The London School of Economics and Political Science

Three Essays in Macroeconomics:

Capital Reallocation, Capital Utilization and Optimal Policy with Partial Information

Andrea Lanteri

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, May 2015.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it - see Statement of conjoint work below).

I declare that my thesis consists of approximately 40,000 words.

Statement of conjoint work

I declare that Chapter 3 was jointly co-authored with Professor Esther Hauk (IAE-CSIC, Barcelona) and Professor Albert Marcet (IAE-CSIC, Barcelona). I have contributed to the derivation of the analytical results and to writing the chapter and I have performed all the numerical computations and simulations.

Three Essays in Macroeconomics: Capital Reallocation, Capital Utilization and Optimal Policy with Partial Information

Andrea Lanteri

May 2015

Abstract

This thesis is composed of three chapters. In the first chapter, I show that capital reallocation is highly procyclical, in contrast to the prediction of existing business-cycle models with firm heterogeneity, where it is highly countercyclical. I argue that endogenizing the price of used capital relative to new solves this puzzle. First I show empirically that in several sectors the price of used investment goods relative to new is procyclical. Then I build a dynamic general equilibrium model with heterogeneous firms facing both aggregate and idiosyncratic productivity shocks. Used capital is an imperfect substitute for new capital because of firm-level capital specificity. In equilibrium both the price of used capital and the volume of reallocation become procyclical.

The second chapter studies the link between investment irreversibility and capital utilization. I show that when it is costly to downsize, firms respond to negative transitory profitability shocks by underutilizing their capital stock. In a partial equilibrium setting I derive both analytical and numerical results on the links between the level of irreversibility, the size and persistence of the shocks and the optimal utilization decision. In an industry-equilibrium model with heterogeneous firms and aggregate shocks, I endogenize the resale price of capital as in the first chapter and show that when this price falls, the option value of idle capital rises and the aggregate utilization rate decreases.

The third chapter, co-authored with Esther Hauk and Albert Marcet, studies optimal policy in a class of models of endogenous partial information. The economy is hit by multiple shocks and the policy-maker cannot observe their realizations, but only aggregate outcomes. In general the solution to this signal extraction problem cannot be separated from the solution to the problem of finding the optimal policy and we show how to solve them jointly. We apply the result to a model of optimal

fiscal policy with incomplete markets and show that the endogeneity of the signal extraction may lead to highly non-linear optimal policies.

Acknowledgements

I wish to thank my supervisor Wouter den Haan for his invaluable advice and encouragement. I am hugely indebted to him for all the fruitful conversations we have had at all stages of the projects presented in this thesis. It was during his reading group, in particular when we read Hopenhayn and Rogerson (1993), that I became excited about production heterogeneity in macroeconomic models.

I also thank Albert Marcet for his great support and for stimulating my interest in optimal policy and inviting me to collaborate with him and Esther Hauk on the project that has lead to chapter 3 of this thesis.

This work has benefitted very significantly from numerous conversations with several people at the LSE, many of whom I will thank individually in person. I am indebted to Alessandro Gavazza for his encouragement and very helpful suggestions. Furthermore, very special thanks go to Alex Clymo and Sergio de Ferra. Discussing Economics (and much more) with them has made these years a fantastic intellectual experience as well as great fun.

Finally, I wish to thank my parents and my wife for their constant support. This thesis is dedicated to them.

Contents

List of Tables

Chapter 1

The Market for Used Capital: Endogenous Irreversibility and Reallocation over the Business Cycle

1.1 Introduction

1.1.1 Motivation

Firms buy and sell large amounts of used investment goods both directly on secondary markets for equipment and plants and indirectly through acquisitions. Over the business cycle, this volume of capital reallocation is volatile and positively correlated with aggregate output.[1] Why is this the case? Can the cyclicality of reallocation be

[1]Eisfeldt and Rampini (2006) show that reallocation of physical capital among US firms, which amounts to approximately 30% of total investment, is strongly procyclical. The cyclical component of their reallocation series, composed of Sales of Plants, Property and Equipment plus Acquisitions from Compustat, is very volatile (about 7 times the volatility of output) and positively correlated with US GDP (with a correlation coefficient of .56). Other measures of capital reallocation point to

efficient? How do equilibrium dynamics in the market for used capital affect macroeconomic variables such as aggregate TFP and investment? This chapter addresses these questions by first showing new evidence on prices of used real assets and then building a dynamic general equilibrium model with heterogeneous firms facing aggregate and idiosyncratic uncertainty.

Importantly, the market for used capital reallocates assets from less productive to more productive firms, as Maksimovic and Phillips (2011) document. Hence, more reallocation in booms means that more capital flows to highly productive firms when the economy is expanding, while downturns are associated with a smaller flow of assets towards their most productive use. This suggests that understanding the cyclicality of capital reallocation may be a step towards a theory of the cyclical movements in aggregate TFP. Particularly in the aftermath of the Great Recession, it seems important to understand the drivers of this reallocative process, which policy-makers in the UK see as an important condition for the onset of a strong recovery in productivity.[2]

Despite its relevance, the procyclicality of capital reallocation has so far been a puzzle for the macroeconomic literature, for at least two reasons. First, existing DSGE models of investment with heterogeneous firms (e.g. Khan and Thomas, 2008, 2013) imply a negative correlation between output and reallocation. In these models, unproductive firms want to disinvest by a larger amount in recessions, because their profitability falls following a negative aggregate shock. As these are one-sector models of the economy, demand for their used capital comes from both consumers and other firms. This demand is perfectly elastic, as the standard assumption is either full reversibility of investment, or a constant level of partial irreversibility, i.e.

the same stylized fact: sales of used corporate assets for the UK from the ONS Capital Expenditure Survey are also procyclical and more disaggregated evidence on the market for used commercial ships shows the same pattern of cyclicality. Section 1.2 and Appendix A present the empirical evidence.

[2]For instance, Ben Broadbent, Deputy Governor of the Bank of England, attributes low labor productivity in the slow recovery post-2008 in the UK to a lack of capital reallocation (Broadbent, 2012, Barnett et al., 2014).

the relative price of used capital is assumed to be constant and less than 1. Hence, an outward shift in supply of used capital from disinvesting firms necessarily leads to more reallocation. This gives more reallocation in recessions and less in booms. Second, as Eisfeldt and Rampini (2006) point out, several measures of dispersion of returns on capital are higher in recessions than in booms, suggesting that benefits from reallocation are countercyclical. Hence we should expect to see more reallocation during downturns, when higher dispersion makes reallocation of capital towards its most productive use more beneficial.

In order to explain the puzzle, this paper starts by presenting a new stylized fact: the relative price of used capital goods, far from being constant, is actually volatile and strongly procyclical, suggesting that partial investment irreversibility is to a great extent a market equilibrium outcome. Recessions are bad times to disinvest, as more firms would like to sell their assets to downsize but the demand side coming from investing firms is weak.

Starting from this observation, I build an equilibrium model of partially irreversible investment, where the resale price of capital is endogenous. I assume a degree of capital specificity at the firm level: after installation, capital becomes a different good with respect to the output (and consumption) good, partially specific to the firm who owns it. Not only is it useless for consumers, but also an imperfect substitute with respect to new investment for other firms. This assumption allows me to rationalize the procyclicality of the resale price and capital reallocation in an otherwise standard business-cycle model. In a recession, used capital is relatively cheap, because more firms would like to disinvest and downsize, while expanding firms cannot fully benefit from the abundance of used capital on the market, because this capital is to an important extent specific to the firms that operated it previously.

The model emphasizes both a static and a dynamic real-options mechanism that induce procyclical reallocation. Let us examine the static mechanism first. A lower resale price associated with a recession increases the target level of capital of a disinvesting firm, hence reducing its desired level of disinvestment. Intuitively, after a negative aggregate productivity shocks there are two opposing forces on the disinvestment decision: both the internal value of capital for the firm and its market value fall. In equilibrium, when new and used capital as sufficiently poor substitutes, the latter effect dominates and sales of used capital fall. Next, let us introduce the dynamic real-options effects. Consider again a firm that is hit by a negative idiosyncratic shock in a recession and evaluates the opportunity to disinvest. In a dynamic environment, this firm needs to compare the price at which it can sell its assets in the current period with the price it would get by waiting one more period. In a recession, the current resale price falls, and so does the future expected resale price. However, if there is a positive probability of exiting the recession in the near future, the future expected resale price falls by less than the current one, and it may be better to wait, hold on to the assets and disinvest later by selling them at a higher price. This dynamic effect holds in general when the underlying stochastic process is mean-reverting and it generates an option value from waiting to disinvest that further decreases and delays the reallocation of capital in bad times.

The procyclicality of reallocation is matched by a countercyclical dispersion of returns from capital, consistently with a growing body of empirical evidence (e.g. Bloom et al., 2012). In the model, this happens because large unproductive firms downsize by less in recessions and hence their marginal product remains low relative to that of more productive firms. Several papers have interpreted the increase in the dispersion of returns associated with downturns as a symptom of the worsening of financial frictions, leading to the policy implication that credit expansions and non-conventional monetary policy can facilitate reallocation and stimulate the recovery. In contrast, the present paper shows that lack of reallocation and high dispersion of returns in recessions can be efficient outcomes in an economy where capital is partially specific at the firm level and hence used assets are imperfect substitutes for new ones.

While theories based on time-varying financial frictions may explain capital misallocation in recessions, they do not have implications for the cyclicality of the relative price of used capital. In contrast, my theory of capital specificity is able to explain both dynamics in prices and quantities traded on secondary markets for capital. The important and challenging question of how much dispersion in marginal products is due to financial or to real frictions is beyond the scope of this paper. However, because procyclical reallocation can be explained as an equilibrium outcome of a model without financial frictions, a policy implication of the paper is that credit expansions may in fact be less relevant for reallocation than previously thought. It should be noted that reallocation has fallen in every recession since we have data for it (1970's), that is also in recessions for which the financial component was arguably less important than in the Great Recession. A contribution of this paper is to present a real model where only one aggregate shock can generate both standard business-cycle facts and procyclical reallocation.

Furthermore, the model highlights important equilibrium real-options effects on investment. Consider again a recession. Used capital becomes cheaper, so that overall investment can be made at a lower cost. However, investment is also expected to be harder to reverse in the future (if the recession is expected to persist). These contrasting effects can either amplify or dampen the response of investing firms to aggregate shocks depending on the properties of the idiosyncratic and aggregate shock processes. In the quantitative section of the paper, I show that one of the aggregate implications of endogenous irreversibility is a significant smoothing of the aggregate investment series, bringing its volatility and autocorrelation closer to the empirical counterparts. Hence, the mechanism presented in the paper can be seen as a plausible microfoundation for an aggregate capital adjustment cost.

1.1.2 Related literature

Using Compustat data, Eisfeldt and Rampini (2006) show that Sales of Plants, Property and Equipment, as well as Acquisitions, are highly procyclical and argue that this is a puzzle given that the benefits from reallocation, as measured for instance by dispersion in TFP or dispersion in utilization rates, appear to be countercyclical. Their conclusion is that there must be a countercyclical degree of reallocation frictions. In this sense, one can see the present paper as microfounding this conclusion by explicitly modelling a market for used capital and showing that the equilibrium resale price falls in bad times.

The empirical evidence I present on the price of used capital fills an important gap in the empirical literature on capital adjustment costs. Inference on investment irreversibility is typically indirect, based on firms' investment and disinvestment rates rather than directly on prices (e.g. Cooper and Haltiwanger, 2006). Furthermore, irreversibility is generally assumed to be a constant technological friction. By looking at sectors that allow a direct comparison of the price of new and used assets, I establish that partial irreversibility is to a large extent a market equilibrium outcome and that it varies significantly with the business cycle.

The most closely related papers are Khan and Thomas (2013) and Cui (2014). Both papers build DSGE models with heterogeneous firms that feature constant partial irreversibility (defined by a constant resale price of capital below one) and collateral constraints. When feeding the model with aggregate TFP shocks, they cannot generate a procyclical response of reallocation, because disinvesting firms face a constant resale price and disinvest by more in recessions and less in booms. However, Cui (2014) shows that the procyclicality of reallocation can be obtained by introducing credit shocks, i.e. an exogenous tightening of the borrowing constraint. After such shocks, unproductive firms hold on to their capital and use its return to pay back their debt and deleverage. Regarding the question on the source of business cycles, Cui (2014) interprets the procyclicality of reallocation as evidence in favor of credit shocks.

The results in the present paper suggest that this conclusion may depend on the assumption of a constant resale price of capital. In fact, I show that the procyclicality of this price can reconcile the Eisfeldt and Rampini (2006) findings without resorting to exogenous credit shocks. However, I would stress that this does not necessarily mean that credit shocks are not important in driving business-cycle fluctuations. It only implies that procyclical reallocation is less of a puzzle, as it can be rationalized in a more standard business-cycle model, where only one aggregate shock drives both standard business-cycle facts and reallocation.[3] Furthermore, the present model provides a useful framework that can be extended to include financial frictions in the form of collateral constraints. Following aggregate shocks, the availability of credit would change endogenously with movements in the price of used capital.

Caunedo (2014) also considers an economy with heterogeneous firms and investment irreversibility and shows that the dispersion in marginal products that arises in equilibrium does not need to be inefficient. In this paper, I show that also the cyclical movements of such dispersion of returns (high dispersion in recessions) are not necessarily a symptom of time-varying financial frictions. Cooper and Schott (2013) consider a similar framework with heterogeneous firms and introduce an exogenous time-varying probability of being able to reallocate capital. They show that exogenous shocks to this probability may induce procyclical reallocation. Gilchrist et al. (2014) treat the resale price of capital as an exogenous shock process and show that a fall in this price, combined with collateral constraints, can replicate a recession associated with a liquidity crisis. My contribution with respect to these papers is to explicitly model the market for used capital and to endogenize the resale price and show that both this price and reallocation respond positively to aggregate TFP shocks in equilibrium.

A related strand of literature is that on real-options theory, starting with the seminal work of Dixit and Pindyck (1994) and Abel et al. (1996). This literature typically assumes exogenous stochastic paths for the prices at which a firm can buy

[3] Section 1.6 shows that the main mechanism is robust to both aggregate and investment-specific productivity shocks.

and sell capital. As the resale price is assumed to be strictly less than the buying price, part of the investment is sunk and uncertainty regarding future productivity (or equivalently the future output price) leads to the presence of option values connected with the opportunity to wait and invest in the future. In this paper, I compute the equilibrium effects of aggregate shocks on these option values and show that with an endogenous resale price the value of the option to resell (put option) can be procyclical, contrary to what arises in partial equilibrium.

The key assumption in this paper is imperfect substitutability between new and used capital. In their seminal work on capital reallocation, Ramey and Shapiro (2001) provide an extreme example of this friction. They report that during the liquidation of an aerospace plant, a wind tunnel that could generate a 270 miles/hour wind was sold to a company that rented it to bicycle helmet designers, who only needed low speeds and did not value it as much as the aerospace firm who sold it. Edgerton (2011) uses evidence from tax depreciation reforms in the US to estimate the elasticity of substitution between new and used capital in the production function and finds values in the range between 1 and 10 for sectors such as farming, construction and aircraft. Jovanovic and Yatsenko (2012) build a vintage capital model to study technology adoption decisions and assume that different vintages of capital enter the production function in a Constant Elasticity of Substitution (CES) form. Eisfeldt and Rampini (2007) show that that in the data firms of all sizes invest both in new and in used capital and build a model assuming that new and used investment goods differ because used capital is cheaper, but requires more maintenance in the future, inducing financially constrained firms to buy a higher ratio of used to new items. In this paper, I abstract from financial frictions and focus on the role of capital specificity.

An alternative attempt to endogenize the resale price of capital relies on asymmetric information, especially lemons problems in secondary markets (Eisfeldt, 2004, Kurlat, 2013, Li and Whited, 2014). In my empirical evidence I focus on the aircraft sector, for which asymmetric information is unlikely to be relevant, as the maintenance history of each aircraft is public information. Furthermore, in models of asymmetric information the fraction of lemons does not necessarily increase in recessions, as would be required to explain a procyclical resale price and procyclical reallocation. As Eisfeldt (2004) argues, one can imagine a case where the fraction of lemons decreases in recessions, because more sellers owning good quality assets are forced to downsize, leading to a higher resale price and more reallocation. Perri and Quadrini (2014) follow a different approach to endogenize the resale price of capital: they assume that the value of used capital depends on whether it is sold to other firms or to consumers. In the latter case, the price is lower. In this paper, I assume that used capital is useless for consumer and focus instead on its imperfect substitutability with new investment.

Furthermore, this paper contributes to the literature on the link between micro-level irreversibility and smooth aggregate investment. Using a partial equilibrium model, Bertola and Caballero (1994) argued that irreversibility at the micro level is a plausible explanation for what in the aggregate looks like a convex adjustment cost. However, this result did not seem to pass the test of general equilibrium. Veracierto (2002) considers a model with constant partial irreversibility and concludes that consumption smoothing forces undo all the effects of irreversibility and the property of the aggregate investment series are almost identical, independently of the level of the resale price of capital. In this paper I show that endogenizing irreversibility reaffirms the result of Bertola and Caballero (1994). What matters is not the average level of the resale price, but its correlation with aggregate shocks: investment becomes more irreversible exactly at the time when disinvesting firms would like to disinvest by more and this induces more caution in investment decisions. Importantly, because the price of used capital is procyclical, the total cost of investment (new and used) is also procyclical and this further smooths investment decisions by making capital cheaper in recessions and more expensive in booms.

Finally, it is worth emphasizing that the literature on DSGE models with heterogeneous firms obtains results that imply a small or insignificant role for heterogeneity and changes in the cross-sectional distribution of firms, especially following aggregate TFP shocks. Most notably, Veracierto (2002) and Khan and Thomas (2008) show that the aggregate behavior of these model is remarkably close to that of representative-firm Real Business Cycle (RBC) models. The present paper is an example of a model where heterogeneity is important in order to microfound and understand an aggregate observation: without firms changing their idiosyncratic productivity levels over time, the market for used capital would not open and we could not rationalize the data on reallocation and the smoothing of the aggregate investment series.

1.2 Empirical evidence

1.2.1 Capital reallocation

Figure 1.1 shows the cyclical components of the Eisfeldt and Rampini (2006) capital reallocation series for the period 1971-2011, composed of annual Compustat data on Sales of Plants, Property and Equipment (SPPE) plus Acquisitions (all deflated using the US GDP deflator) and filtered using a Hodrick-Prescott (HP) filter with smoothing parameter equal to 6.25 (as suggested by Ravn and Uhlig, 2002, for annual data). Capital reallocation is very volatile (approximately 7 times as volatile as US GDP) and positively correlated with output, with a correlation coefficient of .56.[4]

I confirm the evidence on the procyclicality of capital reallocation by looking at two other data sources (both of which exclude acquisitions): UK sales of second-hand

[4]Acquisitions represent the larger component of the capital reallocation series (approximately two thirds of the total). However, each of the two components (SPPE and Acquisitions) is significantly procyclical. In this paper, I will not distinguish between bundled and unbundled sales of used capital and I will refer to the sum of the two as capital reallocation, following Eisfeldt and Rampini (2006). Using a different data source, i.e. the Longitudinal Research Database compiled by the US Census Bureau, Maksimovic and Phillips (2001) find that the fraction of manufacturing plants involved in M&A activity goes from 3.89% in an average year to 6.19% in expansion years.

investment goods and global sales of second-hand commercial ships. In the UK, data on second-hand investment from the Survey of Capital Expenditures of the ONS show positive correlation between sales volumes and GDP. In particular during the recent recession, sales of second-hand investment goods were historically low, as shown in Figure 1.13 in Appendix A.

The market for second-hand commercial ships also provides a useful source of data on second-hand sales. Differently from the above-mentioned data sources, data on trading of used ships are divided into prices and quantities traded (number of sales) and do not depend on aggregation across types of investment goods.[5] By looking at these industry-level data, the following picture emerges: high trading volumes are associated with the period of economic expansion leading to 2007, and an abrupt fall in the number of sales coincides with the start of the Great Recession. This is illustrated in Figure 1.14 in Appendix A.

1.2.2 The price of used capital

A new stylized fact emerges from the analysis of sectoral evidence on the resale price of capital: the price of used investment goods is more volatile and more procyclical than the price of new investment goods. I construct or gather price indices from sectors that allow direct comparison of the value of new and used items in the same asset class. These sectors are

- commercial aircraft
- commercial ships
- vehicles and trucks
- construction equipment.

[5] Price indices and sales numbers are compiled by Clarkson and VesselsValue.

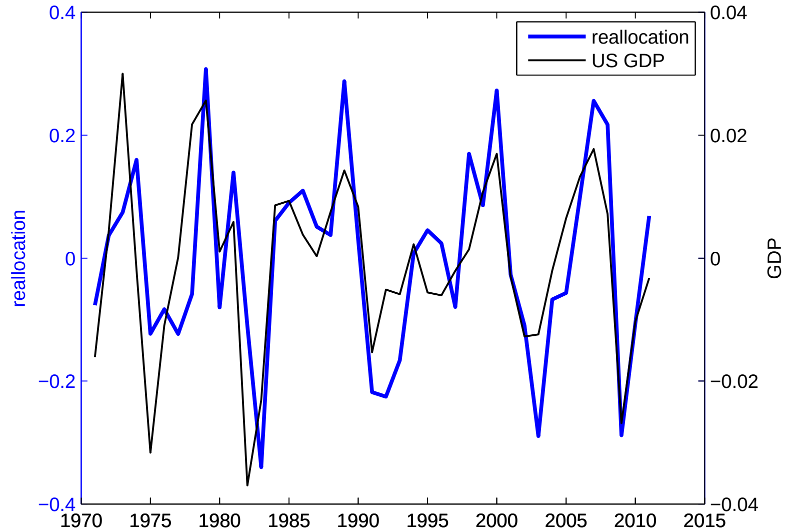


Figure 1.1: Capital reallocation and US GDP (cyclical components)

Log-deviations from HP trend (smoothing parameter = 6.25) of (i) the Eisfeldt and Rampini (2006) capital reallocation series, composed of Sales of Property, Plants and Equipment and Acquisitions from Compustat, deflated using the US GDP deflator, (ii) US real GDP. Yearly frequency.

While this is only suggestive evidence related to these specific sectors, the pattern of cyclicality in these four sectors is remarkably similar, showing a much stronger reaction of the resale price of capital to business-cycle shocks relative to the price of new investment goods. I will now describe the evidence related to the aircraft sector. Appendix A reports the evidence for ships, vehicles and construction equipment.

Starting with a dataset on the value of all Western-built commercial aircraft from 1967 to 2009, I construct a price index of used and new aircraft. This dataset is compiled by a specialized consulting company that evaluates aircraft based on actual transactions prices for which the seller was not bankrupt. It includes prices of all the different vintages of 38 types of aircraft, starting from their first production year onwards. The observation unit is an aircraft of type j, vintage v in year t, with price p_{jvt}. To construct the index, I divide the data into prices of new aircraft ($v = t$) and prices of used aircraft ($v < t$). I deflate all prices using the US GDP deflator. Then I create dummy variables for year, age and type (and interaction terms) and run a regression of $\log(p_{jvt})$ on these dummies. In each subsample (new and used), the coefficients on the time dummies are the quality-age-adjusted price index of aircraft. Finally, I detrend the series using an HP filter, with a smoothing coefficient of 6.25.[6] Figure 1.2 plots the price index of new aircraft, that of used aircraft and US GDP as a measure of the business cycle. It is evident that the cyclical component of the price of used aircraft is more volatile than that of new aircraft. It is also more strongly correlated with GDP. Table 1.1 reports standard deviations and coefficients of correlation of these series.

I interpret variations in the relative price of used assets as evidence in favor of capital specificity and against the standard assumption of perfect substitutability between new and used capital. Consistently with this interpretation, Gavazza (2011b) suggests a reason why capital specificity may be playing an important role in determining the volatility of the price of used aircraft. Carriers typically operate a very

[6] Robustness exercises with different smoothing parameters lead to very similar results.

small number of models in order to exploit economies of scale in maintenance and staff training costs and they are unwilling to substitute into other models when there is an increase in the supply of used aircraft due to aggregate shocks, leading to a fall in the value of used aircraft. By looking at cross-sectional evidence on the prices of different models, he finds support for this theory: the volatility of resale prices of more specific models of aircraft (e.g. Boeing 747, which can operate on a limited range of routes) is significantly higher than the volatility of more flexible models that can be used on a larger range of routes (e.g. Boeing 737).

Evidence on secondary markets for commercial ships, vehicles and trucks and construction equipment is consistent with the main finding: the price of used capital relative to new is volatile and procyclical. Appendix A presents price series for these sectors.

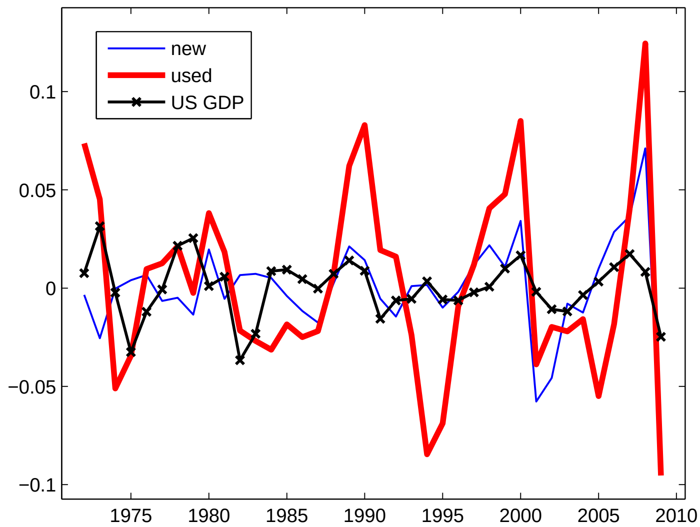


Figure 1.2: Aircraft prices and US GDP (cyclical components)

Log-deviations from HP trend (smoothing parameter = 6.25) of (i) price index of new aircraft, (ii) price index of used aircraft, (iii) US real GDP. Aircraft prices are deflated using the GDP deflator. Yearly frequency.

Series	Standard Deviation	Corr. with *new*	Corr. with *used*	Corr. with GDP
new	0.0342	1	-	-
used	0.0799	0.4781	1	-
GDP	0.0239	0.4090	0.5647	1

Table 1.1: New and used aircraft prices: second order moments

Standard deviations and correlation coefficients of the cyclical components of the price index of new aircraft, the price index of used aircraft and US real GDP. Yearly frequency, HP smoothing parameter = 6.25.

1.2.3 Discussion

Looking again at the Eisfeldt and Rampini (2006) reallocation series (Figure 1.1), in light of this evidence on the cyclicality of resale prices one may ask whether deflating their series with an index of used capital prices (instead of the GDP deflator) would explain the procyclicality of the volume of reallocation. In other words: is the cyclicality in the volume of reallocation only due to the cyclicality of the price of used capital? A back-of-the-envelope calculation suggests a negative answer. Under the assumption that the sectors discussed above (and in Appendix A) are representative of the whole economy, it is possible to compare the cyclical movements in these price series (cyclical deviations from trend in a ballpark of 10%) with that of the Eisfeldt and Rampini series (approximately 20% above and below trend in booms and recessions respectively). This suggests that part of the volatility in the Eisfeldt and Rampini series is certainly due to prices, but approximately half of this volatility may be due to movements in the quantity traded, consistently with the observation on the quantities traded on the market for used commercial ships.

Overall, the empirical evidence suggests that recessions are bad times to disinvest, because the resale price of capital is low. By treating the resale price as a constant parameter, the previous theoretical literature on investment irreversibility has not drawn any distinction between the case of a firm that needs to downsize during an expansion or a firm that needs to downsize in a downturn. However, the price of these two types of transactions can be quite different.

The assumption of perfect substitutability between new and used assets, which is implicit in the literature, is inconsistent with the evidence presented on the relative price of used capital: even an infinitesimal decrease in this price would lead investing firms to jump to a corner solution and demand only used capital, which is counterfactual. For instance, US Census ACES data on capital expenditures show a stable ratio between used and new investment expenditures, with a standard deviation of approximately 1%.[7]

1.3 A simple model of capital reallocation

Building on the empirical evidence presented above, this section introduces a simple static model that features imperfect substitutability between new and used investment goods and allows the derivation of analytical results on the response of capital reallocation to exogenous changes in aggregate productivity.

Section 1.4 extends this simple setup to include dynamic real-options effects and section 1.5 embeds the mechanism in a dynamic general equilibrium model with aggregate and idiosyncratic productivity shocks.

1.3.1 Technological assumptions

There is a continuum of firms $j \in [0,1]$, all of which are endowed with an initial capital level k_0. They produce a homogeneous output good with production function

$$y_j = z s_j k_j^{\alpha}, \tag{1.1}$$

[7] The volatility of the price of used capital could also be partly explained by time-to-build frictions: if used capital is readily available, while new capital takes time to become productive, shocks will affect the relative price of used capital. I will abstract from this in the model, but the introduction of time-to-build in my framework would be very interesting and is left for future work.

where z is an aggregate productivity parameter, s_j is an idiosyncratic shock with cdf $F(s_j)$ and $\alpha \in (0,1)$ is the returns-to-scale parameter.

Each firm uses its specific type of capital in order to produce the output good. Before production, firms can adjust their capital level k_j according to their productivity. Firms that decide to decrease their capital stock can sell some of their capital on the market for used capital. On the other hand, firms that decide to increase their capital level can invest using newly produced capital (supplied inelastically by a representative consumer) or used capital sold by disinvesting firms. New capital can be freely specialized. In contrast, used capital is partially specific to its previous owner. As a consequence, expanding firms cannot make the whole investment buying used capital and they need to bundle it instead with some newly produced output good in order to make it specific to their firms. Hence the substitutability between new and used investment goods is imperfect. This can be rationalized in a world where investment goods needed to build a plant are of different types. Some of them are fairly generic and can be easily purchased as used and put in production in a different plant. Some others have to be specifically designed for the production of a particular business line. In this environment, substitutability is imperfect and firms will only be willing to substitute towards more used capital if this becomes cheaper.[8] However, investing firms always have the choice to buy only new goods.

Formally, the investment technology is given by a perfect substitutes aggregator

[8] The sale of a GM pick-up trucks production plant in Shreveport, Louisiana, to three-wheeled electric car manifacturer Elio Motors provides a recent example of capital reallocation with imperfect susbstitutability. Elio chose to acquire the plant because it was ready for reuse as well as more convenient than building a new plant from scratch. However, they clearly need to substitute part of the GM machinery with specific equipment for the production of their product. (source: cnn.com, April 2014)

of new capital and a CES aggregator of used and new capital.[9]

$$k_j - k_0 = \widetilde{i}_{j,new} + g(i_{j,new}, i_{j,used}) \tag{1.2}$$

$$g(i_{j,new}, i_{j,used}) = [\eta^{\frac{1}{\epsilon}} (i_{j,new})^{\frac{\epsilon-1}{\epsilon}} + (1-\eta)^{\frac{1}{\epsilon}} (i_{j,used})^{\frac{\epsilon-1}{\epsilon}}]^{\frac{\epsilon}{\epsilon-1}} \tag{1.3}$$

where $\widetilde{i}_{j,new}$ and $i_{j,new}$ are new investment goods and $i_{j,used}$ is used capital sold by disinvesting firms. $\eta \in (0,1)$ is a parameter that determines the average ratio between new and used investment, while $\epsilon > 0$ is the elasticity of substitution between new and used investment goods.

The price of a unit of new capital in terms of the output good is 1, while the cost of a unit of used capital is equal to the sum of the price of used capital q_t and a per-unit reallocation cost γ. Hence, the CES price index associated with a composite g of new and used investment goods is

$$Q = [\eta + (1-\eta)(q+\gamma)^{1-\epsilon}]^{\frac{1}{1-\epsilon}}. \tag{1.4}$$

Clearly, firms will choose the cheapest option between a fully new investment $\widetilde{i}_{j,new}$ and a bundle g. As long as $q < 1-\gamma \Rightarrow Q < 1$, the bundle is the cheapest option and firms choose to make a fraction of their investment using used capital and set $\widetilde{i}_{j,new} = 0$. However, if the price of used capital became hypothetically higher then the price of new capital, firms would optimally buy only new capital. Throughout the paper, in equilibrium it will always be the case that $q < 1-\gamma$, hence I will focus on this case in the following analysis.

I interpret the elasticity ϵ as an inverse measure of capital specificity. When $\epsilon \to \infty$, new and used capital are perfect substitutes and the model nests a standard model of partial irreversibility with constant resale price $q = 1-\gamma$. On the other

[9] Differently from Jovanovic and Yatsenko (2012) and Edgerton (2011), and in order to make my model computationally tractable, I assume that the imperfect substitutability is in the investment technology rather than in the production function, which allows to endogenize the resale price while at the same time keeping track of only one type of capital as a state variable in the full dynamic model.

hand, when $\epsilon = 0$, the technology does not allow any substitutability between new and used capital.

With imperfect substitutability, for each price of used capital q, there is a well defined optimal ratio of used to new investment, in contrast to models that assume perfect substitutability. Increasing the ratio of used to new investment goods above this optimal level would be suboptimal, because the investing firm would be buying a larger amount of capital that was specific to another firm, relative to the new capital than can be freely specialized.

Finally, I assume that used capital is useless for consumers and the market for used capital clears between investing and disinvesting firms only. While it is true that some types of capital like cars and computers could be useful for consumers after having been used by firms, for most other kinds of equipment and for plants this is impossible. This assumption is equivalent to the assumption of total irreversibility in the aggregate, as for instance in Sargent (1980). Investment is partially reversible from an individual firm's point of view, because it can be sold to another firm, but in the aggregate, used capital cannot be retransformed into consumption.

1.3.2 Optimal investment and reallocation decisions

The solution to the firms' optimal investment problem can be easily characterized:

- If they are sufficiently productive, they will invest, buying a bundle of new and used capital at price Q. This will happen if $s_j \geq s^I = \frac{Q}{\alpha z k_0^{\alpha-1}}$. In this case, the optimal capital level is given by

$$k_j = \left(\frac{\alpha z s_j}{Q}\right)^{\frac{1}{1-\alpha}}.$$

- If they are sufficiently unproductive, they will disinvest, selling part of their capital at price q. This will happen if $s_j < s^D = \frac{q}{\alpha z k_0^{\alpha-1}}$. In this case, the optimal capital level is given by

$$k_j = \left(\frac{\alpha z s_j}{q}\right)^{\frac{1}{1-\alpha}}.$$

- Firms with intermediate productivity $s^D \leq s_j < s^I$ will choose to keep their capital level at k_0, as their marginal product of capital lies between the purchasing price Q and the selling price q.

1.3.3 Equilibrium in the market for used capital

Given a chosen amount of total investment, investing firms minimize their expenditure by buying a composite of new and used capital. By solving a standard CES expenditure minimization problem and integrating over the measure of investing firms we get total demand for used capital:

$$D_{used} = (1-\eta)\left(\frac{q+\gamma}{Q}\right)^{-\epsilon} \int_{s^I} \left[\left(\frac{\alpha z s}{Q}\right)^{\frac{1}{1-\alpha}} - k_0\right] dF(s). \tag{1.5}$$

On the other hand, total supply of used capital, coming from disinvesting firms is

$$S_{used} = \int^{s^D} \left[k_0 - \left(\frac{\alpha z s}{q}\right)^{\frac{1}{1-\alpha}}\right] dF(s). \tag{1.6}$$

The market-clearing condition $D_{used} = S_{used}$ defines implicitly the equilibrium price as a function of the aggregate productivity parameter z, $q = q(z;\epsilon)$, where I emphasize that this equilibrium price mapping depends on the elasticity of substitution ϵ. The following proposition relates this elasticity to the effect of aggregates shocks on irreversibility and reallocation.

Proposition 1. *(i) $q(z;\epsilon)$ is increasing in z. (ii) There exists an $\bar{\epsilon} > 0$ such that for $\epsilon < \bar{\epsilon}$ the elasticity of q with respect to z is greater than 1 and reallocation is increasing in z.*

The proof, in Appendix B, is based on an application of the Implicit Function Theorem to the market-clearing condition. To derive intuition on the mechanism, consider without loss of generality a marginal decrease in z. At a given resale price, disinvesting firms would like to sell more capital, because their optimal target level is now lower. This implies that the supply schedule of used capital (1.6) shifts out in a standard price-quantity space. Similarly, investing firms want to invest less and demand (1.5) shifts in. The price of used capital q must fall to clear the market, and the price index of investment Q will also decrease, although by less, because it is a CES average of 1 (the price of new capital) and $(q + \gamma)$. This reflects the fact that investing firms cannot fully benefit from the cheaper used capital on offer, because its specificity makes it less valuable for them.

For sufficiently low elasticity of substitution, q will fall by more than the initial fall in z, the threshold $s^D = q/\alpha z k_0^{\alpha-1}$ will decrease and the choice of new capital level conditional on s_j for disinvesting firms will increase, inducing less reallocation. Importantly, this is in contrast to a model where the resale price is constant, which implies that the disinvesting threshold would increase following a fall in z, and total reallocation would necessarily be higher.

We will see in the following sections that the reaction of the resale price potentially leads to amplification of exogenous aggregate productivity shocks. The mechanism works as follows: measured aggregate productivity (the Solow residual obtained assuming an aggregate production function) falls by more that the initial negative shock because the decrease in trading in the market for used capital leads unproductive firms to remain large relative to what they would be in a world with fixed resale price, and as a consequence a larger fraction of aggregate capital is operated by firms with lower productivity.

Furthermore, under a reasonable assumption on the distribution F, the model generates a countercyclical dispersion of marginal products. Assume for simplicity that F is uniform on $[s_{min}, s_{max}]$. Consider again a marginal decrease in z. Out of the inaction region, investing firms set their marginal product equal to Q and disinvesting firms set it equal to q. In the inaction region, the marginal product of each firm stays equal to its initial value. When aggregate productivity falls, the distance between q and Q increases, because q decreases by more than Q (Q is a CES average of $q + \gamma$ and 1). This has two effects. First, the difference between the marginal product of investing firms and that of disinvesting firms increases. Second, the mass of firms in the inaction region increases. Both effects necessarily lead to higher dispersion of marginal products.

The previous literature has often interpreted the countercyclical dispersion of returns from capital as a consequence of a worsening of financial frictions in recessions, with productive firms not getting enough external finance to implement high return projects (see for instance Cui, 2014, Chen and Song, 2013). This model highlights a different channel: part of the increase in the dispersion of returns in recessions is due to imperfect substitutability between new and old capital and hence can be fully efficient and should not be addressed with expansionary credit policies.

Figure 1.3 provides a graphical representation of the solution to the model and its comparative statics. Consider the upper horizontal line. Each point on the line is the marginal product of capital of a firm, evaluated at k_0. Firms with initial marginal product larger than Q invest up to the point where their marginal product equals Q. Likewise, firms with initial marginal product below the resale price q disinvest up to the point where their marginal product equals q. These investment and disinvestment decisions are represented by the curved arrows pointing towards the two prices. Firms with intermediate initial marginal product are in the inaction region and remain at k_0. The lower horizontal line corresponds to a decrease in z. Note that both Q and q decrease, but the former decreases by less than the latter. Hence, the inaction region becomes larger.

In conclusion, this simple static model shows that in an economy with a sufficient degree of capital specificity a fall in aggregate productivity leads to an even larger decrease in the resale price of capital, a decrease in reallocation and an increase in the dispersion of marginal products.

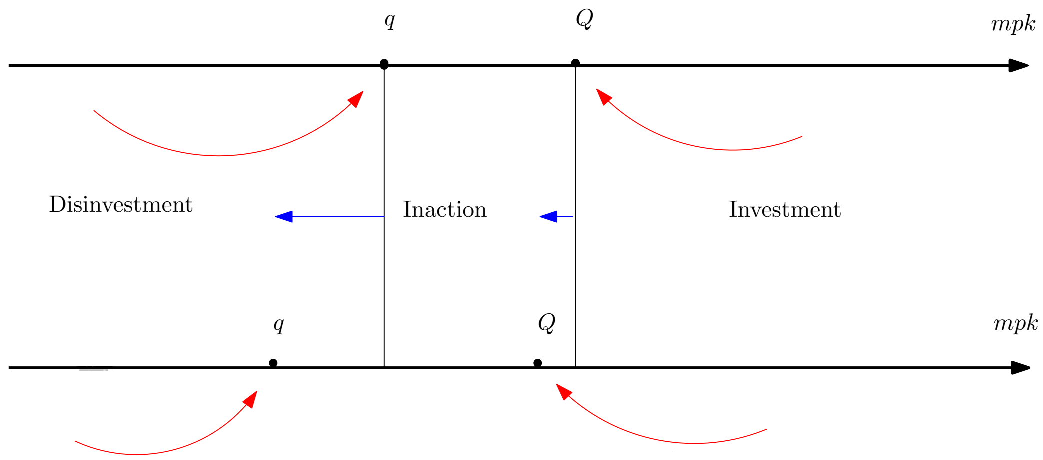


Figure 1.3: Graphical representation of the static mechanism

Firms' investment/disinvestment decisions and inaction region. Upper horizontal line: benchmark solution. Lower horizontal line: marginal decrease in aggregate productivity z.

1.4 Aggregate shocks and equilibrium real option values

This section presents a two-period model with uncertainty and forward-looking firms, where investment and reallocation decisions depend not only on current prices, but also, importantly, on future expected prices at which firms can buy and sell investment goods. The model extends the seminal work of Abel et al. (1996) by imposing equilibrium in the market for used capital. Abel et al. (1996) assume exogenous streams of purchasing and resale prices of capital and focus on solving the individual firm's problem of partially irreversible investment under idiosyncratic uncertainty. They show that with partial irreversibility, as part of the investment is sunk, the firm has an option value from waiting until future productivity is known (or equivalently the future output price).

Here, I consider instead a continuum of firms hit by idiosyncratic and aggregate shocks and I impose that the market for used capital clears in all states. This allows me to obtain results on the equilibrium effects of aggregate shocks on firms' real option values in an environment where all firms make their investment decisions taking these options into account. As the resale price is positively correlated with the aggregate productivity shock, the option value to sell capital is procyclical, contrary to what happens in partial equilibrium, where this option is more valuable in recessions. After a negative persistent aggregate TFP shock, capital is not only less productive, but is also perceived as more irreversible, as its expected future resale price falls. This section ends with a discussion of how such equilibrium effects of aggregate shocks on real option values affect both investment and reallocation.

1.4.1 Two-period model

There is a continuum of firms with idiosyncratic productivity s_{jt} producing with technology

$$y_{jt} = z_t s_{jt} k_{jt}^{\alpha}, \tag{1.7}$$

for $t = 1, 2$. Both z_t and s_{jt} follow a positively autocorrelated process. In particular, aggregate productivity takes either of two values $\left\{z^L, z^H\right\}$ and a Markov transition matrix gives conditional probabilities for time $t = 2$. I assume $Pr\left\{z_2 = z_1\right\} > 1/2$. The idiosyncratic shock at $t = 1$ is drawn from a log-normal distribution with mean $-\sigma_1^2/2$ and variance σ_1^2. At time 2, the shocks satisfies $\log\left(s_2\right) = \rho\ \log\left(s_1\right) + v_2$, with $v_2 \sim N(-\sigma_2^2/2, \sigma_2^2)$.

Firms start the initial period with a common level of capital k_0, observe the realizations of the two productivity shocks, (s_{j1}, z_1), and are allowed to choose their capital level k_{j1} before starting production. If they invest, they can purchase a combination of new capital and used capital, which is being sold by disinvesting firms who decide to decrease their capital level. As in the static model of the previous section, the investment technology is given by (1.2) and the price index of investment goods is given by (1.4) in both periods, except that q and Q will now have a t subscript. The reallocation cost γ is constant.

After the investment/disinvestment activity is concluded, production takes place. Abstracting from physical depreciation for simplicity of exposition, firms start the second period with an initial level of capital k_{j1}, observe the realizations of s_{j2} and z_2 and are again allowed to adjust their capital level before production. Then production takes place again and at the end of the period firms receive a terminal value proportional to their capital level, χk_{j2}, with $\chi \geq 0$.

1.4.2 Value of a firm

Because $q_t < Q_t$, in both periods some firms will invest, some will disinvest and some will keep their capital stock unchanged because the marginal product of their capital lies between the two prices. Let's start by characterizing the value of a firm after its choice of capital at $t = 1$. The next subsection will then move backwards and solve for the optimal choice of k_{j1}. By anticipating optimal behavior at $t = 2$, the value of a firm can be decomposed into a component that assumes no further adjustment in the second period, a component that depends on the opportunity to buy more capital in the second period (call option) and a component that depends on the possibility to sell some capital in the second period (put option). The call option will be exercised only for sufficiently high s_{j2} and the put option for sufficiently low s_{j2}. In the following derivation I will drop the subscript j for notational convenience and consider a generic firm. At $t = 1$, after observing the pair (s_1, z_1) and choosing k_1, the value of the firm is

$$V(k_1, s_1, z_1) = z_1 s_1 k_1^\alpha + \beta \left(\mathrm{E}_1 z_2 s_2 k_1^\alpha + \chi k_1\right) + \beta C(k_1, s_1, z_1) + \beta P(k_1, s_1, z_1) \quad (1.8)$$

where E_1 is a conditional expectation operator that sums over the future realizations of z and integrates over the distribution of s_2, conditional on (s_1, z_1). The value of the firm consists of the value of producing in both periods with k_1, i.e. without doing any further adjustment at $t = 2$, plus the call option value of increasing the capital stock, $C(k_1, s_1, z_1)$ and the put option value of selling part of the capital stock in the second period, $P(k_1, s_1, z_1)$.

The call option value is given by

$$C(k_1, s_1, z_1) = \mathrm{E}_{z_1} \int_{s_2^I(k_1,z_2)} \{[z_2 s_2 (k_2 (s_2, z_2))^\alpha + \chi k_2 (s_2, z_2)] \\ - [z_2 s_2 k_1^\alpha + \chi k_1] - Q_2 [k_2 (s_2, z_2) - k_1]\} \, dF (s_2 | s_1) \quad (1.9)$$

where E_{z_1} sums over realizations of z_2, with probabilities conditional on z_1. This option will be exercised at $t = 2$ if idiosyncratic productivity turns out to be above the threshold $s_2^I (k_1, z_2) = \frac{Q_2 - \chi}{\alpha z_2 k_1^{\alpha-1}}$, in which case the firm will invest and go to a capital level given by

$$k_2 = \left(\frac{\alpha z s_{j2}}{Q_2 - \chi} \right)^{\frac{1}{1-\alpha}} > k_1.$$

Similarly, the put option is

$$P(k_1, s_1, z_1) = \mathrm{E}_{z_1} \int^{s_2^D(k_1,z_2)} \{[z_2 s_2 (k_2 (s_2, z_2))^\alpha + \chi k_2 (s_2, z_2)] \\ - [z_2 s_2 k_1^\alpha + \chi k_1] + q_2 [k_1 - k_2 (s_2, z_2)]\} \, dF (s_2 | s_1) \quad (1.10)$$

and will be exercised at $t = 2$ if idiosyncratic productivity turns out to be below the threshold $s_2^D (k_1, z_2) = \frac{q_2 - \chi}{\alpha z_2 k_1^{\alpha-1}}$ by selling capital up to the level

$$k_2 = \left(\frac{\alpha z s_{j2}}{q_2 - \chi} \right)^{\frac{1}{1-\alpha}} < k_1.$$

It is easy to see that the value of the call option is decreasing in the realizations of Q_2, while the put option is increasing in the realizations of q_2. While a higher Q_2 makes it harder to expand tomorrow, a higher q_2 makes it easier to downsize, should it be desirable.

1.4.3 Optimal investment and reallocation decisions

We can now characterize the optimal choice of capital level in the first period. At $t = 1$, firms compare the marginal value of their initial capital level k_0 with the purchasing price Q_1 and the selling price q_1. Call V_k the partial derivative of V with respect to its first argument.

- Firms who receive an idiosyncratic shock such that $V_k(k_0, s_1, z_1) \geq Q_1$ will choose to invest and their optimal capital level $k_1(s_1, z_1)$ satisfies
$$\frac{V_k(k_1, s_1, z_1)}{Q_1} = \frac{W_k(k_1, s_1, z_1) + C_k(k_1, s_1, z_1) + P_k(k_1, s_1, z_1)}{Q_1} = 1$$
where I have emphasized that the marginal value of capital is composed by the present discounted value of their marginal product assuming no further adjustment, which I call W_k, the marginal call C_k and the marginal put P_k. Note that W_k and P_k are positive, while C_k is negative as increasing the capital level implies exercising part of the call option value.

- For firms with sufficiently low idiosyncratic productivity, $V_k(k_0, s_1, z_1) < q_1$. They will disinvest and choose $k_1(s_1, z_1)$ by solving
$$\frac{V_k(k_1, s_1, z_1)}{q_1} = \frac{W_k(k_1, s_1, z_1) + C_k(k_1, s_1, z_1) + P_k(k_1, s_1, z_1)}{q_1} = 1$$

- Firms with intermediate productivity, such that $q_1 \leq V_k(k_0, s_1, z_1) < Q_1$, are in the inaction region and optimally keep $k_1(s_1, z_1) = k_0$.

As in the static model, firms compare the marginal value of capital with Q if they consider investing and with q if they consider disinvesting. Differently from the static model, however, this marginal value now takes into account the variations in the option values induced by such investment and disinvestment activity.

The market for used capital clears, meaning that total disinvestment from unproductive firms equals investment in used capital coming from investing firms. The market-clearing equation is analogous to that in the previous section.

1.4.4 Put option in booms and recessions

As the evidence presented in section 1.2 suggests, the resale price of capital is more volatile and procyclical than the price of new capital. Hence, I will focus on the put option value and its reaction to shocks, although similar arguments can be made about equilibrium effects of shocks on the call option.

To understand how equilibrium real options affect investment and disinvestment behavior following business-cycle shocks, consider the marginal value of the put option P_k. Differentiating (1.10) with respect to the choice of capital and writing the expectation more explicitly, we get

$$P_k(k_1, s_1, z_1) = \sum_{z_2 \in \{z^L, z^H\}} Pr\{z_2|z_1\} \int^{s_2^D(k_1,z_2)} \left[q_2(z_1, z_2) - \alpha z_2 s_2 k_1^{\alpha-1} - \chi\right] \, dF\left(s_2 \,|s_1\right) \quad (1.11)$$

where I emphasize that the equilibrium resale price depends on both realizations of the aggregate shock. Hence, conditional on z_1, there are two possible outcomes for q_2 depending on the realization of z_2.

Intuitively, P_k is increasing in the expected value of the future resale price q_2, as this price adds value to a marginal unit of capital bought in the first period, in the case this unit has to be resold in the second period. Consider the two elements of the summation over z_2. If one keeps the resale price q_2 constant, it appears that the marginal put value of capital is decreasing in z_2, i.e. a decrease in the value of aggregate productivity leads necessarily to an increase in P_k. This is because P_k depends negatively on the marginal product of installed capital. However, the expected resale price is also endogenous in this model. In numerical examples, for sufficiently low elasticity of substitution between new and used capital ϵ, this equilibrium effect dominates the effect of exogenous changes in z_2, implying that the marginal put option value becomes procyclical. The intuition for this is that the value of a marginal unit of capital purchased in the first period depends positively both on its productivity in the second period and on its resaleability.

Figure 1.4 illustrates the payoff of the put option value in equation (1.10), evaluated at k_0, as a function of the initial idiosyncratic shock s_1 and for each of two values of z_1. I label the low realization of the aggregate state "recession" and the high realization "boom". This figure shows that this option value has a similar payoff function to a financial option. Here the strike price is the resale price q_2 and the underlying asset value is the marginal value of the firm's capital. For a firm with very low s_1, k_0 is a relatively high capital level, so that assuming no adjustment at $t = 1$, it is likely that some disinvestment will be optimal at $t = 2$, given the autocorrelation of s_t. This explains a high put option value. On the other hand, for a firm with very high s_1, the optimal size is higher than k_0, so that if the firm keeps its capital level at k_0 it is unlikely that there will be any need for disinvesting at $t = 2$, which explains a put option value close to 0.

Figure 1.5 shows what happens to the option value after firms choose k_1, both when $z_1 = z^L$ and when $z_1 = z^H$. Low productivity firms exercised some of their initial put value by selling some capital. High productivity firms, on the other hand, invest and purchase some put option value. Firms with intermediate productivity are in the inaction region and optimally choose to keep $k_1 = k_0$, so their put option value after trading equals the initial put option value. This figure illustrates that when aggregate productivity is low, the put option value falls, because of the equilibrium effect on the expected resale price illustrated above, thus making investment more irreversible.

1.4.5 Equilibrium real options and investment

The model has rich implications in terms of the effects of aggregate shocks on investment. In a boom, capital is now attractive for two reasons. First of all, it is directly more productive. Second, it is easier to resell, in case a bad idiosyncratic shock hits the firm in the future. Third, it is more expensive today, because the current price of used assets increases and hence total investment comes at a higher price.

In the quantitative model presented in the next section, I investigate the aggregate effects of all these different incentives on investment and disinvestment decisions. It turns out that in general equilibrium endogenous irreversibility smooths aggregate investment, bringing its volatility and autocorrelation closer to the data, consistently

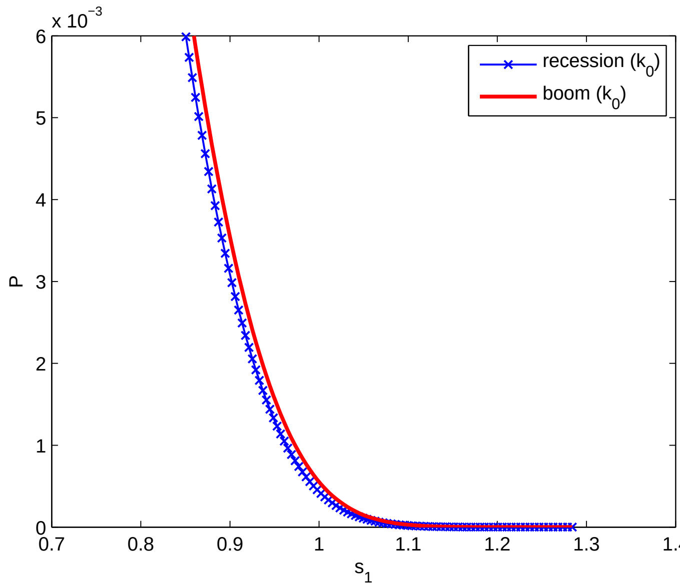


Figure 1.4: Put option value, before trading at $t = 1$

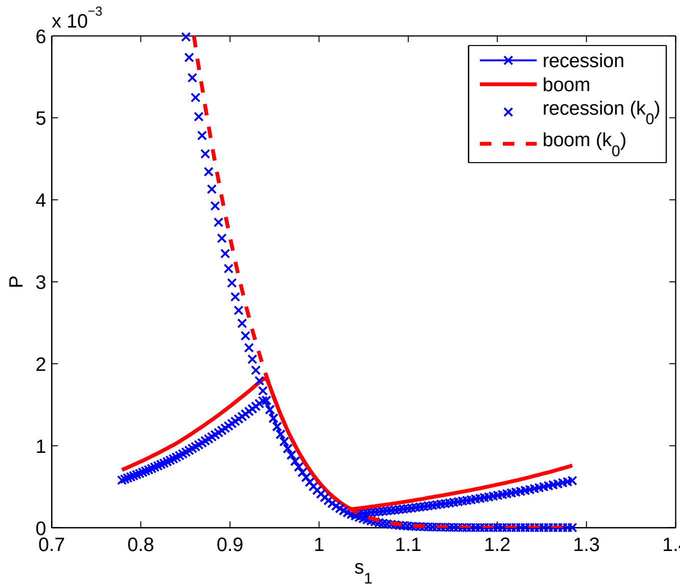


Figure 1.5: Put option value, after trading at $t = 1$

with the original conjecture of Bertola and Caballero (1994) on the aggregate effects of micro-irreversibilities and in contrast to DSGE models where the resale price of capital is assumed to be constant.

1.4.6 Equilibrium real options and reallocation

For disinvesting firms, the first order condition for k_1 suggests a key comparison between the marginal value of capital and the current resale price. The ratio between the present discounted value of marginal returns and q_1 behaves similarly to the static model. Let us disregard variations in the marginal call value as they are small, given the relatively low volatility of Q_t and focus on the ratio $\frac{P_k}{q_1}$.

Unproductive firms compare the price they can get for their assets at $t = 1$, with the value from waiting to disinvest until the second period, which as we have seen is an increasing function of the expected value of q_2. This allows us to identify two forces that act in opposite directions. On the one hand, in a recession the marginal put option falls, because the resale value of capital is going to fall at $t = 2$ with high probability (as the aggregate shock is persistent). This would imply that it is optimal to disinvest by more in the first period. On the other hand, also the current resale price in the first period is low. Importantly, with positive probability z_2 will be high and the resale price will increase, in which case it would be optimal to disinvest by less in the first period and wait until the second period.

Because of mean reversion of the aggregate shock process, following a low realization of the aggregate TFP shock, the fall in the current price of used capital dominates the fall in its future expected value, generating a value of waiting to disinvest in the future, further delaying reallocation and amplifying the static effects analyzed in section 1.3.

1.5 A DSGE model with endogenous irreversibility

This section presents an infinite-horizon general equilibrium model that combines all the static and dynamic effects illustrated in the previous sections and includes risk aversion and endogenous labor supply, allowing for a quantitative evaluation of the mechanism.

1.5.1 Households

There is a representative household who consumes the output good, supplies labor and owns shares in all firms in the economy. Her preferences are described by the utility function

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \left[\log\left(c_t\right) - \psi n_t\right] \tag{1.12}$$

where c_t is consumption and n_t are hours worked.

The representative household's budget constraint is

$$c_t = w_t n_t + \pi_t \tag{1.13}$$

where π_t are aggregate profits.[10]

The labor supply schedule is defined by the first order condition that equates the marginal rate of substitution between hours and consumption to the wage w_t

$$\psi c_t = w_t. \tag{1.14}$$

[10] Alternatively, one could write this budget constraint including the household's choice of buying and selling shares in all firms. In equilibrium, her portfolio would have to coincide with the distribution of firms in the economy and stock prices would be given by the firms' value functions below. The distinction between these two formulations is immaterial in terms of competitive equilibrium allocations and prices.

1.5.2 Firms

Consider now firms' optimization problem. In each period t, productivity of firm j is the product of an aggregate component $z_t \in \left\{z^L, z^H\right\}$ that follows a Markov chain with transition matrix T_z and an idiosyncratic component $s_{jt} \in \left\{s^L, s^H\right\}$ with Markov transition matrix T_s. The firm produces a homogenous output good with technology

$$y_{jt} = z_t s_{jt} k_{jt}^{\alpha} n_{jt}^{\nu} \tag{1.15}$$

with $\alpha + \nu < 1$, and chooses current labor demand and the future level of capital in order to maximize its value for the consumer taking prices q_t and w_t as given.[11] By assuming a flexible labor market with no adjustment costs, I can separate the labor demand choice from the investment decision in a very convenient way. I will first describe the intratemporal labor decision and then derive the implied return on capital in order to formulate the intertemporal investment problem.

Labor demand equates the marginal product of labor to the wage rate:

$$n_{jt} = \left(\frac{\nu z_t s_{jt} k_{jt}^{\alpha}}{w_t}\right)^{\frac{1}{1-\nu}} \tag{1.16}$$

Using (1.16), it is easy to derive an expression for output net of the wage bill as a function of the two productivity shocks, current capital level and wage:

$$y_{jt} - w_t n_{jt} = A\left(w_t\right) z_t^{\theta} s_{jt}^{\theta} k_{jt}^{\alpha\theta}, \tag{1.17}$$

where $A\left(w_t\right) = \left[\left(\frac{\nu}{w_t}\right)^{\frac{\nu}{1-\nu}} - w_t \left(\frac{\nu}{w_t}\right)^{\frac{1}{1-\nu}}\right]$, and $\theta = 1/\left(1-\nu\right)$. This transformation of the production function is used by firms in order to evaluate the return on investment in physical capital. In other words, the flexible labor demand decision is incorporated in their expectations as they know that in every period they will be free to reoptimize their required labor input.

[11] Differently from the simple models presented above, here I assume that capital is chosen one period in advance. This assumption makes the model more easily comparable with standard business-cycle models.

Let m be the distribution of firms over individual capital level and idiosyncratic productivity. Both the price of used capital and the wage will depend on it, so that this distribution is now a state variable with its own law of motion.

$$m_{t+1} = \Gamma(m_t, z_t) \tag{1.18}$$

Let us focus on a recursive solution to the firm's problem, with state vector (k, s, z, m). After observing the state, each firm decides whether to invest or disinvest, and by how much. Switching to recursive notation, the value of an investing firm is

$$V^i(k, s, z, m) = \max_{k' \geq (1-\delta)k} A\,(w)\, z^\theta s^\theta k^{\alpha\theta} - Q\,[k' - (1-\delta)\,k] + \beta \mathrm{E}\left\{\frac{c}{c'} V(k', s', z', m') | s, z\right\} \tag{1.19}$$

and the value of a disinvesting firm is

$$V^d(k, s, z, m) = \max_{k' \leq (1-\delta)k} A\,(w)\, z^\theta s^\theta k^{\alpha\theta} - q\,[k' - (1-\delta)\,k] + \beta \mathrm{E}\left\{\frac{c}{c'} V(k', s', z', m') | s, z\right\} \tag{1.20}$$

At the beginning of each period, the discrete choice between investment and disinvestment gives $V(k, s, z, m) = max\left\{V^i(k, s, z, m), V^d(k, s, z, m)\right\}$. Note that these Bellman equations implicitly define the value of the firm as the present discounted value of profits (i.e. output net of the wage bill and investment expenditure), evaluated using the representative household's stochastic discount factor.

1.5.3 General Equilibrium

Market clearing in the used capital market needs to be imposed in an analogous way to the simpler models in the previous sections. Investing firms demand new capital and used capital by solving a standard CES expenditure minimization problem and market clearing implies that total investment in used capital equals total disinvestment.

I can now define a recursive competitive equilibrium.

Definition 1. *A recursive competitive equilibrium is defined as a set of functions m, Γ, w, q, Q, π , C, N, V^i, V^d, V, n, k', i, i_{new}, i_{used}, d that solve the household's and* *firms' optimization problems and clear markets for the output good, labor and used capital:*

- *Consumption* $C(z,m)$ *and labor supply* $N(z,m)$ *solve the consumer's problem of maximizing (1.12) subject to (1.13)*

- *Firms labor demand* $n(k,s,z,m)$ *satisfies equation (1.16)*

- *The value functions* V^i, V^d *and* V *satisfy the functional equations (1.19), (1.20) and* $V(k,s,z,m) = max\left\{V^i(k,s,z,m), V^d(k,s,z,m)\right\}$

- *For investing firms, i.e. firms such that* $V^i(k,s,z,m) \geq V^d(k,s,z,m)$ *the policy function* $k'(k,s,z,m)$ *solves (1.19), investment is* $i(k,s,z,m) = k'(k,s,z,m) - (1-\delta)k$ *and is allocated to new and used investment goods according to the CES expenditure minimization first order condition:*

$$\frac{i_{used}(k,s,z,m)}{i_{new}(k,s,z,m)} = \frac{1-\eta}{\eta}\left(q(z,m)+\gamma\right)^{-\epsilon}$$

- *For disinvesting firms, i.e.* $V^i(k,s,z,m) < V^d(k,s,z,m)$, *the policy function* $k'(k,s,z,m)$ *solves (1.20) and disinvestment is* $d(k,s,z,m) = (1-\delta)k - k'(k,s,z,m)$

- *Aggregate profits are given by* $\pi(z,m) = z\int sk^\alpha n^\nu \, dm(k,s) - w(z,m)N(z,m)$

$$-Q(z,m)\int i(k,s,z,m)\, dm(k,s) + q\int d(k,s,z,m)\, dm(k,s)$$

- *The market for the output good clears:*

$$C(z,m) = z\int sk^\alpha n^\nu \, dm(k,s) - Q(z,m)\int i(k,s,z,m)\, dm(k,s) +$$

$$q\int d(k,s,z,m)\, dm(k,s)$$

- *The labor market clears:* $N(z,m) = \int n(k,s,z,m)\, dm(k,s)$

- *The market for used capital clears:*

$$\int d(k,s,z,m)\; dm(k,s) = \int i_{used}(k,s,z,m)\; dm(k,s)$$

- *The price functions $q(z,m)$ and $Q(z,m)$ satisfy equation (1.4)*

- *The transition function Γ defines the evolution of the distribution of firms m according to the policy function k' and the Markov transition matrices T_s and T_z*

1.5.4 Calibration

Table 1.2 reports the choice of parameter values. When possible, these choices reflect the attempt to stay close to previous work on firm heterogeneity and investment for comparison purposes (in particular Khan and Thomas, 2013). A period coincides with a year: this choice is motivated by the fact that both data on capital reallocation and on micro-level investment are yearly.

Parameters β, ψ and δ correspond to a yearly interest rate of 4 percent, hours worked equal to .33 and an investment/capital ratio of 6.5%. The capital share α is then set to match a capital/output ratio around 2.5. The labor share ν is 60% as in US postwar data.

Both aggregate and idiosyncratic shocks are initially parametrized as AR(1) processes in logs with autocorrelations ρ_z and ρ_s and standard deviations of innovations $\sigma_{inn,z}$ and $\sigma_{inn,s}$. Then they are discretized following the Rouwenhorst method with two values for each shock.

In particular, aggregate productivity z_t is parametrized as in Khan and Thomas (2013), who estimate a process for the Solow residual in US data. This gives a standard deviation of innovations of .014 and an autocorrelation coefficient of .909.

The standard deviation of the process for the idiosyncratic shock s is calibrated to match the standard deviation of the distribution of investment ratios computed by Cooper and Haltiwanger (2006), which is .33. The autocorrelation of the process is parametrized as in Khan and Thomas (2013) to be equal to .65. This implies a standard deviation of innovations of .084.[12]

The investment technology is defined by two parameters: η and ϵ. The first parameter is calibrated to match the steady-state ratio of used capital to total capital purchased by investing firms. The target chosen is a ratio of 30%, which is an upper bound of the estimates found by Eisfeldt and Rampini (2007), in order to take into account that smaller firms out of their sample are likely to buy a higher ratio of used capital, as this ratio appears to be decreasing in firms' size in their empirical evidence.

The elasticity of substitution between new and used investment goods ϵ is a key parameter of the model. Edgerton (2011) estimates this elasticity using data from construction equipment, aircraft and farming equipment and exploiting a tax-credit reform that affected only new investment. He finds values in a range between 1 and 10. I set $\epsilon = 5$ as my benchmark value. Beside being an intermediate value in this range of estimates, it allows to match the standard deviation of the ratio between used and new capital expenditure from ACES data, which is around 1%. I show the results for different value of ϵ in Appendix C. Note that the baseline choice implies that the relative price of used capital will move less in the model than in the data shown in the empirical section, so that this parametrization is quite conservative. Finally, I set $\gamma = .01$, which implies an average level of irreversibility of .96, close to the constant resale price in Khan and Thomas (2013). Hence, in the stationary equilibrium the baseline economy is a very close match to a version of the Khan and Thomas (2013) economy without financial frictions.

[12] The discretization with a two-state Markov chain implies that the average autocorrelation of investment rates and the frequency of large adjustments (lumpiness) cannot be matched at the same time as the standard deviation of investment rate, differently from Khan and Thomas (2013).

Parameter	Value
β	.96
δ	.065
ψ	2.15
α	.27
ν	.6
$\sigma_{inn,z}$	.014
ρ_z	.909
$\sigma_{inn,s}$	.084
ρ_s	.65
η	.7
ϵ	5
γ	.01

Table 1.2: Parameter values in the baseline model

1.5.5 Computation

I solve the model using an extension of the method of Krusell and Smith (1998) and Khan and Thomas (2008, 2013) that takes care of "non-trivial market clearing" in the market for used capital.[13] I approximate the distribution m with its first moment, aggregate capital. Agents perceive a law of motion $\log(K') = \hat{\phi}_0 + \hat{\phi}_1 \log(K) + \hat{\phi}_2 z + \eta$ and price functions $\hat{q}(K,z)$, $\hat{w}(K,z)$. Given these perceptions, I solve the individual firm's problem by value function iteration and obtain the policy functions, making

[13] By "non-trivial market clearing", Krusell and Smith (2006) mean that a price has to be solved for at each period during the simulation equating total supply and total demand (in this case for used capital), differently from what happens for instance in Krusell and Smith (1998), where the rental rate of capital can be easily solved for analytically given the predetermined level of aggregate capital.

them dependent on the current resale price q_t. Then, I simulate a continuum of firms using the simulation method of Young (2010) and update the price functions by explicitly imposing market clearing in the used capital market (and in the labor market) along the simulation. Finally, I update the laws of motion using standard regression methods up to convergence. The accuracy of the solution is illustrated in Appendix C.

1.6 Results

This section presents the numerical results from the full model, starting from firm dynamics in a stationary equilibrium and then moving on to the business-cycle properties of the model.

1.6.1 Stationary equilibrium: no aggregate uncertainty

This subsection describes the equilibrium of the model when there is no aggregate uncertainty and z is always equal to its mean. Consider first the investment policy function obtained by solving the firms' optimization problem. As in the simpler models presented in the previous sections, the wedge between the price of investment goods Q_t and the resale price q_t generates inaction areas, where firms optimally let their capital depreciate without taking any action. As $q_t < Q_t$, it is always the case that the capital level that solves (1.19) without the inequality constraint of positive investment, call it $k^i(k, s, z, m)$, is strictly less than the capital level that solves (1.20) without the inequality constraint of positive disinvestment, call it $k^d(k, s, z, m)$. It follows that the policy function for future capital will be:

$$k'(k,s,z,m) = \begin{cases} k^i(k,s,z,m), & k \leq k^i(k,s,z,m)/(1-\delta) \\ (1-\delta)k, & k^i(k,s,z,m)/(1-\delta) < k \leq k^d(k,s,z,m)/(1-\delta) \\ k^d(k,s,z,m), & k > k^d(k,s,z,m)/(1-\delta). \end{cases}$$

Figure 1.6 illustrates the policy function for future capital for firms with low produc-

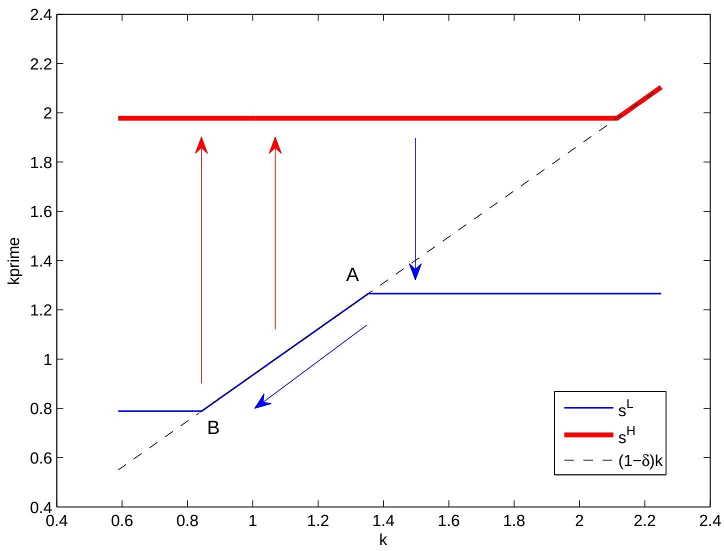


Figure 1.6: Policy function for future capital level

tivity (thin blue line) and high productivity (thick red line) under the parametrization reported in Table 1.2. The variable on the x-axis is the current capital level, while on the y-axis I plot next period capital. In a world without resale frictions, this picture would consist of only two horizontal lines, one at higher level for s^H and one at a lower level for s^L and firms would jump from one level to another depending on the current realization of s and regardless of their size k, given that there would be no adjustment costs. However, partial irreversibility induces disinvesting firms not to sell the whole amount of capital needed to jump to the bottom part of the blue line (point B). This is because they expect to need to reinvest in the future if they receive a positive idiosyncratic shock in the following period and they would clearly incur a loss due to the fact they would repurchase capital at a higher price than the one obtained for their disinvestment. In other words, the wedge between the price paid for investment and the price received for disinvestment creates an option value from waiting and hence an inaction region where firms optimally wait before taking any action and just let their capital depreciate in the hope for a high productivity shock. The inaction region for low productivity firms is the upward sloping part of the thin blue line, between points A and B, which coincides with the depreciation line (dashed black line). Note that there is an inaction region also for high productivity firms (top right in the figure), but it turns out that firms never invest enough to enter this area in equilibrium.

Firm level dynamics in the stationary equilibrium are as follows. As soon as firms get a high idiosyncratic shock, they jump to the horizontal part of the thick red line. They stay there as long as they have high productivity. As soon as they get a bad shock that brings them to s^L, they sell part of their capital and jump down to the thin blue line, close to point A. Then, as long as they have productivity s^L they move down left along this line until they reach point B, where they stay until a further positive shock. Hence, on the market for used capital, supply comes from the firms that have a high level of capital and get a negative idiosyncratic shock, whereas demand comes from firms of all sizes that obtain a positive shock, plus the smallest firms with low productivity that invest to keep their size constant.

These firm-level dynamics give rise to the stationary distribution plotted in Figure 1.7, where the x-axis is again k and the y-axis is the mass of firms m. The thick red line with high mass on the right-hand side of the picture represents firms with productivity s^H. Moving towards the left, the thin blue lines with crosses represent the masses of firms with productivity s^L. Gradually, the mass decreases as some of the firms with those sizes receive a positive shock and only the remaining fraction let their capital depreciate for one more period. At the left end of the picture, there is a mass of low productivity firms that just rebuy their depreciated capital and keep the same small size until they get a positive idiosyncratic shock (point B).

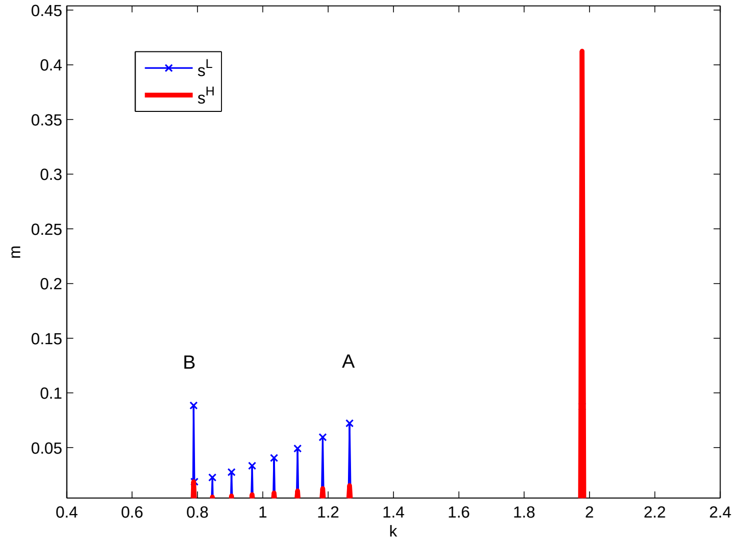


Figure 1.7: Stationary distribution of firms

1.6.2 Business cycles and capital reallocation

I will now turn to describe the properties of the economy when it is hit by aggregate productivity shocks. Table 1.3 shows standard business-cycle statistics as well as statistics for the resale price of capital and the reallocation series, taken from a simulation of the model economy. The first row presents the unconditional mean of the variables of interest. To construct the second and third rows, I HP-filter the data with a smoothing parameter of 6.25 and then compute relative standard deviations and correlations with output.[14] The standard deviation of output is in parentheses. It is instructive to compare these business-cycle statistics with those obtained in an economy with constant resale price (Table 1.4), which closely resembles a version of the Khan and Thomas (2013) economy without financial frictions.

As far as standard business-cycle variables are concerned, endogenous irreversibil-

[14]Khan and Thomas (2013) use a smoothing parameter of 100. In the interest of a comparison with their results, I recompute these two tables with a smoothing parameter of 100 in Appendix C.

ity reduces output and employment volatility slightly and investment volatility quite significantly (this result is discussed in more detail in a following subsection). By comparing the last columns of Tables 1.3 and 1.4, it can be seen that going from a constant q to a market-clearing resale price of capital, reallocation turns from being strongly countercyclical (the puzzle) to being strongly procyclical and approximately three times as volatile as output. In the data, the ratio between the standard deviation of (filtered) reallocation and output is 7.942 and their correlation is .562. Hence, the model cannot quite match the empirical volatility ratio and at the same time it overestimates the correlation with output. However, both statistics are significantly closer to the data than in the model with constant resale price.

In Appendix C, it can be seen that robustness exercises with respect to the elasticity of substitution ϵ lead to very similar qualitative results. The volatility of the reallocation series is decreasing in ϵ, as expected: the lower this elasticity, the more specific used capital, and hence the stronger the effects of aggregate shocks on reallocation. The high correlation with output is robust to different values of this elasticity.

Statistic	Y	C	I	K	N	q	reall
mean	.587	.489	.097	1.490	.335	.9521	.048
$\sigma(.)/\sigma(Y)$	(1.39)	.474	4.102	.257	.546	.198	**3.276**
corr(.,Y)	1	.895	.936	-.271	.988	.963	**.959**

Table 1.3: Business-cycle statistics: baseline model

Statistic	Y	C	I	K	N	q	reall
mean	.587	.489	.097	1.493	.335	.9512	.0484
$\sigma(.)/\sigma(Y)$	(1.54)	.431	5.04	.291	.657	0	**1.184**
corr(.,Y)	1	.764	.914	-.212	.984	0	**-.7937**

Table 1.4: Business-cycle statistics: constant q

A limitation of the model is that the volatility of the resale price q_t is small compared to the volatilities implied by the sectoral data presented in section 1.2 and Appendix A. However, this suggests that even a small amount of volatility and procyclicality in this price can go a long way in explaining the empirical patterns of capital reallocation.

In Figures 1.8 and 1.9, I show the (unfiltered) paths of the resale price and the reallocation series when the economy goes from a long series of high realizations of the aggregate productivity shock to a long series of low realizations. It can be clearly seen that in the baseline model the initial fall in the resale price induces a large decrease in capital reallocation. This is in contrast with a model with constant resale price (black line with crosses), where reallocation actually increases in the first two years of the recession, and then falls gradually as the capital stock and the size of the whole economy decrease.

Figure 1.10 illustrates the effect of endogenous irreversibility on the dispersion of marginal product of capital, as measured by the average marginal product for high productivity (s^H) firms and that for low productivity ones (s^L). Consistently with a growing body of empirical evidence (Bloom et al., 2012), the model implies that returns are more dispersed in recessions than in booms. This is clearly related to the lack of reallocation, because large unproductive firms downsize less than they would do in a model with constant irreversibility and this prevents an equalization of marginal returns. The previous literature has either taken the countercyclicality of dispersion of returns as fully exogenous (e.g. Bloom et al., 2012) or explained it as a consequence of financial frictions (e.g. Chen and Song, 2013). This paper suggests a different explanation, based on partial capital specificity, which bears important consequences for policy in the current recovery. If one interprets the high levels of observed dispersion of returns from capital as due to a worsening of credit frictions, it is possible that a credit expansion or unconventional monetary policy could facilitate reallocation and strengthen the recovery. If instead the high dispersion is fully efficient and due to capital specificity and equilibrium irreversibility, then no policy intervention is in order and credit expansions are not relevant.

The dynamics of the distribution of firms over the business cycle are illustrated in Figure 1.11. It can be seen that the distribution becomes more compressed when the economy moves from a boom to a recession: large unproductive firms downsize by less than they do in booms (compare points A and A'), while highly productive units expand by less. Jointly, these facts explain the fall in reallocation and the increase in the dispersion of returns.

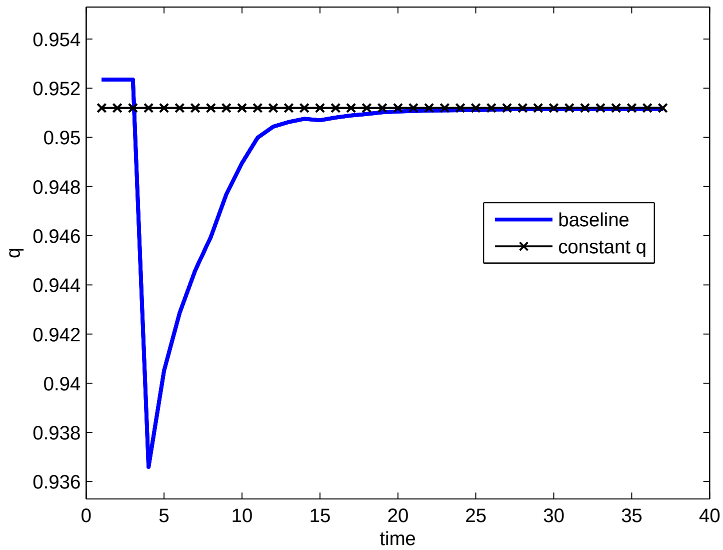


Figure 1.8: Price of used capital

Transition from long sequence $z_t = z^H$ to long sequence $z_t = z^L$. Response of the price of used capital. Comparison between the baseline model and a model with constant resale price. Unfiltered simulated data.

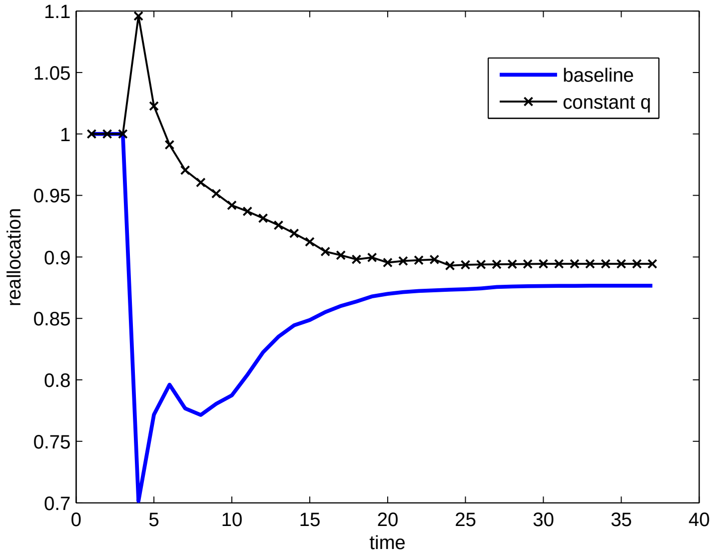


Figure 1.9: Capital reallocation

Transition from long sequence $z_t = z^H$ to long sequence $z_t = z^L$. Response of capital reallocation. Comparison between the baseline model and a model with constant resale price. Unfiltered simulated data.

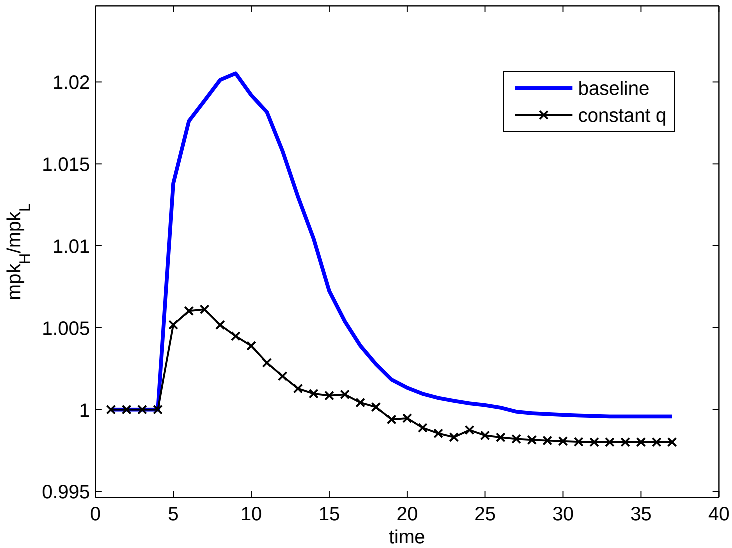


Figure 1.10: Dispersion of returns

Transition from long sequence $z_t = z^H$ to long sequence $z_t = z^L$. Response of the ratio between the average marginal product of high productivity firms (s^H) and that of low productivity firms (s^L). Comparison between the baseline model and a model with constant resale price. Unfiltered simulated data.

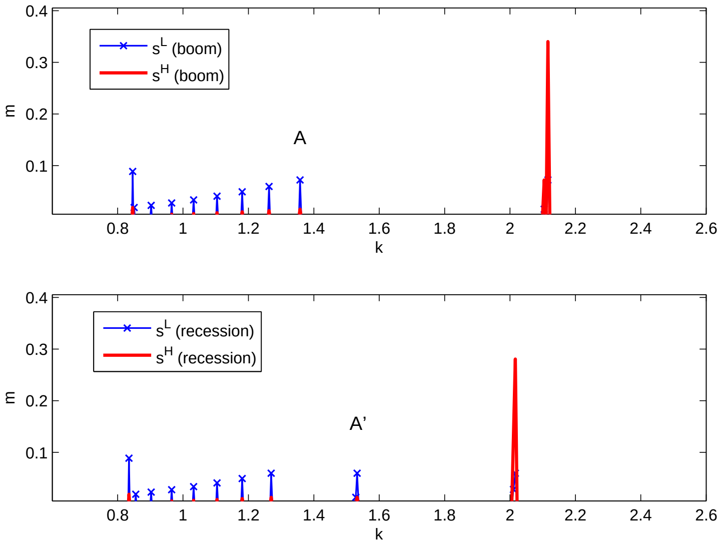


Figure 1.11: Distribution dynamics

Transition from long sequence $z_t = z^H$ to long sequence $z_t = z^L$. Cross-sectional distribution of firms before and immediately after the shock.

1.6.3 Amplification of aggregate TFP

The procyclicality of reallocation generates endogenous movements in aggregate TFP, amplifying the exogenous aggregate productivity shock. Measured TFP, call it Z_t, is the Solow residual that an econometrician would compute by assuming an aggregate production function $Y_t = Z_t K_t^{\alpha} N_t^{\nu}$. A large part of the variation in this variable is due to the exogenous component z_t, while the rest is due to how capital and labor are allocated across the heterogeneous productive units in the economy. In the model, this second component, $TFP_{end} \equiv \log(Z_t) - \log(z_t)$, is of second order, when compared to the exogenous one, so the absolute importance of the allocative component of TFP should not be overemphasized.

However, this component is magnified by endogenous irreversibility, as can be seen in Table 1.5, where both the ratio between its volatility and the volatility of the shock (σ_{TFPend}/σ_z) and the ratio between its volatility and that of output (σ_{TFPend}/σ_Y) increase by more than four times when going from constant to endogenous irreversibility. Importantly, this should be seen as a lower bound for the importance of capital reallocation for aggregate TFP, because the model generates less volatility in both q_t and the reallocation series than we observe in the data.

The amplification mechanism for TFP works as follows: during recessions, reallocation decreases and firms with idiosyncratic productivity s^L are in a sense "too large", which not only implies that capital is less productively used, but also employment is "too high" in these relative less productive firms, as labor demand is an increasing function of a firm's capital stock. Hence the allocation of inputs gets further away from the one that would arise in a model where investment and disinvestment are fully flexible. However, the allocation is always efficient, in the sense that it would coincide with the choice of a planner that faced the same reallocation frictions.

Furthermore, to see how this mechanism is propagated over time, observe again the policy functions illustrated in Figure 1.6. When large firms are hit by a negative idiosyncratic shock, they sell part of their capital once and then they just let their capital depreciate until they become highly productive again. This means that if they sell a small amount of capital in the first period, they will remain "too large" (relative to a model with constant q) for several periods, until they get a positive idiosyncratic shock again. Hence, the negative effects of lower reallocation on aggregate productivity in recessions are propagated over time trough these movements in the distribution of firms. This implication of the model seems consistent with the patterns observed in the current slow recovery of productivity in the UK, that Broadbent (2012) attributes precisely to insufficient capital reallocation.

It is worth emphasizing that the amplification of TFP in the model is an increasing function of the dispersion of idiosyncratic productivity across firms: the more dispersed productivity is, the larger the benefits from reallocation. Hence, the procyclicality of capital reallocation induces larger movements in aggregate productivity when the variance of s_t is higher. For instance, doubling the unconditional variance of s_t leads to $\sigma_{TFPend}/\sigma_z = 0.0278$ and $\sigma_{TFPend}/\sigma_Y = 0.0178$. In this paper, the volatility of idiosyncratic productivity is calibrated to match micro-level investment data following the methodology of Khan and Thomas (2008, 2013). However, in the literature there is no unanimous consensus on this parameter value and in general on the procedure to parametrize the idiosyncratic productivity process. For example, Bloom et al. (2012) estimate time-varying volatilities of firm-level productivity and get values for volatility of up to 12% quarterly in high uncertainty periods. This again suggests that the amplification of aggregate productivity delivered by the present paper is a lower bound for the empirical effect of procyclical capital reallocation on TFP.

Model	σ_{TFPend}/σ_z	σ_{TFPend}/σ_Y
Constant q	0.0034	0.0021
Baseline	0.0134	0.0093

Table 1.5: Amplification of endogenous TFP

1.6.4 Endogenous irreversibility smooths aggregate investment

The previous literature on investment irreversibility has debated whether the observed smoothness of the aggregate investment series can be attributed to irreversibility at the micro level. Bertola and Caballero (1994) affirm this point in a partial equilibrium model and suggest that the fear of not being able to disinvest may act as a smoothing device at the time of investing, making firms more cautious in their investment decisions and generating inaction regions. However, Veracierto (2002) introduces constant partial irreversibility in a general equilibrium model with heterogeneous plants and shows that the properties of aggregate investment are unchanged when moving from totally flexible to totally irreversible investment.[15] This is because the consumption smoothing force makes the interest rate adjust in such a way that aggregate investment has the same desired properties for the representative agent. Furthermore, aggregate shocks are just shifting the inaction region without affecting its size, so that the mass of firms in the inaction region is not changing significantly over time.

The present model reaffirms the original conjecture of Bertola and Caballero (1994) by making irreversibility an equilibrium outcome that moves over the business cycle. Investment becomes more irreversible in recessions, when unproductive firms would like to disinvest by more. This makes investment riskier for firms, hence

[15] Of course, the properties of micro-level investment decisions are very different depending on the level of irreversibility.

making them more cautious at the time of investing. Furthermore, the endogenous prices for investment goods are acting in the direction of smoothing the investment decisions even more: in a recession, when investment falls, used capital is cheaper, which implies that total investment becomes slightly cheaper (Q_t falls) hence dampening the fall in aggregate investment. The opposite happens in booms, when firms want to invest more, but Q_t increases.

As can be seen comparing again Tables 1.3 and 1.4, the volatility of aggregate investment relative to output falls from 5 to 4. Following the previous literature on micro lumpiness and aggregate investment (e.g. Khan and Thomas, 2008), I also report the volatility and autocorrelation of the unfiltered investment/capital ratio. Table 1.6 compares the baseline model with (i) a fully flexible model without irreversibility and (ii) the model with constant q. With endogenous irreversibility, the investment/capital ratio is more persistent and its innovations are less volatile than in the two comparison models, and closer to the US data reported by Khan and Thomas (2008), presented in the last row.

Model	$\sigma_{inn,I/K}$	$\rho_{I/K}$
Frictionless	.010	.675
Constant q	.009	.582
Baseline	.008	.680
Data	.008	.695

Table 1.6: Volatility and autocorrelation of aggregate investment rates

Relatedly, another feature of the model is that following a bad aggregate shock the inaction region becomes larger. This is because the higher wedge between the price of new investment goods and the resale price increases the option value of inaction. This feature of the model resembles the behavior of a model with non-convex adjustment costs and uncertainty shocks (e.g. Bloom et al., 2007). In that case, the freezing of investment activity associated with a widening of the inaction region is driven by exogenous increases in uncertainty. In this model, the same behavior arises in response to first order productivity shocks, via the endogenous reaction of the resale price of capital, without resorting to changes in second order moments. This time-varying wedge may have important policy implications. For example, investing subsidies like those included in the US fiscal stimulus package of 2009 are likely to have procyclical multipliers in this setting, as more firms are in the inaction region in recessions and are thus likely to be less responsive to this kind of stimulus.

1.6.5 Investment-specific shocks

In the baseline version of the model, business cycles were driven by aggregate productivity shocks. However, a large literature emphasize the importance of shocks to the productivity of investment as important drivers of aggregate fluctuations (e.g. Greenwood et al., 2000).

In this subsection, I argue that the main mechanism of endogenous irreversibility and capital reallocation is robust to this different source of business cycles. To see why this is the case, let us first define the modified model. For simplicity, let the aggregate productivity parameter z be constant and equal to 1. Let p_t be the relative price of new investment goods in terms of consumption. Investment goods are produced using the output good as input by a competitive firm. Hence shocks to the marginal cost of production of new investment goods translate into shocks to p_t. Let this shock follow a Markov chain with two values $p_t \in \left\{p^L, p^H\right\}$.

The CES price index for a bundle of new and used investment is now

$$Q_t = [\eta p_t^{1-\epsilon} + (1-\eta)(q_t + \gamma)^{1-\epsilon}]^{\frac{1}{1-\epsilon}} \tag{1.21}$$

where q_t is the relative price of used capital in terms of the consumption good.

Consider a persistent 1% increase from p^L to p^H, illustrated in the first panel of Figure 1.12 (solid black line). When the shock hits the economy, new investment becomes more expensive, inducing a recession as is standard in the literature on investment-specific shock. Importantly, in the present model, the shocks also leads to a fall in capital reallocation (second panel).

To see why this is the case, consider first the buyers on the market for used capital. The increase in the price for new investment goods has two effects of opposite sign on their demand for used capital: on the one hand, total investment is now more expensive (Q_t is higher), leading to a fall in demand for all kinds of investment goods. On the other hand, for a given total investment, firms are willing to partially substitute from new to used investment goods, leading to an increase in demand for used. It turns out that for the calibrated elasticity of substitution between new and used investment, the first effect dominates and demand for used capital falls. Hence q_t (first panel, dashed blue line) increases only gradually with the result that q_t/p_t is below its average for several periods.

For disinvesting firms, given partial irreversibility, when the shock hits it is a bad time to sell assets: they know that the resale price is likely to increase and that they might have to rebuy some capital at a higher Q_t in case they receive a positive idiosyncratic shock in the near future. This implies that also investment-specific shocks induce a procyclical response of capital reallocation.

1.7 Conclusion

This paper shows that the procyclicality of capital reallocation can be rationalized in a model where the resale price of capital is endogenous. According to the sectoral data presented, this price is strongly procyclical, making it harder to reverse past investment decisions during recessions. The model generates this fact as an equilibrium outcome by assuming that new and used investment goods are imperfect substitutes because of partial firm-level capital specificity. Hence, in a recession higher supply of used capital and lower demand lead to a fall in the price, inducing both static and

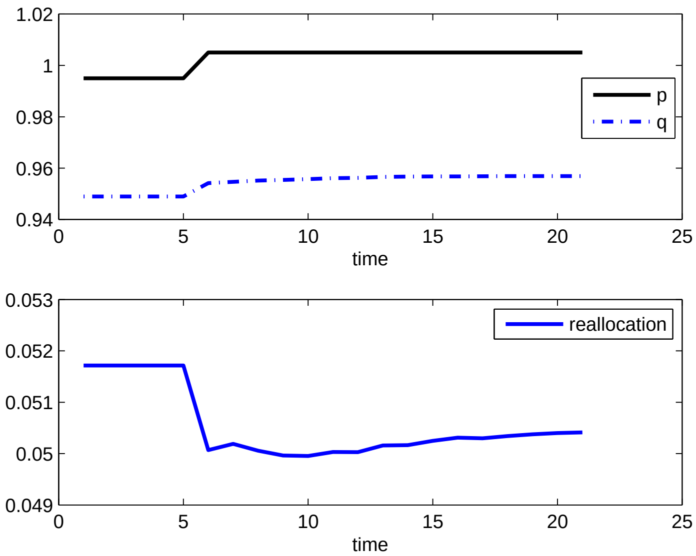


Figure 1.12: Investment-specific shock: investment prices and reallocation

Transition from long sequence $p_t = p^L$ to long sequence $p_t = p^H$. Response of the relative price of used capital (in terms of consumption) and capital reallocation (lower panel).

dynamic effects on investment and reallocation via the equilibrium response of real option values. This mechanism induces endogenous movements in aggregate TFP, because during expansions, when the resale price of capital increases, the allocation of capital and labor gets closer to the one that would arise in a flexible model and vice versa in downturns more capital is operated by unproductive firms.

Endogenous irreversibility is a plausible mechanism behind a smooth aggregate investment series. In this sense the paper provides an explicit microfoundation for what in the aggregate resembles a convex capital adjustment cost. Furthermore, the model generates a countercyclical dispersion of returns, consistently with a growing literature on firm-level uncertainty. Importantly, this result can be fully efficient in a Pareto sense. Previous work has connected a high dispersion of returns in recessions with the malfunctioning of credit markets, hence providing one rationale for expansionary credit policies. This paper suggests that part of this increased dispersion, which has been emphasized both in academic and policy work during the recent recession, may be unrelated to credit conditions and attributable to partial capital specificity.

While this model assumes perfect capital markets, it is clear that financing constraints may play an important role in shaping investment dynamics. Introducing a collateral constraint that ties the borrowing capacity to the resale value of a firm's capital is likely to further amplify the mechanism described in the paper. This has important implications for the question on the source of business cycles. Previous work based on a constant value of collateral has suggested that procyclical capital reallocation is evidence in favor of exogenous credit shocks (Cui, 2014). However, the value of a firm's collateral depends on the resale price of its assets. Hence, an extension of the present model with collateral constraints could potentially generate an endogenous tightening of collateral constraints after a negative TFP shock, reconciling both the cyclicality of reallocation and that of credit availability with standard productivity shocks.

Furthermore, US plant level data suggest that while entry is strongly procyclical, exit is almost acyclical (Lee and Mukoyama, 2013). This evidence on exit is to some extent a puzzle for models with productivity shocks where the exit decision is driven by a fixed cost of production denominated in units of the output good. In such models, after a bad aggregate TFP shock, more firms optimally decide to liquidate their capital and exit. Endogenous irreversibility seems to be a promising explanation for this puzzle. In a recession, on the one hand the value of staying in business falls, so that more firms would like to exit, but on the other hand also the value of exit falls, as it depends of the resale price of capital, so that overall the incentive to liquidate is dampened.

Finally, this paper provides a natural framework to analyze movements in the utilization rate of capital both in the aggregate and at the micro-level. A large firm hit by a negative profitability shock can choose whether to reallocate its assets or to keep them idle for some time, hoping for an improvement in business conditions. The previous literature on heterogeneous firms and capacity utilization has imposed restrictions on the possibility to sell assets after the realization of idiosyncratic productivity shocks in order to justify the choice by unproductive firms to keep some idle capital (e.g. Hansen and Prescott, 2005, Sustek, 2011).

In the context of a model with endogenous irreversibility, no such assumptions are necessary. A version of the model that includes endogenous capacity choice (see chapter 2) yields a natural solution to the question of whether to sell assets or keep them idle. Depending on aggregate conditions, the resale price may be high enough to induce reallocation or low enough to make it optimal to keep capital idle. The key mechanism works through equilibrium changes in two option values: the put option value to resell and the value to keep capital idle and save on production costs. The price of used capital responds to aggregate shocks and makes one or the other option more valuable at different points in the business cycle. Hence cyclical movements in output, reallocation and utilization can be jointly explained in a model of endogenous irreversibility.

Appendix A: Additional empirical evidence and data sources

Further evidence on capital reallocation: UK sales of used equipment (Figure 1.13) and global sales of used commercial ships (Figure 1.14).

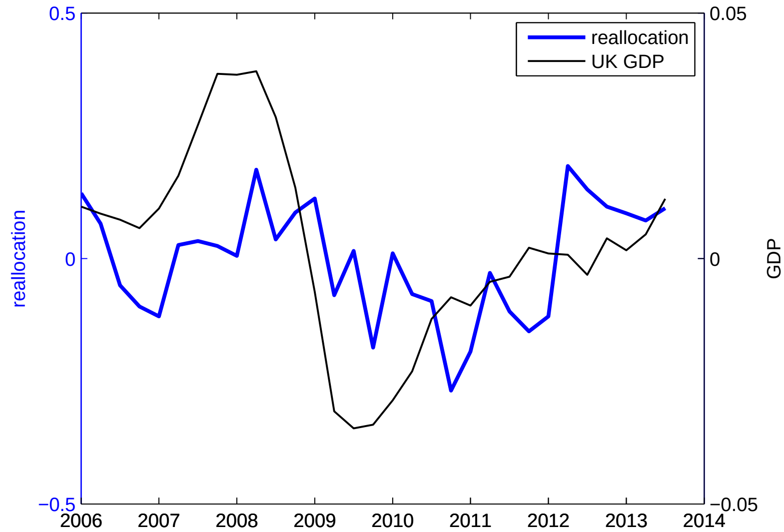


Figure 1.13: Capital reallocation in UK during the Great Recession

Cyclical components of quarterly sales of used equipment and UK real GDP, deflated using the GDP deflator.

Further evidence on the price of used capital over the business cycle:

Ships I gather price series for new and used ships from the mid-90's onwards. It is interesting to observe that prices and quantities traded fall contemporaneously in 2008, and that the price index of used ships is more volatile than the price index of new ships (Figures 1.14 and 1.15). Similarly to what discussed in section 1.2 for the aircraft sector, also in the case of ships the resale price of more specific models in terms of possible routes (e.g. the very heavy and large Capesize bulk carrier) grows more strongly in the period 2006-2008 and then falls by a larger fraction towards the

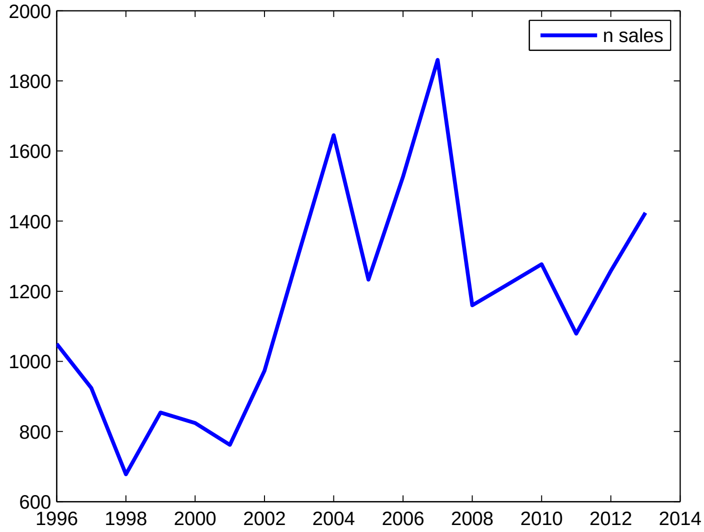


Figure 1.14: Ships: number of second-hand sales

Global yearly sales of second-hand commercial ships. Source: Clarkson

end of 2008 than that of less specific ones (e.g. the more flexible and small Handysize bulk carrier). This is shown in Figure 1.16.

Vehicles In the case of vehicles and trucks one can compare two separate separate CPI series, one for new (CPI new vehicles) and one for used (CPI used cars and trucks). Figure 1.17 shows the cyclical components (HP-filtered) of the CPI for used cars and trucks and the CPI for new vehicles, both relative to the total CPI including all items. It emerges that the price of used vehicles is much more volatile and more procyclical than that of new ones, which is actually acyclical (their correlations with GDP are 0.41 and -0.09 respectively). The volatility of prices of used vehicles is smaller than that of aircraft and ships, arguably because vehicles are a less specific type of asset. Hence, this difference in volatilities is broadly consistent with a theory based on capital specificity.

Construction equipment Edgerton (2011) constructs an index of the price of used construction machinery by collecting data on auctions where this equipment is reallocated across US construction firms. These data are illustrated in Figure 1.18 and show that the price of used construction equipment fell by more than the corresponding PPI (construction machinery) both in the 2001 and in the 2009 recession, and is in general significantly more volatile. In 2009 the index of used construction equipment is more than 15% below trend, while the corresponding PPI of new construction machinery is slightly above trend.

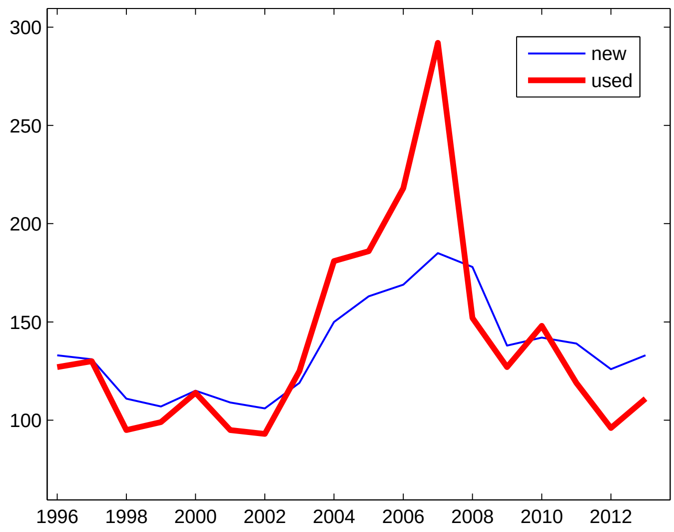


Figure 1.15: Ships: price indices of new and used

Price indices of new and second-hand commercial ships. Yearly frequency.

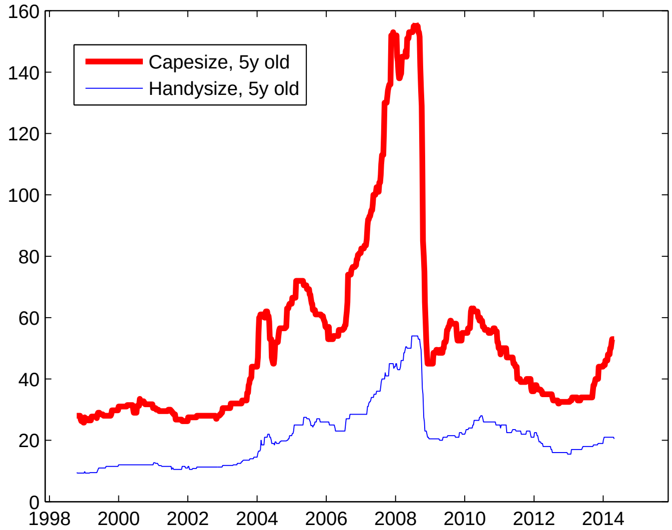


Figure 1.16: Ships: price of used Capesize and used Handysize

Prices in million \$ of second-hand 5 year-old Capesize (more specific) and Handysize (less specific). Weekly frequency: estimated values based on actual transactions and shipping market information.

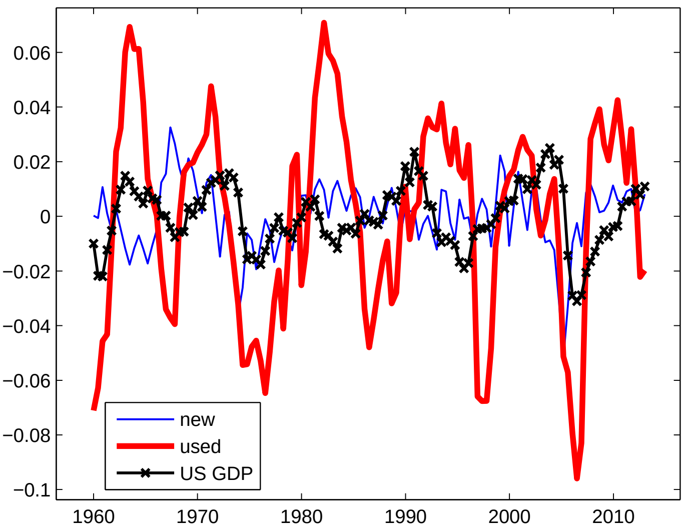


Figure 1.17: Vehicle prices and GDP

Cyclical components of CPI new vehicles, CPI used cars and trucks and US real GDP. Quarterly frequency. Both CPI series are divided by CPI All items.

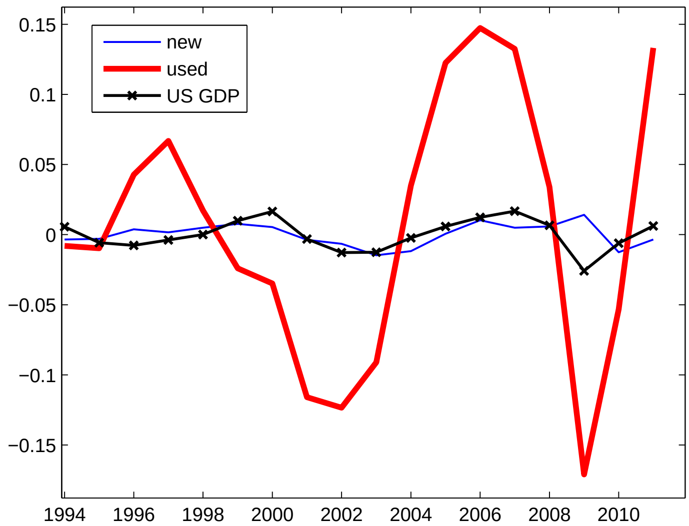


Figure 1.18: Construction equipment prices and GDP

Cyclical components of construction equipment PPI, price index of used construction equipment (Edgerton, 2011) and US real GDP. Yearly frequency.

Data sources Data on capital reallocation in the US come from the Compustat dataset and have been kindly made available by Andrea Eisfeldt on her personal webpage. Data on sales of used equipment in the UK come from the Office for National Statistics (ONS) Survey of Capital Expenditures. Data on aircraft prices are compiled by Aircraft Values. Data on commercial ships are compiled by Clarkson and VesselsValue. The price index for used commercial equipment has been constructed by Edgerton (2011) using auction prices. Data on US GDP, GDP Deflator, CPI, CPI for new and used vehicles, as well as PPI for the construction sector come from the US Bureau of Economic Analysis and the US Bureau of Labor Statistics. Data on UK GDP and GDP deflator come from the ONS.

Appendix B: Proof of Proposition 1

(i) By equating (1.5) and (1.6), the market-clearing condition for used capital can be written as follows:

$$G(q,z,\epsilon) \equiv \theta(q)\int_{s^I}\left[\left(\frac{\alpha z s}{Q(q)}\right)^{\frac{1}{1-\alpha}} - k_0\right]dF(s) - \int^{s^D}\left[k_0 - \left(\frac{\alpha z s}{q}\right)^{\frac{1}{1-\alpha}}\right]dF(s) = 0. \tag{1.22}$$

where $s^I = \frac{Q(q)}{\alpha z k_0^{\alpha-1}}$, $s^D = \frac{q}{\alpha z k_0^{\alpha-1}}$, $Q(q) = [\eta + (1-\eta)(q+\gamma)^{1-\epsilon}]^{\frac{1}{1-\epsilon}}$, $\theta(q) = (\frac{q+\gamma}{Q(q)})^{-\epsilon}(1-\eta)$ is the ratio of used investment to total investment for investing firms and I have left implicit the dependence of θ, q and Q on ϵ. Equation (1.22) defines the market-clearing price q as an implicit function of the aggregate productivity parameter z and the elasticity of substitution between new and used capital ϵ. We can obtain the derivative of q with respect to z by applying the Implicit Function Theorem to function G and we get[16]

$$\frac{dq}{dz} = -\frac{G_z}{G_q} \tag{1.23}$$

with

$$G_z = \frac{\theta}{(1-\alpha)z}\int_{s^I}\left(\frac{\alpha z s}{Q}\right)^{\frac{1}{1-\alpha}}dF(s) + \frac{1}{(1-\alpha)z}\int^{s^D}\left(\frac{\alpha z s}{q}\right)^{\frac{1}{1-\alpha}}dF(s)$$

and

$$G_q = \theta_q\int_{s^I}\left(\frac{\alpha z s}{Q}\right)^{\frac{1}{1-\alpha}}dF(s) - \frac{\theta Q_q}{(1-\alpha)Q}\int_{s^I}\left(\frac{\alpha z s}{Q}\right)^{\frac{1}{1-\alpha}}dF(s) - \frac{1}{(1-\alpha)q}\int^{s^D}\left(\frac{\alpha z s}{q}\right)^{\frac{1}{1-\alpha}}dF(s)$$

Note that in applying Leibniz rule to derive this expression we do not need to worry about the derivatives of the end points s^D and s^I because by their definition, the respective integrands are equal to zero when evaluated at these points.

Hence $\phi_{q,z}(\epsilon) \equiv \frac{dq}{dz}\frac{z}{q}$, the elasticity of q with respect to z, is

$$\phi_{q,z}(\epsilon) = -\frac{F_z}{F_q} = \frac{\frac{\theta}{(1-\alpha)}\int_{s^I}\left(\frac{\alpha z s}{Q}\right)^{\frac{1}{1-\alpha}}dF(s) + \frac{1}{(1-\alpha)}\int^{s^D}\left(\frac{\alpha z s}{q}\right)^{\frac{1}{1-\alpha}}dF(s)}{\theta_q q\int_{s^I}\left(\frac{\alpha z s}{Q}\right)^{\frac{1}{1-\alpha}}dF(s) + \frac{\theta Q_q q}{(1-\alpha)Q}\int_{s^I}\left(\frac{\alpha z s}{Q}\right)^{\frac{1}{1-\alpha}}dF(s) + \frac{1}{(1-\alpha)}\int^{s^D}\left(\frac{\alpha z s}{q}\right)^{\frac{1}{1-\alpha}}dF(s)} \tag{1.24}$$

[16]Notation: Call f_x be the partial derivative of function f with respect to argument x.

Now, note that when $\epsilon = 0$ (Leontief investment technology), the share of used capital to total investment becomes $\theta = 1 - \eta$, so that $\theta_q = 0$, while the price index becomes $Q = \eta + (1-\eta)(q+\gamma)$, so that we get $Q_q = 1 - \eta$. Hence we can write

$$\phi_{q,z}(0) = \frac{(1-\eta)\int_{s^I}\left(\frac{\alpha z s}{Q}\right)^{\frac{1}{1-\alpha}} dF(s) + \int^{s^D}\left(\frac{\alpha z s}{q}\right)^{\frac{1}{1-\alpha}} dF(s)}{(1-\eta)^2\frac{q}{\eta+(1-\eta)(q+\gamma)}\int_{s^I}\left(\frac{\alpha z s}{Q}\right)^{\frac{1}{1-\alpha}} dF(s) + \int^{s^D}\left(\frac{\alpha z s}{q}\right)^{\frac{1}{1-\alpha}} dF(s)} \tag{1.25}$$

and this establishes that $\phi_{q,z}(0) > 1$ as $q < 1 \Rightarrow (1-\eta)\frac{q}{\eta+(1-\eta)(q+\gamma)} < 1$. Standard arguments can be used to show that $\phi_{q,z}$ is continuous.

(ii) It suffices to observe that the equilibrium supply of used capital S^*_{used}, i.e. is total reallocation, is a decreasing function of $\frac{z}{q}$ (as above, we can disregard the derivative of s^D as the integrand is zero when evaluated at s^D):

$$S^*_{used} = \int^{s^D}\left[k_0 - \left(\frac{\alpha z s}{q}\right)^{\frac{1}{1-\alpha}}\right] dF(s) \tag{1.26}$$

Hence, the sign of its derivative with respect to z is the sign of $\phi_{q,z} - 1$. This establishes that in the limit for sufficiently low elasticity of substitution between new and used capital, reallocation is increasing in z, i.e. "procyclical". □

Appendix C: Accuracy and robustness

Figure 1.19 illustrates the accuracy of the solution by showing the simulated series of aggregate capital (solid red line) and a forecast series constructed using the estimated coefficients of the law of motion and iterating on the forecast (blue crosses), as suggested by den Haan (2010). The R^2 of the regression of log capital on constant, its lag and aggregate productivity is .9993.

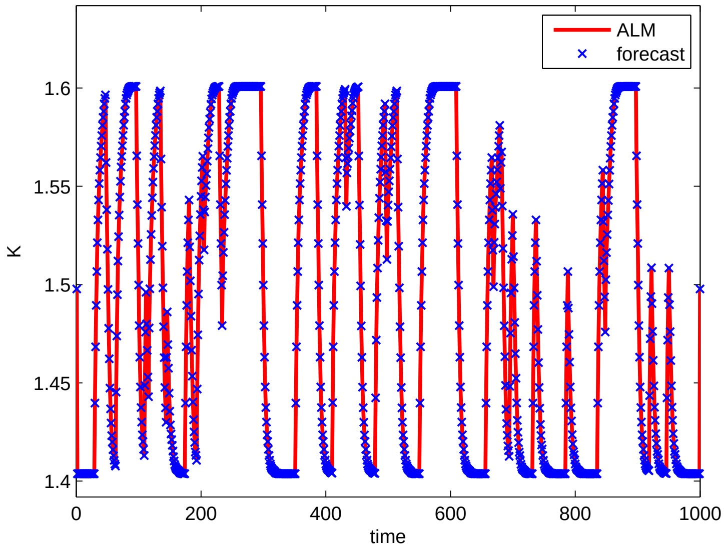


Figure 1.19: Actual law of motion and its forecast

Robustness exercises: business-cycle statistics for different values of ϵ and HP-filter smoothing parameter.

Statistic	Y	C	I	K	N	q	reall
mean	.595	.495	.099	1.518	.335	.948	.042
$\sigma(.)/\sigma(Y)$	1.34	.470	4.25	.257	.539	.192	4.25
corr(.,Y)	1	.838	.950	-.281	.989	.952	.956

Table 1.7: Business-cycle statistics: $\epsilon = 1$

Statistic	Y	C	I	K	N	q	reall
mean	.588	.489	.097	1.490	.336	.959	.055
$\sigma(.)/\sigma(Y)$	(1.40)	.442	4.242	.266	.558	.172	2.617
corr(.,Y)	1	.905	.936	-.277	.984	.970	.960

Table 1.8: Business-cycle statistics: $\epsilon = 10$

Statistic	Y	C	I	K	N	q	reall
$\sigma(.)/\sigma(Y)$	(2.05)	.539	3.786	.403	.532	.183	3.011
corr(.,Y)	1	.914	.930	.048	.963	.916	.9585

Table 1.9: Business-cycle statistics: baseline model, HP smoothing parameter = 100

Statistic	Y	C	I	K	N	q	reall
$\sigma(.)/\sigma(Y)$	(2.25)	.545	4.398	.407	.626	0	1.055
corr(.,Y)	1	.828	.887	.114	.966	0	-.592

Table 1.10: Business-cycle statistics: constant q, HP smoothing parameter = 100

Chapter 2

Sell It or Keep It Idle? Partial Irreversibility and Capital Utilization

2.1 Introduction

2.1.1 Motivation

When is it optimal for unprofitable firms to keep part of their capital idle instead of downsizing? In this chapter, I develop a model of partially irreversible investment with endogenous capacity utilization to address this question. First, I discuss the relation between partial irreversibility and capital utilization for an individual firm and then I embed this problem in an equilibrium model with heterogeneous firms, aggregate shocks and endogenous irreversibility.

Capital utilization has been recognized as an important amplification channel for productivity shocks since the early RBC literature (e.g. Kydland and Prescott, 1988). In a one-sector model of the economy, if the utilization rate of capital is chosen purely based on static considerations such as the wage rate or energy cost, a degree of investment irreversibility is necessary to justify why capital is kept idle in a constrained-efficient equilibrium. Absent any irreversibility, in a representative-firm model, idle capital could be used to increase consumption. In presence of heterogeneous firms with different levels of productivity, unproductive firms could also sell idle capital to more productive firms that may be willing to operate it.

The previous literature on capital utilization has either assumed full irreversibility of capital (Pindyck, 1988), implying that downsizing in response to negative profitability shock is impossible, or time-to-build (Greenwood et al., 1988, Kydland and Prescott, 1988), implying that the current level of capital was chosen in the previous period and hence capital cannot be reallocated back to the consumer before production.

The first contribution of this paper is to consider any level of partial irreversibility. This allows me to clarify the relation between irreversibility and utilization: I derive a necessary condition on the degree of irreversibility that makes underutilization an efficient outcome in response to negative profitability shocks. In a partial equilibrium setting, I allow for investment and reallocation (downsizing) before production and argue that underutilization of capital still arises optimally in response to negative transitory shocks when the resale price of capital is below a threshold that depends only on the discount factor and the depreciation rate.

Idle capital has an option value: the firm may benefit from it in the future, should stochastic profitability improve. When a sufficiently large negative transitory shock hits, I show that underutilization of capital becomes optimal as it allows the firm to save on the cost of downsizing and then re-expanding if profitability improves.

After deriving these analytical results, I solve the model numerically for different values of the volatility and persistence of the profitability shock. Three testable implications of the model emerge. First, only sufficiently large negative shocks lead to underutilization. As a consequence, the unconditional mean of the utilization rate is decreasing in the volatility of the shock. Second, the average level of capital is increasing in this volatility. This is because the underutilization option insulates profits from negative shocks, hence decreasing the downside risk of investment. Third, while transitory shocks lead to underutilization, persistent shocks of the same size call for downsizing. If the shock is sufficiently persistent, it is unlikely that the firm will need the current level of capital in the foreseeable future. Hence the option value of idle capital is low and it is optimal to downsize before production.

Finally, I consider an equilibrium model of asset allocation and utilization with heterogeneous firms hit by both aggregate and idiosyncratic shocks. Because of firm-level capital specificity, used investment goods are imperfect substitutes for new goods and the resale price of capital becomes endogenous. In good times, this price is high and all firms hit by negative shocks choose to reallocate their assets and downsize. In bad times this price falls. Firms with very low productivity sell their assets, while firms with an intermediate level of productivity optimally choose to hold on to their capital and keep part of it idle.

The macroeconomic literature has generally treated utilization and allocation of capital in isolation (see next subsection for details on the literature). This paper is a step towards building a joint theory of capital reallocation and utilization. I show that these decisions are inherently interconnected through the dynamics of the resale price of capital. In future work I will study how these two channels interact to determine the dynamics of output and aggregate productivity in a quantitative model of the business cycle.

2.1.2 Related literature

A starting point for the work presented here is the seminal paper on real options by Pindyck (1988). The author develops a partial equilibrium model of fully irreversible investment with endogenous choice of capacity utilization. A key result is that because of discounting, the intratemporally optimal capacity (which does not take into account the opportunity cost of capital) is generally larger than the actual size of the firm. Hence the firm typically operates at full capacity, unless it is hit by a large negative shock. Relatedly, McDonald and Siegel (1986) build a model with discrete and irreversible investment decisions with a temporary shut-down option.

With respect to this literature, this paper contributes by generalizing the setup in a number of dimensions. First, I allow for partial investment irreversibility. Capital can be purchased and then resold, though at a lower price.[1] I derive both analytical and numerical results that show that depending on the level of the resale price, unprofitable firms choose either to sell part of their capital or to keep it idle. Second, I consider shocks of different persistence and discuss how this affects the optimal utilization choice, while the real options literature focuses only on unit root processes. Finally, I embed the firm's problem in a model of industry equilibrium with imperfect substitutability between new and used investment goods, where the resale price of capital is determined in equilibrium. Firms face both aggregate and idiosyncratic shocks. When the economy enters the bad state, meaning that a fraction of firms become very unproductive and need to downsize, the resale price of capital falls and firms with intermediate levels of productivity find it optimal to hold their capital and keep part of it idle.

A large literature deals with capital utilization in representative-firm RBC models. A key reference is Greenwood et al. (1988), who endogenize utilization by assuming that capital depreciates faster, the more intensively it is used in production. This constitutes a key channel in the transmission of investment specific shocks. In this paper, I abstract from this depreciation-in-use assumption in order to simplify the analysis by making the utilization choice fully intratemporal. On the other hand, I allow for firm heterogeneity and reallocation of capital before production. Other papers on utilization in RBC models include Kydland and Prescott (1988) and Hansen and Prescott (2005).

Idiosyncratic firm shocks generate strong incentives to reallocate capital instead

[1] Full irreversibility arises as the special case with resale price equal to zero.

of keeping it idle. However, few papers study capital utilization in the context of a model with production heterogeneity and all of them make strong irreversibility assumptions, that is assumptions on the possibility to reallocate capital. Both Cooley et al. (1995) and Sustek (2011) build economies with production heterogeneity and capacity choice. Plants observe their idiosyncratic productivity and then choose their capacity utilization level. Low-productivity plants keep some capital idle. Importantly, in both papers underutilization happens only because reallocation cannot take place after the realization of the idiosyncratic shock. In absence of such restriction, idle capital would flow to the more productive firms. In the present paper, I allow both reallocation and utilization decisions to take place after the shocks are realized and show that less-than-full utilization can still be optimal if the equilibrium price of used capital is sufficiently low. More in general, this paper contributes to the growing literature on business cycle models with production heterogeneity (e.g. Khan and Thomas, 2008, 2013). This literature has so far abstracted from capital utilization.

On the empirical side, Gavazza (2011b) uses data on commercial aircraft to study the relationships between trading friction in resale markets, asset allocation and capacity utilization, showing that larger frictions in the resale of capital induce lower utilization rates. Gavazza (2011a) introduces leasing of capital in a model of costly capital reallocation and shows that because of irreversibility, following negative profitability shocks carriers return their leased aircraft, but retain and underutilize their owned aircraft. This leads to a higher utilization rate for leased capital. An empirical analysis and a test of the implications of the present model is left for future work. Furthermore, it would be interesting to study the implications of adding the option to lease capital in the present model.

2.2 Individual firm's problem

In this section I introduce the problem of a firm who produces output using capital with stochastic productivity.[2] Investment is partially irreversible, as only a fraction of its value can be recovered if capital is resold. Production costs depend on the amount of capital effectively used in production, which is at most equal to the total capital owned by the firm. This allows me to endogenize capital utilization.

I will discuss under what conditions on the resale price of capital and under what properties of the profitability shocks underutilization arises as the optimal response to low profitability.

2.2.1 Model

An infinitely-lived risk-neutral firm produces output using capital. Its period profit function (gross of investment) is $\pi(x, s)$ where x is capital used in production and s is a non-negative productivity shock. I make the following assumptions on the profit function:

- A1) π is at least twice differentiable with respect to x and once with respect to s
- A2) $\pi(0, s) = 0, \ \forall\ s$
- A3) $\pi_x(0, s) > 0, \ \forall\ s$
- A4) $\pi_{xx}(x, s) < 0, \ \forall\ (x, s)$
- A5) $\pi_x(x, s) < 0, \ for\ x\ large\ enough, \ \forall\ s$
- A6) $\pi_s(x, s) > 0, \ \forall\ (x, s)$
- A7) $\pi_{xs}(x, s) > 0, \ \forall\ (x, s)$.

[2] Or, equivalently, stochastic demand for the output good, that is, shocks to the output price. Hence I will use productivity and profitability interchangeably in this partial equilibrium model.

Assumptions A1-A4 and A6-A7 are entirely standard and are satisfied in most existing models with productivity shocks and decreasing returns. Assumption A5 is less standard. It states that the profit function becomes decreasing (in capital) for a large enough level of capital. This assumption allows me to obtain a well-defined intratemporally optimal production size, that is the level of utilization that maximizes $\pi(., s)$.[3] In the numerical examples, I will work with the following parametrization: $\pi(k, s) = sk^{\alpha} - ck$ with $\alpha \in (0, 1)$, $c > 0$. One can easily microfound the linear cost function ck by assuming a complementary input in production with constant marginal cost c (e.g. energy), as in Burnside et al. (1995) and Atkeson and Kehoe (1999).[4]

The firm can choose its level of capital before production. This assumption allows me to isolate the effect of partial irreversibility on utilization decisions: the firm is always free to change its level of capital after the shock hits, if it wants to do so, before producing. Hence underutilization is not the direct consequence of an unchangeable "wrong" size inherited from past decisions.

Capital depreciates at rate $\delta \in [0, 1]$ per period. When investment is positive, the relative price of capital in terms of the output good is 1. When investment is negative, however, the firm obtains a price $q \in [0, 1]$ for its disinvestment. In this section, I treat q as an exogenous parameter and discuss how its level affects the optimal investment and utilization choices. In the next section, I endogenize q in an equilibrium model with heterogeneous firms.

Let $\beta \in (0, 1)$ be the discount factor and let s follow a Markov process with transition probabilities given by $G(s'|s)$. The value function of the firm can be expressed as follows.

$$V(k, s) = \max \left\{ V^{i}(k, s), V^{d}(k, s) \right\} \tag{2.1}$$

[3]Pindyck (1988) assumes a quadratic production cost, which is a special case of A5.

[4]To see this, assume that the production function has capital and energy as inputs and is defined by $f(k, e) = s \ (min\{k, e\})^{\alpha}$. The constant marginal cost of energy is c. Clearly, optimality requires $k = e$ which gives the profit function in the main text.

where V^i and V^d are the values conditional on positive (or zero) and negative investment respectively.

$$V^i(k,s) = \max_{k' \geq (1-\delta)k} \left\{ (1-\delta)\,k - k' + \max_{x \leq k'} \{\pi(x,s)\} + \beta \mathrm{E}\left[V(k',s')|s\right] \right\} \tag{2.2}$$

and

$$V^d(k,s) = \max_{k' \leq (1-\delta)k} \left\{ q\left[(1-\delta)\,k - k'\right] + \max_{x \leq k'} \{\pi(x,s)\} + \beta \mathrm{E}\left[V(k',s')|s\right] \right\} \tag{2.3}$$

Notice that the firm can choose its level of capital k' and then can choose to produce using any level of capacity x below and up to k'. Both decisions are based on the same information set, that is the history of s up to time t. Hence, at the moment of investing/disinvesting, the firm can anticipate how much capital will be put in production as a function of k' and s. This is given by the optimal production capacity:

$$X(k',s) = \min\{k', \tilde{x}(s)\} \tag{2.4}$$

where $\tilde{x}(s)$ solves $\pi_x(\tilde{x}(s),s) = 0$. In words, the firm computes the intratemporally optimal size $\tilde{x}(s)$ by maximizing period profits. If this quantity is less than the capital level k', there will be underutilization. If it is greater or equal to k', there will be full utilization.

Accordingly, the value functions can be rewritten as follows.

$$V^i(k,s) = \max_{k' \geq (1-\delta)k} \left\{ (1-\delta)\,k - k' + \pi\left(X(k',s),s\right) + \beta \mathrm{E}\left[V(k',s')|s\right] \right\} \tag{2.5}$$

and

$$V^d(k,s) = \max_{k' \leq (1-\delta)k} \left\{ q\left[(1-\delta)\,k - k'\right] \right] + \pi\left(X(k',s),s\right) + \beta \mathrm{E}\left[V(k',s')|s\right] \right\} \tag{2.6}$$

The solution to this dynamic program is fully characterized by policy functions for capital level K such that $k' = K(k,s)$ and production capacity X such that $x = X(k',s)$.

If $q < 1$, the wedge between the cost of positive investment and the resale price generates inaction areas, where firms optimally let their capital depreciate without doing any investment or disinvestment. It is necessarily the case that the capital level that solves (2.5) without the inequality constraint of positive investment, call it $K^i(k,s)$, is strictly less than the capital level that solves (2.6) without the inequality constraint of positive disinvestment, call it $K^d(k,s)$. It follows that the policy function for future capital will be:

$$K(k,s)=\begin{cases}K^i(k,s), & k\leq K^i(k,s)/\left(1-\delta\right)\\ \left(1-\delta\right)k, & K^i(k,s)/\left(1-\delta\right)<k\leq K^d(k,s)/\left(1-\delta\right)\\ K^d(k,s), & k>K^d(k,s)/\left(1-\delta\right).\end{cases} \tag{2.7}$$

In this subsection, I will establish some results on when underutilization is optimal. Before doing this formally, observe that in absence of partial irreversibility, that is when $q=1$, the firm will never want to underutilize capital. In fact, in this case the firm could always sell any excess capacity and then rebuy it in the next period at the same price in case it is optimal to expand again. Because of discounting and depreciation, this dominates holding idle extra-capacity. Moreover, again because of discounting and depreciation, the intertemporally optimal level of capital is in general less than the level that maximizes period profits. Without partial irreversibility, the problem becomes fully differentiable and the first order condition for capital level is

$$1-\beta\left(1-\delta\right)=\pi_x\left(K(k,s),s\right) \tag{2.8}$$

which can be solved for the policy function. As the left hand side of (2.8) is positive, we have $K(k,s)<\tilde{x}(s)$. From a purely static point of view, the firm would like to maximize π. However, from a dynamic point of view, the cost of capital has to include the opportunity cost due to discounting as well as depreciation.[5] Hence the optimal capital level will be less than the statically optimal capacity, leading to full utilization in all periods and states.

[5]This generalizes a similar point made by Pindyck (1988) in a model without depreciation.

2.2.2 Analytical results

I will now show that there is an upper bound on the resale price of capital q such that below this level of irreversibility it is never optimal to keep capital idle. After I have established this, I will show that with sufficient irreversibility idle capital can arise as optimal in response to a transitory profitability shock.

Proposition 2. *If $q > \beta(1-\delta)$, then full utilization is always optimal, that is $X(K(k,s),s) = K(k,s)$, $\forall\ (k,s)$.*

Proof I will first assume that the policy function involves underutilization for a certain state (k,s) and show that this leads to a contradiction. Let $\tilde{k} \equiv K(k,s)$ and $x \equiv X(\tilde{k},s)$. Assume $x < \tilde{k}$. It is convenient to define the following function:

$$W(k',s) \equiv \max_{y \leq k'} \left\{ \pi(y,s) + \beta \mathrm{E}\left[V(k',s)|s\right] \right\}.$$

Consider the case $\tilde{k} < (1-\delta)k$ (the opposite case will follow a fortiori). The assumed policy function implies

$$V(k,s) = q\left[(1-\delta)\,k - \tilde{k}\right] + \pi(x,s) + \beta\mathrm{E}\left[\max\left\{V^i(\tilde{k},s'), V^d(\tilde{k},s')\right\}|s\right] \tag{2.9}$$

with

$$V^i(\tilde{k},s') = -\left[K(\tilde{k},s') - (1-\delta)\,\tilde{k}\right] + W(K(\tilde{k},s'),s') \tag{2.10}$$

for values of s' that imply non-negative investment and

$$V^d(\tilde{k},s') = -q\left[K(\tilde{k},s') - (1-\delta)\,\tilde{k}\right] + W(K(\tilde{k},s'),s') \tag{2.11}$$

for values of s' that imply negative investment.

Consider now an alternative strategy, that is selling the additional unused amount of capital $(\tilde{k} - x)$ in the first period and then following the same policy function from the second period onwards. This strategy gives full utilization in the first period and I will use letter F to characterize its outcomes. We have $K^F(k,s) = X(k,s)$ and let $V^F(k,s)$ be the associated value of the objective function.

Let $\Delta(k, s, s')$ be the difference between the value associated with strategy F and that associated with the assumed policy as a function of the initial state and the realization of the second period shock. We can partition the set of possible realization of s' as follows. There are

- (i) Values of s' that call for non-negative investment in the second period under both the assumed policy and strategy F, that is $K(\tilde{k}, s') \geq (1-\delta)\,\tilde{k} > (1-\delta)\,x$
- (ii) Values of s' that call for negative investment in the second period under both the assumed policy and strategy F, that is$(1-\delta)\,\tilde{k} > (1-\delta)\,x \geq K(\tilde{k}, s')$
- (iii) Values of s' that call for non-negative investment in the second period under strategy F, but for negative investment under the assumed policy function $(1-\delta)\,\tilde{k} > K(\tilde{k}, s') \geq (1-\delta)\,x$

Let us begin with case (i). We have

$$\Delta(k, s, s') = [q - \beta\,(1-\delta)]\left(\tilde{k} - x\right) > 0 \tag{2.12}$$

For case (ii), we obtain

$$\Delta(k, s, s') = q\,[1 - \beta\,(1-\delta)]\left(\tilde{k} - x\right) > 0 \tag{2.13}$$

For case (iii), we get

$$\Delta(k, s, s') = q\,[1 - \beta\,(1-\delta)]\left(\tilde{k} - x\right) - \beta\,(1-q)\left[K(\tilde{k}, s) - (1-\delta)\,x\right] \tag{2.14}$$

and because

$$\beta\,(1-q)\left[K(\tilde{k}, s) - (1-\delta)\,x\right] \leq \beta\,(1-q)\left[(1-\delta)\,\tilde{k} - (1-\delta)\,x\right]$$

we have

$$\Delta(k, s, s') \geq q\,[1 - \beta\,(1-\delta)]\left(\tilde{k} - x\right) > 0$$

Hence, the alternative strategy gives a higher payoff for all realizations of the second period shock (and then follows the same policy from then onwards). This implies that the difference between the value of the strategy denoted by F and the value function, which is equal to $\mathrm{E}\{\Delta(k,s,s')|s\}$, is positive. This contradicts the fact that K is the policy function for the problem.

Finally, it is easy to see that the same arguments hold a fortiori for the case of non-negative investment in the first period, that is $\tilde{k} \geq (1-\delta)\,k$. $\square$

Proposition 2 establishes that there is a minimum level of irreversibility below which idle capital is never optimal. This level is fully determined by only the discount factor and the depreciation rate. Intuitively, when $q > \beta(1-\delta)$, the expected return on capital is sufficiently low that is optimal to respond to negative shocks by downsizing instead of holding idle capital. A corollary of Proposition 2 is that in the special case of full depreciation $\delta = 1$ it is never optimal to underutilize capital. This is fairly intuitive, as idle capital would be anyway destroyed before next period's production and hence has no option value.

Proposition 2 states a necessary condition for idle capital, but does not say when underutilization is optimal. In order to gain intuition on this question, I will now consider a simplified version of the model under the case $q < \beta(1-\delta)$. This will allow me to show that for sufficiently large negative transitory profitability shocks it is optimal to underutilize capital.

The only modification to the model presented above is the following: productivity s is stochastic in the first period and after that is deterministic and constant. More specifically, I assume that s_0 is drawn from a distribution $G(s)$ on $[0,\infty)$ with $\mathrm{E}\{s\} = 1$. For $t = 1, 2, ...$ we have $s_t = 1$. This assumption defines an extreme case of transitory shock. Let k^{SS} be the steady-state value of capital for this economy, that is the capital level that solves $1-\beta\,(1-\delta) = \pi_x(k^{SS}, 1)$ and assume the initial value $k_{-1} = k^{SS}$. To establish the optimality of idle capital, I will consider the firm's behavior for low productivity shocks in period 0, that is $s_0 < 1$. In this simple model the first period capital choice is a simple differentiable problem and I can prove the following.

Proposition 3. *Let $q < \beta(1-\delta)$ and define s^U as the solution to $\pi_x((1-\delta)k^{SS}, s^U) = 0$. In response to sufficiently low productivity shocks, such that $s_0 < s^U$, it is optimal to underutilize capital, that is $x_0 < k_0$.*

Proof I will first impose full utilization and compute the optimal behavior under this restriction, defined by functions $K_0^F(s_0)$ and $X_0^F(K_0^F(s_0), s_0) = K_0^F(s_0)$, with associated indirect utility $V^F(s_0)$. Then, I will introduce endogenous utilization and show that the associated policy functions $K_0^U(s_0)$ and $X_0(K_0^U(s_0), s_0) < K_0(s_0)$ give rise to higher utility $V(s_0)^U > V(s_0)^F$ whenever $s_0 < s^U$.

First, notice that for $k_0 \leq \frac{k^{SS}}{1-\delta}$ (which will always be the case for $s_0 < 1$), the value from $t = 1$ onwards is

$$V_1(k_0) = (1-\delta)k_0 - k^{SS} + \pi(k^{SS}, 1) + \beta\frac{\pi(k^{SS}, 1) - \delta k^{SS}}{1-\beta} \tag{2.15}$$

so the discounted marginal value of capital at $t = 1$ from the point of view of $t = 0$ is $\beta(1-\delta)$.

Hence, for the range of shocks considered in the proposition, optimal policy under the restriction of full utilization is easily characterized as follows. Let s^D be the lower bound of the inaction region in period 0, that is the productivity level that solves $q - \beta(1-\delta) = \pi_x((1-\delta)k^{SS}, s^D)$. Notice that $s^D < s^U$ as $q < \beta(1-\delta)$.

- if $s_0 \leq s^D$ then the firm disinvests, choosing $K_0^F(s_0) < (1-\delta)k^{SS}$ by solving $q - \beta(1-\delta) = \pi_x(K_0^F(s_0), s^D)$. The associated value is given by

$$V^F(s_0) = q\left[(1-\delta)k^{SS} - K_0^F(s_0)\right] + \pi(K_0^F(s_0), s_0) + W(K_0^F(s_0))$$

- if $s^D < s_0 \leq s^U$ the firm is in the inaction region and chooses $K_0^F(s_0) = (1-\delta)k^{SS}$. The associated value is

$$V^F(s_0) = \pi((1-\delta)k^{SS}, s_0) + V_1((1-\delta)k^{SS})$$

Now consider the possibility of choosing the optimal utilization rate of capital and not downsizing for any value of $s < s^U$. Formally, this means $K_0^U = (1-\delta)k^{SS}$ and

$X_0^U(K_0^U, s_0)$ solves $\pi_x(X_0^U(K_0^U, s_0), s_0) = 0$. The associated value is

$$V^U(s_0) = \pi(X_0^U(K_0^U, s_0), s_0) + V_1((1-\delta)\, k^{SS})$$

which implies the following differences in attained utility:

- if $s_0 \leq s^D$,

  $$V^U(s_0) - V^F(s_0) = \pi(X_0^U(K_0^U, s_0), s_0) - \pi(K_0^F(s_0), s_0) + (1-\delta)^2\, k^{SS} - (1-\delta)\, K_0^F(s_0) > 0$$

  Notice that there are two reasons why this expression is positive: both the difference in period 0 profits and the second part of the expression are positive. This is because underutilization allows to maximize profits in period 0 as well as to save the reinvestment in period 1.

- if $s^D < s_0 \leq s^U$,

  $$V^U(s_0) - V^F(s_0) = \pi(X_0^U(K_0^U, s_0), s_0) - \pi((1-\delta)\, k^{SS}, s_0) > 0.$$

  In this case the choice of capital level is the same under the two scenarios, but underutilization allows to maximize static profits in period 0.

To conclude the proof, it is easy to see that $K^U(s_0) = (1-\delta)\, k^{SS}$ dominates any disinvestment for all $s_0 < s^U$ because the sold capital would have to be repurchased in period 1 at price of 1 and $q < \beta\,(1-\delta)$ so this would be suboptimal. □

Proposition 3 establishes that when downsizing is sufficiently costly, it is optimal to respond to sufficiently large transitory negative shocks by underutilizing capital.

2.2.3 Numerical results

In this subsection I parametrize the model and illustrate its key properties. A period is a quarter. The profit function is $\pi(k, s) = sk^\alpha - ck$.[6] Table 2.1 shows the choice of parameter values. The values of α, β and δ are standard in the macro literature.

[6]In this model, including labor as a second input in the Cobb-Douglas production function amounts to a simple modification of the curvature parameter α and the mean of s as long as the labor decision is fully flexible.

I consider an autoregressive process for productivity: $\log(s_{t+1}) = \rho \log(s_t)+u_t$ with $\rho \in [0,1)$ and u_t iid normally distributed with mean $-\frac{\sigma_u}{2}$ and standard deviation σ_u. I fix $\sigma_u = .068$ to match the yearly volatility of firm's idiosyncratic shocks in chapter 1 and then compare a model with $\rho = 0$ ("iid s" model) with a model with $\rho = .95$ ("persistent s" model).

The parametrization of the production (or energy) cost and the resale price is meant to be illustrative of the properties of the model. I set $c = .2$ and $q = .9 < \beta(1-\delta)$. A full calibration of the model is left for future work.

Parameter	Value
α	.33
β	.99
δ	.025
c	.2
σ_u	.068
ρ	$\in \{0, .95\}$
q	.9

Table 2.1: Parameter values in the baseline model (individual firm)

I solve the model by iterating on the value function defined in (2.1), (2.2) and (2.3). In the following, I discuss and illustrate three numerical results that arise from simulation of the model. Let us consider the "iid s" model first.

Result 1. *Only sufficiently large negative shocks lead to underutilization. The average utilization rate is therefore decreasing in the volatility of the profitability shock.*

The intertemporally optimal capital level is generally smaller than the static-profit-maximizing size, because the former takes into account the opportunity cost of capital due to discounting as well as the depreciation rate. Hence the utilization rate is normally one, and only falls below one when s_t is sufficiently low relative to k_{t-1}. In this case, downsizing is sufficiently costly that the firm chooses to use the underutilization option.

In Figure 2.1 I start the firm from the capital level that arises after a long series of shocks equal to the mean of s and then hit it with a one standard deviation positive (blue line) or negative (red line with crosses) innovation u_t. I construct the impulse responses by computing the difference between the path of the firm hit by each shock and a firm that is hit by no shock (and hence keeps fully utilizing its capital). In the upper panel, we see an asymmetry in the response of capital: a positive shock calls for positive investment, while a negative shock leads to inaction and capital decreases only by the depreciation rate. In both cases, utilization (lower panel) is full and does not respond to the shock. However, for a shock size of three standard deviations, the response is different, as illustrated in Figure 2.2. A negative shock of this size leads to underutilization, as can be seen in the lower panel.

The relationship between volatility and the average utilization rate (obtained from a long simulation) is shown in Figure 2.3. To construct this graph, I solve the model for three different values of σ_u, equal to the baseline value, the baseline value times 2 and the baseline value times 3. Average utilization falls as the shocks become larger, consistently with the impulse response functions illustrated above.[7]

Result 2. *Endogenous utilization leads to higher average capital relative to full utilization. The difference in average capital between endogenous utilization and full utilization is increasing in the volatility of the profitability shock.*

When the firm can choose the utilization rate, the downside risk coming from negative shocks is lower than when utilization is restricted to be full all the time. Profits can be insulated from large negative shocks by underutilizing capital instead of downsizing. Furthermore, large positive shocks lead to larger investment, because

[7] As the model is non-linear, larger volatility changes also the thresholds for underutilization. However this effect is quantitatively small and dominated by the direct effect of larger shocks.

Figure 2.1: Impulse Response Functions (1 standard dev. shock)

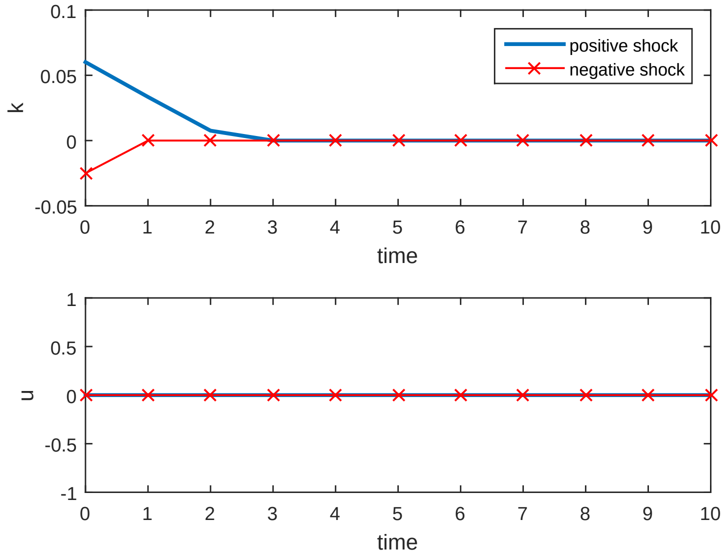


Figure 2.2: Impulse Response Functions (3 standard dev. shock)

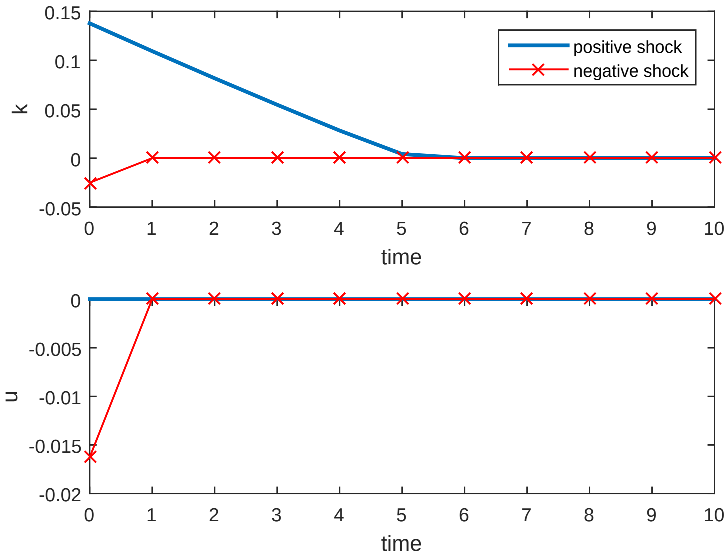
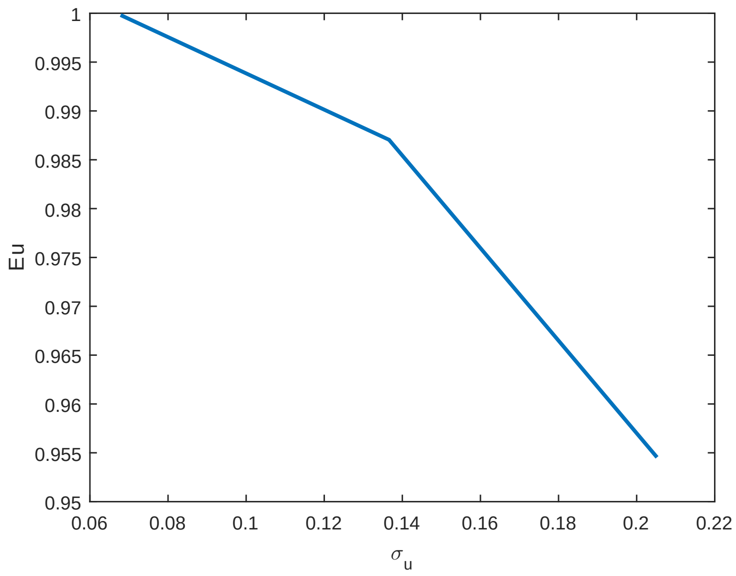


Figure 2.3: Volatility and average utilization rate



the firm is less concerned about a potential reversal of its profitability in the future. These two effects jointly contribute to a larger average size. This is illustrated in Figure 2.4, where I contrast the baseline model with endogenous utilization with a model where utilization is restricted to be full for different values of the volatility parameter σ_u. Average capital increases sharply with higher volatility, but only when utilization is endogenous.

Figure 2.5 illustrates a sequence of capital and utilization taken from a long simulation of the model with high volatility, equal to three times the baseline value. Here we can see that the lowest values for the utilization rate (lower panel) coincide with the periods when the firm restricted to fully utilize its capital (upper panel, red line with crosses) chooses to downsize by more than depreciation.

Consider now the "persistent s" model.

Result 3. *While large negative transitory shocks lead to inaction and underutilization, highly persistent shocks of the same size lead to downsizing and full utilization.*

I solve the model with autocorrelation of shocks $\rho_s = .95$. In Figure 2.6 I start

Figure 2.4: Volatility and the mean of capital

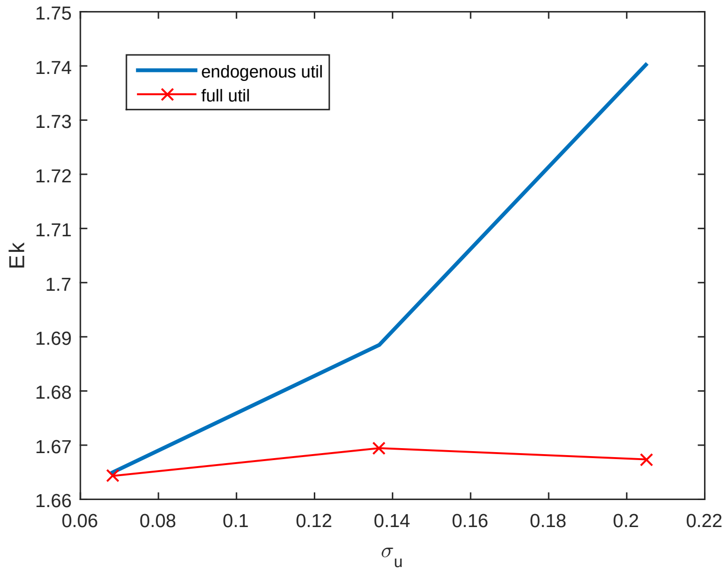
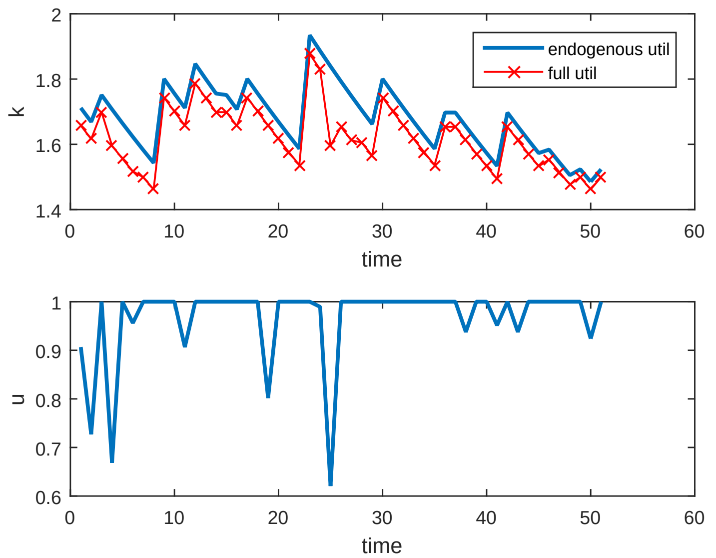
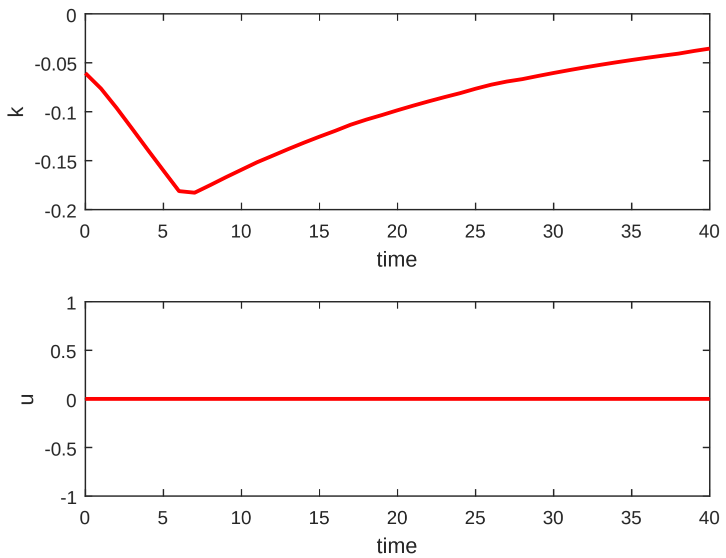


Figure 2.5: Simulation: endogenous v. full utilization



the firm from the same initial capital as in Figure 2.2 and then hit it with the same negative innovation. While in Figure 2.2 s_t was iid over time, in Figure 2.6 the shock is expected to persist for a long time. Hence, the firm chooses to downsize (upper panel) instead of holding extra capital idle. This result is consistent with a key intuition developed with the analytical results in the previous subsection: idle capital not only allows to maximize static profits, but it also saves future investment costs, when the current low profitability is expected to improve in the near future. However, if the current negative conditions are very persistent, the value of current disinvestment dominates this effect and downsizing becomes optimal.

Figure 2.6: Impulse Response Functions (3 standard dev. persistent shock)



2.3 Equilibrium with heterogeneous firms and aggregate shocks

In this section I embed the individual problem described in section 2.2 in a model with heterogeneous firms and aggregate shocks. The level of partial irreversibility q_t becomes endogenous and time-varying, as firms can trade used capital on a competitive market as in chapter 1.

In a numerical example, I show that when some firms are hit by large negative shocks and need to downsize, the resale price of capital falls and firms hit by smaller negative shocks choose to underutilize their capital instead of reallocating it to more productive firms. In this sense, I provide a joint theory of capital allocation and utilization that can be used in future work to study the cyclicality of productivity in a quantitative model.

2.3.1 Environment and firms' values

An infinitely-lived risk-neutral representative household with discount factor $\beta \in (0, 1)$ consumes the output good and owns all firms in the economy.[8]

There is a continuum of firms indexed by $j \in [0, 1]$ producing a homogeneous output good with production function $y_{j,t} = s_{j,t} x_{j,t}^{\alpha}$, where $x_{j,t}$ is utilized capital. As in the previous section, profits are given by $\pi_{j,t} = y_{j,t} - c x_{j,t}$.

Productivity $s_{j,t}$ follows a Markov process with transition matrix $T_s(z)$, where z is an aggregate shock, following another Markov process with transition T_z. Note that the realization of the aggregate shock affects the transition probabilities of the idiosyncratic shock, but there is no direct effect of the aggregate state on current production. This allows me to isolate equilibrium real-options effects following aggregate shocks.

The utilization decision is identical to the previous section and is again denoted by function X. However, the intertemporal investment decision is now modified as follows: if the firm decides to sell capital, the price is q_t, if the firm chooses to buy capital, the price is denoted by $Q_t \geq q_t$.[9]

Let m be the distribution of firms over individual states (k_t, s_t), with transition

[8] Alternatively, one can interpret the model as a small open economy.

[9] This was constant and equal to 1 in the previous section.

equation $m_{t+1} = \Gamma(m_t, z_t)$. This object is a state variable of the problem, because it will help firms to forecast future prices. The firm's problem is recursive in the state vector (k_t, s_t, z_t, m_t). Switching to recursive notation, the value of investing for a firm with previous capital level k and productivity s is

$$V^i(k, s, z, m) = \max_{k' \geq (1-\delta)k} \left\{ Q\left[(1-\delta)k - k'\right] + s\left[X(k', s)\right]^{\alpha} - cX(k', s) + \beta \mathrm{E}\left[V(k', s', z', m')|z, s\right]\right\} \tag{2.16}$$

and the value of a disinvesting firm is

$$V^d(k, s, z, m) = \max_{k' \leq (1-\delta)k} \left\{ q\left[(1-\delta)k - k'\right] + s\left[X(k', s)\right]^{\alpha} - cX(k', s) + \beta \mathrm{E}\left[V(k', s', z', m')|s, z\right]\right\}. \tag{2.17}$$

Overall, the value of a firm is $V(k, s, z, m) = max\left\{V^i(k, s, z, m), V^d(k, s, z, m)\right\}$.

2.3.2 New and used capital

Disinvesting firms sell their used investment goods. As in chapter 1, investing firms can invest either by purchasing new investment goods or a CES aggregator of new and used investment goods

$$k' - (1-\delta)k = \widetilde{i}_{new} + g(i_{new}, i_{used}) \tag{2.18}$$

$$g(i_{new}, i_{used}) = [\eta^{\frac{1}{\epsilon}} (i_{new})^{\frac{\epsilon-1}{\epsilon}} + (1-\eta)^{\frac{1}{\epsilon}} (i_{used})^{\frac{\epsilon-1}{\epsilon}}]^{\frac{\epsilon}{\epsilon-1}} \tag{2.19}$$

where $\widetilde{i}_{new}$ and i_{new} are new investment goods and i_{used} is used capital sold by disinvesting firms. $\eta \in (0, 1)$ is a parameter that determines the average ratio of new to used investment, while $\epsilon > 0$ is the elasticity of substitution between new and used investment goods. One can interpret this elasticity as an inverse measure of capital specificity. When $\epsilon \to \infty$, new and used capital are perfect substitutes and the model nests the standard case of constant resale price $q = 1-\gamma$. On the other hand, when ϵ is 0, the technology does not allow any substitutability between new and used capital. Note that new investment goods can be directly put in the production of new capital $(\widetilde{i}_{j,new})$ or bundled with used investment goods $(i_{j,new})$.

The price of a unit of new capital in terms of the output good is 1, while the unit cost of a unit of used capital is equal to the sum of the price of used capital q_t and a per-unit reallocation cost γ. Hence, the CES price index associated with a bundle g of new and used investment goods is

$$Q = [\eta + (1-\eta)(q+\gamma)^{1-\epsilon}]^{\frac{1}{1-\epsilon}}. \tag{2.20}$$

For a given chosen capital target k', investing firms split their investment between new and used in order to minimize their expenditure. As long as $q + \gamma < 1$ (which will be satisfied in the following example), firms set $\tilde{i}_{new} = 0$ and allocate their expenditure on new goods i_{new} and used capital in order to satisfy the following first order condition:

$$\frac{i_{used}}{i_{new}} = \left(\frac{1-\eta}{\eta}\right)(q+\gamma)^{-\epsilon}. \tag{2.21}$$

The market for used capital clears, meaning the total disinvestment coming from disinvesting firms equals total investment in used capital coming from investing firms.[10]

2.3.3 Numerical example

I show a numerical example that illustrates how aggregate shocks affect utilization decisions through changes in the resale price of capital. Table 2.2 shows the parameter values. Several of them are in common with the individual firm's problem described in section 2.2 and are reproduced here for convenience. The parametrization of the productivity process and the production cost is only meant to be illustrative of the properties of the model. I assume that in the good state there are two values for s_t: s^L and s^H, with iid transition probabilities $p_s = .5$. In the bad state, which happens with probability $p_z = .01$ a third, lower value s^{LL} becomes possible, and a fraction

[10]The market for the new output good also clears: consumption plus new investment equals output net of production costs and reallocation costs. This market-clearing condition is trivially satisfied thanks to risk-neutrality.

$\chi = .03$ of firms transit into this state in every quarter. The values of η, ϵ and γ are chosen following the calibration in chapter 1.

Parameter	Value
α	.33
β	.99
δ	.025
c	.2
(s^{LL}, s^{L}, s^{H})	(.5, 1, 1.3)
p_s	.5
χ	.03
p_z	.01
η	.7
ϵ	5
γ	.01

Table 2.2: Parameter values in the equilibrium model

I solve the model using a version of the Krusell and Smith (1998) algorithm, based on the approximation of the distribution of capital by means of its first moment. Firms hold a subjective belief on the law of motion of aggregate capital and the law of motion of q_t. Given these perceived laws of motion, I solve the individual problem by value function iteration as in section 2.2. Then, I simulate the economy and impose market clearing period by period. At the end of the simulation, I run regressions to update the laws of motion and proceed up to convergence of the actual laws of motion to the perceived ones.

A complication arises because of the necessity to impose market clearing in the market for used capital. Demand and supply on this market are obtained by aggregating the policy function described above, which has kinks because of irreversibility and because of the utilization option: even a small decrease in the resale price may lead to a jump from downsizing to inacting and underutilizing capital. Market clearing requires interpolation of these demand and supply schedules. Appendix A illustrates the potential inaccuracy due to these kinks and describes a simple way to deal with interpolation of kinked functions efficiently in the market-clearing stage of the algorithm.

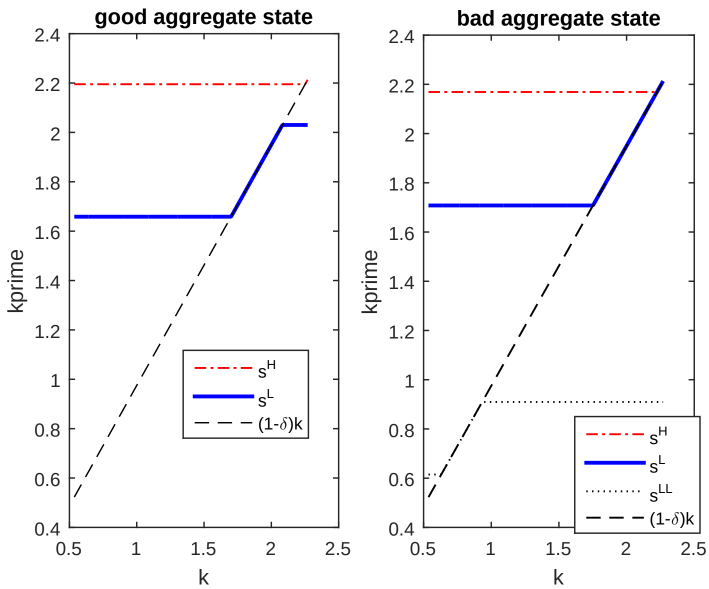
The left-hand panel of Figure 2.7 shows the policy function for capital choice in the good aggregate state. The dashed-dotted red line represents the capital choice for firms with productivity s^H, while the solid blue line represents the capital choice for firms with productivity s^L. For both types of firms, the inaction region coincides with the depreciation line (dashed black line). It can be seen that reallocating some capital is optimal for a firm that goes from high to low productivity: there is a gap between the two lines.

When the bad aggregate shock hits (right-hand panel), a fraction χ of firms have productivity s^{LL} and downsize to the thin black dotted line. Firms going from s^H to s^L stop reallocating and immediately enter the inaction region. The reason for this behavior is the fall in the resale price q_t. Instead of selling part of their capital, these firms wait and keep it idle, hoping for one of two outcomes: a new transition to s^H or a higher resale price.

In order to illustrate this point, I simulate the economy and show a transition from a long sequence of good aggregate state to a long sequence of bad aggregate state. Figure 2.8 shows that the aggregate utilization rate falls when the aggregate shock hits (period 1 in the figure), as both firms with s^L and s^{LL} find it optimal to underutilize their capital.

Figure 2.9 illustrates the dynamics of the resale price, compared with those obtained in a model without option to underutilize capital (full utilization). It can be seen that the fall in q_t is smaller when underutilization is allowed, because firms with low productivity have two ways to save on their production costs: reallocating or keeping capital idle. On the contrary, in the full utilization model the shift in supply is larger because unproductive firms' only way to avoid a large fall in profits is to downsize. This suggests that an implication of endogenous utilization is a reduction in the volatility of the resale price of capital.

Figure 2.7: Policy function



2.4 Conclusion

This paper studies the relation between partial investment irreversibility and capital utilization. In a partial equilibrium setting, I establish conditions on the degree of irreversibility and size of transitory shocks such that firms optimally choose to keep idle capital even when they could downsize before production.

I embed this problem in an industry equilibrium model with endogenous irreversibility. Aggregate shocks affect the utilization rate through equilibrium movements in the resale price of capital. The allocation of capital across heterogeneous

Figure 2.8: Capital utilization rate

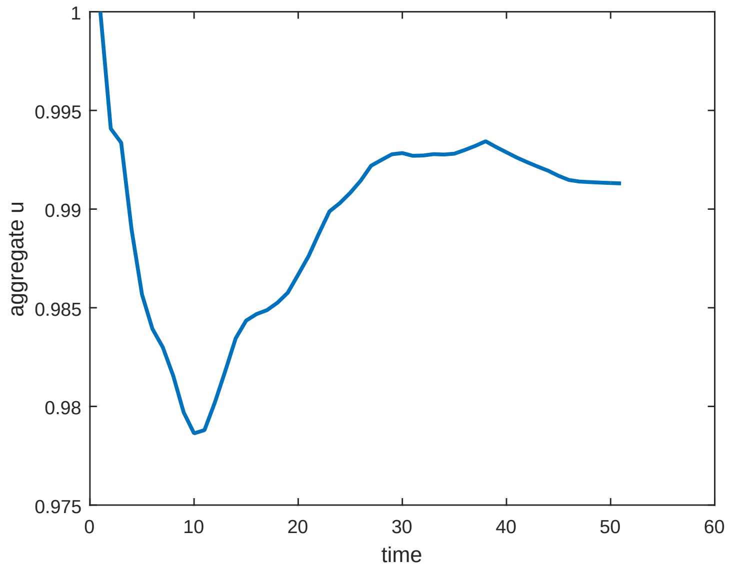
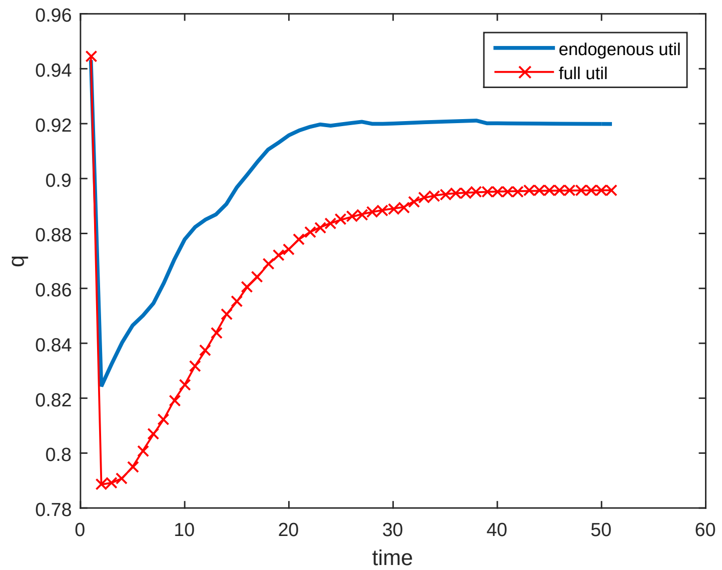


Figure 2.9: Resale price of capital



firms and the distribution of utilization rates are jointly determined in equilibrium.

There are two main directions of future work that follow naturally from this paper. First, using disaggregated data at the industry or firm-level, I will test the results of section 2.2 on the relations between the volatility and persistence of shocks and utilization and investment decisions. Second, I will calibrate the economy described in section 2.3 and analyze how allocation and utilization of capital determine the dynamics of output and aggregate productivity in a quantitative model of the business cycle.

Appendix A: An algorithm to interpolate kinked policy functions accurately and efficiently

The model is solved using a version of the Krusell and Smith (1998) algorithm. The distribution of capital is approximated using its first moment. Firms perceive laws of motion for aggregate capital and the resale price. For a given policy function obtained by value function iteration, the model is simulated in order to impose market clearing and update the perceived law of motion up to convergence with its actual counterpart.

In order to impose market clearing in the market for used capital, it is necessary to evaluate aggregate demand and aggregate supply for used capital at different values for q_t on a grid in each period of the simulation, and then solve for the market-clearing q_t, which in general will not be a grid point. Hence, interpolation is necessary. These aggregate demand and supply functions are derived by aggregating individual policy functions, which have kinks because of the inequality constraints in (2.16) and (2.17). However, interpolation of kinked functions will unfortunately not preserve the properties of the true policy function, and will introduce further kinks, which have no economic rationale and lead to inaccuracy in the computation of the equilibrium price.

Here, I briefly show a way to circumvent this problem by applying the interpolation on a smooth transformation of the kinked policy function, preserving the economic properties of the solution and obtaining an accurate solution efficiently. In order to illustrate first the problem and then my solution, I will introduce a simple static problem. It will be then straightforward to see how the solution to this problem can be applied to the full model specified in the main text.

Consider the problem

$$\max_{k' \leq (1-\delta)k} q\left[(1-\delta)k - k'\right] + sk'^{\alpha}$$

where k and s are treated as fixed numbers and we want to compute the solution as a function of q.

$$k'(q) = \begin{cases} (1-\delta)\bar{k}(q), & \text{if } k \geq \bar{k}(q) \\ (1-\delta)k, & \text{otherwise} \end{cases} \tag{2.22}$$

where the disinvestment threshold is defined by $\bar{k}(q) \equiv (\frac{\alpha s}{q})^{\frac{1}{1-\alpha}}/(1-\delta)$. Equivalently, $k'(q) = min\left\{(1-\delta)\bar{k}(q), (1-\delta)k\right\}$.

Imagine that we have solved for this policy function for two values of q, q^i and $q^j > q^i$ (grid points) and assume that the market clearing price is $q^l \in (q^i, q^j)$ and we want to rely on interpolation to compute the policy function for this price.[11] Figure 2.10 illustrates the problem. The policies for q^i and q^j are the thin blue lines, while the depreciation line is the dashed black line. If we simply interpolate (2.22) on the grid for q and evaluate it at q^l, we obtain the blue dashed-dotted line, which clearly does not have the same properties of the original policy function: it has two kinks instead of one and it is upward sloping with a slope different from $(1-\delta)$ between $\bar{k}(q^j)$ and $\bar{k}(q^i)$. However, observe that the function $\bar{k}$ is smooth in q. Hence, we can interpolate this function, evaluate $\bar{k}(q^l)$, and then reconstruct the policy function as follows:

$$k'(q^l) = \begin{cases} (1-\delta)\bar{k}(q^l), & \text{if } k \geq \bar{k}(q^l) \\ (1-\delta)k, & \text{otherwise.} \end{cases} \tag{2.23}$$

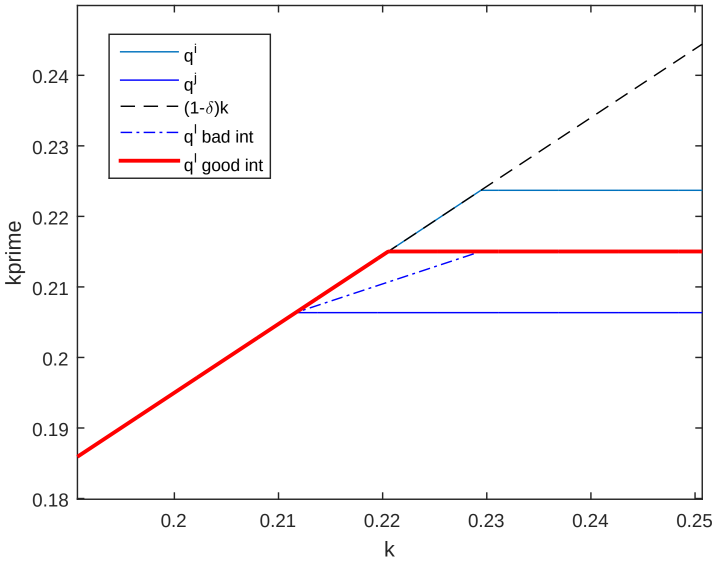
This gives the thick red line, which represents an accurate evaluation of the true policy function, preserving the properties of the solution, with only one kink at the threshold for disinvestment.

In order to implement this algorithm to solve the full model, I solve the individual problem using q as a pseudo-state variable and obtain the thresholds of the inaction region as functions of q. In the simulation stage, in each period I compute the equilibrium price by interpolating these thresholds and then reconstruct the policy functions to compute how much capital each firm buys and sells before moving to the

[11] Clearly, in this simple case, this is not needed, because we have an analytical solution for the policy function, but this is not the case in the dynamic model.

next period.

Figure 2.10: Interpolation of kinked policy functions



Chapter 3

Optimal Policy with Endogenous Signal Extraction

Joint with Esther Hauk and Albert Marcet.

> "In the policy world, there is a very strong notion that if we only knew the state of the economy today, it would be a simple matter to decide what the policy should be. The notion is that we do not know the state of the system today, and it is all very uncertain and very hazy whether the economy is improving or getting worse or what is happening. Because of that, the notion goes, we are not sure what the policy setting should be today. [...] In the research world, it is just the opposite. The typical presumption is that one knows the state of the system at a point in time. There is nothing hazy or difficult about inferring the state of the system in most models" (James Bullard, interview on *Review of Economic Dynamics*, November 2013)

3.1 Introduction

3.1.1 Motivation

The opening quote states that inferring the underlying state of the economy is a key practical difficulty in setting macro policy. One could say that this is not such an insurmountable problem: a policy-maker should choose the optimal policy taking into account the uncertainty ("haziness") about the underlying state given the information available to him. However, information available to policy-makers is often endogenous to policy decisions. Hence, in general the problem of designing optimal policy contingent on endogenous information cannot be separated from the problem of inferring the state of the world from the observables. In this paper we develop a general solution to the problem of finding optimal policy with signal extraction from endogenous variables and we illustrate our results in a simple model of fiscal policy.

To illustrate the importance of this issue, consider the fiscal policy response to the recent financial crisis. In 2008-2009 policy-makers observed a large fall in output and employment, but it was unclear whether the recession was due to a shock to productivity or to a demand shock or some combination of both. Nonetheless, policy-makers had to design a reaction to the recession, based only on a signal extraction on the nature of the shock. Since output itself depends on whether an expansionary fiscal policy or austerity is adopted, the problem of signal extraction is endogenous to the choice of policy.

The G20 decided in its Washington meeting to take aggressive expansionary fiscal policy measures in order to reactivate the economy. Those measures involved spending 2% of world GDP. Policy-makers were uncertain about the fiscal deficits that would ensue. From the relatively optimistic initial estimate many governments have eventually switched to the opposite fiscal policies, implementing tax hikes and austerity programs. In many occasions the switch to austerity has been quite sudden. Austerity had to be even stricter because of the higher spending that the G20 encouraged.[1]

The standard literature on Ramsey-optimal fiscal policy assumes that taxes can be a function of the realizations of the shocks hitting the economy. However in the real world these policies can only depend on observable variables such as income, that in turn responds simultaneously to taxes as well as the underlying shocks. Automatic stabilizers, e.g. income taxes and unemployment benefits, are a leading example of policies that respond to aggregate endogenous information. This paper addresses the question of how to design such instruments optimally.

The existing literature on optimal policy with signal extraction has already pointed out that there can be an endogeneity issue when signals respond to policy. However it has worked only with linear models, where a "certainty equivalence" result arises: under Partial Information it is optimal to apply the same policy as under Full Information to the conditional expectation of the underlying shocks. While the signal extraction can in some cases depend on the policy, the optimal policy itself is independent of the information structure and can be derived easily.[2]

A model of fiscal policy is non-linear in nature. For instance, productivity shocks enter multiplicatively with hours worked in the budget constraint. Furthermore, the existence of a Laffer curve creates non-linearities in tax revenue. This paper derives general non-linear methods to solve for optimal policy when there is no separation of any kind between optimization and signal extraction, leading to a violation of "certainty equivalence".

We first analyze optimal policy in a static (or two-period) model of optimal control with multidimensional uncertainty and an endogenous signal. The density of the signal depends on the optimal policy and vice versa. We derive a first order condition

[1]The Spanish Finance Minister at the time, Pedro Solbes, published an article in El País on the 10th of November 2013 under the title "Cuando decidí salir del Gobierno" explaining how the outcome of this meeting eventually lead to the current Spanish Debt crisis and near-default in summer 2012.

[2]See subsection 3.1.2 for a detailed discussion of the literature.

(FOC) for the optimal policy relying on first principles. This optimality condition is different from the standard FOC found in dynamic stochastic models: the probabilities of each state of nature need to be weighted by a kernel that depends on the effect of policy on the observed signal. The FOC is derived for a general model so that our results can be widely applied. We show how "certainty equivalence" arises in some special cases as an implication of our general theorem.

Our leading example is a two-period version of the standard fiscal policy model of Lucas and Stokey (1983). We introduce two shocks (to demand and to supply) and to make the issue of hidden information relevant we assume incomplete insurance markets as in Aiyagari et al. (2002). Then we solve for optimal Ramsey taxation under the assumption that the government does not observe the realizations of the shocks, but it only observes some endogenous signal, such as output or hours worked.

This model yields interesting insights about the conduct of fiscal policy. First, hidden information can be a driver of tax smoothing: taxes are less volatile relative to a case with Full Information. This is because one implication of the optimal policy is to average all the possible contingencies using "proper" weights for each contingency that depend on how reactive observables are to policy. This identifies a reason for tax smoothing different from the standard one in Ramsey policy.

When we consider a case with very high government spending, capturing a situation of fiscal crisis, the implied policy reaction can be very non-linear. In particular, in a case when future taxes could be close to the maximum of the Laffer curve, as was arguably the case in some European economies in the current crisis, a government may go very quickly from not reacting to low observed output to increasing taxes very strongly as output goes down. This sudden adjustment occurs once policy-makers realize that a fiscal crisis may materialize under the worst possible realizations of the shocks in the next period, but importantly this inference is endogenous to the output observed and the policy chosen. This may rationalize why some countries first reacted slowly to the crisis and then went for austerity quickly but with a delay. We also show the consequences of using the existing linear methods based on “certainty equivalence” when the actual optimal policy is highly non-linear.

Finally, we build an infinite-horizon model that confirms our main results in a dynamic setup: the Ramsey government under Partial Information may react with a delay to a downturn. The delayed fiscal adjustment induces a longer recession relative to the Full Information benchmark, as has arguably happened in the countries that expanded government spending after November 2008 and eventually introduced austerity measures. It may seem that irresponsible politicians amplified the recession by delaying tax increases, but (in the model) this is a feature of an optimal decision.

3.1.2 Related literature

Most of the literature on optimal policy in dynamic models in the last thirty years has disregarded the issue of endogenous signal extraction. However, Partial Information and signal extraction were often present in the early papers on dynamic models with Rational Expectations (RE).

Signal extraction with an *exogenous* signal is well understood; it goes as far back as Muth (1960). Typically, it just requires a routine application of the Kalman filter. Because the signal extraction problem is solved before deciding on policy it is said that there is a “separation principle” between optimization and estimation.

Few papers have studied optimal policy when signals are endogenous. Pearlman et al. (1986), Pearlman (1992) and Svensson and Woodford (2003) consider linear Gaussian models where the policy-maker and the private sector have the same information set. In other words information is partial but symmetric. In this case, they show that the “separation principle” continues to hold. Baxter et al. (2007, 2011) derive an “endogenous Kalman filter” for all these cases which is equivalent to the solution of a standard Kalman filter of a parallel problem where all the states and signals are fully exogenous.

Closest to our work is Svensson and Woodford (2004). They consider optimal policy in a non-microfounded linear Rational Expectations model, where the government's information set is a subset of the private sector's information set. They show that, even though the "separation principle" fails because of asymmetric information, there is a suitable modification of the standard Kalman filter that works, thanks to linearity and additively separable shocks. Moreover optimal policy has the "certainty equivalence" property: under Partial Information the government applies the Full Information policy to its best estimate of the state. In our setup, because of non-linearity, both the "separation principle" and "certainty equivalence" fail. Aoki (2003) applies these results to optimal monetary policy with noisy indicators on output and inflation. Nimark (2008) applies them to a problem of monetary policy where the central bank uses data from the yield curve while at the same time understanding that it is affecting them.

Our contribution is to consider a fully microfounded optimal policy model and to provide a general solution to the endogenous signal extraction and optimization problem when the government (or, more generally, a Stackelberg leader) conditions on a signal simultaneously determined with policy. There is no separation of any kind if the shocks and policy variables are allowed to enter non-linearly in the equilibrium conditions. We find how "certainty equivalence" arises in the linear case using our general theorem and we show a case where the correct solution is highly non-linear in nature and, therefore, a linear approximation can be misleading in this case.[3]

The effect of policy choices on information extraction about unobserved variables is also considered in the armed-bandit problems of decision theory. For some applications to dynamic macro policy see Kiefer and Nyarko (1989), Wieland (2000a, 2000b) and Ellison and Valla (2001), or to monopolist behavior see Mirmann, Samuelson and Urbano (1993). Van Nieuwerburgh and Veldkamp (2006) use a similar learning frame-

[3]Optimal non-linear policies have been found in the literature but for totally different reasons. Swanson (2006) obtains a non-linear policy when he relaxes the assumption of normality in the linear model with separable shocks. He considers a model where the separation principle applies. The non-linearity results entirely from Bayesian updating on the a priori non-Gaussian shocks.

work to explain business-cycle asymmetries. In these papers the planner's decision influences the probability of next period's signal, but the current signal is unaffected by the current policy decision so that separation holds given the past policy decision. It would be interesting to blend the issues in these papers with the one studied in the current paper.

Furthermore, a wide literature considers models where competitive agents use prices as signals of unknown information.[4] Since prices are taken as given by competitive agents the signal extraction problem can be solved with standard filtering techniques and the issue we address does not arise in this literature.

The literature of optimal contracts under private information and incentive compatibility constraints (or the "New Dynamic Public Finance" as in Kocherlakota, 2010) is perhaps less directly related to our work. This literature usually assumes revelation of the private information conditional on the equilibrium actions (the "invertibility" case where we argue below Partial Information is not relevant) but it assumes that agents react strategically to the optimal contingent policy set up by the principal, an issue that we abstract from. In our setup the government conditions on aggregate variables and, since agents are atomistic, the policy function ($\mathcal{R}$ in the text) does not affect agents' decision while the government action (the tax rate τ in our main example) does and it is taken as given. On the other hand, there is an interaction between the signal extraction problem and the optimal policy decision that the literature on incentive compatibility constraints often ignores. Blending the two issues would be of interest but it is left for future research.

[4]Lucas's (1972) seminal paper, and Guerrieri and Shimer (2013) analyze a competitive market in this setup. For an optimal policy problem see Angeletos and Pavan (2010) where those solving a signal extraction problem are the agents (not the government). The whole controversy about whether asymmetric information RE equilibria could be reached as discussed in Townsend (1983) is also within this framework.

3.2 A simple model of optimal fiscal policy

We first present a simple model of optimal policy. This will serve the purpose of illustrating the problem of endogenous signal extraction and it will be of interest on its own. It is a very simple two-period version of Lucas and Stokey (1983). We introduce incomplete insurance markets to be consistent with the Partial Information story.

3.2.1 Preferences and technology

The economy lasts two periods $t = 1, 2$. A government needs to finance an exogenous and deterministic stream of expenditure (g_1, g_2), where subscripts indicate time periods, using distortionary income taxes (τ_1, τ_2) and bonds b^g issued in the first period that promise a repayment in second period consumption units with certainty.

The economy is populated by a continuum of agents. Each agent $i \in [0, 1]$ has utility function

$$\mathrm{E}\left[U\left(c_1^i, l_1^i, c_2^i, l_2^i; \gamma\right)\right] \tag{3.1}$$

where

$$U\left(c_1^i, l_1^i, c_2^i, l_2^i; \gamma\right) = \gamma u\left(c_1^i\right) - v\left(l_1^i\right) + \beta\left[u\left(c_2^i\right) - v\left(l_2^i\right)\right]$$

where c_t^i and l_t^i for $t = 1, 2$ are consumption and hours worked respectively, with $u' > 0$, $u'' < 0$, $v' > 0$, $v'' > 0$.

γ is a random variable with distribution F_γ, we refer to it as a "demand shock". When γ is high, agents like first period consumption relatively more than other goods. At the same time high γ makes them willing to work more in their intratemporal labor-consumption decision and also more impatient in their intertemporal allocation of consumption. Vice versa, for low γ. Given that agents are identical, in the following we drop the subscripts i for notational convenience.

The production function in each period is linear in labor and output is given by

$$y_t = \theta_t l_t \; for \; t = 1, 2. \tag{3.2}$$

The random variable $\theta_1 = \theta$ has distribution F_θ and we will refer to it as the "productivity shock". As far as θ_2 is concerned, we will distinguish two cases, one where $\theta_2 = \theta_1$, in which case the productivity shock is permanent, and one where $\theta_2 = \mathrm{E}\theta$, that is, second period productivity is a known constant, equal to the mean of the first period shock, in which case the productivity shock is temporary. Notice that in both cases θ_2 is known given θ_1.

To summarize, the state of the economy is fully described by a realization of the random variables $A \equiv (\gamma, \theta)$, these variables are observed at the beginning of period $t = 1$ by consumers and firms, but not by the government. γ and θ are assumed to be independent. The distributions F_γ and F_θ represent the government's perceived distribution of the exogenous shocks.

We will consider agents that have rational expectations. To this end, denoting by Φ the space of possible values of A, we assume that agents know that fiscal policy is given by a triplet of functions $(\widetilde{\tau}_1, \widetilde{\tau}_2, \widetilde{b}^g) : \Phi \rightarrow R^3$ and these are actually the equilibrium values of taxes and government bonds for each A.

Consumers' choices and prices are contingent on the state (γ, θ) observed in period $t = 1$. Agents choose $(c_1, c_2, l_1, l_2, b) : \Phi \rightarrow R^5$ knowing the fiscal policy and the bond price function $q : \Phi \rightarrow R$. Obviously, the solution of the agents' problem in this setup coincides with the non-stochastic model where γ, θ are known. Uncertainty will only play a role in the government's problem, to be specified later.

Firms also observe θ at $t = 1$. Profit maximization implies that agents receive a wage equal to θ_t, so that the period budget constraints of the representative agent are

$$c_1 + qb = \theta l_1 (1 - \widetilde{\tau}_1) \tag{3.3}$$

$$c_2 = \theta_2 l_2 (1 - \widetilde{\tau}_2) + b \tag{3.4}$$

where q is the price of the government discount bond b. The above budget constraints have to hold for all realizations γ, θ we leave this implicit.

3.2.2 Competitive Equilibrium

Here we provide a definition of competitive equilibrium. The definition is standard in the literature, it is common to both the Full Information (FI) and the Partial Information (PI) equilibria that we analyze.

Definition 2. *A* ***competitive equilibrium*** *is a fiscal policy $(\widetilde{\tau}_1, \widetilde{\tau}_2, \widetilde{b^g})$, price q and allocations (c_1, c_2, l_1, l_2, b) such that when agents take $(\widetilde{\tau}_1, \widetilde{\tau}_2, q)$ as given the allocations maximize the agents' utility (3.1) subject to (3.3) and (3.4). In addition, bonds and goods markets clear, so that $b^g = b$ and*

$$c_t + g_t = \theta l_t \text{ for } t = 1, 2. \tag{3.5}$$

This definition embeds competitive equilibrium relations insuring that wages are set in equilibrium and that the budget constraint of the government holds in all periods due to Walras' law.

Utility maximization implies for all A

$$\frac{v'(l_1)}{u'(c_1)} = \theta\gamma(1 - \widetilde{\tau}_1) \tag{3.6}$$

$$\frac{v'(l_2)}{u'(c_2)} = \theta_2(1 - \widetilde{\tau}_2) \tag{3.7}$$

$$q = \beta \frac{u'(c_2)}{\gamma u'(c_1)} \tag{3.8}$$

As anticipated, the demand shock enters the first period labor supply decision described by (3.6) as well as the bond pricing equation (3.8). A competitive equilibrium is fully characterized by equations (3.3) to (3.8).

3.2.3 Ramsey Equilibrium

To describe government behavior we now provide a definition of Ramsey equilibrium. As is standard we assume the government has full commitment, perfect knowledge about how taxes map into allocations for a given value of the underlying shocks A and that it chooses the best policy for the consumer.

We first give the standard definition when both government and consumers observe the realization of A.[5]

Definition 3. *A* ***Full Information (FI-) Ramsey equilibrium*** *is a fiscal policy* (τ_1, τ_2, b^g) *that achieves the highest utility (3.1) when allocations are determined in a competitive equilibrium.*

Our interest is in studying optimal taxes under Partial Information. More precisely, we assume that taxes in the first period have to be set before the shock A is known but after observing a **signal** s that depends potentially on aggregate outcomes observed in period 1, $s = G(c_1, l_1, q, A)$ for a given G.

Definition 4. *A* ***Partial Information (PI-) Ramsey equilibrium*** *when government observes a signal* s *is a FI-Ramsey equilibrium satisfying*

1. τ_1 *is measurable with respect to* s
2. *fiscal policy* (τ_1, τ_2, b^g) *achieves the highest utility from among all equilibria satisfying 1.*

Restriction 1 can be expressed as the PI-Ramsey equilibrium having to satisfy

$$\tau_1 = \mathcal{R}(s)\ for\ all\ A \in \Phi,\ \ for\ some\ \mathcal{R} : R \to R \tag{3.9}$$

We are interested in the case when this restriction prevents the PI-Ramsey equilibrium from achieving the FI version.

[5]In these definitions we take for granted that we only consider tax policies for which a competitive equilibrium exists and is unique.

Note that consumers may or may not know that (3.9) holds, in any case they take as given the tax rate that arises from this equation and equilibrium. Even if they knew (3.9), they would not be able to exploit this knowledge in their optimization problem, because they are atomistic and cannot affect the aggregate signal, and hence the tax rate.[6] We assume that s is an aggregate variable so that as long as consumers are infinitesimal they take the tax level τ as given. In this model, as is standard in Ramsey equilibria, the tax level τ and the equilibrium allocations (and therefore s) are determined simultaneously as a consequence of the government's choice for $\mathcal{R}$.

FI-Ramsey equilibria

Using the so-called "primal approach" and standard arguments it is easy to show that an allocation is a competitive equilibrium if and only if, in addition to resource constraints (3.5), the following implementability condition holds

$$\gamma u'(c_1)\, c_1 - v'(l_1)\, l_1 + \beta \left[u'(c_2)\, c_2 - v'(l_2)\, l_2\right] = 0. \tag{3.10}$$

The standard approach to find Ramsey policy under FI is to maximize (3.1) subject to (3.10). We now slightly deviate from this traditional approach in order to obtain a formulation of the FI problem that is as close as possible to the PI problem.

Implicit in the standard definition of FI Ramsey equilibrium is the assumption that the government knows how the economy reacts to a given tax policy given each value of A. We find it convenient to write out this reaction function explicitly since this reaction function is the natural way to write the problem under Partial Information. Using (3.5) for $t = 1$ to substitute out consumption in (3.6), we get

$$\frac{v'(l_1)}{u'(\theta l_1 - g_1)} = \theta\gamma(1 - \tau_1), \tag{3.11}$$

Letting h be the function that gives the l_1 that solves this equation for given τ_1, θ, γ we can rewrite the above equilibrium condition as

$$l = h(\tau, \theta, \gamma) \tag{3.12}$$

[6]This differs from the situation in the New Public Finance where consumers optimize given the policy function $\mathcal{R}$ which is a function of individual choices.

where we have again suppressed the time subscript from first period labor and tax rate. This shows how the signal l reacts to the tax choice.

Now we write the equilibrium utility function as a function of first period equilibrium allocations and shocks only. Using the resource constraints (3.5) to substitute out c_t in (3.10) gives one equation that, for each A, involves only the unknowns l_1, l_2. This defines implicitly a function that maps an equilibrium l_1 into a second period equilibrium labor l_2, call this map $L_2^{imp} : \Phi \times [0,1] \to R$,[7] so that

$$L_2^{imp}(A, l_1) \text{ for all } A \in \Phi \tag{3.13}$$

solves (3.10). The welfare of the planner for each A can be written as

$$\begin{aligned} W(l; A) &\equiv U(\theta l - g_1, l, \theta_2 L_2^{imp}(l, A) - g_2, L_2^{imp}(l, A); \gamma) \qquad (3.14) \\ &= U(c_1, l_1, c_2, l_2; \gamma) \end{aligned}$$

Note that the only arguments in W are the observed variable l and the shock γ. The functions U and L_2^{imp} are embedded in the definition of competitive equilibrium and known by the government.

Given the above discussion, the FI Ramsey Equilibrium reduces to solving

$$\begin{aligned} &\max_{\tau:\Phi\to R} \mathrm{E}[W(l; A)] \qquad (3.15) \\ &\text{s.t. } (3.12) \end{aligned}$$

Obviously the result is the same as to maximize (3.1) subject to (3.10) given A.

The standard approach in Ramsey policy analysis is to not even write the constraint (3.12) in (3.15). This is reasonable as under FI this constraint is not binding, its only role is to give the FI-Ramsey tax policy $\mathcal{R}^{FI} : \Phi \to R$ in the first period, but this constraint will be active in the PI formulation.

[7] Under some specific assumptions on u and v, it will be possible to solve for l_2 as a function of l_1 in closed form. In general, the marginal effect of l_1 on l_2 is easily found by applying the implicit function theorem to (3.10). See our examples below.

PI-Ramsey equilibria

We focus on the case when the signal is just labor so $s = l_1 = l$ (in the robustness section we consider the case when output $y = \theta l_1$ is observed). The only difference under PI is that the additional constraint (3.9) appears and that the choice is over a tax contingent on the signal. Hence a **PI-Ramsey equilibrium** (given a signal l) solves

$$\begin{aligned} &\max_{\mathcal{R}:R\to R} \mathrm{E}[W(l;A)] \\ &\text{s.t. } (3.12) \\ &\tau = \mathcal{R}(l) \end{aligned} \tag{3.16}$$

This gives rise to a non-standard maximization problem. We tackle this issue by application of calculus of variations in section 3.3.

3.2.4 The economic consequences of PI for taxation policy

Before giving a mathematical solution it is worthwhile discussing the economic issues raised by limited information in the fiscal policy example we use.

As is well known the optimal FI policy is one of tax smoothing over time as the government wants to spread the distortions equally in the two periods. In the case of CRRA preferences

$$u(c) = \frac{c^{1+\alpha_c}}{1+\alpha_c},\ v(l) = B\frac{l^{1+\alpha_l}}{1+\alpha_l}$$

for $\alpha_c \leq 0$, $\alpha_l, B > 0$, tax smoothing will be perfect and Ramsey policy under FI involves setting a constant tax rate $\tau = \tau_1 = \tau_2$ that solves the intertemporal budget constraint

$$\tau\theta l_1 - g_1 + \beta\frac{u'(c_2)}{\gamma u'(c_1)}\left(\tau\theta_2 l_2 - g_2\right) = 0. \tag{3.17}$$

It is clear from (3.17) that the government needs to know the realization of both productivity and demand shock in order to implement this policy under FI. In particular, the realization of $\theta = \theta_1$ is a crucial piece of information, as it determines the revenue that a given tax rate is going to raise. The demand shock γ also matters as it affects both the objective function and the interest rate that the government will have to pay on its debt. Furthermore, both shocks clearly contribute to the determination of an allocation (c_1, c_2, l_1, l_2).

Under PI the government can only condition its policy on l, without knowing what combination of the shocks gives rise to a given observation. Depending on the realizations, the government would like to set different tax rates and under some preference assumptions (e.g. log-quadratic, discussed in section 3.4) it may even be the case that a certain increase in hours would call for a tax cut if driven by a high realization of γ, but would call for a tax hike if driven instead by low θ. Since the government does not observe these shocks, this makes this model a very interesting framework to study optimal policy with PI.

Clearly, under PI the choice of constant taxes is not feasible. The government has to fix τ_1 while it is still uncertain about the revenue that this tax rate will generate and it will enter period 2 with an uncertain amount of debt. Once θ and γ are known in period 2 the government will have to set τ_2 so as to balance the budget in the second period in order to avoid default, so to the government τ_2 is unavoidably a random variable at the time of choosing τ_1.

Arguably, uncertain tax revenue is a crucial feature of actual fiscal policy decision, and tax rates are decided based on information from equilibrium outcomes that are observed frequently. In this sense, one can interpret this model as a simple model of optimal automatic stabilizers, as these are fiscal instruments that are designed to respond to endogenous outcomes, such as income or unemployment, independently of the source of fluctuations in these variables. The optimal design of these instruments requires a simultaneous determination of the density of taxable income and the policy. The next section studies a generic problem that allows the determination of taxes under limited endogenous information.

3.3 Optimal Control under Endogenous Signal Extraction (ESE)

Maximization problems such as (3.16) can be characterized as "optimal control under endogenous signal extraction". The key difficulty in (3.16) is that the policy choice $\mathcal{R}$ affects both the policy action τ as well as the distribution of the signal $s = l$. The signal provides information about the two unobserved exogenous shocks (γ, θ), so that the optimal policy depends on the conditional density $f_{(\gamma,\theta)|l}$, but this density depends itself on the tax policy $\mathcal{R}$. Therefore the choice of an optimal $\mathcal{R}$ must be consistent with the implied density $f_{A|s}$. To our knowledge this is the first paper to consider such a difficulty. The literature has considered setups where separation holds in the sense that $f_{A|s}$ does not depend on $\mathcal{R}$, see subsection 3.1.2 for a discussion. Here we provide a solution for a general setup.

Many problems in economics have this form. It can be thought of as a Stackelberg-reaction-function game where the "leader" (in section 3.2, the government) chooses its policy (or reaction) function $\mathcal{R}$ optimally given the reaction of the "follower" h, but h is given and independent of the choice on $\mathcal{R}$. Unlike standard Stackelberg games the actions τ and s are determined jointly in this setup, there is a hierarchy only in the way leader and follower choose the reaction functions $\mathcal{R}, h$. Simultaneity is the standard assumption in Ramsey equilibria, where the equilibrium allocations are influenced only by the policy action τ, not by the whole policy function. This is a natural assumption when, as is standard in Ramsey taxation, followers are atomistic and the signal is an aggregate variable.[8]

We now present a generic problem of optimal control under endogenous signal extraction. This generalizes the PI-Ramsey Problem without adding any difficulty to the proof. The generalization may be useful in other applications.

[8] Simultaneity also occurs in the literature on supply function equilibria (Klemperer and Meyer, 1989). Here firms simultaneously choose a supply function.

Consider a planner/government that chooses a policy variable $\tau \in \mathcal{T} \subset R$, observes endogenous signal $s \in \mathcal{S} \subset R$ at the time of choosing τ, when random variables $A \in \Phi \subset R^k$ have a given distribution F_A.[9]

The planner's objective is to maximize $\mathrm{E}\left[W\left(\tau, s, A\right)\right]$ for a given payoff function W. The planner knows that a value for the policy variable τ maps into endogenous signals through the following equation

$$y = h(\tau, A) \tag{3.18}$$

The government is assumed to know W, h, F_A, it does not observe the value of A.

This nests the case when other endogenous variables enter the objective function, as these can be embedded in W. We did this in section 3.2 through the use of the L_2^{imp} function defined in (3.13) to substitute out l_2 in the objective function.

Optimal behavior under uncertainty implies that the government chooses a policy contingent on the observed variable s, therefore the government's problem is to choose a policy function $\mathcal{R} : \mathcal{S} \to \mathcal{T}$ setting policy actions equal to

$$\tau = \mathcal{R}(s) \tag{3.19}$$

Our interest lies in solving for the case that the exact value of A can not be inferred from the observed s, the chosen τ, knowledge of h and of (3.18). This happens, generically, when the dimension of A is higher than the dimension of s ($k > 1$) and some of the variables in A have a continuous density. In terms of the fiscal policy example when the signal $s = l$ is observed, the values of θ, γ remain hidden even after the choice of τ has been made for a given observed labor l.

To summarize, we wish to solve the following model of **Optimal Control with**

[9] It is possible to generalize the problem to the case of multidimensional policy instruments and signals. However, notation becomes cumbersome, hence we only refer to the univariate case in the following.

Endogenous Signal Extraction:

$$\max_{\{\mathcal{R}:\mathcal{S}\to\mathcal{T}\}} \mathrm{E}\left[W\left(\tau, s, A\right)\right] \tag{3.20}$$
$$\text{s.t.} \qquad (3.18),\ (3.19)$$

To see why a standard first order condition does not apply, we can rewrite the objective function as follows

$$\int \mathrm{E}\left[W\left(\tau, \bar{s}, A\right) \middle| \, \bar{s}\right] \; f_s(\bar{s}) d\bar{s} \tag{3.21}$$

Taking derivatives with respect to τ and *if* $f_s(\bar{s})$ *could be taken as given* we would find the following optimality condition

$$\mathrm{E}\left[W_\tau + W_s h_\tau \middle| \, \bar{s}\right] = 0 \;\; \text{for all } \bar{s}. \tag{3.22}$$

In most applications in dynamic models this would be correct, but it is not the correct FOC in our case because, in general, $\mathcal{R}$ determines the density f_s. To see this notice that since s is determined implicitly by

$$s = h(\mathcal{R}(s), A)$$

$f_{s|A}$ depends on $\mathcal{R}$. Therefore $f_s = \int_A f_{s|A}\, f_A$ is also endogenous to $\mathcal{R}$. The derivative of f_s with respect to $\mathcal{R}$ should be taken into account in deriving optimality conditions, as we do below.

3.3.1 General first order conditions

Let us call $S(\mathcal{R}, A)$ the observable s induced by the shock A and a policy $\mathcal{R}$. Formally, $S(\mathcal{R}, A)$ is defined as follows: define H as

$$H(s, A; \mathcal{R}) \equiv s - h(\mathcal{R}(s), A). \tag{3.23}$$

then $H(s, A; \mathcal{R}) = 0$ gives the equilibrium value of s. We consider $\mathcal{R}$ such that $S(\mathcal{R}, A)$ is uniquely defined for all $A \in \Phi$.

The policy variable that is realized for each value of the shocks A and for a given policy function $\mathcal{R}$ is then given by

$$T(\mathcal{R}, A) = \mathcal{R}\left(S(\mathcal{R}, A)\right)$$

Notice the following distinction between the objects S, T and $\mathcal{R}$: the latter is a function of s while S and T are functions of $\mathcal{R}$ and the realizations of the shocks.

Let $\mathcal{F}$ be the value of the objective function for a given choice for $\mathcal{R}$.[10]

$$\mathcal{F}\left(\mathcal{R}\right) \equiv \mathrm{E}\left[W\left(T\left(\mathcal{R}, A\right), S\left(\mathcal{R}; A\right), A\right)\right] \tag{3.24}$$

We can now re-define the Optimal Control with Endogenous Signal Extraction problem as

$$\max_{\{\mathcal{R}:\mathcal{S}\to\mathcal{T}\}} \mathcal{F}(\mathcal{R}) \tag{3.25}$$

and denote its solution by $\mathcal{R}^*$.

3.3.2 Apparent PI: Invertibility

In some cases the government can still implement the FI policy even if it does not observe the shocks. This occurs whenever the information set of the government is invertible, allowing it to learn the true state of the economy A from observing the signal s.[11]

[10] Notice that $\mathcal{F}$ maps the space of functions into R. The expectation operator integrates over realizations of A using the government's perceived distribution of A, so that the above objective function is mathematically well defined given the above definitions for T, S and under standard boundedness conditions.

[11] In the literature of optimal contracts under private information and incentive compatibility constraints this is the standard assumption, which amounts to assuming full revelation, but those papers concentrate on the difficulties raised by the reaction of the agents' to the choice of $\mathcal{R}$, this reaction is the reason that the Full Information solution is not reached in those papers.

Invertibility also holds in the supply function literature (Klemperer and Meyer, 1989) because uncertainty is onedimensional.

To formally define Invertibility, consider the set of all possible values (τ, s) in the FI case, namely

$$M^{*} \equiv \left\{ (\tau, s) \in R^{2} : \tau = \mathcal{R}_{FI}^{*}\left(\overline{A}\right) \text{ and } s = h\left(\mathcal{R}_{FI}^{*}\left(\overline{A}\right), \overline{A}\right) \text{ for some } \overline{A} \in \Phi \right\} \quad (3.26)$$

Let M_{τ}^{*} (M_{y}^{*}) denote the projection of M^{*} on $\mathcal{T}$ $(\mathcal{S})$

Definition 5. *Invertibility holds if for any* $\overline{s} \in M_{s}^{*}$ *there exists a unique* $\tau \in M_{\tau}^{*}$

Clearly, Invertibility is satisfied if $h\left(\mathcal{R}^{FI*}(A), A\right) = \overline{s}$ defines implicitly a unique value for A for all $\overline{s}$.

Invertibility will often occur when the dimension of τ is the same as the dimension of A. Even if A is high-dimension, Invertibility also obtains if Φ is a finite set, in this case we can expect to be able to map an equilibrium into the shock since there are finitely many realizations, only by coincidence would the same equilibrium point (τ, l) occur for two different realizations of A.

Proposition 4. *Under Invertibility* $\mathcal{R}^{*} = \mathcal{R}^{FI*}$

This follows from the fact that the PI case is a restricted FI case, therefore the value in the PI case is less than or equal than the FI case, and the value of the FI case is achievable under Invertibility.

To illustrate a case of Invertibility, consider the example of section 3.2, assume that $\gamma = 1$ with certainty. The government does not observe the random value of θ and it has to choose taxes observing l. This is only apparently a PI problem, because the government can infer θ from observing the labor choice, hence the government can implement the FI policy.

3.3.3 General case of PI

The case of interest in this paper arises when knowledge of (τ, s) is not sufficient to back out the actual realizations of the shocks from (3.18). Observe that

Remark 1. *Invertibility is generally violated if*

- *A has higher dimension than s and some element of A has a continuous distribution or*
- *if $h\left(\mathcal{R}^{FI*}(\cdot),\cdot\right)$ is non-monotonic.*

In these cases the solution (3.20) should take into account that the distribution is endogenous to policy while, at the same time, the policy depends on the optimal filtering of the fundamental shocks A given the observed signal s.

Let $\mathcal{S}^*$ be the support of the random variable $S(\mathcal{R}^*, A)$. For the general case we will show that

Proposition 5. *Assume there is a solution $\mathcal{R}^*$ to the PI problem (3.20) and Assumptions 1-6 stated in Appendix A hold. Assume in addition that $S(\mathcal{R}^*; A)$ has a density. Let $\mathcal{S}^*$ be the support of $S(\mathcal{R}^*; A)$. The solution $\mathcal{R}^*$ satisfies the following necessary first order condition*

$$\mathrm{E}\left(\frac{W_\tau^* + W_s^* h_\tau^*}{1 - h_\tau^* \mathcal{R}^{*\prime}}\middle| S(\mathcal{R}^*; A) = \overline{s}\right) = 0 \tag{3.27}$$

for $\overline{s} \in \mathcal{S}^*$ in a set of probability one.

To find the optimality condition in Proposition 5 we use a variational argument. In the following paragraphs we show some of the key steps in the proof. For a detailed proof see Appendix A.

Take any function (a variation) $\delta : \mathcal{S}^* \to R$ and a constant $\alpha \in R$. Now consider reaction functions of the form $\mathcal{R}^* + \alpha\delta$. For a given δ consider solving the (one-dimensional) maximization problem

$$\max_{\alpha \in \Re} \mathcal{F}(\mathcal{R}^* + \alpha\delta) \tag{3.28}$$

In other words, now we maximize over small deviations of the optimal reaction function in the direction determined by δ. It is clear that

$$0 \in \arg\max_{\alpha \in \Re} \mathcal{F}(\mathcal{R}^* {+} \alpha\delta) \tag{3.29}$$

In Appendix A we show that $\frac{\partial \mathcal{F}(\mathcal{R}^*+\alpha\delta)}{\partial\alpha}$ evaluated at $\alpha = 0$ is

$$\mathrm{E}\left((W_\tau^* + W_s^* h_\tau^*) \frac{\delta}{1 - h_\tau^* \mathcal{R}^{*\prime}}\right) = 0 \tag{3.30}$$

This can be derived by carefully writing down all the derivatives involved.

The general definition of conditional expectation implies (3.27).

Notice that, as we anticipated, the first order condition (3.27) does not coincide with the FOC when separation holds (3.22). The term $\frac{1}{1-h_\tau^*\mathcal{R}^{*\prime}}$ acts as a kernel, or as a measure change, it is the new term relative to the standard case when $f_{A|s}$ does not depend on $\mathcal{R}$. The term $\frac{1}{1-h_\tau^*\mathcal{R}^{*\prime}}$ captures the effect of the choice of $\mathcal{R}$ on the density f_s that, as we anticipated, had to appear in some way in the optimality conditions.

With Proposition 5 in hand we can derive previous results available in the literature on optimal policy under Partial Information as special cases. The following corollary shows some cases that had been considered in the literature and that gave rise to separation.

Corollary 1. *(separation) Assume for some $\overline{s} \in \mathcal{S}^*$ one of the following hold*

1. *(Invertibility) there is a unique $A \in \Phi$ such that $S(\mathcal{R}^*, A) = \overline{s}$.*

2. *(exogenous signals or linearity) h_τ is given constant for all $A \in \Phi$ such that $S(\mathcal{R}^*, A) = \overline{s}$.*

Then the general FOC (3.27) reduces to the FOC under separation (3.22).

The proof is trivial: in both cases $\frac{1}{1-h_\tau^*\mathcal{R}^{*\prime}}$ is known given s so that this terms goes out of the conditional expectation in (3.27) and it cancels out. Case 1 obtains a generalization of the Invertibility theorem we found before. Case 2 is useful to discuss the various approaches to optimal policy under Partial Information that have been discussed in the literature. Consider the case of exogenous signals, when s is a function of A only, in this case $h_\tau = 0$ so case 2 applies. The armed-bandit problems are also a special case of this, since it is only previously determined policies that influence the signal, so that with respect to the current signal we still have $h_\tau = 0$. Note that the linear models that have been analyzed in the literature (Svensson and Woodford, 2004) arise as a special case of (3.27) when h_τ is a known constant due to linearity.

The condition (3.27) is useful for comparability with the standard FOC in the cases of corollary 1, but it turns out to be less convenient for computations. The reason is that (3.27) contains the derivative of the policy function to be computed $\mathcal{R}^{*\prime}$. An algorithm trying to approximate $\mathcal{R}^*$ numerically will have to ensure that not only $\mathcal{R}^*$ is well approximated but that its derivative is well approximated as well along the iterations. For this purpose it is more convenient to use an envelope condition that delivers the optimality condition that will be stated in Proposition 6.

Note that (3.27) conditions on s. If in addition we condition on some additional variables in A the remaining variables can be taken as given. More precisely, let us partition $A = (A_1, A_2)$ where $A_1 \in R$ and assume for all possible values of $\left(\overline{y}, \overline{A}_2\right)$ one can back out a unique value of A_1 that is compatible with $\left(\overline{s}, \overline{A}_2\right)$, so that

$$H\left(\overline{s}, A_1, \overline{A}_2; \mathcal{R}^*\right) = 0 \tag{3.31}$$

holds. This equation defines an implicit function $A_1 = \mathcal{A}^*(s, A_2)$ that, given the information on s, maps a realization of A_2 into the corresponding A_1 for the optimal policy.

Proposition 6. *The optimality condition (3.27) is equivalent to*

$$\int_{\Theta_2(\overline{s},\mathcal{R})} \frac{W_\tau^* + W_s^* h_\tau^*}{h_{A_1}^*} f_{A_1}(\mathcal{A}^*(\overline{s}, A_2)) f_{A_2}(A_2) dA_2 = 0 \; \textit{for} \;\; \textit{all} \;\; \overline{s} \tag{3.32}$$

where $\Theta_2\left(\overline{s}, \mathcal{R}^*\right)$ is the support of A_2's conditional on observing $\overline{s}$, stars denote that the partial derivatives and A_1 are evaluated at $\mathcal{A}^*\left(\overline{s}, \overline{A}_2\right)$.

This condition is much easier to use for computations than (3.27) because it does not involve the derivative of $\mathcal{R}^*$. The only cost is that it has to assume that the function $\mathcal{A}^*$ is uniquely defined. This will not only simplify computations but it also honors the title of the paper, as its derivation involves explicitly $f_{A|s}$ which is precisely the filter that is determined endogenously by the optimal choice of policy.

We now give the main steps of the proof. Using the notation $f(x)+g(x) = (f+g)(x)$ note that

$$\mathrm{E}\left(\frac{W_\tau^* + W_s^* h_\tau^*}{1-h_\tau^*\mathcal{R}^{*\prime}} \middle| S(\mathcal{R}^*, A) = \overline{s}, A_2 = \overline{A}_2 \right) = \frac{W_\tau^* + W_s^* h_\tau^*}{1-h_\tau^*\mathcal{R}^{*\prime}}(\overline{\tau}^*, \overline{s}, \mathcal{A}^*\left(\overline{s}, \overline{A}_2\right), \overline{A}_2).$$

By the law of iterated expectations (3.27) implies

$$\mathrm{E}\left(\frac{W_\tau^* + W_s^* h_\tau^*}{1-h_\tau^*\mathcal{R}^{*\prime}}(\overline{\tau}^*, \overline{s}, \mathcal{A}^*\left(\overline{s}, A_2\right), A_2) \middle| S(\mathcal{R}^*, A) = \overline{s} \right) = 0 \text{ for any } \overline{s} \in \mathcal{S}^*. \quad (3.33)$$

so that we can integrate over A_2's only as long as A_1 is substituted by $\mathcal{A}^*(s, A_2)$ whenever these shocks appear as arguments of W_τ^*, W_s^*, h_τ^* and f_{A_1}. Therefore this can be rewritten as

$$\int_{\Theta_2(\overline{s},\mathcal{R}^*)} \frac{W_\tau^* + W_s^* h_\tau^*}{1-h_\tau^*\mathcal{R}^{*\prime}}(\overline{\tau}^*, \overline{s}, \mathcal{A}^*\left(\overline{s}, A_2\right), A_2) f_{A_1}(\mathcal{A}^*(\overline{s}, A_2)) f_{A_2|\overline{s}}\, dA_2 = 0 \quad (3.34)$$

To find $f_{A_2|\overline{s}}$ proceed as follows. First we find $f_{s|A_2}$. Notice that given $A_2 = \overline{A}_2$ we have that $s = \mathcal{A}^{*-1}\left(A_1, \overline{A}_2\right)$ where $\mathcal{A}^{*-1}$ is the inverse function of $\mathcal{A}^*(\cdot, \overline{A}_2)$. By the change of variable rule

$$f_{s|A_2}\left(\overline{s}, \overline{A}_2\right) = f_{A_1}\left(\mathcal{A}^*(\overline{s}, \overline{A}_2)\right) \left|\mathcal{A}_s^*(\overline{s}, \overline{A}_2)\right| \quad (3.35)$$

where $\left|\mathcal{A}_s^*(\overline{s}, \overline{A}_2)\right|$ is the Jacobian of $\mathcal{A}^*(\cdot, \overline{A}_2)$. To find this Jacobian we apply once again the implicit function theorem to H and get

$$\mathcal{A}_s^*\left(\overline{s}, \overline{A}_2\right) = \left(h_{A_1}^*\right)^{-1}\left(1 - h_\tau^*\mathcal{R}^{*\prime}\right) \quad (3.36)$$

Plugging (3.36) into (3.35) gives $f_{s|A_2}$. A standard application of Bayes' rule gives $f_{A_2|\overline{s}}$ in terms of $f_{s|A_2}$ namely

$$f_{A_2|s} = \frac{f_{s|A_2} f_{A_2}}{\int f_{s|\overline{A}_2} f_{\overline{A}_2}\, d\overline{A}_2} \quad (3.37)$$

Plugging the above formula for $f_{A_2|\overline{s}}$ in (3.27) and since the denominator $f_s(s)$ drops out in the first order condition gives (3.32).

This proof highlights that $f_{A|\overline{s}}$ depends on $\mathcal{R}^{*\prime}$as it appears in the Jacobian in (3.36) and it also determines $\mathcal{A}^{*}$.Therefore we have that $f_{A|\overline{y}}$depends on the policy function $\mathcal{R}^{*}$and Bayes' rule implies that the signal extraction $f_{A_2|s}$ depends on the optimal policy $\mathcal{R}^{*}$. This justifies that we dub the maximization problem (3.20) and the title of the paper "endogenous filtering".

3.3.4 Linearity and "certainty equivalence"

We now show that if we consider a linear model (quadratic objective function and linear reaction function), we obtain the "certainty equivalence" result of Svensson and Woodford (2004).

Assume the objective function is

$$W(\tau, l) = -\frac{\omega_\tau}{2}\tau^2 - \frac{\omega_l l^2}{2} \tag{3.38}$$

and the reaction function

$$l = h + h_\tau \tau + h_\theta \theta + h_\gamma \gamma \tag{3.39}$$

where the ω's are positive coefficients and the h's are any real numbers.

Using Svensson and Woddford's (2004) notation, let $X \equiv (1, \theta, \gamma)'$. Then the Full Information optimal policy is

$$\tau = FX$$

where

$$F = \left(-\frac{\omega_l h_\tau h}{\omega_\tau + \omega_l h_\tau^2}, -\frac{\omega_l h_\tau h_\theta}{\omega_\tau + \omega_l h_\tau^2}, -\frac{\omega_l h_\tau h_\gamma}{\omega_\tau + \omega_l h_\tau^2} \right).$$

The FOC under PI (3.27) becomes

$$\mathrm{E}\left[\omega_\tau \tau + \omega_l h_\tau \left(h + h_\tau \tau + h_\theta \theta + h_\gamma \gamma\right) | l\right] = 0 \tag{3.40}$$

which can be rewritten as

$$\left(\omega_\tau + \omega_l h_\tau^2\right) \tau \omega_l h_\tau h + \omega_l h_\tau h_\theta \mathrm{E}\left[\theta | l\right] + \omega_l h_\tau h_\gamma \mathrm{E}\left[\gamma | l\right] = 0 \tag{3.41}$$

which implies that optimal policy is given by

$$\tau = F\mathrm{E}[X|l] \tag{3.42}$$

where F is the same vector of coefficients found under Full Information, that is, independently of the information structure and distribution of the shocks. This property of optimal policy with PI in linear models is called "certainty equivalence", as the government forms the best estimate of the state and behaves as if this estimate was certainty, or Full Information.

However, the "separation principle" does not hold, because one cannot compute the expectation of the state $\mathrm{E}[X|l]$ without knowledge of the policy. To see this, consider for example $\mathrm{E}[\theta|l]$, which can be rewritten as

$$\mathrm{E}[\theta|l] = \mathrm{E}[X|h + h_\tau F\mathrm{E}[X|l] + h_\theta\theta + h_\gamma\gamma = l] \tag{3.43}$$

This shows that optimal policy (3.42) and signal extraction (3.43) have to be solved jointly as a system.

3.3.5 Set of possible values and transversality condition

Let us know go back to the general non-linear case. So far we have ignored two issues that often complicate obtaining solutions to dynamic stochastic problems, namely, the ex-ante restriction of the set of possible values of choice variables and the transversality condition.

In many models of optimal choice under uncertainty it is important to know the range of possible values that endogenous variables can take. One reason this is important is that if the true solution is non-linear and approximations to non-linear functions can only be accurate on a compact set, one needs to know, at least roughly, what are the likely values of the equilibrium solution. A second reason is that it is possible that the solution may be simply undefined outside a certain range of the endogenous variable so that the algorithm may break down if it attempts to compute the solution outside that range.[12]

In our case and, in particular, in the two-dimensional uncertainty case posed by the fiscal policy example, the limits of l are easy to find ex-ante by exploiting the fact that the extreme values for l coincide with the FI case. More precisely, assume that l is monotonic in both γ and θ. In particular, assume that l is increasing in γ and decreasing in θ.[13] It is clear that, letting $l_{\min}$ and $l_{\max}$ be the extreme values of l in the PI solution, we have

$$\begin{aligned} l_{\min} &= L(\mathcal{R}^*; \theta_{max}, \gamma_{min}) = L^{FI}(\theta_{max}, \gamma_{min}) \\ l_{\max} &= L(\mathcal{R}^*; \theta_{\min}, \gamma_{\max}) = L^{FI}(\theta_{\min}, \gamma_{\max}) \end{aligned}$$

and the PI solution is in the interval $[l_{\min}, l_{\max}]$, which is trivial to compute.

Another issue is that in most infinite-dimensional optimization problems the first order conditions derived by applying δ variations to the policy function do not determine a unique policy function, one needs to add another condition that holds as a terminal or initial condition, often referred to as a "transversality condition". Since $\mathcal{R}^{*\prime}$ appears in (3.27) it appears that a transversality condition may be needed as the solution in a given value of s may depend on the solution for other values.

But there are two reasons why this is not a problem in our case. First, because there are two natural end conditions: using the above discussion about extreme values of l it is trivial to establish that at the extremes the optimal tax is given by the FI

[12]In fact, for applications of the promised utility approach to solving incentive problems, one needs to know the set of possible values to even be able to formulate a consistent recursive problem. This problem does not arise in our case, where the set of possible τ's can be restricted to a large set in the formalization of the maximization problem.

[13]This will be the case in the log-quadratic case shown in section 3.4. Such monotonicity can be easily proved in the FI case, although it should be checked with a candidate solution for $\mathcal{R}^*$.

solution so that

$$\begin{aligned}\mathcal{R}^*(l_{min}) &= \mathcal{R}^{FI}(\theta_{max}, \gamma_{min}) \\ \mathcal{R}^*(l_{\max}) &= \mathcal{R}^{FI}(\theta_{\min}, \gamma_{\max})\end{aligned}$$

give terminal and initial condition for $\mathcal{R}^*$ that are trivially computed.

Second, because when we rewrite the optimality condition as (3.32) the derivative $\mathcal{R}^{*\prime}$ does not appear so the optimal value for each s can be found independently of the solution in other points.

3.3.6 Algorithm

Given (3.32) it is easy to calculate the PI solution using the following numerical algorithm. Fix a value for $\overline{s}$. We must be able to compute $\mathcal{A}^*(\overline{s}, A_2)$ for a given candidate of $\overline{\tau} = \mathcal{R}^*(\overline{s})$. Then at a possible candidate $\overline{\tau}$ we can evaluate the integrand in (3.32) for each possible A_2 since we have the corresponding $A_1 = \mathcal{A}^*(\overline{s}, A_2)$. We compute the integral by running A_2 from the lowest to highest possible value of A_2 given the candidate $\overline{\tau}$. Note that these limits to the possible values of A_2, defining $\Theta_2(\overline{s}, \mathcal{R}^*)$, are endogenous to $\overline{\tau}$ and they have to keep A_1 within the admissible limits of the support of A_1. This operation maps a value of a candidate $\overline{\tau}$ to the left side of (3.32), the optimal $\tau = \mathcal{R}^*(\overline{s})$ is found by solving this non-linear equation that makes this integral as close as possible to zero.

3.4 Solution of fiscal policy example with ESE

3.4.1 Log-quadratic utility

We now study the optimal fiscal policy model introduced above using the tools developed in the previous section. For simplicity we study a special case that allows for an analytical reaction function h. Assume $u(c) = \log(c)$ and $v(l) = \frac{B}{2}l^2$. Both the productivity and the demand shock are temporary. In the second period, θ_2 is known to be equal to the mean of θ.

We show numerical examples with the following parameter values and distributional assumptions: let θ be uniformly distributed on a support $[\theta_{min}, \theta_{max}]$, γ uniformly distributed on $[\gamma_{min}, \gamma_{max}]$ and assume $\beta = .96$, B calibrated to get average hours equal to a third and government expenditure constant and equal to 25% of average output. The mean of γ is 1 and the mean of θ is 3. The supports of both shocks imply a range of $\pm 10\%$ from the mean.

We first present the FI solution in order to illustrate the optimal response of taxes and allocations to the two different shocks we consider. Notice that the equilibrium condition (3.6) becomes

$$Bl_1 c_1 = \gamma\theta_1(1 - \tau_1) \tag{3.44}$$

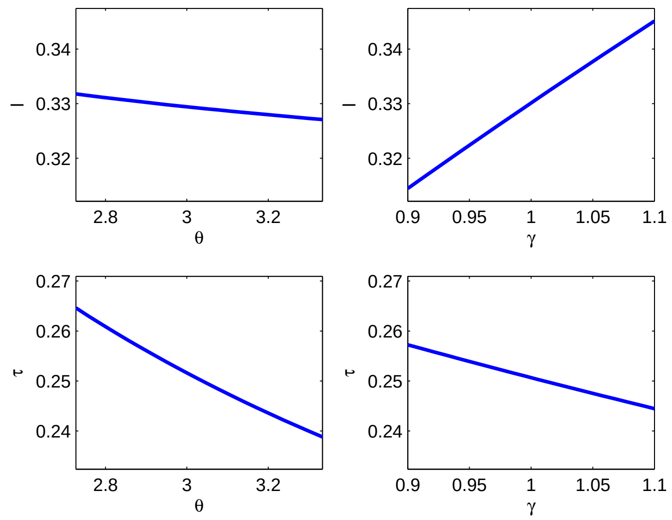
and after substituting out consumption using the resource constraint, we obtain that labor supply is the positive root of a quadratic equation, so that the reaction function (3.12) specializes to

$$l = h(\tau, \theta, \gamma) = \frac{Bg_1\theta^{-1} + \sqrt{(Bg_1)^2\theta^{-2} + 4B\gamma(1 - \tau)}}{2B}. \tag{3.45}$$

It is important to note that in general the productivity shock θ has two opposing effects: the substitution effect between leisure and consumption and the wealth effect, that acts in the opposite direction. With log-quadratic preferences, the second effect dominates and hence high realizations of θ will lead to low labor, ceteris paribus. On the other hand it can be seen from equation (3.45) that hours are increasing in the demand shock γ.

In Figure 3.1 we illustrate how hours and taxes move with the two different shocks under FI. On the left side of the figure, we keep γ constant and equal to its mean and we show that both hours and taxes are decreasing in the productivity shock. On the right side, we keep θ constant and equal to its mean and show that labor is increasing in γ, while taxes are decreasing. This shows that when we introduce PI with only hours being observed, if the government sees an increase in l, it would want to react in opposite directions depending on the source of the shock: this would call for a tax increase, if driven by low θ, or a tax cut if driven by high γ. Hence this model is particularly interesting to analyze optimal policy with endogenous PI since by observing a certain value $\bar{l}$ and imposing a tax rate $\bar{\tau}$ the government cannot infer the value of the shocks.

Figure 3.1: Hours and taxes with Full Information



Under PI, for an intermediate value of $\bar{l}$ the government is uncertain whether say both θ and γ are high, or vice versa, and in general there is a continuum of realizations $\left(\bar{\theta}, \widetilde{\gamma}(\bar{\theta}, \bar{l}; \mathcal{R})\right)$ consistent with the observation of $\bar{l}$ and a policy $\mathcal{R}$. Therefore it cannot choose the policy under Full Information (constant taxes) since the realizations of γ and θ enter separately in (3.17).

The partial derivatives h_τ and h_γ are easily obtained analytically. In particular

$$h_\tau(\tau, \theta, \gamma) = \frac{-1}{\sqrt{(Bg_1)^2 \left(\theta\gamma\right)^{-2} + \gamma^{-1} 4B(1-\tau)}}. \tag{3.46}$$

It is clear that both the productivity shock and the demand shock affect this slope, therefore endogenous signal extraction is an issue. Hence we proceed to find a $\mathcal{R}^*$ that satisfies (3.32) using the algorithm described in subsection 3.3.6

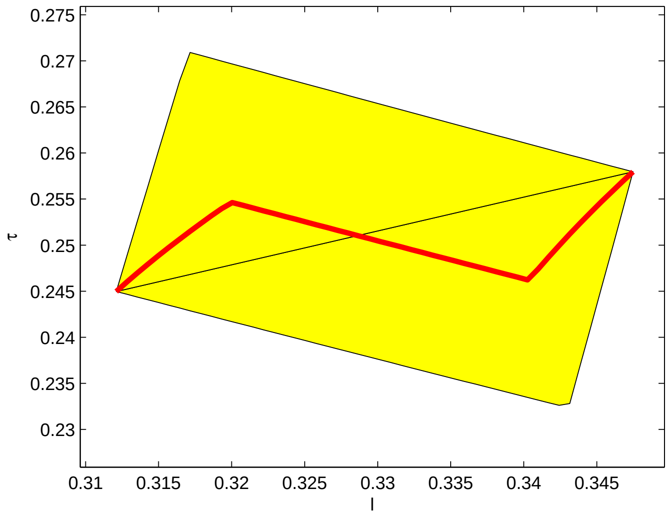
Figure 3.2 illustrates the optimal policy for this case, plotting the tax rate against observed labor. The red line is $\mathcal{R}^*$, while the yellow region is the set of all equilibrium pairs (l^{FI}, τ^{FI}) that could have been realized under Full Information.

As explained in subsection 3.3.5, the lowest (and highest) labor that is realized under FI is also the lowest (highest) value that can occur under PI and the PI tax is the same as the FI tax. In these extremes there is full revelation but anywhere between these two extremes the government has to choose a policy without knowing the values of γ, θ that give rise to equilibrium taxes or labor. It can be seen that the optimal policy calls for a tax rate in between the minimum and the maximum FI policies for each observation (but it is sometimes far from being the average of those tax rates.

For low labor, the government learns that productivity must be high, so the tax rate can be rather low. The lowest labor realization leads to the FI equilibrium for $(\theta_{max}, \gamma_{min})$. Then taxes start to increase: higher l's signal lower expected productivity and hence revenue, as the set of admissible θ's is gradually including lower and lower realizations. This goes on up to a point where the set of admissible θ's conditional on l is the whole set $[\theta_{min}, \theta_{max}]$. From that point on, the tax rate changes slope and becomes decreasing with respect to l. This is because now, with any θ being possible, increasing l signals an increasing expected revenue, hence allowing lower tax rates on average, up to the point where the highest θ's start being ruled out, at which point the policy becomes increasing again, up the full revelation point $l_{\max} = L^{FI}(\theta_{min}, \gamma_{max})$.

To gain further understanding on the implications of PI for the properties of the model, we plot again hours and taxes as functions of each shock individually in Figure 3.3. In all four panels, we reproduce the FI outcomes shown in Figure (FI)

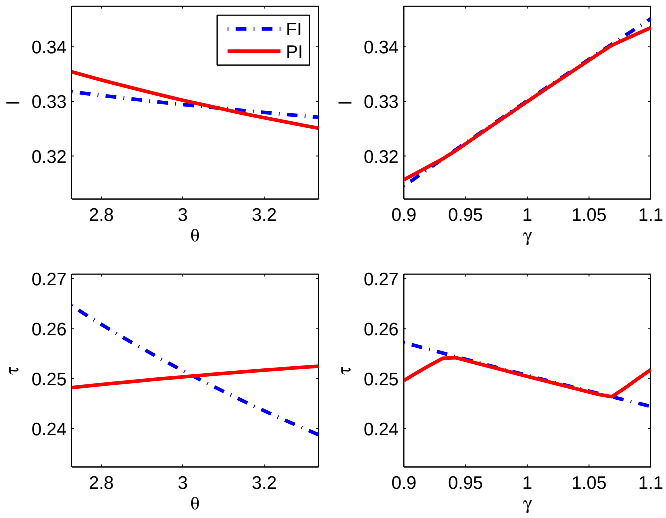
Figure 3.2: Optimal policy with log-quadratic utility



(blue dashed-dotted lines). The red lines represent the PI outcomes. For instance in the left panels we keep γ equal to its mean and we plot hours and taxes as functions of θ. Of course the government does not observe the values of θ and γ, but only hours. Interestingly, it can be seen that hours become more volatile in response to productivity shocks under PI and taxes become smoother and change the sign of their response to θ. This is because under this parametrization the government learns little about the realizations of θ and hence optimally chooses to cut taxes as hours increase.[14] On the right-hand side, we plot again hours and taxes as functions on γ, keeping θ equal to the mean. For intermediate values of γ, the government is relatively confident about the realization of the demand shock, hence the policies under FI and PI are very close. However for extreme realizations the government is fooled about which shock is driving hours, hence it cuts taxes for very low γ's and increases taxes for very high γ's, believing that changes in productivity are responsible

[14]We will see in the next subsection that this property of the solution will change with higher government expenditure.

Figure 3.3: Hours and taxes with Partial Information and Full Information



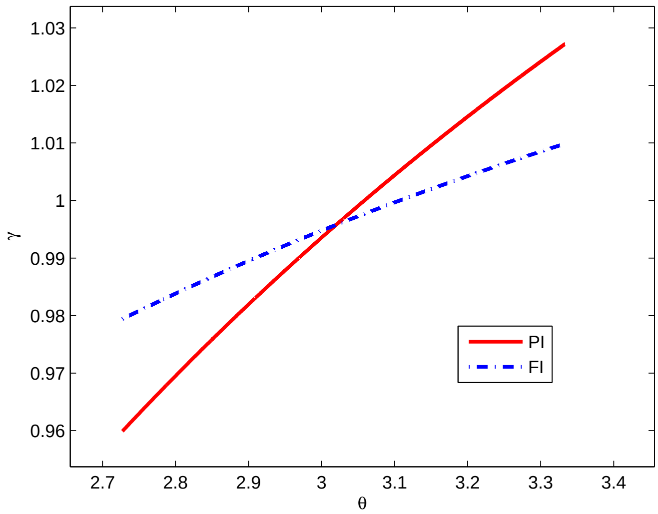
for the observed behavior of hours.

We also plot the locus of admissible realization of shocks for $\bar{l} = .33$ in Figure 3.4. The wealth effect of productivity makes it an increasing function in the (θ, γ) space. Now conditional on l, we can have combinations of high productivity (low wealth effect on labor supply) and high demand or low productivity and low demand.

Optimal policy with PI calls for a substantial smoothing of taxes across states. This can be seen in Figure 3.5, where the equilibrium cumulative distribution function of tax rates under PI (red line) is contrasted with the one obtained under FI (blue dotted line). This result is rather intuitive and it carries a general lesson for optimal fiscal policy decisions under uncertainty: When the government is not sure about what type of disturbance is hitting the economy, it seems sensible to choose a policy that is not too aggressive in any direction and just aims at keeping the budget under control on average.

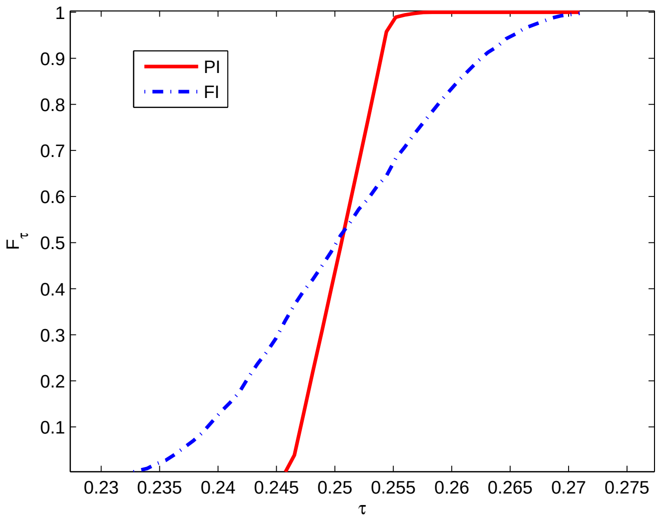
In our model, this smoothing of taxes across states will imply a larger variance of tax rates in the second period with respect to the FI policy. In the second period,

Figure 3.4: Set of admissible shocks



all the uncertainty is resolved and the tax rate will be whatever is needed to balance the budget constraint. This is of course taken into account at the time of choosing a policy under uncertainty, so that we could say that optimal policy is very prudent while the source of the observed aggregate variables is not known and then responsive after uncertainty has been resolved. In this sense, this model can rationalize the slow reaction of some governments to big shocks like the current recession. The Spanish example in the latest recession is a case in point. In 2008, it was far from clear how persistent the downturn would be and also whether is was demand-driven or productivity-driven and the government did not adjust its fiscal stance quickly, only to make large adjustments in the subsequent years. We will discuss this further in the paper.

Figure 3.5: Equilibrium CDF of tax rates



3.4.2 Close to the top of the Laffer curve

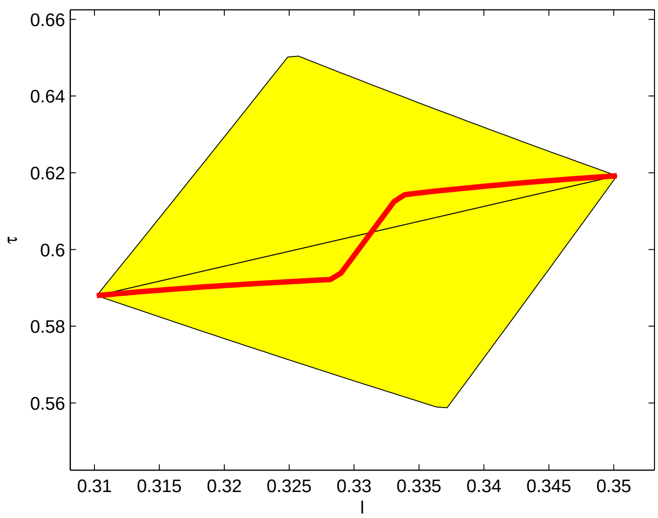
Let us now look at the case where government expenditure is very high, equal to 60% of average output in both periods.[15] We will see that this leads to a very non-linear optimal policy and to an exception to tax-smoothing across states. This example is of interest for several reasons. From an economic point of view, private information is of more importance here: since the government needs to balance the budget in the second period it is thus now very concerned about the possibility of a very low level of productivity θ, as in this case tax revenue is low in the first period and a large amount of debt will need to be issued in this case. A high debt, combined with high future expenditure, may call for very high taxes in the future, it could even mean getting the economy closer to the top of the Laffer curve, where taxation is most distortionary and hence consumption is very low. This example will also be of interest because the PI solution has some very different features from the FI outcome.

Figure 3.6 shows optimal policy for this case (red line), again contrasted with the

[15] All other assumptions on preferences and shocks are the same as in the previous section.

set of tax-labor outcomes under FI (yellow region).

Figure 3.6: Optimal policy with high government expenditure



The figure shows that the optimal solution is highly non-linear. The derivative $\mathcal{R}^{*\prime}$ is positive and relatively high in a middle range of levels of l, but both to the left and to the right of this middle range $\mathcal{R}^{*}$ it is much flatter. Notice that this is the opposite of what happens with a low level of g in the previous subsection. When government expenditure is sufficiently low, the government is very uncertain about the true realization of θ. Hence higher labor does not allow a more precise signal extraction about productivity. On the other hand, when g is sufficiently high, there is an intermediate region of observables where the government becomes confident about low realizations of θ. In Appendix B we prove this result by illustrating how g affects the slope of the loci of realizations of the shocks.

To illustrate how the PI policy involves a relatively precise signal extraction on θ with high government expenditure, consider the sets of possible realizations of θ under FI in Figure 3.7 and under PI in Figure 3.8.

Consider Figure 3.7 first. It can be seen that under FI any realization of θ is

Figure 3.7: Set of admissible θ's with FI

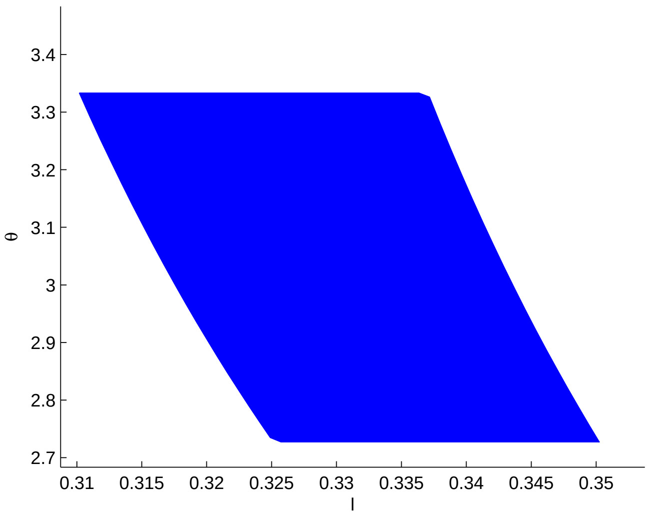
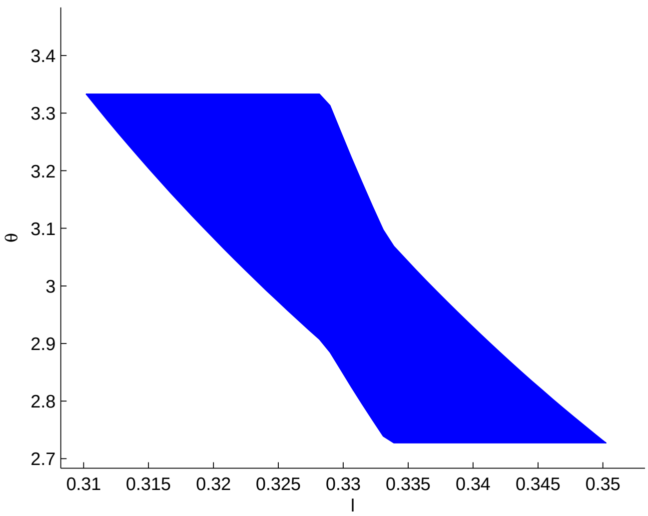


Figure 3.8: Set of admissible θ's with PI

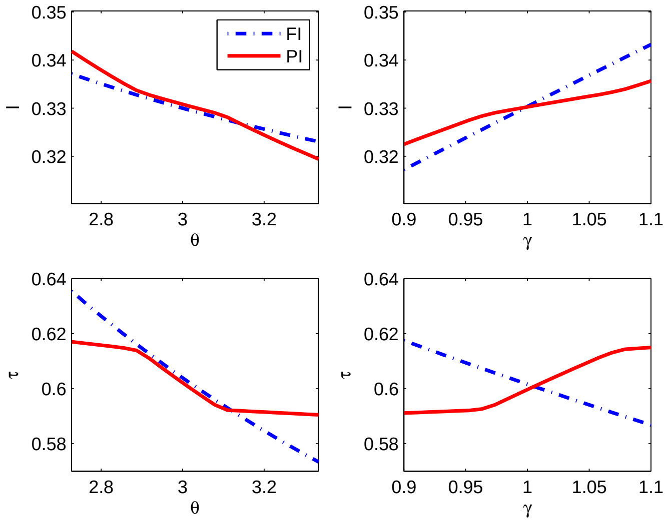


consistent with an intermediate realization of l, but each of these θ's would call for a different tax rate. However, under PI, there can only be one tax rate for each observed l and the government uses this policy to extract information on θ.

To see this, consider now Figure 3.8. The minimum value of l is only consistent with the highest possible θ (and lowest possible γ) because the wealth effect dominates. Under PI, increasing l from this point, the government becomes uncertain and lower realizations of productivity become consistent with the observations. At first, uncertainty is rising with l, but in the intermediate region of l's the government becomes more and more confident about low realizations of productivity. This leads to the sharp increase in the tax rate, which in turn gives rise to feedback effect on the set of possible θ's: high taxes discourage work effort, so higher labor now is an even stronger signal of low θ (high marginal utility from consumption). In this way an optimal policy and a conditional distribution of shocks consistent with it confirm each other in equilibrium.

Consistently with this analysis of the signal extraction, we also plot hours and taxes as functions of each shock individually, and we contrast the PI outcomes with the FI solution in Figure 3.9. On the left-hand side we consider productivity shocks only. As illustrated above, in the intermediate region of l's the government has a precise signal about θ, hence PI and FI policies and allocations are very close to each other. However for extreme realizations of θ the government is fooled about the source of the fluctuations and does hardly respond to productivity. On the right-hand side we consider only demand shocks. It can be seen that the PI government has very imprecise information about γ. Hence it responds to these shocks with the opposite slope with respect to the FI government. The case of high government expenditure shows that optimal policy with PI can be very non-linear in order to avoid the worst outcomes, e.g. in the model hitting the top of the Laffer curve or in the real world a debt crisis. As shown in the previous subsection, when expenditure is low and there are no concerns related to the government budget constraint, policy has to be smooth, but when there are contingencies that are particularly dangerous for agents, then optimal policy calls for being very reactive to observables in order to prevent those cases to materialize. This is exemplified by the optimality of increasing taxes

Figure 3.9: Hours and taxes with PI and FI: high g



steeply in the first period to avoid having to distort the economy too heavily in the second period if realized productivity turn out to be low (and hence the fiscal deficit turns out to be high). This lesson seems relevant for the understanding of the fiscal policy reaction to the financial crisis in 2008 and afterwards, especially in countries like Spain and Italy, that arguably where in danger of getting close to the top of the Laffer curve, as testified by the fact that significant increases in taxes after 2009 did not raise the amount of revenue as much as it was desired by these governments.

3.4.3 Linear approximation

We now compare our solution to existing methods based on linear-quadratic optimization (Svensson and Woodford, 2004). In order to do so, we modify our distributional assumption and we assume that both θ and γ are normally distributed. We then truncate these distributions at three standard deviations from the mean in order to have a bounded support for the shocks in our solution method. The standard deviation of each of the shocks is assumed to be 3% of the mean.

In order to compute the linear approximation, we take a second-order approximation of the objective function and a first-order approximation of the reaction function h around the allocation and policy that arises under Full Information when the shocks take their mean value. Then, we compute the certainty equivalent policy as described in subsection 3.3.4. Importantly, this policy can be found under Full Information and then applied to the Partial Information case by simply computing the conditional mean of the shocks for each value of $\bar{l}$. Figures 3.10 and 3.11 compare the optimal policy and the linear approximation in the case of low g and high g respectively. It can be seen that the approximation is quite accurate for intermediate realizations of labor, but less so for extreme values. This suggests that linear approximations can be misleading when there is endogenous Partial Information and the economy is hit by large shocks.

Figure 3.10: Linear approximation, low g

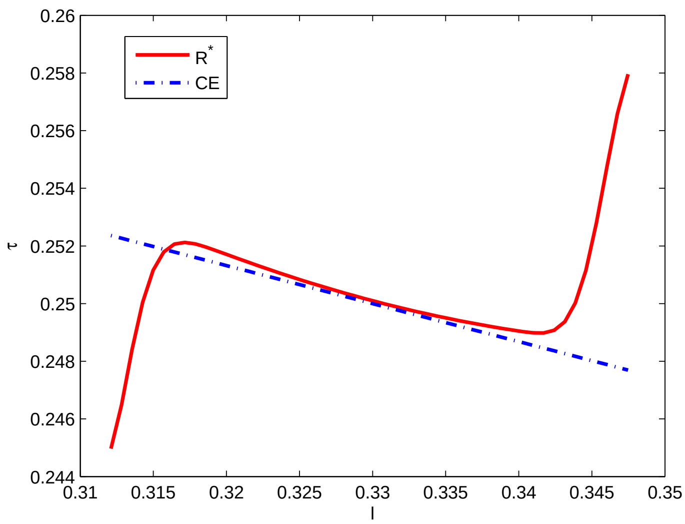
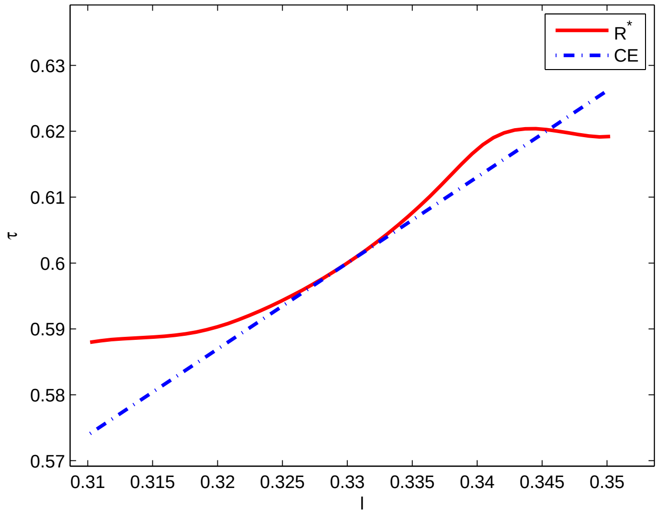


Figure 3.11: Linear approximation, high g



3.5 Robustness

3.5.1 Risk aversion and precautionary fiscal adjustments

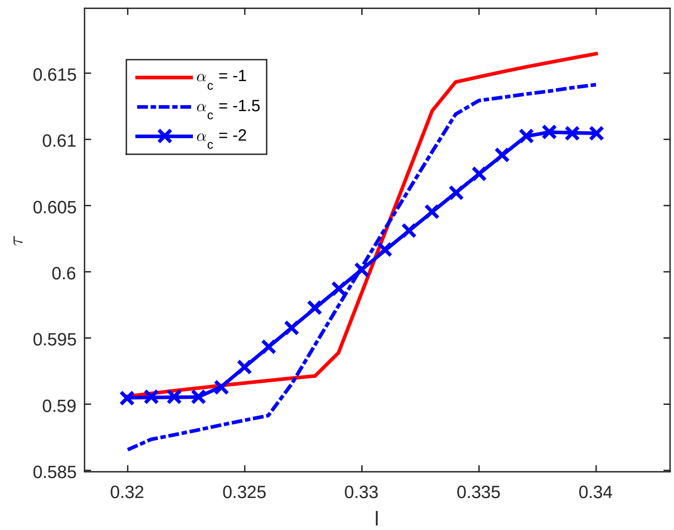
As we have seen in the previous section, when the economy is close to the top of the Laffer curve, optimal policy is very non-linear in the observable variable, creating a region of sharp fiscal adjustments for intermediate realized values of labor. The government raises taxes dramatically in the first period, in order to prevent the worst scenarios with low productivity, high taxes and low consumption in the second period. In this section, we consider a general CRRA utility function from consumption

$$u(c) = \frac{c^{1+\alpha_c}}{1+\alpha_c}$$

and investigate how optimal policy changes with different degrees of the risk-aversion parameter α_c, while keeping quadratic disutility from labor effort as in our baseline parametrization.

Figure 3.12 illustrates the optimal policy for $\alpha_c = -1$ (log utility, that is the baseline case), $\alpha_c = -1.5$ and $\alpha_c = -2$. It can be seen that as risk aversion increases, the area where policy is more reactive of observables becomes wider, and the government reacts strongly even for weaker signals of a recession. Intuitively, this is because the government wants to avoid contingencies with high debt that would lead to high taxes in the second period. The more risk averse the agent is, the more painful it is to be in those states, where consumption has to be cut substantially. However, this larger region of reaction also makes the policy function less steep, as can be seen from the picture.

Figure 3.12: Changing risk aversion



3.5.2 Permanent shocks and linear-quadratic utility

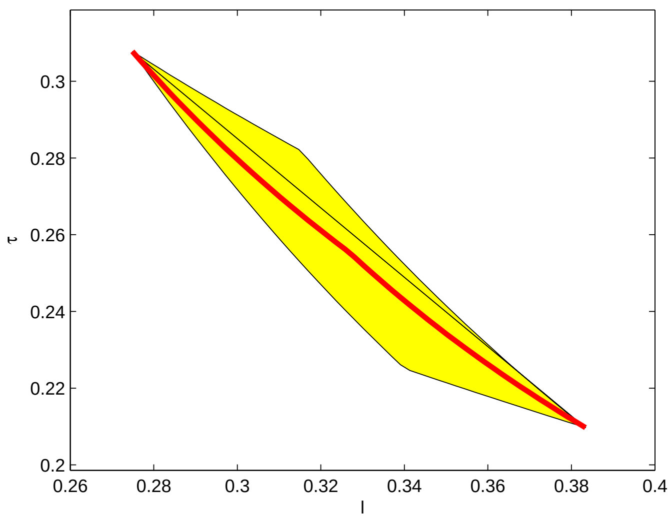
We now consider a permanent productivity shock, that is $\theta_2 = \theta_1$. This would be uninteresting with log-quadratic preferences because hours would become independent of θ as income and substitution effects would cancel out. Therefore we assume linear utility from consumption $u(c) = c$ and quadratic disutility from labor $v(l) = \frac{B}{2}l^2$, which give another case with an analytical solution for the reaction function h and its derivatives.

In particular, it is easy to see from the first order condition (3.6) that the reaction function (3.12) specializes to

$$l = h(\tau, \theta, \gamma) = \frac{\gamma\theta}{B}(1 - \tau), \tag{3.47}$$

The optimal policy is illustrated in Figure 3.13 and is compared with the set of FI tax rates conditional on l. In general, $\mathcal{R}^*$ is decreasing as higher observed labor suggests higher conditional expectation for productivity, hence allowing to balance the intertemporal budget constraint with a lower distortionary tax. The figure also compares the optimal policy with a linear policy obtained connecting the two full revelation points with a straight line. While the optimal policy is not quite linear, in this example a linear approximation would not be too wrong.

Figure 3.13: Optimal policy with linear-quadratic utility



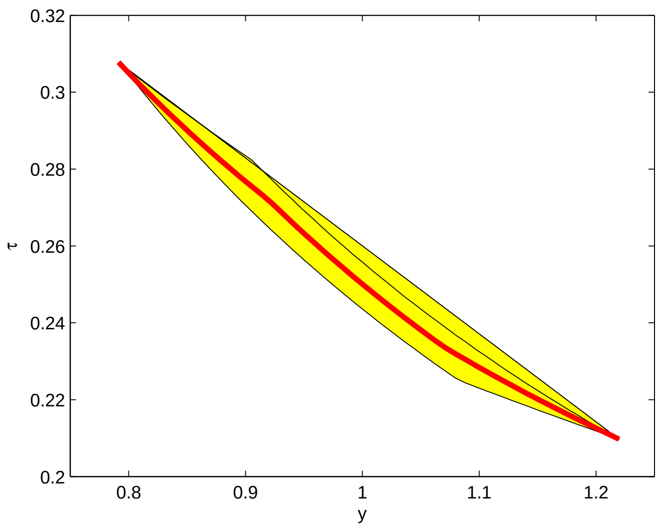
3.5.3 Observable output

Our previous example assumed that labor was observable. This made notation and presentation easier as the shocks are not involved in the signal directly. We now consider the case when the signal observed by the government is the total product, so

$s = y = \theta l$. The government cannot sort out the values of the θ and l independently, hence there still is a problem of signal extraction as the government can not back out the values of (θ, γ). In addition we assume again linear-quadratic utility and a permanent productivity shock.

Figure 3.14 illustrates the optimal policy for this case. It can be seen that the result is remarkably similar to that obtained in the previous subsection with observable labor. However, in this case the government has a lot more information than in the previous case. This is because current revenue $\tau\theta l$ is known, and hence there is no uncertainty about the amount of debt that needs to be issued. The only uncertainty is about the amount of revenue that will be collected in the future, as the value of (permanent) productivity is unknown.[16] As we have seen, uncertainty about debt is key to get large fiscal adjustments as in subsection 3.4.2.

Figure 3.14: Optimal policy with output observed



[16]Note that if we had assumed, as in section 3.4, that θ_2 is not random the model would not be interesting, since in that case there would be no uncertainty about future revenue.

3.6 An infinite-horizon model with debt

In this section we present an infinite horizon version of the optimal fiscal policy model we have considered. We will see that some key intuitions developed in the two-period model are still present and they lead to interesting dynamics. In particular, under PI the government sometimes reacts slowly to recessions and as a consequence needs to raise taxes for a longer time endogenously prolonging slumps. In this case the slow reaction and the ensuing deepening of the recession is fully desirable.

As is well known, in the case of non-linear utility current interest rate is determined by future taxes, hence the optimal policy under full commitment would be time inconsistent, leading to some complications in the solution of optimal policy under full commitment.[17] In order to avoid these difficulties we assume a linear utility of consumption.

We also assume that the shocks are iid over time as this reduces the number of state variables and it abstracts from the issues of government experimentation that has been studied in the armed-bandit literature of optimal policy that we discussed in subsection 3.1.2.[18] In this way we are left with the simplest infinite horizon model of fiscal policy where endogenous signal extraction plays a role.[19]

3.6.1 Full Information

Our model under Full Information is a small variation of Example 2 of Aiyagari et al. (2002), with linear utility from consumption and standard convex disutility from labor effort. Preferences of the representative agent are given by:

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\gamma_t c_t - v(l_t)\right] \tag{3.48}$$

[17]Under Full Information and uncertainty this issue was first addressed in Aiyagari et al. (2002).

[18]For example Wieland (2000a, 2000b), Kiefer and Nyarko (1989), Ellison and Valla (2001)

[19]Combining this with issues of commitment and optimal experimentation is of interest but we leave it for future work.

where γ_t is a demand shock, iid over time.

The period t budget constraint of the representative agent is

$$c_t + q_t b_t = \theta_t l_t (1 - \tau_t) + b_{t-1} \tag{3.49}$$

where θ_t is an iid productivity shock. Note that the government can only issue real riskless bonds b_t.

The standard first order conditions for utility maximization are

$$\frac{v'(l_t)}{\gamma_t} = \theta_t (1 - \tau_t) \tag{3.50}$$

and

$$q_t = \beta \frac{\bar{\gamma}}{\gamma_t}. \tag{3.51}$$

where $\bar{\gamma}$ is the unconditional expectation of the demand shock γ.

The Ramsey government finances a constant stream of expenditure $g_t = g \; \forall t$ and chooses taxes and non-contingent one-period debt in order to maximize utility of the agent subject to the above competitive equilibrium conditions as well as the resource constraint $c_t + g = \theta_t l_t$. Under FI, the government can choose a sequence of taxes conditional on a sequence of shocks $A^t = (A_t, A_{t-1}, ..., A_0)$, where $A_t = (\theta_t, \gamma_t)$.

The period-t implementability constraint is

$$b_{t-1} = c_t - \frac{v'(l_t)}{\gamma_t} l_t + \beta \frac{\bar{\gamma}}{\gamma_t} b_t. \tag{3.52}$$

We now introduce an upper bound on debt, b_{max}. We will assume that whenever debt goes above this threshold, the government pays a quadratic utility cost $\beta \frac{\chi}{2}(b_t - b^{max})^2$ and we will set the parameter χ to be a very high number in order to mimic a model with an occasionally binding borrowing constraint while still retaining differentiability of the problem.

The first order conditions for Ramsey allocations with respect to hours and debt are:

$$\gamma_t \theta_t - v'(l_t) + \lambda_t \theta_t - \frac{\lambda_t}{\gamma_t} \left[v'(l_t) + v''(l_t) l_t \right] = 0 \tag{3.53}$$

and

$$\lambda_t \frac{\bar{\gamma}}{\gamma_t} = \mathrm{E}_t \lambda_{t+1} + \chi(b_t - b^{max}) I_{[b^{\max},\infty)}(b_t). \tag{3.54}$$

where λ_t is the Lagrange multiplier of constraint (3.49) and we denote by $I_{[b^{\max},\infty)}(b)$ the indicator function for the event $b > b^{max}$.

Thanks to the assumption of linear utility from consumption, the Ramsey policy is time-consistent and allocations satisfy a Bellman equation that defines a value function $W^{FI}(b_{t-1}, A_t)$. Thus optimal taxes are given by a time-invariant policy function $\tau_t = \mathcal{R}^{FI}(b_{t-1}, A_t)$

3.6.2 Partial Information

We start the description of the PI problem by specifying its timing. At the beginning of each period t, the Ramsey government observes the realization of the exogenous shocks of last period A_{t-1}, the value of its outstanding debt b_{t-1} and the realization of current labor l_t. Based on this information, but before knowing the value of A_t, it sets the tax rate τ_t. Formally, the choice of taxes at time t is contingent on - i.e. a function of - (A^{t-1}, l_t).

Note that because of the iid assumption on the shocks A_t, information about outstanding debt summarizes all the information about past realizations that is relevant in terms of the objective function and the constraints of the Ramsey problem. In other words, the government cares about past realizations of the exogenous shocks only to the extent that they affect the level of current outstanding debt. As a consequence, debt is a sufficient state variable in addition to the current observed signal l_t. Hence the optimal policy has a recursive structure and taxes are given by a policy function $\tau_t = \mathcal{R}(b_{t-1}, l_t)$.

Define W as the value of the utility (3.48) at the optimal choice for given debt before seeing the realization of l_0. By a standard argument, the choice from period 1 onwards is feasible from period 0 onwards given the same level of debt. Therefore W satisfies the following Bellman equation

$$W(b) = \max_{\mathcal{R}\colon \Re^2 \to \Re_+} \mathrm{E}\left[\gamma(\theta l - g) - v(l) + \beta W\left(\frac{(b + g - \theta l + \frac{v'(l)l}{\gamma})\gamma}{\beta\bar{\gamma}}\right) + \right.$$

$$\left. - \beta\frac{\chi}{2}(\frac{(b + g - \theta l + \frac{v'(l)l}{\gamma})\gamma}{\beta\bar{\gamma}} - b^{max})^2\right] \tag{3.55}$$

where l satisfies $l = h(\mathcal{R}(b, l), \theta, \gamma)$ where $h(\tau, \theta, \gamma)$ is the labor that satisfies (3.50).

The only difference with respect to the reaction function in the two-period model is that now the government should recognize that debt affects labor indirectly through the tax rate.

At this point it is important to pause the maths for a second and discuss how shocks and PI influence the optimal choice of taxes. Under incomplete markets a sequence of adverse shocks (low θ) will lead to an increase in debt. This will be more so under PI than under FI, because under PI the government only learns that a low θ and, therefore, a low tax revenue, occurred with a delay. The reason that b is an argument in $\mathcal{R}$ is that in the presence of incomplete markets, debt piles up after a few bad shocks, much more so than under FI, therefore the government will have to increase the level of taxes for a given l_t to avoid debt from becoming unsustainable.

Note that in (3.55) we have substituted future debt using the budget constraint (3.52). It is important to highlight a key difference with respect to the FI problem: while in that case a choice of τ_t implied a choice of b_t, now, b_t is a random variable even for a given choice of τ_t. In other words, just like in the two-period model, the government is uncertain about how much debt will need to be issued and in particular must take into account that bad realizations of productivity may lead to a debt level above b^{max}, if taxes are not sufficiently high.

In order to solve the model, we exploit its recursive structure, by solving for the PI first order condition at each point on a grid for debt and iterating on the value function of the problem. To see how this works, consider the objective function defined by the right-hand side of (3.55).

For a given guess for the value function, this is just a function of observed labor to which we can apply the main theorem of the paper (Proposition 5) and obtain the general first order condition with PI.[20]

This first order condition involves the derivative $W'(b)$. In standard dynamic programming it is well known that an envelope condition applies that allows the simplification of the derivative of the value function. In Appendix C we show that an analogous envelope condition holds under our PI model so that

$$W'(b) = \mathrm{E}\left[\frac{\gamma}{\bar{\gamma}} W'(b')\right] - \chi(b' - b^{max}) I_{[b^{\max},\infty)}(b'). \tag{3.56}$$

Hence by solving the first order condition using (3.56) and iterating on the Bellman equation (3.55), we can approximate the optimal policy. In the next subsection, we show some numerical results obtained after parametrizing the economy. While the model is not meant to be a quantitative model of fiscal policy, it can nonetheless rationalize important features of the fiscal response to the Great Recession, with slow and large fiscal adjustments inducing protracted slumps.

3.6.3 Numerical results

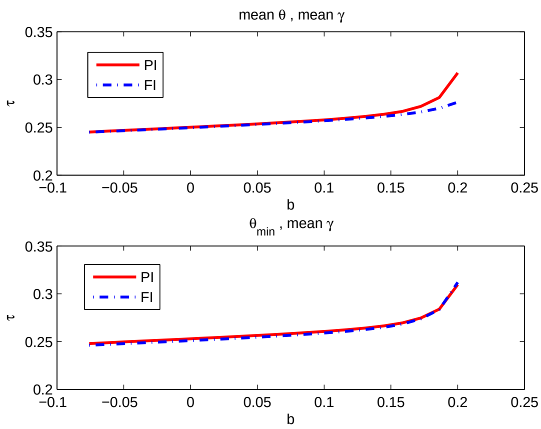
In order to parametrize the economy we assume quadratic disutility from labor, and the other parameters are as in the two-period model. The shocks are uniformly distributed on a support of ± 5 % from their means, implying a volatility of 2.89%. The debt limit is 20 % of mean output.

Figure 3.15 illustrates a key property of optimal policy by showing the tax policy as a function of outstanding debt, for two different realization of the shocks: in both panels we keep γ equal to its means, but while in the upper panel θ is also equal to the mean, in the lower panel θ is equal to its minimum value. We contrast optimal policy under PI with the FI counterpart (dashed-dotted line). In the upper panel, we can observe a non-linearity of optimal policy with respect to debt. When the

[20]The FOC is explicitly shown in Appendix C.

economy gets closer to the debt limit, the PI policy calls for a significantly larger increase in taxes than the FI when the shocks are equal to their mean. This is because observed labor takes an intermediate value for this realization of the shocks. Hence, the government is uncertain about what combination of the shocks has been realized and needs to avoid exceeding the debt constraint for all possible realizations of θ's. On the other hand the FI government knows that a lower tax rate is sufficient to avoid a fiscal crisis for this combination of the shocks. In the lower panel, on the other hand, signal extraction under PI is more precise, as the economy is hit by a large negative productivity shock and both PI and FI call for similar tax responses, independently of the debt level.

Figure 3.15: Tax policy as a function of debt



We now show impulse response functions for our economy.[21] In particular, we engineer two different scenarios that give rise to a 1% fall in observed hours. With linear utility from consumption, hours are increasing in θ, differently from the two-period model. Hence a fall in hours can be driven by either low θ and high γ or vice versa. In the first case (Figure 3.16), the shocks hitting the economy are a negative

[21]These non-linear impulse response functions are computed as percentage deviations from the path that would arise absent all shocks.

Figure 3.16: Impulse response function: low θ, high γ

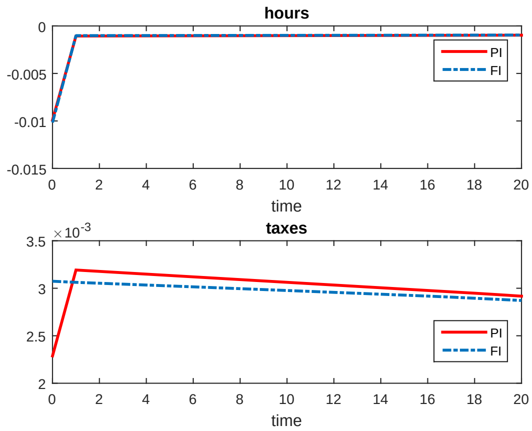
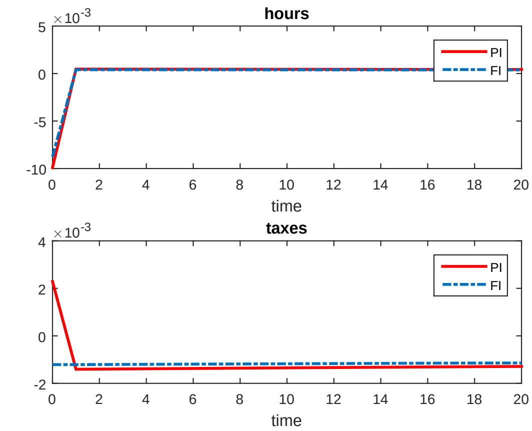


Figure 3.17: Impulse response function: high θ, low γ



θ shock combined with a (smaller) positive γ shock. In the second case, we consider the opposite combination of θ and γ.

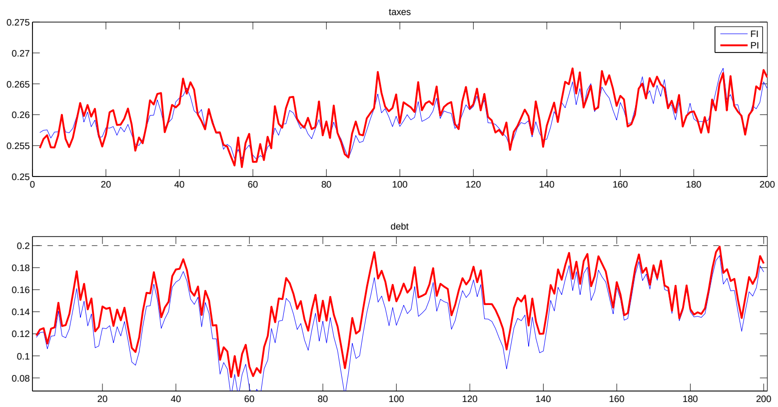
When the true nature of the recession is productivity, the tax response under PI is smaller then under FI. Hence, the government needs to respond by more in the second period, after the nature of the shock becomes known. Eventually taxes increase by more than under FI and stay high for a long time, inducing a (slightly) more persistent fall in hours and output. This behavior of the economy is qualitatively similar to what happened in some European countries after the financial crisis (e.g. Spain), where an initial slow reaction, or even an expansionary policy, has been followed by a necessary large fiscal adjustment and the recovery has so far been very slow and weak.

On the other hand, when the true nature of the fall in hours is a demand shock (Figure 3.17), the PI government reacts in the wrong direction, increasing taxes, while the FI government cuts them. This policy is reversed in the second period, after the past realization of the shocks become known.

While the PI policy is optimal in our setup where only current income can be taxed, the above findings suggest that allowing for retrospective taxation could improve welfare. Society would be better off if the government could adjust taxes on past income after observing the realization of past shocks and consumers knew of this possibility. However, retrospective taxation might not be easily implementable in the real world due to time-consistency issues, as ex-post surprising taxes on past income are non-distortionary.

Figure 3.18 illustrates a long stochastic simulation of the model. It is easy to see that taxes are very responsive to debt. One interesting question is whether taxes are smoother or more volatile under PI with respect to FI. Intuitively, there seem to be two opposing forces. On the one hand, the PI government does not observe the shocks, and hence smooths its policy across states for a given debt level. However, this policy induces necessary fiscal adjustments following the dynamics of debt, so that this pushes towards higher volatility under PI. The results from long simulations is that this second effect seems to dominate (although slightly) and the FI government is more successful than the PI government at smoothing tax rates. It can be seen that often when debt gets close to the borrowing limit (20% of mean output) the PI government imposes larger fiscal adjustments. This can be rationalized in analogy with the example of the two-period economy close to the top of the Laffer curve. Fear of future large required adjustments in the event of low θ lead the PI government to raises taxes significantly.

Figure 3.18: Simulation



3.7 Conclusion

We derive a method to solve models of optimal policy with Partial Information without any separation assumption between the optimization and signal extraction problem. In our model the optimal decision influences the distribution of the shocks conditional on the observed endogenous signal, therefore the signal extraction and optimization problem need to be solved consistently and simultaneously. The method works in general and we show algorithms that solve these problems using standard ideas for solving dynamic models. We also show that Partial Information on endogenous variables matters as some revealing non-linearities appear in very simple models. These non-linearities are due to the fact that in different regions of the observed signal the information revealed about the underlying state changes in a non-linear fashion even if the model is not highly non-linear.

Optimal fiscal policy under endogenous signal extraction calls for smooth tax rates across states when the government budget is under control, and for regions of large response to aggregate data when the economy is close to the top of the Laffer curve or to a borrowing limit. Uncertainty about the state of the economy helps to understand the slow reaction of some European governments to the Great Recession, followed by sharp fiscal adjustments and prolonged downturns: as the signal worsens and it becomes more consistent with a slump in productivity the government becomes certain that a lower productivity has occurred and this certainty accelerates.

Clearly, while we have illustrated the technique in a model of optimal fiscal policy, the methodology can be easily extended to other dynamic models, for example in the analysis of optimal monetary policy in sticky price models (e.g. Clarida et al. 1999) under the assumption of Partial Information. Our optimal policy smoothing result is likely to extend to that setup, potentially leading to a microfoundation for smooth nominal interest rates.

Appendix A: Proof of Proposition 5

To arrive at Proposition 5 we first need to show that for any variation δ and small α the equilibrium $S(\mathcal{R}^* + \alpha\delta, A)$ is well defined with probability one. This will be stated in Lemma 2. We first state a generic result on the existence and uniqueness of solutions.

On the sets $X \subset R$, $Y \subset R^n$ we consider functions $f, f^k : X \times Y \to R$ for $k = 1, 2, ...$

Let $d(f, g)$ be the sup norm

$$d(f,g) = \sup_{X \times Y} |f(x,y) - g(x,y)|$$

We make the following assumptions

- **Assumption L1:** X and Y are compact, $f(\cdot, y)$ and $f^k(\cdot, y)$ are absolutely continuous for each y.

- **Assumption L2:** f^k converge uniformly to f, i.e. $d(f, f^k) \to 0$ as $k \to \infty$

- **Assumption L3:** for each $y \in Y$ there is a unique solution to $f(\cdot, y) = 0$. This solution lies in the interior of X.

  Hence, there is a well-defined mapping $\chi : Y \to X$ with the property $f(\chi(y), y) = 0$ and $\chi(Y) \subset int(X)$.

  Let $X^d(y) \subset X$ be the set containing all points where the derivatives $f_x(\cdot, y)$ and $f_x^k(\cdot, y)$ exist for all k. Note that because of absolute continuity $f_x(\cdot, y)$ and $f_x^k(\cdot, y)$ are differentiable almost everywhere so $X^d(y)$ has measure zero.

- **Assumption L4:** The partial derivatives of f^k converge as follows

$$\sup_{x \in X^d(y), y} \left| f_x(x,y) - f_x^k(x,y) \right| \to 0$$

- **Assumption L5:** f_x is uniformly bounded away from zero near the zeroes. Formally, there is a constant $K > 0$ and an $\varepsilon > 0$ such that for all $y \in Y$ and all x such that $|x - \chi(y)| < \varepsilon$, then $x \in X$ and if $f_x(x, y)$ exists then

$$|f_x(x, y)| > K$$

Furthermore, for all y, the sign of $f_x(x, y)$ is the same for any x such that $|x - \chi(y)| < \varepsilon$ where f_x exists.

Note that once we have uniqueness as in Assumption L3 the last assumption is easily guaranteed, it just requires that the partial f_x is not close to 0 near the zeroes, for example if $f(x, y) = \phi x + x^3$ it excludes $\phi = 0$ but any other ϕ works. The sign restriction excludes, for example, $f(x, y) = |x|$.

Lemma 1. ***(Existence and uniqueness of solutions)*** *Under Assumptions L1-L5 there exists a unique solution to $f^k(\cdot, y) = 0$ for all y for k is sufficiently high. Let $\chi^k(y)$ denote this solution. Furthermore, $d(\chi^k, \chi) \to 0$ as $k \to \infty$.*

Proof

Define the neighborhood of the zeros as the set

$$XY^\varepsilon \equiv \{(x, y) : \text{for all } \; y \in Y \; \text{ and } \; x \; \text{ such that } \; |x - \chi(y)| < \varepsilon \; \text{ for some } \; y \in Y\}$$

where ε is as in assumption L5.

We first prove that $f^k(\cdot, y)$ has a solution for all y for sufficiently high k.

The fundamental theorem of calculus gives

$$\int_{\chi(y)}^{x} f_x(\overline{x}, y) d\overline{x} = f(x, y) \tag{3.57}$$

for all (x, y).[22] Consider a y such that the sign of f_x near $\chi(y)$ is positive. Using assumption L5, if $|x - \chi(y)| < \varepsilon$ and if $x > \chi(y)$ we have

$$\int_{\chi(y)}^{x} f_x(\cdot, y) > K(x - \chi(y)) \tag{3.58}$$

[22] It is understood that for $x < \chi(y)$ upper and lower limit of the integral need to be exchanged.

Plugging $x = \chi(y) \pm \varepsilon/2$ in (3.57) we have

$$\begin{array}{rcl} \frac{\varepsilon}{2}K & < & f(\chi(y) + \frac{\varepsilon}{2}, y) \\ \frac{\varepsilon}{2}K & > & f(\chi(y) - \frac{\varepsilon}{2}, y) \end{array}$$

A similar argument gives the reverse inequalities for y's such that the sign of f_x is negative near $\chi(y)$.

This all implies that $f(\chi(y) \pm \frac{\varepsilon}{2}, y)$ are bounded away from zero, and of oposite signs for all y.

Note that here we have used $\chi(Y) \subset int(X)$ in guaranteeing that f is well defined at $x = \chi(y) \pm \frac{\varepsilon}{2}$.

Applying uniform convergence of f^k there is k sufficiently high so that $d(f^k, f) < \frac{\varepsilon}{4}K$. The last two equations imply that if $K > 0$

$$\begin{array}{rcl} \frac{\varepsilon}{4}K & < & f^k(\chi(y) + \frac{\varepsilon}{2}, y) \\ \frac{\varepsilon}{4}K & > & f^k(\chi(y) - \frac{\varepsilon}{2}, y) \end{array}$$

Therefore $f^k(\cdot, y)$ takes a positive and a negative value. This implies by the intermediate value theorem that a solution to $f^k(\cdot, y) = 0$ exists for all y and for k high enough. As promised, we have shown that $f^k(\cdot, y)$ has a solution for all y for sufficiently high k. Let us call this solution $\chi^k(y)$, since we have not proved uniqueness all we know thus far is that $\chi^k(y)$ is a non-empty set.

We now prove that $\chi^k(y) \subset XY^\varepsilon$. We first show that

$$|f(x, y)| \geq V \text{ for all } (x, y) \notin XY^\varepsilon \tag{3.59}$$

for a constant $V > 0$. To see this consider the infimum of $|f|$ outside XY^ε

$$V \equiv \inf_{X \times Y - XY^\varepsilon} |f(x, y)|$$

Compactness of $X \times Y - XY^\varepsilon$ and continuity of f implies the infimum is attained at some $(x^*, y^*) \in X \times Y - XY^\varepsilon$ such that $f(x^*, y^*) = V$. If $V = 0$ this implies that

$x^* = \chi(y^*)$ and it contradicts the fact that all the zeroes of f are contained in XY^ε. Therefore $V > 0$.

Uniform convergence together with (3.59) implies

$$\left|f^k(x,y)\right| > V/2 > 0 \;\forall (x,y) \in X \times Y - XY^\varepsilon \;\text{ for }\; k \;\text{ large enough}$$

so that $f^k(x,y) = 0$ can only happen for $(x,y) \in XY^\varepsilon$ for k large.

This proves that all solutions of f^k (namely (x,y) where $x \in \chi^k(y)$) must lie in XY^ε for k large enough.

All that is left to show is that $\chi^k(y)$ has only one element for k large enough. Applying (3.57) and using assumption L5 in a similar way as we did above implies that given any y for which f_x is positive in assumption L5, for any elements $x', x'' \in \chi^k(y)$ with $x' \geq x''$ then

$$\int_{x''}^{x'} f_x(\overline{x}, y) d\overline{x} = f(x', y) - f(x'', y) = 0$$

which is impossible since

$$\int_{x''}^{x'} f_x(\overline{x}, y) d\overline{x} > K(x' - x'') > 0$$

Therefore there exists a unique solution of $f^k(\cdot, y)$ for k large enough for all y or, equivalently, $\chi^k(y)$ has one element for all y.

To prove convergence of the solutions to $\chi^k(y)$ note that, given y, for any convergence subsequence $\left\{\chi^{k_j}(y)\right\}_{j=1}^\infty$ where the limit is denote $\lim_{j\to\infty} \chi^{k_j}(y) = L(y)$, uniform convergence of f^k implies that $f(L(y), y) = 0$, so that $L(y) = \chi(y)$ and $\chi^k(y) \to \chi(y)$ as $k \to \infty$. $\square$

We apply Lemma 1 to show that $S(\mathcal{R}^* + \alpha\delta, A)$ is well defined. We take h and W as defined on sets for signals $s \in \mathcal{S}$ and policies $\tau \in \mathcal{T}$.

We need the following assumptions on our model. Recall H has been defined in (3.23).

Assumptions

1. The sets $\mathcal{S}$,$\mathcal{T}$ are both compact. The set of realizations Φ is also compact

2. $h(\cdot; A)$ is differentiable everywhere, $|h_\tau| < Q$ uniformly on (s, A) for a constant $Q < \infty$ and h_τ is Lipschitz continuous with respect to s for $s = S(\mathcal{R}^*, A)$ for almost all A.

3. For each A there is a unique solution $S(\mathcal{R}^*, A) \in int(\mathcal{S})$ setting $H(S(\mathcal{R}^*, A), A; \mathcal{R}^*) = 0$

4. $\mathcal{R}^*$ is continuous in s. Furthermore, $|\mathcal{R}^{*\prime}| < K^R$, for a constant $K^R < \infty$ for $s = S(\mathcal{R}^*, A)$ for almost all A.

5. Either $h_\tau(\mathcal{R}^*(s); A)\mathcal{R}^{*\prime}(s) < 1 - K$ or $h_\tau(\mathcal{R}^*(s); A)\mathcal{R}^{*\prime}(s) > 1 + K$ for all $s = S(\mathcal{R}^*; A)$ where the derivatives exist and for some constant $K > 0$.

Assumptions 1-2 can be imposed before knowing the solution to the PI problem. Assumptions 3-5 depend on the solution $\mathcal{R}^*$ which is not known ahead of time, but they can be checked ex-post with a candidate solution. Typically $\mathcal{R}^*$ will have a bounded derivative almost everywhere and since we consider cases where $S(\mathcal{R}^*, A)$ has a density these assumptions are easily verified.

Lemma 2. *Consider a bounded function (a "variation") $\delta : \mathcal{S} \to R$ differentiable everywhere and with a uniformly bounded derivative. Under assumptions 1-5 $S(\mathcal{R}^* + \alpha\delta, A)$ exists and is unique for α small enough.*

Proof

We apply Lemma 1 when we take a sequence $\alpha_k \to 0$. The objects in our model maps into the notation of Lemma 1 as follows

-s takes the role of x and A takes the role of y in Lemma 1.

-$f(\cdot) \equiv H(\cdot; \mathcal{R}^*)$

-$f^k(\cdot) \equiv H(\cdot; \mathcal{R}^* + \alpha_k\delta)$

We now have to check that assumptions 1-5 imply assumptions L1-L5.

Assumptions 1,2,4 imply Assumption L1.

Absolute continuity of h implies

$$|H(s, A; \mathcal{R}^*) - H(s, A; \mathcal{R}^* + \alpha\delta)| < Q\ \alpha\, |\delta(s)|$$

where $Q < \infty$ is the uniform bound on $|h_\tau|$. Since δ is bounded this implies that $H(.; \mathcal{R}^* + \alpha_k\delta)$ convergence uniformly to $H(\cdot; \mathcal{R}^*)$ as in assumption L2.

Assumption L3 and 3 are equivalent.

Consider s, A where the derivatives $h_\tau(\cdot, A)$ and $\mathcal{R}^{*\prime}$ exist. We have

$$H_s(s, A; \mathcal{R}^* + \alpha_k\delta) = 1 - h_\tau((\mathcal{R}^* + \alpha\delta)(s), A)(\mathcal{R}^{*\prime} + \alpha\delta')(s)$$

Therefore

$$\begin{aligned} |H_s(s, A; \mathcal{R}^*) - H_s(s, A; \mathcal{R}^* + \alpha_k\delta)| &= \\ |h_\tau((\mathcal{R}^* + \alpha\delta)(s), A)(\mathcal{R}^{*\prime} + \alpha\delta')(s) - h_\tau(\mathcal{R}^*(s), A)\mathcal{R}^{*\prime}(s)| &= \\ |\ [h_\tau((\mathcal{R}^* + \alpha\delta)(s), A) - h_\tau(\mathcal{R}^*(s), A)]\,(\mathcal{R}^{*\prime} + \alpha\delta')(s) + \alpha\delta'(s)\ | &\leq \\ Q^L\alpha K^\delta(K^R + \alpha K^{\delta'}) + \alpha K^{\delta'} & \end{aligned}$$

the second equality follows from adding and subtracting $h_\tau(\mathcal{R}^*(s), A)(\mathcal{R}^{*\prime} + \alpha\delta')(s)$ and where $K^{\delta'}$ is the bound on δ', K^δ the bound on δ, and Q^L the Lipschitz constant for h_τ.

This guarantees Assumption L4.

Assumption L5 is given by assumption 5.

Therefore, Lemma 1 implies that $S(\mathcal{R}^*{+}\alpha\delta, A)$ is well defined for α small enough. $\square$

To prove Proposition 5 we now need to add

Assumption 6: W is continuously differentiable with respect to (τ, s) for almost all A and when the derivative exists it is uniformly bounded.

Proof of Proposition 5

Consider a variation $\delta : \mathcal{S} \to R$ such that δ is continuous, differentiable, $|\delta(s)| \leq K^\delta$ and $|\delta'(s)| \leq K^{\delta\prime}$ for all s for some constants $K^\delta, K^{\delta'} < \infty$.

Consider the problem defined by (3.28). For α small enough $S(\mathcal{R}^* + \alpha\delta; A)$ is well defined by Lemma 2, and so is $\mathcal{F}$. It is clear that since $\mathcal{R}^* + \alpha\delta$ is a feasible policy function in the PI problem the solution of (3.28) is attained at $\alpha = 0$. Since (3.28) is a standard one-dimensional maximization problem this implies

$$\left.\frac{d\mathcal{F}(\mathcal{R}^* + \alpha\delta)}{d\alpha}\right|_{\alpha=0} = 0 \tag{3.60}$$

if this derivative exists. We now prove that this derivative exists and that (3.60) implies (3.27). We consider here the one-dimensional case for τ and s, namely $m = n = 1$, the generalization to the multivariate case is left for future work.

The assumptions on W and h imply

$$|W_\tau|, |W_s|, |h_\tau| < K^{Wh}$$

for some finite constant K^{Wh} whenever the derivatives exist.

Take any sequence $\alpha_k \to 0$. For each k we can write

$$\frac{\mathcal{F}(\mathcal{R}^* + \alpha_k\delta) - \mathcal{F}(\mathcal{R}^*)}{\alpha_k} = \int_\Phi M_k(A)\; dF_A(A)$$

for

$$M_k(A) \equiv \frac{W\left(T(\mathcal{R}^*{+}\alpha_k\delta, A), S(\mathcal{R}^*{+}\alpha_k\delta, A), A\right) - W\left(T(\mathcal{R}^*, A), S(\mathcal{R}^*, A), A\right)}{\alpha_k}$$

At all points A where $\mathcal{R}^*, W, h$ are differentiable at $s = S(\mathcal{R}^*, A)$ we have that

$$\begin{aligned} M_k(A) &\to \left.\frac{dW\left(\left(\mathcal{R}^* + \alpha\delta\right)\left(S(\mathcal{R}^*{+}\alpha\delta; A)\right),\; S\left(\mathcal{R}^*{+}\alpha\delta, A\right), A\right)}{d\alpha}\right|_{\alpha=0} && (3.61)\\ &= W_\tau^{*\prime}(A)\left\{\left[\mathcal{R}^{*\prime}(A)\right] S_\delta^{*\prime}(A) + \delta^*(A)\right\} + W_s^{*\prime}(A) S_\delta^{*\prime}(A) && (3.62)\end{aligned}$$

where

$$\begin{aligned} \mathcal{R}^{*\prime}(A) &= \mathcal{R}^{*\prime}\left(S(\mathcal{R}^*; A)\right) && (3.63)\\ \delta^{*\prime}(A) &= \delta'\left(S(\mathcal{R}^*, A)\right) \\ \delta^*(A) &= \delta\left(S\left(\mathcal{R}^*, A\right)\right) \\ W_x^*(A) &= W_x\left(\left(\mathcal{R}^*\left(S(\mathcal{R}^*, A)\right)\right), S\left(\mathcal{R}^*, A\right)\right) \quad \text{for } \; x = \tau, s \\ S_\delta^{*\prime}(A) &= \left.\frac{dS(\mathcal{R}^*{+}\alpha\delta, A)}{d\alpha}\right|_{\alpha=0} && (3.64)\end{aligned}$$

(note there is a slight abuse of notation on $\mathcal{R}^{*\prime}$ and $\delta^{*\prime}$ as we use the symbol for the function of s as well as for the function of A).

The only non-obvious term is the derivative $S_\delta^{*\prime}$, furthermore, this is the only term that depends on δ at $\alpha = 0$. The term $S_\delta^{*\prime}$ can be found by applying the implicit function theorem to

$$S(\mathcal{R}^*{+}\alpha\delta; A) = h((\mathcal{R}^*{+}\alpha\delta)(S(\mathcal{R}^*{+}\alpha\delta; A)), A)$$

Carefully differentiating with respect to α we have

$$\begin{aligned}\frac{dS(\mathcal{R}^*{+}\alpha\delta; A)}{d\alpha} &= h_\tau((\mathcal{R}^*{+}\alpha\delta)(S(\mathcal{R}^*{+}\alpha\delta; A)), A)\left[(\mathcal{R}^{*\prime}{+}\alpha\delta')(S(\mathcal{R}^*{+}\alpha\delta; A))\right]\cdot \\ &\quad \cdot\frac{dS(\mathcal{R}^*{+}\alpha\delta; A)}{d\alpha} + \delta(S(\mathcal{R}^*{+}\alpha\delta; A))\end{aligned}$$

So that

$$S_\delta^{*\prime}(A) \equiv \left.\frac{dS(\mathcal{R}^*{+}\alpha\delta, A)}{d\alpha}\right|_{\alpha=0} = \frac{h_\tau^*(A)\delta^*(A)}{1 - h_\tau^*(A)\mathcal{R}^{*\prime}(A)}$$

Using the boundedness assumptions it is clear that $M_k(A)$ is uniformly bounded on k for almost all A.

The assumption that $S(\mathcal{R}^*, A)$ has a density implies that the limit in (3.62) occurs with probability one in A. Since M_k is uniformly bounded Lebesgue dominated convergence implies

$$\frac{\mathcal{F}(\mathcal{R}^* + \alpha_k\delta) - \mathcal{F}(\mathcal{R}^*)}{\alpha_k} \to \int_\Phi \left(\left[W_\tau^*\mathcal{R}^{*\prime} + W_s^*\right] S_\delta^{*\prime} + W_\tau^*\delta^*\right)\, dF_A$$

Since this holds for any sequence $\alpha_k \to 0$ it proves that the derivative of $\mathcal{F}$ with respect to α exists for any variation δ and from (3.60) we have

$$\int_\Phi \left(\left[W_\tau^*\mathcal{R}^{*\prime} + W_s^*\right] S_\delta^{*\prime} + W_\tau^*\delta^*\right)\, dF_A = 0 \tag{3.65}$$

Using the formula for $S_\delta^{*\prime}(A)$ and rearranging, we conclude that for any variation δ

$$\int_\Phi \left(W_\tau^* + W_s^* h_\tau^*\right) \frac{\delta^*}{1 - h_\tau^*\mathcal{R}^{*\prime}}\, dF_A = 0 \tag{3.66}$$

Since (3.66) holds for any bounded δ with bounded derivative it also holds when δ is any bounded function measurable with respect to s. Therefore, the general definition of conditional expectation implies (3.27).

Appendix B: The effect of government expenditure on the signal extraction

In this Appendix we show that in the log-quadratic case an increase in g can change the sign of optimal policy in the intermediate region of the observables.

First, for a given $\bar{l}$ consider the locus $\tilde{\gamma}(\mathcal{R}(\bar{l}), \theta, \bar{l})$ which is the inverse of the reaction function h with respect to γ, implicitly defined by

$$h(\mathcal{R}(\bar{l}), \theta, \tilde{\gamma}(\mathcal{R}(\bar{l}), \theta, \bar{l})) - \bar{l} = 0. \tag{3.67}$$

Let ζ be the derivative of this function with respect to θ. This can be found using the implicit function theorem, which gives the positive slope

$$\zeta = -\frac{h_\theta}{h_\gamma} = \frac{g\sqrt{(Bg)^2 + 4B\theta^2\gamma(1-\tau)} + Bg^2}{2\theta^3(1-\tau)} \tag{3.68}$$

Now we differentiate ζ with respect to g and get

$$\frac{d\zeta}{dg} = \frac{\partial\zeta}{\partial g} + \frac{\partial\zeta}{\partial\tau}\frac{\partial\tau}{\partial g} \tag{3.69}$$

where all these partial derivatives are positive. Hence higher government expenditure makes the loci of realizations of the shocks steeper.

Now we illustrate how this effect changes the nature of the signal extraction on the shocks and hence the slope of optimal policy in the intermediate region of the observables. For this purpose we will take a first order approximation of the loci $\tilde{\gamma}$.[23]

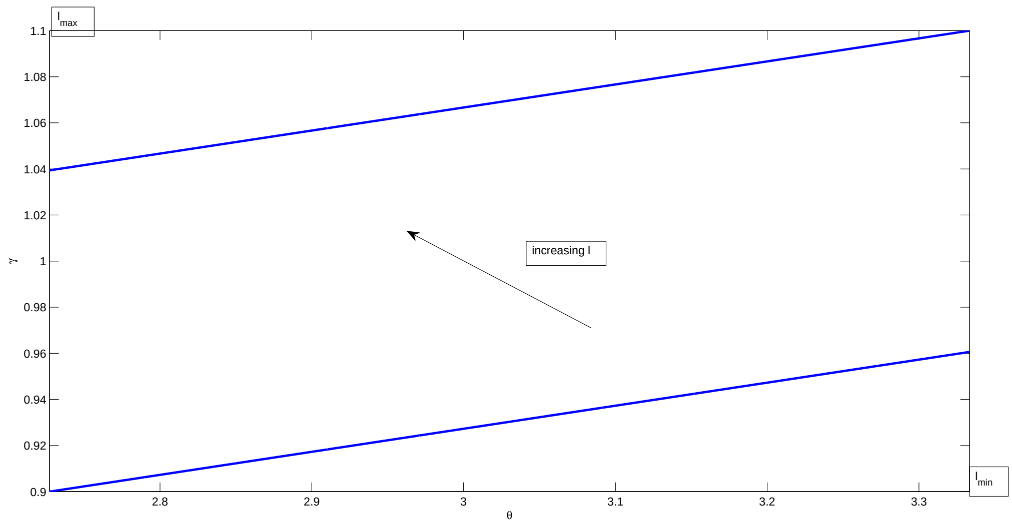
Consider Figures 3.19 and 3.20. When g is sufficiently low, the map of loci (solid blue lines) moving in the direction of increasing l's looks like in Figure 3.19. Starting at l_{min} (bottom right corner) and increasing l the loci first hit the bottom-left corner, where the lowest θ becomes possible, and then the top-right corner, where the highest values for θ start to be inconsistent with the observed l's. Hence in the intermediate region of l's all θ's are possible, but clearly not all $\gamma's$. In this region, the government

[23] In our computed examples these loci are very close to linear.

learns little about productivity. All the government learns is that the agent is working more as l increases so expected output is higher and taxes can be lower. This gives the negative slope of $\mathcal{R}^*$ in subsection 3.4.1 with low government expenditure.

On the other hand, when g is sufficiently high, the slope of the loci becomes higher. Hence, as illustrated in Figure 3.20, in the intermediate region of l's the government learns that only a relatively small set of θ's is possible, whereas any γ is consistent with the observations. This leads to the positive slope of the optimal tax rate in subsection 3.4.2 with high government expenditure.

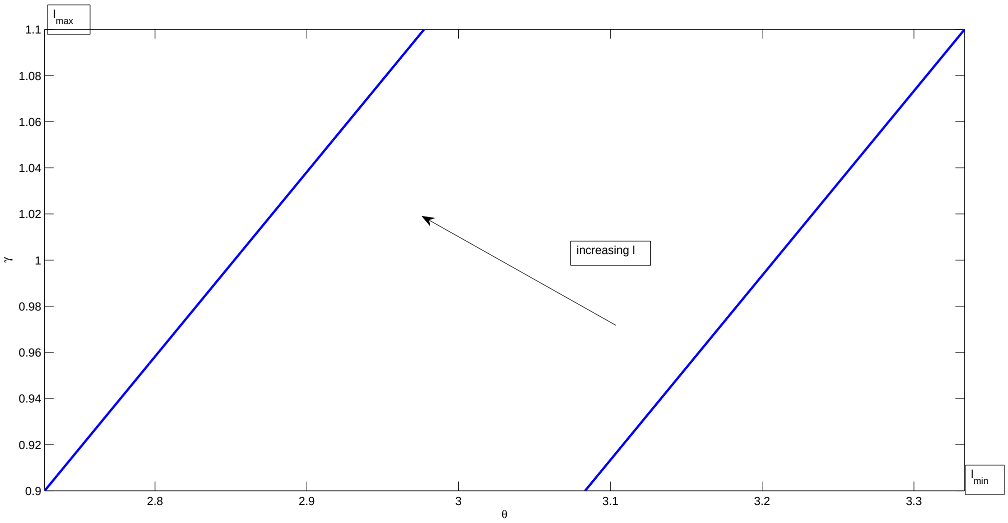
Figure 3.19: Loci of shocks realizations with low g



Appendix C: Derivation of the Envelope Condition (3.56)

In this Appendix we derive the Envelope Condition (3.56). First of all let us introduce the necessary notation. A tax policy is a function of debt and labor $\mathcal{R}(b, l)$ and labor

Figure 3.20: Loci of shocks realizations with high g



is a function of a policy $\mathcal{R}$, outstanding debt and the exogenous shock, $L(\mathcal{R}; b, A)$ defined by the zero of

$$H(l, A, \mathcal{R}) \equiv l - h(\mathcal{R}(b, l), A), \tag{3.70}$$

in analogy with the two-period model. By total differentiation of (3.70), the partial derivative of labor with respect to debt, L_b, is given by

$$L_b(\mathcal{R}, b, A) = \frac{\gamma\theta\mathcal{R}_b(b, l)}{v''(l) + \gamma\theta\mathcal{R}_L(b, l)}. \tag{3.71}$$

Now, for simplicity consider a case without borrowing penalty. In order to derive the envelope condition, we differentiate (3.55) with respect to b and get

$$W'(b) = \mathrm{E}\left[\left(\gamma\theta - v'(l^*)\right) L_b^* + W'\left(b^{*'}\right)\left(\frac{\gamma}{\bar{\gamma}} + \beta b_L^{*'} L_b^*\right)\right]$$

where

$$b_L^{*'} = \frac{-\theta\gamma + [v''(L(\mathcal{R}^*, A, b))L(\mathcal{R}^*, A, b) + v'(L(\mathcal{R}^*, A, b))]}{\beta\bar{\gamma}}$$

$$l^* = L(\mathcal{R}^*; b, A)$$

$$L_b^* = L_b(\mathcal{R}^*; b, A).$$

Using (3.71) we can write

$$W'(b) = \mathrm{E}\left[\left(\gamma\theta - v'(l^*) + \beta W'\left(b^{*'}\right) b_L^{*'}\right) \frac{-\gamma\theta\mathcal{R}_b^*(l,b)}{v''(l) + \gamma\theta\mathcal{R}_L^*(l,b)} + W'\left(b^{*'}\right)\frac{\gamma}{\bar{\gamma}}\right]. \quad (3.72)$$

Using Proposition 5, the FOC of PI Ramsey problem is

$$\mathrm{E}\left[\left(\theta\gamma - v'(l^*) + \beta W'(b'^*) b_L^{*'}\right) \frac{h_\tau^*}{1 - h_\tau^* \mathcal{R}_L^*} | \bar{l}\right] = 0 \quad (3.73)$$

for all $\bar{l}$. Furthermore, we have that the partial derivative of the reaction function h with respect to taxes is

$$h_\tau = \frac{-\gamma\theta}{v''(l)}.$$

So from (3.72) we get

$$W'(b) = \mathrm{E}\left[\left(\gamma\theta - v'(l^*) + \beta W'\left(b^{*'}\right) b_L^{*'}\right) \frac{h_\tau^*\ \mathcal{R}_b^*(l,b)}{1 - h_\tau^* \mathcal{R}_L^*(L,b)} + W'\left(b^{*'}\right)\frac{\gamma}{\bar{\gamma}}\right].$$

Now, applying the law of iterated expectations, using the fact that $\mathcal{R}_b(l,b)$ is known given L, b and using (3.73), we obtain

$$\begin{aligned} W'(b) &= \mathrm{E}\left[\mathrm{E}\left(\left(\gamma\theta - v'(L^*) + \beta W'\left(b^{*'}\right) b_L^{*'}\right) \frac{h_\tau^*\ \mathcal{R}_b^*(L,b)}{1 - h_\tau^* \mathcal{R}_L^*(L,b)}\middle| L\right) + W'\left(b^{*'}\right)\frac{\gamma}{\bar{\gamma}}\right] && (3.74) \\ &= \mathrm{E}\left[0 + W'\left(b^{*'}\right)\frac{\gamma}{\bar{\gamma}}\right] && (3.75) \end{aligned}$$

Finally, adding the marginal cost of excessive debt this becomes

$$W'(b) = \mathrm{E}\frac{\gamma W'(b')}{\overline{\gamma}} - \chi(b' - b^{\max}) I_{[b^{\max},\infty)}(b').$$

Bibliography

[1] Abel, A.B., A.K. Dixit, J.C. Eberly, R.S. Pindyck (1996), Options, the Value of Capital, Investment, *The Quarterly Journal of Economics*, 111(3): 753-777.

[2] Aiyagari, S.R., A. Marcet, T.J. Sargent, J. Seppala (2002). Optimal Taxation without State-Contingent Debt, *Journal of Political Economy*, 110(6): 1220-1254.

[3] Altig, D., L.J. Christiano, M. Eichenbaum, J. Linde' (2011), Firm-specific Capital, Nominal Rigidities and the Business Cycle, *Review of Economic Dynamics*, 14(2): 225-247.

[4] Angeletos, George-Marios and Alessandro Pavan (2009), Policy with Dispersed Information, *Journal of the European Economic Association*, 7(1): 11-60.

[5] Aoki, K. (2003), On the Optimal Monetary Policy Response to Noisy Indicators, *Journal of Monetary Economics*, 50(3): 501-523.

[6] Atkeson, A., P.J. Kehoe (1999), Models of Energy Use: Putty-Putty versus Putty-Clay, *American Economic Review*, 89(4): 1028-1043.

[7] Bachmann, R., R.J. Caballero, E. Engel (2013), Aggregate Implications of Lumpy Investment: New Evidence and a DSGE Model, *American Economic Journal: Macroeconomics*

[8] Backus, D.K., P.J. Kehoe, F.E. Kydland (1994), Dynamics of the Trade Balance and the Terms of Trade: The J-Curve?, *American Economic Review*, 84(1): 84-103.

[9] Barnett, A., B. Broadbent, A. Chiu, J. Franklin, H. Miller (2014), Impaired Capital Reallocation and Productivity, *National Institute Economic Review*, 228:R35, NIESR.

[10] Basu, S., J.G. Fernald, M.D. Shapiro (2001), Productivity Growth in the 1990s: Technology, Utilization, or Adjustment?, *Carnegie-Rochester Conference Series on Public Policy*, 55(1): 117-165.

[11] Baxter, B., L. Graham, S. Wright (2007), The Endogenous Kalman Filter, *Birkbeck School of Economics, Mathematics and Statistics working paper*, BWPEF 0719.

[12] Baxter, B., L. Graham, S. Wright (2011), Invertible and Non-invertible Information Sets in Linear Rational Expectations Models, *Journal of Economic Dynamics & Control*, 35: 295-311.

[13] Bertola, G., R. J. Caballero (1994), Irreversibility and Aggregate Investment, *Review of Economic Studies*, 61(2): 223-246.

[14] Bloom, N., S. Bond, J. Van Reenen (2007), Uncertainty and Investment Dynamics, *Review of Economic Studies*, 74(2): 391-415.

[15] Bloom, N. (2009), The Impact of Uncertainty Shocks, *Econometrica*, 77(3): 623-685.

[16] Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, S.J. Terry (2012), Really Uncertain Business Cycles, *NBER Working Paper*, No. w18245.

[17] Broadbent, B. (2012), Productivity and the Allocation of Resources, Speech given at Durham Business School, 12th September.

[18] Burnside, C., M. Eichenbaum, S. Rebelo (1995), Capital Utilization and Returns to Scale, *NBER Macroeconomics Annual*, Vol. 10, ed. B. Bernanke and J.J. Rotemberg.

[19] Caunedo J. (2014), Industry Dynamics, Investment and Business Cycles, *Working Paper*.

[20] Chen K., Z. Song (2013), Financial Frictions on Capital Allocation: A Transmission Mechanism of TFP Fluctuations, *Journal of Monetary Economics*, 60(6).

[21] Clarida, R., J. Galí, M. Gertler (1999), The Science of Monetary Policy: A New Keynesian Perspective, *Journal of Economic Literature*, 37: 1661-1707.

[22] Cooley, T.F., G.D. Hansen, E.C. Prescott (1995), Equilibrium Business Cycles with Idle Resources and Variable Capacity Utilization, *Economic Theory*, 6(1): 35-49.

[23] Cooper, R.W., J.C. Haltiwanger (2006), On the Nature of Capital Adjustment Costs, *Review of Economic Studies*, 73(3): 611-633.

[24] Cooper, R.W., I. Schott (2013), Capital Reallocation and Aggregate Productivity, *NBER Working Paper*, No. 19715.

[25] Cui, W. (2014), Delayed Capital Reallocation, *Working Paper*.

[26] Cui, W., S. Radde (2014), Search-Based Endogenous Illiquidity, *Working Paper*.

[27] den Haan (2010), Assessing the Accuracy of the Aggregate Law of Motion in Models with Heterogeneous Agents, *Journal of Economic Dynamics and Control*, 34(1): 79-99.

[28] Dixit, A.K., R.S. Pindyck (1994), *Investment under Uncertainty*, Princeton University Press.

[29] Edgerton, J. (2011), The Effects of Taxation on Business Investment: New Evidence from Used Equipment, *Working Paper*.

[30] Eisfeldt, A.L. (2004), Endogenous Liquidity in Asset Markets, *Journal of Finance*, 59(1): 1-30.

[31] Eisfeldt, A.L., A.A. Rampini (2006), Capital Reallocation and Liquidity, *Journal of Monetary Economics*, 53(3): 369-399.

[32] Eisfeldt, A.L., A.A. Rampini (2007), New or Used? Investment with Credit Constraints, *Journal of Monetary Economics*, 54(8): 2656-2681.

[33] Ellison, M., N. Valla (2001), Learning, Uncertainty and Central Bank Activism in an Economy with Strategic Interactions, *Journal of Monetary Economics*, 48(1): 153-171.

[34] Gavazza, A. (2011a), Leasing and Secondary Markets: Theory and Evidence from Commercial Aircraft, *Journal of Political Economy*, 119(2): 325-377.

[35] Gavazza, A. (2011b), The Role of Trading Frictions in Real Asset Markets, *The American Economic Review*, 101(4): 1106-1143.

[36] Gilchrist, S, J.W. Sim, E. Zagrasjek. (2014), Uncertainty, Financial Frictions, and Investment Dynamics, *NBER Working Paper*, No. 20038.

[37] Greenwood, J., Z. Hercowitz, G.W. Huffman (1988), Investment, Capacity Utilization, and the Real Business Cycle, *American Economic Review*, 78(3): 402-17.

[38] Greenwood, J., Z. Hercowitz, P. Krusell (2000), The Role of Investment-specific Technological Change in the Business Cycle, *European Economic Review*, 44(1): 91-115.

[39] Guerrieri, V., R. Shimer (2013), Markets with Multidimensional Private Information , *Society for Economic Dynamics Meeting Papers series 2013*, n. 210.

[40] Hansen, G.D. (1985), Indivisible Labor and the Business Cycle, *Journal of Monetary Economics*, 16(3), 309-327.

[41] Hansen, G.D., E.C. Prescott (2005), Capacity Constraints, Asymmetries, and the Business Cycle, *Review of Economic Dynamics*, 8(4): 850-865.

[42] Hopenhayn, H.A. (1992), Entry, Exit, and Firm Dynamics in Long Run Equilibrium, *Econometrica*, 60(5): 1127-50.

[43] Hopenhayn, H.A., R. Rogerson (1993), Job Turnover and Policy Evaluation: A General Equilibrium Analysis, *Journal of Political Economy*, 101(5): 915-38.

[44] Jovanovic, B., P.L. Rousseau (2008), Mergers as Reallocation, *The Review of Economics and Statistics*, 90(4), 765-776.

[45] Jovanovic, B., Y. Yatsenko (2012), Investment in Vintage Capital, *Journal of Economic Theory*, 147(2), 551-569.

[46] Khan, A., J.K. Thomas (2008), Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics, *Econometrica*, 76(2): 395-436.

[47] Khan, A., J.K. Thomas (2013), Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity, *Journal of Political Economy*, 121(6): 1055 - 1107.

[48] Kiefer, N., Y. Nyarko (1989), Optimal Control of an Unknown Linear Process with Learning, *International Economic Review*, Dept. of Economics, U. of Pennsylvania and Osaka U. ISERA, vol. 30(3), pages 571-86, August.

[49] Klemperer, P.D., M.A. Meyer (1989), Supply Function Equilibria in Oligopoly under Uncertainty, *Econometrica*, Vol. 57, No. 6, pp. 1243-1277

[50] Kocherlakota, N.J. (2010), *The New Dynamic Public Finance*, Princeton University Press.

[51] Krusell, P., A.A. Smith (1998), Income and Wealth Heterogeneity in the Macroeconomy, *Journal of Political Economy*, 106(5): 867-896.

[52] Krusell, P., A.A. Smith (2006), Quantitative Macroeconomic Models with Heterogeneous Agents, in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress.*

[53] Kurlat, P. (2013), Lemons Markets and the Transmission of Aggregate Shocks, *American Economic Review*, 103(4): 1463-89.

[54] Kydland, F.E., E.C. Prescott (1988), The Workweek of Capital and Its Cyclical Implications, *Journal of Monetary Economics*, 21(2-3): 343-360.

[55] Lee, Y., T. Mukoyama (2013), Entry, Exit, and Plant-level Dynamics over the Business Cycle, *Working Paper*.

[56] Li, S., T.M. Whited (2014), Capital Reallocation and Adverse Selection, *Simon Business School Working Paper*, No. FR 14-16.

[57] Ljungqvist, L., T.J. Sargent (2004), *Recursive Macroeconomic Theory*, 2nd Edition, The MIT Press.

[58] Lucas Jr, R.E., Jr. (1972), Expectations and the Neutrality of Money, *Journal of Economic Theory*, 4(2): 103-124.

[59] Lucas, R.E., Jr., N.L. Stokey (1983), Optimal Fiscal and Monetary Policy in an Economy without Capital, *Journal of Monetary Economics* , 12: 55-93

[60] Maksimovic, V., G. Phillips (2001), The Market for Corporate Assets: Who Engages in Mergers and Asset Sales and are there Efficiency Gains? *The Journal of Finance*, 56(6): 2019-2065.

[61] Mehra, R.and E.C. Prescott (1980), Recursive Competitive Equilibrium: the Case of Homogeneous Households, *Econometrica*, 48(6): 1365-1379.

[62] Mirman, L.J., L. Samuelson, A. Urbano (1993), Monopoly Experimentation, *International Economic Review*, 549-563.

[63] Muth, J.F. (1960), Optimal Properties of Exponentially Weighted Forecasts, *Journal of the American Statistical Association*, 55, 290.

[64] Nimark, K. (2008), Monetary Policy with Signal Extraction from the Bond Market, *Journal of Monetary Economics*, 55: 1389-1400.

[65] Pearlman, J., D. Currie, P. Levine (1986), Rational Expectation Models with Private Information, *Economic Modelling*, 3(2): 90-105.

[66] Pearlman, J. (1992), Reputational and Nonreputational Policies under Partial Information, *Journal of Economic Dynamics and Control*, 16(2): 339-357.

[67] Perri, F., V. Quadrini (2014), International Recessions, *Working Paper*.

[68] Pindyck, R.S. (1988), Irreversible Investment, Capacity Choice, and the Value of the Firm, *American Economic Review*, 78(5): 969-85.

[69] Pulvino, T. (1998), Do Asset Fire Sales Exist? An Empirical Investigation of Commercial Aircraft Transactions, *Journal of Finance*, 53(3): 939-978.

[70] Ramey, V.A., M.D. Shapiro (2001), Displaced Capital: A Study of Aerospace Plant Closings, *Journal of Political Economy*, 109(5): 958-992.

[71] Ravn, M.O., H. Uhlig (2002), On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations, *The Review of Economics and Statistics*, 84(2): 371-375.

[72] Rogerson, R. (1988), Indivisible Labor, Lotteries and Equilibrium, *Journal of Monetary Economics*, 21(1), 3-16.

[73] Sargent, T.J. (1980), Tobin's q and the Rate of Investment in General Equilibrium, *Carnegie-Rochester Conference Series on Public Policy*, 12(1): 107-154.

[74] Shleifer, A., R.W. Vishny (2012), Liquidation Values and Debt Capacity: A Market Equilibrium Approach, *The Journal of Finance*, 47(4): 1343-1366.

[75] Sustek, R. (2011), Plant-level Nonconvex Output Adjustment and Aggregate Fluctuations, *Journal of Monetary Economics*, 58(4): 400-414.

[76] Svensson, L.E.O., M. Woodford (2003), Indicator Variables for Optimal Policy, *Journal of Monetary Economics*, 50: 691-720.

[77] Svensson, L.E.O., M. Woodford (2004), Indicator Variables for Optimal Policy under Asymmetric Information, *Journal of Economic Dynamics & Control*, 28: 661-690

[78] Swanson, E.T. (2006), Optimal Nonlinear Policy: Signal Extraction with a Non-normal Prior, *Journal of Economic Dynamics and Control*, 30: 185-203.

[79] Townsend, R.M. (1983), Forecasting the Forecasts of Others, *Journal of Political Economy*, 91(4): 546-588.

[80] Van Nieuwerburgh, S., L. Veldkamp (2006), Learning Asymmetries in Real Business Cycles, *Journal of Monetary Economics*, 53(4): 753-772.

[81] Veracierto, M. L. (2002), Plant-level Irreversible Investment and Equilibrium Business Cycles, *American Economic Review*, 92(1): 181-197.

[82] Wieland, V. (2000a), Monetary Policy, Parameter Uncertainty and Optimal Learning, *Journal of Monetary Economics*, 46: 199-228

[83] Wieland, V. (2000b), Learning by Doing and the Value of Optimal Experimentation, *Journal of Economic Dynamics & Control*, 24: 501-534

[84] Young, E.R. (2010), Solving the Incomplete Markets Model with Aggregate Uncertainty Using the Krusell-Smith Algorithm and Non-stochastic Simulations, *Journal of Economic Dynamics and Control*, 34(1): 36-41.