Firm Dynamics and the Macroeconomy

Economics Department, Cardiff University

Anthony Savagar

May 15, 2016

A Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor of Philosophy of Cardiff University

ABSTRACT

The thesis investigates how firm entry and exit into industry influences macroeconomic productivity. The first contribution is to show that firm entry and exit dynamics cause endogenous productivity movements over the business cycle due to the slow response of incumbent firms to macroeconomic conditions. The second contribution is to show that these productivity effects persist into the long run because of firm dynamics' effect on industry competition. Therefore the thesis argues that slow firm responses cause amplified productivity effects in the short run and that these effects can persist into the long run.

A key distinction of the research is to develop an analytically tractable dynamic general equilibrium model. This provides a precise explanation of productivity movements, without using numerical simulation. A crucial feature of the modelling is that firm dynamics have a time-to-build lag, so entry and exit are noninstantaneous. This causes a short-run period during which shocks to the economy are borne by inert incumbent firms and this is responsible for amplified short-run productivity effects. However, over time firms are able to enter and exit which ameliorates the amplification effect. Thus this process alone does not explain persistent effects on productivity. In order to understand persistent effects, the thesis explains that one must consider the effect of entry and exit on the competitive pressure of incumbents. When this is taken into account it shows that firms change their pricing behaviour in response to entry and exit, and the result is that long-run pricing markups change which in turn affect long-run productivity.

Chapter 1 demonstrates the empirical relevance of the relationship between productivity, firm entry and output in US data. Chapter 2 develops a structural model to explain short-run movements in productivity and firm dynamics. Developing chapter 2, chapter 3 explains the long-run effect of firm dynamics on productivity through entry's effect on competition.

ACKNOWLEDGEMENTS

Special thanks to my supervisor Professor Huw Dixon. He has always been available to discuss research. Also thanks to Professor Patrick Minford who has mentored me since undergraduate studies. Thank you to all faculty members, management and support staff and to my PhD classmates at Cardiff University.

Thank you to KU Leuven and Lisbon University for hospitality during my PhD. I appreciate the hospitality of my hosts at these institutions Vivien Lewis (KU Leuven) and Paulo Brito and Luis Costa (Lisbon University).

Thank you to funders: Economic Social Research Council 2011-2015 and Royal Economic Society Junior Fellowship 2015-2016. Thank you to the Julian Hodge Institute of Applied Macroeconomics for conference funding 2013-2016.

CARDIFF UNIVERSITY

CARDIFF BUSINESS SCHOOL

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ………………………………………………………. (candidate)

Date ………………………..

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of ………………………..(insert MCh, Md, **MPhil,** PhD etc, as appropriate)

Signed ………………………………………………………. (candidate)

Date ………………………..

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.
Other sources are acknowledged by footnotes giving explicit references.

Signed ………………………………………………………. (candidate)

Date ………………………..

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ………………………………………………………. (candidate)

Date ………………………..

CONTENTS

LIST OF FIGURES

ACRONYMS AND NOTATION

IEP Income Expansion Path

EF Euler Frontier

BGM Bilbiie, Ghironi, and Melitz 2012

IO Industrial Organization

I(D)RTS Increasing (Decreasing) Returns to Scale

hod-x Homogeneous of Degree x

1

UNDERSTANDING FIRM DYNAMICS IN THE MACROECONOMY

This dissertation is a theoretical investigation into the effects of firm dynamics on the macroeconomy. I refer to 'firm dynamics' as the process of firm entry and exit into industries. I study how these processes affect macroeconomic behaviour with particular focus on total factor productivity (TFP). The thesis is divided into four chapters: Chapter 1 is an introduction that unifies the work and outlines empirical motivation; Chapter 2 develops the theory that slow firm entry and exit cause short-run productivity movements; Chapter 3 extends this argument to endogenize markups which explains the effect of firm dynamics on long-run productivity movements; Chapter 4 analyzes the implications of the thesis, and proposes future developments.

This introductory chapter 1 sets the thesis in its wider context. First I summarise the main findings and modelling technique. Second I unify the research with other literature. Third I complement the literature with empirical evidence that firm dynamics and productivity are closely correlated with the business cycle.

1.1 MAIN FINDINGS

Chapter 2's main result is to show that firm dynamics cause measured productivity to overshoot underlying productivity whilst firm entry or exit adjusts. The result relies upon firms being slow to adjust. Chapter 3 then extends this short-run theory to explain how measured productivity overshooting in the short run can persist into the long run. The result relies upon entry(exit) increasing(decreasing) competition, and thus having a long-run effect on markups. Hence chapter 2 is about short-run productivity amplification, whereas chapter 3 is about long-run productivity propagation.

Chapter 2 explains how firm dynamics cause short-run endogenous productivity movements. The theory is that firm dynamics determine the number of firms in the economy which determines division of inputs per firm and consequently productivity through returns to scale. Returns to scale are present because of a fixed cost and imperfect competition. Entry increases resource division and reduces incumbents returns to scale; exit decreases resource division and increases incumbents returns to scale. Importantly firm dynamics take place slowly, so there is an adjustment period during which changes in scale arise. For example, a positive shock to the economy, with slow firm adjustment, means that incumbents initially benefit from that positive shock. Thus their production increases (without changing anything they produce more), and by returns to scale are more productive. But over time firms are able to adjust and outsiders recognize the excess profit now being earned by incumbents. Consequently they enter the market eroding per firm output and therefore productivity until these excess profits are arbitraged to zero leaving a long-run zero profit free entry equilibrium. If one assumes instantaneous free entry, as is normal in macroeconomics, then these dynamics are eradicated.

Chapter 2 roadmap of main results: The main result, theorem 2.5.1, emphasizes that initial measured productivity overshoots long-run underlying productivity in response to technology shocks. The effect is increasing in the degree of imperfect competition and the size of the fixed cost, and it disappears if either of these are zero. This is because a higher fixed cost strengthens the degree of returns to scale, and similarly more imperfect competition increases returns to scale because it allows firms to suppress more production (raising their marginal product). Aside from the fixed cost that causes initial increasing returns to scale, the firms' variable (net of the fixed cost) production function has decreasing returns to scale (i.e. rising marginal cost) so the cost curve is U-shaped. Thus there is an efficient level of production at minimum cost that provides a benchmark to gauge firm output against. In theorem 2.2.2 I show that these efficient output outcomes always exceed those that arise under imperfect competition, the difference being a firm's excess capacity. So, given imperfect competition, we interpret increases in firm production as capacity utilization since a firm moves closer to its cost minimizing level, and decreases in firm production are opening in capacity as a firm moves further from its cost minimizing level.

Chapter 3 extends this argument to explain why productivity effects can propagate into the long run. The chapter considers a second firm dynamics factor that determines per firm output and therefore productivity. This second factor is that pricing markups of incumbents reflect the number of competitors in the market. Whereas in chapter 2, firm dynamics solely alter per firm inputs, in chapter 3 they also affect competitive pricing decisions. Entry increases competition and lowers markups;

exit decreases competition and raises markups. For example, firm exit weakens competition among remaining incumbents and they raise markups, in turn they produce less to reach the necessary revenue to cover costs, so per firm output falls. The new mechanism creates a trade-off that is not present in chapter 2: exit expands per firm resources (less division of inputs) increasing production, contrarily it raises markups decreasing production, vice-versa for entry. The result is that in addition to the short-run endogenous productivity movements of chapter 2, firm dynamics have a long-run persistent effect on productivity.

Chapter 3 roadmap of main results: Section 3.4 defines the two competing effects of entry: a *"competition effect"* reducing markups and increasing per firm output and productivity; an *"allocation effect"* reducing per firm output and productivity as inputs are divided among more firms. The effect of endogenous markups is to create endogenous long-run output per firm and productivity (i.e. the long-run outcomes depend on long-run number of firms). Theorem 3.5.1 shows that long-run output and productivity are increasing in number of firms. These feed through to the main result, theorem 3.6.1, which emphasizes that technology shocks will amplify the response of measured productivity on impact and lead to a long-run persistent effect. The concept is investigated in relation to negative shocks which cause exit, and propagate to lower long-run productivity. This is because of the analog to current empirical productivity puzzles facing economies like the UK, France and Italy.

1.1.1 *Modelling Approach*

The thesis develops a Dynamic General Equilibrium (DGE) framework that extends a Cass-Koopmans type model to account for entry, endogenous labour and imperfect competition. The work is analytical, continuous time and deterministic; parametric assumptions are only made for numerical illustrations. At any point in time an economy is defined by its levels of consumption, labour, capital and number of firms.

The thesis recognize that firm entry and exit are a slow process. This means the number of firms in the economy is a state variable, that can be thought of like capital. Similarly it recognizes that imperfect competition teamed with a fixed cost in production and increasing marginal costs (decreasing returns in variable production) leads to locally increasing returns to scale because firms produce below the cost minimizing level of output (to the left hand side of minimum on U-shaped average cost curve). The result is that variations in firm production cause procyclical productivity changes. The shocks I investigate are once-and-for-all unexpected changes in TFP measured by technology. I analyze entry and exit in the broad sense of net entry. So the process is symmetric, entry and exit do not occur concurrently. If the number of firms in the economy increases there has been entry, and if it decreases there has been exit.

1.2 THESIS CONTRIBUTION WITHIN WIDER LITERATURE

New Keynesian theory, beginning with Blanchard and Kiyotaki 1987 and Dixon 1987, extended neoclassical theory to account for product differentiation and imperfect competition in general equilibrium. This development superseded neoclassical RBC models characterized by perfect competition, constant returns to scale and zero fixed costs. In RBC research goods are priced at marginal cost, profits are superfluous and as a result investigating the microeconomic structure that determines number of firms, their interaction, and production is an unrewarding research path. However, the New Keynesian introduction of monopolisitic competition, following Dixit and Stiglitz 1977 and Spence 1976, creates a role for these microeconomic market structures by embedding a separate industry equilibrium within general equilibrium. Monopolistic competition formalizes Chamberlin 1933 theory that firms face downward sloping demand curves (have some market power) but there is free entry so the marginal firm makes zero profits. In this environment the number of active firms becomes an important determinant of output per firm and pricing markups, but New Keynesian literature has initially simplified firm dynamics' effects by assuming both number of firms and markups are exogenous. Similarly to recent papers by Bilbiie, Ghironi, and Melitz 2012 (BGM)[1] and Etro and Colciago 2010, I repeal these two assumptions (1. exogenous number of firms 2. exogenous

1 Widely cited as lengthy working paper Bilbiie, Ghironi, and Melitz 2007.

markup) to endogenise market structures so that the number of firms in the economy and the markup they charge are endogenous[2]. Chapter 2 of the thesis repeals the first assumption, then chapter 3 repeals the second. The next two sections investigate how these two assumptions have been addressed in the literature. The contribution of my research is to interpret their importance analytically, and the reward is to better understand why firm dynamics create endogenous and persistent productivity movements.

1.2.1 *How to endogenize number of firms?*

The introduction of sunk entry costs for prospective firms endogenizes firm entry. In this environment a prospective firm's entry decision is to compare the sunk cost with discounted future profits from incumbency, and in free entry equilibrium these equate so a marginal firm is indifferent between entry and inactivity. This has become quite standard, a more innovative consideration is timing. Is entry static or dynamic? In static models entry is instantaneous so free entry equilibrium arises immediately (profits are zero immediately), whereas in dynamic models there is a short-run phase during which profits are nonzero as they are arbitraged but in the long run free-entry equilibrium arises leading to the same equilibrium outcomes as the static case. Early work in the area by Chatterjee and Cooper 1993[3] and Devereux, Head, and Lapham 1996b,

2 Etro 2009 terms this approach *'endogenous market structures'*.
3 Finally published as Chatterjee and Cooper 2014.

Devereux, Head, and Lapham 1996a focused on the static case and interactions with monopolistic competition, but static models are frictionless, there is no sunk cost to entry and profits are arbitraged immediately (entry is bound by fixed cost of production which is equivalent to a sunk cost if entry occurs instantaneously). Therefore these models avoid the procyclical relationship between profit and entry. Some extensions of the static case, notably Comin and Gertler 2006, Jaimovich and Floetotto 2008 and Jaimovich 2007, have focused on the endogenous markups effect of entry, rather than the time-to-build lag (dynamic entry) that is crucial in this thesis and most recent literature.

The recent popularity of teaming entry with macroeconomics is centred on the advantages of dynamic entry which adds an endogenous propagation mechanism that the standard RBC setup lacks. Dynamic entry means that firms are fixed in the short-run which appeals to intuition and evidence Cook 2001. It can be added to a standard RBC environment in the absence of the New Keynesian extensions of imperfect competition. In an RBC environment movements in capital stock do not create enough variation in state variables. With dynamic entry the number of firms becomes an additional state variable which is mathematically analogous to capital. Brito and Dixon 2013 provide theory to explain the complex topology of dynamics that arises with the addition of this second state variable even without imperfect competition complications. And this is corroborated by nonmonotone (hump-shaped) impulse response functions in more applied work by Bilbiie, Ghironi, and Melitz 2012

(BGM). The two papers have a similar mathematical structure but differ in their construction of the entry process. The first paper is based on the earliest work on dynamic entry by Datta and Dixon 2002, Aloi and Dixon 2003 and Aloi and Dixon 2002. These papers' strength is analytical tractability–a benefit I exploit. The second paper by BGM offers an accessible calibrated, simulated DSGE approach–it has stimulated the recent popularity in this area. They provide a simple additional equation determining number of firms that DSGE modelers can port into larger models. The impulse response functions of this work and research based on it reflect the additional dynamic propagation introduced by the extra state variable and explained in the earlier theoretical line of research. Specifically productivity shocks cause hump-shaped responses in output, consumption and number of firms. My contribution is to deepen our understanding of the productivity propagation in these recent works by using the tools of the theoretical line of research.

In both sets of literature a firm is a product, so firm entry can be thought of as product entry–procylical firm entry is analogous to procyclical product variation. This interpretation has been in place since seminal work by Chatterjee and Cooper 1993, but likely understates product space variations by ignoring multiproduct firms. Minniti and Turino 2013 recognize that entering firms produce multiple products by disentangling the entry decision and product production decision. The extra competition from entry encourages product expansion creating joint procyclicality. The heterogeneous firm literature I discuss in section 1.2.3 is also successful in distinguishing firm from product.

Most of these papers focus on entry and exit as distinct processes. The result is that most papers focus on firm entry with firm exit ignored as an exogenous process. This follows BGM who assume an exogenous exit rate (a percentage die each period, like capital depreciation). Hamano and Zanetti 2014 is the first paper to endogenize exit in this setup, drawing the result that the survival of firms (i.e. lack of exit) is an important dynamic for recovery from recession. The Datta and Dixon 2002 method treats entry as symmetric, so negative entry is exit. This allows me to discuss firm dynamics in a more general business formation sense, and is useful as I focus on negative shocks (related to contemporary productivity puzzles) for which firm exit is more important.

1.2.2 *How to endogenize markups?*

Endogenous markups means the markup of price above marginal cost depends on the number of firms in an industry. In a vanilla New Keynesian model with imperfect competition the price markup is exogenous: it is a fixed constant that depends on the exogenous structural parameter 'intersector substitutability of goods' (it diminishes to nothing when sector goods are highly substitutable). Number of firms affecting markup can arise through alternative setups: a demand-side approach as in BGM and a supply side approach as in Etro and Colciago 2010. Lewis and Poilly 2012 compares the two, finding that the demand-side (translog preferences) approach is more empirically relevant than the supply-side strategic interactions approach. Despite the puzzle that the supply-side approach is weaker, most theoretical researchers prefer it due to its ubiquity in structural IO and evidence for it in empirical IO literature. Furthermore it ties more closely to firm entry, whereas the demand-side approach suits discussion of product entry. The demand-side approach reflects that entry alters the substitutability between goods that characterizes the constant markup. This is due to Feenstra 2003 translog preference specification (when there are more firms there are more goods, substituability increases and markups fall)[4]. The supply-side approach captures strategic interactions: more firms dilute the price setting ability of incumbents which lowers their markups. Therefore both lead to countercyclical markups. Papers that combine the two assumptions to numerically investigate macroeconomic questions have proven very successful in answering questions on monetary policy (Lewis and Stevens 2015, Lewis and Winkler 2015a Winkler and Lewis 2014), fiscal policy Lewis and Winkler 2015b and generally at improving moment-matching of benchmark models Colciago and Etro 2010. An important aspect of my work is to avoid relying on numerical simulation in order to improve our understanding in general of how firm dynamics affect the economy. Of course, this simplicity trade-off precludes answer-

4 Advocates suggest most pressure on markups comes from existing firms switching products and hence is demand-side preference driven, rather than the small effects of firm entry usually by insignificantly sized firms. This however requires a switch to thinking of firms as products, so undermines discussion of firm entry.

ing some of the interesting questions addressed by computational methods, but it offers a preciser narrative to attach to these investigations.

1.2.3 *The View from Outside Macroeconomics: Firm Heterogeneity*

A shortcoming of the literature I have reviewed is absence of firm size and productivity differences, a benchmark feature in structural industrial organization (IO) modeling. It is well documented that firm movements are attributable to small (potentially insignificant) firms Giovanni and Levchenko 2013. This causes skepticism of entry and exit influence over the business cycle. Complementing the literature I have discussed is an equally large body that recognizes firm heterogeneity. Typically this literature defines *'firm dynamics'* by entry, exit, size, age and life cycle features. It builds on the work of Hopenhayn 1992a, Hopenhayn 1992b, Hopenhayn and Rogerson 1993, Veracierto 2008, Campbell 1998. Whereas literature I have explored, like mine, stems from macroeconomists (and trade economists) questioning the micro underpinnings of market structures. The complementary literature with greater emphasis on heterogeneity stems from microeconomists questioning the aggregate implications of IO work. A small number of papers begin to link the two literature. For example, Clementi and Palazzo 2013 shows that allowing for firm heterogeneity (idiosyncratic productivity) and entry-exit improves internal propagation. Exit is endogenously determined by the stochastic idiosyncratic productivity component that alters operating costs and there is capital accumulation. However, the analysis is partial equilibrium–output price is give. This setup, like Campbell 1998, Samaniego 2008 Lee and Mukoyama 2008[5], is perfectly competitive. These studies succeed in distinguishing firms from products; in fact, they abstract from from modeling the household investment and consumption decision so the impact of products on utility is irrelevant, but this is a double-edged sword as it abstracts from general equilibrium. Samaniego 2008 deserves mention as a paper that reports negative findings. He suggests that firm entry and exit are not influential at an aggregate level. But this is debated by Lee and Mukoyama 2015[6] who conclude entry and exit are necessary to match data, and question Samaniego's assumptions on entry costs. To complete discussion on using heterogeneity among firms to explain aggregate puzzles one must mention Gabaix 2011 that has ignited a movement recognising that very large firms can wield influence at the aggregate level in the economy as supported by growing empirical evidence Giovanni, Levchenko, and Mejean 2014 and Stella 2015.

Ottaviano 2011 and Peters 2013 are examples of links in the opposite direction: macro papers at core but recognising firm heterogeneity. Additionally both papers are analytically tractable. Ottaviano 2011 introduces firm heterogeneity and endogenous markup using a linear quadratic demand function in a two-sector model. He seeks to understand how heterogeneity in productivity feeds through to technology shock propagation

5 Hopenhayn and Rogerson 1993 is the mutual feature of these papers.
6 Part of which is from the widely cited Lee and Mukoyama 2008.

by recognizing that most entry-exit activity arises among low productivity firms. Peters 2013 considers the impact of supply-side markup competition on aggregate TFP. His innovation is to consider that there is heterogeneity in production which feeds through to the markup measure. His interest is misallocation of resources across firms. He finds that the TFP gap depends on the marginal distribution of markups only and not the correlation between markups and firm productivity. Therefore more entry reduces markups which reduces misallocation across firms (there is tighter distribution of markups).

Lastly some very recent papers have made the natural connection from firm dynamics in the macroeconomy to stock market dynamics. These finance oriented papers (Loualiche 2014 and Corhay, Kung, and Schmid 2015) recognise the link between firm dynamics and asset price in general equilibrium. Both follow the BGM firm evolution approach.

1.3 DESCRIPTIVE STATISTICS ON ENTRY, PRODUCTIVITY AND THE BUSINESS CYCLE

A fundamental implication of my research is that firm entry is procyclical and correlated with TFP movements. In this section I outline these relationships with US data.

Large volumes of research have sought to understand the empirical relationship between firm dynamics and aggregate movements. The stylized facts are that entry (business formation) and productivity are procyclical, and markups are countercylical (though this is contested, Nekarda and Ramey 2013). An early reference on the topic is Portier 1995 who finds current entry and GDP are correlated 0.53 from 1977-1989 in French data. More recently Lee and Mukoyama 2008 use the US Annual Survey of Manufacturers to show at the plant level that entry correlates 0.413 with output growth. Rotemberg and Woodford 1999 is a good survey of countercylical markup behaviour over the business cycle for US data, though it relates to work on sticky prices. An important benefit of my entry narrative is that profits are procylical and markups countercyclical, a feat the sticky price literature struggled to replicate. Literature in industrial organization strongly suggests markups decline with competitive pressure from entry (Campbell and Hopenhayn 2005), but this does not draw a relationship to the correlation between entry and GDP and thus the business cycle significance. The procyclicaltiy of productivity is well-documented, a common reference is Rotemberg and Summers 1990. Of core interest to this thesis is the negative side of the relationship: the fall in productivity associated with recession that is outlined in UK data by Haskel, Goodridge, and Wallis 2015.

In the following analysis data is logged and Hodrick-Prescott filtered[7]. TFP and GDP data are taken from FRED (Federal Reserve Economic Database). Net Business Formation (NBF) is an index (1967=100) of business formation kept by the US Bureau of Economic Analysis 1948:1 - 1995:3[8]. Figure 1 plots comovements between US net business formation (NBF) and detrended

7 The chapter appendix includes detailed descriptions and the primary source for NBF.

8 Thanks to Vivien Lewis in helping me to access this data.

output. The correlation across the period is 0.59, similar to figures of 0.6 for France 1993-2002 documented by Dos Santos Ferreira and Dufourt 2006 and strengthening the Ambler and Cardia 1998 figure of 0.58 for US 1954-1991 (I use their data but extend the time series). The graph also suggests NBF is more volatile than output over the business cycle which pairs with the well-known business cycle fact of excess investment volatility[9].

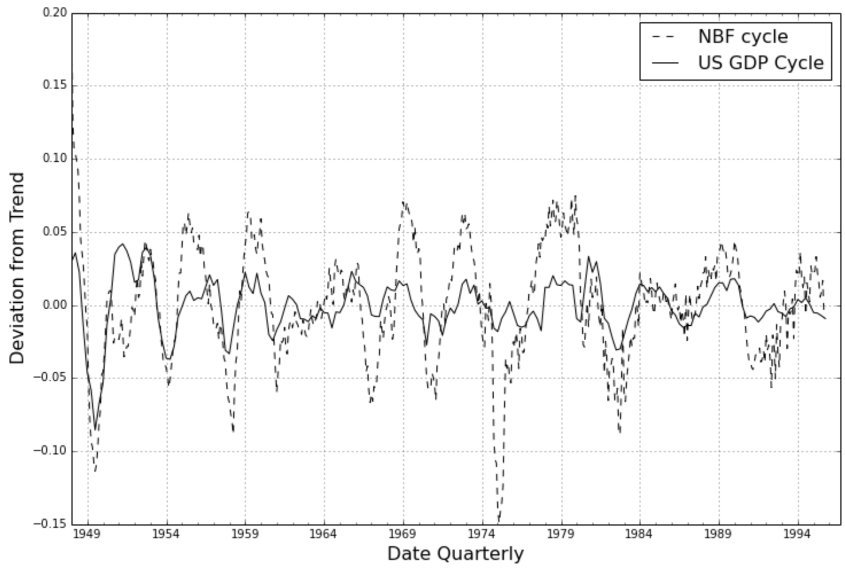


Figure 1.: US Net Business Formation and Output Business Cycle Relationship

A more discerning correlation across the period captures this volatility and confirms the statistical significance of the correlation. Figure 2 plots the beta values of NBF given a movement in output using a simple dynamic volatility model $NBF = \alpha + \beta GDP$. Therefore the y-axis is the the coefficient β from the

9 The other common stylized facts are less volatile consumption than GDP and equally as volatile labour.

dynamic OLS regression. It is dynamic because the regression is run on the previous 7 years, so 28 quarterly observations. For example, in 1994 the beta value on the y-axis is 2.0, so over the 7 years previous (1997-1994) a 1% rise in output above trend is associated with a 2% rise in net business formation above trend. The 95% confidence intervals (dashed) show the relationship is statistically different to 0.0 for the whole time period, excluding the late 60s during which there was spurious volatility in NBF.

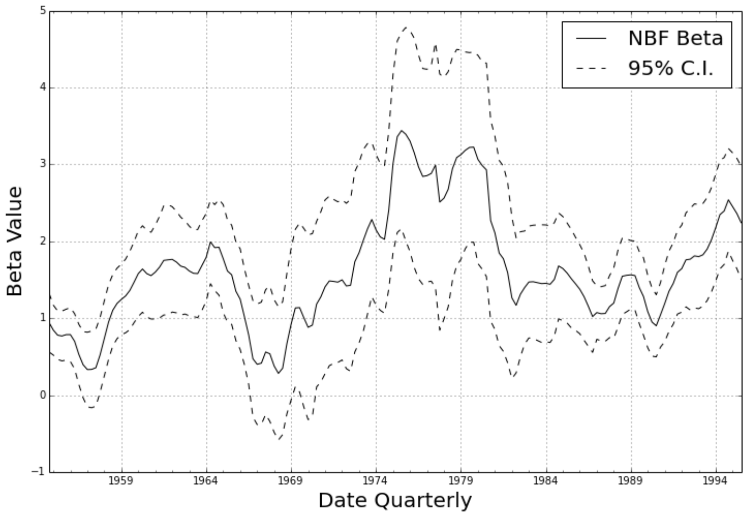


Figure 2.: US Net Business Formation and Output Correlation and Volatility

Figure 3 adds TFP (as measured in the Penn World tables) to figure 1 and supports its positive correlation with output and net business formation. This graph embodies the main relationships of the thesis. Falling output diminishes business formation and worsens measured productivity[10]. Rising output raises business formation and improves measured productivity.

10 Measured productivity since the causal reason for the initial output movement could be a shift in underlying productivity (a technology shock).

This thesis provides a structural explanation to link these stark correlations.

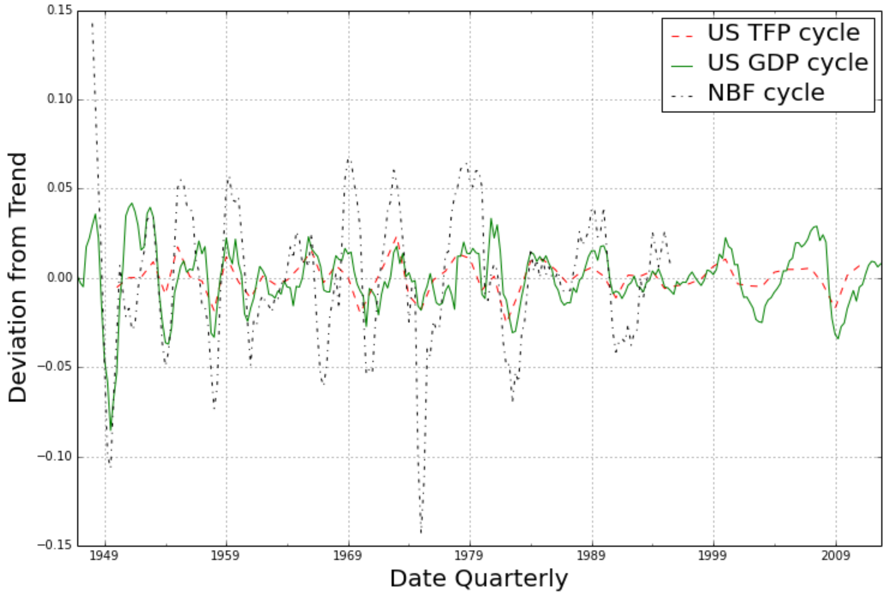


Figure 3.: US Net Business Formation, TFP and Output Correlation

2

THE EFFECT OF FIRM ENTRY ON MACROECONOMIC PRODUCTIVITY WITH FIXED MARKUPS

This chapter argues that firm entry causes endogenous fluctuations in macroeconomic productivity through its effect on incumbent firms' capacity utilization. The analysis shows that imperfect competition causes long-run *excess entry*: too many firms each with excess capacity. Since entry occurs slowly, macroeconomic shocks are initially borne by the excess capacity incumbents who respond by altering their capacity utilization. Incumbents' efficiency changes because of non-constant returns to scale which aggregates to affect the economy's productivity. In the long run, entry occurs and new firms absorb the shock, which alleviates incumbents' alteration in capacity. Therefore the productivity change is ephemeral. In sum, the slow response of firms to economic conditions causes endogenous productivity dynamics.

2.1 INTRODUCTION

How does industry competition affect firm entry and consequently macroeconomic dynamics? Since Chamberlin 1933, economists have understood that in a monopolistically competitive economy there is *'excess capacity'*. Each firm produces below their full capacity, which minimizes costs, because underproduction earns monopoly profits. In terms of firm entry, underproduction causes too many firms each producing too little. This chapter's focus is transition toward the excess ca-

0 This Chapter is an adaptation of a working paper co-authored with Professor Huw Dixon called "The Effect of Firm Entry on Capacity Utilization and Macroeconomic Productivity".

pacity steady state. During transition the capacity utilization of firms fluctuates which causes endogenous productivity variations. The productivity variations arise because capacity levels alter returns to scale and therefore productivity. The firm entry capacity utilization mechanism is unexplored in macroeconomics, hence the aim of the chapter is to provide an analytically tractable theory to connect firm entry, capacity utilization and productivity dynamics. The theory should interest empiricists in light of contemporary 'productivity puzzles' Haskel, Goodridge, and Wallis 2015. For example, our theory implies that a negative shock to the economy is not immediately dissipated by firms exiting, and thus worsens capacity utilization which exacerbates negative productivity until firms have time to exit. This scenario could relate to large falls in productivity experienced by the UK economy post Great Recession[1].

The chapter's main result is that resources are divided between too many firms in the long run, but a movement away from equilibrium causes firms to increase (reduce) capacity so they produce closer to (further from) full capacity in the short run whilst other firms enter to arbitrage monopoly profits (exit to avoid negative profits). For example, a positive production shock causes a short-run productivity gain. Incumbents raise output as other firms cannot enter in the short run to absorb the positive shock. On raising output, increasing returns improve incumbents' productivity which aggregates to the macroeconomy. Since the output of the incumbents exceeds equilibrium and yields monopoly profits prospective firms slowly enter to

1 I investigate this in greater depth in Chapter 3.

arbitrage the profits back to zero profit long-run equilibrium. Therefore the productivity gain is ephemeral, and excess capacity returns in the long run. A clear implication of this model is that capacity utilization is procylical: during expansion capacity utilization increases (less excess capacity), but during booms capacity utilization decreases (more excess capacity.) Figure 4 emphasizes this relationship in quarterly US data 1974-2013[2]. It is clear that the two series co-move and capacity utilization is more volatile. Furthermore the volatility in capacity utilization increases in downturns for example during the recent Great Recession 2007-2009 output falls below trend, but the accompanying fall in capacity utilization (opening up in capacity) is even greater. To capture the relationship more robustly figure

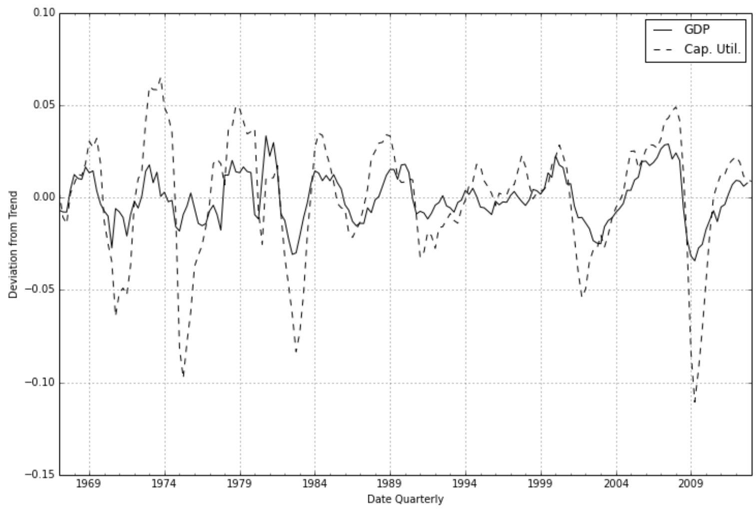


Figure 4.: US Procylical Capacity Utilization

2 Data are logged and HP-filtered at a quarterly frequency. Raw data is taken from Federal Reserve FRED database with unique MNEMONICS GDP and TCU. *"Capacity utilization is the percentage of resources used by corporations and factories to produce goods in manufacturing, mining, and electric and gas utilities for all facilities"*.

5 shows the betas associated with capacity utilization from a dynamic OLS regression over the period. Each value on the y-axis is the β from regression $Cap.Util. = const + \beta GDP$ run for the previous 28 periods (7 years). The plot confirms that the effect of a change in GDP on capacity utilization is positive and statistically significant e.g. the value of 1.5 in 2009 implies over the previous 7 years a 1% rise in GDP caused a 1.5% rise in capacity utilization[3].

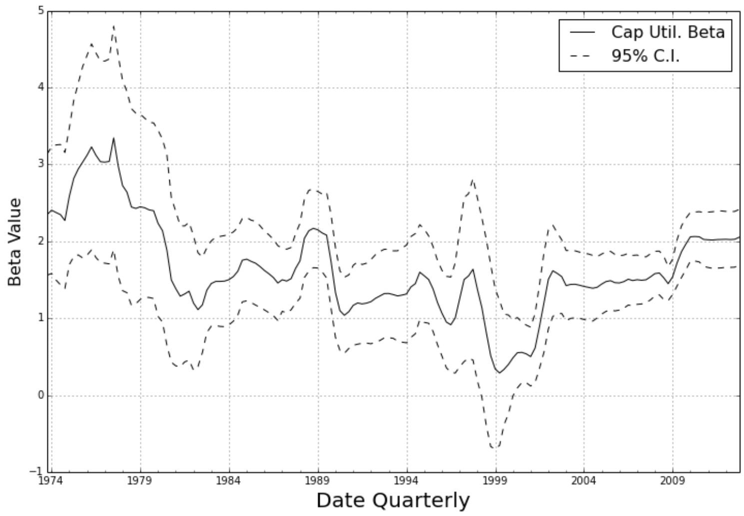


Figure 5.: The Effect of GDP on Capacity Utilization

The focus on theory and analytical dynamics distinguishes this chapter from other research (discussed below) that typically focuses on quantitative replication of empirical results. These quantitative works are stochastic and dynamics are simulated, as opposed to our qualitative deterministic study that avoids specifying functional forms or solving parameterized,

3 Additionally the volatility of capacity utilization is typically twice that of GDP and increases in the periods after a recession, like the early 2000s recession and late 2000s recession. This interesting result is not core to the chapter, but is too striking to ignore comment.

path dependent models. This approach allows us to give a robust narrative of the role entry is playing in the model by pinpointing its interactions with other variables. The model includes important features from the main papers in the literature so far: endogenous labour, and capital accumulation (Bilbiie, Ghironi, and Melitz 2012); imperfect competition (Etro and Colciago 2010) and entry congestion effects that Lewis 2009 finds empirically important and Berentsen and Waller 2009 model structurally[4]. Imperfect competition is the paramount addition since it is ubiquitous in these models but its relation to entry and mapping to dynamics has not been investigated. Although in an open-economy setting without capital Aloi and Dixon 2003 follow a similar line of argument that imperfect competition creates excess capacity that can lead to endogenous fluctuations in productivity, but in this chapter we focus on relating these arguments to a business cycle framework including capital. Excess capacity is a standard feature of models with imperfect competition, but its relationship to entry is often overlooked, since three features–imperfect competition, non-constant returns to scale, slow firm entry–are necessary for capacity utilization effects. The counterfactuals of these three assumptions clarifies their necessity: 1) Without slow entry, there is standard instantaneous free entry[5], short-run capacity utilization by incumbents will not arise because other firms enter instantaneously to meet output, so incumbents do not respond by varying their capacity. 2) If there were perfect competition,

4 Congestion means entry costs increase with number of entrants.
5 The two extremes entry cases–instantaneous free entry and no entry fixed number of firms–are special cases the model.

firms produce efficiently, at minimum average cost; they do not have excess capacity which if used can improve productivity. 3) If there were constant returns to scale, entry is inert: there is no difference between one large firm producing all output versus many small firms producing all output[6].

To understand the relationship between entry, imperfect competition and macroeconomic dynamics, we build on work by Brito and Dixon 2013. The model is a Cass-Koopmans model with labour-leisure choice and capital. We add Datta and Dixon 2002 entry and imperfect competition in the product market, so firms control output price. In the model a firm is a divisor of resources, and with this apparatus we formalize our intuitions to ask: How does imperfect competition affect the division of resources (capital and labour) among firms when there is entry? How does productivity vary as we transition toward this division of capital to firms?

Related Literature The chapter bridges two groups of literature: quantitative endogenous entry DSGE models and qualitative dynamical systems analyses of determinacy, returns to scale, capacity utilization, and imperfect competition. A burgeoning group of endogenous entry DSGE papers show promising quantitative results from including richer industrial organization features in standard RBC models. Recent cornerstones are Bilbiie, Ghironi, and Melitz 2012, Etro and Colciago 2010, Lewis 2009 and Jaimovich and Floetotto 2008. The former two provide initial forays into qualitative dynamics of the systems

6 Furthermore with a fixed cost, like we have, there would be a one firm natural monopoly.

they simulate, mainly these are complementary exercises to establish saddlepath stability, all in discrete time due to simulation. A simplicity trade-off is to reduce the dimensionality to one, so a state variable (capital) is stripped from the system leaving only number of firms state, which rules out the most empirically appealing dynamics of their simulations: those that exhibit nonmonotone impulse responses.

The closest model to ours is Brito and Dixon 2013; the paper studies entry implications for macroeconomic dynamics without imperfect competition. They show that with perfect competition, firm entry causes empirically plausible macroeconomic dynamics in a Ramsey model. The main result is that entry is sufficient to give nonmonotone deviations from equilibrium under fiscal shocks. The research derives sufficient conditions for hump-shaped responses that quantitative DSGE papers observe via simulation. Our addition of imperfect competition creates the capacity utilization mechanism and excess entry[7]. This concept is also present in Aloi and Dixon 2003 which relates entry to capacity utilization in a simpler open-economy framework without a second state variable in capital. They find that demand and technology shocks cause endogenous variations in measured productivity through changed in firm capacity utilization. In these papers and our model the sunk entry cost is endogenous, and firms' technology gives U-shaped average

7 Etro and Colciago 2010 also note that Cournot competition causes inefficiency through excess entry. Etro 2009 provides an excellent survey of macroeconomic models with endogenous entry and endogenous market structures.

cost curves[8], which endogenously generates increasing returns to scale under imperfect competition, as in Jaimovich 2007. The endogenous sunk cost of entry depends on the number of entrants, a so-called congestion effect that recent papers have emphasised the importance of: Lewis 2009 offered initial empirical support for congestion effects of firm entry. It is a VAR study that shows congestion effects can account for observed lags in monetary policy.

Roadmap – Section 2.2 proposes a model of firm entry in the macroeconomy and derives measured productivity. The analysis follows and consists of section 2.4 on comparative statics, stability and comparative dynamics. The final section 2.5 uses the analytical results to reveal the main result that productivity varies endogenously, so causes measured productivity to exceed underlying.

2.2 ENDOGENOUS ENTRY MODEL WITH IMPERFECT COMPETITION

The model follows a Ramsey-Cass-Koopmans setup with additions of imperfect competition, firm entry, and capital accumulation. The model is deterministic, and labour is endogenous as developed by Brock and Turnovsky 1981. There are two state variables: capital and number of firms $(K, n) \in \mathbb{M} \subseteq \mathbb{R}^2$, where $\mathbb{M}$ is the state space of the control problem that later forms a subset of the general dynamical system state (or phase) space.

8 Often assumed in industrial organization literature such as Luttmer 2012 and Rossi-Hansberg and Wright 2007.

We solve the model as a decentralised equilibrium because imperfect competition distorts the optimising behaviour of the firm which does not coincide with centralised equilibrium of a planner optimising behaviour of the economy. [9]

2.2.1 *Household*

The economy is inhabited by a continuum of infinitely-lived identical households. A household seeks policy functions of consumption $\{C(t)\}_0^\infty \in \mathbb{R}$ and labour supply $\{L(t)\}_0^\infty \in [0,1]$ that maximise lifetime utility

Assumption 1 Household Utility.
Aggregate utility $\mathrm{U} : \mathbb{R}^2 \rightarrow \mathbb{R}$ *is composed of individual utility* $u : \mathbb{R} \times [0,1] \rightarrow \mathbb{R}$ *which is jointly concave and differentiable in both of its arguments. It is strictly increasing in C and strictly decreasing in L, so* $u_C > 0, u_L < 0$. *Thus* u_L *should be read disutility from labour (i.e.* $-u_L$ *is utility from labour). Both goods are normal* $u_{CC},\ u_{LL} < 0$, *so marginal utility of consumption and disutility of labour are diminishing. Utility is additively separable* $u_{CL} = 0$.

A representative household solves the utility maximisation problem:

$$\mathrm{U} := \int_0^\infty \mathrm{u}(C(t), 1 - L(t))\, \mathrm{e}^{-\rho t}\, \mathrm{dt} \tag{1}$$

$$\text{s.t.} \quad \dot{K}(t) = rK(t) + wL(t) + \Pi - C(t) \tag{2}$$

9 The conditions we derive could be derived from an optimal control problem with two restrictions, but the economic intuition is less clear. However, the equivalence is important for the theory of dynamics we use later.

where e is the exponential function and $\rho \in (0,1)$ is the discount factor over time $t \in \mathbb{R}_+$. The household owns capital $K \in \mathbb{R}$ and takes equilibrium rental rate r and wage rate w as determined by the market (we determine their levels in the firm equilibrium, whereas households take them as constant). A dot above a variable denotes time derivative $\dot{x} = \frac{\partial x}{\partial t}$. Households own firms and receive firm profits $\Pi \in \mathbb{R}$. There are $n \in \mathbb{R}$ firms in the aggregate economy, or more accurately there are n active firms per industry which translates to n in aggregate under a continuum of symmetric industries of measure 1. Using the Maximum Principle to maximise utility yields six Pontryagin conditions for optimal consumption and labour [10] The conditions simplify to an intertemporal consumption Euler equation (3), intratemporal labour-consumption trade-off (4) and the resource constraint (2).

$$\dot{C}(r,\rho) = \frac{C}{\sigma(C)}(r-\rho), \quad \text{where } \sigma(C) = -C\frac{u_{CC}(C)}{u_C(C)} \tag{3}$$

$$w = -\frac{u_L(L)}{u_C(C)} \tag{4}$$

The solution of the dynamic optimization problem for household consumption will be one of the solutions of this system of two differential equations that satisfy the initial condition $K(0) = K_0$. To complete the boundary value problem we add two transversality conditions on the upper boundary. This completes the unique solution for the boundary value problem: three variables (K, L, C), three equations (2)-(4), three bound-

10 Appendix B.1 sets up the Hamiltonian and solves it.

ary conditions (5). λ is the costate variable associated with the state variable K in the Maximum Principle.

$$\lim_{t\to\infty} K_t e^{-\rho t} \geqslant 0\,, \quad \lim_{t\to\infty} K_t \lambda_t e^{-\rho t} = 0\,, \quad K_0 = K(0) \tag{5}$$

Therefore equations (2)-(5) characterize optimal paths of consumption and labour. In equilibrium these equations continue to hold, with factor prices w and r at their market value, which arises from firms profit maximisation under imperfect competition as we show in section (2.2.2).

2.2.1.1 *Intratemporal Labour*

The intratemptoral condition (4), given wage w at market equilibrium $w(L, K, n)$, defines optimal labour choice in terms of the other variables $L(C, K, n)$, so it can be substituted out of the system, thus reducing the system's dimensionality by one. In section (2.2.2) we derive the market wage in firm equilibrium (proposition 1), but for now we assume it is known in order to understand the general equilibrium behaviour of labour.

Theorem 2.2.1 Optimal labour choice.
Labour supply increases in capital (K), number of firms (n) and decreases in consumption (C).

$$L(\underset{-}{C}, \underset{+}{K}, \underset{+}{n}) \tag{6}$$

Proof. Apply the implicit function theorem to differentiate the intratemporal condition (4) with labour defined implicitly by $L(C, K, n; A, \zeta)$. See Appendix B.3. □

The intratemporal condition shows that the marginal rate of substitution between consumption and labour equates to the wage (negative because labour decreases utility[11]). A rise in consumption causes the marginal utility of consumption to fall (consumption is a normal good) therefore marginal disutility of labour must decrease to maintain the marginal rate of substitution, hence labour increases which reduces disutility from labour u_L. When capital increases the marginal product of labour increases as they are complements, therefore wage increases which increases supply of labour.

A notable corollary to theorem 2.2.1 is that entry (rise in number of firms n) increases labour supply. This is because entry reduces capital and labour per firm which raises the marginal product of production by decreasing returns (decreasing returns means less production causes higher marginal product). Higher marginal product of labour raises wage which creates a greater incentive to supply labour.

Corollary 1 Firm Entry Encourages Labour Supply.

Firm entry causes an increase in the number of firms which raises households desired labour supply[12] $L_n > 0$ *when production has decreasing returns* $\nu \in (0,1)$*; decreases labour supply* $L_n < 0$ *when production has increasing returns* $\nu \in (1,\infty)$*; and does not affect labour supply* $L_n = 0$ *with constant returns* $\nu = 1$.

11 Think of u_L as labour disutility and $-u_L$ as labour utility.
12 I follow the notation that x_z is derivative of x with respect to z.

Proof. From the intratemporal Euler condition (4) with wage w at the market rate (14) implicitly differentiate with respect to n

$$u_{LL}L_n + u_C(1-\zeta)A\left[F_{lk}\frac{-K}{n^2} + F_{ll}\left(\frac{-L}{n^2} + \frac{L_n}{n}\right)\right] = 0,$$

$$\text{therefore } L_n = \frac{u_C(1-\zeta)A(\nu-1)\frac{F_l}{n}}{u_{LL} + u_C(1-\zeta)A\frac{F_{ll}}{n}} > 0$$

Notation is formally defined in firm section. F is the production technology that has decreasing returns; that is, we assume it is homogeneous of degree $\nu \in (0,1)$ $(hod - \nu)$ and therefore its derivative is $hod - (\nu - 1)$. $\zeta \in (0,1)$ represents market power (Lerner index). A is technology. □

The result relies on there being decreasing returns to scale $\nu \in (0,1)$ in the variable production function. With constant returns labour supply is irresponsive $L_n = 0$. This is because the effect of entry reducing capital and labour per firm does not change marginal products by definition of constant returns to scale. Therefore wage is unaffected and there is no channel through which entry affects labour. It is also noticeable that when there is absolute market power $\zeta \to 1$ entry does not affect labour supply. This is because raising market power increases markup between wage and marginal product. Therefore a change in marginal product (induced by entry and returns to scale) has little effect on wage and therefore optimal labour choice. In other words there is a one price monopsonist.

2.2.2 *Firm Production*

Imperfect competition leads to fixed markups of price above marginal cost as in the standard Dixit and Stiglitz 1977 setup. There is a continuum of n one firm industries, and substitutablity across industries is $\theta \in (1, \infty)$ which makes aggregate output a Cobb-Douglas aggregate as $\theta \to 1$. Multiproduct firms do not exist, so a firm is a producer, is a product, is an industry, is a sector. A firm faces a fixed cost and decreasing returns to scale which leads to a U-shaped average cost curve. There is imperfect competition in the product market and perfect competition in the factor market. Product market imperfect competition means that firms can control output price; factor market perfect competition means firms cannot control factor prices. That is, the market determines the interest rate and wage, which are the prices of factors capital and labour, and since each firm faces the same price, then each firm employs the same amount of capital and labour. We shall denote capital per firm and labour per firm in lower case

Definition 1 Per Firm Variables.

By assumption of perfect factor markets, capital and labour are divided equally among firms. Lower case letters denote per firm variables

$$k(t) = \frac{K(t)}{n(t)}, \quad l(t) = \frac{L(t)}{n(t)}, \quad y(t) = \frac{Y(t)}{n(t)}$$

Respectively capital per firm, labour per firm, output per firm.

The aggregate good which I take to be the numeraire is a CES aggregate of each $i \in n$ sector (a sector is a 1 firm industry, so n is the measure of number of firms). $\theta \in (1, \infty)$ is intersector substitutability which tends to a Cobb-Douglas aggregator in the imperfectly substitutable limit $\theta \to 1$. The $n^{\frac{1}{1-\theta}}$ component removes love-of-variety.

$$Y = n^{\frac{1}{1-\theta}} \left[\int_0^n y(i)^{\frac{\theta-1}{\theta}} di \right]^{\frac{\theta}{1-\theta}} \tag{7}$$

With the unit price of the aggregate good as the numeraire the sectoral demand directed at each 1-firm industry takes constant elasticity form

$$y(i) = p(i)^{-\theta} Y, \quad \forall i \in (0, n) \tag{8}$$

Later we shall see that this leads to market power captured by the Lerner Index $\zeta = \frac{1}{\theta}$, which is the difference between price and marginal cost as a proportion of price ($\frac{P-MC}{P}$). As oppose to the Lerner Index of Market Power ζ, macroeconomists often express the equivalent idea in terms of markup (μ), as I do in chapter 2, the markup is defined as price over marginal costs ($\frac{P}{MC}$) therefore $\mu = \frac{1}{1-\zeta} = \frac{\theta}{\theta-1}$ is the relationship between markup (μ) and market power (Lerner Index) (ζ).

Assumption 2 Production.

Firms have the same production technology

$$y(t) = \max\{AF(k(t), l(t)) - \phi, 0\} \tag{9}$$

where $F : \mathbb{R}^2_+ \supseteq (k, l) \to \mathbb{R}_+$ *is a firm production function with continuous partial derivatives which is homogenous of degree* $\nu < 1$ *(hod-*ν*) on the open cone* $\mathbb{R}^2_+$*, and* $\phi \in \mathbb{R}_+$ *a fixed cost denominated in output. F has concavity properties* $F_k, F_l, F_{kl} = F_{lk} > 0 F_{kk}, F_{ll} < 0, F_{kk}F_{ll} - F^2_{kl} > 0$*.*

The production function exhibits a U-shaped average cost curve because there is a fixed cost and decreasing returns to scale. Decreasing returns to scale arise because the variable production function F is convex, $hod - \nu < 1$, in capital and labour which causes increasing marginal cost. ϕ is the fixed cost and is in terms of output; it is a nonconvexity which prevents some firms producing; it occurs each period, and is different to the sunk entry cost which is paid once to enter (see Entry section 2.2.3)[13]. Inada's conditions hold so that marginal products of capital and labour are strictly positive which rules out corner solutions. Since the average cost curve is U-shaped there is an optimal efficient level of production at minimum average cost, where average cost and marginal cost intersect. These efficient levels denoted with an e superscript (x^e) coincide with the optimal firm size $Y_n = 0$ where under symmetry aggregate output is $Y = ny$

13 As in Jaimovich 2007 the role of this parameter is to reproduce the apparent absence of pure profits despite market power. It allows zero profits in the presence of market power.

$$F(k^e, l^e) = \frac{\phi}{A(1-\nu)} \tag{10}$$

$$y^e = AF(k^e, l^e) - \phi = \frac{\phi\nu}{1-\nu} \tag{11}$$

$$\mathcal{P}^e = \frac{y^e}{F(k^e, l^e)^{\frac{1}{\nu}}} = A^{\frac{1}{\nu}}\nu\left(\frac{\phi}{1-\nu}\right)^{\frac{\nu-1}{\nu}} \tag{12}$$

Technology A does not affect optimal net output, but it reduces optimal gross output $F(k, l)$ and raises productivity. *Measured productivity* $\mathcal{P}$ is output per unit of production, where production is normalized to be homogeneous of degree 1 (In the steady state section 2.3 I show that this productivity definition is a natural derivation.). From (10) $\left(\frac{1}{n}\right)^{\nu} F(K, L)^e = \frac{\phi}{A(1-\nu)}$ whence $n^e = \left(\frac{A(1-\nu)}{\phi}F(K, L)\right)^{\frac{1}{\nu}}$ so number of firms is hod-1 in capital and labour. A rise in capital and labour by some proportion will raise number of firms by the same proportion, so capital and labour per firm remain unchanged. Hence the result in (11) that optimal output per firm is a fixed level, independent of model variables.

The efficient levels that correspond to optimal firm size are not achieved when we consider the strategic interactions of firms under imperfect competition.

2.2.2.1 *Strategic Interaction: Monopolistic Competition*

Under imperfect competition, with output price as the numeraire, a firm maximises profits given real wage w and interest rate r by choosing labour and capital such that the following factor market equilibrium holds

Proposition 1 Factor Market Equilibrium.

Factor market equilibrium conditions reflect the profit maximising choices of firms competing under imperfect competition. They show that the marginal revenue product of capital equates to the cost of capital and the marginal revenue product of labour equates to the wage.

$$AF_k(k,l)(1-\zeta) = r \tag{13}$$

$$AF_l(k,l)(1-\zeta) = w \tag{14}$$

$\zeta \in [0,1)$ is a fixed markup known as the Lerner index[14]. It reflects the degree of market power. Under perfect competition $\zeta = 0$, so factor price equals marginal product, but under imperfect competition, firms charge a markup so prices exceed their marginal product. Standard derivation[15] shows that the Lerner Index is the inverse of intersector substitutability $\zeta = \frac{1}{\theta}$, therefore high substitutability corresponds to little market power. For the purpose of this research we focus on variations in ζ, rather than its microfoundations. A result of the markup is that a firm produces less and earns monopoly profits. The imperfectly competitive factor prices r and w create monopoly profits that increase firm operating profits by $\zeta \nu AF(k,l)$ relative to the perfect competition case when $\zeta = 0$. Both returns to scale of the technology and market power increase the extra profit.

14 Notice that I refer to the Lerner Index as the markup as it is useful to bind the parameter between 0 and 1. It is common in macroeconomics (as I do in second chapter) to define markup as price over marginal cost calling this definition of the markup μ we have $\mu = \frac{1}{1-\zeta}$.

15 Maximise firm profits subject to sectoral demand 8.

Proposition 2 Excess Profit.

Market power ζ suppresses factor prices which leads to excess profits π relative to the perfect competition case $\pi(\zeta) > \pi(\zeta = 0)$

$$\pi(L, K, n; A, \zeta, \phi) = (1 - (1 - \zeta)\nu)AF(k, l) - \phi \tag{15}$$

Proof. See appendix B.2. Proof follows from definition of profit $\pi(t) = y(t) - w(t)l(t) - r(t)k(t)$ with factor prices at factor market equilibrium. □

The next subsection uses proposition 2 to show that the additional profit causes firms to have excess capacity by raising entry incentives.

2.2.2.2 *Zero Profit Outcome and Excess Capacity*

The extra profit $\zeta\nu AF(k, l)$ from imperfect competition is crucial to understanding the entry-imperfect-competition relationship. Profits offer entry incentives, so monopoly profits encourage excess entry. In section (2.4.1) we show that profits are zero in steady state so there is no incentive to enter the market and entry is zero. Taking zero profits as given (an asterix x^* denotes zero profit steady state outcome under imperfect competition.), proposition 2 implies

$$F(k^*, l^*) = \frac{\phi}{A(1 - (1 - \zeta)\nu)} \tag{16}$$

$$y^*(\underset{+}{\phi}, \underset{+}{\nu}, \underset{-}{\zeta}) = AF(k^*, l^*) - \phi = \frac{\nu(1 - \zeta)\phi}{1 - (1 - \zeta)\nu} \tag{17}$$

Therefore zero-profit output per firm is increasing in both fixed cost and returns to scale and is decreasing in market power.

More market power raises marginal revenue products of inputs, so less needs to be produced in order to cover fixed costs and attain zero profits. Importantly y^* is unaffected by TFP (A). Firms always produce the same long-run level of output regardless of technology.

With zero profit, the profit definition $\pi = y - wl - rk$ implies

$$y^* = wl^* + rk^* = (1-\zeta)A\nu F(k^*, l^*) \tag{18}$$

$$\begin{aligned} \mathcal{P}^* &= \frac{y^*}{F(k^*, l^*)^{\frac{1}{\nu}}} = F(k^*, l^*)^{\frac{\nu-1}{\nu}} A\nu(1-\zeta) \\ &= A^{\frac{1}{\nu}}\nu(1-\zeta)\left(\frac{\phi}{1-(1-\zeta)\nu}\right)^{\frac{\nu-1}{\nu}} \end{aligned} \tag{19}$$

Measured productivity $\mathcal{P}$ is output per unit of production, where production is normalized to be $hod - 1$ [16]. The definition is a generalization of productivity with constant returns to scale: under constant returns to scale $\nu = 1$ then $\mathcal{P}^{\text{CRTS}} = (1-\zeta)A$, so that market power determines measured productivity. With CRTS and perfect competition $\zeta = 0$ measured productivity is TFP A. Importantly measured productivity is decreasing in market power because it allows firms to suppress output more and benefit less from returns to scale.

As mentioned the outcome is technically efficient under the special case of perfect competition, $\zeta = 0$. Comparing the outcome under imperfect competition to the efficient outcome ((10) and (11) with (16) and (17)) demonstrates excess capacity.

Theorem 2.2.2 Excess Capacity.

In zero-profit steady state variable production, output per firm and

16 since $F(\alpha k, \alpha l)^{\frac{1}{\nu}} = (\alpha^{\nu} F(k,l))^{\frac{1}{\nu}} = \alpha F(k,l)^{\frac{1}{\nu}}, \; \alpha \in \mathbb{R}$

measured productivity are less than the efficient cost-minimizing level that corresponds to optimal firm size.

$$F(k^*, l^*) < F(k^e, l^e), \quad y^* < y^e, \quad \mathcal{P}^* < \mathcal{P}^e \tag{20}$$

As ζ increases long-run output y^* decreases. Hence imperfect competition causes firms to produce less than the efficient level y^e, so AC $>$ MC and AC is decreasing, giving locally increasing returns to scale. Increasing output would improve firm productivity. With entry we shall see that firms can be manipulated to use some of this capacity. This result implies that in equilibrium the trade-off between an additional firm bringing an extra fixed cost (which is underutilized, reducing overall production possibility frontier) outweighs the efficiency gain from smaller firms producing with lower marginal cost.

Remark 1. Zero profit equilibrium only exists if the denominator is positive $\nu < \frac{1}{1-\zeta}$ which will always hold under our assumption of decreasing returns $\nu < 1$. But we note that as $\zeta \to 1$ monopoly power allows ever increasing returns to scale $\nu \to \infty$. If there is little monopoly power $\zeta \to 0$ then decreasing returns are necessary for existence of zero-profit equilibrium since no firm is large enough to make use of the fixed cost. With CRTS $\nu = 1$ then $\zeta > 0$ because with constant returns and a fixed cost the economy cannot be competitive.

2.2.3 *Firm Entry*

The number of firms at time t is determined an endogenous sunk cost of entry and an arbitrage condition that equates entry cost with incumbency profits.

Assumption 3 Sunk Entry Cost (congestion effect).
Entry cost $q \in \mathbb{R}$ increases with the number of entrants $\dot{n}$ in t.

$$q(t) = \gamma \dot{n}, \qquad \gamma \in (0, \infty) \tag{21}$$

The process is symmetric, a prospective firm pays q to enter; an incumbent firm pays $-q$ to exit. If there is exit $\dot{n} < 0$, so $q < 0$ and $-q > 0$– an incumbent must pay a fee to exit. $\dot{n}$ is the change in the stock of firms so represents net business formation; later we define this as 'entry' (definition 2). γ is the marginal cost of entry, and its bounds capture the limiting cases of entry. $\gamma \to 0$ implies instantaneous free entry because the sunk cost is vanishingly small so the outcome is similar to the static case, and $\gamma \to \infty$ implies fixed number of firms because the sunk cost is so high that it prohibits entry.

Assumption 4 Entry Arbitrage.
Gain from entry equals return from investing the cost of entry at the market rate.

$$\dot{q}(t) + \pi(t) = r(t)q(t) \tag{22}$$

A firm's gain from entry is the value of incumbency which is operating profits π plus any change in the original sunk cost $\dot{q}$.

The two assumptions are well supported. The congestion effect assumption is common in industrial organization literature for example Das and Das 1997. Justified by increased advertising costs, or competition for a fixed resource, and variously called "negative network effects" or "entry adjustment cost". In terms of the macroeconomy, Lewis 2009 statistically models congestion externalities in entry and concludes that the mechanism improves model fit because it reduces impact responses of entry. Berentsen and Waller 2009 also model a congestion effect in a DSGE model.

The arbitrage assumption implies that the return to investing in a firm is equal to the return of that investment at the risk free rate r. And an implication of this is that the value of a firm is equal to present discounted value of future profits as in Bilbiie, Ghironi, and Melitz 2012.

The two assumptions form a dynamical system in number of firms and cost of entry $\{n, q\}$ which reduces to a second-order ODE in number of firms

$$\gamma\ddot{n}(t) - r(t)\gamma\dot{n}(t) + \pi(t) = 0 \tag{23}$$

To interpret this second-order ODE consider that profits are high, then to maintain equilibrium the speed of net business formation $\dot{n}$ is high which translates to higher sunk entry cost thus discouraging future entry so net business formation decelerates $\ddot{n} < 0$ to maintain equilibrium. Defining entry can help this intuition, and by defining entry, this second-order ODE is separable into two first-order ODEs

Definition 2 Entry and Exit.

Entry (or exit) is measured by the the change in the number of firms. Negative entry is exit.

$$e(t) = \dot{n} \tag{24}$$

Hence our model of industry dynamics, which determines the number of firms, is defined by two ODEs

$$\dot{n} = e \tag{25}$$

$$\dot{e} = -\frac{\pi(t)}{\gamma} + r(t)e(t) \tag{26}$$

The endogenous entry cost causes a non-instantaneous adjustment path to steady state, which provides an analytical framework to understand short-run dynamics. To observe the importance of the endogenous entry cost consider the contradiction that entry cost is fixed, $q(t) = \gamma$. In which case the the second-order ODE becomes $\pi = r\gamma$ so there is no dynamic entry. Datta and Dixon 2002 show that the level of entry that satisfies the first order differential equation in entry arises when cost of entry is equal to future discounted operating profits.

2.3 AGGREGATION AND GENERAL EQUILIBRIUM

The last section discussed firm level production and showed that imperfect competition caused variable production, output and productivity of a firm to be below the perfect competition case in the long run, so-called excess capacity. Importantly these levels are fixed and independent of model variables in the long run because capital per firm and labour per firm always return to the same level–firms enter or exit until the desired per firm levels arise. That each firm ultimately produces the same output, has implications for aggregation: when there is a rise in firms, aggregate output will increase; when there is a fall in firms, aggregate output will decrease. Both productivity and output per firm will always tend to the same level, whereas aggregate output will change, it has constant returns to scale.

The first firm to enter in a period pays 0, whereas the second firm will pay γ and the third firm to enter 2γ and so on. Therefore the economy wide cost of entry, $Z(t)$, is

$$Z(t) = \gamma \int_0^{e(t)} i \, di = \gamma \frac{e(t)^2}{2} \tag{27}$$

Hence aggregate profits are aggregate operating profits less entry costs

$$\Pi = n\pi - Z(t) \tag{28}$$

$$\Pi = n[AF(k,l)(1-(1-\zeta)\nu) - \phi] - \gamma \frac{e(t)^2}{2} \tag{29}$$

Under symmetry aggregate output is the sum across each industry that produces according to the firm level production function.

$$Y = ny = n \left[A\mathrm{F}\left(\frac{K}{n}, \frac{L}{n}\right) - \phi \right] \tag{30}$$

We previously noted at the firm level output y is homogenous of degree 0 in inputs (K, L, n). This is despite the production technology being hod-ν, as a proportional rise in capital and labour is offset by the rise in number of firms. Hence we view firms as a type of quasi-input in production. Therefore the aggregate production function Y is homogeneous of degree 1 in (K, L, n)[17]. Consider a doubling of all inputs (double capital, labour and number of firms), all firms remain as they were, productivity and output at the firm level are unaffected, so aggregate output also doubles because there are twice as many firms each producing the same amount as before.

2.3.1 *Derivation of Measured Productivity in Aggregate*

Aggregate output provides a natural derivation of measured productivity $\mathcal{P}$ since it is hod-1 in capital and labour. In the long run $Y^* = n^* y^*$ and we can show that output Y^* is hod-1 in capital and labour. If we substitute in zero profit output per firm (17) long-run output is

$$Y^* = n^* \left[AF\left(\frac{K^*}{n^*}, \frac{L^*}{n^*}\right) \nu(1-\zeta) \right] \tag{31}$$

$$= {n^*}^{1-\nu} AF(K^*, L^*) \nu(1-\zeta) \tag{32}$$

substitute out $n^* = \frac{Y^*}{y^*} = \frac{Y^*}{\frac{\phi\nu(1-\zeta)}{1-\nu(1-\zeta)}}$

$$ {Y^*}^{\nu} = \left(\frac{1-\nu(1-\zeta)}{\phi\nu(1-\zeta)}\right)^{1-\nu} AF(K^*, L^*)\nu(1-\zeta) \tag{33}$$

$$Y^* = F(K^*, L^*)^{\frac{1}{\nu}} \underbrace{A^{\frac{1}{\nu}} \nu(1-\zeta) \left(\frac{1-\nu(1-\zeta)}{\phi}\right)^{\frac{1-\nu}{\nu}}}_{\mathcal{P}} \tag{34}$$

---

17 Long-run aggregate output is $hod - 1$ in its factors. Increasing the factors (K, L, n) in Y^* by λ gives $Y^*(\lambda K, \lambda Y, \lambda n) = (\lambda n)^{1-\nu} AF(\lambda K, \lambda L)\nu(1-\zeta) = Y^* = \lambda^{1-\nu} n^{1-\nu} \lambda^{\nu} AF(K, L)\nu(1-\zeta) = \lambda n^{1-\nu} AF(K, L)\nu(1-\zeta) = \lambda Y^*$

Since $\nu < 1$, free entry causes long-run increasing returns to scale between $F(K, L)$ and Y (i.e. Y is hod-$\frac{1}{\nu} > 1$ in F). Furthermore we have removed n so it is clear that Y is hod-1 in (K, L). This leads to a natural definition of measured productivity $\mathcal{P}$ which accounts for this scale effect (so that an economy is not considered more productive simply because it has more K, L resources)

$$\mathcal{P}^* = \frac{Y^*}{F(K^*, L^*)^{\frac{1}{\nu}}} = A^{\frac{1}{\nu}} \nu(1-\zeta) \left(\frac{1-\nu(1-\zeta)}{\phi} \right)^{\frac{1-\nu}{\nu}} \tag{35}$$

Measured productivity $\mathcal{P}$ is an efficiency parameter since it is a function of true TFP A. Incidentally constant returns to scale $\nu = 1$ imply *underlying productivity* (productivity at steady state) is $\mathcal{P}^* = \frac{Y}{F} = A(1-\zeta)$ which is equivalent to TFP if there is perfect competition $\zeta = 0$. Measured productivity is decreasing in ζ. Measured productivity $\mathcal{P}$ is increasing in technological development A, whereas technology does not affect output per firm y^* (17).

2.3.2 *General Equilibrium Effect of Entry on Aggregate Output*

The response of aggregate output to entry is ambiguous under imperfect competition with endogenous labour.

$$Y_n = \pi - \zeta \nu A \text{F} \left(\frac{K}{n}, \frac{L}{n} \right) + F_l L_n \tag{36}$$

Since firms operate with excess capacity, a firm entering the economy can decrease aggregate output. Assuming the positive labour effect $F_l L_n$ is small, this is what happens in zero-profit equilibrium. A firm enters the economy, pays a fixed cost and underutilizes this fixed cost. The extra output added by this new incumbent is more than offset by the fall in output among other firms. And the negative effect is stronger with closer to CRTS $\nu \to 1$. In the perfect competition case there is no negative effect as firms all produce at constant returns to scale so it is irrelevant how the production is split among firms. In this classic Ramsey case with no endogenous labour, and no imperfect competition, $Y_n = 0$ so the steady state number of firms is optimal in the sense that it maximises output.

2.3.3 *General Equilibrium*

General equilibrium determines prices, consumption and labour given the current capital stock and number of firms. The aggregate equation of motion for capital follows from the household resource constraint (2).

$$\begin{aligned}\dot{K} &= wL + rK + \Pi - C \\ &= n(1-\zeta)\nu AF(k,l) + n[AF(k,l)(1-(1-\zeta)\zeta) - \phi] \\ &- \gamma\frac{e^2}{2} - C \end{aligned} \tag{37}$$

$$\dot{K} = n\left[AF\left(\frac{K}{n}, \frac{L}{n}\right) - \phi\right] - C - \gamma\frac{e^2}{2} \tag{38}$$

This is equivalent to rearranging the standard definition of aggregate output $Y = C + I$. Output Y is split between consumption $C(t)$ and investment $I(t) = \dot{K} + Z(t)$.

Definition 3.

Competitive equilibrium is the 'equilibrium' paths of aggregate quantities and prices $\{C(t), L(t), K(t), n(t), e(t), w(t), r(t)\}_{t=0}^{\infty}$*, with prices strictly positive, such that* $\{C(t), L(t)\}_{t=0}^{\infty}$ *solve the household problem.* $\{K(t)\}_{t=0}^{\infty}$ *satisfies the law of motion for capital. Labour and capital* $\{L(t), K(t)\}_{t=0}^{\infty}$ *maximise firm profits given factor prices. The flow of entry causes the arbitrage condition on entry to hold (price of entry equals NPV of incumbent). State variables* $\{K(t), n(t)\}_{t=0}^{\infty}$ *satisfy transversality. Factor prices are set according to (13) and (14), which ensures goods and factor markets clear.*

2.4 DYNAMICAL SYSTEM OF ENDOGENOUS ENTRY ECONOMY

The model economy is a system of four ordinary differential equations (ODEs) in four variables (C, e, K, n): two equations dictate the law of motion of capital and number of firms and two equations restrict these evolutions so that consumption maximises utility and entry arbitrages profit opportunities. There is also a fifth condition (intratemporal Euler), but it is static, which determines labour. The system is sufficient to qualitatively analyse the economy, by which we mean that for a given parameter set $\Omega \in \mathbb{R}^{\mathbb{P}}$, we find the set of solutions that satisfy the system given an initial state on the manifold $m_0 := (K_0, n_0) \in \mathbb{M} \subset \mathbb{X}$, where $\mathbb{X} \in \mathbb{R}^4$ is the four dimensional space of the dynamical system, and $\mathbb{M} \in \mathbb{R}^2$ is the manifold: a subset defined by states (capital and number of firms) in the control problem.

At a point in time $t \in (0, \infty)$ our economy is entirely described by its level of capital and the number of firms. For now we do not specify initial conditions or functional forms that would give specific trajectories of capital and number of firms. The primitives of our boundary value problem are the state of the system defined on an open set $\mathbb{M} := K \times n \in \mathbb{R}^2$, the time $t \in \mathbb{I}$ defined on an open interval of $\mathbb{R}$, the parameterization defined on an open set $\Omega \in \mathbb{R}^{\mathbb{P}}$ and the nonlinear C^1 function $g : \mathbb{M} \times I \times \Omega \supseteq \mathbb{R}^{3+\mathbb{P}} \to \mathbb{R}^2$, or *"time evolution law"*, that maps a given state, time and parameterization into a new state. This rule determines the state of the system at each moment in time from its state at all previous times. For most of the analysis we work with a system where the state space includes C and e, although these variables can be defined by their *policy functions* which map $K \times n$ into $C \times e$, so the system equations $\dot{K}$ and $\dot{n}$ describe the evolution of the state variables along the stable manifold.

Definition 4 Nonlinear system.

The dynamical system is a pair $(\mathbb{X}, g)$, where $\mathbb{X} = (\mathbb{R}^4, \psi)$ is Euclidean space and metric. It defines at a point in time $t \in \mathbb{R}$ the state of the system $x(t) \in \mathbb{X} \subseteq \mathbb{R}^4$. The motion of the system is described by a C^1 vector valued transition map $g : \mathbb{R}^{5+\mathbb{P}} \supseteq \mathbb{X} \times \mathbb{R} \times \Omega \longrightarrow \mathbb{R}^4$

Throughout the system labour is implicitly defined by the static intratemporal condition $w = \frac{-u_L}{u_C}$ and the system is given by the following four ODEs.

$$\dot{K} = Y - \frac{\gamma}{2}e^2 - C, \qquad Y = n(F(k,l) - \phi) \tag{39}$$

$$\dot{n} = e \tag{40}$$

$$\dot{C} = \frac{C}{\sigma(C)}(r - \rho), \qquad \sigma(C) = -\frac{Cu_{CC}}{u_C} \tag{41}$$

$$\dot{e} = re - \frac{\pi}{\gamma}, \qquad \pi = AF(k,l)(1 - (1-\zeta)\nu) - \phi \tag{42}$$

where factor prices

$$r = (1-\zeta)AF_k, \qquad w = (1-\zeta)AF_l \tag{43}$$

The *system equations*, (39) and (40), explain how the state of the system evolves. The *optimization conditions*, (41) and (42), restrict the state evolutions. They impose that households maximise utility and potential entrants maximise profits. The economic reiteration is that the system equations determine how capital and number of firms evolve as the economy moves through time, and the optimization conditions ensure that capital and number of firms move so as to maximise consumers' utility and firms' utility.

Mathematically it is easier to view the system in terms of the four state variables of the dynamical system (C, e, K, n), remembering that labour is also in terms of these states $L(C, K, n)$ via the intratemporal condition $(1-\zeta)AF_l = \frac{-u_L}{u_C}$.

$$\dot{K} = n(AF(k,l) - \phi) - \frac{\gamma}{2}e^2 - C, \tag{44}$$

$$\dot{n} = e \tag{45}$$

$$\dot{C} = -\frac{u_C}{u_{CC}}((1-\zeta)AF_k - \rho), \tag{46}$$

$$\dot{e} = ((1-\zeta)AF_k)e - \frac{AF(k,l)(1-(1-\zeta)\nu) - \phi}{\gamma}, \tag{47}$$

2.4.1 *Steady State Behaviour*

In steady state capital, number of firms, consumption and entry are stationary $\dot{K} = \dot{n} = \dot{C} = \dot{e} = 0$[18].

$$\dot{K} = 0 \quad \Leftrightarrow \quad Y^*(C^*) = C^* + \frac{\gamma}{2}e^* \tag{48}$$

$$\dot{n} = 0 \quad \Leftrightarrow \quad 0 = e^* \tag{49}$$

$$\dot{C} = 0 \quad \Leftrightarrow \quad r^*(C^*, K^*, n^*) = \rho \tag{50}$$

$$\dot{e} = 0 \quad \Leftrightarrow \quad r^*(C^*, K^*, n^*)e^* = \frac{\pi^*(C^*, K^*, n^*)}{\gamma} \tag{51}$$

18 Ignore the trivial steady state that arises when the state vector is the zero vector.

Plug $e^* = 0$ into $\dot{K}$ and $\dot{e}$ and rewrite in terms of variables (C, K, L, n, e)

$$\dot{K}: \quad n^* \left[AF\left(\frac{K^*}{n^*}, \frac{L^*}{n^*}\right) - \phi \right] = C^* \tag{52}$$

$$\dot{n}: \quad e^* = 0 \tag{53}$$

$$\dot{C}: \quad (1-\zeta) AF_k \left(\frac{K^*}{n^*}, \frac{L^*}{n^*}\right) = \rho \tag{54}$$

$$\dot{e}: \quad F\left(\frac{K^*}{n^*}, \frac{L^*}{n^*}\right) = \frac{\phi}{A(1-(1-\zeta)\nu)} \tag{55}$$

with static intratemporal Euler

$$(1-\zeta) AF_l \left(\frac{K^*}{n^*}, \frac{L^*}{n^*}\right) = -\frac{u_L(L^*)}{u_C(C^*)} \tag{56}$$

The system determines $(C^*, L^*, K^*, n^*, e^*)$, where $e^* = 0$ is immediate, and the intratemporal Euler (56) defines $L(C, K, n)$. The results formally confirm that the zero-profit outcome we analyzed in section 2.2.2.2 is a steady state outcome of the system. The entry arbitrage condition, which becomes a zero profit condition in steady state, gives steady state technology $F(\frac{K}{n}, \frac{L}{n})$, which then gives y^*, and a mapping between C^* and n^* via (52). The system is recursive: first find K^*, n^* then the policy rules C^*, L^* as a function of the states. Projecting the optimization conditions (54) and (55) onto the state space determines K^*, n^* for a given L^*, which is fixed via the capital labour ratio. Ceteris paribus, say for the K, L, n at perfect competition steady state levels, (55) implies imperfect competition reduces production per firm and (56) and (54) imply marginal products rise with imperfect competition (because firms underproduce); (52) implies lower consumption. These do not necessarily hold when K, L, n adjust. The intuition is that firms produce less so they benefit from greater increasing returns to scale hence marginal products are higher.

2.4.1.1 *Steady State Graphical Intuition*

The optimization conditions (54) and (55) determine per firm capital k and labour l, and therefore aggregate capital-labour ratio $\frac{k}{l} = \frac{K}{L}$[19]. In k, l space (55) is an isoquant for zero-profit output, and (54) is a locus along which marginal product of capital is at steady state. Figure (6) shows the convex isoquant and linear marginal product of capital in steady state for the functional forms and parameterizations used in the numerical section. Later we shall comment on the unamibiguous decrease in k indicated by the dashed lines which is caused by a rise in imperfect competition. Labour on the other hand can increase or decrease. In the numerical example it increases as market power increases which refelcts that the income effect outweighs the substitution effect.

Figure (7) shows K, n combinations that cause the optimization nullclines ($\dot{C} = 0$ and $\dot{e} = 0$) to hold[20]. At the intersection both hold. The interpretation of the functions is that consumption does not change when interest rates equal the discount rate $r = \rho$, and entry does not change when profits are zero $\pi = 0$. Incentives cause both results: incentive to consume today is the

19 In other words, eliminate n in the two equations to give a single relationship between K and L.

20 Functions (54) and (55) respectively. The loci are equivalent to $\dot{C} = 0$ and $\dot{e} = 0$ only if other vairables are at steady state.

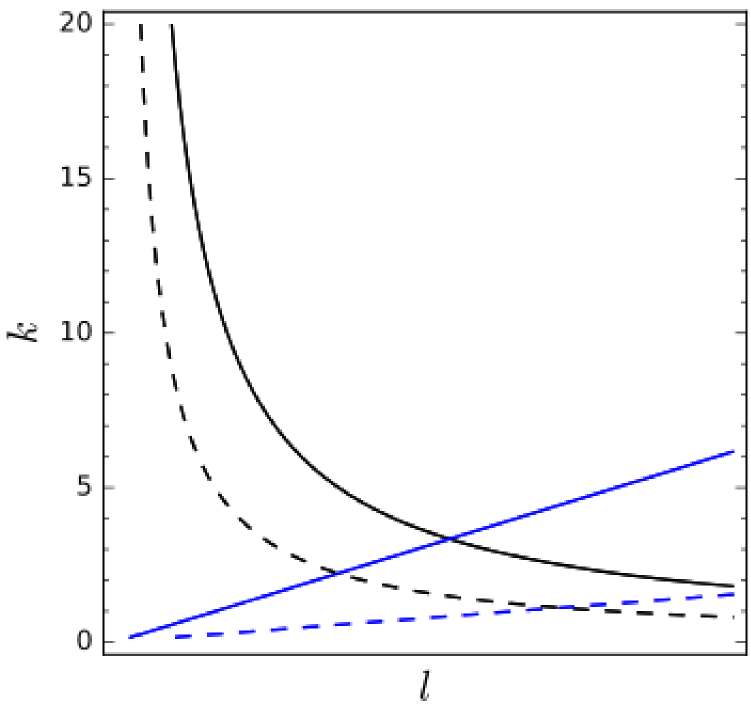


Figure 6.: Optimization Nullcines $\dot{C}$ and $\dot{e}$ (equations (54) and (55)) in k, l space

same as incentive to consume later because discount rate and interest rate are equal, so discount a household suffers from waiting to consume is offset by interest earned whilst waiting. There is no incentive for entry when profits are zero. The interest rate condition is upward sloping. A rise in capital per firm decreases the marginal product of capital[21], and therefore the interest rate, but a rise in firms decreases capital per firm back to its steady state level. The free entry condition ($\pi = 0$) slope is ambiguous when labour varies, and is upward sloping when labour is fixed. Increasing capital increases profits, but capital also raises labour which can reduce profits. If profit falls number of firms decrease until zero profit is restored. Figure (7) shows the case where an increase in capital raises profits and

21 since DRTS F_k is decreasing in its arguments.

number of firms increase to syphon the profit causing a positive slope.

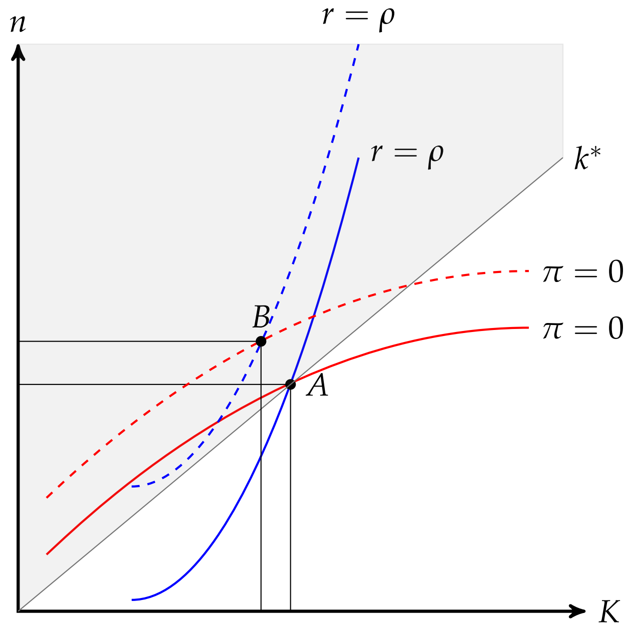


Figure 7.: $\pi = 0$ and $r = \rho$ in K, n space

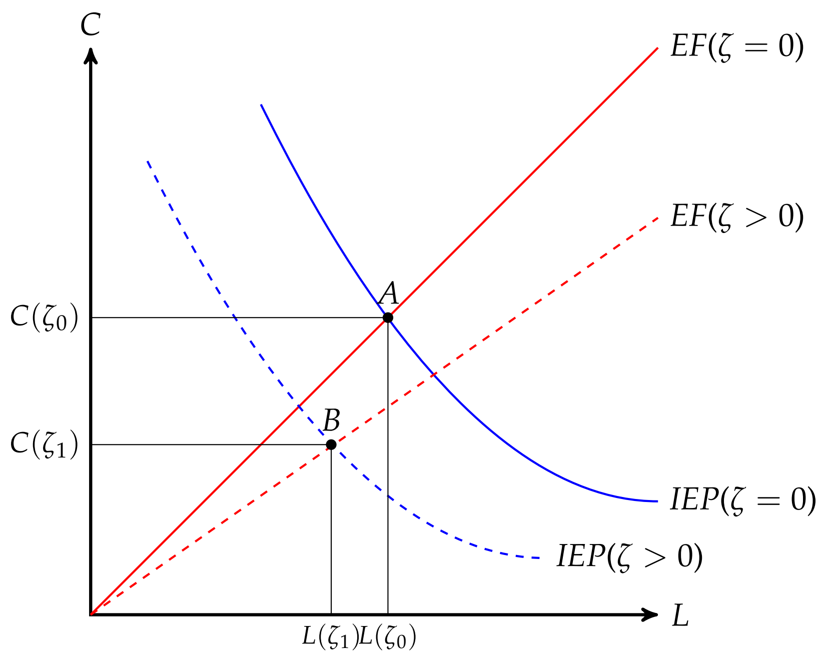


Figure 8.: IEP and EF in L, C space

We have shown that the optimization conditions determine capital and labour per firm (k^* and l^*), therefore we can determine the marginal product of labour, which gives the steady state wage w^*. Steady state wage is fixed so the intratemporal condition (56) describes L, C combinations that keep the ratio of marginal utilities constant. Thus it represents the *income expansion path* (IEP). We call the capital evolution equation (52) the Euler frontier when Y is replaced by zero profit output from (55), equivalent to (17), and number of firms is written in terms of L, so the equations can be plotted in L, C space. $n = \frac{L}{l}$ and l is determined, so when L changes it changes l but n moves to keep l at its steady state level. Then the IEP and EF are

$$\overbrace{(1-\zeta)AF_l\left(k^*, l^*\right)}^{w^*} = -\frac{u_L(L^*)}{u_C(C^*)} \qquad \text{IEP} \qquad (57)$$

$$C^* = \frac{L^*}{l^*}\left(\frac{\nu(1-\zeta)\phi}{(1-(1-\zeta)\nu)}\right) \qquad \text{EF} \qquad (58)$$

The IEP is downward sloping because labour is an inferior good and consumption is a normal good. As income increases consumption increases and labour decreases; specifically non-labour income since w^* is fixed. Income and utility expand North West on the IEP. The slope is convex because utility from consumption diminishes and disutility from labour grows. Under quasi-homothetic preferences the IEP is a straight line; if preferences are homomethic IEP intersects (1,0).

2.4.1.2 *Comparative statics: Imperfect competition*

Theorem 2.4.1 Steady State Capital and Labour per Firm.
In steady state market power decreases capital per firm, and ambiguously affects labour per firm.

$$k_\zeta < 0, \quad l_\zeta \lesseqgtr 0 \tag{59}$$

Proof. Appendix (B.4) □

The result is shown in figure (6) by a downward shift in steady state output isoquant and downward shift in steady state marginal product of capital. The result is also shown by the grey region of figure 7. In the grey region $\frac{K}{n}$ is strictly less, and one can see any shift in the curves caused by a rise in ζ will put the new intersection in the grey region, as the example with dotted lines shows–leading to new equilibrium B.

Imperfect competition reduces the interest rate so it is less than the discount rate. Interest rate will return to parity if the marginal product of capital increases which for given K occurs by increasing n–less capital per firm raises MPK as each firm employs capital more efficiently due to DRTS in variable cost. The $r = \rho$ curve shifts up. A rise in imperfect competition increases profits $\pi > 0$, so $\pi = 0$ shifts up too because given K a rise in n will reduce production per firm, and therefore profits, since each firm has less capital and uses it more efficiently (has less excess capacity lower monopoly profits), which will keep profits at zero. Under both steady state conditions imperfect competition must be offset by a rise in n for a given K, hence capital per firm, k, unambiguously falls. The grey region is $k < k^*$, and any rise in ζ will reduce k, to a point in this region[22].

A rise in imperfect competition shifts the IEP down and decreases the EF slope causing a decrease in consumption and an increase or decrease in labour supply, but a fall in consumption for any L. The IEP shifts down because market power ζ reduces wage, so for given L, according to the intratemporal condition, consumption must decrease to decrease the utility from consumption and equate the fall in wages[23]. It is a pure substitution effect: leisure $(1 - L)$ is cheaper relative to consumption, therefore the household decrease consumption and increase leisure (for a given level of consumption takes less labour). The EF slope is shallower because each unit of L increases income less, which creates an income effect. The household has less income so decreases intake of both normal goods: consumption and leisure. For a given L the household reduces C. Both substitution and income effects reduce consumption, but the substitution effect reduces labour whereas the income effect increases labour (decreases the normal good leisure). Figure 8 shows the case where substitution effect dominates the income effect–labour falls. Under fixed labour EF would still rotate because return on capital and wage decrease, but the IEP is vertical which removes substitution effect, leaving consumption reducing income effect.

---

22 The diagram is schematic; k must be in a region bound above both original curves, which rules out some of the grey, but consider different functions e.g. both are flatter so they tend to the k^* line–this opens up more of the grey region.

23 The ratio of marginal utilities is negative, so we want the ratio to increase which means a decrease when we consider the negation.

In summary, the steady state analysis gives a snapshot of the economy when capital, number of firms, consumption and entry are constant. Entry is zero because profits are arbitraged to zero in the long run. The effect of a rise in imperfect competition is lower output per firm, lower consumption, higher marginal products, and ambiguous change in capital, number of firms and labour. However, the ratios capital per firm and consumption to labour are lower. A decrease in capital per firm is important for our thesis because in steady state firms have excess capacity and employ too few inputs. This result asserts that more imperfect competition will worsen excess capacity.

2.4.2 *Parameterization and Numerical Exercise*

In this section I specify functional forms and parameterizations which allow us to view the bifurcations mentioned above. Further I shall show that the system is a saddle point.

The baseline RBC model assumes isoelastic (constant elasticity) separable subutilties and a Cobb-Douglas production function.

2.4.2.1 *Isoelastic Utility*

$$U(C,L) = \frac{C^{1-\sigma}-1}{1-\sigma} - \xi\frac{L^{1+\eta}}{1+\eta} \tag{60}$$

The derivatives are

$$U_C = C^{-\sigma}, \quad U_{CC} = -\sigma C^{-\sigma-1}, \quad U_L = -\xi L^{\eta} \tag{61}$$

The degree of relative risk aversion is constant $\sigma(C) = -C\frac{U_{CC}}{U_C} = \sigma$. Isoelastic utility implies there is constant elasticity of utility with respect to each good[24]. Conceptually, $\sigma \in (0, \infty) \setminus \{1\}$ is the constant coefficient of relative risk aversion[25]. $\sigma \to \infty$ implies infinite risk aversion, so consumption has little effect on utility, no effect in the infinite limit. $\sigma \to 0$ is risk neutrality so that a % change in consumption has the same % change on utility. The limiting case of $\sigma \to 1$ implies log utility $\ln(C)$; we shall assume this for the numerical exercise–it will lead to fixed labour as income and substitution effects cancel out. $\frac{1}{\eta}$ is Frisch elasticity of labour supply. Inverse elasticity $\eta \in (0, \infty)$ of $\eta = 0$ implies indivisible labour; Mertens and Ravn 2011 estimate it as $\eta = 0.976$.

2.4.2.2 *Cobb-Douglas Production*

$$F(k, l) = k^{\alpha} l^{\beta} = K^{\alpha} L^{\beta} n^{-(\alpha+\beta)} = F(K, L) n^{-(\alpha+\beta)} \tag{62}$$

so derivatives are

$$F_k = \alpha k^{\alpha-1} l^{\beta} = \alpha K^{\alpha-1} L^{\beta} n^{1-(\alpha+\beta)}, \tag{63}$$

$$F_l = k^{\alpha} \beta l^{\beta-1} = K^{\alpha} \beta L^{\beta-1} n^{1-(\alpha+\beta)} \tag{64}$$

24 Mertens and Ravn 2011 is a useful example usage of isoelastic utility with detailed parameter descriptions and accompanying empirical estimates.
25 Although precisely it is only a curvature parameter in the absence of risk.

Cobb-Douglas production conforms to our assumptions on the production function derivatives, and it is homogeneous of degree $\alpha + \beta$, so $\nu = \alpha + \beta$ in our general notation. α and β are capital and labour shares respectively. This implies increasing marginal costs if $\alpha + \beta < 1$.

Table 1 summarizes the parameter values used for simulation exercises. They are benchmark for this simple model, and

Market Power	ζ	$(0, 1)$
Capital Share	α	0.3
Labour Share	β	0.5
Fixed Cost	ϕ	0.3
Entry Cost	γ	3.0
Technology	A	1.0
Risk Aversion	σ	1.0
Discount Rate	ρ	0.025
Labour Weight	ξ	0.01
Labour Elast. (Frisch)	η	0.5

Table 1.: Parameter Values for Numerical Exercises

are replicated from Brito and Dixon 2013. For comparison, Jaimovich 2007 and Jaimovich and Floetotto 2008 calibrate a similar model and parameters taking a 'formal calibration' approach. In Jaimovich 2007 the markup (price over marginal cost definition) is switched between $\mu = 1.05$ and $\mu = 1.10$, so in terms of Lerner Index ($\zeta = 1 - \frac{1}{\mu}$) then $\zeta = 0.047$ and $\zeta = 0.09$. In Jaimovich and Floetotto 2008 it is $\mu = 1.3$ so $\zeta = 0.231$. Jaimovich and Floetotto 2008 calibrate the fixed cost ϕ to be a percentage of sales $\frac{n\phi}{Y} = \frac{\phi}{y} = 0.127$, and use slightly larger proportion of 15% in the appendix. In this paper we assume no intersector substitutability $\theta_I = 1$ (different products across industries i.e. no substitution across industry) and infinite intrasector substitutability $\theta_F = \infty$ (homogeneous goods within industry i.e. perfect substitution within industry). Whereas Jaimovich and Floetotto 2008 emphasize keeping these parameters free, but it gives similar outcomes to our assumptions $\theta_I = 1.001$ and $\theta_F = 19.6$[26]. They assume indivisible labour $\eta = 0$.

Under this parameterization the dynamical system (39)-(42) becomes

$$\dot{C} = \frac{C}{\sigma}\left[(1-\zeta)A\alpha K^{\alpha-1}L^{\beta}n^{1-(\alpha+\beta)} - \rho\right] \tag{65}$$

$$\dot{e} = (1-\zeta)A\alpha K^{\alpha-1}L^{\beta}n^{1-(\alpha+\beta)}e - \frac{1}{\gamma}\left(AK^{\alpha}L^{\beta}n^{-(\alpha+\beta)}(1-(1-\zeta)\nu) - \phi\right) \tag{66}$$

$$\dot{K} = n\left[AK^{\alpha}L^{\beta}n^{-(\alpha+\beta)} - \phi\right] - \frac{\gamma}{2}e^2 - C \tag{67}$$

$$\dot{n} = e \tag{68}$$

where is L is defined in terms of (C, K, n) through the intratemporal condition (4)

$$L = \left(\frac{(1-\zeta)AK^{\alpha}\beta n^{1-(\alpha+\beta)}}{\xi C^{\sigma}}\right)^{\frac{1}{1+\eta-\beta}} \tag{69}$$

26 Specifically they use ω and τ which since their aggregators are written as p-norms and Holder conjugates, rather than elasticities as is common in economics $\theta_I = \frac{1}{1-\omega} = \frac{1}{1-0.001} = 1.001$ and $\theta_F = \frac{1}{1-\tau} = \frac{1}{1-0.949} = 19.6$.

and from which optimal labour behaviour follows (as shown theoretically in theorem 2.2.1).

$$L_n = \frac{1-(\alpha+\beta)}{1+\eta-\beta}\frac{L}{n} \tag{70}$$

$$L_K = \frac{\alpha}{(1+\eta-\beta)}\frac{L}{K} \tag{71}$$

$$L_C = \frac{\xi\sigma}{(1-\zeta)A\beta(\beta-1-\eta)}\frac{C^{\sigma-1}}{K^{\alpha}L^{\beta-2-\eta}n^{1-(\alpha+\beta)}} \tag{72}$$

$$= -\frac{\sigma}{(1+\eta-\beta)}\frac{L}{C} \tag{73}$$

$$L_\zeta = -\frac{1}{(1-\zeta)(1+\eta-\beta)}L \tag{74}$$

$$L_A = \frac{1}{A(1+\eta-\beta)}L \tag{75}$$

2.4.2.3 *Parameterized Steady State*

To derive steady state outcomes under CES utility and Cobb-Douglas production it is easier to use per firm variables k and l. Thus, the intratemporal condition is

$$C^* = \left(\frac{\beta(1-\zeta)Ak^{*\alpha}l^{*\beta-1-\eta}n^{*\eta}}{\xi}\right)^{\frac{1}{\sigma}} \tag{76}$$

the dynamical system is

$$C^* = \frac{\phi(1-\zeta)\nu}{1-(1-\zeta)\nu}n^* \tag{77}$$

$$e^* = 0 \tag{78}$$

$$\alpha k^{*\alpha-1}l(k^*,n^*)^{\beta} = \frac{\rho}{A(1-\zeta)} \tag{79}$$

$$k^{*\alpha}l(k^*,n^*)^{\beta} = \frac{\phi}{A(1-(1-\zeta))\nu)} \tag{80}$$

Therefore substituting $C^*(n^*)$ into the intratemporal condition, where $\nu = \alpha + \beta < 1$, gives

$$l^*(k^*, n^*) = \left(\frac{(1-\zeta)^{1-\sigma} A k^{*\alpha} \beta n^{*-\sigma-\eta}}{\xi \left(\frac{\phi\nu}{1-(1-\zeta)\nu} \right)^{\sigma}} \right)^{\frac{1}{1+\eta-\beta}} \tag{81}$$

Solving (79) and (80), as plotted in figure 6, gives k^* and l^*

$$k^* = \frac{\phi\alpha(1-\zeta)}{(1-(1-\zeta)\nu)\rho} \tag{82}$$

$$l^* = \left[\frac{1}{A} \left(\frac{\rho}{\alpha(1-\zeta)} \right)^{\alpha} \left(\frac{\phi}{1-(1-\zeta)\nu} \right)^{1-\alpha} \right]^{\frac{1}{\beta}} \tag{83}$$

These results capture that capital per firm is decreasing in market power, whilst labour per firm is ambiguous. This verifies the general results of theorem 2.4.1. Furthermore it emphasizes that fixed cost raises both inputs as does returns to scale both raise the output level that minimizes costs.

Rearranging the intratemporal condition (81) gives[27]

$$n^* = \left(\left(\frac{\alpha}{\rho\nu} \right)^{\sigma} \frac{(1-\zeta)A\beta}{\xi k^{*\sigma-\alpha} l^{*1+\eta-\beta}} \right)^{\frac{1}{\sigma+\eta}} \tag{84}$$

where k^* and l^* are in terms of model parameters defined in (82) and (83). Since k^* and l^* are increasing in fixed cost ϕ, then the number of firms is decreasing in fixed cost, which the next equation shows. By substituting in k^* and l^*, simplifying gives

27 Intermediate step from rearranging intratemporal condition is $n^* = \left(\frac{l^{*1+\eta-\beta} \xi \left(\frac{\phi\nu}{(1-(1-\zeta)\nu)} \right)^{\sigma}}{k^{*\alpha}(1-\zeta)^{1-\sigma} A\beta} \right)^{-\frac{1}{\sigma+\eta}}$ and by noting $\left(\frac{\phi\nu}{(1-(1-\zeta)\nu)} \right)^{\sigma} = \left(\frac{\nu\rho k^*}{\alpha(1-\zeta)} \right)^{\sigma}$ simplifies to above.

an expression for the steady state number of firms entirely in exogenous parameters[28]

$$n^* = \left[\frac{\beta}{\xi \nu^\sigma} \left\{ \left(A \left(\frac{\alpha}{\rho} \right)^\alpha \right)^{1+\eta} (1-\zeta)^{\alpha(1+\eta)+\beta(1-\sigma)} \left(\frac{1-(1-\zeta)\nu}{\phi} \right)^{1-\nu+\eta(1-\alpha)+\sigma\beta} \right\}^{\frac{1}{\beta}} \right]^{\frac{1}{\eta+\sigma}} \tag{85}$$

where $\nu = \alpha + \beta < 1$. Whence we have an exact expression for the number of firms in steady state, which is decreasing in the fixed cost ϕ[29]. Given n^* we can now determine all the steady

28 First rewrite as the following, then rearrange

$$n^* = \left[\frac{\beta}{\xi \nu^\sigma} \left(A \left(\frac{1-(1-\zeta)\nu}{\phi} \right)^{1-\alpha} \left(\frac{\alpha(1-\zeta)}{\rho} \right)^\alpha \right)^{\frac{1+\eta}{\beta}} \left(\frac{\phi(1-\zeta)}{(1-(1-\zeta)\nu)} \right)^{1-\sigma} \right]^{\frac{1}{\eta+\sigma}}$$

$$n^* = \left[\frac{\beta}{\xi \nu^\sigma} \left\{ A \left(\frac{1-(1-\zeta)\nu}{\phi(1-\zeta)} \right)^{\frac{1-\nu+\eta(1-\alpha)+\beta\sigma}{1+\eta}} (1-\zeta) \left(\frac{\alpha}{\rho} \right)^\alpha \right\}^{\frac{1+\eta}{\beta}} \right]^{\frac{1}{\eta+\sigma}}$$

29 Later we shall discuss the result that market power ζ enters the expression twice, in one case it increases number of firms and in the other case it decreases number of firms. In the numerical exercise we shall see that this means raising market power can increase and eventually decrease number of firms.

state variables (C^*, e^*, K^*, L^*) which we previously derived in per firm terms.

$$C^* = n^* \frac{\phi(1-\zeta)\nu}{1-(1-\zeta)\nu} \tag{86}$$

$$e^* = 0 \tag{87}$$

$$K^* = n^* \frac{\phi\alpha(1-\zeta)}{(1-(1-\zeta)\nu)\rho} \tag{88}$$

$$L^* = n^* \left[\frac{1}{A} \left(\frac{\rho}{\alpha(1-\zeta)} \right)^{\alpha} \left(\frac{\phi}{1-(1-\zeta)\nu} \right)^{1-\alpha} \right]^{\frac{1}{\beta}} \tag{89}$$

2.4.2.4 *How steady state changes with market power?*

This section uses numerical analysis to reinforce the general graphical description of steady state behaviour in section 2.4.1.1. Figure 9 shows a 2D projection of steady states as market power ζ changes with the values of ζ marked next to the points[30]. This corresponds to the theoretical figure 7 discussed earlier. Figure 10 makes the equivalence clear by showing a panel of two plots that depict how these steady states are determined from the intersection of $\pi = 0$ (curved line) and $r = \rho$ (straight line). Furthermore panel 2 plots two projections of the loci to show their direction of motion from the zero market power lower curves to the higher market power ($\zeta = 0.65$) upper curves. Incidentally this is approximately the market power that maximises number of firms[31]. That is, the graphs show that by raising market power ζ from no market power to dominance $(0 \to 1)$, capital is always decreasing in ζ and number of firms increases

30 Figure 18 in Appendix B.4.1 offers an alternative perspective with ζ on the z-axis.

31 More accurate numerical optimization shows that number of firms is maximised when market power is $\zeta = 0.625$.

then decreases after $\zeta = 0.65$. Therefore after market power is sufficiently high n^* begins decreasing. This is a feature of the parametrization, rather than general, as we noted in the schematic approach figure 7 by commenting that steady state is possible in any of the grey region. This can be understood through the analytic n^* derivation (85) which shows that there is one component

$$\left(\frac{1-(1-\zeta)\nu}{\phi}\right)^{1-\nu+\eta(1-\alpha)+\sigma\beta} = \left(\frac{1}{AF(k^*, l^*)}\right)^{1-\nu+\eta(1-\alpha)+\sigma\beta}$$

that increases n^* as ζ increases, and a secondary component

$$(1-\zeta)^{\alpha(1+\eta)+\beta(1-\sigma)}$$

that decreases n^* as ζ increases (given positive power). The first component is associated with the fixed cost and production per firm. It raises number of firms as market power rises because production per firm falls as higher markups allow firms to cover fixed cost by producing less, this creates room for more incumbents. However, for this calibration, as market power gets large this positive effect is dominated by the secondary effect which arises because of the negative effect of ζ on labour per firm[32] (see equation (83)). Figure 10 makes the relationship to the nullclines clear. The nonlinearity in $\pi = 0$ is responsible

32 This negative effect of ζ on n^* relies on a positive power. However, risk aversion σ is negative in the power which allows for ambiguity. A judicious choice of risk aversion $\sigma = 1 + \frac{\alpha}{\beta}(1+\eta)$ would switch off this effect by making the power zero. Or indeed a σ exceeding this value would make the power negative and thus the effect of ζ on number of firms positive, thus reinforcing the primary effect. More market power would raise number of firms. No ambiguity. Incidentally given the parameter values in table 1 this value of σ would be $\sigma = 1.9$, not incomprehensible.

for relatively fewer, and eventually decreasing number of firms, as market power rises. At low levels of capital, small changes in capital cause large changes in number of firms. A higher σ risk aversion would flatten the profit nullcline (to maintain zero profits a rise in n requires a larger rise in K) so that number of firms continues to increases as the curve shifts upwards.

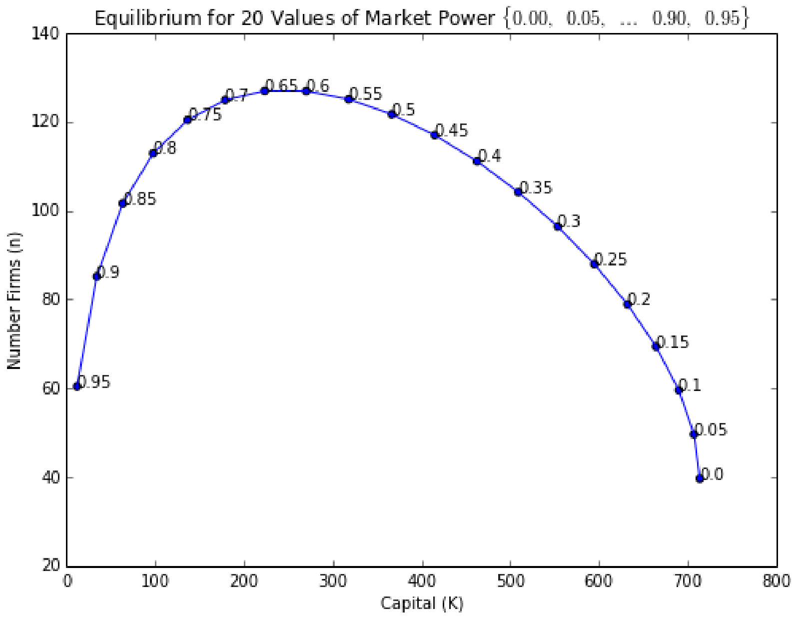


Figure 9.: Capital and No. Firms Equilibrium as Market Power Changes

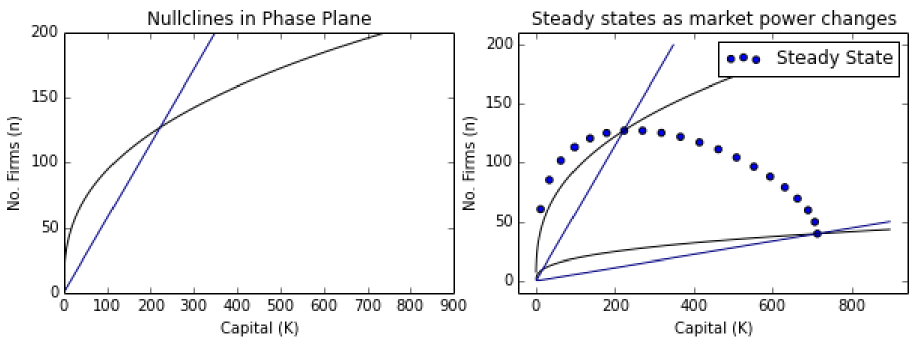


Figure 10.: Relationship between Intersecting Nullclines and Steady State as Market Power Changes

Figure 11 plots the household nonlinear intra- and linear intertemporal conditions which we call the Income Expansion Path (IEP) and Euler Frontier (EF) in steady state as discussed earlier and graphed in figure 8. Since we have set unitary risk aversion $\sigma = 1$ then there is log utility in consumption, and this causes income and substitution effects to cancel out meaning the rise in imperfect competition (solid lines to dashed lines) has no effect on labour. Clearly consumption falls unambiguously.

$$\text{IEP} \qquad C^* = \left(\frac{\beta(1-\zeta)Ak^{*\alpha}l^{*\beta-1}}{\xi L^{*\eta}} \right)^{\frac{1}{\sigma}} \tag{90}$$

$$\text{EF} \qquad C^* = \frac{\phi(1-\zeta)\nu}{1-(1-\zeta)\nu} \frac{L^*}{l^*} \tag{91}$$

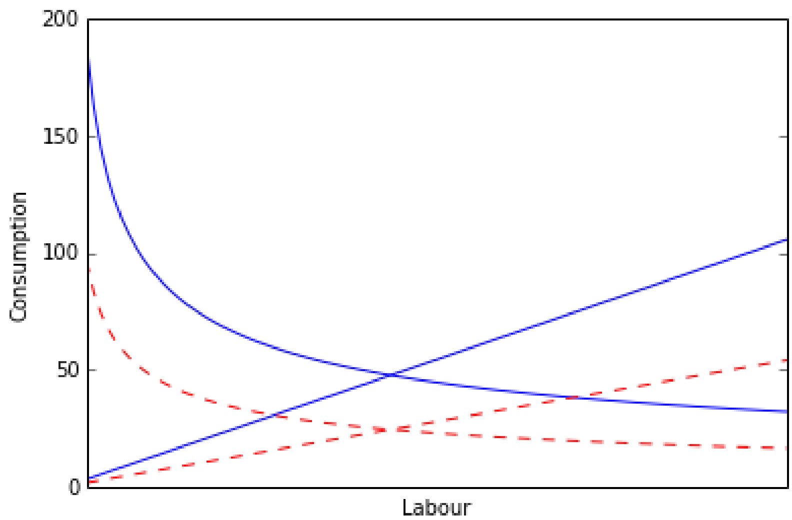


Figure 11.: Determining Household Consumption-Labour Choice

2.4.3 *Local Comparative Dynamics and Solution*

Comparative dynamics explain how the economy transitions to excess capacity steady state. Our interest is how imperfect competition affects this transition, specifically through its effect on firm capacity utilization. To understand this we solve the four dimensional system which gives trajectories of the variables over t.

Our system is of nonlinear form $\dot{x} = g(x)$, as defined in definition 4, but we linearize it to $\dot{x} = \mathbf{J}(x - x^*)$, and analyse the $\mathbf{J}$ operator which is a matrix on the space defined. $\mathbf{J}$ is the Jacobian matrix where each element is a respective derivative. The derivatives treat $\{K, n, C, e\}$ as independent, and labour is a function of these variables through the intratemporal condition $L(K, n, C)$. We solve the system recursively by first deriving C and e as functions on K, n, then plugging these policy rules back in to get K, n as functions of time and initial conditions.

Definition 5 Jacobian Matrix and the Jacobian (determinant). *The Jacobian matrix* $\mathbf{J} : \mathbb{R}^4 \to \mathbb{R}^4$ *is the nonlinear system matrix evaluated at steady state* x^*.

$$\begin{bmatrix} \dot{C} \\ \dot{e} \\ \dot{K} \\ \dot{n} \end{bmatrix} = \overbrace{\begin{bmatrix} \frac{C^*}{\sigma} r_C^* & 0 & \frac{C^*}{\sigma} r_K^* & \frac{C^*}{\sigma} r_n^* \\ -\frac{\pi_C^*}{\gamma} & \rho & -\frac{\pi_K^*}{\gamma} & -\frac{\pi_n^*}{\gamma} \\ Y_C^* - 1 & 0 & Y_K^* & Y_n^* \\ 0 & 1 & 0 & 0 \end{bmatrix}}^{\mathbf{J}} \begin{bmatrix} C - C^* \\ e - e^* \\ K - K^* \\ n - n^* \end{bmatrix} \tag{92}$$

In the linearized model, the state vector is deviation from equilibrium point[33]. Labour behaves according to the optimal intratemporal condition $L(K, n, C)$. Superficially, the elements of the Jacobian appear the same as the perfect competition case in Brito and Dixon 2013, for example element 3×3 is Y_n^* in both cases. But their magnitudes change for any given point due to the mark-up ζ, and as the steady state is also different they are evaluated at different points.

A more granular version of the Jacobian expressed in K^*, n^* terms is

$$\begin{bmatrix} \frac{C^*}{n^*\sigma}(1-\zeta)AF_{kl}L_C & 0 & \frac{C^*}{n^*\sigma}(1-\zeta)A(F_{kk}+F_{kl}L_K) & \frac{C^*}{n^*\sigma}(1-\zeta)\left[(1-\nu)\frac{\rho}{1-\zeta}+AF_{kl}L_n\right] \\ -\frac{(1-(1-\zeta)\nu)}{\gamma n^*}AF_lL_C & \rho & -\frac{(1-(1-\zeta)\nu)}{\gamma n^*}\left(\frac{\rho}{1-\zeta}+AF_lL_K\right) & \frac{1}{\gamma n^*}(\nu\phi-(1-(1-\zeta)\nu)AF_lL_n) \\ AF_lL_C-1 & 0 & \frac{\rho}{1-\zeta}+AF_lL_K & \frac{-\zeta\nu\phi}{1-(1-\zeta)\nu}+AF_lL_n \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{93}$$

which can be simplified even further with specification of functional forms (65 - 68) and using $F_l, F_{kl}, F_{kk}, L_C, L_K, L_n$

$$\begin{bmatrix} -\frac{\rho\beta}{1+\eta-\beta} & 0 & \frac{\rho^2\nu((1+\eta)(\alpha-1)+\beta)}{\sigma\alpha(1+\eta-\beta)} & \frac{\phi(1-\zeta)\nu\rho(1-\nu)(1+\eta)}{(1-(1-\zeta)\nu)\sigma(1+\eta-\beta)} \\ \frac{(1-(1-\zeta)\nu)\beta\sigma}{\gamma n^*(1+\eta-\beta)(1-\zeta)\nu} & \rho & \frac{-(1-(1-\zeta)\nu)\rho(1+\eta)}{\gamma n^*(1-\zeta)(1+\eta-\beta)} & \frac{\phi(\nu(1+\eta)-\beta)}{\gamma n^*(1+\eta-\beta)} \\ \frac{-\beta\sigma}{(1+\eta-\beta)(1-\zeta)\nu}-1 & 0 & \frac{\rho(1+\eta)}{(1-\zeta)(1+\eta-\beta)} & \frac{\phi[-\zeta\nu(1+\eta-\beta)+\beta(1-\nu)]}{(1-(1-\zeta)\nu)(1+\eta-\beta)} \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{94}$$

33 In the terminology of the dynamical system all undetermined variables are state variables, whereas in the terminology of the optimization problem K, n are states and C, e are controls. Actually the dynamical system is only 2-dimensional, K, n is a basis for the whole system. C, e will be shown to be functions of K, n

In the second row I have not substituted out the n^* component which is in terms of model parameters because it does not simplify nicely. Appendix B.6 offers a detailed derivation of the parameterized Jacobian[34].

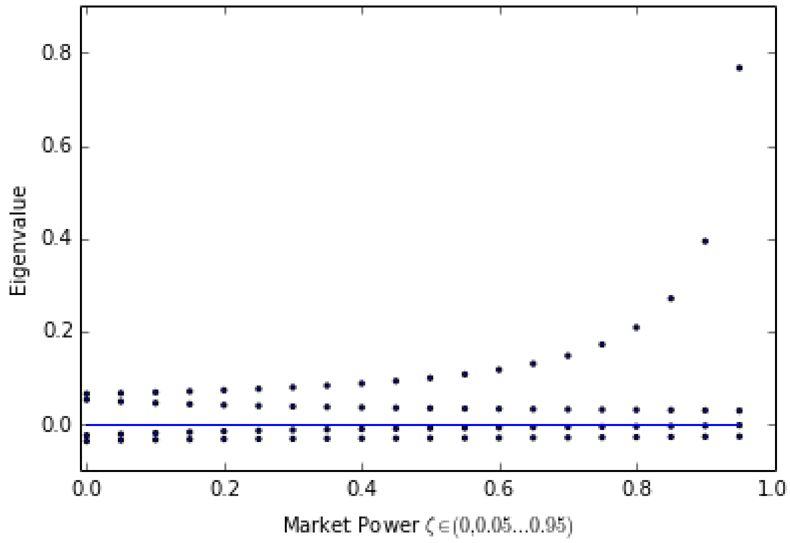


Figure 12.: Sets of 4 Eigenvalues for 20 Values of Market Power

Figure 12 shows that the Jacobian evaluated at steady state is a saddle point (2 positive eigenvalues, 2 negative eigenvalues) for different values of market power ζ. That is, the plot shows 20 values of market power $\zeta \in (0, 0.05...0.95)$ on the x-axis and the corresponding 4 eigenvalues at each of these levels of market power. There are always two unstable (positive eigenvalues) and two stable (negative) eigenvalues. This will allow us to show that the model is able to generate endogenous productivity movements which the next section explains. The numerical illustration is useful for the case of larger market

34 One notable feature is that element $J_{3,4} = Y_n^*$ is 0, so output is maximised with respect to number of firms, when $\zeta = \frac{\beta(1-\nu)}{\nu(1+\eta-\beta)}$. In our numerical example this implies a Lerner Index $\zeta = 0.125$ or a markup $\mu = 1.14$ will maximise aggregate output.

power because the theoretical eigenvalue approximation Brito and Dixon 2013 Dockner and Feichtinger 1991 is less accurate with less symmetry.

2.4.3.1 *Solving the Model*

As shown in figure 12 and supported theoretically by Brito and Dixon 2013[35], the four dimensional system has two positive and two negative eigenvalues, so the system is a saddle which is locally asymptotically unstable. Given the system is a saddlepath, then by the stable manifold theorem, we show saddlepath stability by setting the constant of integration on the two explosive eigenvalues to zero, thus reducing attention to the stable set. This gives a system of saddle-path conditions that may be solved for C and e. These conditions (*policy functions*) for C and e describe the stable manifold $\mathbb{M} \in \mathbb{R}^2$ of the system that ensures the constants on the explosive eigenvalues are zero, and thus solutions asymptotically reach steady state.

The solution in *open-loop* form which is in terms of initial values (K_0, n_0), t, and the parameters that comprise the eigenvectors $P^i_{j,k}$ and eigenvalues λ_j in steady state is as follows [36]:

35 A synopsis of their results is in appendix B.7.
36 Detailed derivations in appendix B.8

$$\begin{bmatrix} c(t) \\ e(t) \\ K(t) \\ n(t) \end{bmatrix} = \begin{bmatrix} c^* \\ e^* \\ K^* \\ n^* \end{bmatrix} +$$
$$\frac{1}{P^s_{1,3}P^s_{2,4} - P^s_{2,3}P^s_{1,4}} \begin{bmatrix} P^s_{1,1}P^s_{2,4}e^{\lambda_1 t} - P^s_{2,1}P^s_{1,4}e^{\lambda_2 t} & -P^s_{1,1}P^s_{2,3}e^{\lambda_1 t} + P^s_{2,1}P^s_{1,3}e^{\lambda_2 t} \\ P^s_{1,2}P^s_{2,4}e^{\lambda_1 t} - P^s_{2,2}P^s_{1,4}e^{\lambda_2 t} & -P^s_{1,2}P^s_{2,3}e^{\lambda_1 t} + P^s_{2,2}P^s_{1,3}e^{\lambda_2 t} \\ P^s_{1,3}P^s_{2,4}e^{\lambda_1 t} - P^s_{2,3}P^s_{1,4}e^{\lambda_2 t} & -P^s_{1,3}P^s_{2,3}e^{\lambda_1 t} + P^s_{2,3}P^s_{1,3}e^{\lambda_2 t} \\ P^s_{1,4}P^s_{2,4}e^{\lambda_1 t} - P^s_{2,4}P^s_{1,4}e^{\lambda_2 t} & -P^s_{1,4}P^s_{2,3}e^{\lambda_1 t} + P^s_{2,4}P^s_{1,3}e^{\lambda_2 t} \end{bmatrix} \begin{bmatrix} K(0) - K^* \\ n(0) - n^* \end{bmatrix} \tag{95}$$

Notably on impact $t = 0$ elements $(2,3)$ and $(1,4)$ [37]are 0, and elements $(1,3)$ and $(2,4)$ are 1. In other words states do not move on impact, jump variables do, which is vital for the endogenous productivity dynamics derived in the next section.

2.5 MAIN RESULT: CAPACITY UTILIZATION

We use the theoretical results of section 2.4.3.1 to prove the endogenous productivity dynamics of the system. As the theory shows in the short run, the state variables, capital and number of firms are predetermined, so they do not adjust immediately when there is a shock. This can be seen from the solution to the system, since at $t = 0$ the state variables equal their initial condition x_0, whereas the controls change depending on the eigenvectors. Labour is one of these free variables that will jump instantaneously. The timing difference between immediate labour adjustment and slow capital and firm adjustment exacerbate the capacity utilization effect which causes measured productivity to exceed underlying productivity. Underlying

37 Which may first be simplified to $P^s_{1,3}P^s_{2,3}(e^{\lambda_2 t} - e^{\lambda_1 t})$ and $P^s_{1,4}P^s_{2,4}(e^{\lambda_1 t} - e^{\lambda_2 t})$

productivity is the steady state value $\mathcal{P}^*$ defined in (35); it depends on $\{A, \nu, \phi, \zeta\}$ and is independent of model variables. Measured productivity $\mathcal{P}$ is productivity observed at any instance.

In steady state firms have excess capacity. Since entry is a slow process, a positive production shock prompts incumbent firms to increase their capacity toward full capacity, making better use of their fixed costs, which improves productivity. A negative production shock causes firms to decrease their capacity, further from full capacity, making less use of their fixed costs and decreasing productivity. In the long run, firms will exit until zero profit returns each firm to producing a long-run level of output which is unchanged. Ignoring labour, with better technology each firm will require less capital to produce the long-run level. Hence capital per firm should fall.

Theorem 2.5.1 Measured Productivity Overshooting.
A change in technology causes a greater change in measured productivity than underlying productivity. There is overshooting.

$$\left|\frac{d\mathcal{P}(0)}{dA}\right| > \left|\frac{d\mathcal{P}^*}{dA}\right|$$

Proof. (Full derivation Appendix B.9)

$$\frac{d\mathcal{P}(0)}{dA} - \frac{d\mathcal{P}^*}{dA} = \left(\frac{\zeta}{1-\zeta}\right)\left[\frac{d\mathcal{P}^*}{dA} + \frac{\mathcal{P}^* F_L L_A}{\nu F}\right] \tag{96}$$

$$= \zeta \left(\frac{\phi}{A(1-(1-\zeta)\nu)}\right)^{1-\frac{1}{\nu}} \left(1 + \frac{A F_L L_A}{F}\right), \quad \nu < 1 \tag{97}$$

□

The degree of measured productivity overshooting depends on the level of industry competition ζ, the fixed cost of production ϕ, and the response of endogenous labour L_A. Marginal cost returns to scale ν are less important as the proposition still holds with constant returns to scale $\nu \to 1$. But notably, the result is undefined if there are constant returns and perfect competition $\zeta = 0$ since equilibrium cannot exist in the presence of a fixed cost, returns are globally decreasing.

Corollary 2 Nature of Productivity Overshooting.
Imperfect competition, fixed costs in production and labour response all reinforce the overshooting effect.

- *Imperfect competition* $\zeta > 0$ *is necessary for measured productivity overshooting, and overshooting is increasing in* ζ. *As proof consider that if* $\zeta = 0$ *then* $\frac{d\mathcal{P}(0)}{dA} = \frac{d\mathcal{P}^*}{dA}$.

- *Fixed cost* $\phi > 0$ *necessary for productivity overshooting because it endogenously creates locally increasing returns to scale that incumbent firms exploit on impact.*

- *Positive elasticity of labour term* $\frac{AF_L L_A}{F}$ *caused by endogenous labour strengthens overshooting. This arises because labour jumps immediately on impact, unlike the slow response of states.*

A positive once-and-for-all technology shock immediately raises output per firm. Firms do not have time to adapt, incumbents absorb the new technology and produce more output given their existing capital. Producing more reduces their excess capacity. Profits increase (π is increasing in A). Over time capital and number of firms adjust, returning production to y^* and profits to zero. Capital per firm will fall because better technology reduces the inputs needed to produce the fixed long-run level y^*. Endogenous labour strengthens the impact response, and causes more overshooting.

2.6 chapter summary

This chapter shows that non-instantaneous entry and exit of firms to macroeconomic conditions causes endogenous productivity dynamics over the business cycle. The result is that productivity shocks are amplified in the short run relative to their long-run level.

We propose the theory through an endogenous firm entry dynamic general equilibrium model with imperfect competition and endogenous entry costs. Imperfect competition creates monopoly profits which cause "too many" entrants in steady state. Each incumbent has excess capacity defined by the amount it underproduces relative to a cost minimizing level of output. Incumbents' excess capacity varies in the short-run because firms cannot enter/exit so incumbents vary production in response to economic shocks. Through increasing returns to scale, variations in production, and therefore capacity utilization, cause variations in firm productivity which aggregate to the macroeconomy. The effect is ephemeral because over time firms enter/exit which ameliorates the shock, causing incumbents to return to their original production.

The next chapter extends this analysis to consider that firm dynamics affect market competition which affects pricing markups (unlike this chapter with fixed markups). The benefit of this approach is to add a long-run productivity narrative to the short-run productivity dynamics studied hitherto.

3

THE EFFECT OF FIRM ENTRY ON MACROECONOMIC PRODUCTIVITY WITH ENDOGENOUS MARKUPS

In this chapter I present a theory of firm entry and exit in the business cycle that links short-run productivity overshooting to long-run persistence, a dynamic observed in contemporary *'productivity puzzles'*. The theory emphasizes two mechanisms: (1) slow firm entry/exit and (2) firm pricing that reflects the number of competitors in the market. Given these mechanisms, economic contraction causes a short-run exacerbated fall in productivity (overshooting) because the negative shock is absorbed by incumbents due to slow exit responses. This weakens incumbents' returns to scale, thus worsening productivity. However, the productivity overshooting recedes over time as firms exit which *dynamically reallocates* resources among incumbents, reviving the remainders returns to scale and thus productivity. This process of exit consolidating the market is not purely beneficial for productivity because the remaining firms face fewer competitors and thus charge higher markups which damages productivity. Therefore despite some reversion from the initial fall, there is a long-run persistent negative effect on productivity due to higher markups responding to the fall in number of firms. To analyze the trade-off between productivity improving dynamic reallocation and productivity degrading endogenous markups, I develop a continuous time, analytically tractable DGE model. The main mechanisms are dynamic entry so firms are slow to respond causing initial overshooting, and endogenous markups so pricing behaviour depends on the number of competitors firms face.

3.1 INTRODUCTION

The chapter proposes a business cycle theory in which firm entry and exit cause endogenous short-run and long-run productivity movements. Interest in endogenous productivity over the business cycle is high in light of Great Recession *productivity puzzles*[1]. The puzzles describe exacerbated productivity falls with weak recovery, and are prominent in several European countries shown in figure 13a for labour productivity[2]. The problem is especially pronounced in the UK, and empirical studies (Barnett et al. 2014) find that up to half of the shortfall in UK labour productivity relative to pre-crisis trend arose because of impaired resource allocation and unusually high firm survival rates[3]. This evidence emphasizes the importance of firm dynamics in explaining macroeconomic productivity, but traditional macroeconomic theory nullifies entry by assuming that the number of firms in an economy adjusts instantaneously to arbitrage profits. If entry is instantaneous, it can only affect productivity through an immediate change in the number of competitors which affects pricing markups, but it ignores the short-run effect of sluggish entry reallocating resources as firms adjust to arbitrage profit. In this chapter I analyze the

1 The term has been used extensively in the media and academia e.g. The Productivity Puzzle Under the Bonnet, The Economist, May 30, 2015; Budget 2015: How do you solve the 'productivity puzzle'?, BBC News, July 8, 2015

2 The source of both figures is Barnett et al. 2014 and they have been reproduced in the press e.g. Emily Cadman's UK Productivity Puzzle: The Bank of England's Answers, Financial Times, August 14, 2014.

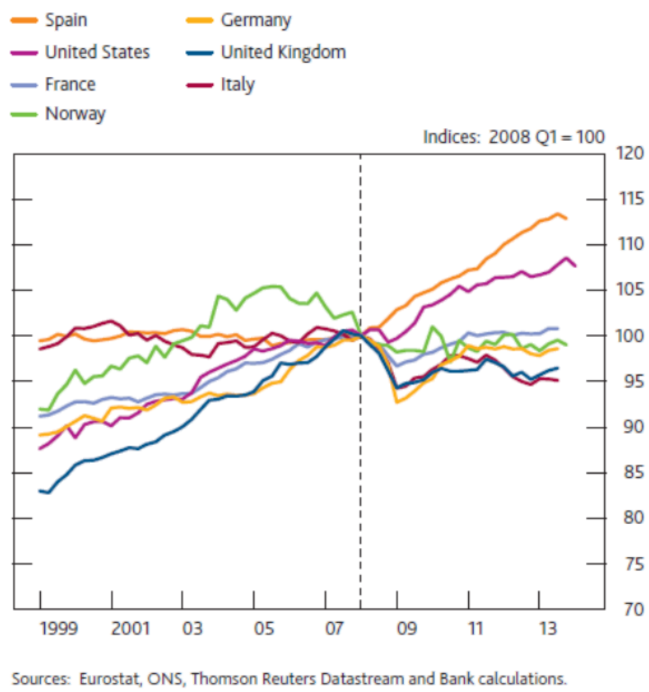
3 Goodridge, Haskel, and Wallis 2014 show that accounting for labour and capital still leaves a TFP puzzle. I focus on TFP.

new trade-off that emerges when noninstantaneous entry is combined with competitive endogenous markups.

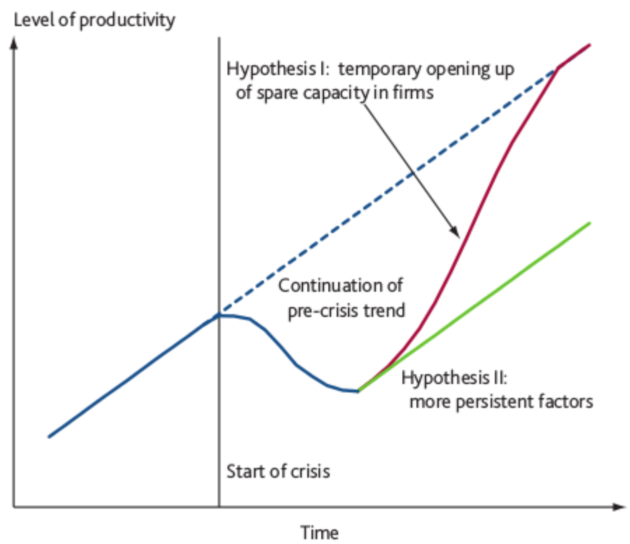
The main result is a theory to explain that shocks initially exacerbate productivity movements but the exacerbation relinquishes as firm entry/exit adjusts. Crucially productivity never regains long-run underlying productivity because of structural changes in competition due to long-run changes in the number of incumbents. To be clear, 'entry' is net entry, so when negative it is exit. Therefore entry and exit the same process–they cannot arise together. A contractionary shock will solely cause exit (negative net entry); an expansionary shock will solely cause entry (positive net entry). Hence the theory is a general explanation of endogenous productivity over the business cycle, with initially exacerbated and persistent positive productivity effects associated with entry in expansion and exacerbated and persistent negative productivity effects associated with exit in contraction. Although general, Great Recession productivity puzzles provide a contemporary view of the theory since they depict a short-run exacerbated fall in productivity followed by some persistence due to structural factors. Therefore my results provide a theory that combines the two hypotheses posed by the Bank of England in figure 13b. I demonstrate that a negative shock to the economy, modeled as a supply-side TFP shock, is first absorbed by incumbent firms because exit cannot arise initially. Therefore productivity falls drastically as the incumbents output falls and they suffer worse returns to scale (hypothesis I in figure 13b). Lower output per firms causes negative profits which leads to exit. As exit occurs productivity improves because resources are reallocated among incumbents and better returns to scale improves productivty, as shown by the hypothesis I reversion. However, this consolidation of resources among fewer firms reduces the competitive pressure on those who remain allowing them to charge higher markups. Higher markups mean each unit sold generates more revenue so that in a long-run zero profit equilibrium firms can produce less to cover their fixed cost of production. By choosing to produce less their scale suffers which creates an offsetting negative productivity effect that persists in equilibrium and hence links to hypothesis II in figure 13b.

I develop a tractable model of dynamic (endogenous) firm entry in the macroeconomy with imperfectly competitive product markets that cause endogenous markups. Dynamic entry means that firms slowly adjust to arbitrage profits, so short-run profits are nonzero. This entry friction arises because a *congestion effect* raises sunk entry costs as entry increases[4]. Imperfect competition creates a markup of factor prices above their marginal products, and the markup is endogenous because it depends on the number of firms. The relationship is negative and occurs because firms are large in their industry so they strategically interact under Cournot competition. With this model setup, I analyze the trade-off between *endogenous markups* and *dynamic reallocation*. Endogenous markups cause entry to increase productivity and exit to decrease productivity. Dynamic reallocation causes entry to decrease productivity and

4 A familiar notion in industrial organization (IO) theory e.g. Ericson and Pakes 1995



(a) Cross-country Labour Productivity



(b) Productivity Puzzle Hypotheses

Figure 13.: Productivity Puzzles (Source Barnett et al. 2014)

exit to increase productivity. For example, with endogenous markups exit (entry) weakens (strengthens) competition which raises (lowers) markups, thus decreases (increases) productivity. In opposition, dynamic reallocation means exit (entry) concentrates (dissipates) resources thus increasing (decreasing) incumbents' scale and therefore productivity. Dynamic reallocation emphasises not the amount of resources, but their division among firms. And entry determines this division. I analyze measured productivity, which is an adjusted measure of total factor productivity (TFP).

The model demonstrates procyclical profits, entry, employment and productivity, whereas markups are countercyclical. For example, a positive shock to technology is initially borne by incumbents who raise their output whilst entry is inert in the short-run. Through greater scale incumbents' productivity increases. However, by raising output incumbents accrue monopoly profits, these non-zero profits incentivise potential firms to begin entering. Entry reallocates resources and reduces output per firm which diminishes scale and therefore productivity. The influx in entry diminishes profits and through congestion raises the sunk entry cost which slows the rate of entry. Eventually the profits from incumbency arbitrage to zero so entry ceases and zero incumbency profits are balanced with zero sunk entry costs because there is no congestion. The long-run effect of a rise in the number of incumbents is that competition in the market is fiercer, so firms charge a lower markup. In order to cover fixed costs, firms with lower markups must raise revenue by increasing output, therefore in zero-profit (free-entry) equilibrium firm scale is increased which means there is a long-run permanent effect on productivity. In summary, the positive shock increases output, profit, employment (an input) and entry, whereas markups decrease because entry increases competition.

Formally the model follows a Ramsey-Cass-Koopmans setup. There is endogenous labour and capital, and the novel additions are firm entry and endogenous markups. A representative household chooses its consumption and labour exertion, but the household is limited by a budget which consists of labour income and investment income. Investment income consists of returns on capital and returns on firm ownership (firm profit). The return on capital is the economy's risk-free rate, which consequently determines the opportunity cost of investing in a firm. This balance between paying a cost to setup a firm and investing that cost at the market rate binds firm entry. It is a dynamic condition because sunk costs depend on the number of entering firms (congestion effect). Hence in free entry equilibrium[5] profits are zero, so a household is indifferent between creating a firm or investing that sunk cost at the risk free rate. For example, if the value of incumbency exceeds the risk free alternative, then there will be entry. Consequently congestion will stifle start-ups and entry will slow. On the firm side of the economy there is imperfect competition and generalized returns to scale (U-shaped cost curves). All firms produce with the same production function (firms are symmetric) which has a fixed overhead cost and nondecreasing marginal cost. The fixed overhead allows for imperfect competition which causes

5 The long-run equilibrium when firms have freely entered to arbitrage positive profits.

pricing markups. Firms are aggregated across two levels. The lowest level of aggregation is the firm level, and aggregating firms gives the industry level. The macroeconomy is the aggregate across all industries. I focus on symmetric equilibria so an industry is representative of the whole economy. Firms have price setting power within their industry, but are small in the aggregate economy. The influence of a firm on industry price causes endogenous markups[6]. Within an industry firms strategically interact with Cournot competition, so they maximise profits by choosing output to produce. This output choice is influenced by the number of competitors in the industry. When there are more competitors demand functions reflect a higher elasticity of market demand and therefore weaker markup setting power.

The model economy includes three core assumptions 1) endogenous entry 2) returns to scale 3) endogenous markups. The counterfactual of each assumption emphasizes its importance. First, in the absence of endogenous entry there is instantaneous free entry[7]. This counterfactual implies that there is no short-run productivity effect as incumbent firms bear shocks.Second, in the absence of increasing returns to scale, returns to scale are constant. This counterfactual makes entry impotent because firms produce at the same productivity regardless of size. Third, in the absence of endogenous markups, markups are

6 If a firm were small in its industry, markups would be fixed as in the status-quo Dixit and Stiglitz 1977 case.

7 This is a limiting case of my model, as is the other extreme a fixed number of firms.

fixed. This counterfactual implies there is no persistent effect on productivity because firms do not alter their markups.

Related Literature This chapter links Etro and Colciago 2010[8] to Jaimovich and Floetotto 2008. The first paper includes sluggish firm entry and endogenous markups, but does not discuss productivity. Their contribution is to improve business cycle moment-matching using Cournot and Bertrand strategic interactions; I use Cournot which the authors advocate. The second paper has endogenous markups and analyzes productivity, but firm entry is instantaneous. Their contribution is to explain the productivity effect of instantaneous entry on markups. This is equivalent to the long-run effect that causes productivity persistence in my chapter. My link combines endogenous entry with endogenous markups to explain productivity over the business cycle. The result is that endogenous entry distinguishes short-run productivity dynamics from long-run productivity dynamics.

The endogenous entry setup of this chapter follows Datta and Dixon 2002 which is close to industrial organization literature by Das and Das 1997. Importantly this differs from most recent endogenous entry literature that uses Bilbiie, Ghironi, and Melitz 2012 (BGM)[9]. However, the interpretation of the two approaches is analogous. Both endogenous entry formulation reduce to an arbitrage condition that equates sunk entry costs to incumbency profits. A strength of the Datta and Dixon 2002

8 Etro and Colciago 2010 also note that Cournot competition causes inefficiency through excess entry. Etro 2009 provides an excellent survey of macroeconomic models with endogenous entry and endogenous market structures.

9 For example, Lewis and Poilly 2012 and Etro and Colciago 2010.

formulation is that dynamics stem from endogenous sunk costs, rather than fixed sunk costs in BGM. These endogenous sunk costs are called congestion effects since they increase as number of entrants (congestion) increases. It is then a lemma that sunk entry costs equate to profit from incumbency, rather than in BGM where this is assumed. The BGM setup is influential in discrete time, simulation exercises[10], whereas the model in this chapter is continuous time and analytically tractable. BGM distinguish entry from exit (exit is exogenous), whereas this chapter treats them as symmetric. Entry measures the change in the number of incumbent firms, so negative entry is exit. Lewis 2009, Lewis and Poilly 2012, Lewis and Stevens 2015 and Berentsen and Waller 2009 all recognise the importance of congestion effects in macroeconomic models with entry.

An important distinction of this chapter is its focus on qualitative dynamical systems, rather than quantitative simulations in the afforementioned works. This follows Brito and Dixon 2013. Rather than productivity, their focus is on theorems to show that firm entry is sufficient for nonmonotone responses to fiscal shocks. Excluding imperfect competition removes the vital mechanism for generating increasing returns to scale that are necessary for productivity dynamics. This mechanism is present in Aloi and Dixon 2003 who use firm entry to explain productivity in an open economy without capital or endogenous markups. This mechanism between imperfect competition, increasing returns to scale and productivity is an estab-

10 And not limited to macroeconomics. Examples include Loualiche 2014 in finance, Peters 2013 in growth and Hamano and Zanetti 2014 in macroeconomics.

lished explanation for procyclical productivity over the business cycle (Hall 1989, Hall 1987, Caballero and Lyons 1992).

There are two competing formulations of endogenous markups. There is a supply-side approach, used in this chapter, Etro and Colciago 2010, Jaimovich 2007, and Jaimovich and Floetotto 2008. There is a demand-side approach used by Bilbiie, Ghironi, and Melitz 2012. The supply-side approach relies on firms strategically interacting which affects market demand. The demand-side approach relies on consumers' elasticity of substitution varying as product variety changes with entry. Lewis and Poilly 2012 compare the methods. Empirical business cycle literature shows many examples of countercylical markups. A cornerstone work is Bils 1987, and there are many contributions by Rotemberg and Woodford surveyed in Rotemberg and Woodford 1999. These traditional explanations of countercyclical markups rely on price stickiness. Whereas, the study of entry provides a new factor to enrichen markup countercylicality. The idea stems from the ubiquity of the relationship in empirical IO. For example Campbell and Hopenhayn 2005 find a negative correlation between markups and entry in many sectors of the US economy. In macroeconomics, Portier 1995 shows that entry is procylical and markups countercyclical over the French business cycle. Other empirical features that relate to this chapter are procyclical productivity Rotemberg and Summers 1990, and procyclical net business formation Bergin and Corsetti 2008. Lastly Brito, Costa, and Dixon 2013 model an economy with endogenous markups that embeds both traditional Dixit-Stiglitz monopolistic competition and entry-driven supply-side markups. This shows that monopolistic competition is a special case of the endogenous markups framework when there is only one firm per industry. They explore the critical bifurcation that arises as an economy moves from a continuum of 1-firm industries competing under monopolistic competition to a continuum of multi-firm industries competing under Cournot.

Roadmap – Section 3.2 explains the intuition behind the model. Section 3.3 outlines a model of firm entry in the macroeconomy where firms compete with strategic interactions. Section 3.4 begins analysis by explaining how the competition effect of entrants reducing markups affects factor prices and profits. Section 3.5 investigates static outcomes showing that long-run output and productivity are endogenous since they depend on the number of operating firms. Section 3.6 is the main result which presents a theorem to explain productivity puzzles, where productivity overshoots on the impact of a shock, then relinquishes but leaving some persistence.

3.2 INTUITION OF EXCESS CAPACITY WITH SHORT-RUN AND LONG-RUN CAPACITY UTILIZATION

Before developing a complex dynamic model with endogenous entry, imperfect competition and endogenous markups, a simple diagram can explain the intuition of how entry causes endogenous and persistent productivity dynamics. Figure 14 shows the cost curves and equilibria of a firm with increasing marginal

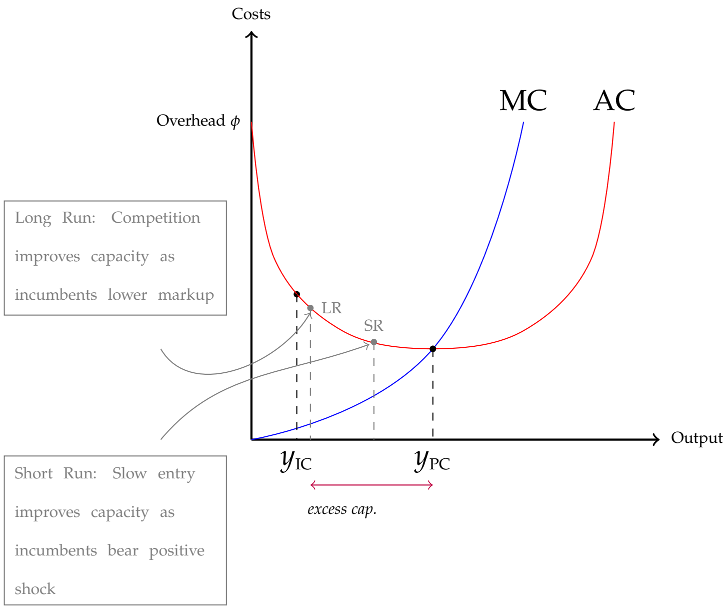


Figure 14.: Excess Capacity, Short-run and Long-run Utilization

costs and a U-shaped average cost due to a fixed overhead cost ϕ. Under imperfect competition a firm produces y_{IC}[11] which is less than the perfect competition outcome y_{PC}, which is also the efficient outcome as it minimizes costs. The difference between y_{IC} and y_{PC} is excess capacity[12] (labelled), and utilizing excess capacity lowers costs which improves firm productivity and in turn aggregate productivity. With an entry mechanism the underproduction of each firm in imperfect competition corresponds to excess entry. This means ther are 'too many' firms each underproducing, so a more efficient outcome is fewer firms but each producing more, hence with y_{IC} closer to minimium cost y_{PC}.

11 This is the long-run Chamberlin-Robinson equilibrium in which marginal revenue equals marginal cost and profits are zero. I omit the MR and MC curves for clarity, and I assume the curves have fully shifted following any shock.

12 Macroeconomists should note this definition of excess capacity which follows Vives 1999. It is distinct from *capital utilization* or any form of input intensity.

If there is a positive shock to the economy, and entry is slow (endogenous), then that shock is initially borne by the incumbents so they utilize capacity and costs lower. A move from y_{IC} to point SR in the diagram. This indicates an immediate increase in productivity in aggregate. But in this new position, incumbents earn monpoly profits which attracts entrants. Over time entrants move into the market, gradually reducing the capacity of incumbents so excess capacity rises back towards the initial level, finally halting at a position like LR where profits are zero again. This mechanism corresponds to the gradual amelioration of the initial boost in productivity as firms adjust. The final part of the story is most important, because it explains why although the initial drastic effect subsides there is still some long-run effect on productivity, so output per firm returns to LR rather than initial y_{IC}. As firms enter to arbitrage profit to zero they must now each produce more $y_{LR} > y_{IC}$ because the positive shock encouraged entry (raised the number of competitors) which put downward pressure on markups. Therefore in long-run zero profit equilibrium firms charge lower markups than in the initial pre-shock position. Consequently each incumbent must raise revenue by increasing output to cover the costs of production ϕ and attain zero profit. So overall there is a small fall in costs per firm due to capacity utilization, and thus a small but persistent improvement in productivity after the initial positive shock.

3.3 ENDOGENOUS ENTRY MODEL WITH IMPERFECT COMPETITION AND ENDOGENOUS MARKUPS

The model follows a Ramsey-Cass-Koopmans setup. Additions are imperfect competition, firm entry, endogenous markups and capital accumulation. The model is deterministic, and labour is endogenous. There are two state variables: capital and number of firms $(K, n) \in \mathbb{M} \subseteq \mathbb{R}^2$, where $\mathbb{M}$ is the state space of the control problem that later forms a subset of the general dynamical system state (or phase) space. I solve the model as a decentralised equilibrium because imperfect competition distorts the optimising behaviour of the firm.

Definition 6 Notation and Terminology.

Y_x denotes the derivative of Y with respect to X, except when $X = t$ which denotes time dependence. For clarity I usually omit the (t) notation that denotes time dependence in ordinary differential equations (ODEs). To be clear, the primitive endogeneous model variables are $C(t), e(t), K(t), n(t)$, defined later. They are the state variables of the four dimensional dynamical system which forms the model economy. The four states depend on time and therefore so do functions of them $L(t), r(t), w(t), \pi(t), Y(t), y(t), \mu(t), \mathcal{P}(t), \Pi(t), Z(t)$. Time dependence is irrelevant in steady state, which I denote with an asterix Y^*. Also for clarity, I often suppress function domains. For example, after first introduction $F : K \times L \to \mathbb{R}$ is written F rather than $F(K, L)$.

3.3.1 *Firm*

In the economy there is a continuum of sectors of measure one. In each industry, there is a finite number of intermediate firms that each produce a homogenenous good. Since the goods are homogeneous, they are perfectly substitutable in the production of an industry good. However, at the next level of aggregation, industry goods are imperfect substitutes for eachother when aggregated into a final good. Entry and exit of firms into existing sectors occurs until profits are zero. This does not happen immediately but occurs in the long run. This is known as the free entry equilibrium. In the short-run profits will diverge from zero as they are arbitraged by entrants. Perfect factor markets mean that each firm faces the same price w for labour and r for capital, and the result is that aggregate capital and labour are divided equally among firms $k = \frac{K}{n}$ and $l = \frac{L}{n}$. A lowercase letter denotes *per firm*, so output per firm $y = \frac{Y}{n}$.

A fixed cost in production allows firms to compete under imperfect competition in the product market. Strategic interactions occur under imperfect competiton because firms are large in their industry so can influence industry price. This is why markups are endogenous (depend on number of incumbent firms), rather than fixed in the traditional Dixit-Stiglitz monopolistic competition case where firms are small therefore do not affect industry price level (in fact this is a special case of my endogenous markup model where there is one firm per industry). I focus on Cournot competition so firms strategically interact through their choice of output to maximise profits given the behaviour of others. The form of strategic interaction determines the markup of factor price above marginal cost. Specifically I focus on the level of factor price markup above the factor marginal product. That is the markup of wage above marginal product of labour and interest rate above marginal product of capital.

It is important to note that in this chapter the measure of market power is the markup μ which is price over marginal cost $\frac{P}{MC}$ which rises from unitary under perfect competition to infinity under market dominance. Whereas, in the previous chapter our measure of market power was the Lerner Index ζ, the difference in price and marginal cost as a proportion of price $\frac{P-MC}{P}$. Therefore the relationship is $\mu = \frac{1}{1-\zeta}$. The Lerner Index was useful because it bound market power between $(0,1)$ which eased taking limits. The change in focus here keeps the chapter comparable to literature on endogenous markups such as Jaimovich and Floetotto 2008 and Lewis and Poilly 2012.

Final output Y is produced by a competitive firm using the output of a continuum of industries (*aka* intermediate goods or sectors) Q_j for $j \in [0,1]$ as inputs in a CES production function with constant elasticity of substitution $\theta_I \in (0,\infty)$.

$$Y(t) = \left(\int_0^1 Q_j(t)^{\frac{\theta_I - 1}{\theta_I}} dj \right)^{\frac{\theta_I}{\theta_I - 1}}, \quad \theta_I \in (0,1) \tag{98}$$

Cost minimization leads to conditional demand for industry j

$$Q_j(t) = \left(\frac{P_j}{P}\right)^{-\theta_I} Y \tag{99}$$

Thus the inverse demand function is $P_j = \left(\frac{Q_j}{Y}\right)^{-\frac{1}{\theta_I}} P$. Substituting the conditional industry demand (99) into the aggregate production function (98) gives the aggregate price index

$$P = \left(\int_0^1 P_j^{1-\theta_I} dj\right)^{\frac{1}{1-\theta_I}} \tag{100}$$

Notice that perfect competition in the final goods market requires equality of price and marginal cost P.

Assumption 5 Firm Production with U-shaped Average Cost Curve.
Firms are symmetric, so each has the same production technology. The ith firm in the jth industry produces output:

$$y_{j,i}(t) := \max\{AF(k_{j,i}(t), l_{j,i}(t)) - \phi, 0\} \tag{101}$$

where $F : \mathbb{R}^2_+ \supseteq (k,l) \to \mathbb{R}_+$ *is a firm production function with continuous partial derivatives which is homogenous of degree* $\nu \in (0,1)$ *(hod-*ν*) on the open cone* $\mathbb{R}^2_+$*, and* $\phi \in \mathbb{R}_+$ *a fixed cost denominated in output. The Hessian matrix of F has a symmetric main diagonal (Young's theorem), negative mixed derivatives (off-diagonal), and its determinant is positive so the concavity properties are*

$$F_{kl} = F_{lk} > 0, \qquad F_{kk}, F_{ll} < 0, \qquad F_{kk}F_{ll} - F_{kl}^2 > 0$$

Inada's conditions hold so that marginal products of capital and labour are strictly positive which rules out corner solutions.

$$F_k, F_l > 0$$

Although we shall focus on the case of U-shaped average costs many of our calculations hold without loss of generality for several cases. Appendix C.6.1 solves the firms static cost minimization problem from which these conclusions are deducible.

- $\phi > 0, \nu \in (0, 1)$ U-Shaped average cost and increasing marginal cost curve compatible with imperfect and perfect competition.
- $\phi = 0, \nu = 1$ Constant returns and no fixed cost so globally constant returns to scale. Average cost and marginal cost are equivalent.
- $\phi > 0, \nu = 1$ A fixed cost with constant marginal cost leads to globally decreasing average cost.
- $\phi \geqslant 0$ and $\nu \in (1, \infty)$ Both average and marginal costs are increasing, so there are globally increasing returns to scale. The extent to which ν exceeds 1 is bounded.

Notice that we view number of firms as a factor of production $F(k, l) = F\left(\frac{K}{n}, \frac{L}{n}\right) = n^{-\nu} F(K, L)$. It is essentially a measure of organization which captures how resources are divided. The production function with a fixed cost and decreasing returns to scale cause a U-shaped average cost curve. Decreasing returns to scale arise because the variable production function $F : \mathbb{R}^2_+ \supseteq (k,l) \rightarrow \mathbb{R}_+$ is convex, $\nu \in (0,1)$, in capital and labour which causes increasing marginal cost. The fixed cost ϕ creates a nonconvexity which prevents some firms producing because an active firm must sell at least enough to cover the fixed cost. The fixed cost occurs each period, and is different to the sunk entry cost which is paid once to enter (see Entry Section 3.3.1.3)[13]. $A \in [1,\infty)$ is a scale parameter reflecting the productivity level. It may be interpreted as total factor productivity (TFP). In Section 3.5.1.1 we derive *measured productivity* which is a function of TFP that captures the fixed cost and returns to scale effect. Since the average cost curve is U-shaped, there is an efficient level of production at minimum average cost, where average cost and marginal cost intersect. This is the Walrasian outcome that would arise under perfect competition.

3.3.1.1 *Strategic Interactions and Endogenous Markups*

Within each industry j there is Cournot monopolisitc competition among a set $\mathcal{I}(j)$ of $n(j) \in (1,\infty)$ firms. So the representative ith firm in industry j chooses output to maximise profits subject to the inverse demand function implicit in (99) and the quantities $y_{j,i'}$ supplied by other firms $i' \in \mathcal{I}(j) \setminus i$. It takes as given the quantitiy of final output Y produced by the competitive sector, the aggregate price level P of the intermediate sector

13 As in Jaimovich 2007 and Rotemberg and Woodford 1996 the role of this parameter is to reproduce the apparent absence of pure profits despite market power. It allows zero profits in the presence of market power.

(it cannot influence this price level) and the factor market prices w and r. Therefore it solves

$$\max_{(y(j,i),l_i,k_i)} P\left(\frac{(y_i + \sum_{i'} y_{i'})}{Y}\right)^{-\frac{1}{\theta_I}} y_i - rk_i - wl_i \tag{102}$$

$$\text{s.t.} \quad y_{j,i}(t) \leqslant AF(k_{j,i}(t), l_{j,i}(t)) - \phi \tag{103}$$

Each firm's technology is symmetric with respect to intermediate inputs that are shared equally due to perfect factor markets $k = \frac{K}{n}$ and $l = \frac{L}{n}$. The result is a symmetric equilibrium outcome, so we can drop i, j indexes and focus on a single representative industry as the whole economy.

Proposition 3 Markups are Endogenous.

Under symmetric equilibrium the first order conditions of the firms profit maximising problem lead to a markup $\mu(n(t)) \in (1, \infty)$ of price above marginal cost.

$$\mu(n(t)) = \frac{\theta_I n(t)}{\theta_I n(t) - 1} \tag{104}$$

Where $\theta_I \in (1, \infty)$ is intersectoral substitutability

Lemma 1 Markup Decreasing in Entry.

The markup is endogenous and decreasing in the number of firms $\mu_n = -\frac{\theta_I}{(\theta_I n - 1)^2} < 0$.

The negativity of the derivative of the markup with respect to number of firms captures the competition effect of entry lowering markup. When there are many firms in the industry $n \to \infty$, the markup disappears $\mu \to 1$ so price equals marginal cost which is the perfect competition outcome. The opposing limit $n = 1$ is the monopolistic competition special case.

Corollary 3 Fixed Monopolistic Competition Markup Special Case.

If $n(t) = 1$ then the economy is populated by a continuum of one firm industries each producing a differentiated product and the resulting fixed markup is the well-known monopolistic competition case (Dixit and Stiglitz 1977).

$$\bar{\mu} = \frac{\theta_I}{\theta_I - 1} \tag{105}$$

It is clear that the Dixit-Stiglitz case is an upper bound on the markup. Therefore endogenous markups will always be lower than the fixed markup case.

$$\mu(n) \leqslant \bar{\mu}$$

In this chapter, the most important feature of the endogenous markup is that it is decreasing in n. However, there are a number of ways to make the markup more complicated by assuming substitutability in industry rather than the homogeneous goods I have, or changing Cournot competition to Bertrand. This leads to various forms of markup that rely on both intra- and inter- sectoral substitutability and therefore provide useful extra degrees of freedom in numerical exercises like Jaimovich 2007 and Etro and Colciago 2010. However, despite possible additions all these papers' markups embody the key feature of a competition effect. In fact, like other theory papers Dos Santos Ferreira and Dufourt 2006 I shall later set $\theta_I = 1$, so the markup is only in terms of n. This is equivalent to a Cobb-Douglas aggregator of industry level goods.

An optimizing firm's choice of labour and capital correspond to an imperfectly competitive factor market equilibrium such that the price of a factor does not reflect its marginal product.

Proposition 4 Factor Market Equilibrium.

Under symmetric inter and intra-industrial equilibrium the optimal price setting rules are a markup of firms' marginal products.

$$AF_k(k(t), l(t)) = \mu(n)r(t) \tag{106}$$

$$AF_l(k(t), l(t)) = \mu(n)w(t) \tag{107}$$

The marginal revenue product of capital (MRPK) $\frac{AF_k}{\mu(n)}$ equates to the price of capital and the MRPL $\frac{AF_l}{\mu(n)}$ equals to the price of labour. As markups increase the marginal revenue from an additional unit of production is less. Because the MRPs are non-monotone functions of n_t there is the possibility of multiple equilibria. Different numbers of firm cause the factor market relationship to hold. I do not investigate these implications, instead I assume a unique solution.

3.3.1.2 *Profit*

Operating profit $\pi(t) : (K, L, n) \to \mathbb{R}$ is the profit of an incumbent firm in a given period. Operating profits exclude the one-time sunk entry cost that is included in aggregate profits, discussed after we cover the entry process.. Therefore operating profit of a firm is $\pi(t) := y(t) - r(t)k(t) + w(t)l(t)$ and by substituting in factor prices and using Euler's homogeneous function theorem we get profit under imperfect competition[14]

$$\pi(L, K, n; A, \phi) = \left(1 - \frac{\nu}{\mu(n(t))}\right) AF(k, l) - \phi \tag{108}$$

Profit is increasing in the markup and is greater than the perfect competition case of $\mu \to 1$. Profits are nonzero in the short run, but in long-run steady state we shall see they are zero (Section 3.5).

3.3.1.3 *Firm Entry*

I use the entry setup developed in Datta and Dixon 2002. The process of entry determines the number of firms $n(t)$ and the amount of entry $e(t)$ in a period. It is important to emphasize that 'entry' is 'net entry', so it measures the change in the stock of firms. If the stock of firms increases then net entry is positive so there has been entry, whereas if the stock of firms decreases then net entry is negative so there has been exit. This emphasizes that entry is a single symmetric process incorporating both entry and exit, and they cannot occur together. This is unlike papers that treat entry and exit as different processes. For example recent macroeconomics literature models a process of firm creation (entry), but treats exit as a fixed exogenous process (analogous to depreciation of capital). The importance of this point is that a positive shock to the economy will always

14 Rearranging the profit function gives the income identity which makes it clearer how markups enter output per firm and is equivalent to the production approach as follows $y(t) := r(t)k(t) + w(t)l(t) + \pi(t) = \frac{AF_l}{\mu(n(t))}l + \frac{AF_k}{\mu(n(t))}k + \left(1 - \frac{\nu}{\mu(n(t))}\right) AF - \phi = \frac{\nu AF}{\mu(n(t))} + \left(1 - \frac{\nu}{\mu(n(t))}\right) AF - \phi = AF - \phi = y(t)$.

cause solely entry and a negative shock solely exit. I shall focus on negative exit-inducing shocks, but the inverse argument would hold for positive shocks.

An endogenous sunk entry cost and an entry arbitrage condition determine the number of firms operating at time t. The sunk entry cost increases with the the number of entrants, and the arbitrage condition equates sunk cost with incumbency profits. Das and Das 1997 term the endogenous sunk cost an entry adjustment cost; in macroeconomics, Lewis 2009 and Berentsen and Waller 2009 use the term *congestion effect*, since more entrants cause congestion in entry that increases the sunk cost. The justification for congestion effects is that resources used to setup a firm are in inelastic supply, so that more entrants raises competition for the resources and therefore increases sunk cost. For example, when introducing a new product, if more firms are entering there is a negative entry externality because it is more costly to differentiate a product. Additionally to evidence for entry externalities and their prevalence in industrial organization literature, the assumption provides an analytical framework to study short-run dynamics away from steady state. It is the sunk entry cost that prevents instantaneous adjustment of firms to steady state[15].

15 The entry adjustment costs theory is analogous to capital adjustment cost models which recognise that investment (deinvestment) in capital is more costly for larger investment (deinvestment). The cost of investment depends on level of investment which is the flow of capital; analogously, the cost of entry depends of the level of entry which is the flow of no. firms. See Stokey 2008 for a modern account of capital adjustment costs.

Assumption 6 Sunk Entry Cost (congestion effect).

Sunk entry cost $q \in \mathbb{R}$ increases with the number of entrants $\dot{n}$ in t.

$$q(t) = \gamma \dot{n}, \qquad \gamma \in (0, \infty) \tag{109}$$

Entry and exit are symmetric for simplicity. A prospective firm pays sunk cost q to enter, and an incumbent firm pays $-q$ to exit. When firms are exiting $\dot{n} < 0 \implies q < 0$, hence $-q > 0$ so the cost of exit is positive. The congestion parameter γ is the marginal cost of entry, and its bounds are the two well-known cases: less sensitivity to congestion $\lim_{\gamma \to 0} q(t)$ implies instantaneous free entry, and more congestion sensitivity $\lim_{\gamma \to \infty} q(t)$ implies fixed number of firms. An extension of the sunk cost assumption to have a fixed cost and the congestion effect, where the fixed cost is paid regardless of the number of entering firms. This setup is closer to Das and Das 1997, and captures the classic case of fixed sunk costs as in Hopenhayn 1992a and Jovanovic 1982, but leads to multiple equilibria in our setup.

The congestion effect assumption is common in industrial organization literature, and has growing support in macroeconomics. Mata and Portugal 1994 show empirically that firm failure and industry entry rates are positively correlated, and theoretically Das and Das 1997 and Ericson and Pakes 1995 both assume sunk entry costs that rise with number of entrants[16]. The intuition for congestion is that there is more competition

16 Ericson and Pakes 1995 assume the sunk cost is non-decreasing in number of entrants. The assumption includes the simple case of fixed cost not responding to entrants, which they assume in the numerical exercise.

for a fixed resource needed to setup. For example, many firms entrants raise initial advertising costs to make consumers aware of the product. If many firms are entering there will be many startup advertising campaigns vying for attention. Congestion effects are also called "entry adjustment costs". In macroeconomics, Lewis 2009 uses a VAR analysis to show that congestion effects in entry weaken the volatility of entry responses which can improve model fit. Both Lewis and Poilly 2012 and Berentsen and Waller 2009 model congestion effects in a DSGE model. They differ slightly to our setup because entry reduces the probability of survival, for example by reducing the likelihood of a sale.

Similarly to Bilbiie, Ghironi, and Melitz 2012 (BGM) the crucial equation that binds entry is an arbitrage condition between returns to entry and the opportunity cost. Despite the literature's differing approaches to attaining endogenous entry (such as congestion sunk cost), all models of this theme ultimately reduce to a condition which equates profits from incumbency to the outside option.

Assumption 7 Entry Arbitrage (intertemporal zero profit). *The return to paying a sunk costs q to enter and receiving profits equals the return from investing the cost of entry at the market rate $r(t)$.*

$$\dot{q}(t) + \pi(t) = r(t)q(t) \tag{110}$$

Therefore there is an *intertemporal zero profit condition* that implies expected profits of an entrant are always zero; if they were ever non-zero, a firm would revise entry to a more profitable time. The zero-profit condition is dynamic rather than static. In the static case current profits, rather than expected future profits, are instantaneously zero (e.g. Jaimovich and Floetotto 2008), so the value of the firm equates to current profits.

Together the congestion effect assumption 6 and arbitrage assumption 7 form a second-order ODE in number of firms

$$\gamma \ddot{n}(t) - r(t)\gamma \dot{n}(t) + \pi(t) = 0 \tag{111}$$

The equation states that if profits are high, then to maintain zero the cost of entry is also high because there will be many entering firms. By defining entry, this second-order ODE is separable into two first-order ODEs

Definition 7 Net Entry and Exit.
Entry (or exit) is measured by the change in the stock of firms, therefore it is net entry, which if negative is called exit.

$$e(t) = \dot{n} \tag{112}$$

Therefore the model of industry dynamics which determines the number of firms is two ODEs

$$\dot{n}(t) = e(t) \tag{113}$$

$$\dot{e}(t) = -\frac{\pi(t)}{\gamma} + r(t)e(t), \quad \gamma \in (0, \infty) \tag{114}$$

With entry defined as the change in number of firms (112), the arbitrage condition's (114) interpretation depends on the rate of change of entry $\dot{e}(t)$, which is acceleration in number of firms $\ddot{n}(t)$. For example the rate of entry is increasing $\dot{e} > 0$ if the outside option $r(t)e(t)$ exceeds the profit from entering $\frac{\pi(t)}{\gamma}$. This is because when households invest in the more attractive outside option, as opposed to setting up firms, the cost of setting up a firm falls because there is less congestion. The result is an increase in the amount of entry. Initially, it is counterintuitive that the rate of entry decreases with profits, but this captures that when profits are high entry is high, so via congestion the cost of entry is high, and thus the rate of entry slows. The dynamic sunk cost causes firms to respond overtime rather than immediately. Intuitively a firm cannot instantaneously know its cost of entry. A propspective entrant must wait an instance in order to observe the amount of entry and therefore its sunk cost. Consider the contradiction that entry cost is fixed so observable in an instance, $q(t) = \gamma \ \forall t$. In which case the second-order ODE that dictate industry dynamics becomes static $\pi = r\gamma$, so there is no dynamic entry. Rather than the intertemporal zero-profit condition, there is an instantaneous alignment of current profit and opportunity cost. As shown in Datta and Dixon 2002 an implication of the model is that net present value of the firm (stock market value) equates to the sunk entry costs. In this sense the model is equivalent to BGM's approach, except the advantage here is that efficient stock market value is a corollary whereas in BGM it is assumed and then firms dynamics follow.

The aggregation of the sunk costs paid by entering firms leads to a deadweight loss that is not accounted for in operating profits $\pi(t)$ (which are period by period profits). Therefore aggregate profits must account for each firm's operating profits, less the aggregate sunk cost of entry. Based on the congestion effect assumption 6, if net entry is 0 then cost of entry is 0, for the next firm entering in that instance the cost now rises by an increment γ, and so on for each additional entrant up to the final e^{th} entrant in that time instance. Therefore the aggregate deadweight loss of entry $Z(t) \in \mathbb{R}$ is

$$Z(t) = \gamma \int_0^{e(t)} i \, di = \gamma \frac{e(t)^2}{2} \tag{115}$$

Therefore aggregating all n firms operating profits and deducting the deadweight loss gives

$$\Pi(t) = n(t)\pi(t) - Z(t) \tag{116}$$

$$\Pi = n \left[AF(k,l) \left(1 - \frac{\nu}{\mu(n)} \right) - \phi \right] - \gamma \frac{e(t)^2}{2} \tag{117}$$

Aggregate profits are an important factor driving capital investment $\dot{K}$. Ceteris paribus entry reduces aggregate profits. It increases the aggregate sunk costs of entry, and diminishes supernormal operating profits through the competition effect lowering markups. Further, this heightened effect of entry on profits will reduce the amount of entry and therefore reduce the size of the aggregate sunk entry cost.

3.3.2 *Household*

In the economy there is a continuum of identical households. This identical household chooses its future series of consumption $\{C(t)\}_0^\infty \in \mathbb{R}$ and labour supply $\{L(t)\}_0^\infty \in [0,1]$ to maximise lifetime utility $\mathrm{U} : \mathbb{R}^2 \to \mathbb{R}$. We assume $u : \mathbb{R} \times [0,1] \to \mathbb{R}$ is jointly concave and differentiable in both of its arguments. It is strictly increasing in C and strictly decreasing in L. A household's choice of consumption and labour is contrained by a its budget constraint which accrues capital income, labour income and profit income. The household owns capital $K \in \mathbb{R}$ and takes equilibrium rental rate and wage $r, w \in \mathbb{R}$ as given. Similarly they own firms and take profits as given $\Pi \in \mathbb{R}_+$. The household maximizes utility subject to a budget constraint rearranged to the law of motion of capital (119). The budget constraint shows that income is earned from capital income, labour income and profit from owning firms and it is spent on consumption or investment in more capital.

$$\mathrm{U} := \int_0^\infty u(C(t), 1 - L(t)) e^{-\rho t} dt \qquad (118)$$

$$\text{s.t.} \quad \dot{K}(t) = rK(t) + wL(t) + \Pi(t) - C(t) \qquad (119)$$

The optimization conditions from the problem reduce to three[17]: an intertemporal consumption Euler equation (120), an intratem-

17 Appendix C.1 derives the Hamiltonian and 6 associated Pontryagin conditions.

poral labour-consumption trade-off (121) and the resource constraint (119).

$$\dot{C}(t) = \frac{C(t)}{\sigma(C(t))}(r(t) - \rho), \tag{120}$$

$$w(t) = -\frac{u_L(L(t))}{u_C(C(t))} \tag{121}$$

where σ represents risk aversion $\sigma(C(t)) = -C(t)\frac{u_{CC}(C(t))}{u_C(C(t))}$. To complete the solution for the boundary value problem, we impose two transversality conditions on the upper boundary and an initial condition on the lower boundary.

$$\lim_{t\to\infty} K(t)e^{-\rho t} \geqslant 0, \quad \lim_{t\to\infty} K(t)\lambda(t)e^{-\rho t} = 0, \quad K_0 = K(0) \tag{122}$$

This completes the unique solution for the boundary value problem that characterizes the optimal path of consumption and labour: three variables (C, K, n), three equations (119)-(121), three boundary conditions (122).

In general equilibrium these equations hold and boundary conditions hold, with factor prices and profit determined endogenously from factor market equilibrium $r, w, \Pi : C \times K \times n \to \mathbb{R}_+$.

3.3.3 *Canonical Model in General Equilibrium*

Combining the equilibrium conditions from the household and the firm side of the economy defines the model economy as a four dimensional dynamical system that determines consumption, entry, capital and number of firms (C, e, K, n). Importantly labour supply L does not enter the system as an independent variable because it can be defined in terms of C, K, n by combining household intratemporal equilibrium condition with the factor market equilibrium from the firm problem. By understanding the trajectories of labour, we can trace how the *competition effect* of entry reducing markups affects the model through factor price equilibrium and consequently profits.

Proposition 5 General Equilibrium labour Supply.
Consumption reduces labour supply, whereas capital and number of firms increase labour supply $L(\underset{-}{C}, \underset{+}{K}, \underset{+}{n})$

$$L_C(C, K, n) < 0, \quad L_K(C, K, n) > 0, \quad L_n(C, K, n) > 0$$

Proof. The effects arise through combining factor market equilibrium (107), which determines wage, with the intratemporal condition (121), which determines consumption-labour choice. Then by the implicit function theorem differentiate the intratemporal condition with labour defined implicitly by $L(C, K, n)$. Derivations in appendix C.2. □

For capital the important determinant of the sign of labour response is labour marginal product which influences wage. Capital complements labour, so a rise in capital improves the marginal product of labour which consequently raises wage and labour supply. Consumption decreases in labour supply because additional consumption reduces the marginal utility of consumption so the value of consumption declines, thus reducing labour to support consumption (in other words leisure–inverse labour–becomes more attractive.)

The effect of entry on labour supply is more complex because it has an effect on markups too. The effect will feed through to the wage response to entry in section 3.4.

Corollary 4 Endogenous Markups Increase Labour Response to Entry.

Endogenous markups strengthen the labour response to entry relative to fixed markup ($\bar{\mu}$) case

$$L_n \geqslant L_n^{\bar{\mu}} \tag{123}$$

Proof. ($\mu(n)$ domain supressed.)

$$L_n = \frac{u_C A F_l \mu^{-1}(\mu_n \mu^{-1} - (1-\nu)n^{-1})}{u_{LL} + u_C A n^{-1} F_{ll} \mu^{-1}}$$

Consider the case of a fixed exogenous markup then $\mu_n = 0$ □

Entry (a rise in n) increases labour supply because it raises marginal product of labour and in turn wage (labour fixed). Wage rises because labour per firm falls which increases its marginal product due to decreasing returns $\nu < 0$. With constant returns there is no effect $L_n = 0$ as firms employ labour at equal productivity regardless of size. However with endogenous markups this does not hold, as even with constant returns there is the negative effect, $\mu_n \mu^{-1}$, which captures that a lower markup increases marginal revenue product of labour. Since this effect of entry diminishing markup brings wage inline with marginal product, it increases labour. Later we shall term this the *competition effect*, and the first effect due to returns to scale will be the *allocation effect*. When we analyse wage behaviour we shall see this extra markup effect will strengthen the labour effect on wage which creates downward pressure, but is offset by the direct effect of lower markup bringing wage closer to marginal product.

An interesting implication of Corollary 4 is that entry and labour supply are positively related. Vice-versa, exit leads to a fall in labour supply. In this chapter, we investigate the influence of entry on measured productivity, whereas much empirical discussion is based on labour productivity. The result shows that our model encapsulates labour productivity arguments as a specific case. For example, if there is a negative shock to the economy, to which firm exit does not respond instantaneously, labour supply will be buoyed which worsens labour productivity. Only when firms exit will employment begin to fall which will raise the productivity of remaining labour.

Definition 8 General Equilibrium.

Competitive equilibrium is the equilibrium paths of aggregate quantities and prices $\{C(t), L(t), K(t), n(t), e(t), w(t), r(t)\}_{t=0}^{\infty}$*, with prices strictly positive, such that* $\{C(t), L(t)\}_{t=0}^{\infty}$ *solve the household problem.* $\{K(t)\}_{t=0}^{\infty}$ *satisfies the law of motion for capital. Labour and capital* $\{L(t), K(t)\}_{t=0}^{\infty}$ *maximise firm profits given factor prices. The flow of entry causes the arbitrage condition on entry to hold (price of entry equals net present value of incumbency). State variables* $\{K(t), n(t)\}_{t=0}^{\infty}$ *satisfy transversality. Factor prices are set according to factor market equilibrium (107) and ensure goods and factor markets clear.*

The dynamic equilibrium conditions from the previous section are the capital accumulation equation, the number of firms definition, the consumption Euler, and the entry arbitrage condition.

Definition 9 Nonlinear System.

The dynamical system defines at a point in time $t \in \mathbb{R}$ *the state of the system* $(C(t), e(t), K(t), n(t)) = x(t) \in \mathbb{X} \subseteq \mathbb{R}^4$ *described by a* C^1 *vector valued transition map* $g : \mathbb{R}^{5+\mathbb{P}} \supseteq \mathbb{X} \times \mathbb{R} \times \Omega \longrightarrow \mathbb{R}^4$*. The parameterization* $(\phi, \nu, \gamma, \rho)$ *is defined on an open set* $\Omega \in \mathbb{R}^{\mathbb{P}}$

The system is

$$\dot{K} = Y(t) - \frac{\gamma}{2}e(t)^2 - C(t), \qquad Y = n(F(k,l) - \phi) \tag{124}$$

$$\dot{n} = e(t) \tag{125}$$

$$\dot{C} = \frac{C(t)}{\sigma(C)}(r(t) - \rho), \qquad \sigma(C) = -\frac{Cu_{CC}}{u_C} \tag{126}$$

$$\dot{e} = r(t)e(t) - \frac{\pi(t)}{\gamma}, \qquad \pi = AF(k,l)(1 - \frac{\nu}{\mu}) - \phi \tag{127}$$

where factor prices

$$r = \frac{AF_k(k,l)}{\mu(n)}, \qquad w = \frac{AF_l(k,l)}{\mu(n)} \tag{128}$$

Substituting in factor prices, profits and output which are all in terms of (C, K, n) and noting that by the intratemporal condition and wage equilibrium L is implicitly defined as $L(C, K, n)$ the model economy is a system of four ODEs in consumption, entry, capital, number of firms (C, e, K, n). Also by Euler's homogeneous function theorem note $F(k,l) = n^{-\nu}F(K,L)$ and $F_k(k,l) = n^{1-\nu}F_K(K,L)$. This gives a more primitive description of the system, with less economic intuition, but easier to understand the underlying dynamics.

$$\dot{K} = n(An^{-\nu}F(K, L(C, K, n)) - \phi) - \frac{\gamma}{2}e^2 - C \tag{129}$$

$$\dot{n} = e \tag{130}$$

$$\dot{C} = -\frac{u_C}{u_{CC}}\left(\frac{An^{1-\nu}F_K(K, L(C, K, n))}{\mu(n)} - \rho\right) \tag{131}$$

$$\dot{e} = \frac{An^{1-\nu}F_K(K, L(C, K, n))e}{\mu(n)} - \frac{An^{-\nu}F(K, L(C, K, n))(1 - \frac{\nu}{\mu(n)}) - \phi}{\gamma} \tag{132}$$

3.4 Competition Effect on Factor Prices and Profit

The competition effect enters the model through factor market equilibrium affecting factor prices r and w and in turn affecting profit π. After outlining these mechanisms in this section, the following section on steady state analysis shows that the mechanism propagates to long-run outcomes, where it raises output per firm and productivity.

Definition 10 Competition Effect of Entry/Exit.

The competition effect is caused by entry's effect on markups. That is, the markup $\mu(n(t))$ decreases in the number of firms competing $\mu_n < 0$. This was shown in lemma 1. The competition effect is zero with exogenous markups $\bar{\mu}_n = 0$.

Definition 11 Allocation Effect of Entry/Exit.

The allocation effect is that entry and exit alter the allocation of resources (capital and labour) among firms. This affects scale of pro-

duction, which is important due to decreasing returns to scale in production. Entry causes 'business stealing' reducing inputs per firm, whereas exit causes 'business consolidation' raising inputs per firm.

There are three effects of an entering firm on factor prices, and analogous three effects on operating profits which are a function of factor prices and form a key result (proposition 6).

$$w_n = \frac{A}{\mu(n)} \left[(1-\nu) n^{-\nu} F_L + n^{1-\nu} F_{LL} L_n - \frac{n^{1-\nu} F_L}{\mu} \mu_n \right] \lesseqgtr 0 \tag{133}$$

$$r_n = \frac{A}{\mu(n)} \left[(1-\nu) n^{-\nu} F_K + n^{1-\nu} F_{KL} L_n - \frac{n^{1-\nu} F_K}{\mu} \mu_n \right] > 0 \tag{134}$$

The three effects are a positive allocation effect, an ambiguous labour effect and a positive competition effect. The allocation effect captures that an extra firm reduces per firm allocation of inputs, but raises aggregate number of firms. The reduction in labour or capital per firm raises their marginal product and therefore price due to returns to scale. Hence with constant returns $\nu = 0$ the effect is not present. The labour effect captures that entry increases labour supply and therefore lowers wage or raises interest rate. Lastly the competition effect that arises from endogenous markups $\mu_n < 0$ captures that an extra firm increases competition and lowers markups which raises the marginal revenue product of labour (capital) and so wage (interest rate) must increase to maintain equilibrium. The interest rate response (134) is unambiguous because the labour effect is positive since labour complements capital, so it raises marginal product of capital. However in the wage result (133) this same labour effect is negative which creates an ambiguity because extra labour (caused by entry) reduces the marginal product of labour so depresses wages. Despite us showing in corollary 4 that the labour effect is stronger with endogenous markups, there is the offsetting positive wage effect that endogenous markups reduce the disparity between wage an marginal product (competition effect), which is reasonable to assume dominates the second-order effect on labour supply. Furthermore labour marginal product is buoyed by entry dividing resources among more firms (allocation effect). I merely acknowledge that entry and the surge in labour it creates can detriment marginal product of labour to the extent that wage indeed falls.

Profits are increasing in markups, which adds an extra effect of a firm entering the market. The result is that profits diminish faster, than if markups were fixed.

Proposition 6 Entry Effect on Profits.

Entry has two negative effects on operating profit.

$$\pi_n = \frac{A}{\mu}\left[(\mu-\nu)\left(-\nu n^{-\nu-1}F + n^{-\nu}F_L L_n\right) + \frac{\nu n^{-\nu}F}{\mu}\mu_n\right] < 0 \tag{135}$$

Proof. By substituting in L_n it can be shown that the negative scale effect dominates the positive labour effect in the second component of π_n. See appendix C.3 □

The three effects of an entering firm on factor prices feed through to profits. The allocation effect decreases profit, the labour effect increases profit, the competition effect decreases profit. The competition (markup effect) and allocation (business stealing) effect reinforce eachother. Making the negative effect of entry on profits larger than the case with fixed markups. However, there is a positive effect on profits from labour, since an entrant raises the market wage leading to a higher supply of labour which raises per firm output and revenues. The effect can be shown to be dominated hence the inequality.

Corollary 5 Profit More Responsive to Entry Under Endogenous Markup.
The responsiveness of profit to firm entry is absolutely larger in the case of endogenous markups.

$$|\pi(\mu(n))_n| > |\pi(\bar{\mu})_n|$$

In sum the competition effect of entry depresses markups which raises wage and interest rate in factor market equilibrium. This effect of higher factor prices (prices closer to their marginal product) causes profits to fall more from each entrant. The result is that zero profit arises when fewer firms have entered and so each firm has a larger market share. Conversely a negative shock that leads to negative profits and exit means each exiter raises incumbents profits less regaining zero profits requires more exit and then remaining firms can produce less. So the mechanisms in this section are responsible for the results in the next section that show output per firm and productivity are increasing in the number of firms.

3.5 EFFICIENCY AND STEADY STATE OUTCOMES

This section first derives the efficient outcomes that correspond to minimum average cost, or the number of firms that maximises output. It then analyzes the fixed point of the dynamical system, which corresponds to the zero profit outcome, often assumed instantaneously in other papers.

3.5.1 *Efficient Output and Productivity*

In symmetric equilibrium aggregate output is the number of firms in a representative sector multiplied by the amount a firm produces. It is homogeneous of degree 1 in K, L, n. This captures that capital and labour per firm do not change if all factors are changed equally, so output per firm is homogeneous of degree 0, but aggregation across all firms causes a proportional increase because of a proportional change in the number of firms that are being aggregated.

$$Y(t) = n(t)y(t) = n(t)^{(1-\nu)}AF(K(t), L(t)) - n(t)\phi \qquad (136)$$

The effect of an entrant on aggregate output is

$$Y_n = (1-\nu)n^{-\nu}AF(K,L) + n^{1-\nu}AF_L(K,L)L_n(C,K,n) - \phi$$
$$\approx (1-\nu)n^{-\nu}AF(K,L) - \phi \quad (137)$$
$$Y_{nn} = (1-\nu)n^{-1}[-\nu n^{-\nu}AF(K,L) + n^{1-\nu}AF_L(K,L)L_n(C,K,n)]$$
$$\approx (1-\nu)n^{-1}[-\nu n^{-\nu}AF(K,L)] \quad (138)$$

with approximations due to small second-order labour effect; alternatively, take labour taken as given at aggregate level. There are three effects of firm entry on aggregate output: an ambiguous allocation (returns to scale) effect, a positive labour effect and a negative fixed cost effect (resource duplication). The scale effect is positive with decreasing returns $\nu \in (0,1)$ and zero with constant returns $\nu = 0$. The effect captures that business reallocation among more firms improves aggregate output when there are decreasing returns.

At the point where the positive returns to scale and labour effect equate the fixed cost effect, there is an optimal efficient number of firms $Y_n = 0|_{\nu<1}$. These outcomes are those that would arise under Walrasian perfect competition, and if there were no markups in our model ($\mu = 1$). In the AC-MC diagram this is where they intersect at minimum average cost. To ensure the outcomes are defined we assume rising marginal cost $\nu \in (0,1)$ and to ensure production is nonnegative we assume fixed cost effect exceeds positive labour effect $\nu\phi \geqslant An^{1-\nu}F_L(K,L)L_n$ so there are initially decreasing returns as costs decrease toward the minimum. Generally we assume the second-order labour effect is small.

Proposition 7.

When output is maximised with respect to number of firms in the economy the efficient levels of output are

$$F(k^e, l^e) = \frac{1}{(1-\nu)}\left(\frac{\phi}{A} - F_l(k,l)L_n\right) \approx \frac{\phi}{A(1-\nu)} \tag{139}$$

$$\begin{aligned} y^e &= AF(k^e, l^e) - \phi \\ &= \frac{1}{1-\nu}\left(\phi\nu - AF_l(k,l)L_n\right) \approx \frac{\phi\nu}{1-\nu} \end{aligned} \tag{140}$$

It is notable that most papers dicussing entry focus on constant returns to scale. As this section has shown this implies there is no optimal firm size; analogously there is no perfect competition equilibrium because the market tends to a natural monopoly due to the fixed cost teamed with constant returns to scale. Firm size, in terms of factors it employs, is unimportant because all firms produce at the same efficiency. This limits the role of entry, so that productivity results arise solely from the competition effect of more firms reducing markups. How output is divided among firms does not matter. We shall see in corollary 7 that when entry is high (the market is competitive), imperfect competition outcomes converge on this section's efficient outcomes ((140) and later (142)).

3.5.1.1 *Homogeneous Degree Zero Productivity*

I call productivity at a point in time *measured productivity*. The measure is equivalent to TFPR (R for revenue) in Peters 2013[18]. Corresponding to the efficient levels of output is a definition of productivity that is also maximised at these efficient output levels, taking labour as given.

Definition 12 Measured Productivity.

Measured productivity $\mathcal{P} : K, L, n \to \mathbb{R}_+$ *is the amount of output an economy produces for a given technology, with technology normalized to be homogeneous of degree 1 to remove scale effects*

$$\mathcal{P}(t) = \frac{Y(t)}{F(K(t), L(t))^{\frac{1}{\nu}}} \tag{141}$$

This aggregate measure is the same as the per firm measure $\mathcal{P} = \frac{n(t)y(t)}{F(K(t),L(t))^{\frac{1}{\nu}}} = \frac{y(t)}{F(k(t),l(t))^{\frac{1}{\nu}}}$.

A more productive economy has larger measured productivity because it combines inputs more efficiently and produces more output with the same technology as another economy. An outcome of this definition of measured productivity is that when it is maximised with respect to number of firms $\mathcal{P}_n = 0$ the corresponding levels of output are the efficient outcomes y^e.

18 Based on Foster, Haltiwanger, and Syverson 2008.

Therefore the maximum attainable productivity $\mathcal{P}^e$ that arises at the efficient level of production is

$$\begin{aligned}\mathcal{P}^e &= \frac{y^e}{F(k^e, l^e)^{\frac{1}{\nu}}} \\ &= \frac{(\phi\nu - AF_l L_n) A^{\frac{1}{\nu}} (1-\nu)^{\frac{1}{\nu}-1}}{(\phi\nu - AF_l L_n)^{\frac{1}{\nu}}}, \quad \phi\nu > AF_l L_n,\ \nu \in (0,1) \\ &\approx \left(A\nu^{\nu} \left(\frac{\phi}{1-\nu} \right)^{\nu-1} \right)^{\frac{1}{\nu}} \end{aligned} \tag{142}$$

And in the constant returns limit the maximum attainable measured productivity is equivalent to TFP $\lim_{\nu \to 1} \mathcal{P}^e = A$.

Since production technology in the denominator is $hod - \nu$ we need to normalize it to be $hod - 1$. Then productivity will be $hod - 0$ in inputs. That means that the scale of inputs K, L, n does not affect productivity. Whereas with a typical non-normalized measure an economy with more inputs would always appear less productive. Hence we capture changes in efficiency of technology use, how effectively the inputs are combined with a given technology, rather than how many inputs there are. Consider an example of two economies $\mathcal{A}$ and $\mathcal{B}$. They are identical in every sense, except economy $\mathcal{B}$ is endowed with $\lambda \in (1, \infty)$ times more factors K, L, n. Since the economies are identical, except for scale of factors, then a good productivity measure should reflect that both economies have the same productivity: they combine factors with the same efficiency to produce output. Now assume the contradiction that we do not normalize technology and use a standard, non-normalized, TFP measure $\hat{\mathcal{P}}$. Then $\hat{\mathcal{P}}^{\mathcal{A}} = \frac{Y(t)}{F(K(t),L(t))}$ and since

Y is $hod - 1$ and F is $hod - \nu$ then $\hat{\mathcal{P}}^{\mathcal{B}} = \frac{\lambda Y(t)}{\lambda^{\nu} F(K(t),L(t))} = \lambda^{1-\nu}\hat{\mathcal{P}}^{\mathcal{A}}$. So $\hat{\mathcal{P}}^{\mathcal{A}} < \hat{\mathcal{P}}^{\mathcal{B}}$ we conclude erroneously that economy $\mathcal{B}$ is more productive simply because it has more factors, not because it combines those factors more efficiently. Under our normalized measure $\mathcal{P}^{\mathcal{B}} = \frac{\lambda Y(t)}{(\lambda^{\nu} F(K(t),L(t)))^{\frac{1}{\nu}}} = \mathcal{P}^{\mathcal{A}}$. With constant returns $\nu = 1$ there are no scale effects, so our measure collapses to the common definition.

3.5.2 *Steady State*

Now I shall show that the steady state of our economy corresponds to zero profits. And leads to levels of output and productivity that depend endogenously on the number of firms, and these levels are strictly less than the efficient levels that would arise under perfect competition defined in section 3.5.1.

Assume that a solution of the system converges to a unique steady state $(K, n, C, e) \to (K^*, n^*, C^*, e^*)$ as $t \to +\infty$[19]. In steady state $\dot{K} = \dot{n} = \dot{C} = \dot{e} = 0$, which immediately implies

19 Ignore the trivial steady state that arises when the state vector is the zero vector. It is possible that with endogenous markups there are multiple values of n^* that allow steady-state to hold. This is an investigation for future research.

entry is zero via (144), which in turn, via 146, implies profits are zero.

$$\dot{K} = 0 \Leftrightarrow Y^*(C^*, K^*, n^*) = C^* \tag{143}$$

$$\dot{n} = 0 \Leftrightarrow e^* = 0 \tag{144}$$

$$\dot{C} = 0 \Leftrightarrow r^*(C^*, K^*, n^*) = \rho \tag{145}$$

$$\dot{e} = 0 \Leftrightarrow \pi^*(C^*, K^*, n^*) = 0 \tag{146}$$

In steady state aggregate output equates to consumption; entry is zero; the interest rate equals the discount factor and profits are zero. Intuitively when profits are zero entry ceases as there is no entry incentive, and when the discount factor and interest rate are equated there is indifference between consumption and saving so all output is consumed. Rewriting the system in terms of underlying variables (C, e, K, n), again with labour defined implicitly $L(C, K, n)$, shows that ouput per firm and therefore measured productivity depend endogenously on the number of firms.

$$n^* \left[A n^{-\nu} F(K^*, L^*) - \phi \right] = C^* \tag{147}$$

$$e^* = 0 \tag{148}$$

$$\frac{A {n^*}^{1-\nu} F_K(K^*, L^*)}{\mu(n^*)} = \rho \tag{149}$$

$${n^*}^{-\nu} F(K^*, L^*) = \frac{\phi}{A\left(1 - \frac{\nu}{\mu(n^*)}\right)} \tag{150}$$

3.5.3 *Steady State Existence*

In this section I provide a condition under which a steady state solution always exists, and in appendix C.4 I calculate a specific solution numerically. For a steady state to exist the following system must be solvable for (C^*, K^*, n^*) (where I have substituted in the trivial $\dot{n}$ condition that $e = 0$).

$$\dot{C}: \quad 0 = r(C^*, K^*, n^*) - \rho \tag{151}$$

$$\dot{e}: \quad 0 = \pi(C^*, K^*, n^*) \tag{152}$$

$$\dot{K}: \quad 0 = Y(C^*, K^*, n^*) - C^* \tag{153}$$

Therefore the determinant of the Jacobian of this three dimensional system with respect to (C, K, n) is[20]

$$\left.\begin{bmatrix} r_C & r_K & r_n \\ -\pi_C & -\pi_K & -\pi_n \\ Y_C - 1 & Y_K & Y_n \end{bmatrix}\right|_{x=x^*} \implies \begin{bmatrix} - & - & + \\ + & - & + \\ - & + & \pm \end{bmatrix} \tag{154}$$

where determining the signs $r_K < 0$ and $\pi_n < 0$ requires extra work[21], and the sign structure is determinable before evaluating at steady state. I assume $Y_n > 0$. Each element is evaluated in a neighborhood of the conjectured steady state, although the sign structures hold regardless[22].

20 This proof of existence follows the approach of Caputo 2005, pp.419.
21 See Appendix C.3 the results involve substituting in labour effects and showing they are dominated by using the second partial derivative test for concavity which the production function is assumed to satisfy.
22 Notice it is important to distinguish between the derivative of a variable x with respect to z when x is in steady state x_z^*, as opposed to the derivative of a variable evaluated at steady state $x_z|_{x=x^*}$.

The determinant is

$$\overbrace{-\pi_K r_C Y_n}^{+} + \overbrace{-\pi_n r_K (Y_C - 1)}^{+} + \overbrace{-\pi_C r_n Y_K}^{+}$$
$$\overbrace{- - \pi_K r_n (Y_C - 1)}^{-} \overbrace{- - \pi_C r_K Y_n}^{+} \overbrace{- - \pi_n r_C Y_K}^{+} \quad (155)$$

Unfortunately the determinant is not clearly nonzero due to the $\overbrace{- - \pi_K r_n (Y_C - 1)}^{-}$ term even though all other terms are positive. However if this negative term can be shown to be dominated by one of the positive terms, it provides a sufficient condition for determinacy I.e. the determinant is strictly positive ensuring that a solution C^*, K^*, n^* exists by the implicit function theorem. The positive term $\overbrace{-\pi_n r_K (Y_C - 1)}^{+}$ proves a good candidate to dominate the negative $\overbrace{- - \pi_K r_n (Y_C - 1)}^{-}$ so their sum is positive. That is, $\overbrace{-\pi_K r_n (Y_C - 1)}^{+} - \overbrace{-\pi_n r_K (Y_C - 1)}^{+} > 0$, or simply

Lemma 2 Steady State Existence.

A steady state solution $\{C^*, K^*, n^*\}$ of the the system (151)-(153) exists if the following sufficiency condition holds

$$r_n \pi_K - r_K \pi_n < 0$$

This is a sufficient condition for determinacy, and it has an intuitive interpretation to support it. The condition states that the combined effect of $r_K \pi_n$ dominates the combined effect of $r_n \pi_K$. Since r_K is a direct effect (i.e. how capital effects its own price) it is sensible to believe it outweighs r_n, and similarly since

π_n is the direct effect of number of firms on value of firms it is sensible to believe it is stronger than the π_K effect. Hence we should believe that the combined positive effect of two second-order effects $r_n \pi_K$ is weaker than the combined positive effect of two first-order effects $r_K \pi_n$. And thus the former minus the latter will be negative. Indeed Appendix C.4 shows that for a standard numerical example with Cobb-Douglas production and Isoelastic utility a solution to the system exists.

Theorem 3.5.1 Endogenous Steady State Output and Productivity.

Steady state output per firm y^* and measured productivity $\mathcal{P}^*$ are endogenous because they depend on the markup which depends on the endogenous variable $n(t)$, the number of active firms.

$$y^*(\underset{+}{n^*}, \underset{-}{\mu(n^*)}) = \frac{\phi \nu}{\mu(n^*) - \nu} \tag{156}$$

$$\mathcal{P}^*(\underset{+}{n^*}, \underset{-}{\mu(n^*)}) := \frac{y^*}{F(k^*, l^*)^{\frac{1}{\nu}}} = \nu \left[\frac{A}{\mu(n^*)} \left(\frac{\mu(n^*) - \nu}{\phi} \right)^{1-\nu} \right]^{\frac{1}{\nu}} \tag{157}$$

Proof. From the zero profit condition variable production becomes $AF(k^*, l^*) = \frac{\phi \mu(n^*)}{(\mu(n^*) - \nu)}$, and then $y^* = AF(k^*, l^*) - \phi$. Then substitute into the expression for productivity. □

Corollary 6.

Steady state output per firm and measured productivity are increasing in the number of firms in the economy.

$$y_n^* > 0, \qquad \mathcal{P}_n^* > 0$$

Proof. The result for output y^* is clear. For productivity consider

$$\frac{\partial \mathcal{P}^*}{\partial n} = -\frac{A\nu}{\mu^2}\frac{(\mu - 1)}{(\mu - \nu)^\nu \phi^{1-\nu}} \cdot \frac{\partial \mu(n^*)}{\partial n} > 0$$

□

Since markups are decreasing in number of firms, output is increasing in number of firms, and similarly productivity is increasing in number of firms. The simpler case of constant marginal cost gives a similar outcome $\mathcal{P}^*|_{\nu=1} = \frac{A}{\mu}$ as in Jaimovich and Floetotto 2008 and Peters 2013. Given markups are negatively related to the number of firms, a single firm needs to sell more output to cover its fixed cost and break even in a free entry equilibrium. Hence with more firms, output per firm rises.

From (157) the fixed cost (ϕ) and decreasing returns to scale ($\nu < 1$) cause a dampening effect that captures that productivity is less sensitive to markups when the fixed cost is high. This is because fixed costs induce higher output per firm and therefore closer to constant returns to scale (nearer minimum AC), thus variations in output around this point caused by the changing markup has less of a productivity effect. This component falls out when there are constant returns to scale $\phi^{1-\nu}|_{\nu \to 1} \to 1$, because the fixed cost is used at equal efficiency regardless of scale.

The markup causes extra profit that helps us to understand the mechanism through which entry is affected. Profits offer entry incentives, and incentives rise when markups are higher which encourages more entry than would arise under perfect competition. However since markups are decreasing in number of firms the excessive amount of profit will diminish faster as firms enter, and zero profit will arise when fewer firms have entered so the allocation effect (business stealing effect) is dampened and larger firms remain. Thus each firm produces more and benefits from returns to scale, which fosters endogenous productivity. That is, long-run underlying productivity is a function of number of firms whereas with fixed markups firms entry always returns the economy to a position with the same productivity. The extra mechanism is important, since a prospective firm now considers how fierce competition in the market is, whereas with fixed markup it took for granted that it could enter and charge a given markup, thus produce a given amount in the long run. This leads to an important corollary

Corollary 7 Efficiency of Imperfect Competition.
When there is a high degree of competition (a large amount of entry) the imperfect competition outcomes ((156) and (157)) converge upon the efficient outcomes ((140) and (142)) because the markup is suppressed.

3.5.4 *Aggregate Output in Steady State*

Using the endogenous steady state output expression (156) we can infer the behavior of number of firms in steady state.

Lemma 3 Firm Procyclicality.

Given aggregate output at steady-state level Y^*, the number of active firms is procyclical.

$$n^* = \left[AF(K^*, L^*(C^*, K^*, n^*))\left(\frac{\mu(n^*) - \nu}{\mu(n^*)\phi}\right)\right]^{\frac{1}{\nu}} = \left(\frac{\mu(n^*) - \nu}{\phi\nu}\right)Y^* \tag{158}$$

Proof. For a given steady-state level of output Y^*, a change in this level on n^* causes $n^*_{Y^*} = \left(\frac{\mu(n^*)-\nu}{\phi\nu}\right) > 0$. The sign can be determined since the markup $\mu(n) > 1$ and returns to scale are decreasing (increasing marginal costs) $\nu \in (0, 1)$, therefore the numerator is positive.

See appendix C.4.2 for extended proof treating Y^* endogenously via implicit function theorem. □

When there is a rise in steady state aggregate output Y^* the number of firms increases since $\left(\frac{\mu(n^*)-\nu}{\phi\nu}\right) > 0$. Closer to constant returns to scale $\nu \to 1$ and higher fixed costs ϕ weaken the procyclicality effect, whereas greater imperfect competition $\mu(n^*)$ strengthens the effect[23]. Furthermore the direct effect of an increase in technology A is to raise number of firms $n^*_A > 0$ through its positive effect on Y^*. I validate this result numerically in section 3.6.2.

With fixed markups output per firm always returns to a constant level that is the amount of sales required to cover the fixed cost. Now μ is endogenous, if n rises then more sales are needed to cover the fixed cost and therefore output per firm

23 This result generalizes eq. 19 Jaimovich and Floetotto 2008, pp. 1245 who show an analogous outcome under constant returns and instantaneous entry.

in equilibrium depends on the number of firms in the market. Therefore in aggregate there is an increase in number of firms, and there is an increase in output per firm

$$Y^* = n^* y^* = (AF(K^*, L^*(C^*, K^*, n^*)))^{\frac{1}{\nu}} \left(\frac{\mu(n^*)\phi}{\mu(n^*) - \nu} \right)^{1-\frac{1}{\nu}} \frac{\nu}{\mu(n^*)} \tag{159}$$

Aggregate output is much simpler with constant returns $Y^*|_{\nu=1} = \frac{AF(K^*,L^*)}{\mu(n^*)}$. There is productive inefficiency from the markup, which reduces Y^*, but the fixed cost component $\left(\frac{\mu(n^*)\phi}{\mu(n^*)-\nu} \right)^{1-\frac{1}{\nu}} \nu$ is unimportant as all firms use the fixed cost with the equal efficiency. It is useful to compare the endogenous markup case, to the better-known case of fixed markups in steady state. With a fixed markup number of firms does not affect aggregate output through y^* which is always fixed exogenously as a function of given parameters, so an extra firm simply contributes this fixed extra amount to output. With endogenous markup an extra firm alters per firm output y^* since a firm needs to produce more to cover fixed costs due to fiercer markup competition. $\frac{\partial Y^*}{\partial n} = y^* + n^* y^*_n$. With fixed markups only the first effect is present (the contribution of an entrant is to add y^* to aggregate output), but with endogenous markup there is also the competition effect which raises output per firm of every incubent because they face more competition $n^* y^*_n$.

The conclusions from the static analysis are that output per firm increases with number of firms, and productivity increases with number of firms. These results arise because number of firms degrade monopoly power, and this effect will always prevail over the dynamic business reallocation effect that causes output per firm to decrease as each firm enters because any shock rises productivity and output per firm too much on impact. From the long-run perspective more firms is better in the sense it raises output per firm, so more aggregate output can be produced from fewer firms.

3.6 MAIN RESULT: PRODUCTIVITY DYNAMICS

The main result shows that the impact effect of a TFP shock causes an exacerbated response in short-run productivity that relinquishes over time but leaves some long-run persistence due to the competition effect. This means that the difference between measured productivity on impact and measured productivity in the long-run is dampened because of a persistent change in productivity draws it closer to the initial level.

Theorem 3.6.1 Permanent Change in TFP.
On impact of a shock productivity overshoots the long-run effect, but there is no reversion to underlying productivity due to a persistent change in degree of competition.

$$\mathcal{P}(0)_A - \mathcal{P}^*_A = \mathcal{P}(0)_A - \mathcal{P}^{*\bar{\mu}}_A - \mathcal{P}^*_\mu \mu(n^*)_A \tag{160}$$

$$\mathcal{P}(0)_A|_{x(0)=x^*} - \mathcal{P}^*_A =$$

$$(\mu^* - 1)\frac{\mathcal{P}^*}{\nu}\left[\underbrace{\left(\frac{1}{A} + \frac{F^*_L L^*_A}{F^*}\right)}_{\text{Alloc. Effect (+)}} + \underbrace{\frac{\nu}{\mu^*(\mu^* - \nu)}\mu_n n^*_A}_{\text{Comp. Effect (-)}}\right] \tag{161}$$

where $\mathcal{P}^* = \nu \left[\frac{A}{\mu(n^*)} \left(\frac{\mu(n^*) - \nu}{\phi} \right)^{1-\nu} \right]^{\frac{1}{\nu}}$

Proof. Details in appendix C.5. From lemma 3, equation (158), $n_A^* > 0$ improved TFP raises the number of firms, see also numerical result in section 3.6.2. □

where $\mathcal{P}(0)_A|_{x(0)=x^*}$ is the response of productivity on impact (at $t = 0$), with all variables x beginning at steady state $x(0) = x^*$.

The positive allocation effect captures that only incumbents bear the change in TFP due to entry/exit inertia thus on impact there is a direct effect on incumbents' productivity from having a different TFP and there is a reinforcing labour effect that also responds immediately. The negative competition effect captures that the long-run level of productivity moves in the same direction as the initial effect which closes the gap between initial impact and long-run productivity. In the absence of a competition effect $\mu_n = 0$, there is no persistent effect on productivity.

Corollary 8.

Entry reduces the size of productivity overshooting, such that as $n \to \infty$

$$\mathcal{P}(0)_A|_{x(0)=x^*} = \mathcal{P}_A^* \tag{162}$$

The result only exists with rising marginal cost $\nu \in (0, 1)$.

Proof. Since the markup disappears as firms increase $\lim_{n\to\infty}\mu(n_t) = 1$, see appendix C.7, then

$$\lim_{n\to\infty}\left(\mathcal{P}(0)_A|_{x(0)=x^*} - \mathcal{P}_A^*\right) = 0$$

□

This corollary can be interpreted as more competitive economies have less productivity volatility. The result implies that productivity puzzles are weakened when there is more competition. Conversly overshooting is greater when there are few firms per sector. This strengthens imperfect competition and therefore markups are higher. It also means that the long-run structural change to competition will be greater. Consequently, the impact of a technology shock causes a large change in measured productivity initially but it then reverts to a similar but weaker level of productivity in the long-run. Contrarily, if the sector is very competitive, there are many firms in the sector and the initial effect on productivity is small, likewise there is little structural change to competition from more firms entering because there are still many firms competing. The implication is that more competition, which is synonymous to more firms, implies less volatile productivity and less persistence in productivity shocks[24]. Importantly this result does not hold under constant returns to scale $\nu \to 1$ because as firms drive markup to unitary no equilibrium exists as there is no cost minimizing level of output.

24 This is an interesting testable implication to expand upon empirically. As number of firms in the economy gets large then the long-run effect arises immediately $\lim_{n\to\infty}\mu(n) = 1$ so $\mathcal{P}(0)_A = \mathcal{P}_A^*$

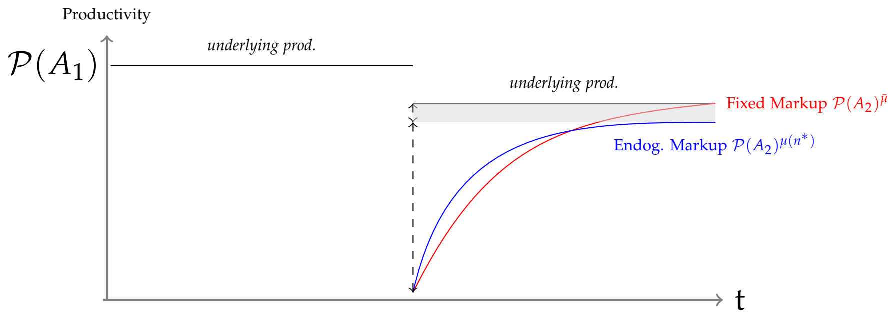


Figure 15.: Exacerbated Productivity Followed by Long-run Persistence

Figure 15 shows that the long-run competition effect tightens the gap between impact and long-run effect. A negative shock to technology from A_1 to A_2 causes an initially big fall in measured productivity (dashed arrow), but it recovers as firms begin to exit. However, the *"Fixed Markup $\mathcal{P}(A_2)^{\bar{\mu}}$"* curve shows that with fixed markups $\bar{\mu}$ productivity recovers to regain the underlying level that incorporates the new worse technology A_2, whereas the *"Endog. Markup $\mathcal{P}(A_2)^{\mu(n^*)}$"* time path shows that despite some recovery there is always persistently worse productivity in the long run (shown by the gray box), and this is because the markup $\mu(n^*)$ rises due to less long-run competition from firms exiting.

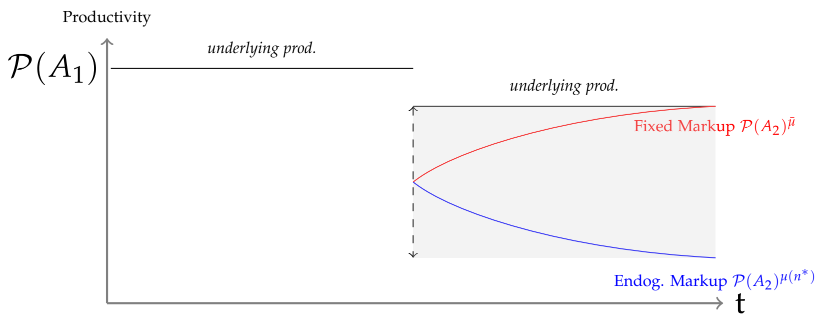


Figure 16.: Short-run Productivity Undershooting

3.6.1 *Competition Effect Strengthens Misallocation Effect*

A special case that may arise is if the negative competition effect is larger than the positive allocation effect. The previous discussion assumed that $\mathcal{P}(0)_A - \mathcal{P}_A^* > 0$. However, if the competition effect is large then $\mathcal{P}(0)_A - \mathcal{P}_A^* < 0$, so the initial movement in measured productivity is less than the long-run change in productivity. In terms of a positive shock to TFP this would mean an increase in measured productivity on impact as incumbent firms benefit from the improved technology, but then as firms begin to enter their negative effect of reallocating business is less than their positive effect reducing markups, so as they enter productivity continues to improve. If there is a negative TFP shock as in figure 15 the result is an initiall fall in productivity, followed by further worsening of productivity to a long-run level below the initial movement. After the initial fall in productivity the further worsening occurs as firms exit and weaker competition reduces productivity more than the reallocation of resources among incumbents.

3.6.2 *Supplementary Numerical Exercise*

The theory of the previous section demonstrates the main result of the chapter, but a numerical exercise is useful to gauge the two effects to gain an intuition for whether undershooting or overshooting in productivity arises. The baseline RBC model assumes isoelastic (constant elasticity) separable subutilties and a Cobb-Douglas production function.

3.6.2.1 *Utility*

$$U(C, L) = \frac{C^{1-\sigma} - 1}{1-\sigma} - \xi \frac{L^{1+\eta}}{1+\eta} \tag{163}$$

The derivatives are

$$U_C = C^{-\sigma}, \quad U_{CC} = -\sigma C^{-\sigma-1}, \quad U_L = -\xi L^{\eta} \tag{164}$$

The degree of relative risk aversion is constant $\sigma(C) = -C\frac{U_{CC}}{U_C} = \sigma$. Isoelastic utility implies there is constant elasticity of utility with respect to each good. $\sigma \neq 1$ is the constant coefficient of relative risk aversion. $\sigma \to \infty$ implies infinite risk aversion, so consumption has little effect on utility. η is Frisch elasticity of labour supply.

3.6.2.2 *Production*

$$F(k, l) = k^{\alpha} l^{\beta} = K^{\alpha} L^{\beta} n^{-(\alpha+\beta)} = F(K, L) n^{-(\alpha+\beta)} \tag{165}$$

Cobb-Douglas production conforms to our assumptions on the production function derivatives,

$$F_k = \alpha k^{\alpha-1} l^{\beta} = \alpha K^{\alpha-1} L^{\beta} n^{1-(\alpha+\beta)}, \tag{166}$$

$$F_l = k^{\alpha} \beta l^{\beta-1} = K^{\alpha} \beta L^{\beta-1} n^{1-(\alpha+\beta)} \tag{167}$$

and it is homogeneous of degree $\alpha + \beta$, so $\nu = \alpha + \beta$ in our general notation. α and β are capital and labour shares respectively. This implies increasing marginal costs if $\alpha + \beta < 1$. Thus the impact versus long-run effect of a change in technology becomes[25]

$$\mathcal{P}(0)_A|_{x(0)=x^*} - \mathcal{P}_A^* = \frac{\mathcal{P}^*}{\nu(n^*-1)}\left[\frac{2}{A} + \beta\frac{n_A^*}{n^*} - \frac{\nu}{\mu(n^*)\left(\mu(n^*)-\nu\right)}\frac{n_A^*}{n^{*2}}\right] \quad (168)$$

If we assume that intersector substitutability is $\theta_I = 1$ then the markup purely depends on number of firms $\mu(n) = \frac{n}{n-1}$

$$\mathcal{P}(0)_A|_{x(0)=x^*} - \mathcal{P}_A^* = \frac{\mathcal{P}^*}{(\alpha+\beta)(n^*-1)}\left[\frac{2}{A} + \left(\beta - \frac{(\alpha+\beta)}{\frac{n^{*2}}{n^*-1}\left(\frac{n^*}{n^*-1} - (\alpha+\beta)\right)}\right)\frac{n_A^*}{n^*}\right] \quad (169)$$

Given this parameterization a sufficient condition for undershooting $\mathcal{P}(0)_A - \mathcal{P}_A^* < 0$ is

$$\frac{2}{A} + \left(\beta - \frac{(\alpha+\beta)}{\frac{n^{*2}}{n^*-1}\left(\frac{n^*}{n^*-1} - (\alpha+\beta)\right)}\right)\frac{n_A^*}{n^*} < 0 \quad (170)$$

therefore a necessary condition is

$$\beta < \frac{(\alpha+\beta)}{\frac{n^{*2}}{n^*-1}\left(\frac{n^*}{n^*-1} - (\alpha+\beta)\right)} \quad (171)$$

where the right-hand side is the competition effect that arises from endogenous markups[26]. The competition effect is zero with exogenous (fixed) markups ($n^* \to 1$), therefore the nec-

25 Details in appendix C.6.
26 I label it the competition effect since it determines the magnitude of the negative component.

essary condition is always violated $\beta \nless 0$ and undershooting cannot arise. This formalizes the logic that if there is no persistent effect on productivity then undershooting of the long-run level cannot arise. Similarly, any effect that weakens the long-run persistent effect of competition (decreases $\mathcal{P}_\mu^*$) will reduce the likelihood of undershooting. For example, the competition effect is increasing in returns to scale $\nu := \alpha + \beta \to 1$, leading to undershooting.

In general equilibrium using previous parameter values (table 1), numerical simulation[27] makes the right-hand side of the necessary condition 0.07 where number of firms is determined endogenously in steady state as $n^* = 44.156$. Clearly the right-hand side value does not exceed the labour share $\beta = 0.5 \nless 0.07$, thus undershooting does not arise in general equilibrium. The conclusion is that the competition effect is small relative to the allocation effect, an important result given most literature focuses on endogenous markups (competition effect) rather than business allocation. So let us ask under what partial equilibrium conditions could undershooting arise, and interpret the plausibility of the economic narrative.

Figure 17 shows a calibration that strengthens the necessary condition by choosing close to CRTS with a low β. The negative region is quantitiatively small, so is unlikely to offset the initial effect required to meet the sufficient condition, unless n_A^* is unrealistically large. For example, we can see there is a region $n^* \in (3, 15)$ when the necessary condition is met, thus

27 This is a purely numerical calculation because the added nonlinearity of endogenous markups precludes an analytic derivation of n^*. The value of firms is high. The purpose here is intuition to aid the analytics.

take $n^* = 5$ and $\nu = \alpha + \beta = 0.6 + 0.2 = 0.8$ meets the necessary condition for negativity, so the sufficient condition, where assuming $A = 1$, is $2 - 2\frac{n_A^*}{5} < 0$. Therefore the condition is $n_A^* > 5 = n^*$, so there must be a 100% change in market size if the competition effect from entry is to exceed the initial misallocation effect.

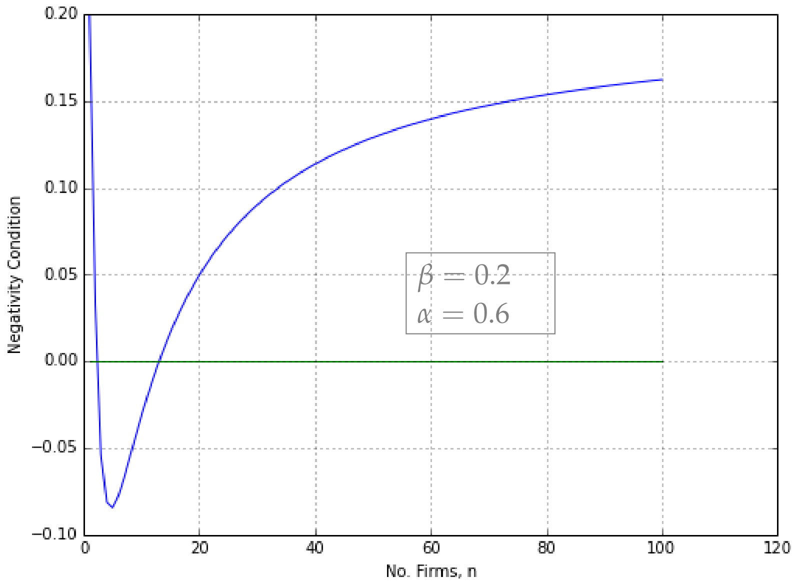


Figure 17.: Initial Overshooting Versus Undershooting

3.7 CHAPTER SUMMARY

The chapter investigates the effect of firm entry on measured productivity over the business cycle. I consider that entry is noninstantaneous and entry affects the price markups that incumbents charge. Together these mechanisms can explain short-run procylical productivity and weaker long-run persistence. Contemporary productivity puzzles provide a lens to view the theory through. In relation to productivity puzzles, the theory explains that productivity is exacerbated on impact, since firms cannot adjust immediately so incumbents bear shocks, and in the long run underlying productivity is not regained because subsequent adjustment of firms causes structural changes in competition. The structural changes in competition reflect that entry strengthens competition which improves productivity in the long run (inversely, exit weakens competition, decreases productivity). Furthermore I show that in highly competitive industries the distinction between short-run and long-run productivity is small, so measured productivity quickly and accurately reflects underlying productivity.

A growing number of DSGE papers show promising quantitative simulation results from adding a firm entry process. Despite these appealing data matching properties, little research has reduced models to minimal state variables to understand the analytical effect of entry. This chapter allows economists to understand how the entry variables interact with the model in a general setup before specifying functional forms or numerical calibrations. We learn that firm entry dynamics can explain short-run dynamic changes in productivity over the business cycle and long-run static changes that persist. The two explanations arise from two different effects of entry, a dynamic reallocation effect that redistributes resources as firms adjust and a static competition effect that alters firms' pricing markup decisions in response to competition from entry. A simple quantitative exercise emphasizes the dominance of the allocation effect over the competition effect, which is an important lesson for researchers who have tended to focus on firm dynamics' effects on markups rather than allocation.

4

LESSONS FROM FIRM DYNAMICS IN THE MACROECONOMY

The thesis argues that firm dynamics (firm entry and exit) are an important determinant of business cycle behaviour. Specifically the changing number of firms that comprise industries in the economy is an important determinant of short-run and long-run measured productivity through its effect on competition and allocation of resources per firm. The thesis is crudely summarized by the following statement

> *"Given the resources in an economy, how they are divided among firms, and how these firms compete with each other, helps us to understand how productively the resources are used."*

4.1 SUMMARY OF RESULTS

An example economic scenario best illustrates the results of the thesis. Chapter 2 develops the short-run mechanisms and chapter 3 develops the long-run mechanisms. Consider a negative shock to the economy, like the 2007 financial crisis, it initially impacts incumbent firms causing amplified productivity falls, but these falls relinquish as firms begin to exit–chapter 2 establishes this. However, there is an opposing effect that exit of firms reduces competitive pressures on those who remain which prevents full reversion of productivity, so there is a persistent productivity fall–chapter 3 establishes this.

Chapter 2 explains that firm entry teamed with imperfect competition causes excessive entry because the monopoly profits that arise under imperfect competition strengthen incentives to enter. The result of excess entry is that incumbents are inefficiently small. They operate with excess capacity: each incumbent pays a fixed cost that it underutilizes. In the short run, incumbents vary their capacity utilization in response to shocks without other firms instantaneously responding. The result is short-run productivity gains as firms utilize fixed costs. However, as firm entry adjusts, business reallocation overcomes the short-run effect. Unlike recent research, this theory has fixed markups (no competition effect); dynamic entry alone is sufficient for endogenous productivity dynamics. **The main finding is** that productivity shocks are amplified on impact due to inert firms, but relinquish in the long run as entry/exit responds.

Despite the promising findings for short-run productivity movements, this research cannot explain persistent effects to productivity after shocks. This is particularly notable in times when policy makers are dealing with so-called productivity puzzles: the persistently low productivity post Great Recession. Chapter 3 attempts to develop a theory that can help to understand these puzzles.

Chapter 3 studies how firm entry determines macroeconomic productivity through division of resources and competitive pressure on markups. Recent research examines markup behaviour under firm entry and its importance for macroeconomic productivity, but the arguments focus on instantaneous firm entry. I argue that this omits a mechanism: the changing allocation of business as firms adjust. My contribution is to combine this dynamic firm entry with competitive markups. The result is a new trade-off: an exiter reduces industry competition which reduces incumbents' productivity, but exit reallocates business among incumbents which improves productivity through returns to scale (vice-versa for entry which raises competition, but steals business). The mechanism helps to clarify 'productivity puzzles'. It explains that economic shocks cause exacerbated productivity responses that weaken as firms adjust, and entry competition prevents reversion to trend productivity. **The main finding is** that competition from firm entry ameliorates short-run productivity volatility but in the long run productivity effects persist because of structural changes to competition.

The mutual argument that the thesis emphasizes is that sunk costs cause firms to respond slowly to economic shocks, hence entry and exit decisions are non-instantaneous. This slow response of firm dynamics to economic conditions teamed with imperfect competition has been observed in quantitative DSGE models of firm entry. These papers observe particular simulations, whereas this work offers an analytical narrative to the 'blackbox' driving simulated dynamics. Therefore I am able to precisely describe the mechanics underlying productivity movements. Throughout the thesis we can identify that the returns to scale parameter and degree of competition can alter and nullify productivity effects.

A clear prescription for applied researchers is that in order to understand macroeconomic productivity notice should be taken of firm behaviour, and this may mean that instantaneous figures for productivity are misleading as firm dynamics take time to adjust.

4.2 FUTURE WORK

The thesis opens several new lines of research. The first is a natural development of the current work to consider its stochastic analog. The second places more emphasis on empirical work and the testable implications of the model, which opens policy relevant questions like the effect of fiscal stimulus. The third explores the mathematical theory of the thesis more deeply to gain a better understanding of imperfect competition's effect on the system's eigenvalues.

4.2.1 *Stochastics and Heterogeneity*

All the discussion in this thesis focuses on once-and-for-all shocks. A relatively simple extension to the work is to consider a stochastic model with a shock process in technology that reverts to its initial level. This research's logic would carry over, and relate to the 'amplification' and 'propagation' keywords that tend to be associated with stochastic business cycle models (Kocherlakota 2000). That is, in a traditional impulse response type analysis, we would expect to observe amplification in productivity on impact followed by a propagating effect due to competition eventually returning to the same initial equilibrium. This would make the work directly comparable with other papers that simulate responses like this, but with the advantage of the tight theoretical description this thesis develops. Another question related to stochastics answers the *FAQ* "Where is the heterogeneity?" that emphasizes the importance of selection effects when discussing firm entry. In international trade it is the status quo that different levels of productivity determine firm survival, entry and exit. The framework I have developed could be extended to recognize that the sunk costs of entry are non-symmetric. For example, a stochastic component in the sunk cost could reflect that exit is less costly than entry. This would strengthen the well-known asymmetries across entry and exit rates, and could address the seemingly heteroskedastic relationship between NBF and output. That is that recession tends to lead to a relatively greater falls in net business formation (NBF) than booms lead to rises in NBF (figure 2).

4.2.2 *Productivity Puzzles and Fiscal Policy*

A benefit of using industrial organization (IO) in macroeconomics is that it opens a trove of firm level IO data to a field reliant on less-meticulous aggregate data. The thesis gives rise to several testable implications. Among them are that productivity effects are weaker in more competitive situations characterized by economies with a large number of firms. Do more competitive economies face less volatile productivity movements? Chapter 3 refers to how the theory could explain productivity puzzles since the Great Recession. As Haskel, Goodridge, and Wallis 2015 evidence, these puzzles of a fall in productivity followed by persistently low levels are profound for UK data. My theory implies that exit leading to market consolidation reduces competitive pressure on incumbents (raises markups) leading to an opening of their capacity underutilization causing a fall in productivity. Loose evidence for market consolidation from exit (and more broadly mergers and acquisitions) in the UK exists the test is to link this to a rise in excess capacity, a variable that could be measured using the population of UK firm level data contained in the *inter departmental business register (IDBR)*. Lastly, this thesis focuses on technology shocks, but a question for policy makers is whether fiscal policy shocks have a similar effect on productivity. If the results of Aloi and Dixon 2003 carry over to the closed economy with capital setting, then initial intuition suggests an increase in government expenditure will stimulate entry and increase input marginal products, thus improving productivity.

4.2.3 *Theory and Speed of Convergence*

Reducing the four-dimensional economic system with entry to a two-dimensional stable manifold allows for analytical solutions for the model within a neighborhood of the hyperbolic fixed point. Initial results that arise from analytical solutions show that imperfect competition reduces the set of complex dynamics, and raises eigenvalues which hastens convergence to steady state. Importantly there is economic intuition to explain the faster convergence, that is that imperfect competition raises firm profits, so an entrant reduces industry profits more thus arbitrage to zero profits quickens.

4.3 closing remark

This section has shown that the results of the thesis open many new questions for future study into *firm dynamics and the macroeconomy*. As the ideas of this field become more embedded into status quo models, an exciting prospect is the natural alignment of industrial organization, finance and macroeconomics. Recent work in macroeconomics has sought to improve understanding of financial frictions, and independently other work, like this thesis, has been improving macro-IO relationships. It seems promising for future researchers to link these two bodies of research: understanding how financially constrained firms alter their strategic behaviour, which impacts macroeconomic behaviour and stock market value would be an eloquent development for the discipline.

A

CHAPTER 1 APPENDIX

A.1 DATA DESCRIPTION

FRED mnemonics are in square brackets.

US GDP Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate. US. Bureau of Economic Analysis, Gross Domestic Product [GDP], retrieved from FRED, Federal Reserve Bank of St. Louis https://research.stlouisfed.org /fred2/series/GDP/, January 4, 2016.

US TFP Annual, Not Seasonally Adjusted. University of Groningen and University of California, Davis, Total Factor Productivity at Constant National Prices for United States [RTFPNAUSA632NRUG], retrieved from FRED, Federal Reserve Bank of St. Louis https://research.stlouisfed.org /fred2/series/RTFPNAUSA632NRUG/, January 4, 2016.

A.1.1 *Net Business Formation (NBF) and New Incorporations (NI)*

Released by the Bureau of Economic Analysis, NBF includes NI but also accounts for failures. NBF was assembled in-house by the BEA, but the NI data it relies upon is from Dunn and Bradstreet consultancy. NBF is an index, 1967=100 base year. NI is absolute numbers. Both are monthly and seasonally adjusted.

Survey of Current Business, Volume 76, (1996) has net business formation (NBF) and New Incorporations (NI) data from January 1948 to December 1994. Specifically it is section C-8 "Historical Data for Selected Series" of the survey, and is available in a clearer PDF via the Bureau of Economic Analysis's (BEA) online archive (I include the primary source in this Chapter's appendix). An additional 9 months up to September 1995 are available in the BEA's Survey of Current Business, Nov-Dec 1995 release, section C-2 (I include the primary source in this Chapter's appendix). Section C-1 The New Incorporations series (NI) was collected by Dun and Bradstreet (a private consultancy) over the entire period, and proprietary versions extend for an additional three years to 1998. I do not use all these proprietary years, which are available via the company *The Conference Board*. However, it can be freely extended by 21 months to September 1996 using the Economagic data repository.

The Survey of Current Business describes New Business Incorporations as

> "V 21, new business incorporations. This series represents the total number of stock corporations is-

sued charters under the general business corporation laws of the various States and the District of Columbia. The statistics include completely new businesses that have incorporated, existing businesses changed from the noncorporate to the corporate form of organization, existing corporations given certificates of authority to operate also in another State, and existing corporations transferred to a new State. Data for incorporations in the District of Columbia are included beginning January 1963."

and Net Business Formation as

"V 22, index of net business formation. This series is compiled from monthly national data on number of new business incorporations, number of business failures, and confidential data on telephones installed. These components are adjusted for seasonal variation and number of trading days before being combined into the index. "

B

CHAPTER 2 APPENDIX

B.1 HOUSEHOLD OPTIMIZATION PROBLEM

To obtain the necessary conditions for a solution to the household's utility maximisation problem I use the Maximum Principle. The current value Hamiltonian is

$$\hat{\mathcal{H}}(t) = u(C(t), L(t)) + \lambda(t)(w(t)L(t) + r(t)K(t) + \Pi(t) - C(t) - G) \quad (172)$$

The costate variable λ_t is the shadow price of wealth in utility units. The Pontryagin necessary conditions are [1]

$$\hat{\mathcal{H}}_C(K, L, C, \lambda) = 0 \implies u_C - \lambda = 0 \quad (173)$$

$$\hat{\mathcal{H}}_L(K, L, C, \lambda) = 0 \implies u_L + \lambda w = 0 \quad (174)$$

$$\hat{\mathcal{H}}_K(K, L, C, \lambda) = \rho\lambda - \dot{\lambda} \implies \lambda r = \rho\lambda - \dot{\lambda}$$

$$\implies \frac{\dot{\lambda}}{\lambda} = -(r - \rho) \quad (175)$$

$$\hat{\mathcal{H}}_\lambda := \dot{K}_t \implies \dot{K} = rK + wL + \Pi - C \quad (176)$$

1

Equations (173)-(175) reduce to two equations. These are a dynamic differential equation called the Consumption Euler equation (intertemporal condition) and a static injective mapping between labour and consumption (intratemporal condition).

B.2 costs and operating profit

Total costs are wages paid on labour and interest on capital

$$
\begin{aligned}
wl + rk &= AF_l(1-\zeta)l + AF_k(1-\zeta)k = A(1-\zeta)[F_l l + F_k k] \\
&= (1-\zeta)\nu AF(k,l) \qquad (177)
\end{aligned}
$$

Operating profit is output less total costs

$$
\pi = y - wl - rk = (AF(k,l) - \phi) - ((1-\zeta)\nu AF(k,l)) \qquad (178)
$$

$$
\pi(L, K, n; A, \zeta, \phi) = AF(k,l)(1-(1-\zeta)\nu) - \phi \qquad (179)
$$

B.3 optimal labour derivatives

Partially differentiate the intratemporal Euler with respect to each variable treating labour as an implicit function. Then by implicit function theorem get labour responses $L(\underset{-}{C}, \underset{+}{K}, \underset{+}{n}; \underset{+}{A}, \underset{-}{\zeta})$

$$
u_L + u_C(1-\zeta)AF_l\left(\frac{K}{n}, \frac{L}{n}\right) = 0 \qquad (180)
$$

Assumptions:

$$F_{ll}(\frac{K}{n},\frac{L}{n}), u_{CC}(C), u_{LL}(L) < 0 \tag{181}$$

$$u_C(C), F_l(\frac{K}{n},\frac{L}{n}), F_{lk}(\frac{K}{n},\frac{L}{n}) > 0 \tag{182}$$

where $l = \frac{L}{n}$ and $k = \frac{K}{n}$

$$u_{LL}L_C + u_{CC}(1-\zeta)AF_l + u_C(1-\zeta)AF_{ll}\frac{L_C}{n} = 0, \tag{183}$$

$$L_C = \frac{-u_{CC}(1-\zeta)AF_l}{u_{LL} + u_C(1-\zeta)A\frac{F_{ll}}{n}} < 0 \tag{184}$$

$$u_{LL}L_K + u_C(1-\zeta)AF_{lk}\frac{1}{n} + u_C(1-\zeta)AF_{ll}\frac{L_K}{n} = 0, \tag{185}$$

$$L_K = \frac{-u_C(1-\zeta)A\frac{F_{lk}}{n}}{u_{LL} + u_C(1-\zeta)A\frac{F_{ll}}{n}} > 0 \tag{186}$$

$$u_{LL}L_n + u_C(1-\zeta)A\left[F_{lk}\frac{-K}{n^2} + F_{ll}\left(\frac{-L}{n^2} + \frac{L_n}{n}\right)\right] = 0, \tag{187}$$

$$L_n = \frac{u_C(1-\zeta)A(\nu-1)\frac{F_l}{n}}{u_{LL} + u_C(1-\zeta)A\frac{F_{ll}}{n}} > 0 \tag{188}$$

$$u_{LL}L_A + u_C(1-\zeta)F_l + u_C(1-\zeta)AF_{ll}\frac{L_A}{n} = 0, \tag{189}$$

Assumptions on the functions, given above, are sufficient to determine the signs in all cases except L_n, which depends on returns to scale of the technology ν.

The denominator is the same in each derivation. It is the intratemporal condition differentiated with respect to L, and it is negative. The negativity reflects that utility is decreasing in labour. Therefore the numerator distinguishes signs. The numerator is the derivative of the intratemporal condition with respect to the variable of interest.

B.4 COMPARATIVE STATICS

From (54) and (55)

$$F_k(k,l) = \Upsilon, \quad F(k,l) = \Xi \tag{190}$$

$$\text{where, } \Upsilon = \frac{\rho}{A\,(1-\zeta)} \quad \Xi = \frac{\phi}{A\,(1-(1-\zeta)\,\nu)} \tag{191}$$

Use Cramer's rule to determine the effect of a change in ζ. Differentiate with respect to ζ

$$F_{kk}k_\zeta + F_{kl}l_\zeta = \Upsilon_\zeta \tag{192}$$

$$F_k k_\zeta + F_l l_\zeta = \Xi_\zeta \tag{193}$$

$$\overbrace{\begin{bmatrix} F_{kk} & F_{kl} \\ F_k & F_l \end{bmatrix}}^{\mathbf{A}} \begin{bmatrix} k_\zeta \\ l_\zeta \end{bmatrix} = \begin{bmatrix} \Upsilon_\zeta \\ \Xi_\zeta \end{bmatrix} \tag{194}$$

$$\begin{bmatrix} k_\zeta \\ l_\zeta \end{bmatrix} = \underbrace{\frac{1}{det(\mathbf{A})} \begin{bmatrix} F_l & -F_{kl} \\ -F_k & F_{kk} \end{bmatrix}}_{\mathbf{A}^{-1}} \begin{bmatrix} \Upsilon_\zeta \\ \Xi_\zeta \end{bmatrix} \tag{195}$$

$$det(\mathbf{A}) = F_{kk}F_l - F_{kl}F_k < 0, \tag{196}$$

$$\Upsilon_\zeta = \frac{\rho}{A(1-\zeta)^2} > 0 \tag{197}$$

$$\Xi_\zeta = -\frac{\phi}{A(1-(1-\zeta)\nu)^2\nu} < 0 \tag{198}$$

Hence,

$$k_\zeta = \frac{1}{det(\mathbf{J})}\left(F_l \Upsilon_\zeta - F_{kl}\Xi_\zeta\right) < 0 \tag{199}$$

$$l_\zeta = \frac{1}{det(\mathbf{J})}\left(-F_k \Upsilon_\zeta + F_{kk}\Xi_\zeta\right) \gtrless 0$$

$$\text{if} \quad \left|\frac{F_k}{F_{kk}}\right| \lesseqgtr \left|-\frac{\phi(1-\zeta)^2}{\rho(1-(1-\zeta)\nu)^2\nu}\right| \tag{200}$$

B.4.1 *Numerical Bifurcation Exercise*

Figure 18 repeats the figures in the paper but with ζ on the z-axis.

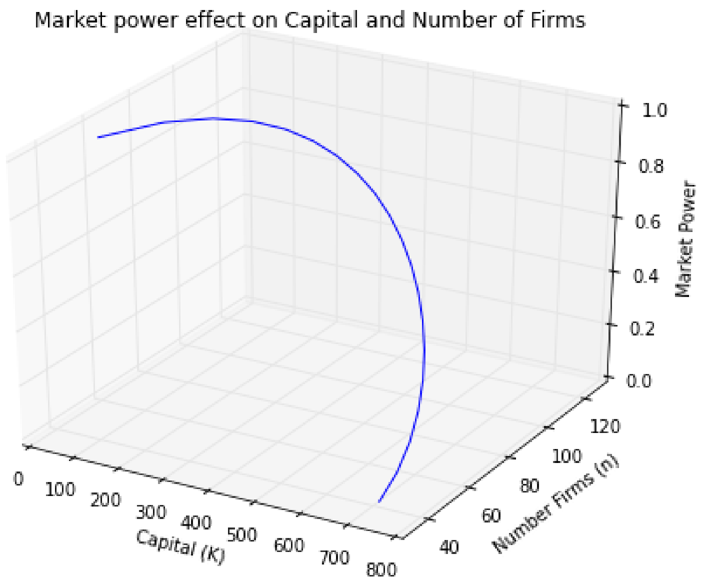


Figure 18.: 3D Projection of Capital and No. Firms Equilibrium as Market Power Changes

B.5 ELEMENTS OF THE JACOBIAN: INTEREST RATE, PROFIT, OUTPUT

Given optimal labour choice we can evaluate how interest rate, profit and output respond.

B.5.1 *Output*

$$Y(L,K,n;A,\phi) = n\left[A\mathrm{F}\left(\frac{K}{n},\frac{L}{n}\right) - \phi\right] \tag{201}$$

$$Y_C = AF_l L_C < 0 \tag{202}$$

$$Y_K = A(F_k + F_l L_K) > 0 \tag{203}$$

$$Y_n = (1-\nu)AF - \phi + AF_l L_n \lesseqgtr 0 \tag{204}$$

Furthermore in steady state when $F(\frac{K}{n},\frac{L}{n})^* = \frac{\phi}{A(1-(1-\zeta)\nu)}$ then $Y_n^* = \frac{-\phi\zeta\nu}{1-(1-\zeta)\nu} + AF_l L_n$ which is positive or negative depending whether the negative component outweighs the positive labour effect.

B.5.2 *Rents*

$$r = (1-\zeta)AF_k \tag{205}$$

$$r_C = (1-\zeta)\frac{A}{n}F_{kl}L_C < 0 \tag{206}$$

$$r_K = (1-\zeta)\frac{A}{n}[F_{kk} + F_{kl}L_K] < 0 \tag{207}$$

$$r_n = (1-\zeta)\frac{A}{n}[(1-\nu)F_k + F_{kl}L_n] > 0 \tag{208}$$

Both r_K and π_n require extra work to derive the signs. They may be found by substituting in L_K and L_n respectively, from which they can be rearranged into a form including $F_{KK}F_{LL} - F_{KL}^2$ which by assumption is greater than zero. The assumption is that it satisfies the second partial derivative test for concavity.

B.5.3 *Profit*

$$\pi = AF(k,l)(1-(1-\zeta)\nu) - \phi \tag{209}$$

$$\pi_C = AF_l\frac{L_C}{n}(1-(1-\zeta)\nu) < 0 \tag{210}$$

$$\pi_K = \frac{A}{n}(F_k + F_l L_K)(1-(1-\zeta)\nu) > 0 \tag{211}$$

$$\pi_n = \frac{A}{n}(-\nu F + F_l L_n)(1-(1-\zeta)\nu) < 0 \tag{212}$$

Notice that for an k,l profit is higher so for any given K,L,n profit is higher, but not necessarily for any given K,n which is why in state space fewer firms can arise–which explains the confusion over why despite higher profits we can have a reduction in firms; the answer is higher profits rely on given L.

B.6 JACOBIAN

The Jacobian before evaluating at steady state is

$$\begin{bmatrix} \dot{C}_C & \dot{C}_e & \dot{C}_K & \dot{C}_n \\ \dot{e}_C & \dot{e}_e & \dot{e}_K & \dot{e}_n \\ \dot{K}_C & \dot{K}_e & \dot{K}_K & \dot{K}_n \\ \dot{n}_C & \dot{n}_e & \dot{n}_K & \dot{n}_n \end{bmatrix} \tag{213}$$

$$\begin{bmatrix} \frac{C^*}{\sigma}(1-\zeta)\frac{A}{n}F_{kl}L_C & 0 & \frac{C^*}{\sigma}(1-\zeta)\frac{A}{n}[F_{kk}+F_{kl}L_K] & \frac{C^*}{\sigma}(1-\zeta)\frac{A}{n}[(1-\nu)F_k+F_{kl}L_n] \\ (1-\zeta)\frac{A}{n}F_{kl}L_Ce-\frac{AF_lL_C(1-(1-\zeta)\nu)}{\gamma n} & (1-\zeta)AF_k & (1-\zeta)\frac{A}{n}[F_{kk}+F_{kl}L_K]e-\Xi & (1-\zeta)\frac{A}{n}[(1-\nu)F_k+F_{kl}L_n]e-\Psi \\ AF_lL_C-1 & -\gamma e & A(F_k+F_lL_K) & (1-\nu)AF-\phi+AF_lL_n \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{214}$$

where $\Xi = \frac{A(F_k+F_lL_K)(1-(1-\zeta)\nu)}{n\gamma}$ and $\Psi = \frac{A(-\nu F+F_lL_n)(1-(1-\zeta)\nu)}{n\gamma}$. Then in steady state we have $e = 0$

$$\begin{bmatrix} \frac{-u_C}{u_{CC}}(1-\zeta)\frac{A}{n}F_{kl}L_C & 0 & \frac{-u_C}{u_{CC}}(1-\zeta)\frac{A}{n}[F_{kk}+F_{kl}L_K] & \frac{-u_C}{u_{CC}}(1-\zeta)\frac{A}{n}[(1-\nu)F_k+F_{kl}L_n] \\ -\frac{AF_lL_C(1-(1-\zeta)\nu)}{\gamma n} & (1-\zeta)AF_k & -\frac{A(F_k+F_lL_K)(1-(1-\zeta)\nu)}{n\gamma} & -\frac{A(-\nu F+F_lL_n)(1-(1-\zeta)\nu)}{n\gamma} \\ AF_lL_C-1 & 0 & A(F_k+F_lL_K) & (1-\nu)AF-\phi+AF_lL_n \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{215}$$

in steady state we also have $F_k = \frac{\rho}{A(1-\zeta)}$, and $F = \frac{\phi}{A(1-(1-\zeta)\nu)}$

Giving the unparameterized Jacobian in steady state in terms of K^* and n^* (where function domain is in per firm form (e.g. F_x is $F_x(k,l)$) and $\nu = \alpha + \beta$) as

$$\begin{bmatrix} \frac{C^*}{n^*\sigma}(1-\zeta)AF_{kl}L_C & 0 & \frac{C^*}{n^*\sigma}(1-\zeta)A(F_{kk}+F_{kl}L_K) & \frac{C^*}{n^*\sigma}(1-\zeta)\left[(1-\nu)\frac{\rho}{1-\zeta}+AF_{kl}L_n\right] \\ -\frac{(1-(1-\zeta)\nu)}{\gamma n^*}AF_lL_C & \rho & -\frac{(1-(1-\zeta)\nu)}{\gamma n^*}\left(\frac{\rho}{1-\zeta}+AF_lL_K\right) & \frac{1}{\gamma n^*}(\nu\phi-(1-(1-\zeta)\nu)AF_lL_n) \\ AF_lL_C-1 & 0 & \frac{\rho}{1-\zeta}+AF_lL_K & \frac{-\zeta\nu\phi}{1-(1-\zeta)\nu}+AF_lL_n \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{216}$$

The signs of the Jacobian may be determined as

$$\begin{bmatrix} - & 0 & - & + \\ + & + & - & + \\ - & 0 & + & \pm \\ 0 & + & 0 & 0 \end{bmatrix} \tag{217}$$

Under the parameterized model, the elements of the Jacobian can be determined entirely in terms of model parameters $\Omega := \{\alpha, \beta, \phi, \gamma, \xi, \rho, \eta, \zeta, \sigma\}$. We have

$$\frac{C^*}{n^*} = \frac{\phi(1-\zeta)\nu}{1-(1-\zeta)\nu}, \tag{218}$$

$$F_l = k^{*\alpha}\beta l^{*\beta-1}, \tag{219}$$

$$F_{kl} = \alpha k^{*\alpha-1}\beta l^{*\beta-1}, \tag{220}$$

$$F_{kk} = \alpha(\alpha-1)k^{*\alpha-2}l^{*\beta}, \tag{221}$$

$$L_C^* = -\frac{\sigma(1-(1-\zeta)\nu)}{(1+\eta-\beta)(\phi(1-\zeta)\nu)}l^*, \tag{222}$$

$$L_n^* = \frac{1-(\alpha+\beta)}{(1+\eta-\beta)}l^*, \tag{223}$$

$$L_K^* = \frac{\alpha}{(1+\eta-\beta)}\frac{l^*}{k^*}, \tag{224}$$

This leads to the 4×4 Jacobian where $J_{r,c}$ is element row r column c.

$$J_{1,1} = -\frac{\rho\beta}{1+\eta-\beta} \tag{225}$$

$$J_{1,2} = 0 \tag{226}$$

$$\begin{aligned} J_{1,3} &= \frac{\rho^2\nu((1+\eta)(\alpha-1)+\beta)}{\sigma\alpha(1+\eta-\beta)} \\ &= \frac{\rho^2\nu(\nu-1+\eta(\alpha-1)}{\sigma\alpha(1+\eta-\beta)} \end{aligned} \tag{227}$$

$$J_{1,4} = \frac{\phi(1-\zeta)\nu\rho(1-\nu)(1+\eta)}{(1-(1-\zeta)\nu)\sigma(1+\eta-\beta)} \tag{228}$$

$$J_{2,1} = \frac{(1-(1-\zeta)\nu)\beta\sigma}{\gamma n^*(1+\eta-\beta)(1-\zeta)\nu} \tag{229}$$

$$J_{2,2} = \rho \tag{230}$$

$$J_{2,3} = \frac{-(1-(1-\zeta)\nu)\rho(1+\eta)}{\gamma n^*(1-\zeta)(1+\eta-\beta)} \tag{231}$$

$$J_{2,4} = \frac{\phi(\nu(1+\eta)-\beta)}{\gamma n^*(1+\eta-\beta)} = \frac{\phi(\alpha+\nu\eta)}{\gamma n^*(1+\eta-\beta)} \tag{232}$$

$$J_{3,1} = \frac{-\beta\sigma}{(1+\eta-\beta)(1-\zeta)\nu} - 1 \tag{233}$$

$$J_{3,2} = 0 \tag{234}$$

$$J_{3,3} = \frac{\rho(1+\eta)}{(1-\zeta)(1+\eta-\beta)} \tag{235}$$

$$J_{3,4} = \frac{\phi[-\zeta\nu(1+\eta-\beta)+\beta(1-\nu)]}{(1-(1-\zeta)\nu)(1+\eta-\beta)} \tag{236}$$

$$J_{4,1} = 0 \tag{237}$$

$$J_{4,2} = 1 \tag{238}$$

$$J_{4,3} = 0 \tag{239}$$

$$J_{4,4} = 0 \tag{240}$$

In the second row I have not substituted out the n^* component which is in terms of model parameters because it does not simplify nicely.

$$J_{2,1} = \left\{ \left[\frac{A}{\phi^{1-\alpha}} \left(\frac{\alpha}{\rho}\right)^{\alpha} (1-(1-\zeta)\nu)^{1-\nu} \right]^{\frac{1+\eta}{\beta}} \frac{\phi^{1-\sigma}\nu^{\eta}(1-\zeta)^{(1+\eta)(1+\frac{\alpha}{\beta})}}{\xi\beta^{\eta+\sigma+1}} \right\}^{-\frac{1}{\eta+\sigma}} \frac{\sigma}{\gamma(1+\eta-\beta)} \tag{241}$$

$$J_{2.3} = \frac{-1}{\gamma} \left[\frac{\beta}{\xi\nu^{\sigma}} \left(\frac{A}{\sigma^{1-\alpha}} \alpha^{\alpha} (1-(1-\zeta)\nu)^{1-\nu} \right)^{\frac{1+\eta}{\beta}} (\phi(1-\zeta))^{1-\sigma} \left(\frac{\rho}{1-\zeta}\right)^{\frac{\beta(\eta+\sigma)+\alpha(1+\eta)}{\beta}} \right]^{\frac{1}{\eta+\sigma}} \frac{1+\eta}{1+\eta-\beta} \tag{242}$$

$$\begin{bmatrix} -\frac{\rho\beta}{1+\eta-\beta} & 0 & \frac{\rho^2\nu((1+\eta)(\alpha-1)+\beta)}{\sigma\alpha(1+\eta-\beta)} & \frac{\phi(1-\zeta)\nu\rho(1-\nu)(1+\eta)}{(1-(1-\zeta)\nu)\sigma(1+\eta-\beta)} \\ \frac{(1-(1-\zeta)\nu)\beta\sigma}{\gamma n^*(1+\eta-\beta)(1-\zeta)\nu} & \rho & \frac{-(1-(1-\zeta)\nu)\rho(1+\eta)}{\gamma n^*(1-\zeta)(1+\eta-\beta)} & \frac{\phi(\nu(1+\eta)-\beta)}{\gamma n^*(1+\eta-\beta)} \\ \frac{-\beta\sigma-(1+\eta-\beta)(1-\zeta)\nu}{(1+\eta-\beta)(1-\zeta)\nu} & 0 & \frac{\rho(1+\eta)}{(1-\zeta)(1+\eta-\beta)} & \frac{\phi[-\zeta\nu(1+\eta-\beta)+\beta(1-\nu)]}{(1-(1-\zeta)\nu)(1+\eta-\beta)} \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{243}$$

B.7 EIGENVALUES AND EIGENVECTORS

To solve for the eigenvalues of the 4-dimensional system, we exploit symmetry that allows us to solve the quartic characteristic polynomial analytically. The process follows Brito and Dixon 2013, where the Jacobian matrix is analytically similar with the same economic interpretations as here. In the presence of imperfect competition is provides a good approximation of dynamics, especially for less extreme values of market power, as supported by numerical simulation figure 12. Eigenvalues have the general structure of Feichtinger, Novak, and Wirl 1994.

The characteristic polynomial of the Jacobian takes a standard form. And the solution of the quartic polynomial is symmetric around $\frac{\rho}{2}$, and non-zero since $\frac{\rho}{2} > 0$.

Proposition 8 Eigenvalues.

There are four eigenvalues

$$\lambda_{1,2}^{s,u} = \frac{\rho}{2} \mp \left[\left(\frac{\rho}{2}\right)^2 - \frac{\mathcal{T}}{2} \mp \underbrace{\left(\left(\frac{\mathcal{T}}{2}\right)^2 - |\mathbf{J}| \right)^{\frac{1}{2}}}_{\Delta} \right]^{\frac{1}{2}} \tag{244}$$

$$\lambda_{1,2}^{s,u} = \frac{\rho}{2} \pm \left\{ \left(\frac{\rho}{2}\right)^2 - \frac{(M_2 - \rho^2)}{2} \pm \frac{1}{2} \left[(M_2 - \rho^2)^2 - 4|\mathbf{J}| \right]^{\frac{1}{2}} \right\}^{\frac{1}{2}} \tag{245}$$

$\mathcal{T} = M_2 - \rho^2$ where M_2 is the sum of the principal minors of order 2 of $\mathbf{J}$, and $|J| = M_4$ is the Jacobian determinant which is the sum of principal minors of order 4 (i.e. the determinant).

$$M_2 = \frac{\pi_n^*}{\gamma} + \rho Y_K^* + \frac{C^*}{\sigma} \left[r_C^*(Y_K^* + \rho) - r_K^*(Y_C^* - 1) \right] \tag{246}$$

$$M_4 = \frac{C^*}{\sigma\gamma} \left[(1 - Y_C^*)(r_K^* \pi_n^* - r_n^* \pi_K^*) + Y_K^*(\pi_n^* r_C^* - \pi_C^* r_n^*) - Y_n^*(r_C^* \pi_K^* - r_K^* \pi_C^*) \right] \tag{247}$$

So the eigenvalues are entirely in terms of parameters since at steady state the variables are in terms of parameters.

Proposition 9 Determinacy and saddlepath equilibrium.

There are always two negative eigenvalues, either real or complex, so *the system is determinate: we can choose state variables freely. The result arises because* $\inf\left\{\left[\left(\frac{\rho}{2}\right)^2 - \frac{\mathcal{T}}{2} \mp \Delta^{\frac{1}{2}}\right]^{\frac{1}{2}}\right\} = \frac{\rho}{2}$*, so the greatest lower bound of the open interval called the outer discriminant (the term in square brackets) is* $\frac{\rho}{2}$ *which will always be some* ϵ *greater/less than 0 when* $\pm$*. Denoting the inner discriminant (inner square root) as* Δ*. The proof follows from* $-\frac{\mathcal{T}}{2} \pm \Delta^{\frac{1}{2}} > 0$*, since* $\mathcal{T} < 0$*.*

The proposition also implies that no eigenvalue has zero real part (system is determinate), therefore by Grobman-Hartman within a neighbourhood of such a fixed point the linearized system is topologically equivalent to the nonlinear system. This verifies that the inference we make for the linearized system holds locally for the nonlinear system.

From the analytic eigenvalues we can derive corresponding eigenvectors and therefore the general solution of the system.

B.7.1 *Eigenvectors*

To calculate the four eigenvectors solve $(\mathbf{J} - \lambda^i_j\mathbf{I})\mathbf{P}^i_j = 0$ for $\mathbf{P}^i_j$, where there are four separate cases to solve for and hence four eigenvectors, since there are four eigenvalues two stable $\mathbf{P}^S_1, \mathbf{P}^S_2$, two unstable $\mathbf{P}^U_1, \mathbf{P}^U_2$. Since the eigenvalues are chosen such that $|\mathbf{J} - \lambda\mathbf{I}| = 0$ and a zero determinant means the matrix $\mathbf{J} - \lambda\mathbf{I}$ is completely linearly dependent (perfectly coupled). Then the eigenvectors are unique only up to a scalar multiple. Hence choose $P_{1,4} = 1$ as the normalization. Then from row four it follows $P_{1,2} = \lambda_j^i$. With $P_{1,2} = \lambda_j^i, P_{1,4} = 1$, we get from row 1 and 3

$$P_{1,1} = \frac{1}{\frac{Cr_C}{\sigma} - \lambda_j^i}\left[\frac{-Cr_n}{\sigma} - \frac{Cr_K}{\sigma}P_{1,3}\right] \tag{248}$$

$$P_{1,1} = \frac{1}{Y_C - 1}\left[-Y_n - (Y_K - \lambda_j^i)v_{1,3}\right] \tag{249}$$

Equating and solving

$$P_{1,3} = \frac{\frac{C}{\sigma}r_n(Y_C - 1) - Y_n(\frac{C}{\sigma}r_C - \lambda_j^i)}{(\frac{Cr_C}{\sigma} - \lambda_j^i)(Y_K - \lambda_j^i) - \frac{C}{\sigma}r_K(Y_C - 1)} \tag{250}$$

Plug back in

$$P_{1,1} = \frac{\frac{C}{\sigma}(r_K Y_n - r_n(Y_K - \lambda_j^i))}{(\frac{C}{\sigma}r_C - \lambda_j^i)(Y_K - \lambda_j^i) - \frac{C}{\sigma}r_K(Y_C - 1)} \tag{251}$$

So our eigenvector

$$\begin{bmatrix} P_{i,1} \\ P_{i,2} \\ P_{i,3} \\ P_{i,4} \end{bmatrix} = \begin{bmatrix} \frac{\frac{C}{\sigma}(r_K Y_n - r_n(Y_K - \lambda_j^i))}{(\frac{C}{\sigma}r_C - \lambda_j^i)(Y_K - \lambda_j^i) - \frac{C}{\sigma}r_K(Y_C - 1)} \\ \lambda_j^i \\ \frac{\frac{C}{\sigma}r_n(Y_C - 1) - Y_n(\frac{C}{\sigma}r_C - \lambda_j^i)}{(\frac{Cr_C}{\sigma} - \lambda_j^i)(Y_K - \lambda_j^i) - \frac{C}{\sigma}r_K(Y_C - 1)} \\ 1 \end{bmatrix} \tag{252}$$

B.8 OPEN LOOP SOLUTION FORM

This section shows how to derive the open-loop solution form.

We may solve our linearized system to get x. The linearized system has the form $\dot{x} = Jx$ and if you factor out the eigenvectors $x = Pq$ we can write in Jordan canonical form

$$\dot{q} = P^{-1}JPq$$

where we denote $P^{-1}JP = \Lambda$ and call it the Jordan normal form of Λ where P is an invertible matrix.

The Jordan form has solution $q = \omega e^{\lambda t}$ where $\boldsymbol{\omega}$ is a 1×4 vector of constants and $e^{\lambda t}$ is a 4×1 vectors corresponding to each eigenvalue of the system.

$$q = \begin{bmatrix} q_2^u \\ q_1^u \\ q_1^s \\ q_2^s \end{bmatrix} = \begin{bmatrix} \omega_2^u e^{\lambda_2^u t} \\ \omega_1^u e^{\lambda_1^u t} \\ \omega_1^s e^{\lambda_1^s t} \\ \omega_2^s e^{\lambda_2^s t} \end{bmatrix} \tag{253}$$

gives the solution $x = pq$

$$\begin{bmatrix} c(t) - c^* \\ e(t) - e^* \\ K(t) - K^* \\ n(t) - n^* \end{bmatrix} = \begin{bmatrix} P_{2,1}^u & P_{1,1}^u & P_{1,1}^s & P_{2,1}^s \\ P_{2,2}^u & P_{1,2}^u & P_{1,2}^s & P_{2,2}^s \\ P_{2,3}^u & P_{1,3}^u & P_{1,3}^s & P_{2,3}^s \\ P_{2,4}^u & P_{1,4}^u & P_{1,4}^s & P_{2,4}^s \end{bmatrix} \begin{bmatrix} \omega_2^u e^{\lambda_2^u t} \\ \omega_1^u e^{\lambda_1^u t} \\ \omega_1^s e^{\lambda_1^s t} \\ \omega_2^s e^{\lambda_2^s t} \end{bmatrix} \tag{254}$$

Next we find what the stable constants ω_1^s, ω_2^s must be in order to suppress the explosive eigenvalues i.e. to ensure we converge. So, set unstable constant of integrations to zero

$$\begin{bmatrix} c(t)-c^* \\ e(t)-e^* \\ K(t)-K^* \\ n(t)-n^* \end{bmatrix} = \begin{bmatrix} P^s_{1,1} & P^s_{2,1} \\ P^s_{1,2} & P^s_{2,2} \\ P^s_{1,3} & P^s_{2,3} \\ P^s_{1,4} & P^s_{2,4} \end{bmatrix} \begin{bmatrix} e^{\lambda^s_1 t} & 0 \\ 0 & e^{\lambda^s_2 t} \end{bmatrix} \begin{bmatrix} \omega^s_1 \\ \omega^s_2 \end{bmatrix} \tag{255}$$

$$\begin{bmatrix} P^s_{1,3} & P^s_{2,3} \\ P^s_{1,4} & P^s_{2,4} \end{bmatrix}^{-1} \begin{bmatrix} K(t)-K^* \\ n(t)-n^* \end{bmatrix} = \begin{bmatrix} \omega^s_1 e^{\lambda^s_1 t} \\ \omega^s_2 e^{\lambda^s_2 t} \end{bmatrix} \tag{256}$$

$$\begin{bmatrix} P^s_{1,3} & P^s_{2,3} \\ P^s_{1,4} & P^s_{2,4} \end{bmatrix}^{-1} \begin{bmatrix} K(t)-K^* \\ n(t)-n^* \end{bmatrix} = \begin{bmatrix} e^{\lambda^s_1 t} & 0 \\ 0 & e^{\lambda^s_2 t} \end{bmatrix} \begin{bmatrix} \omega^s_1 \\ \omega^s_2 \end{bmatrix} \tag{257}$$

$t \rightarrow 0$

$$\begin{bmatrix} P^s_{1,3} & P^s_{2,3} \\ P^s_{1,4} & P^s_{2,4} \end{bmatrix}^{-1} \begin{bmatrix} K(0)-K^* \\ n(0)-n^* \end{bmatrix} = \begin{bmatrix} \omega^s_1 \\ \omega^s_2 \end{bmatrix} \tag{258}$$

which if we plug back in

$$\begin{bmatrix} c(t)-c^* \\ e(t)-e^* \\ K(t)-K^* \\ n(t)-n^* \end{bmatrix} = \begin{bmatrix} P^s_{1,1} & P^s_{2,1} \\ P^s_{1,2} & P^s_{2,2} \\ P^s_{1,3} & P^s_{2,3} \\ P^s_{1,4} & P^s_{2,4} \end{bmatrix} \begin{bmatrix} e^{\lambda^s_1 t} & 0 \\ 0 & e^{\lambda^s_2 t} \end{bmatrix} \begin{bmatrix} \omega^s_1 \\ \omega^s_2 \end{bmatrix} \tag{259}$$

$$\begin{bmatrix} c(t)-c^* \\ e(t)-e^* \\ K(t)-K^* \\ n(t)-n^* \end{bmatrix} = \begin{bmatrix} P^s_{1,1} & P^s_{2,1} \\ P^s_{1,2} & P^s_{2,2} \\ P^s_{1,3} & P^s_{2,3} \\ P^s_{1,4} & P^s_{2,4} \end{bmatrix} \begin{bmatrix} e^{\lambda^s_1 t} & 0 \\ 0 & e^{\lambda^s_2 t} \end{bmatrix} \begin{bmatrix} P^s_{1,3} & P^s_{2,3} \\ P^s_{1,4} & P^s_{2,4} \end{bmatrix}^{-1} \begin{bmatrix} K(0)-K^* \\ n(0)-n^* \end{bmatrix} \tag{260}$$

Multiplying out yields the solution in *open-loop* form

$$\begin{bmatrix} c(t)-c^* \\ e(t)-e^* \\ K(t)-K^* \\ n(t)-n^* \end{bmatrix} = \tag{261}$$

$$\frac{1}{P^s_{1,3}P^s_{2,4}-P^s_{2,3}P^s_{1,4}} \begin{bmatrix} P^s_{1,1}P^s_{2,4}e^{\lambda_1 t}-P^s_{2,1}P^s_{1,4}e^{\lambda_2 t} & -P^s_{1,1}P^s_{2,3}e^{\lambda_1 t}+P^s_{2,1}P^s_{1,3}e^{\lambda_2 t} \\ P^s_{1,2}P^s_{2,4}e^{\lambda_1 t}-P^s_{2,2}P^s_{1,4}e^{\lambda_2 t} & -P^s_{1,2}P^s_{2,3}e^{\lambda_1 t}+P^s_{2,2}P^s_{1,3}e^{\lambda_2 t} \\ P^s_{1,3}P^s_{2,4}e^{\lambda_1 t}-P^s_{2,3}P^s_{1,4}e^{\lambda_2 t} & -P^s_{1,3}P^s_{2,3}e^{\lambda_1 t}+P^s_{2,3}P^s_{1,3}e^{\lambda_2 t} \\ P^s_{1,4}P^s_{2,4}e^{\lambda_1 t}-P^s_{2,4}P^s_{1,4}e^{\lambda_2 t} & -P^s_{1,4}P^s_{2,3}e^{\lambda_1 t}+P^s_{2,4}P^s_{1,3}e^{\lambda_2 t} \end{bmatrix} \begin{bmatrix} K(0)-K^* \\ n(0)-n^* \end{bmatrix} \tag{262}$$

B.9 PRODUCTIVITY EFFECT

Denote $F(K,L)$ by F. On impact $t=0$ the state variables K,n are predetermined so do not adjust; however, L adjusts[2]. By the quotient rule differentiate $\mathcal{P} = \frac{Y}{F^{\frac{1}{\nu}}} = \frac{An^{1-\nu}F+n\phi}{F^{\frac{1}{\nu}}}$ with K,n fixed

$$\begin{aligned} \frac{d\mathcal{P}(0)}{dA} = F^{-\frac{1}{\nu}}(n^{1-\nu}F + An^{1-\nu}F_L L_A) - \\ (An^{1-\nu}F+n\phi)\frac{1}{\nu}F^{-\frac{1}{\nu}-1}F_L L_A \end{aligned} \tag{263}$$

2 See Caputo 2005 p.426

Evaluate impact effect beginning in steady state $\left.\frac{\partial \mathcal{P}(0)}{\partial A}\right|_{K^*,n^*}$. In equilibrium

$$\begin{aligned}\mathcal{P}^* &= (1-\zeta)A\nu {n^*}^{1-\nu} F(K^*,L^*)^{1-\frac{1}{\nu}} \\ &= (1-\zeta)A^{\frac{1}{\nu}}\nu\left(\frac{\phi}{1-(1-\zeta)\nu}\right)^{1-\frac{1}{\nu}} \qquad (264)\end{aligned}$$

so $\frac{\partial \mathcal{P}^*}{\partial A} = (1-\zeta)A^{\frac{1}{\nu}-1}\left(\frac{\phi}{1-(1-\zeta)\nu}\right)^{1-\frac{1}{\nu}} = \frac{\mathcal{P}^*}{A\nu}$. Therefore

$$\frac{d\mathcal{P}(0)}{dA} = n^{1-\nu}F^{1-\frac{1}{\nu}} + \frac{\mathcal{P}^*}{(1-\zeta)F}F_L L_A - \mathcal{P}^*\frac{1}{\nu F}F_L L_A \qquad (265)$$

$$= \frac{\mathcal{P}^*}{(1-\zeta)A\nu} + \frac{\mathcal{P}^* F_L L_A}{\nu F}\left(\frac{\zeta}{1-\zeta}\right) \qquad (266)$$

$$= \frac{1}{(1-\zeta)}\frac{d\mathcal{P}^*}{dA} + \frac{\mathcal{P}^* F_L L_A}{\nu F}\left(\frac{\zeta}{1-\zeta}\right) \qquad (267)$$

$$\frac{d\mathcal{P}(0)}{dA} - \frac{d\mathcal{P}^*}{dA} = \left(\frac{\zeta}{1-\zeta}\right)\frac{d\mathcal{P}^*}{dA} + \frac{\mathcal{P}^* F_L L_A}{\nu F}\left(\frac{\zeta}{1-\zeta}\right) \qquad (268)$$

Substitute out $\mathcal{P}^*$ and $\frac{\partial \mathcal{P}^*}{\partial A}$

$$\frac{d\mathcal{P}(0)}{dA} - \frac{d\mathcal{P}^*}{dA} = \zeta\left(\frac{\phi}{A(1-(1-\zeta)\nu)}\right)^{1-\frac{1}{\nu}}\left(1+\frac{AF_L L_A}{F}\right) \qquad (269)$$

C

CHAPTER 3 APPENDIX

C.1 HOUSEHOLD OPTIMIZATION PROBLEM

Use the Maxmimum Principle to obtain the necessary conditions for a solution to the household's utility maximisation problem. The current value Hamiltonian is

$$\hat{\mathcal{H}}(t) = u(C(t), L(t)) + \lambda(t)(w(t)L(t) + r(t)K(t) + \Pi(t) - C(t)) \tag{270}$$

The costate variable λ_t is the shadow price of wealth in utility units. The Pontryagin necessary conditions are

$$\hat{\mathcal{H}}_C(K, L, C, \lambda) = 0 \implies u_C - \lambda = 0 \tag{271}$$

$$\hat{\mathcal{H}}_L(K, L, C, \lambda) = 0 \implies u_L + \lambda w = 0 \tag{272}$$

$$\hat{\mathcal{H}}_K(K, L, C, \lambda) = \rho\lambda - \dot{\lambda} \implies \lambda r = \rho\lambda - \dot{\lambda} \implies \frac{\dot{\lambda}}{\lambda} = -(r - \rho) \tag{273}$$

$$\hat{\mathcal{H}}_\lambda := \dot{K}_t \implies \dot{K} = rK + wL + \Pi - C \tag{274}$$

The four Pontryagin conditions (271)-(273) reduce to two equations: a differential equation in consumption (*consumption Euler equation* or *intertemporal condition*), and a static injective mapping between labour and consumption (*intratemporal condition*).

C.2 OPTIMAL LABOUR DERIVATIVES

Partially differentiate the intratemporal Euler with respect to each variable treating labour as an implicit function, and with wage set at the imperfect competition market rate $w(K, L, n) = \frac{An^{1-\nu}F_L(K,L)}{\mu(n)}$.

$$u_L(L) + u_C(C)w(K, L, n) = 0 \tag{275}$$

$$u_L(L) + u_C(C)\frac{An^{1-\nu}F_L(K, L)}{\mu(n)} = 0 \tag{276}$$

Recall the utility and production function assumptions:

$$F_{LL}(K, L),\ u_{CC}(C),\ u_{LL}(L) < 0$$
$$u_C(C),\ F_L(K, L),\ F_{LK}(K, L) = F_{KL}(K, L) > 0$$

These can be used to sign the behaviour of labour

$$u_{LL}L_C + u_C\frac{An^{1-\nu}F_{LL}L_C}{\mu(n)} + \frac{u_{CC}An^{1-\nu}F_L}{\mu(n)} = 0 \tag{277}$$

$$L_C = \frac{-u_{CC}An^{1-\nu}F_L\mu(n)^{-1}}{u_{LL} + u_C An^{1-\nu}F_{LL}\mu(n)^{-1}} < 0 \tag{278}$$

$$u_{LL}L_K + u_C \frac{An^{1-\nu}F_{LL}L_K}{\mu(n)} + \frac{u_C An^{1-\nu}F_{LK}}{\mu(n)} = 0 \tag{279}$$

$$L_K = \frac{-u_C An^{1-\nu}F_{LK}\mu(n)^{-1}}{u_{LL} + u_C An^{1-\nu}F_{LL}\mu(n)^{-1}} > 0 \tag{280}$$

$$u_{LL}L_n + \frac{u_C A(1-\nu)n^{-\nu}F_L}{\mu(n)} + \frac{u_C An^{1-\nu}F_{LL}L_n}{\mu(n)} + u_C An^{1-\nu}F_L \frac{-\mu(n)_n}{\mu(n)^2} \tag{281}$$

$$L_n = \frac{u_C An^{1-\nu}F_L\mu_n\mu^{-2} - u_C A(1-\nu)n^{-\nu}F_L\mu^{-1}}{u_{LL} + u_C An^{1-\nu}F_{LL}\mu(n)^{-1}} \tag{282}$$

Therefore if we suppress notation and simplify (e.g. $n^{1-\nu}F_{LL}(K,L) = n^{-1}n^{2-\nu}F_{LL}(K,L) = n^{-1}F_{ll}(\frac{K}{n},\frac{L}{n})$ by Euler's homogeneous function theorem) we get

$$L_n = \frac{Au_C F_l(\mu^{-1}\mu_n - (1-\nu)n^{-1})}{\mu u_{LL} + U_C An^{-1}F_{ll}} > 0, \quad \nu \in (0,1) \tag{283}$$

In all cases the denominator $u_{LL} + u_C An^{1-\nu}F_{LL}\mu(n)^{-1} < 0$ is the intratemporal condition differentiated with respect to labour, and it is negative. Therefore the numerator distinguishes signs. Concavity of the production and utility functions, assumptions above, are sufficient to determine the signs of the numerator except for L_n which depends on returns to scale of the technology ν. With decreasing returns $\nu < 1$ labour increases; with increasing returns labour decreases and with constant returns $\nu = 1$ labour would be irresponsive to entry if there were fixed markups $\mu_n = 0$, but the endogenous markup $\mu_n < 0$ means labour increases with entry even with constant returns. This is because although the marginal product of labour does not change because of constant returns, the fall in markups reduces the wedge between marginal product of labour and wage, so wage increases.

The economic intuition is easier to understand in terms of wages, where $w^{\bar{L}}$ is wage with labour fixed.

$$L_C = \frac{-u_{CC}w}{u_{LL} + u_C w_L} < 0, \quad L_K = \frac{-u_C w_K^{\bar{L}}}{u_{LL} + u_C w_L} > 0, \tag{284}$$

$$L_n = \frac{w\frac{\mu_n}{\mu} - u_C w_n^{\bar{L}}}{u_{LL} + u_C w_L} > 0 \tag{285}$$

C.3 OPTIMAL INTEREST RATE, PROFIT, OUTPUT

Given optimal labour choice $L(C, K, n)$ we can evaluate how interest rate, wage, profit and output respond. The markup μ is a function of number of firms $\mu(n)$, but I suppress the domain for clarity.

C.3.1 *Output*

$$Y(L(C,K,n),K,n) = n^{1-\nu}[AF(K, L(C,K,n)) - \phi] \tag{286}$$

$$Y_C = An^{1-\nu}F_L(K,L)L_C(C,K,n) < 0 \tag{287}$$

$$Y_K = An^{1-\nu}[F_K(K,L) + F_L(K,L)L_K(C,K,n)] > 0 \tag{288}$$

$$Y_n = (1-\nu)AF - \phi + AF_l L_n \lesseqgtr 0 \tag{289}$$

Furthermore in steady state when $F(\frac{K}{n}, \frac{L}{n})^* = \frac{\phi}{A(1-\frac{\nu}{\mu})}$ then $Y_n|_{x*} = \frac{-\phi\mu\nu}{1-\frac{\nu}{\mu}} + AF_l L_n$ which is positive or negative depending whether the negative component outweighs the positive labour effect.

C.3.2 *Wage*

$$w = \frac{1}{\mu} AF_l \tag{290}$$

$$w_C = \frac{1}{\mu}\frac{A}{n} F_{ll} L_C > 0 \tag{291}$$

$$w_K = \frac{1}{\mu}\frac{A}{n}[F_{lk} + F_{ll}L_K] = \frac{1}{\mu}\frac{A}{n}\left[\frac{\mu F_{lk} u_{LL}}{\mu u_{LL} + u_C AF_{ll}}\right] > 0 \tag{292}$$

$$w_n = \frac{1}{\mu}\frac{A}{n}[(1-\nu)F_l + F_{ll}L_n] - \frac{1}{\mu^2}\mu_n AF_l \gtreqless 0 \tag{293}$$

C.3.3 *Rents*

$$r = \frac{1}{\mu} AF_k \tag{294}$$

$$r_C = \frac{1}{\mu}\frac{A}{n} F_{kl} L_C < 0 \tag{295}$$

$$r_K = \frac{1}{\mu}\frac{A}{n}[F_{kk} + F_{kl}L_K] = \frac{1}{\mu}\frac{A}{n}\left[\frac{\mu F_{kk} u_{LL} + Au_C(F_{kk}F_{ll} - F_{kl}^2)}{\mu u_{LL} + u_C AF_{ll}}\right] < 0 \tag{296}$$

$$r_n = \frac{1}{\mu}\frac{A}{n}[(1-\nu)F_k + F_{kl}L_n] - \frac{1}{\mu^2}\mu_n AF_k > 0 \tag{297}$$

Both r_K and w_K require extra work to derive the signs. They are found by substituting in L_K. Then r_K can can be rearranged into a form including $F_{KK}F_{LL} - F_{KL}^2$ which is positive by the second partial derivative test for concavity assumption.

C.3.4 *Profit*

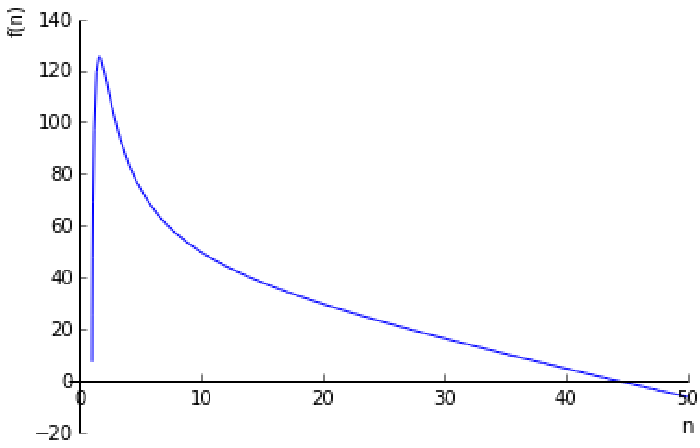
$$\pi = AF(k,l)(1 - \frac{\nu}{\mu}) - \phi \tag{298}$$

$$\pi_C = AF_l \frac{L_C}{n}(1 - \frac{\nu}{\mu}) < 0 \tag{299}$$

$$\pi_K = \frac{A}{n}(F_k + F_l L_K)(1 - \frac{\nu}{\mu}) > 0 \tag{300}$$

$$\pi_n = \frac{A}{n}(-\nu F + F_l L_n)(1 - \frac{\nu}{\mu}) + AF(k,l)\frac{\nu}{\mu^2}\mu_n < 0 \tag{301}$$

For any K, L, n profit is higher when imperfect competition μ increases, but not necessarily higher for any given K, n. This explains that even if imperfect competition increases and therefore higher profits are available, the number of firms can (counterintuitively) decrease. The offsetting factor is the the indirect labour effect: labour supply is discouraged by the rise in imperfect competition and as there is a bigger wedge between wage and marginal product of labour. Therefore labour supply falls such that profits are lower for a given K, n. I typically assume these secondary labour effects to be too small to offset the primary mechanisms. So operating profit increases with imperfect competition, even as L is allowed to adjust.



(a) Steady State Numerical Solution

C.4 STEADY STATE RESULTS

C.4.1 *Existence with Functional Forms Numerically*

Since we have shown an analytical condition for existence, we can move on from existence to ask what that solution is for a set of numerical parameter values? Solving the highly nonlinear number of firms in steady-state equation yields $n^* = 44.156$. The function for number of firms in steady state

$$n^* = \left[\frac{\beta}{\xi \nu^\sigma} \left\{ \left(A \left(\frac{\alpha}{\rho} \right)^\alpha \right)^{1+\eta} \left(\frac{1}{\mu(n^*)} \right)^{\alpha(1+\eta)+\beta(1-\sigma)} \left(\frac{1 - \frac{\nu}{\mu(n^*)}}{\phi} \right)^{1-\nu+\eta(1-\alpha)+\sigma\beta} \right\}^{\frac{1}{\beta}} \right]^{\frac{1}{\eta+\sigma}} \tag{302}$$

is highly nonlinear as shown in figure 19a, where it intersects the x-axis at the solution $n^* = 44.156$. The markup is $\mu = \frac{n}{n-1}$

hence the graph is undefined in the $n = 1$ region. The other solutions are $n^* = 44.156,\ K^* = 713.685,\ C^* = 47.486$.

C.4.2 *Procyclical Firms*

From (156) we have

$$Y^* = \frac{n^*\phi\nu}{\mu(n^*) - \nu} \tag{303}$$

Then by the product rule and the implicit function theorem

$$Y^*_n = \frac{(\mu(n^*) - \nu)\phi\nu - n^*\phi\nu\mu_n}{(\mu(n^*) - \nu)^2} > 0 \tag{304}$$

The quadratic denominator is positive. The first component of the numerator is positive because $\mu(n^*) > 1$ and decreasing returns $\nu \in (0, 1)$ so $\mu(n) - \nu > 0$ and the second component is positive due to the double negative which occurs from endogenous markups decreasing in number of firms $\mu_n < 0$.

C.5 PRODUCTIVITY DYNAMICS

Throughout the derivations remember that the markup is a function of number of firms $\mu(n)$, but for simplicity I write μ.

$$\mathcal{P}(t)_A = \frac{n^{1-\nu}F + yn_A + (n^{1-\nu}A - ny\frac{1}{\nu}F^{-1})(F_K K_A + F_L L_A)}{F^{\frac{1}{\nu}}} \tag{305}$$

The crucial step with dynamic firms and capital is that state variables do not move on impact $K_A = 0$ and $n_A = 0$. This is what causes the distinction between short-run and long-run productivity that is not present with instantaneous free entry. Therefore at $t = 0$ the change in productivity depends on the direct effect of better technology, and its indirect effect on labour, which increases labour supply.

$$\mathcal{P}(0)_A = \frac{n^{1-\nu}F + (n^{1-\nu}A - ny\frac{1}{\nu}F^{-1})F_L L_A}{F^{\frac{1}{\nu}}} \tag{306}$$

$$\mathcal{P}(0)_A = n^{1-\nu}F^{1-\frac{1}{\nu}} + \left(F^{-\frac{1}{\nu}}n^{1-\nu}A - \mathcal{P}\frac{1}{\nu}F^{-1}\right)F_L L_A \tag{307}$$

Assuming that the economy is initially in steady state when the shock occurs, evaluate the expression with all variables x at steady state $x(0) = x^*$. From $\pi = y - rK - wL$ then $y^* = rK^* + wL^*$ so $y^* = \frac{A\nu n^{*-\nu}F}{\mu^*}$ and thus $\mathcal{P}^* = \frac{n^*y^*}{F^{*\frac{1}{\nu}}} = \frac{A\nu n^{*1-\nu}F^{1-\frac{1}{\nu}}}{\mu}$. This expression for productivity makes it easier to represent the impact effect of a TFP shock in terms of steady state productivity $\mathcal{P}^*$ as follows

$$\mathcal{P}(0)_A|_{x(0)=x^*} = \frac{\mu^*\mathcal{P}^*}{A\nu} + \left(\frac{\mu^*\mathcal{P}^*}{\nu F^*} - \frac{\mathcal{P}^*}{\nu F^*}\right)F_L^* L_A^* \tag{308}$$

$$\mathcal{P}(0)_A|_{x(0)=x^*} = \frac{\mu^*\mathcal{P}^*}{A\nu} + (\mu^* - 1)\frac{\mathcal{P}^*}{\nu F^*}F_L^* L_A^* \tag{309}$$

Comparing the short-run impact effect to the long-run steady state effect $\mathcal{P}_A^* = \mathcal{P}_A^{*\bar{\mu}} + \mathcal{P}_\mu^*\mu_A = \frac{\mathcal{P}^*}{A\nu} + \mathcal{P}_\mu^*\mu_n n_A^*$ shows that the endogenous productivity effect dampens the difference between short-run and long-run effects

$$\mathcal{P}(0)_A - \mathcal{P}^*_A|_{x(0)=x^*} = \frac{\mu^*\mathcal{P}^*}{A\nu} + (\mu^*-1)\,\frac{\mathcal{P}^*}{\nu F^*}F^*_L L^*_A - (\mathcal{P}^{*\bar{\mu}}_A + \mathcal{P}^*_\mu \mu_A) \tag{310}$$

$$\mathcal{P}(0)_A - \mathcal{P}^*_A|_{x(0)=x^*} = \frac{\mu^*\mathcal{P}^*}{A\nu} + (\mu^*-1)\,\frac{\mathcal{P}^*}{\nu F^*}F^*_L L^*_A - \left(\frac{\mathcal{P}^*}{A\nu} + \mathcal{P}^*_\mu \mu_n n^*_A\right) \tag{311}$$

$$\mathcal{P}(0)_A - \mathcal{P}^*_A|_{x(0)=x^*} = (\mu^*-1)\frac{\mathcal{P}^*}{A\nu} + (\mu^*-1)\,\frac{\mathcal{P}^*}{\nu F^*}F^*_L L^*_A - \mathcal{P}^*_\mu \mu_n n^*_A \tag{312}$$

$$\mathcal{P}(0)_A - \mathcal{P}^*_A|_{x(0)=x^*} = (\mu^*-1)\frac{\mathcal{P}^*}{\nu}\left(\frac{1}{A} + \frac{F^*_L L^*_A}{F^*}\right) - \mathcal{P}^*_\mu \mu_n n^*_A \tag{313}$$

The expression for $\mathcal{P}^*_\mu$ can simplyify the expression further

Lemma 4.

$$\mathcal{P}^*_\mu = -\frac{\mathcal{P}^*(\mu-1)}{\mu(\mu-\nu)} < 0 \tag{314}$$

Proof.

$$\mathcal{P}^*_\mu = \left[\frac{A}{\mu}\left(\frac{\mu-\nu}{\phi}\right)^{1-\nu}\right]^{\frac{1}{\nu}-1} \times$$
$$\left(\frac{-A}{\mu^2}\left(\frac{\mu-\nu}{\phi}\right)^{1-\nu} + \frac{A}{\mu}(1-\nu)\left(\frac{\mu-\nu}{\phi}\right)^{-\nu}\frac{1}{\phi}\right) \tag{315}$$
$$= \left[\frac{A}{\mu}\left(\frac{\mu-\nu}{\phi}\right)^{1-\nu}\right]^{\frac{1}{\nu}}\left(-\frac{1}{\mu} + \frac{1-\nu}{\mu-\nu}\right) \tag{316}$$
$$= \left[\frac{A}{\mu}\left(\frac{\mu-\nu}{\phi}\right)^{1-\nu}\right]^{\frac{1}{\nu}}\left(\frac{\nu(1-\mu)}{\mu(\mu-\nu)}\right) = -\frac{\mathcal{P}^*(\mu-1)}{\mu(\mu-\nu)} < 0 \tag{317}$$

□

Notice that with constant returns the expression is simply

$$\mathcal{P}^*_\mu\Big|_{\nu\to 1} = -\frac{A}{\mu^2}$$

and with many firms the markup tends to unity, therefore long-run underlying productivity reflects true TFP.

$$\mathcal{P}^*_\mu = -\frac{\mathcal{P}^*(\mu^*-1)}{\mu^*(\mu^*-\nu)} \tag{318}$$
$$\mathcal{P}(0)_A - \mathcal{P}^*_A|_{x(0)=x^*} =$$
$$(\mu^*-1)\frac{\mathcal{P}^*}{\nu}\left(\frac{1}{A} + \frac{F^*_L L^*_A}{F^*}\right) + \frac{\mathcal{P}^*(\mu^*-1)}{\mu^*(\mu^*-\nu)}\mu_n n^*_A \tag{319}$$
$$\mathcal{P}(0)_A - \mathcal{P}^*_A|_{x(0)=x^*} =$$
$$(\mu^*-1)\mathcal{P}^*\left[\frac{1}{\nu}\left(\frac{1}{A} + \frac{F^*_L L^*_A}{F^*}\right) + \frac{1}{\mu^*(\mu^*-\nu)}\mu_n n^*_A\right] \tag{320}$$

C.6 PRODUCTIVITY DYNAMICS WITH FUNCTIONAL FORMS

First let us restate the short-run versus long-run productivity effect, and focus attention on the square bracketed component Γ that represents allocation versus competition effect.

$$\begin{aligned} &\mathcal{P}(0)_A|_{x(0)=x^*} - \mathcal{P}^*_A = \\ &(\mu^* - 1)\frac{\mathcal{P}^*}{\nu}\left[\left(\frac{1}{A} + \frac{F^*_L L^*_A}{F^*}\right) + \frac{\nu}{\mu^*(\mu^* - \nu)}\mu_n n^*_A\right] \qquad (321) \\ &\text{Define } \Gamma = \frac{1}{A} + \frac{F^*_L L^*_A}{F^*} + \frac{\nu}{\mu^*(\mu^* - \nu)}\mu_n n^*_A \qquad (322) \end{aligned}$$

From the state-state labour per firm l^* we have

$$L^* = n^* \left[\frac{1}{A}\left(\frac{\rho}{\alpha}\right)^{\alpha}\left(\frac{1 - \frac{\nu}{\mu}}{\frac{\phi}{\mu}}\right)^{1-\alpha}\right]^{\frac{1}{\beta}} \text{ thus } L^*_A = L^*\left(\frac{n^*_A}{n^*} + \frac{1}{\beta A}\right) \qquad (323)$$

We can also substitute out the following simplifications

$$\frac{F^*_L}{F^*} = \frac{\beta K^{\alpha} L^{\beta-1} n^{-(\alpha+\beta)}}{K^{\alpha} L^{\beta} n^{-(\alpha+\beta)}} = \beta L^{*-1} \qquad (324)$$

$$\mu = \frac{n}{n-1} \text{ therefore } \mu_n = -\frac{1}{n^2} \qquad (325)$$

Therefore the allocation versus competition effect component becomes

$$\Gamma = \frac{2}{A} + \left(\beta - \frac{\nu}{\left(\frac{n^*}{n^*-1}\right)^2 (n^* - (n^* - 1)\nu)}\right)\frac{n^*_A}{n^*} \qquad (326)$$

And short-run versus long-run productivity dynamics simplify to

$$\mathcal{P}(0)_A|_{x(0)=x^*} - \mathcal{P}_A^* = \left(\frac{1}{n^*-1}\right)\frac{\mathcal{P}^*}{\nu}\left[\frac{2}{A} + \left(\beta - \frac{\nu}{\left(\frac{n^*}{n^*-1}\right)^2 (n^* - (n^*-1)\nu)}\right)\frac{n_A^*}{n^*}\right] \tag{327}$$

$$\mathcal{P}^* = \nu\left[\frac{A(n^*-1)^\nu}{n^*}\left(\frac{n^*(1-\nu)+\nu}{\phi}\right)^{1-\nu}\right]^{\frac{1}{\nu}} \tag{328}$$

C.6.1 *Cost function*

Static optimization problem so drop time subscripts

$$C(r, w, y) = \min_{l,k} wl + rk + \phi \quad \text{s.t.} y \leqslant Ak^\alpha l^\beta - \phi \tag{329}$$

With Cobb-Douglas production the total cost function from substituting Lagrangean obtained conditional input demands $k(r, w, y) = \left[\left(\frac{w\alpha}{r\beta}\right)^\beta \left(\frac{y+\phi}{A}\right)\right]^{\frac{1}{\alpha+\beta}}$ and $l(r, w, y) = \left[\left(\frac{r\beta}{w\alpha}\right)^\alpha \left(\frac{y+\phi}{A}\right)\right]^{\frac{1}{\alpha+\beta}}$ into the cost function is

$$C(r, w, y) = (\alpha+\beta)\left(\frac{y+\phi}{A}\right)^{\frac{1}{\alpha+\beta}}\left(\frac{r}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}}\left(\frac{w}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} + \phi \tag{330}$$

Where the firm takes factor prices as given. The average cost $AC := \frac{C}{y}$ is U-shaped and the marginal cost $MC := \frac{dC}{dy}$ is increasing in output with $\alpha + \beta < 1$.

$$MC = \frac{\partial C(r,w,y)}{\partial y} = \frac{(y+\phi)^{\frac{1}{\alpha+\beta}-1}}{A^{\frac{1}{\alpha+\beta}}} \left(\frac{r}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}} \left(\frac{w}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} \quad (331)$$

$$\frac{\partial MC}{y} = \left(\frac{1}{\alpha+\beta} - 1\right) \frac{(y+\phi)^{\frac{1}{\alpha+\beta}-2}}{A^{\frac{1}{\alpha+\beta}}} \left(\frac{r}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}} \left(\frac{w}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} \quad (332)$$

The leading multiplier $\frac{1}{\alpha+\beta} - 1$ determines how marginal cost responds to changing output. This shows that it is increasing when $\alpha + \beta < 1$ but is zero with constrant returns to scale $\alpha + \beta = 1$ which reflects a flat marginal cost curve.

C.7 MARKUP PROPERTIES

If $\theta_I = 1$ industry goods are imperfectly substitutable, and the aggregate good is a Cobb-Douglas composite of industry goods. Thus the markup is a common asymptotic function[1].

Remark 2 (Endogenous markup). With many firms per industry the markup is 1

Proof.

$$\lim_{n \to +\infty} \mu(n(t)) = \lim_{n \to +\infty} \frac{n(t)}{n(t)-1} \quad (333)$$

$$= \lim_{n \to +\infty} \left(\frac{1}{n(t)-1}\right) + 1 = 1, \quad n(t) \in (0, \infty] \quad (334)$$

□

1 See Wolfram Alpha for eloquent properties.

REFERENCES

Aloi, Marta and Huw Dixon (2003). "Entry Dynamics, Demand Shocks and Induced Productivity Fluctuations". In: *Ekonomia* 6.2, pp. 115–146.

Aloi, Marta and Huw David Dixon (2002). *Entry Dynamics, Capacity Utilisation and Productivity in a Dynamic Open Economy*. Discussion Papers 02/05. Department of Economics, University of York.

Ambler, Steve and Emanuela Cardia (1998). "The Cyclical Behaviour of Wages and Profits under Imperfect Competition". In: *Canadian Journal of Economics* 31.1, pp. 148–164.

Barnett, Alina, Sandra Batten, Adrian Chiu, Jeremy Franklin, and Maria Sebastia-Barriel (2014). "The UK productivity puzzle". In: *Bank of England Quarterly Bulletin*, Q2.

Berentsen, Aleksander and Christopher J. Waller (2009). *Optimal stabilization policy with endogenous firm entry*. Working Papers 2009-032. Federal Reserve Bank of St. Louis.

Bergin, Paul R. and Giancarlo Corsetti (2008). "The extensive margin and monetary policy". In: *Journal of Monetary Economics* 55.7, pp. 1222–1237.

Bilbiie, Florin, Fabio Ghironi, and Marc J. Melitz (2007). *Endogenous Entry, Product Variety, and Business Cycles*. NBER Working Papers 13646. National Bureau of Economic Research, Inc.

Bilbiie, Florin O., Fabio Ghironi, and Marc J. Melitz (2012). "Endogenous Entry, Product Variety, and Business Cycles". In: *Journal of Political Economy* 120.2, pp. 304 –345.

Bils, Mark (1987). "The cyclical behavior of marginal cost and price". In: *The American Economic Review*, pp. 838–855.

Blanchard, Olivier Jean and Nobuhiro Kiyotaki (1987). "Monopolistic competition and the effects of aggregate demand". In: *The American Economic Review*, pp. 647–666.

Brito, Paulo and Huw Dixon (2013). "Fiscal policy, entry and capital accumulation: Hump-shaped responses". In: *Journal of Economic Dynamics and Control* 37.10, pp. 2123–2155.

Brito, Paulo B., Luís F. Costa, and Huw Dixon (2013). "Nonsmooth dynamics and multiple equilibria in a Cournot-Ramsey model with endogenous markups". In: *Journal of Economic Dynamics and Control* 37.11, pp. 2287 –2306.

Brock, William A and Stephen J Turnovsky (1981). "The analysis of macroeconomic policies in perfect foresight equilibrium". In: *International Economic Review*, pp. 179–209.

Caballero, Ricardo J and Richard K Lyons (1992). "External effects in US procyclical productivity". In: *Journal of Monetary Economics* 29.2, pp. 209–225.

Campbell, Jeffrey R (1998). "Entry, exit, embodied technology, and business cycles". In: *Review of economic dynamics* 1.2, pp. 371–408.

Campbell, Jeffrey R. and Hugo A. Hopenhayn (2005). "Market Size Matters". In: *Journal of Industrial Economics* 53.1, pp. 1–25.

Caputo, M.R. (2005). *Foundations of Dynamic Economic Analysis: Optimal Control Theory and Applications*. Cambridge University Press.

Census, United States. Bureau of the, United States. Bureau of Foreign, Domestic Commerce, United States. Office of Business Economics, and United States. Bureau of Economic Analysis (1996). *Survey of Current Business*. v. 76. U.S. Department of Commerce.

Chamberlin, Edward (1933). *The theory of monopolistic competition*. Harvard University Press.

Chatterjee, Satyajit and Russell Cooper (1993). *Entry and Exit, Product Variety and the Business Cycle*. NBER Working Papers 4562. National Bureau of Economic Research, Inc.

— (2014). "Entry And Exit, Product Variety, And The Business Cycle". In: *Economic Inquiry* 52.4, pp. 1466–1484.

Clementi, Gian Luca and Berardino Palazzo (2013). *Entry, Exit, Firm Dynamics, and Aggregate Fluctuations*. Working Paper (Forthcoming AER: Macro) 19217. National Bureau of Economic Research.

Colciago, Andrea and Federico Etro (2010). "Real business cycles with Cournot competition and endogenous entry". In: *Journal of Macroeconomics* 32.4, pp. 1101–1117.

Comin, Diego and Mark Gertler (2006). "Medium-Term Business Cycles". In: *American Economic Review* 96.3, pp. 523–551.

Cook, David (2001). "Time to enter and business cycles". In: *Journal of Economic Dynamics and Control* 25.8, pp. 1241–1261.

Corhay, Alexandre, Howard Kung, and Lukas Schmid (2015). "Competition, Markups and Predictable Returns". In: *Available at SSRN 2550981*.

Das, Sanghamitra and Satya P Das (1997). "Dynamics of entry and exit of firms in the presence of entry adjustment costs". In: *International Journal of Industrial Organization* 15.2, pp. 217–241.

Datta, Bipasa and Huw Dixon (2002). "Technological Change, Entry, and Stock-Market Dynamics: An Analysis of Transition in a Monopolistic Industry". In: *American Economic Review* 92.2, pp. 231–235.

Devereux, Michael, Allen Head, and Beverly Lapham (1996a). "Aggregate fluctuations with increasing returns to specialization and scale". In: *Journal of Economic Dynamics and Control* 20.4, pp. 627–656.

Devereux, Michael B, Allen C Head, and Beverly J Lapham (1996b). "Monopolistic Competition, Increasing Returns, and the Effects of Government Spending". In: *Journal of Money, Credit and Banking* 28.2, pp. 233–54.

Dixit, Avinash K. and Joseph E. Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity". English. In: *The American Economic Review* 67.3, pp. 297–308.

Dixon, Huw (1987). "A simple model of imperfect competition with Walrasian features". In: *Oxford Economic Papers*, pp. 134–160.

Dockner, EngelbertJ. and Gustav Feichtinger (1991). "On the optimality of limit cycles in dynamic economic systems". English. In: *Journal of Economics* 53.1, pp. 31–50.

Dos Santos Ferreira, Rodolphe and Frédéric Dufourt (2006). "Free entry and business cycles under the influence of animal spirits". In: *Journal of Monetary Economics* 53.2, pp. 311–328.

Ericson, Richard and Ariel Pakes (1995). "Markov-Perfect Industry Dynamics: A Framework for Empirical Work". English. In: *The Review of Economic Studies* 62.1, pp. 53–82.

Etro, Federico (2009). *Endogenous market structures and the macroeconomy*. Springer Science & Business Media.

Etro, Federico and Andrea Colciago (2010). "Endogenous Market Structures and the Business Cycle*". In: *The Economic Journal* 120.549, pp. 1201–1233.

Feenstra, Robert C. (2003). "A homothetic utility function for monopolistic competition models, without constant price elasticity". In: *Economics Letters* 78.1, pp. 79–86.

Feichtinger, Gustav, Andreas Novak, and Franz Wirl (1994). "Limit cycles in intertemporal adjustment models: Theory and applications". In: *Journal of Economic Dynamics and Control* 18.2, pp. 353 –380.

Foster, Lucia, John Haltiwanger, and Chad Syverson (2008). "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" In: *American Economic Review* 98.1, pp. 394–425.

Gabaix, Xavier (2011). "The granular origins of aggregate fluctuations". In: *Econometrica* 79.3, pp. 733–772.

Giovanni, Julian di and Andrei A. Levchenko (2013). "Firm entry, trade, and welfare in Zipf's world". In: *Journal of International Economics* 89.2, pp. 283 –296.

Giovanni, Julian di, Andrei A Levchenko, and Isabelle Mejean (2014). "Firms, Destinations, and Aggregate Fluctuations". In: *Econometrica* 82.4, pp. 1303–1340.

Goodridge, PR, Jonathan Haskel, and Gavin Wallis (2014). "The UK productivity puzzle is a TFP puzzle: current data and future predictions". In:

Hall, Robert E (1987). "Productivity and the business cycle". In: *Carnegie-Rochester Conference Series on Public Policy*. Vol. 27. Elsevier, pp. 421–444.

— (1989). *Invariance properties of Solow's productivity residual*. Tech. rep. National Bureau of Economic Research.

Hamano, Masashige and Francesco Zanetti (2014). "Endogenous Establishment Destruction and Macroeconomic Dynamics". In:

Haskel, J, P Goodridge, and G Wallis (2015). *Accounting for the UK productivity puzzle: a decomposition and predictions*. Working Papers 21167. Imperial College, London, Imperial College Business School.

Hopenhayn, Hugo and Richard Rogerson (1993). "Job turnover and policy evaluation: A general equilibrium analysis". In: *Journal of political Economy*, pp. 915–938.

Hopenhayn, Hugo A. (1992a). "Entry, Exit, and firm Dynamics in Long Run Equilibrium". English. In: *Econometrica* 60.5, pp. 1127–1150.

Hopenhayn, Hugo A (1992b). "Exit, selection, and the value of firms". In: *Journal of Economic Dynamics and Control* 16.3, pp. 621–653.

Jaimovich, Nir (2007). "Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations". In: *Journal of Economic Theory* 137.1, pp. 300–325.

Jaimovich, Nir and Max Floetotto (2008). "Firm dynamics, markup variations, and the business cycle". In: *Journal of Monetary Economics* 55.7, pp. 1238 –1252.

Jovanovic, Boyan (1982). "Selection and the Evolution of Industry". In: *Econometrica* 50.3, pp. 649–70.

Kocherlakota, Narayana (2000). "Creating business cycles through credit constraints". In: *Federal Reserve Bank of Minneapolis Quarterly Review* 24.3, pp. 2–10.

Lee, Yoonsoo and Toshihiko Mukoyama (2008). *Entry, exit and plant-level dynamics over the business cycle*. Working Paper 0718. Federal Reserve Bank of Cleveland.

— (2015). "Entry and exit of manufacturing plants over the business cycle". In: *European Economic Review* 77, pp. 20 – 27.

Lewis, Vivien (2009). "Business Cycle Evidence On Firm Entry". In: *Macroeconomic Dynamics* 13 (05), pp. 605–624.

Lewis, Vivien and Céline Poilly (2012). "Firm entry, markups and the monetary transmission mechanism". In: *Journal of Monetary Economics* 59.7, pp. 670–685.

Lewis, Vivien and Arnaud Stevens (2015). "Entry and markup dynamics in an estimated business cycle model". In: *European Economic Review*.

Lewis, Vivien and Roland Winkler (2015a). "Fiscal policy and business formation in open economies". In: *Research in Economics* 69.4, pp. 603 –620.

— (2015b). "Product Diversity, Demand Structures, And Optimal Taxation". In: *Economic Inquiry* 53.2, pp. 979–1003.

Loualiche, Erik (2014). *Asset pricing with entry and imperfect competition*. Tech. rep. Working paper.

Luttmer, Erzo G.J. (2012). "Technology diffusion and growth". In: *Journal of Economic Theory* 147.2. Issue in honor of David Cass, pp. 602 –622.

Mata, Jos and Pedro Portugal (1994). "Life Duration of New Firms". English. In: *The Journal of Industrial Economics* 42.3, pp. 227–245.

Mertens, Karel and Morten Overgaard Ravn (2011). "Understanding the Aggregate Effects of Anticipated and Unanticipated Tax Policy Shocks". In: *Review of Economic Dynamics* 14.1, pp. 27–54.

Minniti, Antonio and Francesco Turino (2013). "Multi-product firms and business cycle dynamics". In: *European Economic Review* 57.C, pp. 75–97.

Nekarda, Christopher J and Valerie A Ramey (2013). *The cyclical behavior of the price-cost markup*. Tech. rep. National Bureau of Economic Research.

Ottaviano, Gianmarco I.P. (2011). *Firm Heterogeneity, Endogenous Entry, and the Business Cycle*. Working Paper 17433. National Bureau of Economic Research.

Peters, Michael (2013). "Heterogeneous mark-ups, growth and endogenous misallocation". In:

Portier, Franck (1995). "Business Formation and Cyclical Markups in the French Business Cycle". English. In: *Annales d'conomie et de Statistique* 37/38, pp. 411–440.

Rossi-Hansberg, Esteban and Mark L. J. Wright (2007). "Establishment Size Dynamics in the Aggregate Economy". In: *American Economic Review* 97.5, pp. 1639–1666.

Rotemberg, Julio J and Lawrence H Summers (1990). "Inflexible Prices and Procyclical Productivity". In: *The Quarterly Journal of Economics* 105.4, pp. 851–74.

Rotemberg, Julio J and Michael Woodford (1996). "Real-Business-Cycle Models and the Forecastable Movements in Output, Hours, and Consumption". In: *American Economic Review* 86.1, pp. 71–89.

Rotemberg, Julio J. and Michael Woodford (1999). "The cyclical behavior of prices and costs". In: *Handbook of Macroeconomics*. Ed. by J. B. Taylor and M. Woodford. Vol. 1. Handbook of Macroeconomics. Elsevier. Chap. 16, pp. 1051–1135.

Samaniego, Roberto M. (2008). "Entry, exit and business cycles in a general equilibrium model". In: *Review of Economic Dynamics* 11.3, pp. 529–541.

Spence, Michael (1976). "Product selection, fixed costs, and monopolistic competition". In: *The Review of Economic Studies*, pp. 217–235.

Stella, Andrea (2015). “Firm dynamics and the origins of aggregate fluctuations”. In: *Journal of Economic Dynamics and Control* 55, pp. 71 –88.

Stokey, Nancy L (2008). *The Economics of Inaction: Stochastic Control models with fixed costs*. Princeton University Press.

Veracierto, Marcelo (2008). “Firing Costs and Business Cycle Fluctuations*”. In: *International Economic Review* 49.1, pp. 1–39.

Vives, X. (1999). *Oligopoly Pricing: Old Ideas and New Tools*. MIT Press.

Winkler, Roland and Vivien Lewis (2014). “Government Spending, Entry and the Consumption Crowding-In Puzzle”. In: