t\to\infty} E_0 \beta_{f,t} p_t^k k_{t+1} = 0.$$

$\tau_{k,t}$ **is an investment wedge that represents regulation, taxes, financing frictions and other changes in the actual or perceived costs of investments.** The depreciation rate δ_t is often assumed constant or subject to endogenous capital utilisation (e.g Greenwood et al (2000) [101]). Here we allow for a possible depreciation shock to δ_t that captures time varying technological and economic obsolescence. This is particularly relevant if we take the extended definition of k_t that includes intangible capital.

The transversality condition implies that under RE there are no bubbles in the price of capital evaluated using the firm's discount factor $\beta_{f,t}$. Here a bubble is defined as the difference between the price of capital and the present value of buying and holding capital forever without reselling it. This present value is often called the fundamental value. Note that the transversality condition is always satisfied if there are no capital adjustment costs ($p_t^k = 1$), and the growth rate of the capital stock is lower than that of discount rate $\left(\frac{1}{\beta_{f,t}} - 1\right)^{1/t}$.**While this no bubbles condition may sound restrictive in a world with big fluctuations in asset prices, it does allow for prices above RE fundamentals due to misperceptions about the returns to capital. It also allows for the possibility that p_t^k may look like it contains bubbles when evaluated under other discount factors**, for example under the complete markets assumption that $\beta_{f,t} = \beta_t \lambda_t$. This is always the case when capital can be used as collateral.

Just as for households, we can reformulate this dynamic optimisation problem recursively as

$$V^f(k_t, \tau_t, S_t) = \max_{k_{t+1}, m_t, n_t} d_t + E_t \frac{\beta_{f,t+1}}{\beta_{f,t}} V^f(k_{t+1}, \tau_{t+1}, S_{t+1}), \text{ where}$$

$$d_t = g(A_t m_t, e_{m,t}) z_t f(k_t, A_t n_t) - w_t(1+\tau_{nd,t})(n_t + m_t) - p_t^k(1+\tau_{k,t})[k_{t+1} - (1-\delta_t)k_t],$$

so that the firm optimises by taking into account the effect of its decisions on current payoffs and the future value of the firm starting next period.

8.1.1 Tangible versus intangible capital

In most existing models business capital is just tangible capital corresponding to the investment variable in NIPA. Recent research has emphasized that intangible capital such as R&D, some marketing expenses, firm specific training and investment in improving total factor productivity may be almost as important for both long run growth and business cycles. By some estimates, the intangible capital investment is as big as tangible capital investment (e.g Corrado et al, 2006). Parente and Prescott (1999) [156] develop a theory of income differences across countries through distortions that affect the accumulation of of intangible capital. Johri et al (2009)[110] find that taking intangible capital into account can explain hump shaped responses to shocks without the typical appeal to more controversial habit formation or investment adjustment costs.

One interpretation of our baseline model is that we lump these two sorts of capital together and treat them as perfectly substitutable from the firm's perspective. This assumption underlies the vast majority of current business cycle models, though of course taking adjustable intangible capital into account would require them to use a higher share of capital in the production function. If the model includes intangible capital we also need to modify the definition of investment in the data that we use to match that in the model: In particular model GDP now includes intagible capital investment. MacGrattan and Prescott (2012) [147] explicitly take into account intangible capital in an RBC model. They find it reverses some of the usual conclusions about business cycles. For example, a common finding with standard calibrations based only on tangible capital is that the correlation between labour productivity and output has been negative in recent US recessions, implying a more prominent role for labour wedge shocks in business cycles. Once we take into account the extra decline in output in a recession due to lower intangible investment labour productivity becomes more procyclical and the role of TFP shocks in business cycles increases.

8.1.2 More on the production function

The production function is usually specified to be Cobb-Douglas in capital and labour:

$$\begin{aligned} f(k, An) &= [k^{\alpha}(An)^{1-\alpha}]^{\gamma}, \\ 0 &\leq \alpha \leq 1, 0 \leq \gamma \leq 1. \end{aligned}$$

This formulation has the advantage that it no longer matters whether labour augmenting, capital augmenting or neutral technological change have long run trends. In contrast, more general production function require the long run growth to be labour augmenting in order to have a balanced growth path. **Within the constant elasticity of substitution in k, n production functions, it can be shown that only the Cobb-Douglas function is compatible with a balanced growth path** (Gomme and Rupert (2006) [97]). Jones (2005) derives a Cobb-Douglas production function as the the result of optimal choices by firms over different fixed input proportions production techniques. If different techniques have a Pareto distribution, then optimal firm choices for different production locations lead to an aggregate production function for the firm that is Cobb-Douglas.

Another way to think of a production function is as coming from a matching process inside the firm between workers and different physical production inputs like capital. Under this interpretation the production function describes the number of successful matches between different production inputs, with each match resulting in one unit of the good. In this case, a Cobb-Douglas matching function (see Stevens (2005) [169] or Lagos (2005) for a possible microfoundation) implies a Cobb-Douglas production function. Long run shifts in the capital intensity of production can be captured by changes in α.

The standard RBC model assumes constant returns to scale, so that the number and size of firms is indeterminate. But in some applications decreasing returns to scale are more realistic as a simple way of incorporating procyclical profits. With decreasing returns to scale, we must specify some entry cost or some congestion cost (such that a larger number of firms reduces productivity) or a fixed resource required to operate a firm such as an entrepreneur (in which case production has constant returns to scale when including the entrepreneur's labour, and the measure/number of entrepreneurs determines the number of firms) in order to pin down a finite number of firms. Otherwise, free entry would lead to a very large (infinite) number of very small (infinitely small) firms in the long run.

To take the Cobb-Douglas case

$$y = k^{\alpha} n^{\gamma}$$

with symmetric firms, aggregate output $Y = My$ is invariant to the number of firms M if we modify the production function to

$$\begin{aligned} y &= M^{\alpha+\gamma-1} k^{\alpha} n^{\gamma}, \text{ so that} \\ Y &= (Mk)^{a} (Mn)^{\gamma}. \end{aligned}$$

$M^{\alpha+\gamma-1}$ is a congestion factor which disappears under constant returns to scale. Alternatively, consider adding a manager as a production factor such that

$$y = k^{\alpha} n^{\gamma} n_m^{1-\alpha-\gamma}, n_m = 1$$

for each firm. The firm's profits are the same as before if we add back the salary/rent of the manager w_m. Decreasing returns to scale production functions are common in models with heterogeneous firms (e.g Atkeson and Kehoe (2007) [13], Restuccia and Rogerson (2007) [160]) since they generate a non-degenerate distribution of productivity across firms. In steady state, their effects are similar to the effects of adding monopolistic competition with a constant markup as in Rotemberg and Woodford (1995) (Cooley et al (1995) [62],ch. 9). They also provide a simple way to introduce positive profits beyond the return to capital, delinking the value of firms from the value of the capital stock.

8.1.3 Assumptions on long run growth

As before we assume an exogenous long run balanced growth path on which output per worker grows at a constant rate $G_y - 1$. The key feature of exogenous growth models such as this one is that changes in the levels of distortions or policy variables cannot affect the long run growth rate of the economy. They only affect the detrended level of economic activity in the long run. Policy changes can have significant effects on the growth rate of the economy for many years or even decades if there are elements that slow down the transition of the economy to a new balanced growth path (see Buera and Shin (2012) [36] for example). Since most distortions or policy variables are bounded, this places a limit on the ability of policy makers to affect the growth rate of the economy in the long run. If one finds an unbounded policy variable, then a process of continual reforms could have permanent effects on growth.

Endogenous growth models such as those in Jones et al (2005) [118] or Ljunqvist and Sargent (2004, ch. 14) ([140]) can generate long run growth effects of level policy changes. For example, some endogenous growth models predict that higher saving rates lead to higher long run growth rates. However, they make the unrealistic prediction that countries with larger populations grow faster or have higher levels of income (Jones 1999). Furthermore, they predict long run, almost permanent differences in growth rates between countries that are hard to find (as opposed to long run differences in the levels of income predicted by exogenous growth models with TFP differences, and they counterfactually predict that very fast growth rates (growth miracles) are actually more likely for rich countries with better institutions than for developing countries. Exogenous growth models can easily explain why growth miracles are more likely in developing economies, and they do not predict any systematic link between population size and growth rates (see Parente and Prescott (1999) [156]).

The modern literature on the macroeconomics of development has focused on exogenous growth models, in which the key questions are explaining the long run differences in income levels across countries and explaining the transition path of economies in response to occasional discrete regime changes such as major reforms or technological revolutions (e.g the shift from communism to capitalism, the IT revolution, the development of electric power) as opposed to endogenising a continuous growth process. See for example Buera and Shin (2012),[36], Atkeson and Kehoe (2007) [13], Restuccia and Rogerson (2007) [160], Jones (2011a) [117].

At the same time, exogenous growth models have trouble explaining why some very poor countries (e.g in Africa) seem to have not only failed to converge towards rich countries but seem to have diverged. This divergence can be explained in the context of a model in which a minimum threshold of development needs to be reached in order to enter a modern economic growth regime, with a constant growth trend. Below this threshold an economy still follows a Malthusian regime with barely any growth (e.g Parente and Prescott (2004)). For our purpose of business cycle analysis, we can approximate this difference in regimes by assuming a balanced growth path with approximately zero growth for very poor countries, while allowing for long run growth of around 1.5% per working age person in other economies.

8.1.4 The supply of capital and the aggregate resource constraint

To complete the description of the economy, we add a large number of identical capital producers that transform output (i.e. labour and capital in the same proportion as final output) into new capital

$$\begin{aligned} x_t^k &= k_{t+1} - (1-\delta_t)k_t = a_t \tilde{G}(\frac{I_t}{k_t})k_t \\ , G'(.) &> 0,\ \tilde{G}''(.) \leq 0. \end{aligned}$$

a_t **represents investment specific technology shocks. A strictly concave** $\tilde{G}()$ **represents capital adjustment costs. The special case of** $\tilde{G''} = 0$ **represents the common case in RBC research of no adjustment costs.** We can think of these adjustment costs as representing extra frictions and bottlenecks in the process of selling and distributing capital goods, that make it harder to expand sales and installation of new goods quickly. Capital producers make profits

$$\pi_{k,t} = p_{k,t} x_t^k - I_t.$$

These profits go to the households. **Profit maximisation yields the first order condition for** I_t,

$$a_t p_{k,t} \tilde{G}'(\frac{I_t}{k_t}) = 1.$$

This defines a supply curve for capital. Finally, the resource constraint becomes

$$c_t + I_t = y_t.$$

8.1.5 The firm's capital investment decision

Now return to the representative firm's problem. Optimal firm decisions imply

$$\begin{aligned} n_t &: \ g_t z_t f_{n,t} = w_t(1+\tau_{nd,t}), \\ m_t &: \ g_{m,t} z_t f_t = w_t(1+\tau_{nd,t}), \\ k_{t+1} &: \ p_{k,t}(1+\tau_{k,t}) = E_t \frac{\beta_{f,t+1}}{\beta_{f,t}}[g_{t+1} z_{t+1} f_{k,t+1} + (1-\delta_{t+1})(1+\tau_{k,t+1})p_{k,t+1}]. \end{aligned}$$

The optimality condition for capital has the standard interpretation of decisions equating marginal costs and benefits. On the left

hand side, the cost of a small increase in the capital stock is the price of capital adjusted by the investment wedge. On the right hand side, the benefit of a small increase in the capital stock is the discounted return to capital.

As for household Euler equations, we can use this to get an asset pricing relation. Iterating foreward, using the law of iterate expectations ($E_t E_{t+j} x_{t+j+1} = E_t x_{t+j+1}$ for $j \geq 0$) and imposing the transversality condition on capital,

$$p_{kt}(1+\tau_{k,t}) = \Sigma_{j=0}^{\infty} E_t \frac{\beta_{f,t+j+1}}{\beta_{f,t}} \Pi_{i=1}^{j}(1-\delta_{t+i}) g_{t+1+j} z_{t+1+j} f_{k,t+1+j},$$

, where by convention $\Pi_{i=1}^{0} x_i = 1$.

That is the price of capital is equally to a discounted value of all future marginal returns to capital.

Non convexities and aggregate investment Investment projects often

come in discrete sizes, and often require significant fixed costs to implement. The optimality condition for capital above can be reinterpreted as the choice of a firm over a large number of fixed size projects with varying levels of profitability, and a cost per project of $p_{k,t}(1+\tau_{k,t})$. At the optimum, new projects are implemented up to the point where the remaining projects are unprofitable at a cost of $p_{k,t}(1+\tau_{k,t})$. The litterature on non convex adjustment costs emphasizes that this is not a realistic model of investment at the level of individual firms. But large idiosyncratic shocks to individual investment project returns and adjustment costs reduce the correlation between individual firms' investment decision and make aggregate investment smooth, though much more volatile than aggregate consumption. While firms face significant adjustment costs, these costs can be swamped by the volatility of (mostly idiosyncratic) shocks hitting firms. Furthermore, while the proportion of firms adjusting under fixed adjustment costs is lower than in the frictionless models, firms that do adjust may make bigger changes in their capital stock. As a result, the aggregate adjustment rate to shocks can be large despite even if each firm is subject to significant adjustment costs.

Elsby and Michaels (2014) [74] study a canonical partial equilibrium model of firm decisions under fixed adjustment costs, with both aggregate and idiosyncratic shocks. They show that if the fixed cost is independent of the size of the firm, aggregate variables are independent of the adjustment costs to a 1st order approximation around the frictionless, zero adjustment cost, limit. They show for realistic calibrations in a labour demand application that this approximate irrelevance of fixed costs is often a good guide in practice, even in partial equilibrium. More generally aggregate behaviour in the frictionless (no adjustment costs) limit becomes a better approximation as the level of fixed costs declines relative to the volatility of idiosyncratic shocks. Adjustment costs reduce the inflow of firms into any point the new frictionless cross section distribution of optimal firm decisions $f(x^*)$ after an aggregate shock. But they also reduce the outflow of firms away from the same point in $f(x^*)$. As the size of the fixed adjustment cost falls relative to the magnitude of idiosyncratic shocks, the inflow and outflow tend to balance out so that the distribution of firm level decisions $f((x)$ converges to the frictionless $f(x^*)$.

Khan and Thomas (2008) [124] generalise the RBC model to include heterogenous production units and nonconvex adjustment costs. They examine a calibration where fixed adjustment costs do matter for aggregate investment in partial equilibrium, so the approximate irrelevance result of Elsby and Michaels (2014) [74] does not hold. For example, the standard deviation of investment in the fixed adjustment costs model is around 2 thirds that in the frictionless model in partial equilibrium. Nevertheless, they find that absent other frictions the aggregate investment dynamics of this more complex model in response to aggregate TFP shocks are very close to those of the baseline RBC model with a representative firm and low capital adjustment costs. They trace this result to general equilibrium effects, with price adjustments reducing the effects of adjustment costs at the aggregate level.

Intuitively, consumption smoothing by households that own firms reduces movements in the frictionless optimal capital stock of firms, which is close to the target capital stock of adjusting firms. The lower volatility of the frictionless capital stock decisions lead to lower variation in the proportion of firms deciding to adjust their capital in response to aggregate shocks. But without time varying adjustment rates on the extensive margin much of the nonlinear partial equilibrium investment dynamics that are supposed to characterise the fixed adjustment cost model is eliminated.

On the other hand, the response of aggregate investment to 2nd moment,uncertainty, shocks can differ significantly with fixed adjustment costs. Higher idiosyncratic firm-level uncertainty lower investment, because it increases the likelihood of paying the fixed cost and raises the ex-ante cost of investment (recall the discussion of uncertainty shocks in the presence of fixed labour adjustment costs earlier and the paper by Bloom et al (2012) [34]).

External financing frictions and differences in the discount factors of households and firms **Define the ratio of firm and household discount factors**

$$v_{f,t} = \frac{\beta_{f,t}}{\lambda_t \beta_t}.$$

In the baseline complete markets RBC model $v_{f,t} = 1$. In models with limited participation of households in equity markets (so there is a significant divide between firm owners and workers) fluctuations in $v_{f,t}$ capture differences between the marginal utility of consumption, external finance premia or misperceptions about the future of firm owners and other households. To see the link to external finance premia, introduce as in Khan and Thomas (2013) [125]a credit market in which firms lend and borrow from each other or from saving households. Firm borrowing is subject to an external finance premium $efp_{e,t+1}$. For simplicity, we assume this external finance premium is taken as given by firms. The firm's dividend is now defined as

$$\begin{aligned} d_t &= g(A_t m_t, e_{m,t}) z_t f(k_t, A_t n_t) - w_t(1+\tau_{nd,t})(n_t + m_t) - p_t^k(1+\tau_{k,t})[k_{t+1} - (1-\delta_t)k_t] \\ &\quad -b_{f,t+1} + b_{f,t} R_t efp_{e,t}, \end{aligned}$$

where $b_{f,t}$ is the firm's stock of lending to other firms. The firm now also chooses sequences of $\{b_{f,t+1}\}$ to maximise

$$V_0^f = E_0 \Sigma_{t=0}^{\infty} \beta_{f,t} d_t.$$

The first order condition for bonds is

$$\begin{aligned} b_{f,t+1} \quad &: \quad \beta_{f,t} = R_{t+1} E_t efp_{e,t+1} \beta_{f,t+1}, \Longleftrightarrow \\ \lambda_t &= R_{t+1} E_t efp_{e,t+1} \lambda_{t+1} \frac{\beta_{t+1}}{\beta_t} \frac{v_{f,t+1}}{v_{f,t}}. \end{aligned}$$

The other first order conditions are the same. Debt market clearing implies that

$$\theta_f b_{f,t+1} + \theta_{sa} b_{sa,t+1} + \theta_{bo} b_{bo,t+1} = 0.$$

First, note that if there are no differences between the discount factors of firms and households ($v_{f,t} = 1$) and if external finance premia are predetermined, it must be the case that

$$efp_{e,t+1} = efp_{h,t+1}$$

and we get the the same first order condition as for households' bond holding decision. In this case inter firm borrowing is equivalent to inter household borrowing used to fund the firm via negative dividends (inflows of cash from the shareholders).

Differences in the external finance premia of households and firms can only occur if there is a difference between their discount factors,

$v_{f,t} \neq 1$. **Otherwise, firms could always circumvent any extra financing frictions by borrowing through households (or vice versa households could use firms to borrow for them and fund them if firms face lower financing frictions). The difference in discount factors can arise from limited participation in firm ownership (e.g entrepreneurs versus workers as in Iacoviello's (2005) [111] model) or because there are restrictions on having negative dividends or on reducing dividends when the firm wants to borrow more. The latter restrictions could be due to potential conflicts of interests between managers and shareholders that are aggravated when the firm keeps its cash flows or when it tries to get cash from shareholders (see Arellano et al (2012) [12] for a similar argument).** Once we allow $v_{f,t} \neq 1$, differences between $efp_{e,t+1}$ and $efp_{h,t+1}$ will contribute to movements in $v_{f,t}$ and distort investment decisions of the firm.

We will develop 2 complementary versions of the firm's optimal investment condition using either $epf_{e,t+1}$ or $efp_{h,t+1}$. The 1st version emphasizes the link between investment and borrowing costs, either in the form of interest rates or other restrictions that can be mapped into an external finance premium. Using a certainty equivalence approximation and combining the firm' bond and capital first order conditions, we get

$$\mathbf{v1}\ , k_{t+1} : p_{k,t}(1+\tau_{k,t}) = \frac{1}{R_{t+1}E_t efp_{e,t+1}} E_t[g_{t+1}z_{t+1}f_{k,t+1}+(1-\delta_{t+1})(1+\tau_{k,t+1})p_{k,t+1}].$$

A similar equation is at the heart of many models with credit constraints on investment (e.g Bernanke et al (1999) [28], Iacoviello (2005) [111], Gertler and Kiyotaki (2010) [94] and Christiano et al (2012) [56]). An increase in the cost of external financing $efp_{e,t+1}$ increases the required return on an extra unit of capital, the marginal product of capital. By the concavity of the production function, this implies a reduction in investment as firms cut off funds to less profitable, marginal investments. Here, efp_e is seen as an exogeneous wedge. Models with explicit credit constraints endogenise part of this wedge, linking it for example to the amount of debt to equity of firms (the leverage ratio).

The 2nd version of the optimality condition for capital investment decomposes the stochastic discount factor of the firm into its household component and a firm specific component. It emphasizes the relation between firms' and households stochastic discount factors. Using a certainty equivalence approximation again, we get

$$p_{k,t}(1+\tau_{k,t}) \simeq E_t \frac{\beta_{t+1}\lambda_{t+1}}{\beta_t\lambda_t} E_t \frac{v_{f,t+1}}{v_{f,t}} E_t[g_{t+1}z_{t+1}f_{k,t+1}+(1-\delta_{t+1})(1+\tau_{k,t+1})p_{k,t+1}],$$

which combining with the households' Euler equation becomes

$$\mathbf{v2},\ k_{t+1} : p_{k,t}(1+\tau_{k,t}) \simeq \frac{1}{R_{t+1}E_t efp_{h,t+1}} E_t \frac{v_{f,t+1}}{v_{f,t}} E_t[g_{t+1}z_{t+1}f_{k,t+1}+(1-\delta_{t+1})(1+\tau_{k,t+1})p_{k,t+1}].$$

To fully understand the relation between these two optimal investment conditions we need to link these two external finance premia. Combine the optimality conditions of the firm and household sectors, and use another certainty equivalence approximation,

$$E_t \frac{\beta_{t+1}}{\beta_t} \frac{\lambda_{t+1}}{\lambda_t} \frac{v_{f,t+1}}{v_{f,t}} \simeq E_t \frac{\beta_{t+1}}{\beta_t} \frac{\lambda_{t+1}}{\lambda_t} E_t \frac{v_{f,t+1}}{v_{f,t}}$$

to obtain

$$E_t \frac{v_{f,t+1}}{v_{f,t}} \simeq \frac{E_t efp_{h,t+1}}{E_t efp_{e,t+1}}.$$

This equation shows that at least for low enough levels of aggregate uncertainty there is a one to one relation between shifts in the relative discount factors between firms and households and the relative external finance premia. So an increase in the firm's discount factor $\frac{v_{f,t+1}}{v_{f,t}}$ will typically be associated with a reduction in the relative external finance premium for firms. For example an increase in firm's precautionary saving that is not specifically tied to idiosyncratic investment risk is actually associated with a lower relative external finance premium for firms and a higher demand for capital as a form of saving. In contrast, an increase in uninsured idiosyncratic investment risk lowers the risk adjsted return on capital, which can reduce investment. This effect is captured in our framework through a higher investment wedge $\tau_{k,t}$.

In fact, in $v1$ of the optimal investment decision, any deviation from certainty equivalence that does not affect $efp_{e,t+1}$ will be captured by $\tau_{k,t}$. At the same time by raising the desire for precautionary saving, higher idiosyncratic investment risk could actually under certain circumstances increase investment as captured by a higher $\frac{v_{f,t+1}}{v_{f,t}}$ (see Angeletos (2007) [8] and Covas (2005) for a steady state analysis of the conditions under which higher idiosyncratic investment risk reduces investment). More generally, shocks to $\tau_{k,t}$ capture changes in the perceived riskiness of capital, or misperceptions about the returns to capital that deviate from RE (similar to some deviations in the discount factor for households β_t).

Some basic partial equilibrium comparative dynamics of the firm's investment decision **Both versions of the optimal investment equation establish an important link between interest rates on debt and the return to capital investment. Ceteris paribus an increase in the marginal productivity of capital will through general equilibrium effects tend to raise real interest rates. In the opposite direction an increase in interest rates R_{t+1} or in borrowing premia $efp_{h,t+1}$ will discourage investment in partial equilibrium.** It may seem surprising at 1st that the household external finance premium can affect firm investment. But this is quite natural if funding for the firm comes from financially constrained households, that have borrowed to finance the firm (presumably giving them a higher rate of return than the cost of the loan). To the degree that firm owners are unaffected by changes in the household external finance premium $efp_{h,t+1}$ (for example if they are wealthy enough such that borrowing constraints are irrelevant), this is captured by shifts in $\frac{v_{f,t+1}}{v_{f,t}}$ that offset movements in $efp_{h,t+1}$ so that the firm's external finance premium $efp_{e,t+1}$ does not change.

Other intuitive comparative dynamics can be derived based on the diminishing marginal productivity of capital:

- **A higher current investment wedge $\tau_{k,t}$ relative to $\tau_{k,t+1}$ increases the required return to capital and discourages investment.**

- **If $\tau_{k,t}$ and $\tau_{k,t+1}$ both increase by the same amount, investment declines because the future investment wedge increase is discounted by $\frac{\beta^f_{t+1}}{\beta^f_t}$.**

- **An increase in future productivity $g_{t+1}z_{t+1}$ raises investment.**

- **An increase in current productivity $g_t z_t$ raises investment if the shock is persistent. Higher shock persistence leads to a bigger effect on investment, since it leads to a bigger rise in future productivity.**

- **Lower returns to scale in production** γ (or higher market power in a model with imperfect competition) reduce the impact of any change in productivity on the marginal product of capital. As a result, they **reduce the change in investment in response to productivity shocks.**

- **Concavity in the production function leads to an assymetric response of investment to productivity shocks**: because of diminishing marginal productivity of capital a positive productivity shock requires a higher absolute value increase in capital to get the same increase in production than the absolute value of the decline in capital for a negative shock. As a result, **investment is more responsive to positive productivity shocks than to negative ones** (Strebulaev and Whited, 2012 [6]).

- **An increase in the ratio of firms' to households' discount factor $\frac{v_{f,t+1}}{v_{f,t}}$ decreases the required return to capital and increases investment.** This is associated with a decline in the firm's external finance premium $efp_{e,t+1}$ relative to that of households'.

- **An increase in the current price of capital $p_{k,t}$ or a decrease in $p_{k,t+1}$ reduce investment.** The first effect just says that a higher cost of new capital reduces desired investment. The second effect is best interpreted using our asset pricing equation linking $p_{k,t}$ to the future marginal returns to capital. A lower $p_{k,t+1}$ is equivalent to a reduction in future returns to capital. This reduces desired investment.

Discount factor or external finance premium shocks and investment Two key shocks often thought to be important for business cycles are the household discount factor shock to $\frac{\beta_{t+1}}{\beta_t}$ and the external finance premium shock to $efp_{h,t+1}$. From the household perspective in partial equilibrium these shocks are very similar. However, these shocks have different effects on investment. For a fixed $\{v_{f,t+j}\}_{j=0,\ldots\infty}$ sequence, discount factor shocks that increase $\frac{\beta_{t+1}}{\beta_t}$ also encourage investment as a form of saving. These shocks on their own cannot be the main factor behind business cycles in an RBC model, because they tend to generate negative comovement of consumption and investment. Households become more eager to save, pushing them to reduce consumption and hold more capital.

A shock to the external finance premium $efp_{h,t+1}$ is also incapable of generating realistic business cycles without some spillover into the labour wedge or into TFP. If $v_{f,t} = 1$ is and there is no compensating factor that reduces the return to capital, an increase in $efp_{h,t+1} = efp_{e,t+1}$ would lower investment. But if the same increase in the household external finance premium also decreases consumption then in a closed economy with standard income effects on labour supply we get a violation of the resource constraint in the short run, when the capital stock is almost fixed. Similar reasoning in the opposite direction establishes that a pure $efp_{h,t+1}$ shock that does not affect $\frac{v_{f,t+1}}{v_{f,t}}$ or the marginal return to capital at a fixed k_{t+1} simply leads to an opposite movement in interest rates R_{t+1} such that output, consumption and investment do not change in the end.

An increase in $efp_{h,t+1}$ will behave more like a higher household discount factor shock to $\frac{\beta_{t+1}}{\beta_t}$ if $\frac{v_{f,t+1}}{v_{f,t}}$ increases or labour supply increases sufficiently due to a negative income effect for households (since household consumption declines). These movements in the relative discount factor of firms or in labour supply increase the return to capital and encourage investment. In a model with entrepreneurs owning production, the higher $\frac{v_{f,t+1}}{v_{f,t}}$ means firm owners' consumption falls by less in response to a higher household external finance premium $efp_{h,t+1}$ so that the marginal utility of consumption of entrepreneurs decreases relative to that of workers (the households in our model). Equivalently $\frac{v_{f,t+1}}{v_{f,t}}$ increases because the firms' external finance premium $efp_{e,f,t+1}$ falls relative to that of households.

Note that shocks to $efp_{h,t+1}$ would produce positive comovement of consumption and investment even with $v_{f,t} = 1$ if nominal prices are sticky (see

Smets and Wouters (2007) [168] for the efp_{h+1} "risk premium" shock, and compare to the β_{t+1}/β_t "preference" shock in Smets and Wouters (2002) [167]). With sticky nominal prices, a shock that simultaneously increases consumption and investment demand is accomodated in the short run by rising production at fixed prices.

Given our earlier discussion, the question arises about how to map shifts in the level of precautionary saving into discount factor or external finance premium shocks. The following simplified setup may provide some guidance in terms of the key difference between shifts in precautionary saving and other credit constraints. Consider an economy where households are subject to uninsured idiosyncratic income risk. Households can save or borrow in a risk free asset or they can invest in a well diversified portfolio of capital. The only credit constraint is the natural borrowing constraint which does not bind. We assume perfect foresight on aggregate shocks so that investment in the capital portfolio earns a risk free return of R^k_{t+1}. with the usual properties that $\frac{\partial R^k_{t+1}}{\partial k_{t+1}} > 0, \frac{\partial^2 R^k_{t+1}}{\partial^2 k_{t+1}} < 0$. Household $i's$ optimisation problem leads to the following first order conditions:

$$\begin{aligned} u_{c_{i,t}} &= \beta R_{t+1} E_t u_{c,_{it,+1}}, \\ u_{c_{i,t}} &= \beta R^k_{t+1} E_t u_{c,_{it,+1}}. \end{aligned}$$

Regardless of the presence of uninsured idiosyncratic income risk, these first order conditions imply the classic result from the perfect foresight complete markets RBC model,

$$R_{t+1} = R^k_{t+1}.$$

It is clear in this economy that any increase in precautionary saving, for example due to uninsured income risk will only affect investment through changes in the risk free interest rate R_{t+1}. Higher precautionary saving will tend to increase investment by reducing R_{t+1}, unless we explicitly introduce a channel linking higher precautionary saving to lower TFP (as in the section on aggregate demand effects). In contrast, explicit binding borrowing constraints or external finance premia would lead to an equation of the form

$$R_{t+1} efp_{h,t+1} = R^k_{t+1}.$$

This suggests that changes in aggregate precautionary saving will in general map into changes in the household discount factor.

8.2 Search frictions in capital markets

The market for capital is believed by many to be subject to significant search frictions (e.g Kurman and Petrosky-Nadeau 2007, Rios Rull et al (2012) [17]). Here we provide an extension that allows for search costs by firms buying new capital. Searching for new capital requires the firm to post "vacancies" V_t^k, such that

$$x_t^k \leq \varphi_t^k V_t^k \ ,(\mu_t^k).$$

φ_t^k is the rate at which "vacancies" V_t^k are filled. There is a cost to posting V_t^k, $\Gamma(V_t^k)$, with $\Gamma_{V^k} > 0, \Gamma_{V^k V^k} \geq 0$. This modifies the firm's cash flow equation to

$$\begin{aligned} d_t &= g(A_t m_t, e_{m,t}) z_t f(k_t, A_t n_t) - w_t(1+\tau_{nd,t})(n_t + m_t) - p_t^k(1+\tau_{k,t})[k_{t+1} - (1-\delta_t)k_t] \\ &\quad -\Gamma(V_t^k). \end{aligned}$$

Finally we modify the first order condition for capital, and add a first order condition for V_t^k :

$$\begin{aligned} k_{t+1} &: p_{k,t}(1+\tau_{k,t}) + \mu_t^k = E_t \frac{\beta_{f,t+1}}{\beta_{f,t}} (g_{t+1} z_{t+1} f_{k,t+1} + (1-\delta_{t+1})[(1+\tau_{k,t+1})p_{k,t+1} + \mu_{t+1}^k]), \\ V_t^k &: \Gamma_{V_t^k} = \mu_t^k \varphi_t^k. \end{aligned}$$

Increasing investment today now has an additional search cost, but it also reduces search costs in the next period. As for search frictions in the market for consumption goods, we can express the effects of the search frictions in terms of an extra investment wedge in the optimality condition for capital, by defining

$$\hat{\tau}_{k,t} = \tau_{k,t} + \frac{\Gamma_{V_t^k}}{p_{k,t}\varphi_t^k}.$$

In more microfounded matching frictions models, both φ_t^k and g_t are functions of V_t^k. From the first order conditions and using certainty equivalence approximations, an increase in $\tau_{k,t}$ or a decrease in

$$E_t \frac{v_{f,t+1}}{v_{f,t}} \simeq \frac{E_t efp_{h,t+1}}{E_t efp_{e,t+1}}$$

would tend to reduce V_t^k which through the aggregate matching function in the capital markets reduces investment specific TFP. This maps into a lower marginal efficiency of investment shock in models omitting this channel. Finally note that the aggregate resource constraint is modified to

$$c_t + I_t = y_t - \Gamma(V_t^k) = GDP_t.$$

8.3 Aggregate fluctuations with a variable capital stock

8.3.1 Front loaded TFP and labour wedge shocks

The presence of capital provides agents in the economy with another channel for separating changes in the output of the economy from

changes in consumption that they actually care about through saving. This can significantly affect the dynamics of certain variables. For example, in a world with fixed capital and a closed economy, output equals consumption and real interest rates adjust sufficiently to ensure savings demand equals the zero net supply. In a model with adjustable capital, agents can smooth consumption in response to shocks by changing their investment in capital. [28] To understand the difference made by flexible capital adjustment, it helps to start with the case in which the price of capital $p_{k,t}$ is unaffected by aggregate fluctuations This is the case without capital adjustment costs such that $p_{k,t} = 1$ for all t.

Consider a positive front loaded productivity shock such that $z_{t+j} > z_{t+j+1} > z$ **for all** j (or $g_{t+j} > g_{t+j+1} > g$).

1) The increase in TFP causes higher output both directly as productivity increases but also indirectly by raising labour demand, since the marginal product of labour is higher.

2) To attract more workers, wages have to increase. The wage increase has two effects. First there is a substitution effect: a higher return to working encourages higher labour supply. Second, there is an income effect. Since leisure is a normal good, higher income encourages leisure and discourages labour supply.

3) With fixed capital and preference that are compatible with a long run balanced growth path these two effects offset each other perfectly in the economy with fixed capital stock. As a result, output and consumption only increase due to the direct effect of higher productivity. Unless, the TFP shock also leads to a shift in the labour wedge, equilibrium labour supply does not change.

4) With fixed capital, the Euler equation says that the real interest rate must decline. Households would like to smooth consumption by increasing saving in response to the higher current productivity of the economy relative to the future. Since net saving is zero in equilibrium, this higher demand for saving reduces interest rates. The reduction in interest rates leads households to accept an increase in their current consumption relative to consumption next period.

5) With flexible capital, the higher desire to save of households can be channeled into higher investment. Households no longer have

[28] In the current model this is done indirectly by firms to reflect the desire saving of their shareholders. In an environment with a clear separation between households and firm owners, this is achived by households adjusting the quantity of lending to the business sector.

to increase their consumption by as much as output. Therefore, consumption will be smoother than output and investment, and interest rates will not have to decline so much to clear credit markets.

6) Typical TFP shocks are quite persistent. As a result, the marginal product of capital will increase. By the optimality condition for capital investment, the higher marginal product of capital will push up real interest rates. The rise in investment propagates the initial effect of the TFP shock by raising the capital stock, which increases future labour productivity for several periods.

7) Smaller increase in consumption and the higher real interest in the flexible capital economy imply that the substitution effect of higher wages dominates the income effect. As a result, labour supply now actually increases even with BGP preferences, amplifying the effect of the productivity shock.

A similar analysis applies to labour wedge shocks, though with a stronger response of employment and hours:

Consider a reduction in the labour demand wedge $\tau_{nd,t}$:

1) The reduction in the labour demand wedge $\tau_{nd,t}$ raises labour demand, pushing up wages.

2) Higher wages raise labour supply through a substitution effect, but at the same time they reduce labour supply through the income effect. However, in contrast to the TFP shock, the income effect of the higher wage is countered by the reduction in transfers T_t. As a result, the substitution effect dominates and equilibrium labour supply increases. With fixed TFP and capital, the higher equilibrium labour supply raises output.

3) With fixed capital, the Euler equation says that the real interest rate must decline. Households would like to smooth consumption by increasing saving in response to the higher current output of the economy relative to the future. Since net saving is zero in equilibrium, this higher demand for saving reduces interest rates. The reduction in interest rates leads households to accept an increase in their current consumption relative to consumption next period.

4) With flexible capital, the higher desire to save of households can be channeled into higher investment. Households no longer have to increase their consumption by as much as output. Therefore, consumption will be smoother than output and investment, and interest rates will not have to decline so much to clear credit markets.

5) Typical labour wedge shocks are quite persistent. As a result, the marginal product of capital will increase. By the optimality condition for capital investment, the higher marginal product of capital will push up real interest rates. The rise in investment propagates the initial effect of the labour wedge shock by raising the capital stock, which increases future labour productivity for several periods.

6) Smaller increase in consumption and the higher real interest in the flexible capital economy further strengthen the substitution effect of higher wages relative to the income effect. As a result, labour supply and output now increase more than with fixed capital.

$$
\begin{aligned}
\Delta z_{t+j} &> 0,\ \Delta g_{t+j} > 0\ ,\ \Delta \tau_{nd,t+j} < 0 \Rightarrow \Delta n^d_{t+j} > 0, \Delta k^d_{t+j+1} > 0 \Rightarrow \Delta w_{t+j} > 0,\ \Delta R_{t+j} > 0 \\
&\Rightarrow \Delta n^s_{t+j} > 0 \Rightarrow \Delta n_{t+j} > 0 \\
&\Rightarrow \Delta c_{t+j} > 0,\ \Delta I_{t+j} > 0,\ \Delta y_{t+j} > 0. \\
\Delta I_{t+j} &> 0 \Rightarrow \Delta k_{t+j+1} > 0 \\
&\Rightarrow \Delta n^d_{t+j+1} > 0 \Rightarrow \Delta w_{t+j+1} > 0 \\
&\Rightarrow \Delta n^s_{t+j+1} > 0 \Rightarrow \Delta n_{t+j+1} > 0 \\
&\Rightarrow \Delta c_{t+j+1} > 0,\ \Delta I_{t+j+1} > 0,\ \Delta y_{t+j+1} > 0
\end{aligned}
$$

If the shocks are transitory, then eventually TFP and labour wedges return back to their long term trend. As a result, the initial increases in outpu, consumption, investment and labour are reversed. On the other hand, a permanent increase in TFP leads to a permanent increase in output, consumption and investment. A permanent reduction in the labour wedge also leads to a permanently higher labour input.

Note that the intertemporal substitution effect of interest rates on labour supply also applies to households with negative net assets: in this case higher interest rates on borrowing also encourage labour supply in order to reduce debt repayments. For a more comprehensive analysis of the interaction between investment, interest rates, labour supply and the persistence of productivity shocks in the RBC model with variable capital see King and Rebelo (2000)[128] .

If investment is flexible but capital or investment adjustment costs are large enough, then even highly persistent front loaded TFP shocks tend to generate countercyclical interest rates (e.g Smets and Wouters (2002) [167]). This occurs because an increase in $\frac{p_{k,t}}{p_{k,t+1}}$ reduces the rate of return on capital, pushing down real interest rates through arbitrage between different asset markets.

The model with fixed capital stock is a special case with infinite adjustment costs. In contrast to the extreme case of a fixed capital stock where hours of work do not respond to TFP shocks for preferences compatible with a BGP, large but finite capital adjustment costs can actually lead to a countercyclical labour input for sufficiently persistent TFP shocks. This occurs because a positive TFP shock encourages a shift in the composition of GDP towards consumption, which is associated with a stronger income effect on labour supply (and vice versa for a negative TFP shock).

To see this more formally, we return to a model with cobb douglas production and utility functions. The household's labour supply optimality condition is

$$\frac{\xi_n}{l_t} = \frac{w_t(1-\tau_{ns,t})}{c_t(1+\tau_{c,t})}.$$

Labour demand satisfies

$$\begin{aligned} \gamma\frac{y_t}{n_t} &= w_t(1+\tau_{nd,t}) \\ &\Longrightarrow \quad y_t = \frac{w_t(1+\tau_{nd,t})n_t}{\gamma}. \end{aligned}$$

Using $c_t = a_{c,t}y_t$ for $0 < a_{c,t} \leq 1$ implies

$$\begin{aligned} \frac{\xi_n}{1-n_t} &= \gamma\frac{w_t(1-\tau_{ns,t})}{a_{c,t}w_t(1+\tau_{nd,t})(1+\tau_{c,t})n_t}, \quad \Longleftrightarrow \\ \xi_n\frac{n_t}{1-n_t} &= \frac{\gamma(1-\tau_{ns,t})}{a_{c,t}(1+\tau_{nd,t})(1+\tau_{c,t})}. \end{aligned}$$

This shows that a shift towards higher consumption as a proportion of output (a higher $a_{c,t}$) reduces equilibrium labour supply. Thus, if investment is hard to adjust in the short run in response to higher TFP, households will spend most of the extra income through higher consumption and reduce their work hours. Empirically, investment is more sensitive to business cycles. This means that the consumption share of output $a_{c,t}$ is usually countercyclical. From this, we conclude that if TFP shocks are important, capital adjustment costs should be moderate. [29]

Note that these results rely critically on the TFP shock affecting the consumption goods and investment goods sectors symmetrically. Under BGP compatible utility functions, random walk TFP shocks that only affect the consumption goods sector would simply shift up consumption while leaving investment and hours unaffected (Basu et al (2009)). Meanwhile investment specific TFP shocks can easily generate negative comovement between consumption and investment, making them a poor candidate as the main source of business cycles.

[29] With different household types, exact aggregation into these conditions is unlikely except under the special conditions outlined in the section on household labour supply decisions. But the general result that higher investment adjustment costs reduce the procyclicality of labour hours in response to TFP shocks is likely to remain.

8.3.2 Discount factor, household external finance premium and news shocks

Consider a shock that increases the household discount factor, raising the demand for savings. Or consider an increase in the household external finance premium,due to financial market shocks or due to other shocks that reduce the income of credit constrained households. These are intratemporal shocks, affecting household optimal consumption/saving conditions.

1) The intratemporal shock reduces consumption in the current period for a given interest rate.

2) In a model with fixed capital, labour market equilibrium implies that output does not respond to this shock (as long as there is no relation to for example the efficiency of product markets search). Therefore, the real interest rate must drop significantly to counter the increase in desired savings and to maintain equilibrium in credit markets.

$$\boldsymbol{\Delta}\mathbf{k}_{t+j+1}=\mathbf{0}, \boldsymbol{\Delta}\mathbf{efp}_{h,t+j}>\mathbf{0}, \boldsymbol{\Delta}\tilde{\boldsymbol{\beta}}_{t+j+1}>\mathbf{0} \Rightarrow \boldsymbol{\Delta}\mathbf{c}^{d}_{t+j}<\mathbf{0} \Rightarrow \boldsymbol{\Delta}\mathbf{R}_{t+j}<\mathbf{0}$$
$$\Rightarrow \boldsymbol{\Delta}\mathbf{c}_{t+j}=\mathbf{0}, \boldsymbol{\Delta}\mathbf{y}_{t+j}=\mathbf{0}.$$

3) In a model with flexible capital, the rise in desired savings of households can be accommodated through higher investment. Consumption falls, which raises labour supply for a given wage.

4) The higher investment raises the capital stock for a few periods and increase labour productivity. As a result, as long as the slope of the aggregate labour demand function is lower than that of the aggregate labour supply function, the negative consumption preference shock actually increases output and we get a comovement problem: consumption falls while investment increases contrary to what we usually observe for business cycles.

5) Interest rates decline due to the increased supply of savings, but because of the ability of investment to adjust the decline in interest rates is typically much smaller than in models with a fixed capital stock.

Note how different this is from the popular Keynesian idea that higher desired savings can generate a recession in the short run. Here, consumption declines but investment increases. There is a shift in the composition of GDP, but output itself increases.

$$\Delta efp_{h,t+j} > 0, \Delta\tilde{\beta}_{t+j+1} > 0 \Rightarrow \Delta c^d_{t+j} < 0 \Rightarrow \Delta R_{t+j} < 0$$
$$\Rightarrow \Delta c_{t+j} < 0, \Delta k_{t+j+1} > 0$$
$$\Rightarrow \Delta n^s_{t+j} > 0 \Rightarrow \Delta w_{t+j} < 0 \Rightarrow \Delta n^d_{t+j} = \Delta n_{t+j} > 0$$
$$\Rightarrow \Delta y_{t+j} > 0.$$

Similar results occur with TFP news or growth rate shocks (Beaudry and Portier (2007) [23]). News shocks are exogenous changes in people's expectations about future economic conditions (e.g optimism about future technological developments, greater concern about higher future taxes etc...). Coupled with the possibility that agents in the economy overestimate the precision of the signals they acquire about future economic conditions, these shocks are perhaps the most straightforward way of formalising many people's intuition that changes in confidence or in beliefs about the future are one of the key drivers of business cycles (e.g excess optimism or pessimism). Formally, agents are too optimistic when they assign too much weight to an imprecise positive signal that they get about the future state of the economy. Reversing the sign of the signal formalises the notion of excessive pessimism.

Most researchers still work with a rational expectations equilibrium in which agents in the economy correctly perceive the true precision of the signals they get about the future. However, current solution methods for DSGE models can also handle the case in which agents are overconfident in the precision of the signals they get, in which case we can get business cycles generated by overoptimism and overpessimism.

A positive news shock generates a wealth effect on consumption similar to a positive consumption demand shock. Therefore, in many baseline business cycle models, news shocks cannot be a main factor in business cycles: if it was, it would cause a similar counterfactual negative comovement of consumption and investment and negative comovement of consumption and output.

There is a large literature on generating business cycles in response to news shocks in RBC models with capital. Viewed from a business cycle accounting perspective, these models create an endogenous relation between intertemporal consumption or news shocks and either labour wedges $\tau_{ns,t}$, $\tau_{nd,t}$ or the TFP wedge z_t.

Habit formation can lead to an increase in labour supply in response to positive news, because higher expected future consumption increases the marginal utility of wealth, raising the marginal benefit of working (Christiano et al (2007) [55]) . This allows output to expand such that both consumption and investment increase for a positive news shock.

Credit constraints on flexible production inputs can be relaxed if a positive news shock raises asset prices, and assets serve as collateral

(Inaba et al [132], Walentin (2009) [176]). **This encourages firms to increase output ahead of a positive news shock.** For this story to work, there must be a mechanism encouraging higher demand for the collateral assets in response to positive news, such as investment adjustment costs, or a high intertemporal elasticity of substitution. Alternatively, the value of the collateral must be strongly tied to the higher future profits of firms.

Convex adjustment costs (e.g bottlenecks caused by having to train too many employees or post too many vacancies simultaneously, or costs of increasing output too fast) can encourage a firm to increase its hiring and output in advance of the realisation of a positive news shock (Den Haan and Kaltenbrunner (2009) [106]). Countercyclical markups can also occasionally generate business cycles in response to news shocks, though without other mechanisms this may require an upward sloping labour demand curve (see Wang (2010) [178]).

One promising channel that is yet to be formally explored is the link between expectations shocks and search frictions in product markets, using for example the framework in Bai et al (2012) [17]. Suppose an increase in optimism stimulates consumer spending. This stimulates higher shopping effort by consumer which increases the probability for firms of finding buyers in product markets. As a result of the increased efficiency in the matching process, output can expand to meet the extra consumption demand.

RBC models can generate realistic business cycles in response to correlated new shocks, with less reliance on real frictions such as labour adjustment costs or habit formation. This is especially important given the limited empirical support at the micro level for the high levels of habit formation and labour adjustment costs that are used in some macroeconomic models.

For example consider the ARMA(1,8) shock process for some wedge τ_t,

$$\begin{aligned} \ln \tau_t - \ln \tau &= \rho(\ln \tau_{t-1} - \ln \tau) + \Sigma_{j=0}^{J} \theta^j \varepsilon_{t-j}, \\ 0 &< \rho < 1, 0 < \theta \leq 1. \end{aligned}$$

Suppose that $\rho = 0.95$ and $\theta = 0.75, J = 7$.This process has a hump shape with a peak impact τ_{t+8} at 8 quarters of around $2.8\varepsilon_t$. We can think of this process as representing the diffusion of transitory changes in sentiment or news through the economy. At period t, an optimism shock ε_t increases τ_t. In the next period, the measure of optimists dissipates by a depreciation factor ρ, but the new optimists from the previous period ε_t infect another $\theta\varepsilon_t$ agents with their optimism. At $t + 2$, the stock of optimists again depreciates at a rate of $1 - \rho$, but the new optimists from the previous period infect $\theta^2\varepsilon_t$ new agents with their optimism. In each period up to $t+J$ the stock of optimists equals to the sum of the remaining previous optimists ($\rho \ln \tau_{t+j-1}$) and the newly infected optimists $\theta^j\varepsilon_t$. The same pattern applies to a pessimism shock.

Processes such as this one allow the RBC model to generate news driven business cycles with much smaller real frictions (Leeper and Walker (2010)[138]). The key difference is that the news component in this process follows an instantaneous shock that raises firms' desired production immediately, so the dynamics become closer to those of a contemporaneous productivity or labour wedge shock. Christiano et al (2012) [56] find that correlated news about idiosyncratic risk can account for around 60% of GDP growth fluctuations in a medium sized New Keynesian DSGE model.

Is there any plausible microfoundation for the MA components in this process? Models with imperfect information, technological diffusion and learning can generate what looks like a gradual boom-bust in certain wedges (see for example Angeletos and Lao (2012) [11]). For example, a gradual spread of uncertainty through the financial system, followed by an eventual reduction in uncertainty lead to hump shaped dynamics in financing wedges. A gradual diffusion of higher uncertainty across firms in an environment where factor inputs cannot be fully adjusted to idiosyncratic shocks as in Arellano et al (2012) [12] can lead to the appearance of hump shaped TFP and labour wedges. To be more concrete, suppose each firm's labour demand follows

$$\gamma \frac{y_{i,t}}{n_{i,t}} = w_t(1+\tau_{nd,i,t}),$$

then aggregate labour demand is

$$n_t = \Sigma_i n_{i,t} = \frac{\gamma}{w_t}\Sigma_i \frac{y_{it}}{1+\tau_{nd,i,t}} = \Sigma_i \frac{y_{i,t}}{y_t}\frac{1}{1+\tau_{nd,i,t}}\frac{\gamma y_t}{w_t} = \frac{1}{1+\tau_{nd,t}}\frac{\gamma y_t}{w_t},$$

where the $\tau_{nd,i,t}$ is the firm specific wedge in labour demand caused by the extra uncertainty, and the average labour wedge is

$$1+\tau_{nd,t} = \frac{1}{\Sigma_i \frac{y_{i,t}}{y_t}\frac{1}{1+\tau_{nd,i,t}}} \simeq \frac{1}{\Sigma_i \frac{y_i}{y}\frac{1}{1+\tau_{nd,i,t}}},$$

with the last approximation using the BGP output shares of different firms. Taking a first order Taylor approximation in percentage changes twice,

$$\tau_{nd,t} - \tau_{nd} \simeq \Sigma_i \frac{y_i}{y}(\tau_{nd,i,t} - \tau_{nd,i}),$$

so that the change in the aggregate labour demand wedge is approximately equal to a weighted average of firm specific wedges.

Angeletos and Lao (2012) [11] show how gradual diffusion across firms of pessimistic expectations can lead to to a gradual increase in the proportion of firm's with a high labour demand wedge, and an increase in the average labour demand wedge $\tau_{nd,t}$ (with the opposite dynamics when expectations become more optimistic).Eventually the wave of increased pessimism across firms dies out, generating a hump shaped aggregate labour wedge shock process. The next section explores these "sentiment" shocks in greater detail.

8.3.3 Imperfect information and sentiment shocks

News shocks are usually modelled as being about changes several periods into the future. But one can also imagine intratemporal shocks to expectations. These intratemporal expectations shocks or sentiment shocks provide a formalisation of the Keynesian idea of animal spirits driven business cycles (Angeletos and Lao (2012) [11] , Benhabib et al (2012) [26], Angeletos et al. (2014) [10]).

In these models, a recession can be explained by correlated (across firms), persistent, pessimistic signals received by firms about demand for their products that cause them to reduce output (with the reverse optimism leading to output booms). The pessimistic signals are compatible with RE: we cannot just drop in a smart economist who would eliminate the recession by showing decision makers that their pessimism is unjustified, because the pessimism is in fact justified. [30] Sentiment shocks also imply that the economy can fluctuate a lot in response to hard to predict and quantify changes in expectations. As a result, they can rationalise a situation where even if we were to understand the structure of the economy reasonably well, accurate forecasting in the traditional sense of providing a single number minimising forecast mean squared error would be difficult.

The main features of these models are:
1) heterogeneity between firms,
2) imperfect information,
3) inflexible adjustment of production input purchases.

Heterogeneity between firms could be due to imperfect substitutability between different products, differences in firm productivity or due to randomeness in the matching process between buyers and sellers as in Angeletos and Lao (2012) [11]. Inflexible adjustment could come in the form of partial irreversibility in production decisions, or in the form of other production input adjustment costs.

While we have emphasized 3 elements, there is a sense in which *2) imperfect information* only matters because of *3) inflexible adjustment*: if firms could

[30] Black (1986) [32] suggests a unified theory of financial markets and business cycles revolving around imperfect information and misperceptions, which he labels "noise".

always perfectly adjust their decision quickly to any available information, then imperfect information would be irrelevant. Note the similarity of these elements to those in the earlier model of the labour wedge with heterogeneous firms and uncertainty shocks in Arellano et al (2012) [12] and Bloom et al (2012) [34]. Heterogenity across firms is not strictly necessary in these models (see Benhabib et al(2013) for a model with imperfect information about an aggregate demand shock), but by allowing for idiosyncratich shocks across the firms it greatly expands the scope for realistic levels of imperfect information.

Suppose there are many firms each affected by its own idiosycratic profitability shock. These profitability shocks could be to TFP, relative demand or cost of production inputs shocks. Firms get imperfect signals about their idiosyncratic shock, based on which they have to make hiring and other production input purchase decisions. These decisions are (partially) irreversible with respect the idiosyncratic shock (otherwise if firms could flexibly readjust their decision when information is revealed, the imperfect information would be irrelevant).

A typical input demand function under perfect foresight is

$$x_t^i = f_x(p_{x,t}, z_t^i),$$

where $p_{x,t}$ is the price of input x relative to firm $i's$ output and z_t is a vector of other demand or cost variables, where

$$\frac{\partial x_t^i}{\partial p_{x,t}} < 0, \frac{\partial x_t^i}{\partial z_{j,t}^i} > 0$$

for each element $z_{j,t}^i$ of z_t^i. Now consider the situation where demand for x_t^i must be set under imperfect information, before uncertainty is fully resolved. In Angeletos and Lao (2012) [11] or Benhabib et al (2012) [26] labour demand is simply predetermined inside the period. Perhaps more realistically, there are labour adjustment costs that make the labour demand decision dynamic: the firm decides how many workers to hire or fire in the current period, taking into account that its decision will also affect costs of hiring and firing in the uncertain future (e.g Bloom et al (2012) [34]).

In general solving for the full effect of this uncertainty is nontrivial (see for example Arellano et al (2012) [12]). But if we are willing to make a certainty equivalent approximation, or take a linear approximation of the demand function, we have

$$\begin{aligned} x_t^i &= f_x(E_{i,t}p_{x,t}, E_{i,t}z_t^i), \\ \frac{\partial x_t^i}{\partial E_{i,t}p_{x,t}} &< 0, \frac{\partial x_t^i}{\partial E_{i,t}z_{j,t}^i} > 0, \end{aligned}$$

where $E_{i,t}(.)$ is firm $i's$ expectation about the prices it will be facing and other demand factors during period t at the beginning of period,

and a subset of these variables $(p_{x2,t}, z^i_{2,t})$ is typically known in advance. Now imagine firms get a negative signal $s_{i,t}$ about the cost of their production inputs p^i_x, t or about demand for their product, their TFP or other factors in z^i_t, with a lower $s_{i,t}$ indicating a more pessimistic signal. We can solve for a a reduced form demand function

$$x^i_t = f_x(E_{i,t}p_{x,t}, E_{i,t}z^i_t) = x^i_t = f_x(s_{i,t}, p_{x2,t}, z^i_{2,t}).$$

Under what conditions do production decisions respond to the signal with $\frac{\partial x^i_t}{\partial s_{i,t}} > 0$? If we allow for arbitary expectation formation mechanism, we can clearly find some examples in which the negative signal worsens firm expectations and causes them to reduce demand for inputs and production. Firms reduce labour and intermediate inputs demand. Lower wages lead to a reduction in labour supply so that overall employment declines. If the negative signal is persistent then the higher pessimism about future profits and the reduction in future labour and intermediate inputs demand also reduce the expected return to capital. As a result, investment also falls.

While, there are a lot of concerns that deviations from RE can justify any result, there are several behavioural economics models of under-reaction and over-reaction to news that can impose some discipline on irrational expetations (e.g due to inattention to certain signals below a certain threshold, followed by excessive extrapolation from short-term trends once agents become aware of news- for example Gabaix (2014)[88], Shleifer et al (2015)).

Can this recession scenario due to negative signals be sustained under RE? Benhabib et al (2012) [26] and Angeletos and Lao (2012) [11] show that the answer is yes under certain conditions. A firm getting an imperfect negative signal about the profitability of its product will have more pessimistic expectations under RE. As a result it will scale down its production by hiring fewer workers and purchasing fewer intermediate goods as long as it perceives the signal to contain some true information. In this way correlated shocks to firms expectations about their TFP or their terms of trade (the price of their output relative to the cost of inputs) can cause business cycles.

In a loglinear approximation, we can write aggregate production in these economies as

$$\ln X_t = A + B \ln Z_t + C \ln S_t,$$

,where Z_t is a vector of fundamental state variables and S_t is an aggregate measure of the average signal perceived by firms in the economy.

A higher average signal represents greater business optimism, while a more negative S_t represents greater pessimism. A negative aggregate sentiment shock to S_t can generate persistent recessions, just like a TFP or labour wedge shock. From a business cycle accounting perspective these sentiment shocks map into production factor demand wedges. For example, Angeletos and Lao (2012) [11] show how they can help explain the labour wedge. They also provide an example in which the spread of optimistic signals through different firms and the gradual revelation that the signal is false as firms communicate with each other leads to a hump shaped sentiment shock, similar to the ARMA(1,q) process in the section on news shocks.

There are some limits to the noisy signals explanation of business cycles under RE:

1) A general result in these models is that the aggregate volatility resulting from noisy signals is bounded from above by the volatility of the source of the fundamental uncertainty on which we get the signal (Angeletos and Lao (2012) [11]). So for these shocks to generate significant aggregate fluctuations we need a source of large fundamental uncertainty. This criterion can be easily satisfied by idiosyncratic firm level risk, which is typically much larger than aggregate risk.

2) There is a trade-off in increasing the volatility of signals relative to that of the underlying fundamental source of uncertainty. Higher volatility of signals on 1 hand increases fluctuations in the expectation of the variables affecting demand such as $E_{i,t}p_{x,t}$ or $E_{i,t}z^i_{j,t}$. On the other hand under RE a higher signals volatility for a given fundament volatility reduces the information content of the signals and encourages agents to filter out more of the signal as noise.

The best case for sentiment shocks driven business cycles under RE is based on finding volatile signals with high signal to noise ratios, because the signals themselves are about variables with high fundamental volatility. Alternatively, we need to appeal to significant deviations from RE.

Angeletos et al (2014) [10] estimate several variants of a DSGE model in which sentiment shocks compete with more traditional news shocks (in which the news on e.g productivity is actually realised), productivity shocks and other frequently used business cycle shocks. They find that in the US their sentiment shocks can account for 40-60% of the variance of output and almost 70% of the variance of hours at business cycle frequencies. Furthermore, a simple RBC model with transitory sentiment shocks and permanent productivity shocks can successfuly reproduces most of the business cycle moments of GDP components, hours and labour productivity. TFP shocks help to match the positive correlation between labour productivity and output. Meanwhile, the sentiment shocks help match the volatility of employment that is usual much too low in RBC models driven by TFP shocks. The sentiment shock extracted in their estimation is also highly correlated with changes in consumer and business confidence indices.

Sentiment shocks can provide a microfoundation for financial shocks. To do this, we can apply Angeletos and Lao's (2012) [11] model of a production network in the nonfinancial sector to the financial system. Imagine a network of financial intermediaries (FI). Each period some FI's will need to borrow funds from other FI's, while some FI's will have excess funds to lend. Trade is decentralized with asymmetric information about the financial health of the partner FI's. Lending decisions to FI's and non financial firms or households have to be made before fully observing the state of trade partners in financial markets.

Again we have the 3 key elements of *1) heterogeneity between financial intermediaries,2) imperfect information and 3) inflexible adjustment of borrowing and lending decisions to new information.* In this setup a negative signal about the financial health and lending terms of one's partner causes FI's to reduce borrowing and lending in financial markets. This reduction in trade volume is accompanied by tighter collateral requirements, or the outright refusal to trade some types of financial contracts. This lower trade volume in financial markets and greater difficulty in obtaining funds then spill over into tighter credit constraints on loans to households and non financial firms.

Finally, we can imagine models where there are imperfect signals on the amount of idiosyncratic risk facing firms, e.g on the standard deviation or skewness of idiosyncratic shocks. Suppose each firm faces a stochastic volatility of its productivity or relative demand for its product. Idiosycratic volatility shocks wash out in the aggregate, but not from the perspective of an individual firm. The firm gets an imperfect signal about the volatility it will face, and it has to make production decisions beforehand. In the presence of non convex adjustment costs as in Bloom et al (2012)[34], risk averse managers facing undiversified business risk as in Angeletos (2007) [8] or nontrivial default costs as in Arellano et a (2012)[12], a negative signal of higher volatility reduces firms' willingness to hire, invest and pay for intermediate production inputs.

Imperfect signals on idiosyncratic risk are likely to reproduce many of the effects on labour demand and production of actual changes in idiosyncratic risk (as in the model of Arellano et al (2012) discussed earlier). Unlike the realised uncertainty shocks in Arellano et al (2012) [12] or Bloom et al (2012)[34], imperfect signals about uncertainty are less dependent on the actual amount of change in the amount of idiosyncratic risk. This opens up the route for a much larger role for shifts in uncertainty in generating business cycles, in the form of sentiment shocks.

8.3.4 Investment shocks

Investment shocks can capture various financial frictions that are biased towards investment. These can include loans that are specifically for capital expenditure, countercyclical uninsured business risk (as in Christiano et al (2012) [56]) or interaction between higher riskiness of investment in a recession and fixed adjustment costs (e.g Bloom and Jaimovich (2012) [34]). They also represent shifts in the riskiness of investment or misperceptions about the returns to investment that are not captured by the (usually misspecified) endogenous mechanisms in the model or by RE. We model this shock in our setup by changes in the investment tax $\tau_{k,t}$ or in the relative discount factor of the firm $v_{f,t}$ (See Buera and Moll (2012) [35] for a credit shock that can be mapped into a combination of a change in the relative discount factor of the firm and a TFP news shock).

Investment shocks often lead to the opposite comovement problem of discount factor shocks, with procyclical investment and countercyclical consumption (see for example Greenwood et al (2000)[101] or in Liu et al's (2011) [139] or Khan and Thomas (2013) [125] models of credit shocks, or Bloom et al's (2012) model of uncertainty shocks in DSGE [34]).

A positive investment shock is associated with a lower investment wedge $\tau_{k,t}$ or a lower firm relative discount rate $v_{f,t}/v_{f,t+1}$.

1) The higher return to capital investment in response to a positive shock increases saving and investment.

2) Absent any shift in aggregate TFP or the labour demand wedge, output in the short term can only increase if labour supply increases. Without any other labour supply shock, this can only happen if consumption declines initially. In the case of a positive investment shock, the higher return to saving and investing leads to lower consumption and higher labour supply.

3) This is partially mitigated by factors that act like endogenous aggregate TFP shocks in response to higher marginal efficiency of investment, such as variable capital utilisation rates.

$$
\begin{aligned}
\Delta\tau_{k,t+j} &< 0, \Delta n_{d,t+j} = 0 \Rightarrow \Delta k_{t+j+1} > 0, \Delta c_{t+j} < 0, \\
&\Rightarrow \Delta n_{s,t+j} > 0 \Rightarrow \Delta w_{t+j} < 0 \Rightarrow \Delta n^{d}_{t+j} = \Delta n_{t+j} > 0 \\
&\Rightarrow \Delta y_{t+j} > 0.
\end{aligned}
$$

The difficulty in explaining comovement of consumption and investment implies that a theory of business cycles that focuses on

investment shocks must also simultaneously generate aggregate TFP or labour wedge shocks to reproduce the comovement of GDP components in the data.Otherwise, in the standard RBC model an investment shock is best seen as complementary to TFP or labour wedge shocks, e.g as part of a more fundamental financial or confidence shock that simultaneously affects labour and investment demand.

Capital utilisation rates increase in response to a positive investment shock if the cost of higher utilisation is greater capital depreciation, and if the shock increases the return to capital (Greenwood et al (2000) [101]). For this we need the direct effect of a lower $\tau_{k,t}$ on the return on capital to be stronger than the indirect effect of a higher price of capital $p_{k,t}$ due to higher investment demand. A more realistic and potentially more important mechanism reducing aggregate TFP in the presence of negative investment shocks is the cost of adjusting the composition of production between the investment goods sector and other sectors.

The basic RBC model with constant returns to scale production functions and without intersectoral adjustment costs assumes that it is easy to reallocate production to other sectors if investing in capital becomes more expensive. But in an economy with matching frictions, it may take some time for the new buyers and sellers in the higher TFP sectors to form new longer term relations and adjust their search strategy and effort to the new sectors. Or it may take time for the for firms in the higher TFP sector to expand their customer base (e.g Gourio and Rudanko (2013)[99]). While this adjustment occurs, matching in the economy is less efficient (the simplest way to capture this is by assuming decreasing returns to scale in sectoral production functions). Equivalently, there is a higher degree of mismatch between buyers and sellers (see for example Black (1986) [32]). Lower matching efficiency reduces the profitability of production for a given quantity of labour and capital, similarly to an aggregate TFP shock. [31]

The impact of investment shocks on real interest rates is more ambiguous. On one hand, a lower $\tau_{k,t}$ increases the price of capital by raising investment demand. This tends to reduce interest rates. On the other hand, the direct effect of lower $\tau_{k,t}$ is to increase the return on capital, raising interest rates. The second effect dominates initially if the higher return to capital is associated with a fall in consumption.

[31] We can also get comovement of investment and consumption for marginal efficiency of investment shocks in the flexible price model without variable capital utilisation due to a composition effect when employed households consume more than the non employed as in Euseppi Preston (2009) [77]. But with only this extra feature we would still predict negative comovement of aggregate investment with consumption of employed households. Finally, if higher desired investment stimulates search efforts by firms making new investments in the capital goods market, this can raise the efficiency of the production process through the distribution and matching function g_t.

8.4 Adding Durables: home production capital

Until now we have only modelled non durable consumption goods such as services or food. Durable consumption goods such as housing or consumer electric goods are another form of saving available to households in the economy. By providing utility over many periods, they allow households to smooth consumption across time similarly to financial assets. They can also be seen as a form of capital used in non-market production, also known as home production. From that perspective, their behaviour may be closer to that of business investment than to that of non-durable consumption. The size of durable goods investment (including housing) can be larger than that of business investment (see Gomme and Rupert (2006) [97] for the US). Durables investment behaviour has significant differences from that of nondurables consumption over the business cycle. For example, it is much more volatile. Durables, particularly in the form of housing play an important role as borrowing collateral for households.

Adding durables to the analysis as another form of household saving can have major implications for our understanding of comovement patterns between consumption and investment and of the cyclicality of wages and labour supply. For all these reasons, business cycle models with explicit treatment of durables have become increasingly common (e.g, Davis and Heathcote (2005) [64], Campbell and Hercowitz (2006) [41] and Iacoviello and Neri (2009) [112], Favilukis et al (2011) [82], Iacoviello and Pavan (2013) [113] .

8.4.1 The household's decision problem with durable goods

Households in period t choose the durables stock D_{t+1} that they will hold at the beginning of period $t+1$. Unlike business capital, we assume that durables D_{t+1} already yield utility in period t when they are chosen. This is common in business cycle models, and it is reasonable for most durable goods with the typical model period length of a quarter. The representative household problem becomes

$$V_0^h = \max_{\{c_t, n_t^s, D_{t+1}, b_{t+1}\}_{t=0}^{\infty}} E_0 \Sigma_{t=0}^{\infty} \beta_t u(c_t, D_{t+1}, l_t)$$

subject to a sequence of constraints

$$c_t(1+\tau_{c,t}) + p_{D,t}(1+\tau_{D,t})\left[D_{t+1} - (1-\delta_D)D_t\right] + b_{t+1} \leq b_t R_t e f p_{h,t} + w_t(1-\tau_{nst})n_{s,t} + d_t^h + T_t,\ (\lambda_t),$$

$$n_{s,t} = 1 - l_t,$$

$$u_c > 0, u_{cc} < 0, u_D > 0, u_{DD} < 0, u_l > 0, u_{ll} < 0, u_{c,l} \leq 0.$$

$\tau_{D,t}$ **is a wedge affecting durable purchases, for example due financial frictions that are specifically affect buying durable goods.** $p_{D,t}$ **is the relative price of durables and** d_t^h **are total household dividends from firms including profits of installed capital and durables producers.**

To complete the model, we also have a representative durables producer owned by households. It solves

$$\max_{D_t} \pi_{D,t} = \max_{D_t} p_{D,t} x_t^D - I_t^D.$$

$$\begin{aligned} x_t^D &= D_{t+1} - (1-\delta_D) D_t = a_t^D \tilde{G}(\frac{I_t^D}{D_t}) D_t \\ , \tilde{G}'(.) &> 0,\ \tilde{G}''(.) \leq 0. \end{aligned}$$

As for capital, a_t^D represents durables investment specific technology shocks, and a concave $G()$ represents capital adjustment costs. This defines a supply curve for capital. The resource constraint becomes

$$c_t + I_t + I_t^D = y_t.$$

The households' optimality conditions for risk free bonds, non durable consumption and labour supply have the same form as before. We now add an optimality condition for durables:

$$\mathbf{p}_{D,t}(\mathbf{1} + \boldsymbol{\tau}_{D,t}) = \frac{u_{D_{t+1}}}{\lambda_t} + \mathbf{E}_t \frac{\beta_{t+1}}{\beta_t} \frac{\lambda_{t+1}}{\lambda_t} \mathbf{p}_{D,t+1}(\mathbf{1} + \boldsymbol{\tau}_{D,t+1})(\mathbf{1} - \boldsymbol{\delta}_D).$$

Using the bond first order condition and a certainty equivalence

approximation we get

$$p_{D,t}(1 + \tau_{D,t}) \simeq \frac{u_{D_{t+1}}}{\lambda_t} + \frac{E_t p_{D,t+1}(1 + \tau_{D,t+1})(1 - \delta_D)}{R_{t+1} E_t e f p_{h,t+1}}.$$

These equations have some similarity to those we got earlier for business capital. On the left there is the cost an extra unit of durables. on the right is the marginal return to holding durables composed of an immediate gain from holding the durable and from the future resale value. However, there are some important differences. First because durables contribute to non market activity, the marginal product of capital is replaced by the marginal utility of durables. Second, because durables are owned by households, the effective discount factor for durables is the same as that for nondurable consumption. In contrast, we argued that differences between the firm owners or managers and the households' discount factors could play an important role in business investment decisions.

The comparative dynamics of the last equation are relatively straightforward, at least under common utility functions for which $\frac{u_{D_{t+1}}}{\lambda_t}$ is a

one to one function of $\frac{c_t}{D_{t+1}}$. A higher price of durables $p_{D,t}$ or a higher durables wedge $\tau_{D,t}$ relative to next period reduce durables investment relative to non durables consumption. Higher interest rates or a higher external finance premium $efp_{h,t+1}$ reduce durables investment relative to non durables consumption. The conclusion that durables investment is more sensitive to financing frictions is common to many models. This suggests that credit constraints may be an important explanation for the much greater volatility of durables investment relative to consumption over the business cycle. We can almost verbatim here our comments on business investment with nonconvex adjustment costs to the relation between discrete durables investment decisions at the household level and continuous aggregate durables.

The representative household's durables investment optimality condition can also be interpreted as the result of aggregating individual household discrete choices of a single unit of the durable good. Households purchase a new TV or a new house as long as the marginal benefit on the right hand side of our optimality condition exceeds the marginal cost on the left hand side. Under some conditions (e.g Caplin and Leahy (2002)[42]), aggregating the discrete choices of many households leads to the same aggregate durables investment function as that of the representative household in our model. The most important necessary condition for this exact equivalence is to have complete financial markets at the household level, which equalises the shadow value of wealth λ_t across households. A similar result holds for continouous household choices subject to non convex adjustment costs (for example a fixed cost of searching for and moving to a new house).

With incomplete financial markets and uninsured idiosyncratic income risk, the equivalence breaks down (Iacoviello and Pavan (2013) [113]) because of the interaction between discretenss or nonconvex adjustment costs and the distribution of households' shadow values of wealth $\lambda_{i,t}$. However the representative household approximation is still a useful simplification once we allow for stochastic wedges capturing some of the effects of financial market incompleteness in the durables investment optimality condition.

8.4.2 Business cycle dynamics with durable goods

Models without significant movements in the external finance premium $efp_{h,t+1}$ often generate a counterfactual negative comovement between durables and consumption or business investment. Consider a front loaded positive TFP shock. With moderate capital adjustment costs, this tends increase interest rates. From the first order condition for durables investment, a higher interest rate reduces the durables stock to consumption ratio. Since the increase in consumption is usually not that large in comparison to the durables stock,

this usually requires a decline in the stock of durables. As a result durables investment declines in the short run. This is a problem, since empirically durables investment are strongly procyclical and the real risk free rate R_{t+1} is also procyclical. One way to deal with this comovement problem is to assume high capital adjustment costs, leading to countercyclical interest rates. But this leads to a volatility of investment relative to output that is too low in comparison to the data.

Mechanisms that increase the response of output and consumption to TFP shocks or shifts in the relative tax on durables $\tau_{D,t}$ or in the external finance premium $efp_{h,t+1}$ can help generate positive comovement between durables investment and other GDP components. We will discuss some of these mechanisms below. Adding household level financial frictions can be particularly helpful. **Countercyclical external finance premia $efp_{h,t+1}$ can reduce the procyclicality of the effective interest rate** $R_{t+1}efp_{h,t+1}$, or even make it countercyclical. On top of that, other financing frictions that are specific to purchasing durables lead to a **countercyclical** $\tau_{D,t}$. **Due to the sensitivity of aggregate durables investment to interest rates and wedges, these mechanisms can be enough to generate strongly procyclical investment in durables.**

The effect of durables investment shocks such as $\tau_{D,t}$ has some similarity to that of investment in business capital shocks. A decline in $\tau_{D,t}$ encourages durables spending, but may discourage spending on non durables consumption by increasing the return on saving. To the degree that similar changes in financing frictions or in investment specific technology affect durables and business capital, the introduction of durables allows us to generate positive comovement between firm investment and consumption including durables more easily [32]. Positive comovement is also more likely if nondurables consumption is complementary with the durables stock. This would tend to encourage a positive link between consumption of durables and nondurables goods.

In general equilibrium, a simultaneous increase in durables and nondurables spending is incompatible with higher business investment in the short-run, unless it raises the labour input in production. Positive comovement with business

[32] Including durables in the consumption variable is common in many models and empirical exercises (e.g Smets and Wouters (2007) [168]). This is only a good approximation if durables depreciation is implausibly high. An intermediate aproach in models without durables is to add the implicit rental value of durables to the consumption measure in estimation or forecasting. Finally, some researchers treat durables investment as part of business investment (e.g Justiniano et al (2011) [120]). In models with firm and household specific credit constraints, this can significantly bias the assessment of the role of financial frictions. The last remaining approach is to redefine GDP to exclude durables and then use an auxiliary reduced form model to forecast durables when necessary. None of these methods is perfect. Given the advances in the computational ability to handle larger state spaces, and the importance of durables investment in business cycles, explicit inclusion of durables is probably the best option for medium-large scale policy models.

investment is more likely to happen if durables spending and hours of work are complements in the utility function or if there is correlation between $\tau_{D,t}$ and the labour supply or nondurable consumption wedges $\tau_{ns,t}, \tau_{c,t}$ (for example, a general improvement in credit conditions in an environment where part of households' labour supply is to pay downpayments on mortgages or durable goods loans (e.g Tomura(2009) [174]).

Allowing for incomplete financial markets and non-convex adjustment costs can also generate countercyclical movements in the durables purchase wedge $\tau_{D,t}$ through several channels. These include countercyclical durables price idiosyncratic risk and interaction between uninsured income risk and durable adjustment costs (Iacoviello and Pavan (2013) [113]). The first channel is clearly most relevant for housing. Higher housing price risk makes housing less attractive as an investment in a recession if it is not fully diversified. One suspects that this risk is countercyclical, like idiosyncratic wage or business income risks, though the empirical evidence on this point has been lacking for housing.

The interaction between uninsured idiosyncratic income risk and non convex durables adjustment costs is probably more robust. In a recession fixed costs of adjustment, such as difficulty in reselling durables or the existence of a minimum scale for durable purchases, together with the greater importance of uninsured income risk when wealth is lower discourage spending on durables. While this is true for any negative shock, this effect is stronger in the empirically plausible case of countercyclical idiosyncratic income risk. In that case the higher risk directly reduces desired durables investment.For example, a risk averse household faced with higher unemployment risk is less willing to invest in a large house when scaling back the investment in case of unemployment is costly.

The adjustment costs also discourage some existing homeowners from switching back to a smaller residence in a recession. But even though fewer households adjust in the presence of fixed costs of adjustment, the overall cross-section distribution of desired adjustment without fixed costs shifts to lower durables investment. As a result the proportion of adjusting households who want to lower their durables stock increases, leading to a reduction in durables investment. In reverse, a reduction in idiosyncratic household income risk increases the desired stock of durables as households are less concerned about having to downsize and pay a significant adjustment cost in case of a negative shock.

The combination of idiosyncratic income risk with non convex durables adjustment costs attenuates the response of durables investment to shocks in recessions. This mechanism applies mostly to housing, which is typically estimated to have a fixed adjustment cost equal to around 5% of its value. The fixed cost of moving to a new house is more painful to pay when wealth is low in a recession due to diminishing marginal utility of wealth. As a result, households are more cautious in changing their housing stock in a recession.

Recall that the main factor that complicates explaining the comovement between durables and business capital investment: the procyclicality of real interest rates for most shocks. The combination of fixed adjustment costs with idiosyncratic income risk makes households more reluctant to respond to lower interest rates in a recession. In reverse, a boom when wealth and income increase is a good time to spend the fixed cost of adjusting housing, even if interest rates are higher. Therefore, the countercyclical effect of interest rates on durables investment is weaker with realistic adjustment costs and income risk. This makes durables investment more procyclical when we combine uninsured income risk and non convex adjustment costs (Iacoviello and Pavan (2013)[113]).

Accounting for durables in the economy gives precautionary savings effects on consumption a more prominent role. Holdings of durables can help explain a higher level of financial wealth inequality (see Kaplan and Violante (2014) [122], Gruber and Martin (2003) [103], Diaz and Luengo-Prado (2010) [66]). Diminishing marginal utility of durables implies that wealthier households have a lower weight of durables in their overall wealth portfolio. As result, the distribution of durables is more concentrated, and there is more inequality in the distribution of financial assets for a given overall wealth inequality level.

In an economy with uninsured idiosyncratic household risk, higher financial wealth inequality increases the size of precautionary savings changes in response to shocks, and raises the average MPC out of current (or more recent periods) income (see for example Carroll (2000) [45]). Furthermore, because of the adjustment cost durables cannot be used as effectively to smooth consumption in response to transitory shocks (Martin and Gruber (2003), Kaplan and Violante (2014) [122]). In combination with the higher financial wealth inequality in the economy with durables, this tends to make non durables consumption more procyclical.

Finally, including durable goods can also improve our ability to match employment changes over the business cycle. Financial market shocks affecting $\tau_{D,t}$ (or for that matter $\tau_{k,t}$, and $\tau_{c,t}$ as well) can help explain higher employment volatility and lower wage volatility. An rise in these wedges due to higher financing constraints can significantly reduce households' work incentives. From the household first order condition for durables investment, an increase in $\tau_{D,t}$ or in $epf_{h,t+1}$ can lower the marginal value of income λ_t by causing substitution from durables to nondurables expenditure. This reduces the marginal value of wage income, leading to lower labour supply.

In contrast, in the model without flexible durables investment a higher $efp_{h,t+1}$ would always tend to increase λ_t and labour supply. In the empirically plausible case of complementarity between durables and nondurables (see the evidence and discussion in Flavin and Yamashita (2008) [85]), lower durables consumption due to higher $efp_{h,t+1}$ or $\tau_{D,t}$ directly reduces the shadow cost λ_t and labour supply. From the perspective of models that omit durable goods these mechanisms act like negative labour supply shocks in a recession.

8.4.3 The savers and borrowers model with durable goods

This is a standard way of modeling borrowing frictions in DSGE (e.g Iacoviello (2005) [111], Iacoviello and Neri (2009) [112]). It is an extension of our earlier model of a representative borrower and a representative saver with an exogenous debt limit, and provides an easy to implement micro foundation for the effects of the external finance premium on durables dynamics discussed earlier. Now borrowers also have access to durables and they can use some of them as collateral for borrowing. Even if durables cannot be used as collateral, durables still help the borrowing household smooth its consumption relative to the case without durables. For example a negative income shock can be partially smoothed by reducing durables investment. A positive income shock will encourage saving in the form of a higher durables stock. Of course, because the durable good yields utility immediate, a reduction in durable goods will be felt by households quite differently than a reduction in financial net worth.

The saver household has the same budget constraint and optimality conditions as before, now indexed by $i = sa$. The representative borrower solves

$$V_0^{bo} = \max_{\{c_{bo,t}, n_{bo,t}, d_{bo,t+1}, b_{bo,t+1}\}} E_0 \Sigma_{t=0}^{\infty} \beta_{bo,t} u(c_{bo,t}, D_{bo,t+1}, l_{bo,t})$$

subject to a sequence of constraints

$$c_{bo,t}(1+\tau_{c,t}) + p_{D,t}(1+\tau_{D,t})\left[D_{bo,t+1} - (1-\delta_D)D_{bo,t}\right] + b_{bo,t+1} \leq b_{bo,t} R_t efp_{bo.h,t}$$
$$+ w_t(1-\tau_{nst})n_{bo,t} + T_t,\ (\lambda_{bo,t}),$$
$$-b_{bo,t+1} \leq \bar{b}_{t+1} + m_{bo,t} p_{D,t} D_{bo,t+1},$$
$$(\mu_{t+1}),\ \bar{b}_{t+1|} \geq 0,\ 0 \leq m_{bo,t} \leq 1.$$
$$n_{bo,t} = 1 - l_{bo,t},$$
$$u_c > 0, u_{cc} < 0, u_l > 0, u_{ll} < 0, u_{c,l} \leq 0.$$

The optimality conditions for bonds and durables change to

$$\begin{aligned}
b_{bo,t+1} \quad &: \quad \lambda_{bo,t} = E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} ef p_{bo,h,t+1} R_{t+1} \lambda_{bo,t+1} + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \mu_{t+1} \\
&\equiv \quad E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} ef \bar{p}_{bo,h,t+1} R_{t+1} \lambda_{bo,t+1} \\
p_{D,t}(1+\tau_{D,t}) &= \frac{u_{D_{bo,t+1}}}{\lambda_{bo,t}} + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \frac{\lambda_{bo,t+1}}{\lambda_{bo,t}} p_{D,t+1}(1+\tau_{D,t+1})(1-\delta_D) + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} m_{bo,t} p_{D,t} \mu_{t+1}.
\end{aligned}$$

Note the extra term on the right hand side of the durables optimality condition. This term reflects the additional marginal benefit of the durable good as collateral.

Market clearing conditions become

$$\begin{aligned}
\Sigma_i \theta_i c_{it} &= c_t, \\
\Sigma_i \theta_i D_{it} &= D_t, \\
\Sigma_i \theta_i n_{i,t} &= n_{st}, \\
\Sigma_i \theta_i b_{i,t} &= 0.
\end{aligned}$$

.

8.4.4 Durable goods demand with costly consumer search

Following the ideas of Rios-Rull et al (2012) [17] about the effect of search frictions on business cycles, we introduce consumer search frictions in the market for durable goods. Housing and other durables are notorious for requiring significant consumer research and shopping effort before making a purchase. In comparison, an increase in non durable goods and services purchases in a boom does not necessarily require more shopping effort: for example in some case it is enough to buy higher quality goods and services during regular shopping time or visits to restaurants and other entertainment venues (that count as leisure time). We stick to a representative household approach in which purchasing x_t^D units of new durable goods requires a shopping effort S_t^D, satisfying

$$x_t^D \le \phi_t^D S_t^D, \ (\mu_t^S).$$

The new durables finding rate ϕ_t^D is taken as given by households. It is determined in equilibrium by the joint search effort of households and firms. We modify the household's utility function to

$$\begin{aligned}
& u(c_t, D_{t+1}, l_t) - v(S_t^D), \\
v_{S^D} &> 0, v_{S^D S^D} \ge 0.
\end{aligned}$$

$v(S^D)$ is search effort. We assume again that the marginal cost of search is independent of the other variables affecting household preferences. The first order condition for durables becomes

$$p_{D,t}(1+\tau_{D,t})+\frac{\mu_t^S}{\lambda_t}=\frac{u_{D_{t+1}}}{\lambda_t}+E_t\frac{\beta_{t+1}}{\beta_t}\frac{\lambda_{t+1}}{\lambda_t}\left[p_{D,t+1}(1+\tau_{D,t+1})+\frac{\mu_{t+1}^S}{\lambda_{t+1}}\right](1-\delta_D).$$

We also add a first order condition for search effort

$$v_{S_t^D}=\mu_t^S\phi_t^D.$$

This allows us to rewrite the first order condition for durables as

$$\begin{aligned}
p_{D,t}(1+\tau_{D,t})+\frac{v_{S_t^D}}{\phi_t^D\lambda_t} &= \frac{u_{D_{t+1}}}{\lambda_t}+E_t\frac{\beta_{t+1}}{\beta_t}\frac{\lambda_{t+1}}{\lambda_t}\left[p_{D,t+1}(1+\tau_{D,t+1})+\frac{v_{S_{t+1}^D}}{\phi_{t+1}^D\lambda_{t+1}}\right](1-\delta_D),\Longleftrightarrow \\
p_{D,t}(1+\hat{\tau}_{D,t}) &= \frac{u_{D_{t+1}}}{\lambda_t}+E_t\frac{\beta_{t+1}}{\beta_t}\frac{\lambda_{t+1}}{\lambda_t}\left[p_{D,t+1}(1+\hat{\tau}_{D,t+1})\right](1-\delta_D),\text{ where} \\
\hat{\tau}_{D,t} &= \tau_{D,t}+\frac{v_{S_t^D}}{\phi_t^D\lambda_t}.
\end{aligned}$$

Using a certainty equivalence approximation,

$$p_{D,t}(1+\tau_{D,t})+\frac{v_{S_t^D}}{\phi_t^D\lambda_t}\simeq\frac{u_{D_{t+1}}}{\lambda_t}+\frac{E_t\left[p_{D,t+1}(1+\tau_{D,t+1})+\frac{v_{S_{t+1}^D}}{\phi_{t+1}^D\lambda_{t+1}}\right](1-\delta_D)}{R_{t+1}E_t efp_{h,t+1}}.$$

These equations show that in the new model we essentially modify the marginal cost of durables to include the shopping effort. Consider an increase in $p_{D,t}(1+\tau_{D,t})$ or in the external finance premium $efp_{h,t+1}$. Ceteris Paribus, these lead to a lower $v_{S_t^D}$, implying lower shopping effort. In equilibrium, the lower shopping effort for durables reduces the probability of making a sale for firms, which from a perspective of a model without these channels amounts to lower TFP in the durables sector. This is similar to an endogenous $\tau_{D,t}$ shocks. In combination with a similar change in the capital goods sector, we would get dynamics similar to those of a joint investment shock in the business capital and consumer durables sectors.

8.5 Adding intermediate production goods: the input-output structure, growth and business cycles

Until now we have assumed that capital and labour are the only production factors. Modern economies have a complex input-output structures in which part of the production of many firms is sold to other firms as an input into their production. These inputs range from material inputs such as fuel or iron to various business services. Under certain conditions we can recover a value added production

function in terms of only capital and labour for GDP net of intermediate good expenditures . The value added production function already incorporates optimal intermediate input choices.

This simplification is commonly used in most business cycle analysis, but it is not innocuous. Explicit modeling of intermediate goods allows us to examine the robustness of conclusions from simpler 1 or 2 sector models to a more realistic treatment of different production sectors. It reveals new sources of efficiency/TFP wedge shocks and of long run income differences between countries. And it suggests a powerful amplification mechanism for shocks hitting the economy, reducing our reliance on large exogenous sources of business cycles. This section shows how to extend our framework to include these intermediate goods.

The economy consists of N sectors, each producing a distinct good y_{it}. $y_{i,t}$ can be turned into a final output good y_{it}^f or into an intermediate good $y_{it}^x = \Sigma_{j=1}^N x_{ij,t}$ sold to other sectors $j = 1...N$ (j includes i, but the sector's use of its own product just nets out in its profits). Final output y_t can be costlessly transformed into nondurable consumption, or it can be transformed into durable goods and capital subject to the previously specified adjustment costs. The representative firm in sector i produces output using labour, capital and intermediate goods $\{x_{ji,t}\}_{j=1}^N$. Note the indexing convention: x_{ij} is produced by i and sold to j. The production function of sector i is

$$\begin{aligned}
y_{i,t} &= g(A_t m_{it}, e_{mi,t}) z_{it} f(k_{it}, A_t n_{it}, X_{it}),\\
X_{it} &= h(\{x_{ji,t}\}_{j=1}^N),\\
y_{it} &= y_{it}^f + y_{i,t}^x,\\
y_{it}^x &= \Sigma_{j=1}^N x_{ij,t}.\\
g_m &> 0,\ g_{mm} < 0,\ g_{e_m} > 0,\ f_n > 0,\ f_{nn} < 0, f_k > 0, f_{kk} < 0, h_{xj} > 0, h_{xjxj} < 0\\
,\ g(am, .)f(ak, an, aX) &\le ag(m, .)f(k, n, X) \text{ and } h(\{ax_{ji,t}\}) = ah(\{x_{ji,t}\}) \text{ for all } a \ge 0.\\
A_t &= G_y A_{t-1}, G_y \ge 1.
\end{aligned}$$

The representative firm in sector i solves

$$\max_{\{k_{t+1}, m_t, n_t, \{x_{ji,t}\}_{j=1}^N\}} E_0 \Sigma_{t=0}^\infty \beta_{fi,t} d_t,$$

$$\begin{aligned}
d_{it} = \; & p_{i,t}^x y_{it} - w_t(1 + \tau_{ndi,t})(n_t + m_t) - \Sigma_{j,t} p_{j,t}^x (1 + \tau_{xi,t}) x_{ji,t}\\
& - p_t^k (1 + \tau_{ki,t})[k_{it+1} - (1 - \delta) k_{it}],
\end{aligned}$$

Note that we have allowed for sector specific wedges, and a common wedge on the purchase of intermediate inputs by sector i. The first order conditions for capital and labour are the same as before except that the marginal revenue term is now multiplied by the price of sector i output relative to the numeraire, $p^x_{i,t}$. For each intermediate input we have the first order condition

$$x_{ji,t} : p^x_{i,t} g_t z_t f_{xji,t} = p^x_{j,t}(1 + \tau_{xi,t}).$$

Under standard assumptions (e.g Mas Collel et al (1995)[146], Mathematical Appendix) we can find a unique solution to this system of N equations for optimal $\{x_{ji,t}\}_j$ demands as a function of capital, labour, prices and the wedges. Substituting these demand functions into $d_{i,t}$ gives us the value added revenue function for the firm

$$VA(k_{it,} n_{i,t}, m_{i,t}, \bullet) = p^x_{i,t} y_{it} - \Sigma_{j,t} p^x_{j,t}(1 + \tau_{xi,t}) x_{ji,t}.$$

This leads to an equivalent problem where the firm pick optimal sequences of $\{k_{it,} n_{i,t}, m_{i,t}\}$ to maximise

$$E_0 \Sigma^{\infty}_{t=0} \beta_{fi,t} \tilde{d}_t,$$

with

$$\tilde{d}_{it} = VA(k_{it,} n_{i,t}, m_{i,t}, \bullet) - w_t(1+\tau_{ndi,t})(n_{it}+m_{it}) - p^k_t(1+\tau_{ki,t})[k_{it+1} - (1-\delta_t)k_{it}].$$

The use of the value added function above is implicit in both macroeconomic and microeconomic studies that assume output is only a function of labour and capital. In the special case of the Cobb-Douglas production function

$$\begin{aligned} y_{it} &= g(A_t m_{it}, e_{mi,t}) z_{it} \left(\left[k^{\alpha}_{it} (A_t n_{it})^{1-\alpha} \right]^{\gamma}, X^{1-\gamma}_{it} \right)^b, \\ X_{it} &= \Pi^N_{j=1} x^{\theta_{ji}}_{ji,t}, \\ 0 &< \alpha < 1, \Sigma_j \theta_{ji} = 1, 0 < \gamma \leq 1,\ 0 < b \leq 1, \end{aligned}$$

the value added function also has a Cobb-Douglas form with exponents less than or equal to 1 (if we further assume $b = 1$, the sum of the exponents in the value added function is again 1).

At this stage we can imagine a large number of identical aggregator firms that produce final output from the intermediate goods at zero profit, or we can assume households, capital and durable goods firms purchase the intermediate inputs directly. The two formulations yield equivalent results. The 1st approach is closer to the model without explicit treatment of intermediary inputs, so we follow it here. To do this, we must specify the link between the intermediate goods and the final output aggregate:

$$\begin{aligned} y_t &= F\left(\left\{ y^f_{it} \right\}^N_{i=1} \right) = c_t + I_t + I^D_t, \\ aF\left(\left\{ y^f_{it} \right\}^N_{i=1} \right) &= F\left(\left\{ a y^f_{it} \right\}^N_{i=1} \right), \text{ for any } a \geq 0. \end{aligned}$$

The final output aggregator's optimisation problem implies the first order conditions for intermediate inputs used

$$\frac{\partial y_t}{\partial y_{i,t}^f} = p_{i,t}^x + \mu_{i,t}^f.$$

$\mu_{i,t}^f$ is the Lagrange multiplier on the constraint $y_{it}^f \geq 0$. This formulation allows for the possibility that an intermediate inputs is only used to produce other intermediate inputs. In this case $y_{i,t}^f = 0$ for that input. The system of equations above has a solution $y_{it}^f = h_i^{yf}(\{p_{j,t}^x\}_j)$. If we want to focus on the effect of intermediate inputs in production, a common simplifying assumption is that intermediate goods are perfect substitutes so that

$$\begin{aligned} F\left(\left\{y_{it}^f\right\}_{i=1}^N\right) &= \Sigma_{i=1}^N a_{i,t} y_{i,t}^f \\ , a_{it} &\geq 0. \end{aligned}$$

[33].

Market clearing conditions for this economy are

$$\begin{aligned} y_{it} &= y_{it}^f + \Sigma_{j=1}^N x_{ij,t} \text{ for } i = 1, ..., N. \\ x_t^k &= \Sigma_{i=1}^N \left[k_{it+1} - (1-\delta)k_{it}\right], \\ x_t^D &= D_{t+1} - (1-\delta_D)D_t, \\ y_t &= c_t + I_t + I_t^D, \\ n_{s,t} &= \Sigma_{i=1}^N n_{i,t} + \Sigma_{i=1}^N m_{i,t}. \\ d_{h,t} &= \Sigma_{i=1}^N d_{i,t} + \pi_{k,t} + \pi_{D,t}. \\ b_{t+1} &= 0, \end{aligned}$$

Note that y_t is the aggregate value added GDP, as reported in national income and product accounts. By using the market clearing conditions for sectors $i = 1, ..., N$ in combination with our previous expressions for $\{x_{ij,t}\}_{i,j}, \left\{y_{i,t}^f\right\}$ as functions of prices, wedges and other production factors , we can solve for the intermediate goods prices $p_{i,t}^x$ also as a function of other prices in the economy, other production factors, and the different wedges. Combining these results with our sectoral value added functions, we can derive and expression for y_t as a function of the sectoral capital stocks, labour inputs and wedges.

Horvath (2000) [109] and Jones (2011b) [116] show that when sectoral production function are Cobb-Douglas in k, n and X and X itself

[33] In that case, under constant returns to scale only the lowest cost intermediate goods sector would produce final output. All other produces would simply sell their output as inputs into the production process of the most cost efficient. With decreasing returns to scale, we can have a nondegenerate distribution of sectors contributing directly to final output.

is a constant elasticity of substitution function in intermediate inputs, then aggregate value added can itself be written as a Cobb-Douglas function of aggregate capital and labour. The total factor productivity term in the aggregate value added production function is a decreasing function of the wedges on intermediate inputs' purchase and the level of sectoral efficiency and labour demand wedges. Under the plausible case in which intermediation inputs are complements, aggregate TFP is also decreasing in the cross-sectional dispersion of the sectoral wedges and productivities. DSGE models usually assume Cobb-Douglas production functions. Since these are a special case of the constant elasticity of substitution function, this result justifies the common practice of focusing on value added for most circumstances.

However, Intermediate production inputs can play an important role in endogenising part of the TFP wedge. This is easiest to see for the Cobb-Douglas production function with constant returns to scale. The result also applies with decreasing returns to scale, introducing a fixed production factor such as a manager or entrepreneur. Consider the static problem of choosing intermediate inputs

$$VA(k_{it}, n_{i,t}, m_{i,t}, \bullet) = p^x_{i,t} y^{va}_{i,t} = \max_{\{x_{ji,t}\}} p^x_{i,t} g_{it} z_{it} \left[k^{\alpha}_{it} (A_t n_{it})^{1-\alpha}\right]^{\gamma} \left(\Pi^N_{j=1} x^{\theta_{ji}}_{ji,t}\right)^{1-\gamma} - \Sigma_{j,t} p^x_{j,t}(1+\tau_{xi,t}) x_{ji,t}.$$

The N first order conditions can be written as

$$x_{ji,t} : (1-\gamma)\theta_{ji} p^x_{i,t} y_{it} = p^x_{j,t}(1+\tau_{xi,t}) x_{j,t},$$

implying

$$p^x_{i,t} y^{va}_{i,t} = \gamma p^x_{i,t} y_{i,t}.$$

Combining this with the first order conditions gives

$$p^x_{i,t} y^{va}_{i,t} = \left[\gamma p^x_{i,t} \left(\frac{1-\gamma}{1+\tau_{xi,t}}\right)^{1-\gamma} g_{it} z_{it} \Pi^N_{j=1} \left(\frac{\theta_{ji}}{p^x_{j,t}}\right)^{\theta_{ji}(1-\gamma)}\right]^{\frac{1}{\gamma}} k^{\alpha}_t n^{1-\alpha}_t.$$

Assume a symmetric equilibrium in which all sectors have the same productivity level, the same production function and the same wedges and in which $y_t = \Sigma^N_{i=1} y^f_{i,t}$. This leads to

$$\frac{y_t}{N} = \mathbf{y}^{va}_{i,t} = \left[\gamma \left(\frac{1-\gamma}{1+\tau_{xt}}\right)^{1-\gamma} \theta^{(1-\gamma)} g_t z_t\right]^{\frac{1}{\gamma}} \mathbf{k}^{\alpha}_t \mathbf{n}^{1-\alpha}_t.$$

The last expression highlights some of the key effects of intermediate inputs.

First, any change in the efficiency wedge z_t or in g_t is amplified by a factor of $\frac{1}{\gamma}$, which is increasing in the weight of intermediate

inputs in the production function $1-\gamma$**. A reduction in the efficiency of production or capacity utilisation encourages firms to reduce intermediate inputs purchases. In terms of a value added production function this reduces production efficiency (TFP) even more. The strength of this reduction in the TFP of capital and labour is larger when intermediate inputs have a bigger role in production (higher** $1-\gamma$**).** The nonlinear amplification effect comes from the complementarity between intermediate inputs and z_t or g_t, so that a firm operating with one half the efficiency of another firm ends up with one quarter of the TFP in terms of value added for a typical $\gamma = 0.5$.

Second, any distortions in the purchase of intermediate inputs, measured by $1+\tau_{x,t}$ **(e.g higher costs of short term financing, greater business uncertainty or lower competition in intermediate input markets) will lower measured TFP. Once again, this effect is stronger the higher the weight of intermediate inputs** $1-\gamma$. For example, imagine as in Arellano et al's (2012)[12] labour wedge model that purchases of intermediate inputs must be decided before gaining full knowledge of a firm's idiosyncratic demand or TFP. Suppose also that these purchases must be paid for before getting any revenue and they must therefore be financed through short term financial loans or trade credit, subject to costly default. In this case higher idiosyncratic uncertainty, or a signal of higher idiosyncratic uncertainty would discourage the purchase of intermediate inputs. From the perspective of a model without heterogeneity this would map into a higher wedge on intermediate inputs $1+\tau_{x,t}$. From the perspective of a value added production function this would amount to a reduction in TFP.

These mechanisms amplify common shocks hitting most sectors simultaneously. A plausible value for γ **is** 0.5, **yielding a multiplier of** 2 **(Jones (2011) [117])**. Jones shows how this multiplier helps explain long run income differences between rich and poor countries with reasonable shares of capital in production. [34] His arguments can also be applied to the amplification of TFP shocks. Horvath (2000) [109] explores the possibility that shocks to a single sector may propagate through the input-output structure to generate an aggregate TFP shock. He finds that for the input-output structure of the US economy, sectoral shocks can be a significant source of fluctuations. Key to this result is to have several critical sectors whose outputs are important inputs into most other sectors' production.

[34] He emphasizes that in the BGP, intermediate inputs are very similar to capital. In general, the neoclassical growth model is capable of accounting for large cross-country income differences with small differences in physical capital investment rates if the share of overall tangible and intangible capital or of capital and intermediate inputs is around 2/3. $\gamma = 0.5$ delivers this share. See also Parente and Prescott (1999) [156] who make similar argument by appealing to differences in intagible capital between countries. Once again, lower productivity reduces intangible capital investment, which reduces the TFP of physical capital and labour even more. The 2 theories are complementary in terms of explaining large income per capita differences with plausible differences in production efficiency.

9 Search frictions in labour markets

Here we allow for search frictions in labour labour markets. This can be seen as a possible microfoundation for labour adjustment costs. We reinterpret labour demand as employment times a fixed number of hours $\bar{h}$, or we can imagine there are part time jobs as well with hours $h' < \bar{h}$, and the cost of posting a vacancy is lower for part time work by a factor of $\frac{h'}{\bar{h}}$ (full time and part time jobs are perfect substitutes for the firm and for the representative household in this interpretation). Without loss of generality normalise $\bar{h} = 1$ and assume all jobs are full time.

To hire n_t^d hours the firm must post vacancies V_t^n that are filled at a rate $\varphi_{n,dt}$ such that

$$\begin{aligned} n_t^d - n_{t-1}^d(1-\delta_{n,t}) &\leq \varphi_{nd,t} V_t^n,\ (\mu_t^{nd}). \\ V_t^n &\geq 0,\ (\eta_t^{nd}), \\ 0 &\leq \varphi_{nd,t}, \delta n, t \leq 1. \end{aligned}$$

Posting vacancies costs $F^N(V_t^n)$, with $F^N_{V^n} > 0, F^N_{V^n V^n} \geq 0$. $\delta_{nd,t}$ is a separation rate taken as exogenous by the firm. For simplicity we assume these costs only affect production workers. Hiring m_t does not incur extra costs. The firm's cash flow becomes

$$\begin{aligned} d_t &= g(A_t m_t, e_{m,t}) z_t f(k_t, A_t n_t) - w_t(1+\tau_{nd,t})(n_t + m_t) - p_t^k(1+\tau_{k,t})[k_{t+1} - (1-\delta_t)k_t] \\ &\quad - F^N(V_t^n). \end{aligned}$$

Note that if the vacancy posting constraint binds we can rewrite cash flows as

$$\begin{aligned} d_t &= g(A_t m_t, e_{m,t}) z_t f(k_t, A_t n_t) - w_t(1+\tau_{nd,t})(n_t + m_t) - p_t^k(1+\tau_{k,t})[k_{t+1} - (1-\delta_t)k_t] \\ &\quad - F^N\left(\frac{n_t^d - n_{t-1}^d(1-\delta_{n,t})}{\varphi_{nd,t}}\right). \end{aligned}$$

This expression emphasizes that these search frictions are similar to other forms of labour adjustment costs. An increase in the vacancy filling rate $\varphi_{nd,t}$ acts like a reduction in these adjustment costs. More generally, we modify the first order condition for n_t^d and add an optimality condition for vacancies:

$$\begin{aligned} n_t^d &: \quad g_t z_t f_{n,t} + E_t \frac{\beta_{t+1}^f}{\beta_t^f} \mu_{t+1}^{nd}(1-\delta_{n,t+1}) = w_t(1+\tau_{nd,t}) + \mu_t^{nd}. \\ V_t^n &: \quad F^N_{V_t^N} = \mu_t^{nd} \varphi_{nd,t} + \eta_t^{nd}. \end{aligned}$$

Combining these 2 equations when vacancies are positive now and

in the future,

$$n_t^d : g_t z_t f_{n,t} + E_t \frac{\beta_{f,t+1}}{\beta_{f,t}} \frac{F^N_{V^N_{t+1}}}{\varphi_{nd,t+1}}(1-\delta_{n,t+1}) = w_t(1+\tau_{nd,t}) + \frac{F^N_{V^n_r}}{\varphi_{nd,t}}.$$

The interpretation of this first order condition is similar to that of other labour demand models with continuous labour adjustment costs. The marginal cost of employment is increased by cost of posting new vacancies, but it is reduced by the saving in terms of lower future vacancy costs. There are several new factors affecting labour demand with search frictions. First, **an increase in the vacancy filling rate $\varphi_{nd,t}$ reduces the cost of hiring and increases labour demand**. Second, iterating foreward and using the boundedness of $F^N_{V_n}$

$$\begin{aligned}
& g_t z_t f_{n,t} + E_t \Sigma_{j=0}^{\infty} \Pi_{k=0}^{j} \frac{\beta_{f,t+1+k}}{\beta_{f,t+k}} (1-\delta_{n,t+1+k})(g_{t+1+j} z_{t+1+j} f_{n,t+1+j} - w_{t+1+j}(1+\tau_{nd,t+1+j})) \\
= \; & \frac{F^N_{V^n_r}}{\varphi_{nd,t}} + w_t(1+\tau_{nd,t}), \Longleftrightarrow \\
& g_t z_t f_{n,t} + E_t \Sigma_{j=0}^{\infty} \Pi_{k=0}^{j} \frac{1}{R_{t+1+k} efp_{e,t+1+k}} (1-\delta_{n,t+1+k})(g_{t+1+j} z_{t+1+j} f_{n,t+1+j} - w_{t+1+j}(1+\tau_{nd,t+1+j})) \\
\simeq \; & \frac{F^N_{V^n_r}}{\varphi_{nd,t}} + w_t(1+\tau_{nd,t}).
\end{aligned}$$

As for capital, we can interpret this as an asset pricing equation: the value of an extra vacancy is equal to a discounted sum of expected future profits from hiring. With adjustment costs, the firm earns extra rents and the discounted marginal profits per employee

$$\mathbf{g}_t \mathbf{z}_t \mathbf{f}_{n,t} - \mathbf{w}_t(\mathbf{1}+\boldsymbol{\tau}_{nd,t}) + \mathbf{E}_t \boldsymbol{\Sigma}_{j=0}^{\infty} \boldsymbol{\Pi}_{k=0}^{j} \frac{\beta_{f,t+1+k}}{\beta_{f,t+k}} (\mathbf{1}-\boldsymbol{\delta}_{n,t+1+k})(\mathbf{g}_{t+1+j} \mathbf{z}_{t+1+j} \mathbf{f}_{n,t+1+j} - \mathbf{w}_{t+1+j}(\mathbf{1}+\boldsymbol{\tau}_{nd,t+1+j}))$$

.**An increase in the future marginal profits per employee, or in the firm's discount factor used to value future profits raises the marginal benefit of hiring workers. This raises labour demand. Similarly, a reduction in future external finance premia of the firm or in risk free interest rates raises labour demand.**

On the household, side, we assume a simple search process with an exogenous matching rate, so that in order to match with n_t the representative household must expand a search effort $e_{ns,t}$ to find new jobs for its members at a rate of $\varphi_{ns,t}$, so that

$$n_{s,t} - (1-\delta_{n,t}) n_{s,t-1} \leq \varphi_{ns,t} e_{ns,t}, \;\; \mu_t^{ns}$$

This holds as an equality in equilibrium. Search effort has a disutility cost of $a_{ns,t}e_{ns,t}$, modifying the utility function to

$$v(c, l, e_{ns}) = u(c, l) - a_{ns,t}e_{ns,t}.$$

The labour supply optimality conditions for this model are

$$\begin{aligned} u_{l,t} &= \lambda_t w_t(1 - \tau_{ns,t}) - \mu_t^{ns} + E_t \frac{\beta_{t+1}}{\beta_t} \mu_{t+1}^{ns}(1 - \delta_{n,t+1}), \\ a_{n,t} &= \mu_t^{ns} \varphi_{ns,t}. \end{aligned}$$

Combining these conditions leads to a new labour supply equation taking into account search costs:

$$u_{l,t} + \frac{a_{n,t}}{\varphi_{ns,t}} = \lambda_t w_t(1 - \tau_{ns,t}) + E_t \frac{\beta_{t+1}}{\beta_t} \frac{a_{n,t} + 1}{\varphi_{ns,t+1}}(1 - \delta_{n,t+1}).$$

The marginal disutility of labour now includes the cost of searching for jobs. On the other side of the equation, searching for a job now has the extra benefit of reducing future search costs. Comparative statics are quite straightforward. For example an increase in the job finding rate $\varphi_{ns,t}$ reduces the labour supply wedge and raises job search.

Normalising the number of households to 1 as before, aggregate consistency requires

$$n_{s,t} = n_{d,t}.$$

That is the number of jobs successfully found by households must equal the number of jobs succesfuly filled by firms. Closing the model requires us to specify more details on either the wage formation process or the matching rates $\varphi_{ns}, \varphi_{nd}$. For example, we could specify exogenous matching rates, and solve the equilibrium for wages ensuring that aggregate consistency holds. This would be similar to a Walrasian, market clearing, equilibrium extended to allow for job rationing (see for example Krusell et al, 2012). More realistically perhaps, we can follow the typical specification in matching models of labour markets and use some combination of Nash bargaining, perhaps with some wage adjustment costs (e.g Shimer, 2009 [165]). Assuming that the households' search effort is in terms of their utility function, the aggregate resource constraint of the economy with search frictions is

$$c_t + I_t = y_t - F^N(V_t^n) = GDP_t.$$

9.1 Involuntary unemployment

The addition of search frictions to the job finding process allows us to account for the distinction between voluntary and involuntary unemployment. In the classic RBC model the job finding rate $\varphi_{ns,t} = 1$.

There is unemployment because some potential workers' reservation wage is above their efficiency units ajusted market wage $\varepsilon_{j,t}w_t > 0$. But in theory those who are unemployed could always find a job with a lower but still positive wage. A qualified job seeker could find a less qualifed or less pleasant job, but at his reservation wage he prefers to keep on looking for a better job. This "voluntary" unemployment is well captured in standard DSGE models with indivisible labour but without explicit search frictions. In this theory, unemployment went above 20% in the Great depression in the US or in Spain during the great recession of 2008, because the wages that would be offered to those who ended up unemployed was too low to make it worthwhile for them to work.

With the job finding rate $\varphi_{ns,t} < 1$, unemployment is "involuntary", in the sense that the job seeker may stay unemployed because he cannot find any work opportunities with a wage above zero. For example, the more highly skilled worker may have trouble in finding a less qualified job, even if he wanted to, because he is considered overqualified and unlikely to be productive in a less skilled job. Voluntary unemployment can be thought of as coming from the fundamental calculus of whether or not working is desirable given job conditions compared to the benefits of unemployment. Involuntary unemployment is associated with genuine inefficiency in the labour market in terms of matching workers and jobs, similar to a productivity shock in the matching function.

9.2 Search frictions and the labour wedge

Search frictions in labour market are often seen as a prime candidate for explaining fluctuations in the labour wedge. Formally, we can rewrite the first order conditions in the costly vacancy model as

$$\frac{g_t z_t f_{n,t}}{w_t} = (1+\tau_{nd,t}) + \frac{F^N_{V^n_r}}{\varphi_{nd,t}w_t} - E_t\frac{\beta_{f,t+1}}{\beta_{f,t}}\frac{F^N_{V^N_{t+1}}}{\varphi_{nd,t+1}w_t}(1-\delta_{n,t+1}) = 1+\tilde{\tau}_{nd,t}.$$

Search frictions add the term $\frac{F^N_{V^n_r}}{\varphi_{nd,t}w_t} - E_t\frac{\beta_{f,t+1}}{\beta_{f,t}}\frac{F^N_{V^N_{t+1}}}{\varphi_{nd,t}w_t}(1-\delta_{n,t+1})$ to the labour demand wedge. This shows, that search frictions increase the labour demand wedge in a recession if the vacancy filling rate is lower or if the marginal cost of posting a vacancy is higher at time t. The marginal cost of posting a vacancy is often assumed constant or increasing in the number of vacancies. This tends to reduce the labour demand wedge in a recession, as the number of vacancies declines. The vacancy filling rate also tends to increase in most recessions, as the pool of unemployed people increases. Again, this makes the contribution of search frictions to the labour wedge procyclical.

Shimer (2009) [165] argues that search frictions per se act like procyclical labour wedges (essentially because search frictions are a form of labour adjustment costs), and are therefore incapable of explaining the significant countercyclical labour wedges in the data. He emphasizes real wage rigidity as a possible cause of countercyclical labour wedges with search frictions. Note that he reaches this conclusion assuming front-loaded TFP shocks. For news shocks, labour adjustment costs (and therefore search frictions) may stimulate labour demand in advance of a positive shock, acting like a countercyclical labour wedge. This can be seen in the equation above by noting that $\frac{F^{N}_{V^{N}_{t+1}}}{\varphi_{nd,t+1} w_t}$ enters the labour wedge with a negative sign. Consider a positive news shocks that leads to an increase in future vacancy postings. A future increase in the marginal cost of posting vacancies relative to the current cost encourages firms to post vacancies earlier. Similarly a lower probability of filling a vacancy in the future relative to the current period encourages hiring now rather than later (see also Jaimovich and Rebelo (2009) [115]).

Several other mechanisms have been proposed for amplifying labour wedge movements through search frictions. More realistic recruitment costs with fixed components can amplify unemployment fluctuations, generating what looks like higher labour wedges in a recession (Pissaride (2009) [157]). Petrosky-Nadeau (2009) and Chugh (2009) add financing frictions affecting recruitment costs to search and matching DSGE models. They find that these financing frictions have significant potential to explain labour wedges. Both of these factors can be mapped into our model as more countercyclical marginal costs of posting vacancies $F_{V^{n}_{t}}$. The matching process can be less efficient in a recession, for example if a higher proportion of vacancies is posted for harder to match skilled workers.

Hall (2014) [107] finds that matching became significantly less efficient during the great recession of 2008, contributing significantly to the rise in unemployment in the US. Cheremukhin et al (2012) augment a standard RBC model with search frictions by adding worker bargaining power, separation and matching efficiency shocks. They find that matching efficiency shocks are responsible for much of the labour wedge in the US. Finally, matching frictions can lead to higher fluctuations in the labour wedge through shifts in the firm's discount factor $\frac{\beta_{f,t+1}}{\beta_{f,t}}$. Essentially, with labour adjustment costs hiring decisions become similar to investment decisions. A fall in the discount factor (e.g due to a higher firm external finance premium $efp_{e,t+1}$) lowers the value of future profits from hiring workers, and reduces labour demand (see Hall (2014) [107] for an elaboration on this effect and a quantitative exercise that shows its potential importance in the 2008 financial crisis and the ensuing Great Recession).

10 An explicit mapping between credit frictions, imperfect competition frictions and distortionnary wedges

Here, we provide an example of how financing and market power frictions show up as wedges in a neoclassical growth model. We start by imagining that firms and households in the product, labour, intermediate input and investment goods markets must finance a proportion $0 < a_{l,t} \leq 1$ of their expenditure using intratemporal loans, where $l = c, D, n_d, k, x$. Because the loans are intratemporal , they cancel out inside the representative household's budget constraint [35]. The intratemporal nature of the loans also means that the initial gross interest rate is 1, but due to financing frictions there is an extra proportional cost to the loan leading to an overall repayment of $R_{l,t}L_{l,t} \geq L_{l,t}$.

We can think of these intratemporal loans as very short term loans for which the benchmark risk free interest has almost no effect, and the overwhelming component of the financing cost comes from external financing spreads or from the shadow cost of credit constraints (the Lagrange multipliers). Most of the credit card debt, lines of credit, commercial paper and trade credit can be treated in this way. We keep assuming that the costs are rebated to households through lump-sum taxes: in other words the financing cost is in the form of a Lagrange multiplier on a credit constraint or loss of resale value due to asymmetric information in secondary loan markets than due to actual physical loan processing costs.

Relative to an economy without distortions, a firm must now spend an additional

$$
\begin{aligned}
&a_{k,t}(R_{k,t}-1)p_{k,t}[k_{it+1}-(1-\delta)k_{it}],\\
&a_{nd,t}(R_{nd,t}-1)w_t(n_t+m_t),\\
&a_{x,t}(R_{x,t}-1)p^x_{j,t}x_{ji,t}.
\end{aligned}
$$

A household must now spend an additional $a_{c,t}(R_{c,t}-1)c_t$ and $a_{d,t}(R_{d,t}-1)x^D_t$ to acquire consumption goods.

Equating $\tau_{l,t} = a_{l,t}(R_{l,t}-1)$ for $l = c, D, n_d, k, x$ gives the relation between changes in financing costs and the wedges. It is immediate that higher financing costs lead to higher wedges. While here we have

[35] For the household intratemporal loans we have to introduce at least 2 types of households. One group has to fund some of its consumption spending before getting any revenue using intratemporal loans. To preserve aggregation into a single representative household, allow households to trade consumption insurance contracts among themselves. Alternatively, with Cobb-Douglas CRRA preferences but without consumption insurance, aggregation holds to a first order approximation just like for $efp_{h,t+1}$ (despite the fact that the $\tau_{c,t}$ wedge also affects labour supply).

assume intratemporal loans for simplicity, the effects would be similar if we had intertemporal loans.

A simple mapping from imperfect competition to wedges can be obtained using the standard Dixit-Stiglitz model of monopolistic competition (as in Walsh (2003) ch. 5 [177] and Gali (2007) , ch. 3 [91]). This is the workhorse model for much of the time varying markups and sticky prices literature. For each sector $l = c, k, x$ imagine there is a continuum of measure 1 of different goods $x_{l,j}$. What actually enters household utility and firm production functions is an aggregate

$$X_l = \left[\int_0^1 x_{l,j}^{\frac{\xi_l - 1}{\xi_l}} dj \right]^{\frac{\xi_l}{\xi_l - 1}}, \xi_l > 1.$$

ξ is the elasticity of substitution among goods, with higher ξ_l indicating higher substitution and a more price elastic demand function ($\xi_{l,t} \to \infty$ leads to perfect substitutes). Each product variety is produced by a single monopolistic producer from the final good of the economy, where for simplicity we have assumed a linear aggregator of sectoral final output into aggregate GDP. The only cost of producing the variety is the cost of purchasing the initial final output good. The profits of the producers go to the representative household. From the first order condition

$$\begin{aligned} p_{l,t}^{\xi} &= \frac{\xi_{l,t}}{\xi_{l,t} - 1} p_{l,t} \\ p_{l,t} &= p_{k,t}, p_{D,t}, \{p_{j,t}^x\}, 1. \end{aligned}$$

The markup $\frac{\xi_{l,t}}{\xi_{l,t}-1} \geq 1$, and is decreasing in the elasticity of substitution $\xi_{l,t}$. Equating $1 + \tau_{l,t} = \frac{\xi_{l,t}}{\xi_{l,t}-1}$ for $l = c, D, x_j, k$ we get a mapping from imperfect competition markups to time varying wedges on consumption, durables and capital investment and intermediate inputs purchases. An increase in the markup (lower competition) is equivalent to a higher wedge.

Finally, we can combine financing and imperfect competition wedges, and we can allow for interactions between the imperfect competition and financing wedge components. for example, higher costs of financing usually reduce entry by new firms, investment in new retail locations or new product lines that offer substitutes to existing products, and investment in advertising that can improve buyer information about substitute products. Through all of these channels higher financing shocks can reduce competition and increase markups.

Our mapping of financing frictions into wedges raises important questions on the identification of the effects of changes in financing supply conditions from changes in the demand for financing due to shifts in confidence, expectations or uncertainty. Recall how these expectations shocks can also be mapped into the same wedges. Recall also the difficulties we had in distinguishing between increases in the households' external finance premium and in the desire for precautionary saving for households that did not participate in the stock market or that owned firms. This suggests that without a specific link between the statistical processes for wedges and financial variables such as credit spreads or credit conditions indices or without further theoretical restrictions from more explicit models of frictions, it is quite difficult to separate financial shocks from expectations or uncertainty shocks.

10.1 The relation between heterogeneous financing wedges on production inputs and TFP shocks

Several papers have found an important effect of negative financial shocks in lowering aggregate TFP when firms are heterogeneous (e.g Buera and Moll (2012)[35], Khan and Thomas (2013) [125], Bloom et al (2012)[34]). Here we illustrate this effect in a simple context.

Consider an economy with a measure 1 of potentially heterogeneous firms indexed by i that use the production function

$$y^i = zf(\{x_k^i\}_{k=1,,K}),$$

where $f()$ is concave, z is a common TFP level and each x_k is a production input that can be adjusted flexibly inside the period (for example labour, materials or intermediate goods).

Each firm picks $\{x_k^i\}$ to maximise profits

$$\pi^i = y^i - \Sigma_k(1+\tau_{xk}^i)p_{xk}x_k^i.$$

The first order conditions for this problem are of the form

$$zf_{x,ik} = (1+\tau_{xk}^i)p_{xk}$$

Comparing two firms i and j we see that

$$\frac{f_{x,ik}}{f_{x,jk}} = \frac{1+\tau_{xk}^i}{1+\tau_{xk}^j}.$$

For most commonly used production functions (e.g the CES production function), this relation implies that $\frac{x_k^i}{x_k^j}$ is decreasing in the relative wedge $\frac{1+\tau_{xk}^i}{1+\tau_{xk}^j}$. We assume this relation holds for the rest of this section. Suppose at $t=0$, the wedges $\tau_{xk}^i = \tau_{xk}$ are common across firms. By the concavity of the production function, all firms then make the same input choices. At $t=1$ a financial shock raises the cost of financing disproportionately for some firms so the dispersion of the wedges increases in the cross-section of firms. Consider the effect of raising the dispersion of the wedges for some of the inputs x_k^i while holding the total aggregate demand $x_k = \int_i x_k^i di$ for each production input constant. The higher dispersion in the wedges increases dispersion in the demand for inputs. **By Jensen's inequality aggregate output at $t = 1$ will be lower than at $t = 0$:**

$$\begin{aligned} y_0 &= \int_i y_0^i di = zf(\{x_k\}_k) = zf\left(\left\{\int_i x_{k,1}^i di\right\}_k\right) > z\int_i f(\{x_{k,1}^i\}_k)di = \frac{\int_i f(\left\{x_{k,1}^i\right\}_k)di}{f(\{x_k\}_k)} zf(\{x_k\}_k) = y_1, \\ \Rightarrow & \quad \frac{\int_i f(\left\{x_{k,1}^i\right\}_k)di}{f(\{x_k\}_k)} < 1. \end{aligned}$$

Defining aggregate TFP as $\bar{z} = \frac{\int_i f(\left\{x_{k,1}^i\right\}_k)di}{f(\{x_k\}_k)} z$, we see that the rising dispersion in financing frictions affecting firms reduces aggregate TFP. So from the perspective of a model with a singly representative firm a financial shock that disproportionately worsens credit conditions for a subset of firms will look like a loss in the production efficiency of the economy. Other authors such as Khan and Thomas (2013) [125] and Buera and Moll (2012) [35] focus instead on aggregate TFP losses caused by misallocation of capital in response to a financial shock. But the general idea is the same: higher financing frictions typically increase the dispersion of marginal products of production inputs in the cross section of firms. With diminishing marginal productivity this leads to lower aggregate production for a given aggregate amount of production inputs: in other words aggregate TFP declines.

11 A large open economy extension of the RBC model

Until now we have assumed a closed economy. This section cannot possibly cover all of the issues raised by an open economy (see Uribe and Schmitt-Grohe (2014) for a comprehensive textbook on open-economy DSGE models). The objective is a more modest extension of the closed economy model to allow us to consider current account fluctuations. Much has been written about the business cycle dynamics of a small open economy with little control over interest rates it faces and extensions to small or moderate frictions generating systematic differences between domestic and foreign risk free rates. The small open economy setup can be seen as having a perfectly elastic supply of world lending to the country.

Here we consider the opposite extreme of an exogenous current account which may be more relevant for large countries or blocks like the US, the Eurozone, Japan and maybe even China (though many would argue that the large

Chinese current account surplus is an important factor in China's growth path and business cycle). For most of the post 2nd world war period, current accounts have been much smaller than would be predicted by DSGE models with unrestricted international financial markets. Heathcote and Perri (2002) compare international RBC models with complete international financial markets, incomplete markets and financial autarky, in which the current account cannot be adjusted. They find that the financial autarky model provides a better empirical fit to the data in several dimensions. From this perspective, the exogenous current account model may be seen as a useful simplification.

Suppose there is a perfectly inelastic world demand for domestic debt, as in Favilukis et al (2011) [82]. This changes the bond market equilibrium condition to

$$b_{t+1} = nfa_{t+1},$$

where nfa_{t+1} is the exogenous process for net foreign assets. Alternatively we can express this as

$$b_{t+1} + b^f_{t+1} = 0,$$

where b^f_{t+1} are exogenous domestic bond holdings by foreigners. In this case movement in domestic demand for loans will simply shift interest rates, just like in the closed economy. If this is a good approximation, then the closed economy framework may not be too inaccurate. At least for current accounts which are not too large in magnitude, the previous analysis of the reaction of the economy to different shocks remains true. But allowing for a non-zero current account does introduce an extra source of shocks.

What happens if world demand for domestic debt shifts for some time (a stationary shock), for example because China desires to hold more US debt? Consider first the real business cycle economy in the case where the economy is a net borrower from global financial markets. The current account balance decreases by assumption (it becomes even lower than zero). Intuitively the increased supply of debt to the US causes real interest rates to go down, encouraging investment and consumption. Households feel wealthier and reduce their labour supply. If there is no feedback from the better borrowing conditions to the demand for flexible production inputs such as labour or to the utilisation rate of the inputs (due to e.g. lower financing or search frictions), domestic production goes down despite the increase in consumption and investment. This is compatible with equilibrium because of the decline in net exports. On the other hand if there are positive feedback effects on utilisation rates and labour demand, domestic production can increase despite the decline in net exports.[36]

The analysis changes for a country that is a net lender to the world. In that case we would still expect a decline in interest rates and an increase in

[36] In theory lower labour supply could reduce investment if it lasts several periods, by reducing the marginal product of capital. In practice, the effect of a lower interest rate probably dominates.

investment. The effect on consumption may be different because lower interest rates generate a negative income effect for savers, going against the direction of the substitution effect and the wealth effect. So we would expect any increase in consumption to be lower, or we could even have a decline in consumption.

If we know that the shock above actually reduces net exports

$$y_t - c_t - x_t^k = b_{t+1} - b_t(1 + r_t)$$

, then the conclusions are clearer. In this case, Chari et al (2007) [50] note that this shock is identical to a lower government spending shock financed by cuts in lump-sum taxes. So we can apply the analysis from government spending shocks in a closed economy to argue that if net exports go down, absent any effects on TFP or effects on the household marginal utility of consumption through changes in durables investment, then consumption and investment increase while labour supply and output decline. The same analysis with reversed signs applies to a "sudden stop" where international financial markets are less willing to lend to the domestic economy. This shows that in a baseline RBC model, a pure sudden stop would actually lead to higher output! Of course, sudden stops are usually accompanied by negative terms of trade shocks and financial shocks that worsen financing conditions for firms. Terms of trade shocks are similar to a reduction in TFP, and we have already seen how worse financing conditions lower TFP and increase labour demand wedges. These factors can explain the decline in output during a sudden stop (see for example Mendoza (2008)).

12 Bringing the DSGE model to the data

Our analysis so far constitutes a relatively complete example of a DSGE model of the economy. Here we discuss a framework for estimating and using such a DSGE model for policy analysis and private sector forecasting or scenario analysis (conditional forecasting). The following setup is inspired by Caldara et al (2012)[40], and Schorfheide et al (2009)[163].

Our setup aims to be a pragmatic compromise between the stronger theoretical and logical coherence of DSGE models and the greater flexibility of reduced-form time series models in fitting the data. We will also try to provide a compromise between the usefulness of having a core macro model to organise the analysis and the forecasts, and the greater robustness of model combination.

For this purpose, we will sketch a forecasting and macro analysis system that combines

1) a core DSGE or hybrid DSGE-VARX model and

2) a set of auxiliary models.

12.1 The core DSGE model

For convenience we repeat the standard state space representation of a solution to a DSGE model,

$$\begin{aligned} s_{t+1} &= g(s_t, e_{t+1}), \\ y_t &= h(s_t). \end{aligned}$$

We partition the state vector $s_t = [k_t; \tau_t]$ where k_t are variables that are always endogenous states, and τ_t are wedges.

The business cycle accounting framework (e.g Chari et al (2007)[50]) assumes that τ_t follow exogenous processes. It emphasizes that there is no reason to expect these wedges to be independent as assumed in many policy DSGE models (e.g Smets and Wouters (2007) [168]). For example, a loss of confidence in financial market can affect labour wedges $\tau_{nd,t}$ and $\tau_{ns,t}$, efficiency wedges $z_t, e_{m,t}$, investment wedges $\tau_{kt}, \tau_{D,t}$ and consumption wedges $\tau_{c,t}, efp_{h,t+1}$ simultaneously. In general we should allow τ_t to depend on k_t, so τ_t is no longer part of the minimal state vector. For example higher output (which is partly a function of k_t) can reduce financing costs through a financial accelerator effect. [37]

Without any further restrictions, we are faced with the possibility of arbitrary patterns of contemporaneous and dynamic correlations between the wedges τ_t. This situation can easily generate identification and estimation problems. A possible solution is, following Caldara et al (2012)[40], to link the wedges to a set of more fundamental truly structural factors F_t. For example, F_t could represent a fundamental change in business confidence or financial markets beliefs about certain assets' quality, or to go even deeper some common news or confidence factor that generates both higher business and household uncertainty and higher asymmetric information in financial markets.

This leads to a system of the form

$$\begin{aligned} k_{t+1} &= g(k_t, \tau_t, e^k_{t+1}), \\ y_t &= h(k_t, \tau_t), \\ \tau_t &= \tau(k_t, F_t, e^\tau_t), \\ F_{t+1} &= \kappa(F_t, e^F_{t+1}), \\ e^\tau_t, e^F_t &\backsim \text{IID shocks.} \end{aligned}$$

[37] An interesting example of wedges that are functions of other states or other endogenous variables can be found in the fiscal policy DSGE litterature (e.g Leeper et al (JE2010)), where tax rates are allowed to respond to variables such as the past stock of government debt or current gdp.

The key simplification in this structure is that the dimension of F_t can be significantly lower than that of τ_t. Furthermore, if F_t truly contains only structural shocks we can assume its elements are orthogonal to each other, simplifying estimation. We can turn this into a state-space system in $\tilde{s}_t = [k_t; F_t]$ by substituting out the equations for τ_t.

We allow for extra randomeness e_t^τ in the τ_t equations to avoid the stochastic singularity that occurs when the number of observables exceeds the number of shocks. Stochastic singularity means that certain linear combinations of y_t are deterministic according to the model, making it impossible to define a likelihood function for the data (since the model is automatically falsified by any data set violating the deterministic linear combination). This can be avoided by having at least one "tax" wedge for each quantity variable and one cost or productivity shock associated with each price variable, and allowing each τ_t component to have a stochastic error term e_t^τ. Alternatively, we can add measurement errors to the observables or embed the DSGE model into a hybrid semi-structural model. This is discussed in greater detail in the next section.

DSGE modelers have typically not dedicated enough attention to the possibility of richer correlation and feedback patterns among the shocks. The best specification of the path of the shocks has also been underexplored, with excessive reliance on AR(1) processes, even though there is evidence that moving average components can significantly improve model fit and the economic interpretation of the shocks (Leeper and Walker (2010) [138]). In particular MA(q) components can capture some effects of the gradual diffusion of shocks or information about shocks, making it easier to generate more realistic hump-shaped dynamics for GDP and its components. For example the process with correlated news (discussed in the section on news shocks)

$$z_t - z = \rho(z_{t-1} - z) + \Sigma_{j=0}^{q} \theta^j \varepsilon_{t-j}$$

can generate more realistic dynamics while only adding a single parameter θ to estimate relative to the typical AR(1) process.

A useful source of priors on the parameters of $\tau(.)$ are macroeconomic models with more detailed frictions, such as heterogenous agent models or more generally models with more explicit credit constraints. For example we can estimate the parameters of $\tau(.)$ on simulated data from an auxiliary model with credit frictions. Because the sample is simulated, it can be much longer than a real world sample and we can treat k_t as observed, allowing much better identification of the parameters of $\tau(.)$. The estimates on simulated data will not be converge to the population estimates of the parameters in the representative agent model, since the heterogenous agent model will typically omit other frictions and shocks that are part of the representative agent model. But the estimates from simulated data should be useful in generating priors for estimating $\tau(.)$.

Formally, we have a set of auxiliary models defined by

$$\left\{ \begin{array}{c} s_{m,t+1} = g_m(s_{m,t}) + e_{m,t+1}, \\ y_{m,t} = h_m(s_{m,t}) + v_{m,t}. \end{array} \right\}_{m=1,,,M}$$

As long as the the auxiliary model variables $(y_{m,t}, s_{m,t})$ include (y_t, s_t) from our core DSGE model, we can form the wedges τ_t using simulated data from the auxiliary model. We can use these simulation based $\tau_{m,t}$ to estimate a process for the wedges, e.g using OLS.

Caldara et al (2012)[40] suggest we pick parameters in the process for τ_t to minimise the distance between the IRF's in response to e_t^F in the main DSGE model and the auxiliary model. They simplify estimation with a 2 step approach. In the 1st step, they estimate parameters that are not related to the wedge processes by Bayesian methods under the assumption that the wedges are independent of each other, and follow simple ARMA(p,q) processes as in typical DSGE models. In the 2nd stage they reestimate parameters θ_τ of the τ_t process to minimise the distance between model IRFs and those from auxiliary models that contain F_t in response to an e_t^F shock, while keeping the other parameters at their estimated value from the the 1st stage. Formally we define the IRF vector in our main model for the jth shock $e_{j,t}^F$ $IRF_t = (\frac{dy_t}{de_{j,t}^F}...; \frac{dy_{t+k}}{de_{j,t}^F}, ...)$. IRF_t^m is defined in a similar way for the auxiliary model m. Then using some weighting matrix W (e.g the identity matrix), we pick the parameters θ_τ for the process τ_t to minimise the quadratic distance

$$(IRF_t - IRF_t^m)^T W (IRF_t - IRF_t^m).$$

For example, we can estimate the effect of credit shocks on the TFP and investment or $\frac{v_{f,t+1}}{v_{f,t}}$ wedges by matching IRF's in response to a credit shock in the Khan and Thomas (2013) [125] model.

The set of auxiliary models $m = 1, ...M$ can include other DSGE models, structural partial equilibrium models as in Strebulaev and Whited (2012) [6],semi structural simultaneous equations macro models ,and reduced form/time series models such as ARMAX or VAR's. We can use a combination of shocks to the wedges in our DSGE model to mimic a weighted average of the response from the auxiliary models either by picking θ_τ as in Caldara et al (2012) [40]. Or we can follow the more traditional approach of estimating independent ARMA(p,q) processes for the components of τ as part of the main DSGE estimation, and then finding a sequence of shocks to τ_t that mimic the IRF's of the auxiliary models. Formally, we use standard DSGE estimation methods to estimate a process

$$\ln \tau_t - \ln \tau = \rho_\tau(\ln \tau_{t-1} - \ln \tau) + e_t^\tau$$

, where ρ_τ is a diagonal matrix, and we focus on AR(1) shock processes without loss of generality. Then, we pick a sequence of shocks $(e_t^\tau ... e_{t+k}^\tau ...)$ to minimise

$$(IRF_t - IRF_t^m)^T W (IRF_t - IRF_t^m).$$

Note that in this procedure the parameters of the wedge processes θ_τ are replaced by realisations of the shocks to the wedges e_{t+k}^τ. Both approaches allow us to adjust DSGE model forecasts to account for shocks and transmission mechanisms that are not formally part of our DSGE. The reader is referred to the documentation of Bank of England's new forecasting and policy analysis system (2013) [39] for a detailed discussion of these issues.

More generally, we may be interested in adjusting the forecasts from a core DSGE model to take into account forecaster or policy-maker judgement, or forecasts from other models. We can think of this as minimising the distance between some desired target and actual DSGE forecasts for some function $m(.)$ of the IRF's. This leads to the following problem:

$$\min_{(e_t^\tau ... e_{t+k}^\tau ...)} (m(IRF_t) - m(IRF_t^m))^T W (m(IRF_t) - m(IRF_t^m)).$$

A common example is fixing a target path for a subset of the observable variables $y_{1,t}$. In the (log)-linear state-space system approximation, this problem has an exact solution. Consider the system

$$\begin{aligned} y_t &= h s_t, \\ s_t &= g s_{t-1} + e_t. \end{aligned}$$

, where the steady state is normalised to $s = 0$.

Substituting the state variables equations into the observable variables equations,

$$\begin{aligned} y_t &= f s_{t-1} + h e_t, \\ s_t &= g s_{t-1} + e_t, \end{aligned}$$

$$, f = hg.$$

We will use the state variables $s_{1,t}$ to fix the values of $y_{1,t}$.Without loss of generality, we assume that $y_t = (y_{1,t}; y_{2,t})$ and $s_t = (s_{1,t}, s_{2,t})$. Partition the matrices in the y_t equations such that

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{22} & f_{22} \end{bmatrix} \begin{bmatrix} s_{1,t-1} \\ s_{2,t-1} \end{bmatrix} + \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}.$$

The objective is to have

$$\begin{aligned} y_{1,t} &= f_{11}s_{1,t-1} + f_{12}s_{2,t-1} + h_{11}e_{1,t} + h_{22}e_{2,t} = m_t, \\ t &= 0,...,T. \end{aligned}$$

In each period t, we take as given s_{t-1} and $e_{2,t}$. If h_{11} is invertible, this allows us to hit our target for $y_{1,t}$ by setting

$$e_{1,t} = h_{11}^{-1}\left(m_t - f_{11}s_{1,t-1} - f_{12}s_{2,t-1} - h_{22}e_{2,t}\right).$$

Given the solution for $e_{1,t}$ we can update our forecast for s_t and $y_{2,t+1}$ and then use the equation above to solve for $e_{1,t+1}$. By repeating these steps for $t+j \leq T$ we get the full forecast for all variables, while satisfying the target values for $y_{1,t}$. A necessary condition for the existence of h_{11}^{-1} is that then number of shocks in $e_{1,t}$ matches the number of targetted variables in $y_{1,t}$. Otherwise, if the number of shocks is too small, we can only approximately hit our targets by solving the original minimisation problem. If the number of shocks is too large, then there are several possible paths of $e_{1,t}$ that satisfy our targets. We can choose among them by adding extra conditions, such as using only the shocks which have the highest contribution to the historical variance of $y_{1,t}$.

12.2 The set of auxiliary models

The set of auxiliary models

$$\left\{ \begin{array}{c} s_{m,t+1} = g_m(s_{m,t}) + e_{m,t+1}, \\ y_{m,t} = h_m(s_{m,t}) + v_{m,t}. \end{array} \right\}_{m=1,,,M}$$

can serve either as a way to better identify certain structural shock IRF's, to incorporate missing channels or mechanisms into scenario analysis, or as a way to improve forecasts through model combination.

We can make our forecasts more robust for a subset y'_t of y_t by combining some of the models in the suite indexed by m' to form the combined density function

$$\begin{aligned} p(y'_{t+1}|y'^t) &= \Sigma_{m'=1}^{M'}\omega_{m',t}p_{m'}(y'_{t+1}|y'^t), \text{ where} \\ 0 &\leq \omega_{m',t} \leq 1 \text{ for each } m' \text{ and} \\ \Sigma_{m'}\omega_{m',t} &= 1. \end{aligned}$$

Amisano and Geweke (2013) review different methods of selecting the model weights $\{\omega_{m',t}\}$. When the number of model in the combination is small (e.g a representative DSGE model, a representative VAR and a representatitve simultaneous equations macro model), they find that by standard critieria equal weighting with $\omega_{m'.t} = \frac{1}{M'}$ is a reasonable and simple approximation to the optimal weights.

Alternative macro models or models of aggregate demand:

- Keynesian sticky price models such as Smets and Wouters (2007) [168], the IMF's GIMF model (2013) [7], Eggertson and Krugman (2012)[71], Mertens and Ravn (2014) [149].

https://www.imf.org/external/pubs/ft/wp/2010/wp1034.pdf

https://www.imf.org/external/pubs/ft/wp/2013/wp1355.pdf

- Semi-structural simultaneous equation models, such as NIGEM, FRB-US or the Oxford Economics Global Economic Model:

https://nimodel.niesr.ac.uk/

http://www.federalreserve.gov/econresdata/notes/feds-notes/2014/a-tool-for-macroeconomic-policy-analysis.html

http://www.oxfordeconomics.com/forecasts-and-models/countries/scenario-analysis-and-modeling/global-economic-model/overview

- Behavioural Macroeconomics models such as those in Gabaix (2015) [88]or the Eurace ABM (2011,2014).

http://pages.stern.nyu.edu/~xgabaix/papers/brdp.pdf

http://www.wiwi.uni-bielefeld.de/lehrbereiche/vwl/etace/Eurace_Unibi/

- RBC models with indeterminacy, matching frictions in product markets, and confidence/sentiment shocks, such as Angeletos et al (2014) [10], Bai et al (2012) [17], Benhabib and Wen (2012) [26], Farmer (2011) [80].

Credit constraints and the financial accelerator:

- This section contains models that are more aggregated than heterogenous agent models of financing frictions, but in return they have more realistic feedback effects between credit conditions, asset prices and economic activity. Examples of such models include Christiano et al (2012) [56], Del Negro et al (2011) [152], Gertler and Kiyotaki (2010) [94], Iacoviello (2005) [111], Liu and Zha (2011) [139].

Consumption and saving decisions:

- Baseline precautionary micro-level saving/permanent income models: e.g Carroll (1997,2001) [44][46], the extension to imperfect information in Guvenen and Smith (2014) [105], extensions to housing and other illiquid assets as in Gruber and Martin (2003) [103], Iacoviello and Pavan (2013) [113] or Kaplan and Violante (2014) [122]. These models have to be aggregated by simulating a large number of households starting from a realistic income and wealth distribution. This makes them more complicated than aggregate consumption models, but they allow for a richer, more realistic aggregate consumption function built from the bottom up.

- Behavioural models such those using prospect theory as in Barberis (2012) or Barberis et al. (2008), or the sparse max consumption/saving model in Gabaix (2015) [88].

- Time series models of aggregate consumption of the form

$$\Delta \ln c_t = \alpha_c + \Sigma_{j=1}^{J} \beta_{jc} \Delta \ln c_{t-j} + \Sigma_{k=0}^{K} \gamma_{kc} x_{t-k}^{c} + e_t^c.$$

x_t^c can include disposable income, interest rates, credit condition indices etc...These models could be estimated with instrumental variables, or by OLS if we are willing to take them as purely reduced-form equations and ignore endogeneity bias.

Labour supply decisions:

- Indivisible labour models such Erosa et al (2014) [75], or Chang and Schorfheide [49].

Firm investment and hiring decisions:

- Structural, optimisation based models of firm decisions at the micro level such as those discussed in Strebulaev and Whited (2012) [6], or Bloom (2009) [33].

- General equilibrium models with heterogeneous firms models, non-convex adjustment costs, and financing frictions, such as Arellano et al (2012) [12], Bloom et al (2012) [34], Buera and Moll (2012) [35] or Khan and Thomas (2013) [125].

- Time series models of aggregate investment and hiring of the form

$$\Delta \ln I_t = \alpha_i + \Sigma_{j=1}^{J} \beta_{ji} \Delta \ln I_{t-j} + \Sigma_{k=0}^{K} \gamma_{ki} x_{t-k}^{I} + e_t^{I},$$

$$\Delta \ln n_t = \alpha_n + \Sigma_{j=1}^{J} \beta_{jn} \Delta \ln n_{t-j} + \Sigma_{k=0}^{K} \gamma_{ni} x_{t-k}^{n} + e_t^{n}.$$

For investment, x_t^I can include aggregate output, stock market to output ratios, interest rates, measures of credit conditions etc...For labour demand, x_t^n can include wages, aggregate output, measures of credit conditions etc...Again, since these equations are meant as simple reduced-form benchmarks rather than structural equations, we can estimate them by OLS (possibly alleviating some of the endogeneity concerns by using only lags of the x_t^I, x_t^N variables).

Long term growth:

- Extensions of the basic neoclassical growth model with a representative firm to include intangible versus tangible capital, human capital as in Parente and Prescott (1999) [156].

- Endogenous growth models as in Ljunqvist and Sargent (2004, ch 14) [140], or Jones and Manuelli (2004) [118].

- Heterogeneous firms models such as Atkeson and Kehoe (2007) [13], Buera and Shin (2013) [36], Restuccia and Rogerson (2007) [160], Restuccia (2012) [159].

12.3 Bridge equations

We will typically not be able to include all the variables of interest to a decision maker in y_t due to computational constraints. For example, the model may not distinguish between residential investment and other durable goods or between equipment and structures business investment. Or it may be missing a distinction between unemployed and out of the labour force. Call the unmodelled variables z_t. Schorfheide et al (2009)[163] suggest that after solving and estimating the model specified above, we estimate equations linking z_t to $\tilde{s}_t$ as a reduced form

$$z_t = l(\tilde{s}_t, e_t^z).$$

The last equations (often called bridge equations) suffer from many of the same flaws as the old style macroeconomic models from the 1970's. They ignore potentially important cross-equation restrictions. However, by using the states $\tilde{s}_t$ as regressors we have a better chance of avoiding Sims' (1980)[166] critique of the older models for having "incredible" identifying restrictions . In fact, under our assumed DSGE data generating process the equation above is immune to endogeneity bias. So we do not need to find instruments, that are often unreliable for the typical macroeconomic sample size. Given the estimation of the DSGE model, the auxiliary equations for z_t are also easy to estimate if $l(.)$ is linear, involving nothing more than OLS.

Relative to older simultaneous equations,"structural", macroeconomic models these equations for z_t are more robust to regime changes or uncertainty about the values of structural parameters(they are less subject to the Lucas critique). For example, suppose there is a fundamental shift in policy that is not accounted for in the standard policy reaction functions. After re-solving the DSGE model to account for the regime change, we can reestimate the $l(.)$ functions linking z_t to the states on simulated data from the new DSGE system to take into account the effect of the new regime on the dynamics of z_t. A similar technique applies in parameter scenarios in which we examine the robustness of our forecasts and analysis to alternative parameter values. [38] In some contexts we may also be interested in the more flexible specification

$$z_t = a\tilde{s}_t + \rho z_{t-1} + e_t^z.$$

This amounts to an ARMAX model where the X's are estimates of the DSGE state vector.

12.4 Hybrid DSGE-VARX and DSGE-ECM models

This section discusses several hybrid models combining the higher theoretical coherence and economic interpretability of DSGE models with more flexible time series models that may improve short run forecasting and the statistical fit of the overall model. The idea of hybrid models is to nest a DSGE core within a more flexible, less restricted time series model. This has several advantages relative to a pure DSGE model:

[38] Traditional simultaneous equations macroeconomic models that do not relate reduced form parameters to "deep" parameters cannot be adjusted this way.

1) The strength of the DSGE based restrictions on the forecasts and analysis can be varied depending on the objective. For example, we can have looser restrictions for unconditional forecasts, while imposing the DSGE restrictions more tightly to forecast the effect of specific shocks - i.e. for conditional forecasting. The best model in terms of unconditional forecasting and in terms of matching the data typically involves only a partial incorporation of the DSGE restrictions. But for conditional forecasting, imposing DSGE restrictions more tightly can help with the identification of various scenarios and shocks.

2) Hybrid models that mix DSGE dynamics with reduced form models such as VAR's can also be seen as a way of mixing rational choice theory with simpler bounded rationality rules. A simple bounded rationality rule is to set a vector of variables x_t to some habitual level x^h up to some random decision shifter that could reflect various environmental signals that cause the agent to deviate from her habitual decision. At the same time, agents are partially rational and have some awareness of the optimal actions identified for example in a DSGE model.

One could easily imagine that these extra signals are persistent, leading for example to a VAR(1) This leads to a model for the simple rule of thumb x_t^{rot}, mixing with the DSGE dynamics x_t^{dsge}

$$
\begin{aligned}
x_t &= \omega x_t^{dsge} + (1-\omega) x_t^{rot}, \\
x_t^{rot} &= x^h + e_t, \\
e_t &= \rho e_{t-1} + v_t, \\
0 &< \omega < 1, \\
&\Rightarrow \\
x_t^{rot} &= x^h + \rho(x_{t-1}^{rot} - x^h) + v_t, \\
x_t &= x^0 + a x_t^{dsge} + b(x_{t-1}^{rot} - x^h) + u_t,
\end{aligned}
$$

for appropriate definitions of $x^0, a.b, u_t$. The hybrid model can also be reformulated as a version of the popular partial adjustment model

$$x_t = x_t^{rot} + \omega(x_t^{dsge} - x_t^{rot}).$$

In this interpretation actual behaviour is governed by the rule of a thumb with some adjustment each period towards the rational decision.

3) The reduced form time series components of these hybrid models can be reestimated more frequently than DSGE models, allowing a faster incorporation of breaks in the DGP into forecasts. These

models also allow easier incorporation of new shock propagation and amplification channels than the wedge based approach in the previous section. This is mainly because we do not use the information in estimating the new parameters to improve the estimates of the state vector process s_t. The cost of these approaches is related to the same point. Since the estimation of the state vector is separate, we are not using all the information that could help improve our knowledge of s_t.

12.4.1 DSGE-VARX

A prominent example of this class of models is the DSGE-VAR framework of Del Negro et al (2004)[153]. Their main idea is to improve the accuracy Vector Autoregression (VAR) forecasts by using priors based on a DSGE model. To accomplish this, we start with a linear projection of the VAR in (log)- levels

$$y_t = \Sigma_{j=1}^{J} A_j y_{t-j} + v_t$$

on the DSGE data generating process. This gives us a prior $\{A_{j,0}\}_j$ **for the estimation of the BVAR in the 2nd stage. For example, one could simulate** $s = 1...S$ **long samples of the DSGE model under different parameter vectors from the posterior distribution. For each sample** s **we can estimate the VAR. This produces distribution for the VAR parameters based on the DSGE model, which can be used as a prior distribution for the VAR.** Intuitively, the procedure is equivalent to increasing the sample size available for the estimation of the VAR by augmenting the original real world sample with artificial observations coming from a simulation of the DSGE model. A higher number of simulated DSGE based observations amounts to putting more prior weight on the DSGE prior $\{A_{j,0}\}_j$.

The original DSGE-VAR framework uses a time series model which is not the same as that of the original DSGE state space model. The DSGE model in its linearized approximation amounts to a VAR in the state variables. Conditioning on the actual DSGE state variables may improve our forecasts (Fernandez de Cordoba and Torres (2011)). The problem from the perspective of a simple VAR estimation is that part of the DSGE state vector is unobserved. But we can use the estimation of the DSGE model to recover estimates of the unobserved DSGE state vector

$$\begin{aligned} s_{t|T} &= E(s_t|y^T), \\ \text{where } y^T &= y_1, ..., y_T \end{aligned}$$

is the real world sample of observed variables. Now we can estimate the **DSGE-VARX model**

$$\begin{aligned} y_t &= \Sigma_{j=1}^{J} A_j y_{t-j} + B s_t + v_t, \\ s_t &= C s_{t-1} + e_t, \end{aligned}$$

replacing the unobserved s_t with $s_{t|T}$ and normalising $s=0$ at the steady state wihout loss of generality. As before we can use priors from the DSGE model, setting A_0 to 0 and B_0, C_0 to the corresponding matrix from the DSGE equations

$$\begin{aligned} y_t &= hs_t, \\ s_t &= gs_{t-1} + e_t. \end{aligned}$$

For observed variables z_t that were not part of the DSGE model we can again use the methods of Schorfheide et al (2009)[163]. The DSGE-VARX model is similar to the factor augmented VAR model (Reichlin et al (2008)), with DSGE model states acting as factors. In both cases the idea is similar: we can improve the forecasting performance of a low dimension VAR and improve the ability to conduct conditional forecasting/scenario analysis by including a small number of key factors or state variables that drive other variables in the economy.

If we substitute out the state variables in the equations for y_t we can rewrite the DSGE-VARX as

$$\begin{aligned} y_t &= \Sigma_{j=1}^{J} A_j y_{t-j} + Ds_{t-1} + u_t, \\ s_t &= Cs_{t-1} + e_t, \\ D &= BC, \; u_t = Be_t + v_t. \end{aligned}$$

This is a DSGE-VAR in $z_t = (y_t, s_t)$. If we use $s_{t|T}$ as an estimate of the unobserved state variables, we can apply the same methods as Del Negro and Schorfheide (2004) [153] to estimate this system. For forecasting, we can replace s_t by $s_{t|t} = E(s_t|y^t)$.

We can extend the idea behind the DSGE-VARX to nonlinear models by using a nonlinear NVARX of order k,

$$\begin{aligned} y_t &= \Sigma_{j=1}^{J} \Sigma_{k=1}^{K} A_{j,k} \left[y_{t-j}\right]^{\otimes[k]} + \Sigma_{k=1}^{K} D_k \left[s_{t-1}\right]^{\otimes[k]} + u_t \\ s_t &= \Sigma_{k=1}^{K} C_k \left[s_{t-1}\right]^{\otimes[k]} + e_t. \end{aligned}$$

We can use the kth order perturbation solution of a DSGE model to form priors for the coefficient matrices. Here, given the extremely large potential number of parameters, the use of the restrictions from a DSGE prior may be important for incorporating more nonlinearity into models.

12.4.2 Core/Non-Core and DSGE-ECM

The DSGE-VARX model is closely linked to the core/non-core models approach used for example by the BoE (Harrison et al (2005)). The core model is a standard DSGE model. The actual forecast for variables of interest in log- (levels) is

$$y_t = ay_t^{core} + bx_t + cy_{t-1} + v_t,$$

where x_t is a set of other explanatory variables that are not part of the DSGE model. Note that this is essentially an augmented DSGE-VARX with some extra explanatory X variables and degenerate priors for the coefficients on s_t matrix B.

The common practice of combining DSGE model forecasts y_t^{core} with those from another reduced form time series model y_t^{nc}, setting

$$\begin{aligned} y_t &= \theta y_t^{core} + (1-\theta) y_t^{nc}, \\ 0 &< \theta < 1, \end{aligned}$$

can also be cast as a special case of the core/non core framework. Also note that we can rewrite the last equation as

$$y_t = y_t^{nc} + \theta(y_t^{core} - y_t^{nc}),$$

which is similar to the popular error correction (ECM) model. The difference compared to the usual ECM is that in the core/non core approach the analyst puts more effort into modelling the target value y_t^{core} as a dynamic variable that can also change significantly in the short run.

We can combine the core/target values for y_t from a DSGE model with ECM dynamics, to get a DSGE-ECM. The ECM model is usually written in first differences to make it more robust to non-stationnary time series, and the deviations from the core enter with a period lag. This leads to the following specification in (log) differences $y_t - y_{t-1} = \Delta y_t$:

$$\boldsymbol{\Delta}\mathbf{y}_t = \boldsymbol{\alpha} - \boldsymbol{\beta}(\mathbf{y}_{t-1} - \mathbf{y}_{t-1}^{core}) + \boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{y}_t^{core} + \boldsymbol{\Sigma}_{j=1}^{J}\mathbf{A}_j\boldsymbol{\Delta}\mathbf{y}_{t-j} + \mathbf{b}\mathbf{x}_t + \boldsymbol{\varepsilon}_t^{y}.$$

A common simplification is to assume that β, γ and $\{A_j\}$ are diagonal matrices, so that each equation is a univariate regression in its lags and the in the response to its core/target value.

12.5 A comparison to Keynesian structural equation models and structural VAR's

Old style "structural" macroeconomic models, otherwise known as macroeconomic Structural Equations Models (SEM), are still common in the private sector and in some policy applications (e.g WEFA-DRI, Moody's Analytics, Macro Advisers, CBO, OECD). Despite their "structural" label, these models are closer to reduced form models than DSGE models, in the sense of directly specifying economic relations without trying to derive them directly from beliefs, preferences and constraints of agents.

The SEM's in their typical linear form can be seen as a form of structural VAR. Consider the dynamic SEM

$$Ay_t = By_{t-1} + e_t.$$

, where the shocks in e_t are uncorrelated among themselves.This can be turned into the reduced form VAR

$$y_t = A^{-1}By_{t-1} + A^{-1}e_t$$

By imposing restrictions on A, SEM's can identify and estimate certain structural relations such as the effect of household income on consumption. In reverse, taking the VAR

$$y_t = Cy_{t-1} + v_t,$$

the SVAR litterature tries to identify uncorrelated shocks e_t with a better economic/structural interpretation by rewriting the VAR as

$$y_t = Cy_{t-1} + De_t.$$

Multiplying this equation by D^{-1} we get the SEM

$$D^{-1}y_t = D^{-1}Cy_{t-1} + e_t.$$

Setting $A = D^{-1}$, we have found an equivalence between the SEM and the SVAR. In SEM analysis we specify restrictions on A.Typically these are exclusion restrictions, setting certain coefficients to zero. Sims (1980)[166]) has argued that the theory underlying traditional SEM's is not strong enough to justify these exclusion restrictions, making identification fragile. He pioneered the SVAR approach as an alternative with more robust identfifying restrictions. In the SVAR approach, the analyst first estimates the unrestricted VAR or a Bayesian VAR with statistically motivated priors. This is the model that is used for short term unconditional forecasting. For scenario and policy analysis, the VAR is then further restricted by putting more structure on D.

For example, a common strategy is to assume a lag in the response of certain residuals in v_t to some of the components in e_t. This is called recursive identification (Killian, 2013 [127]). Consider the bivariate system

$$\begin{aligned}
z_t &= A_z y_{t-1} + v_{z,t}, \\
x_t &= A_x y_{t-1} + v_{x,t}, \\
y_t &= (z_t; x_t)\,,\ \text{which can be written as} \\
y_t &= A y_{t-1} + v_t.
\end{aligned}$$

A structural decomposition of the shocks yields

$$\begin{aligned}
v_t &= D e_t, \\
e_t &= (e_{z_t}; e_{x,t}),\ Eee' = I.
\end{aligned}$$

We are interested in the impact of the impact of a structural shock $e_{x,t}$ (for example x_t could be a monetary policy tool such as a short run interest rate, or it could be an index of credit condition, in which case $e_{x,t}$ is a credit shock). A common assumption is that

$$D = \begin{pmatrix} D_{zz} & 0 \\ D_{zx} & D_{xx} \end{pmatrix}$$

Identification comes from assuming that z_t does not initially respond to a shock to $x_t : D_{xz} = 0$. This allows to identify $e_{z,t}$. Since this component of e_t is now observable, we can now identify the effect of $e_{x,t}$ on x_t e.g through OLS. The effect on z_t just involves simulating the VAR using our estimates of A and D.

How plausible is the assumption that $D_{xz} = 0$? In many cases it is actually quite restrictive. For example there is no reason to assume that shocks credit conditions or interest rate conditions don't affect GDP or investment inside the quarter. Certainly, it is hard to justify this assumption in a DSGE model. This example shows the promise and limitations that are common in SVAR analysis. On one hand, we want to let "the data speak "by imposing minimal plausible restrictions. On the other hand, such restrictions are hard to find.

SEM's and SVAR's have similar advantages and disadvantages relative to DSGE models. They are more data driven and less restricted, with the SVAR being less restricted than the SEM. This can improve unconditional forecasting especially in the short -term, but it may reduce the ability to answer certain questions that require more economic theory based identifying restrictions, such as the effect of certain shocks or changes in economic regimes and institutions. In general, there is a tradeoff in which imposing less economic structure on models can improve forecasting and , but may reduce the reliability of scenario and policy analysis which require a better understanding of the structure of the economy.

In the previous section we took the perspective of starting from a theoretical DSGE analysis, and extending the theoretical structure with extra data-driven, atheoretical, dynamics to improve the statistical fit of the model to the actual economy. Here, we proceed in the opposite direction by assuming the analyst starts with an SVAR or a SEM. From this perspective, we discuss how more theoretical DSGE analysis can enhance the SEM or SVAR.

First, DSGE models are better at distinguishing agents' expectations about the future from adjustment costs. In contrast in a SEM or SVAR it can be hard to distinguish adjustment costs from the usually assumed backward looking expectations (though some models allow for rational expectations, especially in the UK (e.g the NIGEM model)). In practice the adjustment costs in DSGE models may not be structural: they may simply be proxies for adaptive expectations or other forms of delays in absorbing information (e.g due to learning), but the DSGE model should still have an advantage in identifying agents expectations' effects on the economy.

Second, DSGE models suggest new variables that should enter into the macro model and that should drive the dynamics, the state variables s_t. These typically include capital stocks and shock processes such as for TFP and business or consumer confidence. The DSGE state variables have better economic interpretation than those in a SVAR. Furthermore, from the DSGE model perspective, the correct linear approximation to the economy is

$$
\begin{aligned}
y_t &= Ay_{t-1} + Ds_{t-1} + u_t, \\
s_t &= Cs_{t-1} + e_t, .
\end{aligned}
$$

SVAR's and SEM's typically ignore s_t, setting $D = 0$. But omitting these extra variables can lead to significant misspecification bias if the DSGE is a reasonable approximation of key aspects of the economy (see for example MacGrattan et al. (2008)[51] and Killian (2013) [127].The state variables or factors driving the economy may be unobservable, but DSGE models can help estimate them.These estimates can be incorporated into SVAR or SEM analysis through the hybrid DSGE-VARX models discussed above, for example by adding the state variables' in-sample estimates from a DSGE model into the estimation of the SVAR or SEM.

Third, DSGE models provide cross equation restrictions on coefficients that come from the imposition of aggregate consistency/market clearing and rational expectations (RE). SEM's are less restricted in estimation. More formally, taking the linearised DSGE model

$$
\begin{aligned}
y_t &= g(\theta)s_t, \\
s_t &= f(\theta)s_{t-1} + e_t.
\end{aligned}
$$

the matrices g and f depend on a smaller number of structural parameters θ that affect simultaneously the coefficients across different equations. The SEM and SVAR approaches typically do not impose such a link between the coefficients in different equations.

A smaller number of restrictions can improve unconditional forecasting if the restrictions are misspecified. On the other hand a smaller number of restrictions is likely to generate more imprecise parameter estimates increasing forecast error variance. In some cases the lack of cross equation restrictions can even increase misspecification if the modelers respond to the large parameter confidence intervals in the unrestricted equation system by imposing more zero restrictions based on t-statistics below a certain threshold value. More generally, ignoring cross equation restrictions can lead to much wider confidence bands, encouraging users to ignore parameter uncertainty altogether. For conditional forecasting, without cross equation restrictions the user could accidentally change coefficients in one equation (possibly setting it to zero) as a part of scenario analysis while forgetting that the resulting change in the economic structure would probably change the behaviour of other equations.

Fourth, DSGE models by definition are consistent in terms of satisfying aggregate accounting constraints. Unconstrained SEM's or SVAR's may violate these constraints. DSGE models can suggest restrictions that improve the aggregate consistency of SEM's or SVAR's, or they can help investigate the robustness of SEM/SVAR results to aggregate consistency requirements.

Finally DSGE models are useful in clarifying the medium and long run equilibrium that SEM's dynamics are supposed to converge to over time. The typical wage or price equation in a SEM is based on partial adjustment of prices in response to a disequilibrium between supply and demand. A generic example of this theory is to specify wages as

$$\begin{aligned} w_t &= w_t^0 + \lambda(n_t^d - n_t^s), \\ \lambda &> 0. \end{aligned}$$

w_t^0 is the wage level that would prevail in a competitive equilibrium with

$$n_t^s = n_t^d.$$

A complete derivation of the market clearing w_t^o would require posing and solving an RBC model. Most SEM's in operational use have a much simpler underlying model of the flexible price equilibrium. To the degree that movements in the flexible price equilibrium only matter in the long run, this may be a reasonable simplification. But DSGE analysis suggests that flexible price or potential output can be quite volatile even in the short run. In that case, analysing the RBC model underlying the flexible price equilibrium should improve forecasting and scenario analysis. Again the DSGE predictions can be partially incorporated into the analysis through hybrid models such as a DSGE-VAR or the related partial adjustment or core/non core model

$$\begin{aligned} x_t &= x_t^{sem} + \theta(x_t^{dsge} - x_t^{sem}), \\ 0 &< \theta < 1. \end{aligned}$$

A key advantage of SEM's is that they are easier to solve and estimate than DSGE models. This advantage has diminished significantly for linear DSGE models with current computer speeds and solution algorithms. Nonlinear DSGE models are significantly harder to handle, at least if we also want to estimate the model. But traditional macroeconomic models are themselves linear, so they cannot be directly compared to nonlinear DSGE models. Further increases in computing power should make this increasingly a moot point, with nonlinear models becoming more commonly used. Programs such Dynare have now made simulation of calibrated nonlinear DSGE model approximations almost routine (using perturbation methods for now, though higher accuracy projection methods are likely to become much faster in coming years with more recent developments in solution methods (e.g Judd et al (2011)) and greater accessibility of multi core computers.

As for estimation, with modern computers linear DSGE model estimation can be achieved reasonably well within the decision making cycle of a policy or private sector consulting institution, as long as we do not insist on reestimating the model every quarter (the core DSGE models at policy institutions are usually reestimated once a year). But a model whose parameters change significantly based on adding an extra quarter of estimation is arguably highly misspecified, and the parameter reestimates should be treated with caution, especially if the goal is to use this as a structural model for scenario and policy analysis (as opposed to unconditional short term forecasting). Furthermore, the available evidence suggests that more frequent re-estimation of DSGE models leads to only minor increases in forecast accuracy (Kolasa and Rubaszek, 2014).

13 A monetary structure underlying the cashless economy

The real business cycle model (and many of its New Keynesian offspring) have been criticised for ignoring money. The description of the cashless economy above is in fact compatible with the existence of money required for transactions, as long as the creation of that money is unrestricted (e.g. perfect electronic private money). Suppose we add a constraint that says a part of expenditures must be financed by showing an evidence of available purchasing power either through notes or electronic evidence using a bank card. Call the previously cumulated stock of these (electronic) notes money, M_t. The cash in advance constraint is

$$\begin{aligned} s_{c,t}c_t(1+\tau_{c,t}) + s_{b,t}(b_{t+1}-b_t) &\leq M_{t+1}p_t^m \\ 0 &\leq s_{c,t}, s_{b,t} \leq 1. \end{aligned}$$

$s_{c,t}$ or $s_{b,t}$ are often known as the velocity of money. Since money is required in some transactions, it has a positive value in an equilibrium with trade. [39] The stock of money held by households during the period is

$$M_{t+1} = M_t + T_t^m.$$

Perfectly competitive financial intermediaries and the government create new money $p_t^m T_t^m$ on demand and distribute it to households for free. The budget constraint of households is modified to

$$c_t(1+\tau_{c,t})+b_{t+1}+p_t^m M_{t+1} \leq b_t R_t efp_{h,t}+p_t^m M_t+w_t(1-\tau_{nst})n_{s,t}+d_t+p_t^m T_t^m+T_t,$$

Since the household knows that $M_{t+1} = M_t + T_t^m$, this budget constraint reduces to the previous one in the cashless economy. The cash in advance constraint does not bind, and therefore the economy reduces to the Real Business Cycle economy except that we have in parallel another asset called money and a definition of inflation. Without any further structure, the process for inflation is indeterminate. For example, we could determine expected inflation in an economy where the central bank exercises complete control over private sector nominal short run interest rates and follows a an interest rate rule as in modern New Keynesian models (e.g. Gali ch. 3 [91]). Alternatively, we can determine inflation in an environment where budget surpluses do not adjust to ensure fiscal solvency of the government, but we exclude explicit government default. In that case, if debt is nominal then the price level adjusts to ensure solvency (see Cochrane (2005) [57] for more details on the fiscal theory of the price level).

How do we get such a radically different conclusion from the classical cash in advance model (e.g. Walsh ch. 3 (2003)[177]) ? The key is that the textbook cash in advance model assumes that T_t^m is exogenous or that households do not perceive any endogenous response of financial institutions or the government in adjusting T_t^m to satisfy money demand M_{t+1}. Here,we have the other extreme of completely endogenous inside money. The real world is an intermediate case, but many would argue that the modern payments system is much closer to the endogenous inside money case. Formal empirical investigation of the importance of imperfect payments systems for business cycles is subject to big endogeneity problems. But if we do not think the cost of electronic transactions relative to the use of paper currency is a major cause of business cycles fluctuations in

[39] Whether this must be a legal imposition or self fulfilling belief of agents that money is acceptable as a proof of purchasing power in daily transactions while other assets are not is not so important in the current setup. A legal requirement does eliminate the bad sunspot equilibrium in which people believe money has no value and trade collapses.

households' and firms desired expenditure or an important factor in long run income differences between countries, then it follows that the cashless economy is a good approximation for our purposes. [40]

13.1 Government versus private issued money, and seigniorage

We can extend the model above to distinguish between private and government money. This extension allows us to introduce the government's budget constraint and to discuss the possibility of seigniorage, the revenue gain of the government from issuing from financing itself with interest free money instead of bonds. Let M^p_{t+1} be private money and M^g_{t+1} be government (central bank) issued money. Unlike the private sector, we assume that the government uses changes in its money supply to finance its spending (if government money supply has a component that simply adjusts passively to demand like private money, then M^g_{t+1} represents the portion of government money supply that is used as a source of financing). The government's budget constraint is

$$S^g_t + b^g_{t+1} + p^m_t(M^g_{t+1} - M^g_t) = b_t R^g_t,$$

where b^g_t are government bonds and S^g_t is the government's primary surplus (taxes minus non interest rate spending and transfers). Using for simplicity $b_{t+1} = 0$ for private sector bonds, the representative household's budget constraint is now

$$\begin{aligned} c_t(1+\tau_{c,t}) + b^g_{t+1} + p^m_t[M^g_{t+1} + M^p_{t+1}] &\leq b_t R^g_t + p^m_t[M^g_t + M^p_t] + w_t(1-\tau_{nst})n_{s,t} + d_t + p^m_t T^m_t + T_t, \\ M^p_{t+1} &= M^p_t + T^m_t. \end{aligned}$$

[40] This statement suggests that the cashless economy model will be misleading in important dimensions for economies with very high inflation rates (for example Latin America in the 1980's, the former USSR in the early 1990's, Germany in the early 1920's). There is little doubt that the transaction and information costs due to hyperinflations cause significant economic damage. Intuitively, unless the money creation technology is completely flexible (e.g perfect electronic money creation), at very high levels of inflation it becomes harder to fulfill the need for M_t. At this point cash in advance constraints may start to bind. At the same time, the difficulty of determining the price level in the fully endogenous money economy highlights the potential for high inflation to emerge from self fulfilling prophecies. It is quite possible that some if not most of the highest observed inflation levels were not just a matter of the government trying to raise seignorage revenue as in the standard analysis of seignorage, but also caused by the money supply responding to people's expectations of higher prices.

The completely endogenous money limit in can also be derived in models of price rigidities due to search frictions as in Williamson and Wright (2010) [182]. First, one has to change the timing of Arrow Debreu and Burdett-Judd markets in their model so that agents get to carry M_{t+1} into time t BJ market . Second one assumes as here that financial intermediaries can create new money on demand. In that case, we get a theory of price rigidity that matches micro pricing facts well while preserving monetary neutrality and many of the conclusions of RBC models.

[41] The cash in advance constraint is

$$s_{c,t}c_t(1+\tau_{c,t}) + s_{b,t}(b_{t+1} - b_t) \leq [M^g_{t+1} + M^p_{t+1}]p^m_t.$$

The first order condition with respect to private money M^p_{t+1} still implies that the cash in advance constraint does not bind: its Lagrange multiplier is zero. Assuming an interior solution, the first order condition for M^g_{t+1} and b^g_{t+1} are

$$\begin{aligned} 1 &= E_t \frac{\beta_{t+1}\lambda_{t+1}}{\beta_t\lambda_t}\frac{p^m_{t+1}}{p^m_t}, \\ 1 &= E_t \frac{\beta_{t+1}\lambda_{t+1}}{\beta_t\lambda_t} R^g_{t+1} \end{aligned}$$

This condition leads to a contradiction whenever $R^g_{t+1} > \frac{p^m_{t+1}}{p^m_t}$, implying that $M^g_{t+1} = 0$ and seigniorage revenue is also zero. This is always the case with nominal risk-free bonds and positive nominal interest rates. When nominal interest rates are zero, money and safe nominal bonds are perfect substitutes. In this case expansion in government money supply can reduce its reliance on debt, but because the nominal interest rate is zero there is still no seigniorage revenue. When the central bank pays interest on reserves, money in the form of reserves and short-term government bonds are again (almost) perfect substitutes. In this case again, expanding central bank reserves reduces dependence on government debt financing, but there is no seignorage revenue.

The situation is more complicated when government default risk has reached levels that make even short run government debt risky, so that we cannot take R^g_{t+1} out of the expectations in the first order condition. However, in this case the inequality $E_t R^g_{t+1} > E_t \frac{p^m_{t+1}}{p^m_t}$ is still quite likely to hold. Using the standard covariance formula, $M^g_{t+1} > 0$ if and only if $cov(\frac{\beta_{t+1}\lambda_{t+1}}{\beta_t\lambda_t}, \frac{p^m_{t+1}}{p^m_t}) > cov(\frac{\beta_{t+1}\lambda_{t+1}}{\beta_t\lambda_t}, R^g_{t+1})$. A priori, there is no reason for the inverse of inflation to covary more with the stochastic discount factor than the real return on short term government debt.

Governments in economies with highly advanced payments systems obtain a small amount of seigniorage revenue, since no modern economy is yet fully at the cashless limit (see Walsh [177], ch. 4). There could also be some agents that are limited in their access to bond markets, for whom money and bonds are not perfect substitutes as saving instruments. Those agents would be willing to hold additional government money despite the lower rate of return. In a modern financial system, one would suspect this channel has limited quantitative effects.

[41] Here we assume $efp_{h,t+1} = 1$, otherwise we would have to explicitly introduce a wedge in the representative household's government bond Euler equation to capture the underlying heterogeneity in the economy between households with positive and zero government bond holdings.

Overall, the results above cast doubt on the possibility of using seigniorage as a significant funding source in modern economies. At the very least they suggest that attempts to raise anything beyond rather small amounts of seigniorage revenue are bound to lead to extremely high inflation, which may cause more damage to the economy than the problems higher seigniorage is supposed to solve (see Reis (2013) [158]for a formal discussion of the limits to the central bank's ability to generate revenue in a modern economy, with an application to central bank asset purchases and the lender of last resort role). [42]

13.2 Phillips curves with flexible prices

The Phillips curve in its most basic form is a positive correlation between inflation and the level of economic activity. Recent data has put into question the robustness of this relation (Uhlig (2011) [175]), and its ability to forecast inflation better than purely time series models such as a random walk has also been subject to strong debate (Stock and Watson (2008)). In this section we take the Phillips curve relation as given. The most popular explanation for the Phillips curve is based on nominal price rigidity. We review that explanation in a latter section, but for now it is worthwhile to ask if there are modifications of the flexible price framework that can lead to a Phillips curve, or a similar relationship. I start with a simple mechanism that generates Phillips curve effects from the interaction of typical monetary policy rules, flexible price real interest rates generated by the RBC model and the Fisher relation linking nominal and real interest rates. Suppose the central bank controls the nominal short run

interest rate and follows a Taylor rule linking the interest rate to output and inflation,

$$\begin{aligned} i_{t+1} &= i_{ss} + a y_t + b\pi_t, \\ a, b &> 0. \end{aligned}$$

Under rational expectations,

$$\begin{aligned} i_{t+1} &= r_{t+1} + E_t\pi_{t+1} + rp^{\pi}_{t+1}, \\ \pi_{t+1} &= E_t\pi_{t+1} + v^{\pi}_{t+1}. \end{aligned}$$

rp^{π}_{t+1} is an inflation risk premium compensating investors for the inflation risk of non-indexed nominal bonds. A common approximation is the Fisher relation that assumes we can ignore the inflation risk premium and set $rp^{\pi}_{t+1} =$

[42] A government that is desperate for seignorage could restrict the payments system and make it less efficient on purpose in order to create binding cash in advance constraints. This would allow it to increase its seignorage revenue, but such a policy of distorting payments systems on purpose can cause significant damage to the economy. At the very least, this policy would have to be carefully evaluated in welfare terms before using it to raise government funding.

0.Equating the 2 equations under this assumption, we get that

$$\pi_{t+1} = i_{ss} + ay_t + b\pi_t - r_{t+1} + v^{\pi}_{t+1}.$$

The equation above implies that inflation will be positively related to lagged output and past inflation. While this "Phillips Curve" is quite different from its Keynesian counterpart, Dittmar et al (2005) show it provides a good fit to the data. In comparison to the tradional Keynesian sticky prices Phillips curve, it provides quite a different interpretation of the response of inflation to shocks. In Keynesian analysis, a recession causes producers to respond by gradually reducing nominal prices, causing a positive correlation between inflation and output. In the flexible price "Phillips curve", central bank reductions in the policy interest rate i_{t+1} in response to a recession lead to lower inflation expectations and eventually lower inflation if the equilibrium real interest rate r_{t+1} declines by less than i_{t+1}.

Another explanation for the Phillips curve with flexible prices relies on countercyclical demand for more liquid assets such as money. For simplicity we go back to the RBC model without capital. We complicate the environment by assuming there are transaction costs in adjusting bond holding and by introducing another financial instrument that can be adjusted without any costs but yields a lower rate of return $R^m_t < R_t$. Call the second asset "money", or deposits. The budget constraint of the representative household becomes

$$\begin{aligned} c_t(1+\tau_{c,t}) + b_{t+1} + g(b_{t+1} - b_t) + M_{t+1} &\leq b_t R_t efp_{h,t} + M_t R^m_t + w_t(1-\tau_{nst})n_{s,t} + d_t + T_t, \\ g'(.) &> 0, g''(..) > 0. \end{aligned}$$

The first order conditions are modified by

$$\begin{aligned} b_{t+1} &: \lambda_t[(1+g'(b_{t+1}-b_t)] = E_t \frac{\beta_{t+1}}{\beta_t}\lambda_{t+1}[efp_{h,t+1}R_{t+1} + g'(b_{t+2}-b_{t+1})] \\ M_{t+1} &: \lambda_t = E_t \frac{\beta_{t+1}}{\beta_t}\lambda_{t+1}R^m_{t+1}. \end{aligned}$$

Under certain conditions both money and bonds will be desired by households despite lower return on money. This is due to the lower transaction costs of money relative to bonds. Most monetary models focus on outside money issued by the government, but in modern developed economies this is probably less relevant than inside money that is in zero net supply (keeping with our closed economy, no government baseline). Note that this notion of money as a liquid store of value, in the sense that there are no costs to changing your net portfolio position $m_{t+1} - m_t$, is different from another common definition in deep microfoundations models of money such as those in Williamson and Wright (2010) [182] or in more traditional cash in advance models. Those models define money as an asset whose only value is from the possibility of resale: i.e. it is a rational bubble which is compatible with transversality conditions because it loosens credit constraints in some states of the world. [43] Here, we do not have to appeal to a resale market. "Money" in our model has value because of a guarantee of a private agent (e.g. a bank) to always make available $-(M_{t+1} - M_t) > 0$ for other purposes to any household that desires to spend that amount on other goods or assets. [44]

In a monetary economy neither R^m_{t+1} nor R_{t+1} are risk-free in real terms anymore, because of inflation risk. While this is potentially important, one can understand many issues while ignoring this risk and using our certainty equivalence approximation. This yields

$$\lambda_t \simeq \frac{E_t\beta_{t+1}}{\beta_t} E_t\lambda_{t+1} E_t R^m_{t+1}.$$

Define the price of m_t in terms of goods as p^m_t, such that $E_t R^m_{t+1} = \frac{1}{p^m_t}$. The price of goods in terms of m_t is $p_t = \frac{1}{p^m_t}$. The inflation level is $\frac{p_t}{p_{t-1}} = 1+\pi_t$. Using the commonly used isoelastic utility function over consumption $u(c_t) = \frac{c_t^{1-\sigma}}{1-\sigma}$, we obtain

$$\begin{aligned} E_t R^m_{t+1} &= p_t \simeq \frac{\beta_t}{E_t\beta_{t+1}} \frac{\lambda_t}{E_t\lambda_{t+1}} \simeq \frac{\beta_t}{E_t\beta_{t+1}} \left(\frac{E_t c_{t+1}}{c_t}\right)^{\sigma}, \text{ i.e.} \\ p_t &\simeq \frac{\beta_t}{E_t\beta_{t+1}} \left(\frac{E_t y_{t+1}}{y_t}\right)^{\sigma}. \end{aligned}$$

This equation says that ceteris paribus inflation increases with higher expected output, decreases with current output, increases with positive current intertemporal shocks (either higher β_t or lower $E_t\beta_{t+1}$). If a boom is associated with a big increase in $\frac{\beta_t}{\beta_{t+1}}$, which is a proxy for an aggregate demand shock, and a gradual increase in output so that $\frac{E_t y_{t+1}}{y_t}$ increases, we'll get higher inflation as predicted by a classical Phillips curve. If the current increase in output dominates, we can get lower inflation in a boom, as predicted for certain supply shocks by modern New Keynesian Phillips curves. If we allow for capital investment, inflation increases as long as the boom raises the growth rate of

[43] In modern economic theory a bubble is defined as any asset whose value exceeds the buy and hold forever value. A bubble is rational to the degree that it emerges in a Rational Expectations equilibrium. The pure fiat money studied in modern monetary theory can be seen as the limit of a very long horizon zero coupon bond whose main value comes from the possibility of resale to other agents. The transversality condition limits the growth rate of the value of money, or equivalently the deflation in the price of goods in terms of money.

[44] Ljunqvist and Sargent (2004) [140], chapter 17, have a similar interpretation of perfectly liquid risk-free debt as inside money. Barlevy (2007) [18] has a particularly clear exposition of money as a rational bubble, and related issues with asset price bubbles. Kiyotaki and Moore (2008) [129] develop a model with money as a bubble, where we also get what looks like a Phillips curve based on a flight to liquidity in response to financial shocks. Perez (2010) explores similar arguments with government debt as a source of liquidity.

consumption. Vice versa, in a recession this model can generate lower inflation as agents in the economy increase their desired holdings of the safest and most liquid assets. This phenomenon often goes under the name of a "flight to liquidity" or a "flight to quality".

One problem with a direct interpretation of the inflation rate here as the inflation rate of goods in terms of money is that even transaction deposits have usually offered positive real rates of return, whereas money has usually offered negative real rates of return (inflation is usually positive). Nevertheless, to the degree that there's some link between the rates of return on money and other liquid assets such as deposits, the arguments for a Phillips curve relation above would also apply to money.

14 Nominal Price Rigidity and New Keynesian Economics

In a real business cycle model only real quantities and relative prices are important. Conventional monetary policy through interest rate announcements is neutral. [45] **To give a prominent role to conventional monetary policy we need to introduce frictions slowing down the adjustment of nominal prices to shocks. New Keynesian economics is the standard framework for studying these nominal price rigidities. It can be seen as a microfoundation for the standard central banker view of the macroeconomy, and for the common perception of business cycles as being caused by disequilibria between supply and demand due to sluggish price adjustment.** It also implies a more important role for intertemporal or aggregate demand shocks in business cycles than the RBC framework, and can justify a higher positive effect on output of temporary increases in government spending.

The key assumptions of New Keynesian economics is that
1) firms post prices in terms of a numeraire good (called money),
2) firms have significant market power, and
3) prices posted in terms of money (nominal prices) are only partially readjusted in response to demand and supply conditions.

As result, demand shocks (intertemporal shocks) play a much more important role in business cycles than in an RBC economy.

[45] Central bank policies do matter in an RBC economy with financing frictions if they alleviate credit constraints. This will be discussed in the section on unconventional monetary policy.

14.1 The microfoundations of the positive relation between inflation and the level of economic activity (the Phillips curve)

Consider a boom that raises aggregate demand. Sticky nominal prices mean that the rising aggregate demand will be partially accommodated by firms,leading to rising real marginal costs for all firms (for example because wages increase as labour demand increases). Total nominal price rigidity implies zero inflation. Partial price rigidity implies that firms that flexibly adjust their prices will want to raise nominal prices relative to the fixed price firms (for which there's zero inflation by definition) in an attempt to preserve the desired markup over the increasing marginal cost. As a result we get positive inflation for firms that reset prices. Since the aggregate price level is a weighted average of fixed price firms and firms that reset prices, aggregate inflation will rise. This establishes a positive relation between inflation and marginal costs that is at the heart of the New Keynesian Phillips curve.

Forward looking firms that adjust their prices take into account that they may not readjust their prices again for some time. So price adjustment in a boom also responds positively to future marginal costs increases. Since future marginal cost increases lead to higher future prices we get a positive relation between current inflation and future expected inflation rates. Under commonly used assumptions, we can link the marginal cost measure in the economy to an output gap

$$\ln y_t - \ln y_t^n.$$

Here the output gap is defined as output relative to the "natural" output level under flexible prices and wages, y_t^n. This leads to the new Keynesian microfoundation for an expectations augmented Phillips curve

$$\begin{aligned} \pi_t &= f^\pi(\{E_t\pi_{t+j}\}_{j>0}, \ln y_t - \ln y_t^n) = f^\pi(\{E_t(\ln y_{t+j} - \ln y_{t+j}^n)\}_{j\geq 0},), \\ &\quad \text{where } \pi_t \text{ is inflation, } y_t^n \text{ is the flexible price output level,} \\ f_1^\pi &> 0,\ f_2^\pi > 0,\ f_3^\pi > 0. \end{aligned}$$

It may helpful at this stage to derive this relation from another direction. Consider a negative aggregate supply shock. To be more specific, suppose firms feel more pessimistic in the flexible price equilibrium so that the labour demand wedge $\tau_{nd,t}$ increases (business sentiment deteriorates as modelled in Angeletos and Lao (2012) [11]). Now flexible price firms are tempted to decrease hiring and production. To lower their sales, they want to raise prices relative to sticky price firms. As a result, inflation increases. Furthermore, since sticky price firms do not reduce their production as much as flexible price firms, the output gap increases, while the natural output level y_t^n declines. This once again establishes a positive relation between inflation and current or future output gaps.

Under commonly used models of nominal price adjustment frictions (e.g Gali ch. 3 (2007) [91])and a loglinear approximation around 0 long run inflation, the Phillips curve simplifies to

$$\begin{aligned}\pi_t &= aE_t\pi_{t+1} + b(\ln y_t - \ln y_t^n) \\ 0 &< a \leq 1, b > 0.\end{aligned}$$

[46].

In most models, the natural output level is affected by various frictions and it can be quite volatile. This makes it different from a long run trend or from output in a frictionless neoclassical model. New Keynesian models usually take a cashless economy limit and assume that the central bank can fully control the most relevant short run nominal interest rate in the economy.

14.2 Determining interest rates with sticky prices: monetary policy enters the scene

To close the model, we specify an interest rate rule for the central bank. This rule typically assumes that the central bank adjusts the nominal interest rate i_{t+1} to stabilise the output gap $\ln y_t - \ln y_t^n$ and inflation. More formally, the interest rate rule leads to a unique stationary solution for the output gap and inflation. A typical example of such a rule is

$$\begin{aligned}i_{t+1} - i_{ss} &= a_i(\ln y_t - \ln y_t^n) + b_i(\pi_t - \pi_{ss}) + \varepsilon_{i,t} \\ \text{where } a_i, b_i &> 0,\end{aligned}$$

$\varepsilon_{i,t}$ **is an IID monetary policy shock,and** x_{ss} **is the steady state of balanced growth path value of any variable** x_t.

To solve the model, we first find the flexible price equilibrium. Then we solve the original system of equations with rigid prices, taking as given the previous solution for the flexible price equilibrium. If the central bank responds

[46] From a practical perspective, in a model with a large set of real frictions like the one we have been working with, it is common to add another group of retailing firms that are only subject to nominal price rigidity. The other firms sell their output to these retailers, taking the price of their output as given, while being subject to all the other frictions. This generats a Phillips curve with a markup of retail price over wholesale prices which are a function of the output gap (see Bernanke et al's [28]financial accelerator model for one of the 1st uses of this setup). Purely foreward looking Phillips Curves frequently imply too little inflation persistence. To deal with this, many models add lags of inflation by assuming that non optimising firms adjust prices as a function of past inflation. This assumption has been criticised as ad-hoc and unrealistic. Lagged inflation related terms emerge more naturally in models of imperfect information such as the sticky information model of Mankiw and Reis (2010).

aggressively enough to inflation or output gaps, higher price flexibility is stabilising in the sense that output gaps dynamics become less important relative to the flexible price output dynamics. In this case we have a unique equilibrium in which deviations between sticky price output and flexible price output die out over time in response to a shock (formally, the output gap is stationary). If the central bank doesn't adjust nominal rates sufficiently to inflation or the output gap we can have stationary sunspot shocks affecting the output gap. With a weak response of interest rates or if we've hit the zero lower bound on nominal interest rates, price flexibility can lose its stabilising power in the New Keynesian model: it can even become destabilising at intermediate levels (see Eggertson et al (2012) [29]). We will have more to say about this possibility in the next section.

Another version of the interest rate rule allows the central bank to react positively to changes in the natural interest rate, defined as the rate of interest that would emerge absent price and wage rigidities. This leads to

$$\begin{aligned}
i_{t+1} - i_{ss} &= a_i(\ln y_t - \ln y_t^n) + b_i(\pi_t - \pi_{ss}) \quad + c_i(r_{t+1}^n + E_t\pi_{t+1} - [r^n + \pi_{ss}]) \\
\text{where } a_i, b_i, c_i &> 0.
\end{aligned}$$

This extension is frequently dismissed because of of doubts regarding the central bank's ability to observe the natural interest rate. But a priori the central bank is not less capable of tracking this variable (with some measurement error) than other agents in the private sector, and potential output can be even harder to measure. Discussions in monetary policy circles about the natural rate of interest r_t^n are quite common, and a lot of the information gathering and analysis of central bank economists can be understood as trying to measure r_t^n. At any rate, Curdia et al (2011)[63] provide some informal evidence that central bankers have a notion of the natural interest rate in mind when taking monetary policy decisions. The also find that interest rate rules with natural interest rate tracking provide a better fit to the data in the US.

14.3 The effect of sticky prices on business cycles

14.3.1 The baseline model with fixed capital

The New Keynesian economy would be almost indistinguishable from the flexible price economy if the central bank set sufficiently high a_i and b_i in the interest rate rule of the previous section. In that case policy would virtually eliminate any differences between the flexible price and the sticky price equilibria. Otherwise, we can get significant gaps between sticky price and flexible price output levels (the

output gap). [47] The difference between the flexible price RBC environment and the New Keynesian environment is easiest to see in the extreme case of a fixed central bank nominal interest rate target, fixed nominal prices (complete nominal rigidity) and a fixed capital stock and durables stock (as in the standard textbook exposition of New Keynesian economics) such as Gali (2007) [91]). [48]

Consider shock to households' discount factor β_{t+1} or to the external finance premium $efp_{h,t+1}$ that increases desired savings. Absent any link between this shock and the efficiency of the search process in product markets or other intratemporal wedges, we have already seen that this shock will not affect flexible price output. Since $c_t = y_t$ and households want to save more and consume less, in equilibrium real interest rates will fall such that households desired consumption matches the unchanged y_t. A crude way of describing this is that aggregate supply determines output and financial markets adjust to ensure equilibrium.

$$\begin{aligned}
\Delta\beta_{t+1+j} &> 0, \Delta efp_{h,t+1+j} > 0,\ \Delta n^d_{t+j} = 0,\ \frac{\Delta n^s_{t+j}}{\Delta c_{t+j}} \leq 0, \\
&\Rightarrow \Delta n_{t+j} = 0 \Rightarrow \Delta y_{t+j} = 0, \Delta c_{t+j} = 0, \Delta R_{t+1+j} < 0.
\end{aligned}$$

Now consider the effect of the same shock in the fixed price and the fixed nominal interest rate economy.

1) By assumption, the real interest rate is fixed. Meanwhile, in the New Keynesian economy firms agree to satisfy any demand at a fixed nominal price.

2) Since the real interest rate cannot adjust, households reduce their consumption and firms adjust by reducing production such that $y_t = c_t$.

3) Labour demand declines, and wages go down to induce the representative household to reduce labour supply.

4) The decline in output reduces desired saving such that credit markets remain in equilibrium. Here, intertemporal shocks

[47] This implies that a RBC model can be a good approximation if the central bank successfully eliminates most of the output gap, which is a desirable policy objective under some conditions. In this case, identifying price rigidity and interest rule parameers from aggregate dynamics may be quite difficult(see Cochrane [58]).

[48] Our analysis below would be similar with variable durables investment, as long as the level of price rigidity in durable goods production is similar to that in nondurables production. In reality, house prices for example are significantly more flexible than services' prices, generating comovement problems between durables and non durables. For further discussion and model extensions that solve this issue see Carlstrom and Fuerst (2006), Bouakez et al (2008).

determine aggregate demand, and aggregate supply adjusts to ensure equilibrium.

This mechanism can be seen as a modern version of the old Keynesian paradox of thrift, in which higher desired saving can reduce output in the short run.

$$\begin{aligned}\Delta\beta_{t+1+j} &> 0, \Delta efp_{h,t+1+j} > 0,\ \Delta R_{t+1+j} = 0 \\ &\Rightarrow \Delta y_{t+j} < 0,\ \Delta c_{t+j} < 0 \Rightarrow \Delta n_{t+j} < 0.\end{aligned}$$

Similar reasoning explains why sticky prices can generate business cycles in response to **news shocks**. Positive news about higher future income act like a positive demand shock, increasing desired spending now. With imperfect real interest rate adjustment due to sticky prices, this leads to higher current output.[49] Note that the response of the economy under sticky prices can be decomposed into its response under a monetary policy that mimics the flexible price outcome by targeting the flexible price real interest rate and a monetary policy shock that causes a deviation of monetary policy from targeting the flexible price real interest. This perspective explains the importance for Keynesian economists of examining monetary policy shocks, even if pure monetary policy shocks typically account for a small share of the variance of output.

- Consider on the other hand a positive front loaded productivity shock in the benchmark fixed capital economy. In the flexible price economy output increases and the real interest rate falls as households try to smooth consumption through higher savings. In the sticky price economy, if the central bank does not reduce interest rates sufficiently, the real interest rate is too high and output is lower in the short term relative to its flexible price level.

- The general idea emerging from the New Keynesian literature is that such productivity shocks generate countercyclical output gaps, because the central bank rarely adjusts monetary policy sufficiently to completely stabilise the output gap. This generates countercyclical inflation for TFP shocks, which allows New Keynesian theory to explain stagflation episodes such as the 1970's in the developed economies.

[49] However, the response of inflation to a positive news shock need not be positive in the sticky price model. The news shock increases the output gap which tends to raise inflation. For the case of a trend stationnary news shock the expectation of a future increase in TFP also generates expectation of lower future inflation. Because of forward looking price setting behaviour, this reduces firms' desired prices today, lowering inflation (see Kobayashi (2008)).

14.3.2 The New Keynesian model with flexible capital

In contrast to the central role of flexible capital in RBC analysis, the standard New Keynesian analysis of business cycles often assumes a fixed capital stock as a benchmark. This is unfortunate, since flexible investment is important in quantitative New Keynesian monetary policy models and in traditional Keynesian thinking about business cycles. This section will attempt a more theoretical analysis of the interaction between nominal price rigidity and investment.

Relative to a flexible price economy, the key difference is that in Keynesian sticky prices economy investment is affected by aggregate demand constraints on future output,

$$g_t z_t f(k_t, n_t) \leq \bar{y}_t,\ (\mu_t^y).$$

Keeping the consumption good as the numeraire, the first order condition for investment becomes

$$k_{t+1} : p_{k,t}(1+\tau_{k,t}) = E_t \frac{\beta_{f,t+1}}{\beta_{f,t}}[g_{t+1} z_{t+1} f_{k,t+1}(1-\mu_{t+1}^y)+(1-\delta)(1+\tau_{k,t+1})p_{k,t+1}].$$

A tighter aggregate demand constraint (for example due to a Keynesian aggregate demand recession) raises the Lagrange multiplier μ_{t+1}^y, reducing desired k_{t+1}. Rewriting this equation in terms of the risk-free interest rate R_{t+1} as before, we also see that any rigidity in R_{t+1} induced by sticky prices will amplify the response of k_{t+1} to demand shocks. In contrast, in a flexible price environment, drops in R_{t+1} in a demand driven recession would encourage investment.

Consider a front loaded positive total factor productivity (TFP) shock. Using a log utility function, the consumption Euler equation

$$\frac{c_{t+1}}{c_t} = \frac{\beta_{t+1}}{\beta_t} efp_{h,t+1} R_{t+1}$$

can be rewritten as

$$\begin{aligned} \frac{y_{t+1}}{y_t}\frac{cy_{t+1}}{cy_t} &= \frac{\beta_{t+1}}{\beta_t} efp_{h,t+1} R_{t+1} \text{ where} \\ cy_t &= c_t/y_t. \end{aligned}$$

Combining this with the Euler equation for the flexible price economy, and ignoring for now the feedback from output and price rigidity to β_t or $efp_{h,t}$, we get

$$\left(\frac{y_{t+1}}{y_t}\frac{cy_{t+1}}{cy_t}\right) / \left(\frac{y_{t+1}^n}{y_t^n}\frac{cy_{t+1}^n}{cy_t^n}\right) = \frac{R_{t+1}}{R_{t+1}^n},$$

where x_t^n is the flexible price value of x_t (often called the "natural" x, for example the natural interest rate R_{t+1}^n). With flexible capital investment we know R_{t+1}^n will rise above the steady state value for a positive TFP shock. Since the output gap is stationary, $\left|\frac{y_{t+j}}{y_{t+j}^n}\right|$ will diminish over time. Suppose as a drastic approximation that the output gap vanishes after 1 period. Then the above equation reduces to

$$\frac{y_t^n}{y_t}\frac{cy_t^n}{cy_t} = \frac{R_{t+1}}{R_{t+1}^n}.$$

Under what conditions can we still say that the output gap

$$\ln y_t - \ln y_t^n$$

will be negative for positive TFP shocks? We need some combination of the interest rate R_{t+1} rising more than the flexible price interest rate R_{t+1}^n or a bigger increase in the investment to output ratio of the flexible price economy relative to the sticky price economy. In other words, investment must be more sensitive to Total Factor Productivity shocks under flexible prices. The analysis for the Total Factor Productivity shock can also be applied to demand shocks. For a given $\frac{R_{t+1}}{R_{t+1}^n}$ a positive demand shock in our simplified setup will cause a higher output gap if $\frac{cy_t^n}{cy_t}$ rises, that is investment must increase more relative to output in response to the demand shock under sticky prices. If $\frac{R_{t+1}}{R_{t+1}^n}$ declines (a typical situation), the condition for a higher output gap is relaxed to requiring that any fall in $\frac{cy_t^n}{cy_t}$ is small enough.

If we take for example an intertemporal consumption shock that raises current consumption, in many flexible price economies this shock will actually lead to a fall in investment due to a negative income effect on labour supply and a fixed labour demand curve in the short run (see the discussion of the response to shocks of the flexible price economy with flexible capital). Sticky prices weaken these effects so that the change in the investment to output ratio is higher for this shock relative to the flexible prices economy (see Smets and Wouters (2002) [167] for an example). Note that in the presence of feedback effects between credit constraints and the level of economic activity, $efp_{h,t+1}\frac{\beta_{t+1}}{\beta_t}$ will no longer be the same for the flexible price and the sticky price economies. This term is usually more sensitive to demand shocks in the presence of price rigidities. In this case, it acts as an amplifier of any output gap dynamics that would occur in its absence. More generally if the output gap (approximately) vanishes after T periods,

$$\frac{y_t^n}{y_t}\frac{cy_t^n}{cy_t} = \frac{R_{t,t+T}}{R_{t,t+T}^n}$$

where $R_{t,t+T} = \Pi_{j=0}^{T-1} R_{t+j,t+j+1}$ is the product of gross interest rates between periods t and $t+T$. The analysis above still applies with the new longer term interest rate ratio $\left(\frac{R_{t,t+T}}{R_{t,t+T}^n}\right)^{1/T}$ replacing $\frac{R_{t+1}}{R_{t+1}^n}$.

To gain more intuition, we take a 2 period economy with a representative household and a representative firm where prices are flexible in the 2nd period, and in the first period prices are fixed at the 2nd period price level. Investment only occurs in the 1st period. We can think of the 1st period as the short run and the 2nd period as the longer run or the medium run.This model cannot capture the effect of future aggregate demand shifts with sticky prices on current investment. It can nevertheless provide a good approximation, if due to investment adjustment costs, most of investment affects production only by the time prices have had the time to adjust to aggregate conditions. We take a perfect foresight approximation and ignore Jensen's inequality.

- Suppose the central bank targets a constant nominal interest rate $i_{t+1} = i$, so it does not respond at all to demand shocks. As a result, the real gross interest rate R_2 is constant. Consider a shock to the discount factor between the 1st and 2nd periods β_2. An increase in β_2 does not affect c_2 for a given k_2. Since $\frac{\Delta R_2}{\Delta \beta_2} = 0$, the Euler equation implies a decline in c_1 if $\frac{\Delta k_2}{\Delta \beta_2} \leq 0$ (equivalently if $\frac{\Delta c_2}{\Delta \beta_2} \leq 0$).

- Solving out the labour market for equilibrium hours and combining the bond and capital first order conditions to get a no arbitrage condition between capital and bonds, we find that k_2 is a function of 2nd period productivity z_2, k_1 (which is exogenous) and R_2. But since neither productivity z_2 nor R_2 are affected by the discount factor shock, $\frac{\Delta k_2}{\Delta \beta_2} = 0$. As a result period 1 output y_1 declines.

- In contrast, with flexible 1st period prices, households' higher desire to save will reduce consumption and increase investment for realistic levels of intertemporal elasticity of substitution below 1, with real interest rates declining to increase demand for capital.

Now consider a **temporary positive productivity shock** that lasts only for the 1st period. In the flexible price economy this raises output in the 1st period. Households also try to smooth consumption by increasing investment (or more realistically with firms holding the capital stock, firms increase investment in the interest of their shareholders). As a result 2nd period output also increases. This is the standard RBC analysis of a temporary increase in productivity.

- Now imagine prices are sticky the 1st period. Similar arguments as for the discount factor shock lead to the conclusion, that with a fixed real interest rate y_2 and k_2 do not change in response to a higher 1st period productivity. By the Euler equation for bonds, 1st period consumption c_1 also does not change. Since k_2 does not change, neither does investment, so $\frac{\Delta y_1}{\Delta z_1} = 0$. As a result, 1st period employment goes down. These results are similar to the earlier ones obtained with a fixed capital stock.

Marginal efficiency of investment shocks also have different effects when nominal prices are sticky. To continue with our 2 period model where 1st period prices are fixed, suppose that there is a negative marginal efficiency of investment τ_{k1} shock. With a fixed real interest rate, this shock reduces investment and therefore next period's capital stock. As a result, output and consumption go down in the next period when prices are flexible.

- By the Euler equation, consumption in the 1st period also declines. So similarly to a a negative news shock effect, a negative marginal efficiency of investment shock generates a simultaneous decline in both investment and consumption. In contrast, with flexible prices, this shock frequently generates negative consumption and investment comovement (see Justiniano et al (2011) [4] for a more detailed analysis of this point in a quantitative New Keynesian model).

$$\begin{aligned}\Delta\tau_{k,t+j} &< 0,\ \Delta R_{t+j} = 0 \Rightarrow \Delta k_{t+j+1} > 0, \Delta c_{t+j} \geq 0,\\ &\Rightarrow \Delta y_{t+j} > 0.\end{aligned}$$

The 2 period model does not allow us to study the effect on investment of future aggregate demand changes due to sticky prices. To analyse this requires at least a 3 period model (the short run, the medium run and the long run), in which prices are fixed for the 1st 2 periods.

- Suppose there is a **negative aggregate demand shock** (either $\tau_{k,t+j}$ or β_{t+j} or a combination of both) during the 1st 2 periods. In the next to last period $T-1$, output goes down by the same arguments as for the 2 period model. Let households own the capital stock and rent it to firms, and suppose price rigidities are at the level of retailers that buy goods wholesale and sell it at a markup. We also simplify by setting the depreciation rate $\delta = 1$.

Assuming the usual Cobb-Douglas production function, profit maximisation by the wholesale producer implies that

$$\alpha\frac{p_{y,2}y_2}{k_2} = r_2,$$

where $p_{y,t}$ is the price at which produces sell their output to retailers. As before, price rigidity implies that r_2 is fixed. By the same arguments as for the 2 period model y_2 goes down, and therefore $p_{y,2}$ declines as well. This implies a lower k_2 and therefore a fall in period 1 investment. [50] By the Euler equation the decline in c_2 and the higher β_2 reduces c_1. With investment and consumption both declining, y_1 goes down as well. If the model has $T > 3$ periods, a similar argument holds when prices are fixed for $T-1$ periods, using backward induction to establish $\Delta y_{T-j-1} < 0$,$\Delta c_{T-j-1} < 0$ when $\Delta y_{T-j} < 0, \Delta c_{T-j} < 0$.

$$\begin{aligned} \Delta\beta_{t+1+j} &> 0, \Delta efp_{h,t+1+j} > 0,\ \Delta R_{t+1+j} = 0 \\ &\Rightarrow \Delta y_{t+j} < 0, \Delta c_{t+j} < 0 \Rightarrow \Delta n_{t+j} > 0,\ \Delta I_{t+j} < 0. \end{aligned}$$

The analysis above assumed fixed prices and a constant nominal interest monetary policy for all periods before the last one, in order to highlight the distortions caused by price stickiness. To the degree that monetary policy responds by targeting the output gap or the natural interest rate (either directly or indirectly through inflation targets), the magnitude of the output gap will decline over time in response to stationary shocks, and in the medium-run outcomes in the economy should be similar to those of the flexible price economy.

14.3.3 Sticky prices and wedges

The sticky price mechanism can be mapped into a combination of countercyclical markups and external finance premia for demand shocks. The sticky price economy behaves as if firms prefer to charge higher markups and reduce production in the face of a negative demand shock. To achieve this lower production, they pay lower real wages which in equilibrium reduce labour supply. Consider the simple fixed capital model with the production function

$$y = zf(n),\ f_n > 0, f_{nn} < 0.$$

In a flexible price economy without labour demand wedges, the firm's hiring decision follows

$$zf_n = w$$

A negative aggregate demand shock reduces labour demand and wages. Starting from the flexible price output, this implies

$$\frac{zf_n}{w} > 1, \Longleftrightarrow zf_n = w(1 + \tau_{nd})$$

[50] Alternatively, if the producers themselves have market power and are subject to nominal price rigidities, cost minimisation implie the following optimal capital to labour ratio:

$$\frac{k_t}{n_t} = \frac{\alpha}{\gamma}\frac{w_t}{r_t}.$$

n_2 will go down only if w_2 declines (labour demand must decline). But then the cost minimising capital to labour ratio implies that k_2 declines as well. If $\Delta n_2 \geq 0$ (because the negative income effect dominates labour supply), then since $\Delta y_2 < 0$ and z_2 is the same, k_2 again declines. Either way, $\Delta k_2 < 0$ so 1st period investment goes down.

for some $\tau_{nd} > 0$. More generally, the aggregate demand shock looks like a higher labour demand tax.

Now suppose the firm also uses intermediate inputs in its production process. Making the same assumptions on the marginal productivity of intermediate inputs as before, and using similar reasoning as for labour demand we see that the negative aggregate demand shock increases the wedge on intermediate inputs $1 + \tau_{x,t}$. From the perspective of a value added function for output we know that the increase in $\tau_{x,t}$ will be like a reduction in TFP. Similarly, if the firm can vary the utilisation rate of capital and labour, or if there are search frictions in product markets, lower aggregate demand would show up as a decline in TFP. For demand shocks, this implies that in a flexible price model omitted price rigidity frictions can show up in the efficiency wedge, or in a mixture of production input wedges. See Goodfriend (2004) [98] and Sustek (2011)[172] for more on this point.

Increases in market power can also be mapped into an increase in production input wedges. Therefore, the negative effect of sticky prices on output in a recession is mathematically equivalent to an increase in market power. In both cases firms set prices that are too high, leading to output which is too low. In reverse, combinations of production input wedges can be mapped into a sticky price model subject to aggregate demand shocks.

At the ZLB, we may need to use the external finance premium/wedge in order to replicate the rise in the actual real interest rate during a recession. In this case, we interpret R_{t+1} in the BCA model as the flexible price/natural real interest rate. The actual real interest rate is $R_{t+1}efp_{h,t+1}$. The ZLB is reflected in a higher $efp_{h,t+1}$ during a recession.

How do sticky price based wedges compare to flexible price theories of aggregate demand or confidence shocks that also map into labour demand and TFP wedges, such as the search frictions model of Rios Rull et al (2012) [17] or the imperfect information model of Angeletos and Lao (2012) [11]? The sticky price perspective is that a recession happens or is amplified because firms systematically underadjust their prices to economic conditions so prices are set too high. The flexible price imperfect information model perspective is that a recession happens or is amplified because firms become pessimistic and think demand for their products is lower even if they adopt the best possible pricing strategy given their information set. The matching frictions model argues similarly that a recession happens or is amplified because it is indeed harder to sell products due to lower search effort by buyers even if firms adopt the best possible pricing strategy.

The equivalence results between the sticky price mechanism and wedges, in combination with questions about the actual allocational importance of sticky prices highlight the difficulty of identifying flexible price versus Keynesian sticky price economies using standard data sets on business cycles, at least if we only rely on aggregate quantities. The fundamental problem is that if we allow sufficient (and arguably quite plausible) flexibility in the shock processes, then real business cycle and sticky price models differ mainly in terms of their conditional forecasts for aggregate quantities given a specific shock. But the data themselves can only be directly mapped into an unconditional forecast and unconditional moments such as standard deviations or correlations. Any attempt to decompose the data into more fundamental shocks must be based on a model imposing some structure, which is itself subject to debate.

Data on inflation and nominal interest rates may provide better identification. But even with nominal variables, there are several reasons to suspect identification is complicated: flexible price models can also generate Phillips curves using similar central bank interest rate rules as in sticky price models (see our discussion of Phillips curves in flexible price economies), the Phillips curve relation is often hard to find in more recent data (Uhlig (2011) [175]), and interest rate rules are themselves subject to severe identification problems (Cochrane (2011)) [58].

It would seem that certain episodes can provide identification conditional on relatively weak auxiliary assumptions. For example the combination of high ex-post real interest rates and a severe decline in investment in the US during the early years of the Great Depression are hard to reconcile with flexible price and wage models if the high real interest rates were anticipated. Alternatively, the Great Depression in the US is compatible with a real business cycle model if agents' inflation expectations were significantly higher than the actual deflation over 1929-1933. In this case, ex ante real interest rates were actually low despite the high ex-post real interest rates. This hypothesis is hard to falsify given the lack of data on inflation expectations during that period and the rare nature of this episode. It is compatible with evidence that inflation was almost IID for most of the pre World War 2 period in the US, so agents might not have expected the deflation of 1929-1933 to persist for several years.

As for the countercyclicality of markups, the empirical evidence is inconclusive. A simple measure in many models is the inverse of the share of wages in output (the labour share). The labour share is usually countercyclical, suggesting procyclical markups. Adjustments for other factors of production and overtime labour costs sometimes imply countercyclical and sometimes procyclical markups (Ramey and Nekarda (2013)[154]). Bils et al (2012)[31] use the fact that durables goods are more procyclical than nondurables to test the cyclical behaviour of markups. A key prediction of countercyclical markup models is that as long as cost curves are upward sloping industries with more procyclical output will have more countercyclical markups. After adjusting for differences in TFP and the capital share, they find significantly more countercyclical markups for durable goods industries. At the same time, some of their evidence is inconsistent with sticky price theories of countercyclical markups. For example, they do not find more countercyclical markups for luxury products, even though output of those industries is highly procyclical. Also, industries with more frequent price adjustments do not have more procyclical markups, suggesting other theories of countercyclical markups besides nominall price rigidity are also important.

A significant complication in measuring markup cyclicality is the frequency of long term contracts. Long term employment or busines-to-business supplies contracts imply smoother and less procyclical average production costs than the marginal cost that's relevant for assessing markups. This biases existing studies that do not take this into account towards finding procyclical markups (though this bias may be less applicable to Bils et al (2012) [31], to the degree that it applies equally across durable and non durable goods industries).

14.3.4 Nominal wage rigidity

In modern New Keynesian DSGE models, the combination of both price and wage nominal rigidity plays an important role. This section explores the effect of nominal wage rigidity for aggregate demand shocks. Imagine a 2 period closed economy with CRRA utility in consumption. There is a continuum of perfectly competitive identical firms that use the production function

$$\begin{aligned} y &= zf(n), \\ f_n &> 0, f_{nn} \leq 0. \end{aligned}$$

The firm's optimisation yields the familiar first order condition

$$zf_n = w = \frac{W^n}{P},$$

where W^n is the nominal wage and P is the nominal price level and w is the real wage. The representative household has a log/log utility function in consumption and leisure. Credit markets open only in the 1st period. The 2nd period is a static flexible price and wage economy in which labour supply is determined by the standard neoclassical labour market first order condition

$$u_{l_2} = u_{c,2}w_2.$$

In the 1st period there is a credit market providing the risk free return $R_2 = \frac{1+i_2}{P_2/P1}$. Households' optimal choices satisfy the usual Euler equation

$$\frac{c_2}{c_1} = \beta_2 R_2.$$

The 2nd period P_2 price level is indeterminate. Suppose it is fixed. i_2 is the nominal interest rate controlled by the central bank. To highlight the inefficiency caused by wage rigidity, we assume that the central bank does not adjust i_2. The first period wage is subject to downwards nominal wage rigidity, that is

$$\Delta W_1^n \geq 0$$

With downward nominal wage rigidity, households' labour supply can no longer satisfy the standard first order condition as in the 2nd period. In this case, households agree to supply whatever amount of labour is demanded by firms at the given real wage w_1. β_2 is subject to shocks, with $\Delta\beta_2 > 0$ reducing desired 1st period consumption for a given interest rate and 2nd period consumption. Recall that the discount factor shock has no effect on output in a flexible price and wage economy.

We examine the effect of nominal wage rigidity on the response of the economy to an increase in the discount factor β_2. Start with the case of a linear production function where $f_{nn} = 0$. In this case $\Delta W_1^n \geq 0$ implies that $\Delta P_1 \geq 0$. As a result, the change in inflation $\Delta(\frac{P_2}{P_1} - 1) \leq 0$, increasing the real interest rate $R_2 - 1$. Since the shock to β_2 has no effect on $c_2 = y_2$, then using the Euler equation a higher β_2 combined with higher real interest rates reduces 1st period consumption and output. Firms are willing to produce less since they make zero profits regardless of the level of production (more generally with monopolistic competition they target a constant markup). The constant marginal product case highlights that one of the key effects of nominal wage rigidity is to increase nominal price rigidity in product markets. We will return to this point latter when we discuss elements that amplify the effects of price adjustment costs.

We can extend this argument to the more realistic case with diminishing marginal productivity of labour. Suppose the higher discount factor increases 1st period output. By $f_{nn} < 0$ and the firm's labour demand condition, w_1 must decline: real wages must fall to make hiring more workers attractive to firms. With downward nominal wage rigidity this requires an increase in the 1st period price level P_1. But this reduces inflation and raises the real interest rate $R_2 - 1$. From the Euler equation and $\Delta\beta_2 \geq 0$ this implies a decline in c_1 and 1st period output. But this contradicts the initial hypothesis that 1st period output increases. Therefore the discount factor shock must reduce 1st period output.

A scenario which is consistent is that in response to the higher discount factor, 1st period nominal prices fall. This increases real wages and depresses hiring and production by firms. Meanwhile looking at the Euler equation, while the decline in P_1 attenuates the effect of the rise in β_1 through a lower real interest rate, this is not enough to prevent a decline in desired 1st period consumption expenditure and output.

Finally, as in Chari et al (2007) [50], we can map nominal wage rigidity into a labour wedge in a business cycle accounting flexible prices and wages model. Suppose a negative aggregate demand shock (e.g a higher β_2) reduces 1st period output and consumption. For a standard utility function of the representative household , the reduction in work hours reduces the marginal utility of leisure, while the decline in consumption increases the marginal utility of consumption. Meanwhile the real wage is constant or higher. This implies

$$\frac{u_{l_1}}{u_{c_1}} < w_1,$$

or equivalently

$$\frac{u_{l_1}}{u_{c_1}} = w_1(1 - \tau_{ns,1})$$

for some $0 < \tau_{ns,1} < 1$. So nominal wage rigidity behaves like a labour supply tax, which is part of the labour wedge.

14.4 Inflation shocks and trade-offs between stabilising output and inflation

In the previous sections we assumed that the output gap targeted by monetary policy is relative to the natural level of output y_t^n, defined as output with flexible prices and wages. The implicit hypothesis was that the only distortion in the economy that monetary policy could fix was nominal price rigidity. This is a common assumption in baseline New Keynesian analysis (e.g Gali (2007) [91], ch. 3). But our analysis of RBC models showed how this natural output level was itself distorted by various "real", time-varying, frictions. In these circumstances, targeting the natural output level is suboptimal for monetary policy.

The central bank may prefer to stabilise the output gap relative to a level of output that is Pareto optimal, y_t^o, so the interest rate rule is for example

$$\begin{aligned} i_{t+1} - i_{ss} &= a_i(\ln y_t - \ln y_t^o) + b_i(\pi_t - \pi_{ss}), \\ \text{where } a_i, b_i &> 0. \end{aligned}$$

In this case, it faces a trade-off between stabilising the output gap and inflation.

To see this, go back to a simple linear New-Keynesian Phillips curve,

$$\begin{aligned} \pi_t - \pi &= aE_t(\pi_{t+1} - \pi) + b(\ln y_t - \ln y_t^n), \\ 0 &< a \leq 1, b > 0. \end{aligned}$$

Here stabilising the output gap $y_t - y_t^n$ is equivalent to stabilising inflation around π. A pure inflation target with literally no concern about deviations in $y_t^o - y_t^n$ and perfect implementation would eliminate the output gap relative to y_t^n. But now assume the central bank is interested in stabilising a more efficient level of flexible price and wage output y_t^o. In many textbooks such as Gali

(2007, ch. 5) [91] y_t^o is Pareto optimal, but more generally y_t^o simply has fewer distortions than y_t^n. We can rewrite the Phillips curve as

$$\begin{aligned}\pi_t - \pi &= aE_t(\pi_{t+1} - \pi) + b(\ln y_t - \ln y_t^o) + b(\ln y_t^o - \ln y_t^n) \\ &= aE_t(\pi_{t+1} - \pi) + b(\ln y_t - \ln y_t^o) + \varepsilon_t^\pi.\end{aligned}$$

If monetary policy aims to completely stabilise the output gap

$$\ln y_t - \ln y_t^o,$$

then $y_t^o \neq y_t^n$ implies that inflation cannot be completely stabilised. There is a trade-off between eliminating the output gap relative to y_t^o and eliminating inflation fluctuations. In general, y_t^o is less volatile than y_t^n. For a given $\ln y_t - \ln y_t^o$, inflation will rise when y_t^n falls by more than y_t^o in response to a shock and vice versa in a boom. From a modeling perspective, any misspecification of the inflation relevant output gap $\ln y_t - \ln y_t^n$ will generate a residual

$$\varepsilon_t^\pi = f(\ln y_t^o - \ln y_t^n),$$

where y_t^o is the misspecified measure of flexible price and wage output. Models often use a measure y_t^o that is too smooth, leading to residuals in the inflation equation that move countercyclically. These residuals, often called inflation shocks or markup shocks, play a large role in explaining inflation volatility in New Keynesian models (e.g Smets and Wouters (2007) [168]).

14.5 Zero lower bound dynamics: is higher price flexibility stabilising for aggregate demand shocks:?

A common intuition in many older Keynesian models (of the kind that are still at the core of most undergraduate macroeconomics courses and private sector macro analysis) is that higher price flexibility reduces the impact of demand shocks. From a modern macroeconomics perspective, this is still true if the central bank interest rate rule reacts aggressively enough to some combination of inflation, the output gap or the natural interest rate (some combination of sufficiently high positive coefficients on these variables). This is the usual assumption of Keynesian models about central bank policy in ordinary times. In this case higher price flexibility implies that a negative aggregate demand shock leads to a bigger decline in real interest rates. Lower real interest rates boost aggregate demand, countering the effect of the negative shock. This conclusion may be reversed when nominal interest rates hit the zero lower bound (ZLB).

A sufficiently negative aggregate demand shock can push the central bank to reduce nominal interest rates to zero(e.g an increase in consumer uncertainty or in the external finance premium). Now imagine that having already hit the ZLB, the economy is hit by another negative demand shock (e.g a further increase in uncertainty, further rise in external finance premia, or government spending cuts). Relative to the fixed prices benchmark, the negative demand shock can have a bigger negative effect on output in the economy with more flexible (but still partially rigid) prices. The shock generates persistently lower output gaps. Through the Phillips curve, these lower expected output gaps generate more expected deflation with more flexible prices. The expected deflation raises real interest rates, which depress aggregate demand further. The bigger reductions in aggregate demand lead to more deflation and a bigger drop in the output gap. Under some conditions, this feedback process is stable with higher price flexibility leading to a bigger drop in output in response to negative aggregate demand shocks.

At the same time, if prices are very close to the flexible price extreme, we can have a large drop in current inflation in response to lower aggregate demand, followed by higher expected inflation and lower real interest rates in the next periods (this is intuitively what happens in the flexible price economy). In this case higher price flexibility can reduce the impact of the aggregate demand shock

The analysis is complicated because in modern macroeconomic models aggregate demand depends on the whole path of expected real interest rates r_{t+j}, and therefore the whole path of expected inflation rates at the ZLB (since $r_{t+j} = i_{t+j} - E_t\pi_{t+j}$). To clarify things, we can examine the two key equations of the basic New Keynesian model: the consumption Euler equation and the Phillips curve. We make the usual simplifying assumption of zero steady state inflation. In log deviations from the steady-state,

$$\begin{aligned} \hat{y}_t &= E_t\hat{y}_{t+1} - \frac{1}{\sigma}(i_{t+1} - E_t\pi_{t+1}) + v_t, \\ \pi_t &= aE_t\pi_{t+1} + b(\ln y_t - \ln y_t^n). \end{aligned}$$

v_t is an aggregate demand shock such a shift in the external finance premium or a change in government spending. The flexible output level y_t^n is assumed to be exogenous. Suppose the economy is at the ZLB for T periods ($i_{t+j} = 0$ for $t + j < T$), and monetary policy successfuly closes the output gap by period $\tilde{T}$. Private sector agents have perfect foresight about T and $\tilde{T}$. Iterating forward on these two equations, we get

$$\begin{aligned} \hat{y}_t &= \frac{1}{\sigma}\Sigma_{j=1}^{T} E_t\pi_{t+j} - \frac{1}{\sigma}\Sigma_{j=T+1}^{\tilde{T}}(i_{t+j} - E_t\pi_{t+j}) + \Sigma_{j=0}^{\tilde{T}} E_t v_{t+j}, \\ \pi_t &= b\Sigma_{j=0}^{\tilde{T}} a^j E_t(\ln y_{t+j} - \ln y_{t+j}^n). \end{aligned}$$

The first equation clearly shows that at the ZLB anything that increases inflation expectations raises output, by reducing real interest rates. Future commitments by the central bank to reduce interest rates once the economy exits the ZLB can also raise output today to the degree that they are not fully offset by lower inflation expectations (lower inflation expectations can occur for example, if these commitments communicate to the private secto greater pessimism about future output gaps).

Inflation today is a weighted average of future expected output gaps. Rewriting this as

$$\pi_{t+j} = b\Sigma_{k=0}^{\tilde{T}} a^j E_t(\ln y_{t+j+k} - \ln y^n_{t+j+k}),$$

and substituting into the Euler equation, we have

$$\begin{aligned}\hat{y}_t &= \frac{b}{\sigma}\Sigma_{j=1}^{T}\Sigma_{k=0}^{\tilde{T}} a^j E_t(\ln y_{t+j+k} - \ln y^n_{t+j+k}) - \frac{1}{\sigma}\Sigma_{j=T+1}^{\tilde{T}} i_{t+j} + \frac{b}{\sigma}\Sigma_{j=T+1}^{\tilde{T}}\Sigma_{k=0}^{\tilde{T}} a^j E_t(\ln y_{t+j+k} - \ln y^n_{t+j+k}) \\ &\quad + \Sigma_{j=0}^{\tilde{T}} E_t v_{t+j}.\end{aligned}$$

The effect of increasing price flexibility (higher b) depends on two opposing forces. On one hand, any given decline in expected future output gaps $E_t(\ln y_{t+j+k} - \ln y^n_{t+j+k})$ is more deflationary with higher price flexibility. On the other hand, higher price flexilibility can reduce the magnitude of changes in expected future output gaps, which reduces expected future deflation. Using loglinear approximations, Eggertson et al (2012) [29] and Werning (2012) [181] find that under the assumption that inflation is at or below target when it exits the ZLB (so that $\tilde{T} = T$), and that the ZLB is hit due to a shock that reduces flexible price real interest rates (e.g an increase in external finance premia or in uncertainty), higher price flexibility worsens the decline in output at the zero lower bound. That is, the effect of higher price flexibility in terms of increasing the expected inflation impact of any change in the output gap dominates. Eggertson and Krugman (2012) [71] find that this effect is amplified when financing frictions are tighter, e.g in a financial crisis.These dynamics have a discontinuity in the behaviour of the model: higher price flexibility is destabilising for any intermediate level, even though with fully flexible prices demand shocks are completely stabilised.

Cochrane (2014) [60] questions the assumption that inflation and the output gap are expected to be at or below target upon exiting the ZLB episode. He finds that there are multiple equilibria in the basic New Keynesian model with sticky prices, which differ by the assumptions about inflation expectations towards the beginning and the end of the ZLB episode. Equilibria which allow for inflation and the output gap to be above target after the economy exits the ZLB converge continuously to the flexible price/frictionless limit ($\tilde{T} > T$ and $\ln y_{t+j+k} - \ln y^n_{t+j+k} > 0$ at $T+1$) , so that higher price flexibility is stabilising for aggregate demand shocks. The key to this result is that the magnitude of expected future output gap movements in response to demand shocks declines as price flexibility increases.

Cochrane also shows that those other "local to frictionless" equilibria are compatible with Taylor rules if we allow more flexibility in the behaviour of the monetary policy error $\varepsilon_{i,t}$ in

$$i_{t+1} - i_{ss} = a_i(\ln y_t - \ln y_t^n) + b_i(\pi_t - \pi_{ss}) + \varepsilon_{i,t}.$$

Different selections of the path of $\varepsilon_{i,t}$ implement different equilibria. These alternate equilibria may see empirically implausible for explaining the 2008 financial crisis, since we have not seen inflation above target in most advanced economies (the only notable exception is the UK). But, inflation may also be low in these local to frictionless equilibria if the reason we hit the ZLB is that the central bank overreacted and reduced interest rates by more than is necessary to eliminate a negative output gap (recall the analysis of flexible price Phillips curves). In the flexible price economy, if the central bank reduces interest rates this leads to lower inflation expectations since the real interest rate is not affected. By continuity, a similar result holds for local to frictionless sticky price equilibria if they're close enough the flexible price limit. Nevertheless, Cochrane's suggested interpretation of that output gaps were actually close to zero or even positive in 2008-2013 in advanced economies is highly controversial.

Mertens and Ravn (2014) [149] find that even if we assume inflation returns to target from below at the end of the ZLB episode (equivalent to assuming that $\varepsilon_{i,t} = 0$) higher price flexibility can be stabilising for aggregate demand shocks if prices are sufficiently flexible to begin with, or if the ZLB episode is expected to be long enough. The alternative ZLB equilibria with stabilising price flexibility in Mertens and Ravn's paper are due to confidence shocks, in which the economy moves to the neighbourhood of an alternative steady state with long-term deflation(or long-term inflation below target more generally). This is in contrast to the implicit assumption in the equations above that the economy stays close to the zero (or positive) long term inflation target steady state. In the confidence shock equilibria the ZLB episode occurs purely because agents expect deflation, an expectation that is fulfilled in a rational expectations equilibrium. These shocks have no effect on the flexible price (natural) real interest rates of the economy.

In contrast, the shocks driving the economy to the ZLB in Eggertson et al (2012) [29](e.g higher demand for precautionnary saving) also reduce the economy's flexible price real interest rates. In terms of their effects on output and inflation the confidence shock in Mertens and Ravn's equilibrium and the more fundamental shocks in Eggertson et al's equilibrium are similar. In both cases we get persistent below trend inflation and negative output gaps due to higher real interest rates. But in the Mertens and Ravn confidence shock equilibrium the effect of a negative aggregate demand shock conditional on being at the ZLB are much smaller. The negative aggregate demand shock has a smaller negative effect because it paradoxically leads to expectations of higher future output gaps and consequently lower expected deflation. This lowers real interest rates and increases private consumption.

The confidence shock ZLB equilibria in Mertens and Ravn's model are also local to frictionless: they converge continously to the flexible price equilibrium as we reduce the amount of price rigidity. While this property may seem intuitive, the result that negative aggregate demand shocks can actually increase output may seem strange. If we insist that the equilibrium that is closest to the actual economy is local to frictionless, and at the same time we insist that changes in aggregate demand move output in the same direction at the ZLB, then Mertens' and Ravn's results suggest that this combination is probably incompatible with RE. There must be some systematic misperceptions on the part of agents in the local to frictionless equilibrium that affects the flexible price equilibrium, e.g through sentiment shocks a la Angeletos and Lao (2012) [11].

Finally, Kiley (2014) [126] finds that higher price flexibility is stabilising at the ZLB if we assume sticky information instead of sticky prices. In the sticky information model costs of price adjustment are low, but firms adjust prices infrequently due to significant costs of updating their information about the state of the economy. Overall, the state of the art research has shown that the behaviour of the economy at the ZLB under Keynesians price or information stickiness is subject to multiple equilibria (or regimes) and significant model uncertainty. From positive perspective, this is a rich area for future research. On a more negative note, monetary or fiscal policy decisions at the ZLB are more complex than was initially thought.

14.6 Zero lower bound equilibria: a simple model

This section illustrates the differences between the destablising price flexibility ZLB equilibrium of Eggertson and Krugman (2012) [71],Eggertson et al (2012) [29], and the stabilising price flexibility ZLB equilibrium of Mertens and Ravn (2014) [149] using a simple 2 equations model. Imagine an economy described by the basic New Keynesian model without capital. As a short-hand, we shall call these the Eggertson-Krugman equilibrium and the Mertens-Ravn equilibrium, reflecting some of the key exponents of these models.

Due to either sunspot or more fundamental shocks, the economy has entered the ZLB regime. The economy stays at the ZLB in the next period with probability q. With a probability $1-q$, the economy exits the ZLB regime forever. Once outside the ZLB, the central bank sets the interest at its flexible price level, the output gap $\tilde{y}$ is zero and inflation is at its target, normalised to zero. We also allow for an exogeneous demand component x, such as government spending or investment (modelled here as exogenous and ignoring its implications in terms of changes to the capital stock). This setup is similar to that of Mertens and Ravn (2014) [149]

Our key simplifying assumption is that there are no transition dynamics: the economy is either in a ZLB steady state or in a zero inflation steady state. This approximates many of the results from more accurate numerical solutions, while allowing for a simple graphical analysis.

The economy in the ZLB regime is described by a Phillips curve for inflation π,

$$\pi_{zlb} = q\beta\pi_{zlb} + q\gamma\tilde{y}_{zlb}$$

and by a consumption Euler equation

$$\tilde{y}_{zlb} - x_{zlb} = q(\tilde{y}_{zlb} - x_{zlb}) + \frac{q}{\sigma}(\pi_{zlb} + r_{n,zlb})$$

, where r_n is the flexible price real interest rate. This model can be represented as two equations for inflation as a function of the output gap:

$$\begin{aligned} PC &: \pi_{zlb} = \frac{q\gamma}{1-q\beta}\tilde{y}_{zlb}, \\ AD &: \pi_{zlb} = -\frac{\sigma}{q}\left[(1-q)\,x_{zlb} + \frac{q}{\sigma}r_{n,zlb}\right] + \frac{\sigma(1-q)}{q}\tilde{y}_{zlb}. \end{aligned}$$

The solution of the model is at the intersection of these two curves. We can get different equilibria depending on the relative slopes of the PC and AD (aggregate demand) equations.

1) If

$$\frac{q\gamma}{1-q\beta} > \frac{\sigma(1-q)}{q}$$

(the PC curve is steeper than the AD curve) we get the Mertens-Ravn equilibrium. In this case, a shock that increases the flexible price interest rate $r_{n,zlb}$ (e.g lower uncertainty, a lower efp_h) or that increases x_{zlb} (e.g a temporary increase in government spending) leads to lower expected ZLB inflation π_{zlb} and a worsening output gap $\tilde{y}_{zlb}$.

2) If

$$\frac{q\gamma}{1-q\beta} < \frac{\sigma(1-q)}{q}.$$

we have the Eggertson-Krugman equilibrium. The same shocks lead to higher inflation and a higher output gap.

In the Mertens-Ravn equlibrium a higher slope of the PC equation (e.g higher γ) reduces the absolute value of changes in π_{zlb} and $\tilde{y}_{zlb}$ in response to AD shocks. In this sense, price flexibility is stabilising in this equilibrium. In contrast, in the Eggertson-Krugman equilibrium

the AD curve is steeper than the PC curve, and higher price flexibility γ amplifies the effects of shocks to AD in absolute value.

Note that the Mertens-Ravn equilibrium is more likely to occur the greater the level of price flexibility γ, the greater the expected length of the ZLB episode (higher q) and the higher the responsiveness of consumption to interest rates (lower σ). The Eggertson-Krugman equilibrium is more likely for higher levels of price-rigidity, short or moderate length ZLB episodes and when consumption is quite insensitive to interest rates.

14.7 Central bank interest rate control in New Keynesian models

One question left unanswered in most New Keynesian models is how exactly the central bank controls the short run nominal interest rate in the economy. There is a long tradition of more explicitly modeling central bank interest rate control by positing a demand for a monetary aggregate for which the central bank has a monopoly. In such a setup, central bank manipulations of the quantity of this "base" money will change interest rates to ensure money demand equals money supply. Associating the definition of money in such a model with something like M2 or even M1 leads to misleading conclusion about the conduct of monetary policy, and is inconsistent with the functioning of a modern payments system.

A more modern version of this theory is that financial institutions have a highly interest rate inelastic demand for central bank reserves, so that very small shifts in reserves supply can shift the short run nominal rate that is the opportunity cost of holding these reserves. The central bank targets the short run nominal rate by elastically supplying these reserves at the desired interest rate. The reserves market graph with the interest rate on the vertical axis has a flat supply curve and at the limit a vertical reserve demand function. This allows the central bank to determine a nominal short run interest rate, at least in theory (see Woodford (2000) [183] for discussion of monetary policy in a modern almost cashless economy. See Disyatat (2008) [67] for a detailed discussion of these points and the consequences of misconceptions about monetary policy operating procedures, such as the overemphasis on open market operations and the frequent assumption that the central bank uses control of a broader monetary aggregate to change interest rates). In combination with nominal price rigidity this allows the central bank to control short term real interest rates. In conjunction with the expectations hypothesis that long run interest rates are the sum of expected short run interest rates, this allows the central bank to have a strong influence on longer run real interest rates.

Note that that financial market imperfections are key here: the central bank must have a monopoly (more generally strong market power) in an asset (central bank reserves) whose role cannot be duplicated by other private sector assets like Arrow securities. Of course, this somewhat contradicts the spirit of many New Keynesian analysis that claim to assume fully complete markets. The role of central bank reserves is also never explicitly modeled in the vast majority of New Keynesian models. This makes them inappropriate for studying changes in the implementation process of monetary policy, and it leaves them vulnerable to the possibility that future financial developments will make reserves less important.

In practice, different short run safe interest rates are not perfect substitutes, due for example to liquidity considerations, so that the 3 month nominal T-bill rate does not mechanically follow the federal funds rate. Alternatively, if prices are relatively flexible the combination of real interest rates and inflation premia mostly determined by financial markets together with partially exogenous inflation expectations means that short run interest rates can be in large part determined by financial markets. There is some evidence for the US of persistent differences between market interest rates such as the 3 month T-bill rate and the central bank target interest rate (e.g Thornton (2010)[173], Fama (2013)[79]).

Fama (2013)[79] argues that the econometric evidence is inconclusive, remaining compatible both with an active central bank control of short run interest rates and a passive central bank whose target interest rates mostly respond to private financial market interest rate movements. In addition the evidence for the expectations hypothesis as a good approximation is mixed.[51] To the degree that short term rates are strongly related to long term rate movements, a lot of this may reflect the central bank reacting to long rates (see Cochrane and Piazzesi (2002) [61]).

14.8 Do sticky prices at the firm level imply aggregate price stickiness?

The critical part of the traditional Keynesian story is aggregate price rigidity. Micro level price rigidity is compatible with significant aggregate price flexibility as long as the price changes of of those firms readjusting their price are large enough and move in the right direction. In this case price adjustment by some firms can compensate for the rigidity of other firms' choices. This is similar to the way in which aggregate investment or labour supply can be quite flexible despite individual agent level fixed adjustment costs (see Khan and Thomas (2008) [124] for investment, and the basic indivisible labour RBC model (Ljunqvist and Sargent ch. 26 (2004) [140])).

[51] See Gurkaynak and Wright [104] for a survey of the recent evidence about the expectations hypothesis, mostly suggesting it does not hold. In contrast, De Graeve et al [100] find that within a medium scale NKDSGE model à la Smets and Wouters [168], the expectations hypothesis is a good approximation.

The effect is probably most familiar from discrete choice econometric models, where individuals can only choose a few discrete values $j = 1, ..J$, and as a result their choice may exhibit significant inertia. For example a household may not buy a car for the next 2 year, or it may only change its television set or fridge infrequently. In this sense, durable goods (or business capital goods) are sticky. Yet the probability of buying a durable good is a continuous function, and aggregate durable goods or capital demand functions are generally smooth and quite volatile due to continuous variation in the proportion of households purchasing new durables. In fact, durable goods and housing investment usually have much higher standard deviations than nondurable and services consumption. So aggregate investment demand is actually quite flexible.

The general idea is that while individual decisions may adjust only in discrete values or infrequently, because idiosyncratic shocks typically dominate aggregate shocks in importance, individual agents' decisions are usually weakly correlated in the cross-section. Therefore, despite the potentially significant fixed costs of adjusting, there are alwasy some agents adjusting and responding to aggregate shocks. In fact, while fixed costs of adjustment will prevent some agents from reacting to a shock in comparison to the costless adjustment case, they will also increase the proportion of adjusting agents whose idiosyncratic shock leads them to change their decisions in the same direction as in response to the aggregate shock alone. Agents whose idiosyncratic shock would lead them to change decisions in the opposite direction to that dictated by the aggregate shock end up with a smaller magnitude of desired frictionless adjustment. As a result they are less willing to pay the fixed adjustment cost, and they are less likely to adjust.

This selection effect increases the response of the economy to a shock, and partially compensates for the lower adjustment probabilities caused by the adjustment costs (see for example Golosov and Lucas (2007) [95] or Elsby and Michaels (2014) [74] for demonstrations of the quasi-irrelevance of non convex adjustment costs under certain conditions). Consequently, the aggregate index over those who adjust and those who don't is not necessarily sticky.

Golosov and Lucas (2007) [95] examine the effects of nominal shocks in a canonical model of firm pricing decisions, with idiosyncratic shocks and fixed price adjustment costs. They find small effects of monetary shocks when the model is calibrated to match the mean frequency of price adjustment in the US. Kehoe and Midrigan (2012) [123] find that if one calibrates the idiosyncratic shock process to match the frequency of small price changes in the data, there can be significant monetary non neutrality with fixed price adjustment costs. Eichenbaum et al (2013) [72] cast some doubt on this finding by suggesting most of the recorded small price changes are due to measurement error. Overall, this suggests that price rigidity need not imply large monetary non-neutrality in the short-run.

Another realistic nonconvexity is the discreteness of price decisions by firms. Discrete choice restrictions on prices are a special case of nonconvex adjustment costs in which adjustment costs are infinite over some range and then drop to 0 once a certain level of adjustment is reached (see for example the models with price plans of Eden and Jaremsky (2009 [69],2010 [70])). Similar to fixed costs, they induce a form of selection effect. Some firms do not adjust to an aggregate shocks when choices are discrete, because the next value on the discrete grid is too high relative to the optimal continuous adjustment. But some firms will adjust, and relative to the continuous choice case they will adjust by more, because they will be forced to choose a higher value on the discrete choice grid (relative to the optimal continuous choice). [52]

Finally, from the individual buyer's perspective price rigidity may not affect the quantity of a good they want to purchase if due to bounded rationality they do not distinguish between paying a higher price within some band (for example the car purchase decision may be insensitive to differences in price of 500$, and consumers may very well buy the same amount of chocolate bars whether the price is 2.50$ or 2.99$). In this case moderate amounts of price rigidity need not affect sales (see Knotek 2009 for a model of this effect for price points). Formally, this partial irrelevance of price rigidity is similar to our earlier model of wage rigidity with long term contracts. The message is similar: price rigidity does not increase the volatility of aggregate demand if it is accompanied by sufficient quantity rigidity at the individual level.

Much of the price adjustment occurs through sales. Kehoe and Midrigan (2012) [123] model sales as a transitory price change that incurs a smaller adjustment cost. They find that sales provide little aggregate price flexibility, with the rigidity of regular prices dominating the overall response to shocks of the aggregate price level. They suggest that in simpler price adjustment models we can ignore sales when calibrating the amount of price rigidity. In contrast Hernaiz (2010) [108] formulates a model where sales are used as a form of price discrimination. He finds small effects of nominal shocks in such a model, because sales are effective in increasing aggregate price flexibility. Eden and Jaremski (2009 [69],2010[70]) focus on the distribution of prices across a chain of stores, finding it to be quite flexible. They argue that despite the discreteness in price choices and rigidity in the store specific regular price, the flexibility of the price level across stores can generate significant aggregate price flexibility. They also suggest that the higher rigidity of regular prices may be specific to the relatively mild business cycles in the US during the great moderation. A larger shock such as the great recession of 2009 may have increased the frequency of regular price adjustment, though this has yet to be confirmed.

There are several arguably realistic models of aggregate flexibility of the nominal price level despite individual level rigidity. Here we focus on one

[52] Note the similarity to the flexibility of aggregate labour supply in models with indivisible household labour supply model (Kydland (1995) [62]).

of the most realistic and recent examples,The Burdett Judd model of price dispersion, described in Williamson and Wright (2010) [182]. In each market some consumers sample prices from several stores, while other consumers only examine a single price. A higher price discourages sales to consumers who have compared several prices, but it may raise revenue from those consumers who did not compare prices. As a result, firms can charge prices above marginal cost. In fact the profit maximising price is indeterminate over a certain range, which means the same homogeneous product can sell at different prices even without product differentiation. Firms can make the same profit through a high price and lower sales volume strategy or through a lower price and high sales volume strategy.

As long as firms charge a price within the profit maximising range they can keep it fixed. However, the profit maximising range of prices that defines the price distribution adjusts as a function of shocks so that for example monetary policy is neutral. Therefore, even though this model can match the frequency of micro level price adjustment and other facts about the price distribution quite well (and in fact better than most commonly used Keynesian sticky price models), its aggregate dynamics are like those of a flexible price level economy.[53] Burdett and Menzio (2014) [38] add fixed price adjustment costs to the Burdett and Judd price dispersion model and calibrate the extended model to US micro level data. They estimate that price adjustment costs account for at most 35% of the observed price rigidity. The remaining price rigidity is due to search frictions. These estimates are still quite preliminary, but they suggest that price rigidity may not be the main factor in explaining monetary non neutrality.

Regardless of one's opinion on the realism of the Burdett Judd sticky price model, the analysis highlights the critical role of complementarity between the pricing decisions of firms holding fixed prices and firms resetting their prices. Given the observed frequency of price adjustment in many economies, a significant quantitative role for Keynesian price rigidity requires price resetters to want to set a price close to that of other firms that are not adjusting their price. This explains the large amount of research effort dedicated by New Keynesians to finding such sources of pricing complementarity (also known in the literature as real rigidities). These range from firm specific production factors (Altig et al 2009), to kinked demand curves (Smets and Wouters (2007) [168]).

The effect of these real rigidities is highly dependant on the form of price stickiness. For example, firm specific production factors can increase aggregate price rigidity in the Calvo model with exogenous probability of price adjustment, but decrease it in state dependant pricing models (Dotsey and King (2005)). More elastic demand increases real rigidity in the Calvo model(Gali ch. 3 (2007) [91]) because the dominant effect is for higher elasticity to reduce the

[53] In fact, in the baseline calibration of the model, the average time between price changes is 11.6 months, which is similar to the typical duration of prices in the Eurozone. Yet monetary policy is neutral.

desired deviation of price adjusters from the price set by non adjusting firms. But in models where the probability is endogenous (state dependent pricing) or the price rigidity affects the size of the adjustment (the Rotemberg partial adjustment model) more elastic demand encourages more frequent and larger price adjustments in response to aggregate shocks (see for example the IMF's GEM (2008) model).

Other sources of pricing complementarity are sticky production input prices which reduce movements in marginal costs. Smaller movements in marginal costs in turn reduce desired price adjustments. The most important of these input price rigidities is wage rigidity, either due to contractual frictions or because of a highly elastic labour supply. Again, New Keynesians have done a lot of research on the level of wage rigidity, either due to rigid contract terms or elastic aggregate labour supply. Whether the observed amount of wage rigidity has a significant effect on costs in a way that affects pricing is controversial (Barro (1977) [20], Pissarides (2009) [157], Chetty et al (2011)[53]). The problem is that while wages of workers with tenure are rigid, wages of new hires, which is what matters most for pricing decisions (if pricing is close to optimal) are actually quite flexible. Therefore, observed wage rigidity inside a contract can be compatible with flexible marginal costs.

As for other explanations of elastic labour supply, Chetty et al (2011) [53]suggest that aggregate labour supply elasticity is much lower than the estimate used typical DSGE models: many macro models which match low elasticity on the intensive margin but assume a highly elastic extensive margin overpredict the response of hours of work to past historical events such as tax reforms. Rogerson and Keane (2011) argue that labour supply is quite elastic if we take into account effort and human capital formation. Erosa et al (2014)[75] find that a model combining several realistic features such as idiosyncratic income risk, matching frictions, partially indivisible labour supply and life cycle saving can match the response to historical tax changes while delivering an aggregate labour supply elasticity of around 1.3-1.75. This is still much lower than the aggregate labour supply elasticity of 2-4 often assumed in macroeconomic models. [54]

Finally, recent micro level evidence for the US (Bils et al (2012) [145]) find that reset price inflation (in the price index of those firms adjusting their price in a given period) is actually quite volatile and has low persistence. This suggests a low level of pricing complementarity between price resetters and fixed

[54]A low aggregate labour supply elasticity is a challenge for both Keynesian and RBC models, to the degree that one of the goals is to endogenise aggregate fluctuations as much as possible and to reduce dependence on large exogeneous shocks. Of course this is less of a problem to the degree that the large shocks (the wedges in our model) can actually be endogenised through some of the mechanisms we have discussed (e.g credit frictions, changes in the efficiency of the search and distribution processes in goods markets, shifts in the level of competition etc...), or to the degree that we can generate shifts in the labour supply curve (e.g from changes in demand for durables or the cost of financing nondurable consumption $\tau_{c,t}$).

price firms, posing a challenge to standard New Keynesian models relying on a high level of aggregate price stickiness and inflation persistence (e.g the Smets and Wouters model). It remains to be seen how these results apply in other important economies such as the Eurozone or emerging markets.

14.9 Do we need large price adjustment costs for aggregate price rigidity?

We can get significant aggregate nominal price rigidity in response to certain shocks despite low price adjustment costs if there are delays in absorbing new information about the economy (e.g Mankiw and Reis' (2010) [144] sticky information model) or if there are costs to simultaneously paying the same level of attention to aggregate and idiosyncratic shocks. In the last situation, a firm will typically prefer to focus on adjusting its price in response to more volatile and significant idiosyncratic shock and adjust more infrequently to aggregate shocks (see Mackowiak and Wiederholt' (2011) [141] DSGE model with rational inattention[55]). The ability to generate nominal effects even with low costs of price adjustment costs is important for Keynesian sticky price theory to the degree that recent evidence has found price adjustment in the US and many other countries is not so infrequent after al, e.g once every 4-7 months on average (Klenow and Malin, 2010)[131].

Limited attention models also predict that when unusually large aggregate shocks affect the economy (especially negative crises) firms will dedicate relatively more effort to understanding the impact of the aggregate shock. In theory, it is even possible that firms pay more attention in this case to the aggregate shock than to the idiosyncratic shocks. Price adjustment may be significantly more flexible for example in a financial crisis than in an ordinary recession. Firms will also pay more attention to more important aggregate shocks such as productivity or financial shocks than to monetary shocks. As a result, prices are likely to be much more flexible for these shocks.

Mackowiak and Wiederholt (2011) [141] find in their preferred calibration that the output gap is almost completely eliminated after 1 year in response to a technology shock. In contrast, monetary policy shocks affect output significantly for more than a year and a half. These mechanism are missing from the most commonly used time dependent price adjustment models, which assume the frequency of price adjustment is (approximately) independent of economic conditions. For example, a similar calibration with Calvo price adjustment costs implies that productivity shocks lead to output gaps that are significant for more than 3 years after the shock. To the degree that real shocks dominate business

[55] They use the Rational Inattention model in which agents optimise subject to noisy measurements of the variables affecting their decisions. Unlike in the standard state-space filtering framework, agents can reduce the measurement errors in the signals they get, subject to information processing constraints.

cycles, these results suggest a much larger role for the RBC dynamics towards which the sticky price economy converges. At the same time, sticky prices are still important in understanding and tracking business cycles if demand shocks are large and frequent, so that in any given quarter the output gap is significant. And rational inattention models can also rationalise strong output effects in response to disinflations, such as the one in the US in the early 1980's or in Canada in the early 1990's, because of imperfect credibility of the changes in the inflation target or because it takes time for people to realise there was a change in the monetary policy regime.

15 Government Spending Shocks in Real Business Cycle and New Keynesian Models

This section examines the impact of changes in government spending on the business cycle. In the baseline analysis, we follow most of the literature in assuming that government spending does not directly increase productivity in order to highlight the pure effect on output. Government spending simply represents a stream of expenditures that have to be financed somehow. This does not mean that it is a complete waste. Another interpretation is that it is essential spending that cannot be avoided, or it may accomplish some redistributional goals but with only minor direct effects on productivity.

The baseline assumption is that government goods are produced using the same production function and under the same profit maximisation considerations as in the private sector. Essentially this says that the equilibrium is the same if the government produces its own goods or subcontracts to the private sector We also assume perfect substitutability in labour supply to either the government or the private sectors. These are strong assumptions, but they allow to highlight the main effects of changes in government spending without worrying about the costs of intersectoral reallocation of production. We will examine the consequences of relaxing these assumptions later in this section.

The resource constraint of the economy with government spending is

$$c_t + I_t + G_t = y_t.$$

In the single representative household framework, Ricardian equivalence holds as long as the government uses only non-distortionnary taxation and we ignore the effect of the timing of taxes on the wedges. That is, for any fixed stream of government sending plans $\{G_{t+j}\}_{j=0,..\infty}$, we get the same path for other non fiscal variables regardless of the timing of taxes or debt changes. This occurs because the higher demand for debt by the government is met by an increase in the supply of savings demanded by the private sector. These hypotheses are a frequent starting point for discussion of these issues despite their lack of realism. We also start with lump sump taxes and Ricardian equivalence as a benchmark in order to emphasize some of the key issues in a simplified way. This allows us to impose a balanced budget every period without loss of generality,

$$T_t^g = G_t,$$

where T_t^g are government taxes [56]. Note that that our analysis of government spending shocks applies more generally to sectors of the economy for which output is taken as exogenous. For example, net exports can be modelled as exogenous in closed economy models (see. for example Smets and Wouters (2002), [167]).

15.1 A digression on Ricardian equivalence and the crowding out effects of government debt: how important is the timing of government taxes and deficits?

Ricardian equivalence has been particularly contentious in many circles (Elmendorf and Mankiw (1999) [73]and Romer (2001) [161], ch. 10)). To some analysts the idea that households increase saving in response to higher future tax liabilities is far fetched. For households that are not borrowing constrained Ricardian equivalence requires that on average households realise that significant increases in government deficits are likely to lead to future fiscal austerity measures if the government avoids default, or more plausibly if the probability of default is low. Alternatively, households would be concerned about the possibility of costly default. In either case, such concerns should lead to higher household saving when government deficits increase. This is a natural consequence of the RE assumption that on average households estimate their expected future income net of taxes sequence correctly.

From an RE perspective, important deviations from Ricardian equivalence have to be justified by appealing to more fundamental frictions such as credit constraints. In fact, Evans et al (2010) argue that without these frictions under some assumptions Ricardian equivalence may hold even under learning dynamics. [57] For a tax cut, Ricardian equivalence is a good approximation as long as households save most of the tax cut to smooth their consumption.

From a general equilibrium perspective, Ricardian equivalence forces us to take into account that a priori, whether the private sector funds new government

[56] Recall that T_t includes other lump sum taxes transfers to households to compensate for deadweight losses from the wedges.

[57] The key assumptions for Ricardian equivalence to hold under learning dynamics are that forecasting rules for interest rates are independent of fiscal variables (which is unrealistic if we allow for default risk, but Evans et a (2010) ignore default risk and anyways default risk would break Ricardian equivalence under RE) and government spending is exogenous (though again, Ricardian equivalence can fail with RE as well if changes in government deficits influence agents' beliefs on the path of government spending).

spending through higher taxes of through higher government bond purchases, in either case this constitutes a reduction of the private sector's disposable income available for consumption or investment. This is true even if Ricardian equivalence itself fails.

A priori, it is not clear why reducing disposable income through government bond purchases should lead to a smaller negative effect on consumption and investment than paying higher taxes. Both types of funding can reduce private sector spending. Ricardian equivalence determines the interest rate effects of changes in government debt. The closer we are to full equivalence, the less interest rates have to rise to induce the private sector to increase its bond holdings. Conventional old Keynesian analysis that assumes only taxes reduces the private sector's ability to spend simply ignores this point, precisely because it is inconsistent in its treatment of the economy's resource and budget constraints.

Ricardian equivalence requires lump sum taxation and no government default risk. Government debt levels matter even under RE and perfect private sector financial markets when they affect the timing of distortionnary taxes or when they can lead to costly sovereign default. Default becomes unavoidable once debt levels become so high that economically and politically feasible tax increases or spending cuts are insufficient to ensure repayment. For example, this is a major concern with the scale of unfunded pension and healthcare liabilities facing many developed economies over the next decades (see Leeper and Walker 2011, Trabandt and Uhlig (2012)).

Finally, Ricardian equivalence is a statement about the effects of the sequence of taxes and deficits hold the sequence of government expenditures constant. Forward looking households would rationally consume a large portion of a tax cut if it leads them to expect significant cuts in future government spending.[58] The point of Ricardian equivalence is that these effects are distinct and can have quite a different from the old Keynesian argument that government deficits crowd out private saving.

The complete markets with intertemporal optimisation consumption/saving model underlying the initial version of Ricardian equivalence is clearly false. But the recent evidence suggests moderate marginal propensities to consume out of tax rebates when taking an average over many households, with the best estimates somewhere around the range of 20%-25%. Kaplan and Violante (2014) [122] survey the evidence from reduced form regressions and develop an extension of the optimising PIH with precautionary savings and illiquid assets that matches the empirical evidence. One of their key findings is that the MPC out of tax rebates depends on the size of the rebate: as transfers become larger, it can go down by up to a factor of 3. This suggests caution in generalising from small tax rebates to larger ones. There is also evidence of significant insurance

[58] See Simon Wren Lewis' blog post for a discussion of this point (http://mainlymacro.blogspot.com/2013/02/ricardian-equivalence-and-political.html).

against individual income shocks (e.g Kaplan and Violante (2010) [121] and the references they cite). This suggests that the other extreme of assuming household saving does not respond at all to changes in the government deficit is probably an even worse approximation.

In general, Ricardian equivalence will fail in the presence of credit constraints if the government can borrow and lend to credit constrained agents more efficiently than the private sector (Romer (2001), ch. 11 [161]). For example the government may have easier access to credit if it can provide a stronger commitment to repay its debt than the private sector. In this case government borrowing affects aggregate macroeconomic quantities by increasing the effectivenes of financial intermediation. A special case of this result is the positive effect on aggregate consumption of tax cuts in models with rule of thumb households who simply consume their after tax income each period (a common assumption in many fiscal policy DSGE models, e.g Gali et al (2007) [92] and the IMF's GIMF model (2013) [7]). These households are unable to borrow even though they would like to. A reduction in taxes relaxes their borrowing constraint.

Alternatively, because transfers to credit constrained agents can affect consumption and interest rates with incomplete markets, Ricardian equivalence fails if the government uses the debt issue to implement a transfer from unconstrained agents (typically the government bondholders) to the credit constrained agents . The government does this by cutting current taxes for credit constrained agents, and paying the debt by raising future taxes for unconstrained agents. This is similar to a transfer from unconstrained to credit constrained agents. In this case a rise in government debt at moderate levels (so sovereign default risk concerns are minimal) behaves similarly to a lower discount factor $\tilde{\beta}_{t+1}$ shock. In a flexible price economy this leads to higher loan demand, pushes up consumption and interest rates while reducing investment. In a sticky price economy, interest rate effects are smaller, and instead output increases. Similarly, in an open economy, suppose the government has more credibility in repaying its debt to foreigners than the private sector in a financial crisis. In this case, a financial crisis like the one in the US in 2008 will be accompanied simultaneously by higher government borrowing from foreigners that is not fully offset by lower private borrowing.

Elmendorf and Mankiw (1999) [73] and Romer (2001) [161] provide excellent discussions of these issues and a review of more reduced form empirical evidence. Elmendorf and Mankiw (1999) conclude that the vast number of empirical studies trying to test Ricardian equivalence are bogged by deep endogeneity and measurement issues, making it difficult to draw any strong conclusions.

Perez (2010) develops a DSGE model where non Ricardian effects of government debt are significant due to heterogenous liquidity constraints on firms. He finds that if credit markets are tight enough, higher government debt raises interest rates and crowds in investment by increasing the returns on savings by firms. On the other hand for more realistic levels of credit constraints higher government debt raises interest rates and crowds out private investment by making private borrowing more expensive. His model makes several extreme assumptions such as no saving by workers. It remains to be seen how these results extend to more general setting and more serious estimation or calibration exercises.

15.2 The government spending multiplier

The key question in this section is how effective are increases in government spending in raising GDP? Define the government spending multiplier for horizon J as

$$\gamma_{g,y}^{J} = \frac{\Sigma_{j=0}^{J} \frac{\Delta y_{t+j}}{R_{t,t+j}^{g}}}{\Sigma_{j=0}^{J} \frac{\Delta G_{t+j}}{R_{t,t+j}^{g}}},$$

where ΔG_{t+j} is the change in government spending in period $t+j$, Δy_{t+j} is the change in GDP in period $t+j$ and $R_{t,t+j}^{g}$ is the gross discount rate between t and $t+j$. The main question for the evaluation of fiscal stimulus plans is whether or not $\gamma_{g,y}^{J}$ is bigger or less than 1. A multiplier below 1 implies that while GDP increases, private spending has actually gone down so that the increase in government spending can only be justified as an unavoidable necessity or as a potentially inefficient redistribution mechanism. A multiplier above 1 means that private spending has actually increased. While this does not imply higher private sector welfare, it makes it much more likely that welfare has increased. A low multiplier coupled with deficit financing also increases the likelihood that fiscal stimulus spending reduces the sustainability of government debt. A high multiplier increases the probability that a fiscal stimulus program may actually lower debt to GDP ratios and improve the government's fiscal position.

Keynesian economists often argue for large fiscal stimulus plans in a recession based on the belief in large fiscal multiplier above 1 (e.g the American fiscal stimulus plan of 2009 was premised on a short run multiplier of 1.5). Of course, if the multiplier is always above 1, then the government could always boost GDP and private spending by increasing the proportion of government spending in GDP. The fact that many analysts are wary of the conclusion that higher government spending as a proportion of output is always better, and the frequent arguments to the contrary that high government spending as a proportion of output can lead to lower output suggests that there are doubts about the linearity of the effects of government spending. Perhaps the same large multipliers that apply for small stimulus programs do not apply to larger ones. To analyse this issue requires a more formal, systematic investigation: precisely what the DSGE approach tries to accomplish.

15.3 Two cases when the government spending multiplier is (almost) irrelevant

Before proceeding we should note the implications of another extreme assumption on how government spending enters the model, that delivers stunningly different conclusions from the standard analysis in the next sections. Suppose that government spending and private consumption are perfect substitutes, so that the utility function is $u(c_t + G_t, 1 - n_t, D_{t+1})$. This may be a good approximation to government spending programs that duplicate goods and services that are usually produced by the private sector with an equal level of efficiency. Since c_t and G_t are perfect substitutes, the household's problem for a given sequence of prices only depends on $c_t + G_t$. The firms' problem for a given sequence of prices is also independent of the composition of $c_t + G_t$. Furthemore, the split of $c_t + G_t$ between its two components is indeterminate without specifiying the process for g_t. Therefore, we can solve the model for all variables and for $c_t + G_t$ independently of G_t as long as the government's choice of G_t does not imply $c_t < 0$ (this would almost never happen for plausibly low G_t).

In this case the equilibrium of the economy is independent of ΔG_t and the government spending multiplier is zero. Intuitively, this case is analogous to the government increasing its spending simply by taking over a private business. In this case, there is no reason for economic activity to change unless there is a change in the efficiency of running the business (output would increase if the government improves the business' efficiency and vice versa output would decline if the government is less effective at managing the business). While households have less income to spend on private consumption, there is no negative wealth effect because higher government consumption is a perfect substitute for the reduced private consumption. Furthermore, under this hypothesis, we can model the economy without explicitly discussing the government sector after redefining the relevant consumption index for the purpose of estimation and forecasting to be $c_t + G_t$.

In reality, activities that are usually performed by the public sector are quite different from those performed by the private sector. However, the analysis with perfect substitutability of government and private consumption does suggest that to the degree that large fiscal stimulus plans may increasingly spend on activities that are already performed well by the private sector, they will lose some of their effectiveness in raising output.

In the opposite direction, to the degree that government spending is well directed towards projects that complement private sector consumption and investment (e.g infrastructure spending or education and healthcare spending), this will increase the government spending multiplier (see for example Coenen et al's (2012) analysis using a fiscal policy version of the ECB's NAWM). In some cases with highly productive government investment even flexible price models can generate large government spending multipliers. At the same time, recognizing the importance of productive government spending suggests that we should be careful in linear extrapolation of government spending multipliers in certain episodes. In the most plausible case of diminishing returns to government investment, government spending multipliers observed in economies with a small government sector are likely to be larger than those in economies with an already large government sector with lower returns on additional spending.

There is another reason for why the effectiveness of government spending increases in stimulating economic activities may be severely overestimated. Consider the classic caricature Keynesian thought experiment in which higher government spending is simply used to pay salaries to the unemployed while engaging them in completely unproductive activities (e.g digging holes and filling them up again). Since the unemployed are not really producing anything, this is equivalent to a transfer from employed workers to the unemployed. But in a model with complete financial markets this transfer has no effect on economic quantities. The government spending multiplier is zero in this case. National income statistics frequently evaluate the output of the government sector at the cost of the labour and other production inputs it uses. In this case, they would actually report a government spending multiplier of 1, even though value added in the economy has not really change.

The case of completely unproductive government spending is extreme. On the other hand, the majority of increases in government expenditure in fiscal stimulus packages during the 2009 recession were in the form of transfers (Oh and Reis (2012) [155]). With more realistic incomplete financial markets, transfers to agents whose credit constraints are binding or more likely to bind (due to lower asset levels) can significantly raise output. But it remains the case that in current DSGE models with credit constraints, the multiplier effects of transfer shocks are significantly lower than those for government spending shocks (see GIMF (2013) [7] for results from a state of the art New Keynesian DSGE model, see Oh and Reis (2012)[155] for an analysis of the effect of transfers with price rigidities and uninsured household idiosyncratic risk).

The positive effects of transfers on consumption and possibly on output are likely to be more important in a financial crisis when many agents are credit constrained, if the transfers are well targetted to alleviate financing frictions. In this case, the government essentially acts as a replacement for the missing private sector financial intermediation during the crisis. Whether or not actual government transfer programs during the 2008-2010 financial crisis were well targetted in this sense is still an open question. On the other hand, a pure transfer financed by lump sum taxes is a priori neutral in its effect on private spending, while government spending crowds out private spending if the multiplier is less than 1. Transfers can have a negative output multiplier if they are financed by distortionnary taxes. [59] Note that regardless of the low multi-

[59] In an open economy, some of the crowding out can take the form of lower availability of domestic goods to foreigners - lower net exports.

plier effects on output, increases in transfers in a recession may increase welfare by providing valuable insurance to poorer households against job loss or more general income loss.

15.4 Government spending multipliers in models with fixed capital stock

We now return to the standard analysis, in which government spending uses productive resources and is not a perfect substitute for private spending. We start with a fixed capital stock model, where labour is the only input and

$$y_t = c_t + G_t.$$

As in earlier analysis, we start with the flexible price and wage economy. In the benchmark RBC model government spending does not affect labour demand. Therefore, changes in output in response to changes in government spending must come from shifts in labour supply.

Suppose first that there is no income effect on labour supply. In this case equilibrium in the labour market is unaffected by changes in government spending. Therefore, output is unaffected and the first period multiplier must be 0. The increase in government spending is completely offset with a fall in private spending.

Now return to the more common case in which leisure is a normal good (whose demand increases with income). An increase in government spending must lower private consumption, leading to a first period multiplier less than one.

To see this note that if consumption does not change, then output must increase. The output increase requires a rise in labour supply which by the income effect requires lower consumption. This contradicts the initial hypothesis that consumption does not change. If consumption rises, labour supply falls by the income effect. This implies a fall in output, contradicting the rise in output due to the increase in consumption and government spending. The only remaining possibility is that consumption falls, labour supply rises and output increases by less than the increase in government spending.

- The result reflects the common intuition that higher government spending reduces private consumption and increases labour supply in an RBC economy because of a reduction in household wealth. It explains the difficulty of such models in reproducing the positive comovement between output, government spending and consumption that appears in some of the reduced form empirical studies. Note that this result of a goverment spending multiplier below 1 with flexible prices is derived from purely static considerations. It would also apply if households in our model consumed all their income instead of trying to smooth their consumption across periods.

In order to understand the difference that sticky prices will make for the strength of the multiplier it helps to examine the difference between purely transitory and persistent government spending shocks. We can gain significant intuition about this distinction using a 2 period economy with labour as the only input. One interpretation of this model is that the 1st period is the short-term in which prices are sticky, while the 2nd period is the medium-long term when prices are flexible.

If the **government spending shock only occurs in the 1st period**, 2nd period output is unaffected. To see this, note that any change in second period consumption would imply that output must change in the same direction, from the resource constraint. At the same time, by the income effect on labour supply consumption and output must move in the opposite direction. This contradiction establishes the result.

We have already found that increased government spending in the 1st period must lower 1st period consumption. Since 2nd period output does not change, the consumption Euler equation implies a higher real interest rate as households try to smooth consumption against what is essentially a negative temporary income shock. In contrast, if productivity and the increase in government spending are the same across periods, for standard utility functions we get that

$$\Delta c_t = \Delta c_{t+1}.$$

Therefore, there is no effect on the real interest rate.More generally, a more persistent positive government spending shock in this economy will generate a smaller real interest rate increase.

Finally, consider a **change in government spending only in the second period.** Previous arguments have shown that 2nd period consumption declines, 2nd period output increases while 1st period variables remain unchanged. By the consumption Euler equation real interests decline, as households try to smooth consumption across periods by saving more in period 1.

The analysis also shows that the standard reasoning about higher government spending affecting labour supply through wealth effects is not completely accurate. The timing of government spending also matters. In general equilibrium the response of consumption to future or past changes in government spending can be zero (in our special case of fixed capital) or quite small more generally. In a flexible price model without capital and without labour demand curve and government spending interactions, consumption is in fact invariant to news shocks on government spending. Finally the response of interest rates to a government spending shock can have different signs depending on the path of the shock.

Nominal price rigidity has a major impact on the analysis of government spending shocks.. In this case, the government spending multiplier critically depends on the central bank's interest rate policy. We continue with the same 2 period model. Start with a special case in which the 1st period price level must equal the 2nd period price level, the central bank follows a fixed nominal interest rate target , and prices after the 1st period can adjust flexibly. This is close to the common undergraduate textbook analysis of government spending shocks with fixed prices in the short run and flexible prices in the long run. Our assumptions imply a fixed real interest rate in the 1st period, so absent any intertemporal demand shocks 1st period consumption must be equal to 2nd period consumption.

Suppose **government spending increases only in the 1st period**. 2nd period consumption and output are unaffected due to price flexibility. By the Euler equation with a fixed real interest rate, 1st period consumption is also unchanged. Therefore, output increases. This is compatible with the rigid price equilibrium because firms are willing to satisfy any demand from the private sector or government at the fixed price. Because consumption does not respond the 1st period government spending multiplier is now 1. In contrast, the flexible price multiplier was below 1.

Now consider a **permanent increase in government spending**, where 2nd period government spending increases by the same amount as that in the 1st period. Second period consumption falls. By the Euler equation with a fixed real interest rate, 1st period consumption will fall by the same amount, as households attempt to smooth consumption through higher saving at a fixed interest rate. Therefore, in this case the government spending multiplier is below 1. In fact it is the same as in the flexible price case, because in that case as well consumption in both periods falls by the same amount.

Finally, if **government spending increases only in period 2**, consumption in period 1 will also fall by the Euler equation. In contrast, in the flexible price economy first period consumption is unaffected by the future government spending shocks. The future higher government spending shock is similar to a negative news shock. This has a bigger effect in the sticky price economy (see the previous discussion of news shocks in the New Keynesian models for more on this point).

All of these conclusions are conditional on a passive central bank interest rate policy. In many setups it is in fact optimal for the central bank to target the same real interest rate as in flexible price models, that is the natural interest rate. If the central successfully tracks this natural interest rate, our conclusions about the importance of sticky prices for the government spending multiplier are completely reversed: now the sticky price and flexible price economies have the exact same government spending multipliers. In intermediate cases, such as an interest rate rule responding positively to inflation and an output gap measure with finite coefficients, the 1st period multiplier for transitory 1st period government spending shocks is below 1 but above the flexible price multiplier.

Much of the analysis from the 2 period model survives in a more realistic infinite horizon economy. Woodford (2010) [184] starts with a benchmark sticky price environment in which the central bank targets a fixed real interest rate and any government spending shocks are stationary so the economy eventually converges back to the steady state. His analysis of this case reproduces our conclusions in the 2 period model with a rigid 1st period price and fixed central bank nominal interest rate target. In particular, the fixed real interest rate pins down consumption as fixed in response to the increase in government spending. As a result the multiplier is equal to 1. [60]

With full output gap stabilisation, the sticky price multiplier is identical to that flexible price multiplier. In the intermediate case of partial response to inflation, the output gap or the natural rate of interest (e.g through a Taylor rule) the multiplier is below 1 but above the flexible price multiplier. The negative effect of higher government spending news shocks in the presence of price rigidity and the lower multiplier for more persistent government spending shocks also survive in infinite horizon economies. In fact Woodford finds that the sticky price multiplier can be significantly negative in the presence of permanent government spending shocks.

The infinite horizon environment allows for a more realistic analysis of the spending multiplier at the zero lower bound (ZLB) on the nominal interest rate. Start with the standard analysis that assumes inflation is below or above target when the economy exits the ZLB episode (e.g Werning (2012) [181]. Here the central bank is constrained in its ability to lower interest rates in order to track the natural interest rate. With a fixed price level, the multiplier would be 1. But with partial price flexibility, higher government spending over several periods during which the zero lower bound constraint is binding can lead to higher expected inflation (through higher output gaps) which lowers real interest rates. As a result, the government spending multiplier may exceed 1 significantly.

Under some parameter values Christiano et al (2010) [54] find a multiplier above 2 at the ZLB, using a medium scale New Keynesian DSGE model. In contrast, other studies (e.g Drautzenberg and Uhlig (2013) [68]) find lower multipliers below 1 for government spending shocks at the ZLB. The key difference in the results is that studies finding a low multiplier assume that the government spending increase lasts significantly longer than the duration of the zero lower bound problem. The expectation of future increases in government spending

[60] Woodford himself notes that it may seem strange that this result is independent of the level of price rigidity. He emphasizes that as prices become more flexible, the government spending shock generates bigger fluctuations in inflation for a fixed real interest rate. As a result the analysis is really only valid for high levels of price rigidity. Otherwise the central bank would almost certainly adjust interest rates to control inflation.

and taxes when the economy is no longer at the ZLB reduces private sector income expectations, lowering consumption and investment. [61] These results match our earlier analysis of how higher government spending news shocks depress output with sticky prices, lowering the government spending multiplier.

The overall picture emerging from the initial analysis using New Keynesian policy models is that a rise in government spending can be quite effective in reducing the effects of a deep recession when nominal interest rates are close to zero, as long the extra spending is mostly accomplished when the economy is still in the ZLB regime (see Coenen et al (2011) [76] for a survey of multipliers from several policy models). The positive effect on output of an early stimulus can be enhanced by promises of future reductions in government spending and taxes, that act like a positive news shock. However, given the multiplicity of equilibria and the importance of nonlinearity at the ZLB, these conclusions need to be interpreted with some caution, especially since they are based on (log)linear approximations of the main model equations and are not robust to different assumptions on monetary policy rules.

Recent research (e.g Mertens and Ravn (2014) [149], Cochrane (2014) [60], Kiley (2014) [126]) has questioned the effectiveness of government spending increases at the ZLB. It has shown that there are other ZLB equilibria in which the government spending multiplier is less than 1. In particular the government spending multiplier is likely to be less than 1 for relatively high price flexibility, high duration of the ZLB and if monetary policy allows inflation to be above target when exiting the ZLB . In Mertens and Ravn (2014) [149], these ZLB equilibria are generated by self fulfilling expectations of deflation and zero interest rates. In Cochrane (2014) [60], government spending multipliers are less than 1 if agents expect the central bank to allow inflation to be above target when exiting the ZLB regime.

These equilibria have the intuitive property that as prices become more flexible the dynamics of the economy converge to the flexible price limit. They have the counterintuitive property that higher government spending may reduce the private sector output gap. For a more detailed analysis, see the earlier section on the conditions for higher price to be stabilising for aggregate demand shocks (noting that an increase in goverment spending is a positive aggregate demand shock).

15.5 Differences in the production function of the private and government sectors, limited intersectoral mobility and the government spending multiplier

The standard DSGE models assume the same production function for the goverment and private sectors. Essentially, our analysis so far (and most standard

[61] Note that Christiano et al report static multipliers $\frac{dY_{t+j}}{dG_{t+j}}$.

analysis) assumes that it is irrelevant whether or not the government itself produces G_t or it subcontracts it out to the private sector. Here, we examine the consequences of the more realistic assumption of separate production functions for government goods and private goods. To keep things simple, we ignore income effects on labour supply, for example by using the utility function

$$u(c,l) = c + v(l),\ v_l > 0, v_{ll} \le 0.$$

Suppose that there is a government production function

$$G_t = z_{g,t} f(n_{g,t}),\ f_{n_g > 0},\ f_{n_g n_g} \le 0,$$

and a private production function

$$c_t = z_{p,t} f(n_{p,t}),\ f_{n_p > 0},\ f_{n_p n_p} \le 0.$$

The private sector decides how much to produce based on standard profit maximisation considerations. The government sector simply produces whatever is demanded by the government.

We start by illustrating a case where this production structure is exactly equivalent to the standard assumption of a common production function across sectors. Continue assuming that government and private sector labour supplies are perfect substitutes with $l = 1 - n_p - n_g$. The private sector labour market always satisfies

$$v_l = w = z_p f_{n_p}.$$

Suppose

$$\begin{aligned} f(n_j) &= n_j^{\alpha},\ j = g, p, \\ ,0 &< \alpha \le 1. \end{aligned}$$

Start with the constant returns to scale case of $\alpha = 1$, so that

$$v_l = w = z_p.$$

This labour market equilibrium condition implies the same amount of labour supply regardless of any change in government spending ΔG. As a result

$$\begin{aligned} \frac{\Delta n}{\Delta G} &= 0,\ \text{and} \\ \frac{\Delta y}{\Delta G} &\ge 0 \iff z_g \ge z_p. \end{aligned}$$

In this case of constant returns to scale and perfect substitutability between working in the private and government sectors, we get identical results as for the case with a single production function if $z_p = z_g$.

Now, we will consider a case where despite flexible prices the government spending multiplier $\frac{\Delta y}{\Delta G} = 1$ just like in the Keynesian fixed price model. Suppose the representative household has the utility function

$$u(c, n_g, n_p) = c - v^g(n_g) - v^p(n_p),$$

with both $v^p(.)$ and $v^g(.)$ increasing and convex functions.

The private labour market equilibrium condition

$$v^p_{n_p} = z_p f_{n_p}$$

establishes the same equilibrium private sector labour supply n_p regardless of g. As a result

$$\frac{\Delta c}{\Delta G} = 0, \text{and } \frac{\Delta y}{\Delta G} = 1.$$

Note that this equilibrium requires a different wage in the government and private sector. For example if $v^g(.) = v^p(.)$ higher government spending requires an increase in $\frac{w_g}{w_p}$ to attract more labour into the public sector.

How did we get this "Keynesian" result despite perfect price flexibility? The key assumption here was that labour supplies to the private sector and the government sector were completely independent of each other. We would get the same result if we assumed two kinds of workers, each of which could only work in either the private or the government sector. If the two labour markets are segmented, the government can increase GDP by hiring more workers and paying them higher wages. Since the two markets are segmented, there are no pressures for higher wages in the private sector that reduce private employment (see Beaudry and Portier (2013)[24] for a more elaborate version of this argument in a richer model of labour mobility costs between sectors). This extreme case highlights a simple RBC theory of the negative effects of fiscal consolidation on output in the short run. Government spending cuts reduce economic activity in the short run even with flexible prices and no income effects on labour supply if it is hard to switch workers and other production factors from the government sector to the private sector.

In the more realistic case in which there is some substitution between government and private sector labour supplies, an increase in government labour demand bids up private sector wages and discourages private sector labour demand. This reduces the government spending multiplier to below 1. To see this, we go back to the model where government and private sectors labour supply are perfect substitutes, with

$$u(c, l) = c + v(1 - n_p - n_g),$$

with $v(l)$ increasing and concave. Define private leisure $l_p = 1 - n_p$. Private labour markets equilibrium implies

$$v_{l_p} = z_p f_{n_p}.$$

Suppose $Gg > 0$. This leads to $\Delta n_g > 0$, which increases v_{l_p} for a fixed n_p. If $\Delta n_p = 0$, then there is no change in the right hand side of this equation. But then we need $\Delta n_p < 0$ to reduce v_{l_p}, leading to a contradiction. $\Delta n_p > 0$ would increase v_{l_p} and reduce f_{n_p}, again contradicting labour market equilibrium. The only way for the private labour market equilibrium condition to hold is for labour supply to fall, $\Delta n_p < 0$. Thus, substitutability between working in government or the private sector leads to crowding out of private labour supply in response to higher government spending.

Our formal derivation of this result assumes perfect substitutability in labour supply. But it may not be such a bad approximation for the marginal worker who is contemplating whether to work in government or the private sector at least for moderate increases in G_t. Meanwhile, income effects on labour supply can make the wage elasticity of existing government workers to wage increases quite small. In this case, the case of high substitutability of labour supply is likely to be a better guide to the size of the government spending multiplier.

15.6 Adding distortionnary taxes

In general government spending must be financed with distortionnary taxes. Higher taxes act like higher wedges, discouraging labour supply, hiring and investment. As a result, with distortionnary taxes flexible price government spending multipliers are much lower and usually negative at least in the long run (Baxter and King (1993) [22],Leeper et al (2011) [137]). With sticky prices, the effect of distortionnary taxes depend on interest rate policy (Woodford (2010) [184])). If the central bank targets a fixed real interest rate and the government spending shocks are stationary, distortionnary tax financing does not affect the earlier conclusion of a multiplier equal to 1. If the central bank follows a Taylor rule (which in its extreme version involves complete output gap stabilisation), sticky price multipliers with distortionnary taxes are lower than lump sum tax multipliers, though they are usually higher than with flexible prices. The intuition for this result is that an active monetary policy typically tries to eliminate output gaps caused by price or wage rigidity, so the response of output to government spending is closer to that under price and wage flexibility than with a constant interest rate policy.

If the zero lower bound constraint binds, then financing government spending through contemporaneous increases in for example labour taxes can have the perverse effect of actually increasing the multiplier(see Christiano et al (2010[56]). This occurs because the rise in labour taxes depresses flexible price output, increasing the output gap and inflation expectations. As a result, real interest rates fall by more than under lump sum taxes. Finally, if the increase in distortionnary taxes extends beyond the period when the zero bound constraint binds and the central bank follows a Taylor rule, then the government multiplier is significantly lower even in the short run because of the negative news effect of future distortionnary taxes depresses current aggregate demand.

Distortionnary taxation also makes the timing of taxes important, breaking Ricardian equivalence. In the flexible price economy, because a high distortionnary tax rate is more costly when output is low, a balanced budget policy can worsen aggregate fluctuations relative to a policy of lowering the tax to GDP ratio in a recession and raising it in a boom (See Ljunqvist and Sargent (2004) [140], chapter 15, for more on the optimality of tax rate smoothing). In general, analysis of models with distortionnary taxes shifts some of the debate on countercyclical fiscal policy towards questions of the efficiency of the overall tax mix used to finance government spending instead of just looking at the pure effect of changes in g_t.

15.7 Flexible capital adjustment, credit constrained households and other extensions

Flexible capital has the same effect as for other shocks with flexible prices of allowing for an aggregate consumption smoothing channel and reducing real interest rate volatility. Households can now partly compensate for lower wealth in response to temporarily higher government spending through lower investment. For transitory shocks, investment usually declines. **With flexible prices, a key difference in comparison to the fixed capital economy is that now more persistent government spending shocks have larger multipliers** (Aiyagari et al (1992) [3], Baxter and King (1993)[22] and Leeper et al (2011) [137], because investment declines by less or even increases.

A more persistent increase in government spending has a bigger negative impact on permanent income. This reduces the scope for consumption smoothing, leading to a bigger drop in consumption and higher aggregate saving through capital.The larger drop in permanent income also implies a stronger negative income effect, leading to a bigger increase in labour supply over several periods. This leads to a persistent and potentially large increase in the marginal product of capital, further boosting investment. If the shock is persistent enought, investment must increase.

To see this, consider a **permanent shock financed by lump sum tax increases (or transfer cuts).** The long run interest rate is determined by the household discount factor which is independent of lump sum tax changes. So higher government spending has no impact on the real interest rate in the long run.

Meanwhile the permanent increase in labour supply due to higher taxes increases the return on capital, and raises the optimal capital stock. In the short run, firms gradually build up the capital stock to its new long run level through higher investment. In fact permanent higher government spending shocks financed with lump sum taxes can generate government spending multipliers for output above 1 in the RBC model, if labour supply elasticity is high enough

(Baxter and King (1993)[22]).For highly persistent but transitory shocks, investment is still likely to increase, and multipliers above 1 are still possible if labour supply is elastic enough.

This analysis highlights an important difference between flexible price and sticky price economies for government spending shocks. Sticky price models predict a higher government spending multiplier in the short run for less persistent shocks. In contrast the baseline RBC model predicts that more persistent government spending shocks lead to bigger short run multipliers if the government uses lump sum taxes.

For realistic distortionnary taxes and shock persistence, the RBC model still generates impact multipliers significantly below 1 and negative long run multipliers. For New Keynesian economies the model with flexible capital still predicts that we can have impact multipliers above 1 (though typically below 1 with active monetary policy and distortionnary taxes), and that long run multipliers are below the short run multipliers with either negative or positive signs (but above the RBC long run multipliers) (Leeper et al 2011) [137]. Once again, the multiplier can be much larger than 1 in sticky price economies when the ZLB is binding.

Finally there are several other important elements that can affect government spending multipliers. A higher proportion of financially constrained households tends to increase multipliers in sticky price or wage economies, with a higher probability of a multiplier above 1 at least on impact. This is due to the traditional aggregate demand multiplier effect: higher government spending raises aggregate demand, which raises the current income of credit constrained households, which then increase their consumption. Complementarity between consumption and labour effort also increases the multiplier above 1 in New Keynesian models, since the higher output in response to higher government spending again stimulates consumption. In the other direction, in an open economy where the current account is endogenous government spending multiplier will tend to decrease (see Leeper et al (2011) [137])

16 Unconventional Monetary Policy

This section studies central bank intervention in credit markets through measures such as asset purchases or discount window lending in an attempt to reduce credit market frictions. We focus on central bank lending to financial institutions or direct central bank lending to the private sector through purchases in private debt markets. These operations are frequently known as credit easing.

Suppose there is a significant worsening in borrowing conditions, represented in our economy as some combination of higher $efp_{h,t+1}, v_{f,t}, \tau_{k,t}, \tau_{nd,t}$ and $\tau_{c,t}$. In economies with credit frictions, transfers of resources from financially unconstrained to financially constrained agents can reduce credit frictions. So a government program that subsidizes credit constrained agents at the expense of taxing other agents can increase output, though the welfare consequences of this transfer need to be carefully assessed, since it is not necessarily a Pareto improvement.

Can central bank intervention in private debt markets reduce these frictions and stimulate economic activity without hurting some agents (by forcing them to accept losses in order to implement a transfer to other more credit constrained agents) ? The baseline answer in an RBC model without heterogeneity across financial institutions is no: the distribution of assets and liabilities across financial intermediaries is irrelevant, except possibly through the specification of the stochastic process of the wedges.

When engaging in unconventional monetary policy the central bank essentially behaves like a financial intermediary, borrowing from some agents in the private sector (either through selling government debt or by expanding the supply of reserves) and lending to other private sector agents. Without specifying any advantage for the central bank in solving the moral hazard or asymmetric information problems that affect private financial intermediaries, the central bank cannot do any better than them. In that case credit easing is irrelevant. Central bank lending to final borrowers simply crowds out private lending to final borrowers.

If the central bank lends to private financial intermediaries, these intermediaries cut back their own lending to other private intermediaries, or they redeposit the central bank loans back at the central bank as reserves. This fundamental point holds in general for models of unconventional monetary policy in which the financing frictions are explicitly modelled (for example Gertler and Kiyotaki (2010) [94]).

For central bank credit easing to reduce credit frictions while maintaining Pareto optimality, the central bank must have an advantage in terms of solving the fundamental causes of the worsening borrowing conditions. More formally, it must have a better borrowing or lending technology. For example, central bank liabilities are often implicitly backed by the government's ability to tax or by seignorage revenues, which reduces moral hazard or asymmetric information problems concerning the quality of loans made by the central bank (note from this perspective, the fiscal authorities could also engage in credit easing operations - see for example Kocherlakota (2009), Del Negro et al (2011) [152] , Gertler and Karadi (2011) [93]). To the degree that private sector agents are concerned that the government may have hit a fiscal limit on its ability to tax, so that the government debt itself is subject to default risk or inflation risk (in order to devalue nominal debt), this reduces the ability of the central bank or the finance department to reduce credit frictions. [62]

[62] In a monetary union, the fiscal backing of the central bank is more uncertain, in which

In the presence of asymmetric information or ambiguity aversion problems in debt markets, there may be a role for a large financial intermediary to coordinate a pooling of other financial institutions' assets. The pooled portfolio diversifies risk and reduces asymmetric information or ambiguity about financial returns [63]. For example, the financial intermediary could swap other financial institutions' assets that are subject to asymmetric information in return for shares in a well diversified portfolio of these assets. This is the essence of several proposals to solve sovereign default risk related asymmetric information in European debt markets in the autumn of 2011 (see for example Brunnermeier et al (2011), Uhlig et al (2011)). A central bank or a fiscal authority may have some advantages in coordinating such an asset swap.

Central bank or in general government lending to financial institutions in distress can also be more effective than private sector lending if it is easier for the central bank or government to enforce measure such as audits or stress testing that directly reduce asymmetric information and moral hazard concerns. There is an exception to these results if the credit crunch was caused by the bursting of an asset price bubble. In that case, it may be possible for the government to counter the collapse of the bubble by introducing another asset in its place as a new bubble (e.g Kocherlakota (2009)). A successful introduction of a new bubble asset would restore the old equilibrium with a higher economic activity. But without fiscal backing , this intervention is subject to multiple equilibria, including negative ones where the new asset is (close to) worthless. In other circumstances, the government cannot finance credit market interventions by introducing a pure bubble at all (Miao and Wang (2011)).

Finally, financial frictions can under certain circumstances generate multiple equilibria (often called sunspot equilibria in the literature). In that case, it is possible that if (almost) everyone in the economy believes that central bank intervention can avoid or stabilise a crisis, then an intervention which is otherwise neutral could indeed avoid a crisis by coordinating agents' beliefs on the more positive, optimistic equilibrium. Here unconventional monetary policy simply serves as a coordination device for eliminating a bad equilibrium.

One way to formalise this is in a Diamond Dybvig style framework (Freixas and Rochet (1997), ch. 7 [86]), in which there are two types of banks. Due to

---

case credit easing is significantly less likely to be successful (this may explain the much greater reluctance of the ECB in 2011 to engage in credit easing relative to the Federal Reserve in 2008). In general, the backing of government debt through taxes in an environment with credit constraints breaks Ricardian equivalence. Under certain circumstances, higher government debt levels can alleviate liquidity constraints by allowing agents to save more easily (see Perez (2010) for an insightful analysis using the Kiyotaki and Moore (2008) [129] model of liquidity constraints).

[63] See Caballero and Krishnamurty (2008) about ambiguity aversion and financial crises. See Diamond (1984) [65] for a similar argument in a costly state verification environment about how a financial institution offering claims to a well diversified asset portfolio can reduce assymetric information problems.

illiquidity or default risk, and imperfect liquidation procedures,Type A banks are subject to a self fulfilling belief by depositors that the bank will become insolvent. As a result, absent further intervention depositors all want to withdraw their funds, and type A banks would in fact become insolvent. Type B banks are for some unspecified reason not subject to this belief: depositors in those banks have somehow coordinated on the alternative equilibrium in which no one wants to withdraw their deposits and the bank does not default. In this case, if depositors in A banks share these beliefs about B banks, they may agree to redeposit their withdrawn funds in type B banks. Type B banks could then lend out these new deposits to type A banks, making them solvent again. If we call the type B bank a central bank, we get a theory of how monetary policy can prevent sunspot based bank runs.

Of course, there are no guarantees in this story that the central bank is itself not subject to a sunspot based bank run. Something combination of fiscal backing of the central bank through taxation, the use of seignorage revenues or surprise inflation (to reduce the real value of government debt) are required to eliminate the possibility of a run on the central bank. [64] The situation is more complicated if there are other reasons for potential bank defaults, such as asymmetric information, moral hazard or ambiguity aversion. Furthermore, it is possible that the central bank intervenes but fails to coordinate expectations on the positive equilibrium initially. Then the perception of disappointing results of the intervention can make agents more pessimistic and worsen a crisis. In that sense, policy interventions that rely purely on multiple equilibria and self fulfilling prophecies are fundamentally more fragile than other interventions that also work in environments with unique equilibria. This makes the latter class of policies preferable whenever they are available.

17 Appendix A, a benchmark DSGE model equations

17.1 Households:

17.1.1 The representative saver

Representative saver optimisation problem:

[64] Note the parallel to our earlier discussion in which a legal requirement to use money was required to eliminate the self fulfilling belief that money has no value.

$$V_0^{sa} = \max_{\{c_{sa,t},n_t^s,D_{t+1},b_{t+1}\}_{t=0}^{\infty}} E_0\Sigma_{t=0}^{\infty}\beta_{sa,t}u(c_{sa,t},D_{sa,t+1},l_{sa,t})$$

$$\text{subject to a sequence of constraints}$$

$$c_{sa,t}(1+\tau_{c,t})+p_{D,t}(1+\tau_{D,t})\left[D_{sa,t+1}-(1-\delta_D)D_{sa,t}\right]+b_{sa,t+1}\leq$$

$$b_{sa,t}R_t efp_{sa,h,t}+w_t(1-\tau_{nst})n_{sa,t}+d_t^h+T_t,\ (\lambda_{sa,t}),$$

$$n_{sa,t}=1-l_{sa,t},$$

$$u_c>0,u_{cc}<0,u_D>0,u_{DD}<0,u_l>0,u_{ll}<0,u_{c,l}\leq 0.$$

Representative saver optimisation conditions

$$\begin{aligned}
c_{sa,t} &: \quad \lambda_{sa,t}(1+\tau_{c,t})=u_{c,t}\\
n_{sa,t} &: \quad u_{l,sa,t}=\lambda_{sa,t}w_t(1-\tau_{ns,t})\\
b_{sa,t+1} &: \quad \lambda_{sa,t}=E_t\frac{\beta_{sa,t+1}}{\beta_{sa,t}}efp_{h,t+1}R_{t+1}\lambda_{sa,t+1}.\\
D_{sa,t+1} &: \\
p_{D,t}(1+\tau_{D,t}) &= \frac{u_{D_{sa,t+1}}}{\lambda_{sa,t}}+E_t\frac{\beta_{sa,t+1}}{\beta_{sa,t}}\frac{\lambda_{sa,t+1}}{\lambda_{sa,t}}p_{D,t+1}(1+\tau_{D,t+1})(1-\delta_D).
\end{aligned}$$

17.1.2 The representative borrower

Representative borrower problem:

$$V_0^{bo} = \max_{\{c_{bo,t},n_{bo,t},d_{bo,t+1},b_{bo,t+1}\}} E_0\Sigma_{t=0}^{\infty}\beta_{bo,t}u(c_{bo,t},D_{bo,t+1},l_{bo,t})$$

$$\text{subject to a sequence of constraints}$$

$$c_{bo,t}(1+\tau_{c,t})+p_{D,t}(1+\tau_{D,t})\left[D_{bo,t+1}-(1-\delta_D)D_{bo,t}\right]+b_{bo,t+1}\leq b_{bo,t}R_t efp_{bo.h,t}$$

$$+w_t(1-\tau_{nst})n_{bo,t}+T_t,\ (\lambda_{bo,t}),$$

$$-b_{bo,t+1}\leq \bar{b}_{t+1}+m_{bo,t}p_{D,t}D_{bo,t+1},$$

$$(\mu_{t+1}),\ \bar{b}_{t+1|}\geq 0,\ 0\leq m_{bo,t}\leq 1.$$

$$n_{bo,t}=1-l_{bo,t},$$

$$u_c>0,u_{cc}<0,u_l>0,u_{ll}<0,u_{c,l}\leq 0.$$

Representative borrower optimality conditions:

$$
\begin{aligned}
c_{bo,t} &: \quad \lambda_{bo,t}(1+\tau_{c,t}) = u_{c,bo,t} \\
n_{bo,t} &: \quad u_{l,bo,t} = \lambda_{bo,t} w_t (1-\tau_{ns,t}) \\
b_{bo,t+1} &: \quad \lambda_{bo,t} = E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} efp_{bo,h,t+1} R_{t+1} \lambda_{bo,t+1} + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \mu_{t+1} \\
&\equiv \quad E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} ef\bar{p}_{bo,h,t+1} R_{t+1} \lambda_{bo,t+1} \\
D_{bo,t+1} &: \\
p_{D,t}(1+\tau_{D,t}) &= \frac{u_{D_{bo,t+1}}}{\lambda_{bo,t}} + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \frac{\lambda_{bo,t+1}}{\lambda_{bo,t}} p_{D,t+1}(1+\tau_{D,t+1})(1-\delta_D) + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} m_{bo,t} p_{D,t} \mu_{t+1}.
\end{aligned}
$$

17.2 Representative firm

Optimisation problem:

$$
\begin{aligned}
V_0^f &= \max_{\{k_{t+1}, m_t, n_t\}} E_0 \Sigma_{t=0}^{\infty} \beta_{f,t} d_t, \\
d_t &= g(A_t m_t, e_{m,t}) z_t f(k_t, A_t n_t) - w_t(1+\tau_{nd,t})(n_t+m_t) - p_t^k (1+\tau_{k,t})[k_{t+1} - (1-\delta_t)k_t] \\
&\quad - b_{f,t+1} + b_{f,t} R_t efp_{e,t}, \\
g_m &> 0,\ g_{mm} < 0,\ g_{e_m} > 0,\ f_n > 0,\ f_{nn} < 0, f_k > 0, f_{kk} < 0 \\
,\ g(am,.)f(ak,an) &\leq ag(m,.)f(k,n) \text{ for all } a \geq 0.\ A_t = G_y A_{t-1}, G_y \geq 1.
\end{aligned}
$$

Optimisation conditions:

$$
\begin{aligned}
n_t &: \quad g_t z_t f_{n,t} = w_t(1+\tau_{nd,t}), \\
m_t &: \quad g_{m,t} z_t f_t = w_t(1+\tau_{nd,t}), \\
b_{f,t+1} &: \quad \beta_{f,t} = R_{t+1} E_t efp_{e,t+1} \beta_{f,t+1}, \\
k_{t+1} &: \quad p_{k,t}(1+\tau_{k,t}) = E_t \frac{\beta_{f,t+1}}{\beta_{f,t}} [g_{t+1} z_{t+1} f_{k,t+1} + (1-\delta_{t+1})(1+\tau_{k,t+1}) p_{k,t+1}].
\end{aligned}
$$

Link between discount factors and external finance premia of firms and households:

$$
\begin{aligned}
v_{f,t} &= \frac{\beta_{f,t}}{\lambda_{sa,t}\beta_{sa,t}}, \\
E_t \frac{v_{f,t+1}}{v_{f,t}} &\simeq \frac{E_t efp_{sa,h,t+1}}{E_t efp_{e,t+1}}.
\end{aligned}
$$

17.3 Capital production sector

17.3.1 Business capital

The representative capital producer solves

$$\begin{aligned} \max_{k_{t+1}} \pi_{k,t} &= p_{k,t} x_t^k - I_t \\ x_t^k &= k_{t+1} - (1-\delta_t) k_t = a_t^k \tilde{G}(\frac{I_t}{k_t}) k_t \\ ,\tilde{G}'(.) &> 0,\ \tilde{G}''(.) \leq 0. \end{aligned}$$

Capital supply:

$$a_t^k p_{k,t} \tilde{G}'(\frac{I_t}{k_t}) = 1.$$

17.3.2 Durable goods

Representative durable goods producer problem

$$\max_{D_t} \pi_{D,t} = \max_{D_t} p_{D,t} x_t^D - I_t^D.$$

$$\begin{aligned} x_t^D &= D_{t+1} - (1-\delta_D) D_t = a_t^D \tilde{G}(\frac{I_t^D}{D_t}) D_t \\ ,\tilde{G}'(.) &> 0,\ \tilde{G}''(.) \leq 0. \end{aligned}$$

Durables supply:

$$a_t^D p_{D,t} \tilde{G}'(\frac{I_t^D}{D_t}) = 1.$$

17.4 Market clearing conditions

$$\begin{aligned} \Sigma_i \theta_i c_{it} &= c_t, \\ \Sigma_i \theta_i D_{it} &= D_t, \\ c_t + I_t + I_t^D + G_t &= y_t \\ \Sigma_i \theta_i n_{i,t} &= n_{st} = n_t + m_t, \\ \Sigma_i \theta_i b_{i,t} &= 0,\ i \in \{sa, bo, f\}, \\ \theta_{sa} d_t^h &= d_t + \pi_{k,t} + \pi_{D,t} \\ i &= sa, bo. \end{aligned}$$

18 Appendix B, a benchmark DSGE model with search and matching frictions

This is an extension of the baseline model with pervasive matching frictions in all goods and labour markets. We leave the different matching rates (job finding rate, vacancy filling rate, product finding rate, etc...) as exogeneous wedges that measure the inefficiency of markets relative to the perfect matching Walrasian equilibrium benchmark. These matching rates can be partially endogenised by assuming for example Cobb-Douglas matching functions and various combinations of Nash bargaining, directed search or price and wage rigidity.

18.1 Households:

18.1.1 The representative saver

Representative saver optimisation problem:

$$V_0^{sa} = \max_{\left\{c_{sa,t}, S^c_{sa,t},, S^d_{sa,t}, e_{nsa,t}, n^s_t, D_{t+1}, b_{t+1}\right\}_{t=0}^{\infty}} E_0 \Sigma_{t=0}^{\infty} \beta_{sa,t} [u(c_{sa,t}, D_{sa,t+1}, l_{sa,t}) - v^c(S^c_{sa,t}) - v^d(S^D_{sa,t}) - a_{nsa,t} e_{nsa,t}]$$

$$\text{subject to a sequence of constraints}$$

$$c_{sa,t}(1+\tau_{c,t}) + p_{D,t}(1+\tau_{D,t})\left[D_{sa,t+1} - (1-\delta_D)D_{sa,t}\right] + b_{sa,t+1} \leq$$

$$b_{sa,t} R_t efp_{sa,h,t} + w_t(1-\tau_{nst}) n_{sa,t} + d^h_t + T_{sa,t},\ (\lambda_{sa,t}),$$

$$c_{sa,t} \leq \phi^c_t S^c_{sa,t},\ (\mu^{SC}_{sa,t}).$$

$$D_{sa,t+1} - (1-\delta_D) D_{sa,t} \leq \phi^D_t S^D_{sa,t},\ (\mu^{SD}_{sa,t}).$$

$$n_{sa,t} - (1-\delta_{n,t}) n_{sa,t-1} \leq \varphi_{nsa,t} e_{nsa,t},\ (\mu^{nsa}_t).$$

$$n_{sa,t} = 1 - l_{sa,t},$$

$$u_c > 0, u_{cc} < 0, u_D > 0, u_{DD} < 0, u_l > 0, u_{ll} < 0, u_{c,l} \leq 0.$$

$$v_S > 0, v_{S^c S^c} \geq 0$$

Representative saver optimisation conditions

$$
\begin{aligned}
c_{sa,t} &: \quad \lambda_{sa,t}(1+\tau_{c,t}) = u_{c,t} - \mu^{SC}_{sa,t}, \\
S^{c}_{sa,t} &: \quad v_{S^{c}_{sa,t}} = \phi^{c}_{t}\mu^{SC}_{sa,t}, \\
S^{d}_{sa,t} &: \quad v_{S^{D}_{sa,t}} = \mu^{SD}_{sa,t}\phi^{D}_{t}, \\
e_{nsa,t} &: \quad a_{n,t} = \mu^{ns}_{sa,t}\varphi_{ns,t}, \\
n_{sa,t} &: \quad u_{lsa,t} = \lambda_{sa,t}w_t(1-\tau_{ns,t}) - \mu^{ns}_{sa,t} + E_t\frac{\beta_{sa,t+1}}{\beta_{sa,t}}\mu^{ns}_{sa,t+1}(1-\delta_{n,t+1}), \Rightarrow \\
u_{lsa,t} + \frac{a_{n,t}}{\varphi_{ns,t}} &= \quad \lambda_{sa,t}w_t(1-\tau_{ns,t}) + E_t\frac{\beta_{sa,t+1}}{\beta_{sa,t}}\frac{a_{n,t+1}}{\varphi_{ns,t+1}}(1-\delta_{n,t+1}) \\
b_{sa,t+1} &: \quad \lambda_{sa,t} = E_t\frac{\beta_{sa,t+1}}{\beta_{sa,t}}efp_{h,t+1}R_{t+1}\lambda_{sa,t+1}. \\
D_{sa,t+1} &: \quad p_{D,t}(1+\tau_{D,t}) + \frac{\mu^{SD}_{sa,t}}{\lambda_{sa,t}} = \frac{u_{D_{t+1}}}{\lambda_{sa,t}} \\
&\quad\quad + E_t\frac{\beta_{sa,t+1}}{\beta_{sa,t}}\frac{\lambda_{sa,t+1}}{\lambda_{sa,t}}\left[p_{D,t+1}(1+\tau_{D,t+1}) + \frac{\mu^{SD}_{sa,t+1}}{\lambda_{sa,t+1}}\right](1-\delta_D) \\
&\Rightarrow \\
D_{sa,t+1} &: \quad p_{D,t}(1+\tau_{D,t}) + \frac{v_{S^{D}_{sa,t}}}{\lambda_{sa,t}\phi^{D}_{t}} = \frac{u_{D_{t+1}}}{\lambda_{sa,t}} \\
&\quad\quad + E_t\frac{\beta_{sa,t+1}}{\beta_{sa,t}}\frac{\lambda_{sa,t+1}}{\lambda_{sa,t}}\left[p_{D,t+1}(1+\tau_{D,t+1}) + \frac{v_{S^{D}_{sa,t+1}}}{\lambda_{sa,t+1}\phi^{D}_{t+1}}\right](1-\delta_D)
\end{aligned}
$$

18.1.2 The representative borrower

Representative borrower problem:

$$V_0^{bo} =$$
$$\max_{\{c_{bo,t},S^c_{bo,t},,S^d_{bo,t}e_{nbo,t},,n_{bo,t},d_{bo,t+1},b_{bo,t+1}\}} E_0\Sigma_{t=0}^{\infty}\beta_{bo,t}[u(c_{bo,t},D_{bo,t+1},l_{bo,t}) - v^c(S^c_{bo,t}) - v^d(S^D_{bo,t}) - a_{nbo,t}e_{nbo,t}]$$

subject to a sequence of constraints

$$c_{bo,t}(1+\tau_{c,t}) + p_{D,t}(1+\tau_{D,t})\left[D_{bo,t+1} - (1-\delta_D)D_{bo,t}\right] + b_{bo,t+1} \leq b_{bo,t}R_t efp_{bo.h,t}$$
$$+w_t(1-\tau_{nst})n_{bo,t} + T_{bo,t},\ (\lambda_{bo,t}),$$
$$-b_{bo,t+1} \leq \bar{b}_{t+1} + m_{bo,t}p_{D,t}D_{bo,t+1},$$
$$(\mu^b_{t+1}),\ \bar{b}_{t+1|} \geq 0,\ 0 \leq m_{bo,t} \leq 1.$$
$$c_{bo,t} \leq \phi^c_t S^c_{bo,t},\ (\mu^{SC}_{bo,t}).$$
$$D_{bo,t+1} - (1-\delta_D)D_{bo,t} \leq \phi^D_t S^D_{bo,t},\ (\mu^{SD}_{bo,t}).$$
$$n_{bo,t} - (1-\delta_{n,t})n_{bo,t-1} \leq \varphi_{nbo,t}e_{nbo,t},\ (\mu^{nbo}_t).$$
$$n_{bo,t} = 1 - l_{bo,t},$$
$$u_c > 0, u_{cc} < 0, u_l > 0, u_{ll} < 0, u_{c,l} \leq 0,$$
$$v_S > 0, v_{S^cS^c} \geq 0.$$

Representative borrower optimality conditions:

$$
\begin{aligned}
c_{bo,t} &: \quad \lambda_{bo,t}(1+\tau_{c,t}) = u_{c,bo,t} - \mu^{SC}_{bo,t} \\
S^c_{bo,t} &: \quad v_{S^c_{bo,t}} = \phi^c_t \mu^{SC}_{bo,t}, \\
S^d_{bo,t} &: \quad v_{S^D_{bo,t}} = \mu^{SD}_{bo,t}\phi^D_t, \\
e_{nbo,t} &: \quad a_{n,t} = \mu^{ns}_{bo,t}\varphi_{ns,t}, \\
n_{bo,t} &: \quad u_{l,bo,t} = \lambda_{bo,t} w_t(1-\tau_{ns,t}) - \mu^{ns}_{bo,t} + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \mu^{ns}_{bo,t+1}(1-\delta_{n,t+1}), \Rightarrow \\
u_{l,bo,t} + \frac{a_{n,t}}{\varphi_{ns,t}} &= \quad \lambda_{bo,t} w_t(1-\tau_{ns,t}) + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \frac{a_{n,t+1}}{\varphi_{ns,t+1}}(1-\delta_{n,t+1}) \\
b_{bo,t+1} &: \quad \lambda_{bo,t} = E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} efp_{bo,h,t+1} R_{t+1}\lambda_{bo,t+1} + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \mu^b_{t+1} \\
&\equiv \quad E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} ef\bar{p}_{bo,h,t+1} R_{t+1}\lambda_{bo,t+1} \\
D_{bo,t+1} &: \quad p_{D,t}(1+\tau_{Dt}) + \frac{\mu^{SD}_{bo,t}}{\lambda_{bo,t}} = \frac{u_{D_{bo,t+1}}}{\lambda_{bo,t}} \\
&\quad + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \frac{\lambda_{bo,t+1}}{\lambda_{bo,t}} p_{D,t+1}\left[(1-\delta_D)(1+\tau_{D,t+1}) + \frac{\mu^{SD}_{bo,t+1}}{\lambda_{bo,t+1}}\right] \\
&\quad + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} m_{bo,t} p_{D,t} \mu^b_{t+1}. \\
&\Rightarrow \\
D_{bo,t+1} &: \quad p_{D,t}(1+\tau_{Dt}) + \frac{v_{S^D_{bo,t}}}{\lambda_{bo,t}\phi^D_t} = \frac{u_{D_{bo,t+1}}}{\lambda_{bo,t}} \\
&\quad + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} \frac{\lambda_{bo,t+1}}{\lambda_{bo,t}} p_{D,t+1}\left[(1-\delta_D)(1+\tau_{D,t+1}) + \frac{v_{S^D_{bo,t+1}}}{\lambda_{bo,t+1}\phi^D_{t+1}}\right] \\
&\quad + E_t \frac{\beta_{bo,t+1}}{\beta_{bo,t}} m_{bo,t} p_{D,t} \mu^b_{t+1}.
\end{aligned}
$$

18.2 Representative firm

Optimisation problem:

$$
\begin{aligned}
& V_0^f = \max_{\{k_{t+1}, V_t^k, V_t^n, m_t, n_t\}} E_0 \Sigma_{t=0}^{\infty} \beta_{f,t} d_t, \\
d_t &= g(A_t m_t, e_{m,t}) z_t f(k_t, A_t n_t) - w_t(1+\tau_{nd,t})(n_t + m_t) - p_t^k(1+\tau_{k,t})[k_{t+1} - (1-\delta_t)k_t] \\
& -b_{f,t+1} + b_{f,t} R_t efp_{e,t} \\
& -F^N(V_t^n) - \Gamma(V_t^k), \\
& \text{subject to} \\
k_{t+1} - (1-\delta_t)k_t &\leq \varphi_t^k V_t^k \ , (\mu_t^k), \\
n_t^d - n_{t-1}^d(1-\delta_{n,t}) &\leq \varphi_{nd,t} V_t^n, \ (\mu_t^{nd}). \\
g_m &> 0, \ g_{mm} < 0, \ g_{e_m} > 0, \ f_n > 0, \ f_{nn} < 0, f_k > 0, f_{kk} < 0, \\
\Gamma_{V^k} &> 0, \Gamma_{V^k V^k} \geq 0, F_{V^n}^N > 0, F_{V^n V^n}^N \geq 0. \\
, \ g(am, .)f(ak, an) &\leq ag(m, .)f(k, n) \text{ for all } a \geq 0. \ A_t = G_y A_{t-1}, G_y \geq 1.
\end{aligned}
$$

Optimisation conditions:

$$
\begin{aligned}
V_t^k &: \quad \Gamma_{V_t^k} = \mu_t^k \varphi_t^k, \\
V_t^n &: \quad F_{V_t^N}^N = \mu_t^{nd} \varphi_{nd,t}. \\
n_t^d &: \quad g_t z_t f_{n,t} + E_t \frac{\beta_{t+1}^f}{\beta_t^f} \mu_{t+1}^{nd}(1-\delta_{n,t+1}) = w_t(1+\tau_{nd,t}) + \mu_t^{nd}, \Rightarrow \\
g_t z_t f_{n,t} + E_t \frac{\beta_{t+1}^f}{\beta_t^f} \frac{F_{V_{t+1}^N}^N}{\varphi_{nd,t+1}}(1-\delta_{n,t+1}) &= w_t(1+\tau_{nd,t}) + \frac{F_{V_t^N}^N}{\varphi_{nd,t}} \\
m_t &: \quad g_{m,t} z_t f_t = w_t(1+\tau_{nd,t}), \\
b_{f,t+1} &: \quad \beta_{f,t} = R_{t+1} E_t efp_{e,t+1} \beta_{f,t+1}, \\
k_{t+1} &: \quad p_{k,t}(1+\tau_{k,t}) + \mu_t^k = \\
& \quad E_t \frac{\beta_{f,t+1}}{\beta_{f,t}} (g_{t+1} z_{t+1} f_{k,t+1} + (1-\delta_{t+1})[(1+\tau_{k,t+1}) p_{k,t+1} + \mu_{t+1}^k]), \\
& \Rightarrow \\
p_{k,t}(1+\tau_{k,t}) + \frac{\Gamma_{V_t^k}}{\varphi_t^k} &= \\
& \quad E_t \frac{\beta_{f,t+1}}{\beta_{f,t}} (g_{t+1} z_{t+1} f_{k,t+1} + (1-\delta_{t+1})[(1+\tau_{k,t+1}) p_{k,t+1} + \frac{\Gamma_{V_{t+1}^k}}{\varphi_{t+1}^k}])
\end{aligned}
$$

Link between discount factors and external finance premia of firms and households:

$$
\begin{aligned}
v_{f,t} &= \frac{\beta_{f,t}}{\lambda_{sa,t}\beta_{sa,t}}, \\
E_t\frac{v_{f,t+1}}{v_{f,t}} &\simeq \frac{E_t efp_{sa,h,t+1}}{E_t efp_{e,t+1}}.
\end{aligned}
$$

18.3 Capital production sector

18.3.1 Business capital

The representative capital producer solves

$$
\begin{aligned}
\max_{k_{t+1}} \pi_{k,t} &= p_{k,t}x_t^k - I_t \\
x_t^k &= k_{t+1} - (1-\delta_t)k_t = a_t^k \tilde{G}(\frac{I_t}{k_t})k_t \\
,\tilde{G}'(.) &> 0,\ \tilde{G}''(.) \leq 0.
\end{aligned}
$$

Capital supply:

$$
a_t^k p_{k,t}\tilde{G}'(\frac{I_t}{k_t}) = 1.
$$

18.3.2 Durable goods

Representative durable goods producer problem

$$
\max_{D_t} \pi_{D,t} = \max_{D_t} p_{D,t}x_t^D - I_t^D.
$$

$$
\begin{aligned}
x_t^D &= D_{t+1} - (1-\delta_D)D_t = a_t^D \tilde{G}(\frac{I_t^D}{D_t})D_t \\
,\tilde{G}'(.) &> 0,\ \tilde{G}''(.) \leq 0.
\end{aligned}
$$

Durables supply:

$$
a_t^D p_{D,t}\tilde{G}'(\frac{I_t^D}{D_t}) = 1.
$$

18.4 Market clearing conditions

$$\begin{aligned}
\Sigma_i \theta_i c_{it} &= c_t, \\
\Sigma_i \theta_i D_{it} &= D_t, \\
c_t + I_t + I_t^D + G_t &= y_t - F^N(V_t^n) - \Gamma(V_t^k). \\
\Sigma_i \theta_i n_{i,t} &= n_{st} = n_t + m_t, \\
\Sigma_i \theta_i b_{i,t} &= 0,,\ i \in \{sa, bo, f\}, \\
\theta_{sa} d_t^h &= d_t + \pi_{k,t} + \pi_{D,t} \\
i &= sa, bo. \\
0 &\leq \phi_i \leq 1 \text{ for } i = c, d \\
0 &\leq \varphi_j \leq 1 \text{ for } j = ns, nd, k.
\end{aligned}$$

19 State space representation and extension to hybrid models

Here $\tilde{k}_t$ is a vector or endogenous, predetermined, state variables (note that these may include a subset of y_{t-1}, for example lagged consumption or investment). y_t is a vector of endogenous non state variables. τ_t is a vector of stochastic wedges. F_t is a vector of exogenous factors driving the wedges. Note that a subset of the variables τ_t could itself be part of F_t. Higher order lags are included by expanding the state vactor with new auxiliary variables, e.g $k_{t-1} = k_{-1,t}$.

$$\begin{aligned}
\tilde{k}_{t+1} &= g(\tilde{k}_t, \tau_t, e_{t+1}^k), \\
y_t &= h(\tilde{k}_t, \tau_t, v_{y,t}), \\
\tau_t &= \tau(\tilde{k}_t, F_t, e_t^\tau), \\
F_{t+1} &= \kappa(F_t, e_{t+1}^F), \\
e_t^\tau, e_t^F, v_{y,t} &\backsim \text{IID shocks.}
\end{aligned}$$

For the DSGE model in the previous section:
$\tilde{k}_t = (k_t, n_{sa,t-1}, n_{bo,t-1}, D_{sa,t}, D_{bo,t}, b_{sa,t}, b_{bo,t}, b_{f,t})'$.
$\tau_t = (g_t, z_t, \tau_{nd,t}, \tau_{k,t}, \varphi_{nd,t}, \varphi_{k,t}, efp_{e,t}, \tilde{\beta}_{f,t+1}, \tilde{\beta}_{sa,t+1}, \tilde{\beta}_{bo,t+1}, efp_{sa,h,t+1}, efp_{bo,h,t+1}, m_{bo,t}, \tau_{c,t}, \tau_{D,t}, \tau_{ns,t}, \phi_{D,t}, \phi_{c,t}, \varphi_{ns,t}, a_t^k, a_t^D, G_t)'$.

Let $s_t = (\tilde{k}_t; F_t)$. A general hybrid model represents y_t as

$$y_t = H(s_t, y_{t-1}, v_t),$$

where $H(.)$ encompasses $h(.)$ as a special case. We can use the DSGE function $h(.)$ to form priors on h. In the special case of a linear approximation to the DSGE model, this reduces to a DSGE-VARX model:

$$
\begin{aligned}
y_t &= Ay_{t-1} + Bs_t + v_t, \\
s_t &= Cs_{t-1} + e_t,
\end{aligned}
$$

replacing the unobserved s_t with $s_{t|T}$. As before we can use priors from the DSGE model, setting A_0 to 0 and B_0, C_0 to the corresponding matrix from the DSGE equations

$$
\begin{aligned}
y_t &= hs_t, \\
s_t &= gs_{t-1} + e_t.
\end{aligned}
$$

Alternatively ,we can combine the core/target values for y_t from a DSGE model with ECM dynamics, to get a DSGE-ECM. The ECM model is usually written in first differences to make it more robust to non-stationnary time series. This leads to the following specification in (log) differences $y_t - y_{t-1} = \Delta y_t$:

$$
\Delta y_t = \alpha - \beta(y_{t-1} - y_{t-1}^{core}) + \gamma \Delta y_t^{core} + \Sigma_{j=1}^{J} A_j \Delta y_{t-j} + bx_t + \varepsilon_t^y.
$$

A common simplification is to assume that β, γ and $\{A_j\}$ are diagonal matrices, so that each equation is a univariate regression in its lags and the in the response to its core/target value.

References

[1] S. Rao Aiyagari. Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, 109:659–684, 1994.

[2] S. Rao Aiyagari. Comments on farmer's and guo's the econometrics of indeterminacy: an applied study. Federal Reserve Bank of Minneapolis staff report 196, 1995.

[3] S. Rao Aiyagari, Lawrence J. Christiano, and Martin Eichenbaum. The output,employment and interest rate effects of government consumption. *Journal of Monetary Economics*, 30:73–86, 1992.

[4] Giorgio E. Primiceri Alejandro Justiniano and Andrea Tambalotti. Investment shocks and business cycles. working paper, 2009.

[5] Yann Algan, Olivier Allais Wouter Den Haan, and Pontus Rendahl. Solving and simulating models with heterogeneous agents and aggregate uncertainty. working paper, 2009.

[6] Ilya Strebulaev an Toni Whited. Dynamic models and structural estimation in corporate finance. draft of chapter for handbook, 2012.

[7] Derek Anderson, Benjamin Hunt, Mika Kortelainen, Michael Kumhof, Douglas Laxton, Dirk Muir, Susanna Mursula, and Stephen Snudden. Getting to know the gimf: The simulation properties of the global integrated monetary and fiscal model. IMF working paper, 2013.

[8] Marios Angeletos. Uninsured idiosyncratic investment risk and aggregate saving. working paper, 2007.

[9] Marios Angeletos and Laurent Emanuel Calvet. Idiosyncratic production risk,growth and the business cycle. working paper, 2005.

[10] Marios Angeletos, Fabrice Collard, and Harris Dellas. Quantifying confidence. working paper, 2014.

[11] Marios Angeletos and Jennifer Lao. Sentiments. working paper, 2012.

[12] Christina Arellano, Yan Bai, and Patrick J. Kehoe. Financial frictions and fluctuations in volatility. Federal Reserve Bank of Minneapolis staff report, 2012.

[13] Andrew Atkeson and Patrick J. Kehoe. Modeling the transition to a new economy: Lessons from two technological revolutions. *The American Economic Review*, 97(1):64–88, March 2007.

[14] Orazio Attanasio and Guglielmo Weber. Consumption and saving: Models of intertemporal allocation and their implications for public policy. NBER working paper, 2010.

[15] Jan Babecky, Philip du Caju, Theodora Kosma, Martina Lawless, Julian Messina, and Tairi Room. Downward nominal and real wage rigidity:survey evidence from european firms. working paper, 2009.

[16] Jan Babecky, Philip du Caju, Theodora Kosma, Martina Lawless, Julian Messina, and Tairi Room. The margins of labour adjustment: Survey evidence from european firms. working paper, 2009.

[17] Yan Bai, Jose Victor Rios Rull, and Kjetil Storesletten. Demand shocks as productivity shocks. working paper, 2012.

[18] Gadi Barlevy. Economic theory and asset bubbles. *Federal Reserve Bank of Chicago Economic Perspectives*, 31, 2007.

[19] Alessandro Barratieri, Susanto Basu, and Peter Gottschalk. Some evidence on the importance of sticky wages. working paper, 2010.

[20] Robert J. Barro. Long term contracting, sticky prices, and monetary policy. *Journal of Monetary Economics*, 3:305–316, 1977.

[21] Susanto Basu and Miles Kimball. Long run labour supply and the elasticity of intertemporal substitution for consumption. working paper, 2002.

[22] Marianne Baxter and Robert King. Fiscal policy in general equilibrium. *The American Economic Review*, 83(3):315–334, June 1993.

[23] Paul Beaudry and Franck Portier. When can changes in expectations cause business cycles in neoclassical settings? *Journal of Economic Theory*, 135:458–477, 2007.

[24] Paul Beaudry and Franck Portier. Understanding non-inflationary demand driven business cycles. working paper, 2013.

[25] Adriana Bellou and Baris Kaymak. Wages, implicit contracts and the business cycle: Evidence from a european panel. working paper, 2011.

[26] Jess Benhabib, Pengfei Wang, and Yi Wen. Sentiments and aggregate demand fluctuations. St Louis Fed Working paper, 2012.

[27] Jess Benhabib and Yi Wen. Indeterminacy, aggregate demand, and the real business cycle. *Journal of Monetary Economics*, 51(3):503–530, 2004.

[28] Ben Bernanke, Mark Gertler, and Simon Gilchrist. The financial accelerator in a quantitative business cycle framework. Chapter in Handbook of Macroeconomics volume 1, 1999.

[29] Saroj Bhattarai, Gauti Eggertson, and Raphael Schoenle. Is increased price flexibility stabilising? redux. working paper, 2012.

[30] Florin Bilbiie. Non separable preferences, fiscal policy puzzles and inferior goods. working paper, 2006.

[31] Mark Bils, Peter Klenow, and Benjamin Malin. Testing for keynesian labor demand. working paper, 2012.

[32] Fischer Black. Noise. *Journal of Finance*, pages 529–543, 1986.

[33] Nicholas Bloom. The impact of uncertainty shocks. working paper, 2009.

[34] Nicholas Bloom, Max Floetotto, Nir Jaimovich, Itay Saporta-Ekstein, and Stephen J. Terry. Really uncertain business cycles. working paper, 2012.

[35] Francisco Buera and Benjamin Moll. Aggregate implications of a credit crunch. working paper, 2012.

[36] Francisco Buera and Yongseok Shin. Financial frictions and the persistence of history: A quantitative exploration. working paper, 2012.

[37] Kenneth Burdett and Kenneth L. Judd. Equilbrium price dispersion. *Econometrica*, 51:955–969, 1983.

[38] Kenneth Burdett and Guido Menzio. The (q,s,s) pricing rule. working paper, 2014.

[39] Stephen Burgess, Emilio Fernandez-Corugedo, Charlotta Groth, Richard Harrison, Francesca Monti, Konstantinos Theodoridis, and Matt Waldron. The bank of engldand's forecasting platform:compass, maps, ease and the suite of models. Bank of England working paper, 2013.

[40] Dario Caldara, Richard Harrison, and Anna Lipinska. Practical tools for monetary policy analysis in dsge models with missing channels. working paper, 2012.

[41] Jeffrey Campbell and Zvi Hercowitz. The role of collateralized household debt in macreconomic stabilisation. working paper, 2006.

[42] Andrew Caplin and John Leahy. How important is discrete adjustment in aggregate fluctuations? working paper, 2002.

[43] Anabela Carneiro, Paulo Guimaraes, and Pedro Portugal. Real wages and the business cycle: Accounting for worker and firm heterogeneity. working paper, 2009.

[44] Christopher D. Carroll. Buffer-stock saving and the life cycle/permanent income hypothesis. *Quarterly Journal of Economics*, 112:1–55, 1997.

[45] Christopher D. Carroll. Requiem for the representative consumer? aggregate implications of microeconomic consumption behavior. working paper, 2000.

[46] Christopher D. Carroll. A theory of the consumption function, with and without liquidity constraints. working paper, 2001.

[47] Christopher D. Carroll and Patrick Toche. A tractable model of buffer stock saving. working paper, 2009.

[48] Edouard Challe and Xavier Ragot. Precautionary saving in the business cycle. working paper, 2010.

[49] Yongsung Chang, Sun-Bin Kim, and Frank Schorfheide. Labor market heterogeneity, aggregation and the policy (in)variance of dsge model parameters. working paper, 2011.

[50] V.V Chari, Patrick J. Kehoe, and Ellen R. McGrattan. Business cycle accounting. *Econometrica*, 75:781–836, 2007.

[51] V.V Chari, Patrick J. Kehoe, and Ellen R. McGrattan. Are structural vars with long-run restrictions useful in developing business cycle theory? *Journal of Monetary Economics*, 55:1337–1352, 2008.

[52] Raj Chetty. A new method of estimating risk aversion. *American Economic Review*, 96:1821–1834, 2006.

[53] Raj Chetty, Adam Guren, Day Manoli, and Andrea Weber. Does indivisible labour explain the difference between micro and macro elasticities? a meta-analysis of extensive margin elasticities. Working paper, 2011.

[54] Lawrence Christiano, Martin Eichenbaum, and Sergio Rebelo. When is the government spending multiplier large? working paper, 2010.

[55] Lawrence Christiano, Cosmin Ilut, Roberto Motto, and Massimo Rostagno. Monetary policy and stock market boom-bust cycles. working paper, 2007.

[56] Lawrence Christiano, Roberto Motto, and Massimo Rostagno. Risk shocks. working paper, 2012.

[57] John Cochrane. Money as stock. *Journal of Monetary Economics*, 52:501–528, 2005.

[58] John Cochrane. Determinacy and identification with taylor rules. *Journal of Political Economy*, 119:565–615, 2011.

[59] John Cochrane. Discount rates. *Journal of Finance*, pages 1047–1109, 2011.

[60] John Cochrane. The new keynesian liquidity trap. working paper, 2014.

[61] John Cochrane and Monika Piazzesi. The fed and interest rates - a high frequency identification. *American Economic Review*, 92:90–95, 2002.

[62] Thomas F. Cooley. *Frontiers of Business Cycle Research*. Princeton University Press, 1995.

[63] Vasco Curdia, Andrea Ferrero, Ging Cee Ng, and Andrea Tambalotti. Evaluating interest rate rules in an estimated dsge model. Federal Reserve Bank of New York staff report no. 510, 2011.

[64] Morris A. Davis and Jonathan Heathcote. Housing and the business cycle. *International Economic Review*, 46(3):751–784, 2005.

[65] Douglas W. Diamond. Financial intermediation and delegated monitoring. *Review of Economic Studies*, pages 393–414, 1984.

[66] Antonia Diaz and Maria Jose Luengo Prado. The wealth distribution with durable goods. *International Economic Review*, 51(1):143–170, 2010.

[67] Piti Disyatat. Monetary policy implementation:misconceptions and their consequences. BIS working paper no. 269, 2008.

[68] Thorsten Drautzenberg and Harald Uhlig. Fiscal stimulus and distortionary taxation. working paper, 2013.

[69] Benjamin Eden and Matthew Jaremski. Rigidity, discreteness and dispersion in chain prices. Working paper, 2009.

[70] Benjamin Eden and Matthew Jaremski. Discreteness and nominal rigidity: Do supermarket prices move too much? Working paper, 2010.

[71] Gauti Eggertson and Paul Krugman. Debt, deleveraging, and the liquidity trap: A fisher-minsky-koo approach. *Quarterly Journal of Economics*, 127:1469–1513, 2010.

[72] Martin Eichenbaum, Nir Jaimovich, Sergio Rebelo, and Josephine Smith. How frequent are small price changes? working paper, 2013.

[73] Douglas Elmendorf and Gregory Mankiw. Government debt. draft of chapter for handbook of Macroeconomics, 1999.

[74] Michael Elsby and Ryan Michaels. Fixed adjustment costs and aggregate fluctuations. working paper, 2014.

[75] Andres Erosa, Luisa Fuster, and Gueorgui Kambourov. Towards a micro-founded theory of aggregate labor supply. working paper, 2014.

[76] Gunter Coenen et al. Effects of fiscal stimulus in structural models. working paper, 2011.

[77] Stefano Eusepi and Bruce Preston. Labor supply heterogeneity and macroeconomic comovement. Federal Reserve Bank of New York staff report no. 399, 2009.

[78] Stefano Eusepi and Bruce Preston. Expectations, learning and business cycle fluctuations. working paper, 2011.

[79] Eugene F. Fama. Does the fed control interest rates? working paper, 2013.

[80] Roger Farmer. Confidence, crashes and animal spirits. *the Economic Journal*, pages 1–16, 2011.

[81] Roger Farmer and Vadim Khramov. Solving and estimating indeterminate dsge models. working paper, 2013.

[82] Jack Favilukis, Sidney C. Ludvigson, and Stijn Van Nieuwerburgh. The macroeconomic effects of housing wealth, housing finance, and limited risk-sharing in general equilibrium. working paper, 2011.

[83] Ernst Fehr and Antonio Rangel. Neuroeconomic foundations of economic choice- recent advances. working paper, 2011.

[84] Jesus Fernandez-Villaverde. The economstrics of dsge models. *Series: Journal of the Spanish Economic Association*, 1:3–49, 2010.

[85] Marjorie Flavin and Shinobu Nakagawa. A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence. *American Economic Review*, 98(1):474–494, 2008.

[86] Xavier Freixas and Jean Charles Rochet. *Microeconomics of Banking*. MIT, 1997.

[87] Xavier Gabaix. A sparsity based model of bounded rationality. working paper, 2014.

[88] Xavier Gabaix. Behavioral macroeconomics via sparse dynamic programming. working paper, 2015.

[89] Douglas Gale. Bargaining and competition part i: Characterisation. *Econometrica*, 54:785–806, 1986.

[90] William Gale, John Karl Scholz, and Ananth Seshadri. Are all americans saving optimally for retirement? working paper, 2009.

[91] Jordi Gali. *Monetary Policy, Inflation and the Business Cycle*. Princeton University Press, 2007.

[92] Jordi Gali, David Lopez-Salido, and Javier Valles. Understanding the effects of government spending on consumption. *Journal of the European Economic Association*, pages 227–270, 2007.

[93] Mark Gertler and Peter Karadi. A model of unconventional monetary policy. *Journal of Monetary Economics*, 58:17–34, 2011.

[94] Mark Gertler and Nobu Kiyotaki. Financial intermediation and credit policy in business cycle analysis. working paper, forthcoming in Handbook of Monetary Economics, 2010.

[95] Mikhail Golosov and Robert Lucas. Menu costs and phillips curves. *Journal of Political Economy*, 115(21):171–199, 2007.

[96] Paul Gomme and Paul Klein. Second order approximation of dynamic models without the use of tensors. *Journal of Economic Dynamics and Control*, pages 1–12, 2010.

[97] Paul Gomme and Peter Rupert. Theory, measurement and the calibration of macroeconomic models. *Journal of Monetary Economics*, 54:460–497, 2006.

[98] Marvin Goodfriend. Monetary policy in the new neoclassical synthesis: A primer. *Federal Reserve Bank of Richmond Economic Quarterly*, 90:21–45, 2004.

[99] Francois Gourio and Leena Rudanko. Customer capital. working paper, 2013.

[100] Ferre De Graeve, Marina Emiris, and Raf Wouters. A structural decomposition of the us yield curve. *Journal of Monetary Economics*, 56:545–559, 2009.

[101] Jeremy Greenwood, Zvi Hercowitz, and Per Krusell. The role of investment specific technological change in the business cycle. *European Economic Review*, 44:91–115, 2000.

[102] Stephanie Schmitt Grohe and Martin Uribe. Solving dynamic general equilibrium models using a second order approximation to the policy function. *Journal of Economic Dynamics and Control*, 28:755–775, 2004.

[103] Joseph Gruber and Robert Martin. Precautionary savings and the wealth distribution with illiquid durables. FED working paper, 2003.

[104] Refet S. Gurkaynak and Jonathan H. Wright. Macroeconomics and the term structure. working paper, 2010.

[105] Fatih Guvenen and Anthony Smith. Inferring labor income risk and partial insurance from economic choices. working paper, 2014.

[106] Wouter Den Haan and Georg Kaltenbrunner. Anticipated growth and business cycles in matching models. working paper, 2009.

[107] Robert Hall. High discounts and high unemployment. working paper, 2014.

[108] Daniel Hernaiz. A monetary model with price promotios. working paper, 2010.

[109] Michael Horvath. Sectoral shocks and aggregate fluctuations. *Journal of Monetary Economics*, 45:69–106, 2000.

[110] Kequiang Hou and Alok Johri. Intangible capital, corporate earnings and the business cycle. working paper, 2009.

[111] Matteo Iacoviello. House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review*, 95(3):739–764, June 2005.

[112] Matteo Iacoviello and Stefano Neri. Housing market spillovers:evidence from an estimated dsge model. working paper, 2009.

[113] Matteo Iacoviello and Marina Pavan. Housing and debt over the life cycle and over the business cycle. working paper, 2013.

[114] Nir Jaimovich and Max Floetotto. Firm dynamics, markup variations, and the business cycle. working paper, 2008.

[115] Nir Jaimovich and Sergio Rebelo. Can news about the future drive the business cycle? working paper, 2008.

[116] Charles I. Jones. Intermediate goods and weak links in the theory of economic development. *American Economic Journal: Macroeconomcis*, 3:1–28, 2011.

[117] Charles I. Jones. Misallocation, economic growth and input-output economics. working paper, 2011.

[118] Larry E. Jones and Rodolfo E. Manuelli. Neoclassical models of endogenous growth:the effects of fiscal policy, innovation and fluctuations. draft of Handbook of Economic Growth chapter, 2004.

[119] Kenneth Judd. *Numerical Methods in Economics*. MIT Press, 1998.

[120] Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti. Investment shocks and the relative price of investment. working paper, 2009.

[121] Greg Kaplan and Giovanni L. Violante. How much consumption beyond self insurance. *American Economics Journal: Macroeconomics*, 2:53–87, 2010.

[122] Greg Kaplan and Giovanni L. Violante. A model of the consumption response to fiscal stimulus payments. working paper, 2014.

[123] Patrick Kehoe and Virgiliu Midrigan. Prices are sticky after all. Working paper, 2012.

[124] Aubik Khan and Julia K. Thomas. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, 76:395–436, 2008.

[125] Aubik Khan and Julia K. Thomas. Credit shocks and aggregate fluctuations in an economy with production heterogeneity. working paper, forthcoming in the Journal of Political Economy, 2013.

[126] Michael Kiley. Policy paradoxes in the new keynesian model. Federal Reserve working paper, 2014.

[127] Lutz Killian. Structural vector autoregressions. draft of chapter for the Handbook of Research Methods and Applications in Empirical Macroeconomics, 2013.

[128] Robert King and Sergio Rebelo. Resusciting real business cycles. working paper, draft of Handbook of Macroeconomics chapter, 2000.

[129] Nobuhiro Kiyotaki and John Moore. Liquidity, business cycles and monetary policy. working paper, 2008.

[130] Paul Klein. Using the generalised schur form to solve a multivariate linear rational expectations models. *Journal of Economic Dynamics and Control*, 24:1405–1423, 2000.

[131] Peter Klenow and Benjamin Malin. Microeconomic evidence on price setting. draft of chapter for Handbook of Monetary Economics, 2010.

[132] Keiichiro Kobayashi, Tomoyuki Nakajima, and Masaru Inaba. Collateral constraints and news driven business cycles. working paper, 2007.

[133] Per Krusell and Anthony A. Smith. Income and wealth heterogeneity in the macroeconomy. *The Journal of Political Economy*, 83(1):78–97, October 1998.

[134] Per Krusell and Anthony A. Smith. Quantitative macroeconomic models with heterogeneous agents. working paper, 2006.

[135] Marianna Kudlyak. The cyclical price of labor when wages are smoothed. working paper, 2010.

[136] Hong Lan and Alexander Meyer-Gohde. Existence and uniqueness of perturbation solutions to dsge models. working paper, 2012.

[137] Eric M. Leeper, Nora Traum, and Todd B. Walker. Clearing up the fiscal multiplier morass. NBER working paper no. 17444, 2011.

[138] Eric M. Leeper and Todd B. Walker. Information flows and news driven business cycles. working paper, 2010.

[139] Zheng Liu, Pengfei Wang, and Tao Zha. Do credit constraints amplify macroeconomic fluctuations? working paper, 2010.

[140] Lars Ljungqvist and Thomas Sargent. *Recursive Macroeconomic Theory*. MIT Press, 2004.

[141] Bartosz Mackowiak and Mirko Wiederholt. Business cycle dynamics under rational inattention. ECB working paper, 2011.

[142] Lilia Maliar and Serguei Maliar. The representative consumer in the neoclassical growth model with idiosyncratic shocks. *Review of Economic Dynamics*, 6:362–380, 2003.

[143] Gregory Mankiw. The savers-spenders theory of fiscal policy. working paper, 2000.

[144] Gregory Mankiw and Ricardo Reis. Imperfect information and aggregate supply. draft of chapter for Handbook of Monetary Economics, 2010.

[145] Peter Klenow Mark Bils and Benjamin Malin. Reset price inflation and the impact of monetary policy shocks. *American Economic Review*, 102(6):2798–2835, 2012.

[146] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

[147] Ellen McGrattan and Edward Prescott. The labor productivity puzzle. working paper, 2012.

[148] Enrique Mendoza. Sudden stops,financial crises and leverage: A fisherian deflation of tobin's q. FRB International Finance Discussion Paper no. 960, 2008.

[149] Karel Mertens and Morten Ravn. Fiscal policy in an expectations driven liquidity trap. working paper, 2014.

[150] Espen R. Moen. Competitive search equilibrium. *Journal of Political Economy*, 105:385–411, 1997.

[151] Kevin Mulligan. Aggregate implications of indivisible labor. NBER working paper no. 8159, 2001.

[152] Marco Del Negro, Gauti Eggertson, Andrea Ferrero, and Nobuhiro Kiyotaki. The great escape? a quantitative evaluation of the fed's liquidity facilities. New York Federal Reserve staff reprot no. 520, 2011.

[153] Marco Del Negro, Frank Schorfheide, Frank Smets, and Raf Wouters. On the fit and forecasting performance of new keynesian models. working paper, 2004.

[154] Christopher Nekarda and Valerie Ramey. The cyclical behaviour of the price-cost markup. NBER working paper, 2013.

[155] Hyungseung Oh and Ricardo Reis. Targeted transfers and the fiscal response to the great recession. working paper, 2012.

[156] Stephen Parente and Edward C. Prescott. Barriers to riches. 3rd Walras-Pareto Lecture, 1999.

[157] Christopher Pissarides. The unemployment volatility puzzle: Is wage stickiness the answer? working paper, 2009.

[158] Ricardo Reis. The mystique surrounding the central bank's balance sheet, applied to the european crisis. working paper, 2013.

[159] Diego Restuccia. The latin american development problem: An interpretation. working paper, 2012.

[160] Diego Restuccia and Richard Rogerson. Policy distortions and aggregate productivity with heterogenous plants. working paper, 2007.

[161] David Romer. *Advanced Macroeconomics*. McGraw-Hill, 2001.

[162] John Rust. Dynamic programming. draft of entry for the New Palgrave Dictionnary of Economics, 2006.

[163] Frank Schorfheide, Keith Sill, and Maxym Kryshko. Dsge model based forecasting of non-modelled variables. working paper, 2009.

[164] Hersh Shefrin. Risk and return in behavioral sdf-based asset pricing models. working paper, 2007.

[165] Robert Shimer. Labor markets and business cycles. Draft of Book, 2009.

[166] Christopher Sims. Macroeconomics and reality. *Econometrica*, 48:1–48, 1980.

[167] Frank Smets and Raf Wouters. An estimated stochastic dynamic general equilibrium model of the euro area. ECB working paper no. 171, 2002.

[168] Frank Smets and Raf Wouters. Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 97(3):586–606, 2007.

[169] Margaret Stevens. New microfoundations for the aggregate matching function. working paper, 2004.

[170] James Stock and Mark Watson. Dynamic factor models. draft chapterf for the Handbook of Economic Forecasting, 2010.

[171] Nancy L. Stokey, Robert E. Lucas Jr., and Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, 1989.

[172] Roman Sustek. Monetary business cycle accounting. *Review of Economic Dynamics*, 14:592–612, 2011.

[173] Daniel Thornton. The unusual behavior of the federal funds rate and treasury yields: a conundrum or an instance of goodhart's law? working paper, 2010.

[174] Hajime Tomura. Heterogeneous beliefs and housing-market boom-bust cycles in a small open economy. Bank of Canada working paper, 2009.

[175] Harald Uhlig. Economics and reality, 2011.

[176] Karl Walentin. Expectation driven business cycles with limited enforcement. Sveriges Riksbank working paper, 2009.

[177] Carl E. Walsh. *Monetary Theory and Policy*. MIT Press, 2003.

[178] Pengfei Wang. Understanding expectations driven fluctuations: A labor-market approach. working paper, 2010.

[179] Ryan Webb. Dynamic constraints on the distribution of stochastic choice: Drift diffusion implies random utility. working paper, 2013.

[180] Yi Wen. Demand shocks and economic fluctuations. *Economic Letters*, 90(3):373–383, March 2006.

[181] Ivan Werning. Managing a liquidity trap: Monetary and fiscal policy. working paper, 2012.

[182] Stephen D. Williamson and Randall Wright. New monetarist economics: Models. working paper, forthcoming chapter in Handbook of Monetary Economics, 2010.

[183] Michael Woodford. Monetary policy in a world without money. NBER working paper no. 7853, 2000.

[184] Michael Woodford. Simple analytics of the government expenditure multiplier. working paper, 2010.