Essays in International Macroeconomics

Alessandro Ferrari

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 15 May 2020

European University Institute
Department of Economics

Essays in International Macroeconomics

Alessandro Ferrari

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Ramon Marimon, EUI, Supervisor
Prof. Philipp Kircher, EUI
Prof. Pol Antràs, Harvard University
Prof. Isabelle Mejean, Ecole Polytechnique

Researcher declaration to accompany the submission of written work

I Alessandro Ferrari certify that I am the author of the work *Essays in International Macroeconomics* I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

The copyright of this work rests with its author. [quotation from it is permitted, provided that full acknowledgement is made.] This work may not be reproduced without my prior written consent. This authorisation does not, to the best of my knowledge, infringe the rights of any third party.

Statement of inclusion of previous work (if applicable):

I confirm that chapter 2 was jointly co-authored with Francisco Queirós.
I confirm that chapter 3 was jointly co-authored with Ramon Marimon and Chima Simpson-Bell.

Signature and Date:

Alessandro Ferrari, 29/04/2020

Acknowledgements

This thesis marks the end of 5 long years. Through the ups and downs a number of people have helped me and without them I am not sure I would have been able to go through it.

I am incredibly grateful to my supervisors. Firstly, Ramon Marimon, for teaching me how to think like an economist and for giving the constant support I needed to pursue what I thought was interesting and important. I am not positive that these years would have been as interesting and, at times, downright fun under someone else's supervision.

I am thankful to Philipp Kircher for giving me the opportunity to learn from him and for having the uncanny ability of saying a word of encouragement exactly whenever I needed it the most.

In these years I have been extremely lucky with the people I have been working with. I have had the privilege of learning from them. Most of the smart things in this thesis are either their ideas or things I came up with while picking their brains. I definitely owe a number of drinks to Carmen, Matic, Anna, and Andreas for that. I probably owe a couple dozens to Francisco and Chima for having to tolerate me during the last months of writing this thesis.

I am thankful to Juan Dolado for making these years much more interesting and being always there for me.

I want to thank my family, Federica, Gabriella and Marco, for all of their support. At the time I am writing this I have been in some version of university education for 11 years. That's something I do not take for granted.

I am thankful to Dalila. I hear I can be exhausting, thanks for withstanding it every day. 🚡

In these years I had the fortune of meeting some of the smartest people I know. They have made my PhD much more interesting and definitely more fun. In particular Andrea, Christian, Francisco, Chima, Egon and Andreas. I also have to thank my teammates, it's hard to think of a better way to end this run.

A special mention goes Matteo, Carmen, Giorgos and Matic. With many of you I have shared a house and with all of you I have shared a considerable chunk of my life. I would not have it any other way. You are a large part of why I will always remember this time fondly.

Abstract

In the first chapter I investigate the role of position in global value chains in the transmission of final demand shocks and the cyclicality and volatility of trade. Relying on a production network model with propagation via procyclical inventory adjustment, I show how shocks can magnify or dissipate upstream. I test the theoretical results empirically using input-output data. I find that industries far from consumers respond to final demand shocks up to twice as much as final goods producers. I also document the critical role of the position in the global value chain for countries' cyclical macroeconomic response: i) controlling for bilateral similarity in global value chain position eliminates the standard correlation between similarity in industrial structure and bilateral output comovement; ii) two indicators, measuring the number of steps of production embedded in the trade balance and the degree of mismatch between exports and imports, explain between 10% and 50% of the volatility and the cyclicality of net exports.

In the second chapter we develop a multi-industry growth model with oligopolistic competition and variable markups. Our model features a complementarity between capital accumulation and competition, which can give rise to multiple competitive regimes – regimes characterized by a large capital stock and strong competition and regimes featuring low capital and weak competition (*low competition traps*). Negative transitory shocks can trigger a transition from a high to a low competition regime. We also show that, as the firm size/markup distribution becomes more dispersed, the economy is increasingly likely to enter a *low competition trap*. In a calibrated version of our model, a transition from a high to a low competition regime rationalizes important features of the US great recession and its aftermath, such as the persistent drop in output and aggregate TFP, the decline of the labor share, the increase in the profit share, and the decline in the number of firms.

In the third chapter we study how countries which share a common currency potentially have strong incentives to share macroeconomic risks through a system of transfers to compensate for the loss of national monetary policy. However, the option to leave the currency union and regain national monetary policy can place severe limits on the size and persistence of transfers which are feasible inside the union. In this paper, we derive the optimal transfer policy for a currency union as a dynamic contract subject to enforcement constraints, whereby each country has the option to unpeg from the common currency and default completely on any payment obligations. Our analysis confirms that the lack of independent monetary policy is an

important obstacle to risk sharing within a currency union; however, under certain conditions, it is still possible to support substantial macroeconomic stabilization through state contingent international transfers within the union.

Contents

Chapter 1

Global Value Chains and the Business Cycle

Alessandro Ferrari[1]

1.1 Introduction

In the last decades, global production and trade have shifted towards more integrated production chains. The rising importance of global value chains (GVC) increased the degree of interconnectedness among economies. This shift has changed both production and trade. Production now involves a larger fraction of foreign inputs.[1] As a consequence, trade now features predominantly intermediate goods.[2]

[1]European University Institute, Italy. E-mail: *alessandro.ferrari@eui.eu*.I am thankful to my advirsors Ramon Marimon and Philipp Kircher for their invaluable guidance and support. I thank Pol Antràs, Andrea Ciani, Andrea Cintolesi, Russell Cooper, Juan Dolado, Matteo Escudé, Mathijs Janssen, Dalila Figueiredo, Philipp Kircher, Georgios Manalis, Ramon Marimon, Isabelle Méjean, Mathieu Parenti and the participants to various seminars at EUI, Collegio Carlo Alberto, ASSET Conference, RIEF AMSE, WIEM, the EEA 2019 meeting, ETSG 2019, PSE Macro Workshop, Econometric Society European Winter Meeting, CERGE-EI Prague, CSEF Naples, University of Zurich, Queen Mary University London, HEC Montreal, and the ECB Research Department for useful comments and suggestions.

[1]Between 1994 and 2014 the share of foreign value added of gross exports increase from 17 to 27% (see Andrews et al., 2018).

[2]The share of final manufacturing goods trade over total trade has been steadily declining and is now stable below 30% (OECD, WTO and World Bank, 2014).

The increased fragmentation of both domestic and cross border production and its implications for the propagation of shocks has been the subject of a large literature stemming from Acemoglu et al. (2012). The structure of the production network and its frictions have been shown to be a determinant of volatility and absorption of shocks (see Miranda-Pinto, 2019; Huneeus, 2019). Less is known about whether shocks amplify in sequential production setups.

In light of the observed fragmentation of production, in this paper I study how demand shocks amplify upstream (meaning further away from consumers) in production chains. I analyse how this phenomenon, coupled with countries' industrial composition (distribution of sectoral output shares) can partially explain the heterogeneous behaviour (cyclicality and volatility) of trade along the business cycle.

This paper starts from two empirical observations: i) different sectors and different countries position themselves at different stages of global production chains;[3] ii) the length of production chains has been significantly increasing over time. [4] Furthermore, an extensive literature has highlighted how inventories might be a source of amplification of shocks across firms. [5]. Combining these observations, firstly I ask whether shocks travelling in production networks with inventories amplify or dissipate for sectors located at different positions in production chains. Secondly I investigate whether similar patterns can be found at the more aggregate country level, in particular in terms of the cyclical properties of trade.

The key intuition can be summarized as follows: if inventories amplify shocks travelling upstream in production chains we should observe upstream industries responding more to changes in final demand than downstream ones. Secondly, by aggregation, countries whose industries are located more upstream should have different cyclical properties from countries whose industries are relatively close to final consumers.

To study this problem, I build a model of network propagation of exogenous shocks to final demand through procyclical inventory adjustment. The model features a flexible production network structure and inventory adjustment. The theoretical analysis provides a condition under which final demand shocks amplify or dissipate upstream in a production chain. In the model firms face stochastic final demand and hold a fraction of expected demand in inventories. As final demand changes, firms adjust production to meet demand. However, whenever final demand shocks are not independent across time a change in demand today

[3]See Antràs and Chor (2013), Miller and Temurshoev (2017) and the Appendix B for evidence of this empirical regularity.

[4]See Antràs and Chor (2013) and Appendix 1.A.

[5]This phenomenon has been labeled bullwhip effect.See Lee et al. (2004), Metters (1997) and Chen et al. (1999) for theoretical analyses.

provides information on the expected demand tomorrow. This implies that firms will also adjust production to change the stock of inventories. Using inventory data I observe that this adjustment is procyclical, hence I assume so in the model. These changes in production propagate through the network potentially magnifying or dissipating. In particular the pattern of propagation is characterized by the interplay of an inventory (amplification) and a network (amplification/dissipation) effect.

The model shows how the amplification/dissipation patterns depend crucially on two forces: the degree of procyclicality of inventories and the strength of a sector's outward connections. To better exemplify the mechanics, start from a vertically integrated economy. In this setting shocks are passed one to one to input suppliers, hence the network effect is absent. In line production networks the inventory channel is the only one active, which implies that shocks magnify upstream. This intuition carries through in general networks. However, when every node is characterized by an arbitrary number of inward and outward links the network effect can fully undo the inventory amplification channel. In particular, networks characterized by weak linkages will be able to dampen final demand shocks as they travel upstream. On the other hand, networks with few very strong links may amplify them if the connections become stronger further away from consumption.

I empirically test this relationship by means of a shift-share instrument design. The instrument is based on estimated destination specific demand shocks and the fraction of industry output consumed in that destination. This design allows me to study the causal effect of these sector specific shocks and how they generate differential output response depending on the industry's position in production chains. I show that, for a given final demand shock, more upstream sectors display larger output responses than less upstream producers. In particular, I find that sectors located 4 production steps from consumption respond 50-80% more than final goods producers.[6] This finding is robust to the inclusion of network importance measures, past output, a rich set of fixed effect and alternative ways of estimating demand shocks. A similar result is obtained when comparing the variances of output growth for a given variance of demand shocks. Furthermore, leveraging end of the year inventory data for US manufacturing industries I can decompose the effects of distance from consumption and inventories. I find that being further away from final consumers, per se, reduces the responsiveness to shocks; however when high distance is coupled with high inventories along the chain, industries respond more to final demand shocks. The latter channel dominates the former, generating higher responsiveness for more upstream sectors.

Additionally, I document that the higher responsiveness of output of more upstream

[6]Such a shift (from 1 to 4 steps away) corresponds to moving from the 20^{th} to 90^{th} percentile of the distance from consumption distribution.

industries is verified independently of the sign of the shock. However, industries respond more to negative demand shocks than to positive ones.

In light of this amplification result, I study how countries' industry structure and composition can determine cyclical behaviour at the country level. If distance from consumption generates heterogeneous response at the industry level, then countries with different industry compositions should behave differently over the business cycle. I start by addressing a result of multiple papers in the international economics literature (Clark and van Wincoop, 2001; Imbs, 2004; Ng, 2010) showing that economies with similar production structure have higher output comovement. I build on this result by showing that, controlling for the similarity of GVC positioning, the importance of industrial structure is significantly reduced. The interpretation of this result is that industrial structure (measured as sector shares of output) masks a large heterogeneity in the location in value chains, which is unaccounted for by the previous literature.

Furthermore, I show that the observed cross-sectional heterogeneity in countries' volatility and cyclicality of net exports (see Uribe and Schmitt-Grohé, 2017) can be rationalized by studying the GVC position of their trade flows. In particular, countries that import and export upstream show higher volatility of trade. Furthermore, countries that tend to export upstream and import downstream tend to have a higher procyclicality of the trade balance. These measures of GVC position explain between 10% and 50% of the volatility and cyclicality of net exports.

1.2 Literature Review

First, this paper is related to the growing body of research on global value chains and their structures. Notably Antràs and Chor (2013), Antràs et al. (2012) and Antràs and Chor (2018) study both theoretically and empirically the recent developments in the structure of global production. They also provide a set of measures to compute the upstreamness of a sector, defined as the expected distance from final consumption. I build on their findings and on Alfaro et al. (2019b) by extending the measure of upstreamness (distance from consumption) to disentangle bilateral differences and composition effects. Such measure provides information on a sector's upstreamness with respect to a specific destination market. This further decomposition is key to compute the upstreamness of countries' trade flows. If one were not to separate the position versus different destinations, it would not be possible to identify the upstreamness of exports from the one of output. This measure also allows the study of how the same industry is positioned differently depending on the trade counterpart.

Secondly, this paper is close to the literature on inventories as an amplification device. This problem has been studied at several levels of aggregation. Altomonte et al. (2012) use firm level transaction data to show that firms producing intermediate goods have a more pronounced response to the crisis than final goods producers. Zavacka (2012) shows that industry trade flows from US trading partners display volatility which is increasing in the industries' distance from final consumers in response to the crisis. Finally, there is a large body of literature discussing the macroeconomic effect of inventories as a trigger of amplification. Alessandria et al. (2010) show that procyclical inventory adjustments significantly contributes to the propagation and amplification of macroeconomic fluctuations. These papers all consider exogenous variation given by the financial crisis to evaluate the responsiveness of different sectors or firms to the shock, depending on whether they produce intermediate or final goods. They do so by assuming that the crisis is a shock of the same magnitude for all sectors and, hence, any difference in output response is due to the position in the production chain. The methodological approach of this paper allows to dispense with this assumption as the shift-share design constructs sector specific shocks. This implies that one can study the response to shocks of equal magnitude without having to assume the same exposure to a single shock.

The existence and the implication of upstream amplification has been extensively analysed by work in the management and operation research literatures. The underlying mechanism is often thought to be generated by either technology (shipping lags and order batching) or information (compounding forecasting error) frictions. These frictions imply that firms optimally hold stocks of finished or unfinished products.[7] In this paper I borrow the kernel of this literature by modelling the inventory choice in reduced form. This assumption implies that final demand shocks may amplify upstream through procyclical inventory adjustment.

Thirdly, this paper relates to the growing literature on shocks in production networks. From the theoretical standpoint, this line of research, stemming from Carvalho (2010) and Acemoglu et al. (2012) studies the role of network structure in the propagation of idiosyncratic industry level shocks. More recently Baqaee and Farhi (2019) extend the benchmark model to study propagation of idiosyncratic shocks in more general economies. These contributions build models in which, provided input substitutability, shocks always diffuse and dampen. This feature stems from the existence of labor as an outside input. This paper builds a similar model, explicitly allowing for forces generating potential amplification in the network. Furthermore I investigate a related but separate problem. I study how aggregate final demand shocks travel through the network and whether they amplify or dampen. From an empirical standpoint Acemoglu et al. (2016) show that demand shock propagate through linkages between firms and that the extent of sectoral response depends on the centrality of the network. Carvalho et al. (2016), Barrot and Sauvagnat (2016) and Boehm et al. (2019) provide evidence of propagation

[7] For some of the theoretical contributions in this area, see Lee et al. (2004), Metters (1997) and Chen et al. (1999).

of natural disaster shocks in production networks. These papers, due to the identification strategy, have to make assumptions of symmetry across firms or industries in terms of the magnitude of the shock.

At a more aggregate level, di Giovanni and Levchenko (2010) and di Giovanni et al. (2018) show that international trade and vertical linkages are a driver for shocks across borders, thereby increasing business cycle comovement. Relative to this literature, this paper provides an additional channel through which sectoral heterogeneity and trade can be amplification devices for shocks both domestically and across borders. Furthermore, the role of production specialization in explaining business cycle behaviour has been studied, among others by Kohn et al. (2017). They show that cross-country differences in sectoral specialization patterns can explain half of the difference in GDP volatility between emerging and developed economies.These findings are related to a set of the results I present in this paper, particularly on how the mismatch in the production chain positions of exports and imports can explain part of the observed cyclical behaviour of a country's net exports.

Lastly, the cross-country heterogeneity in the volatility and cyclicality behaviour of trade balances has been discussed in the literature. Possible rationales for these differences normally rely on the inability of developing countries to access insurance devices or on the different nature of the shocks they are subject to (see Aguiar and Gopinath, 2007). My analysis shows that the heterogeneity in the position of industries within countries can partially explain the observed differences, providing a complementary rationalization to the existing theories.

The rest of the paper is structured as follows: Section 1.3 provides motivating evidence. Section 3.2 describes the model of amplification through inventories along a value chain. Section 1.5 describes the data used for the empirical analysis. Section 1.6 provides the methodology. Section 1.7 presents the main results of the paper, while Section 3.4 provides a set of robustness checks. Finally, Section 1.9 discusses future steps and concludes.

1.3 Motivating Evidence

In this Section I provide key empirical observations on production chains and inventories that will be critical in disciplining the theoretical model of this paper.

The two main ingredients I study in this paper are position in production chains and inventories. What motivates such analysis is a set of empirical observations on their behaviour and evolution in recent years.

In the last decades modes of production have seen a significant mutation. As highlighted by the World Bank Development Report 2020 a growing share of production is now across boarders. Figure 1.1 shows, on average, how many steps of production a good goes through before reaching final consumption. Throughout the time period covered by the WIOD dataset this measure has been steadily increasing from 2.6 in 2000 to 3.3 in 2014. As shown in Figure 1.8 this is in equal part driven by longer chains growing in importance and by more important chains growing in length. As mentioned in the discussion of the existing literature, one salient feature of our current models of production networks is that as distance from the source of the shock increases so does their dissipation. This theoretical result, combined with the increasing length of production chains, would imply that we are moving towards a world that is more resilient to shocks. This result can be overturned by the existence of inventories as a source of upstream amplification.

Secondly, as shown in Figure 1.2, sectors are very heterogeneous in their position in value chains. This, combined with heterogeneity in countries' industry composition, implies that countries themselves are positioned at largely different points of production chains, as shown in Figure 1.3. In what follows I relate such heterogeneity in position to differences in the cyclical properties of trade at the country level.

Thirdly, this radical shift in the production structure implies an increase in frictions occurring when distinct entities need to coordinate. The literature on supply chain management has highlighted how firms use inventories as a buffer against unexpected changes in demand or input providers missing deliveries. During the same period inventories to sales ratios (computed as end of the year inventories over yearly value of shipments) for US manufacturers have been relatively stable at around 10.5% (NBER CES Manufacturing Industries Dataset). As the literature has noted , inventory investment is procyclical while the inventory to sales ratio is countercyclical.[8] These features imply that inventories can be a source of amplification of shocks in production chains. Estimating non-parametrically the mapping between inventories and sales (after applying the Hodrick-Prescott filter) I obtain an average derivative of .1, as reported in Table 1.1. This estimate suggests that end of the period inventories are an increasing function of yearly sales. This parameter will be a key object in the theoretical model of the next section.

In the next section I outline a model of production chains where firms hold inventories with the goal of studying propagation and amplification/dissipation patterns of changes in final demand.

1.4 Theoretical Framework

I start by building an extended example of demand shock propagation in vertically integrated economies with inventories. The basic intuition is an extension of Zavacka (2012). I then combine the framework with the standard model of production network (see Acemoglu et al.,

[8] See Wen (2011) for a recent contribution on this topic.

2012) to evaluate under which conditions we observe amplification or dissipation depending on the features of the network and the inventory response.

1.4.1 Vertically Integrated Economy

The setup consists of a partial equilibrium model with one final good, whose demand is stochastic, and $N-1$ stages that are sequentially used to produce the final good. Throughout I will use industry, sector and firm interchangeably. The structure of this production network is a line, where stage N provides inputs to stage $N-1$ and so on until stage 0 where goods are consumed.

The demand for each stage n in period t is D_t^n with $n \in \{0, N\}$ and stage 0 demand, which is the final stage, is stochastic and follows an AR(1) with persistence $\rho \in (-1, 1)$ and a positive drift $\bar{D}$. The error terms is distributed according to some finite variance distribution F on a bounded support. $\bar{D}$ is assumed to be large enough relative to the variance of the error so that the demand is never negative.[9] Formally, final demand in period t is

$$D_t^0 = (1-\rho)\bar{D} + \rho D_{t-1}^0 + \epsilon_t,\, \epsilon_t \sim F(0, \sigma).$$

The production function is linear so that for any stage n, if production is Y_t^n, this also represents the demand for stage $n+1$, D_t^{n+1}. This implies $Y_t^n = D_t^{n+1}$.

Stage 0 production is the sum of the final good demand and the change in inventories. Inventories at time t for stage n are denoted by I_t^n.

Firms at stage n form expectations on future demand $\mathbb{E}_t D_{t+1}^n$ and produce to end the period with some target level of inventories $I_t^n = I(\mathbb{E}_t D_{t+1}^n)$. Where $I(\cdot)$ is some non-negative differentiable function that maps expectations on future demand into end of the period inventories. In what follows it will be convenient to discuss inventories in terms of their ratio to sales. To do so denote $I_t^n == \alpha(\mathbb{E}_t D_{t+1}^n)\mathbb{E}_t D_{t+1}^n$, where $= \alpha(\mathbb{E}_t D_{t+1}^n)$ is a positive, differentiable function that represents the inventories to future sales ratio.

Given this setup it is possible to derive how output behaves at every step of production n by solving the economy upward from final demand.

$$\begin{aligned} Y_t^0 &= D_t^0 + I(\mathbb{E}_t D_{t+1}^n) - I(\mathbb{E}_{t-1} D_t^n) \\ &= D_t^0 + \alpha(\mathbb{E}_t D_{t+1}^n)\mathbb{E}_t D_{t+1}^n - \alpha(\mathbb{E}_{t-1} D_t^n)\mathbb{E}_{t-1} D_t^n. \end{aligned} \tag{1.1}$$

Noting that Y_t^0 represents the demand for output of stage 1 goods and that this relationship

[9]The inclusion of the positive drift does not change the inventory problem since for storage the relevant statistic is the first differenced demand.

holds for all n, one can solve the model recursively until stage N.

Asking whether exogenous changes in final demand amplify upstream is effectively comparing $\frac{\partial Y_t^n}{\partial D_t^0}$ and $\frac{\partial Y_t^{n+1}}{\partial D_t^0}$. In particular, amplification occurs if $\frac{\partial Y_t^n}{\partial D_t^0} < \frac{\partial Y_t^{n+1}}{\partial D_t^0}$.

The following proposition formalises the sufficient condition for amplification in this economy.

Proposition 1.1 (Amplification in Vertically Integrated Economies)
A vertically integrated economy with inventories features upstream amplification of positively autocorrelated final demand shocks if the inventory function satisfies

$$0 < \alpha'(x)x + \alpha(x) < \frac{1}{1-\rho}$$

Or, in terms of the $I(\cdot)$ function

$$0 < I'(x) < \frac{1}{1-\rho}$$

Proof. See Appendix 1.B. ■

The first inequality requires the inventory function is increasing. This ensures that as demand increases so do inventories, thereby generating higher demand for a sector's output. The second inequality requires that the function is not "too increasing" relative to the persistence of the process. The second inequality arises because a positive change of demand today implies that the conditional expectation of demand tomorrow is lower than demand today, due to mean reversion. Note that a positive change in demand at t increase the expectation of output at $t+1$ but reduces the expectation at $t+2$. The condition ensures that the first effect dominates the second one. Intuitively, as shocks become arbitrarily close to permanent, the second condition is trivially satisfied and it is enough for inventories to be increasing in expected demand.

The intuition of this result can be summarized as follows. In vertically integrated economies without labour and inventories, changes in final demand are transmitted 1-to-1 upstream as no substitution is allowed across varieties. When such an economy features inventories this result needs not to hold. If inventories are used to smooth production, meaning that $\alpha(\cdot)$ is a decreasing function, shocks can be transmitted less than 1-to-1 as inventories are used as an absorber. However this result is not necessarily true for any $\alpha(\cdot)$ such that $\alpha' < 0$. Shock smoothing occurs when the net inventory change is negative any time demand increases. When the production smoothing motive is strong enough to generate countercyclical inventory investment, i.e. $I' < 0$, shocks dissipate upstream.

The result holds a fortiori in economies in which inventory to expected sales ratios are constant or increasing in expected demand.

As discussed in Section 1.3 the estimated average derivative I' is approximately .10. Given an empirical estimate of the autocorrelation of HP-filtered sales at around .6, the data suggests that this condition is empirically verified.

To retain tractability of the model I henceforth assume that $I(x_t^n) = \alpha x_t^n, \forall n, t$ with $\alpha > 0$. This assumption implies that inventory policies are identical across sectors and that inventories represent a constant fraction of expected demand. Given this assumption, output has a closed form solution:

Lemma 1.1 (Sectoral Output in Vertical Economies)
In a vertical economy with inventory shares α, industry output for a generic sector at distance n from final consumption is

$$Y_t^n = D_t^0 + \alpha\rho \sum_{i=0}^{n} (1 + \alpha(\rho - 1))^i \Delta_t^0.$$

Where $\Delta_t^0 = D_t^0 - D_{t-1}^0$.

Proof. See Appendix 1.B. ■

First note that the term in brackets in the summation is positive if $\alpha < 1/(1-\rho)$. This implies that the second inequality in Proposition 1.1 is satisfied, while the first one is trivially true by $\alpha > 0$. These conditions are typically satisfied in the data, where yearly values of α range between 0 and 50% of next year sales, with an average of 12%.[10] I therefore assume this condition is verified. Furthermore, I assume that $1 + \alpha(\rho - 1) \in [0, 1]$ as this naturally follows from $\rho > 0$ and $\alpha \in [0, 1]$.[11]

In this context the responsiveness of industry n output to a change in final demand is given by

$$\frac{\partial Y_t^n}{\partial D_t^0} = 1 + \alpha\rho \sum_{i=0}^{n} (1 + \alpha(\rho - 1))^i. \tag{1.2}$$

This result states that any shock to the final demand traveling upstream gets magnified at rate $(1 + \alpha\rho)$ for each stage. The operations literature labels this result the *bullwhip effect*.

---

[10] These estimates are from the NBER CES Manufacturing Industries Database for the years 2000-2011.

[11] This assumption is not needed in the context of a vertically integrated production economy but it will become important as the network is generalized to potentially infinitely long chains.

In this model amplification is clearly increasing upstream provided that the final demand process is positively autocorrelated. From Lemma 1.1 it is also apparent than if shocks are i.i.d. no amplification occurs as the second term is zero. Note that in this setting, due to production taking place on a line with only one endpoint, the structure of the network does not play a role in determining the degree of amplification.

In the next section, I extend the model by including labour and allowing for a more general production structure, such that the network itself shapes the degree of propagation of demand shocks.

1.4.2 Network Structure and Amplification

In this section I extend the model to study how the structure of the production network interplays with the inventory amplification mechanism.

In this model the network is characterized by an input requirement matrix A, in which there are possible cycles and self-loops.[12] The network has a terminal node given by final consumption

Assume consumers demand a stochastic number of consumption baskets D_t. This follows the process[13]

$$D_t = (1-\rho)\bar{D} + \rho D_{t-1} + \epsilon_t,\, \epsilon_t \sim F(0,\sigma).$$

The composition of the consumption basket is generated through a Leontief aggregator over varieties

$$D_t = \min_{s \in S} \left\{ \frac{D_{s,t}}{\beta_s} \right\},$$

where S is a finite number of available products and β_s the consumption weight of good s. This formulation implies that $D_{s,t} = \beta_s D_t$ for $D_{s,t}$ solving the expenditure minimization problem.

Firms produce using recipes with fixed input requirements. This is generated through

[12] An example of a cycle is if tires are used to produce trucks and trucks are used to produce tires, formally $\exists r : [A^n]_{rr} > 0, n > 1$. An example of a self loop is if trucks are used in the production of trucks, technically this is the case if some diagonal elements of the input requirement matrix are positive, i.e. $\exists r : [A]_{rr} > 0$.

[13] This is equivalent to having a stochastic income process and linear preferences over the consumption basket.

Leontief production functions

$$Y_{s,t} = \min_{l_{s,t}, M_{s,t}} \left\{ \frac{l_{s,t}}{1-\gamma_s}, \frac{M_{s,t}}{\gamma_s} \right\},$$

where l_s is the labour used by industry s, M_s is the input bundle and γ_s is the input share for sector s. The input bundle is aggregated as

$$M_{s,t} = \min_{r \in R} \left\{ \frac{Y_{rs,t}}{a^{rs}} \right\},$$

where Y_s is the value of output of sector s, Y_{rs} is the value of output of industry r used in sector s production and a^{rs} is an input requirement, namely, the value of Y_{rs} needed for every unit of Y_s in value terms. R is the set of industries directly supplying inputs to sector s. The aggregator function is assumed to have constant returns to scale ($\sum_r a^{rs} = 1$). Input requirements are non-negative in general, meaning $a^{rs} \geq 0$, $\forall r, s$ and positive when the sector r serves as input provider of sector s, $a^{rs} > 0, r \in R$. These conditions imply $a^{rs} < 1$, $\forall r, s$ any time R is not a singleton.

I maintain throughout that firms want to hold a fraction α of expected demand as end of period inventory. This implies that output of final goods producers, denoted by the superscript 0, is

$$Y^0_{s,t} = \beta_s[D_t + \alpha\rho\Delta_t].$$

This also represents the input demand of sector s to its suppliers, once it is rescaled by the input requirement. Hence output of producers 1 step of production removed from consumption is

$$Y^1_{r,t} = \sum_s \gamma_s a^{rs} \left[D^0_t + \alpha\rho \sum_{i=0}^{1} (1 + \alpha(\rho - 1))^i \Delta_t \right].$$

Where $\sum_s \gamma_s a^{rs} = \sum_s \tilde{a}^{rs}$ is the outdegree on a node r, namely the sum of the shares of output of all industries s coming from input r. Iterating forward to generic stage n, and then defining the following object for industry k

$$\chi^n_k \equiv \underbrace{\sum_v \tilde{a}^{kv} \sum_q \tilde{a}^{vq} ... \sum_r \tilde{a}^{or} \sum_s \tilde{a}^{rs} \beta_s}_{n \text{ sums}},$$

then output at stage n is given by

$$Y_{k,t}^n = \chi_k^n \left[D_t + \alpha\rho \sum_{i=0}^{n} (1+\alpha(\rho-1))^i \Delta_t \right]. \tag{1.3}$$

In equation 1.3 the structure of the network is summarized by χ_k^n, while the rest of the equation represents the inventory effect.

In this setup, the effect of a change in contemporaneous demand on output is

$$\frac{\partial Y_{k,t}^n}{\partial D_t} = \chi_k^n \left[1 + \alpha\rho \sum_{i=0}^{n} (1+\alpha(\rho-1))^i \right]. \tag{1.4}$$

Where the first term summarizes the network effect and the second term represents the inventory amplification. Equations 1.4 is a generalization of equation 1.2, which accounts for the network structure.

Finally, assume that firms produce at multiple stages of production, such that

$$Y_{k,t} = \sum_{n=0}^{\infty} Y_{k,t}^n.$$

With these definitions it is possible to characterize sectoral output as a function of the inventory channel and the features of the network

Lemma 1.2 (Sectoral Output)
The sectoral output of a generic industry k is given by

$$Y_{k,t} = \sum_{n=0}^{\infty} \chi_k^n \left[D_t + \alpha\rho \sum_{i=0}^{n} (1+\alpha(\rho-1))^i \Delta_t \right]. \tag{1.5}$$

This can be written in matrix form as

$$Y_{k,t} = \tilde{L}_k B D_t + \alpha\rho \left[\sum_{n=0}^{\infty} \tilde{A}^n \sum_{i=0}^{n} (1+\alpha(\rho-1))^i \right]_k B\Delta_t, \tag{1.6}$$

where B is the $S \times 1$ vector of demand shares and $\tilde{L}_k$ is the k^{th} row of the Leontief inverse, defined as

$$\tilde{L} = [I + \tilde{A} + \tilde{A}^2 + ...] = [I - \tilde{A}]^{-1}.$$

Where $\tilde{A} \equiv A\hat{\Gamma}$ and $\hat{\Gamma} = \text{diag}\{\gamma_1, ..., \gamma_R\}$.
Sectoral output exists non-negative for any α, ρ such that $\alpha(\rho-1) \in [-1, 0]$.

Proof. See Appendix 1.B. ■

A number of features of Lemma 1.2 are worth discussing. First, a direct implication of the Lemma is that this model collapses to the standard characterization of output in production networks when there is no inventory adjustment as the second term in the summation vanishes, i.e. $Y_{k,t} = \tilde{L}_k B D_t$. This occurs whenever there is no inventory adjustment by construction ($\alpha = 0$) or when current shocks do not change expectations on future demand ($\rho = 0$). Secondly note that output might diverge as $n \to \infty$ if $\alpha(\rho - 1) > 0$. Lastly, note that by the assumption made on $\tilde{A}$,[14] as long as $\alpha(\rho - 1) \in [-1, 0]$ additional distance from consumption implies ever decreasing additional output, hence output converges.

Before characterizing the main result on amplification patterns in this economy it is useful to describe the metric of distance from consumption in the general network. In this context one can think of the production structure as a collection of vertical chains of different lengths. Industries have a potentially infinite collection of paths to reach consumption. Distance can then be characterized as weighted average of such production lines. Antràs et al. (2012) label this measure upstreamness. Lemma 1.3 formalises this concept

Lemma 1.3 (Upstreamness)
In a general production network characterised by the Leontief inverse discussed above, with $\alpha(\rho - 1) \in [-1, 0]$, *distance from consumption for some industry* k *is*

$$U_k = \sum_{n=0}^{\infty} (n+1) \frac{Y_k^n}{Y_k}, \tag{1.7}$$

with $U_k \in [1, \infty)$.

Proof. See Appendix 1.B. ■

Note that while $n \in \mathcal{N}$, $U \in [1, \infty)$.

From Lemma 1.2 it is possible to characterize the following comparative statics on the output responsiveness to final demand shocks.

Proposition 1.2 (Comparative Statics)
This proposition formalises the comparative statics on the responsiveness of output to final demand shocks. Denote $\omega = 1 + \alpha(\rho - 1)$.

a) *The effect of change in contemporaneous aggregate demand on sectoral output is given by*

$$\frac{\partial Y_{k,t}}{\partial D_t} = \tilde{L}_k B + \alpha\rho \sum_{n=0}^{\infty} \tilde{A}_k^n \sum_{i=0}^{n} \omega^i B. \tag{1.8}$$

[14] In particular the fact that $\sum_k \tilde{a}^{kv} < 1$, i.e. the assumption that the firm labour share is positive.

b) *Furthermore, a change in the composition of demand, defined as a marginal increase in the s^{th} element of the vector B (β_s), paired with a marginal decrease of the r^{th} element (β_r), changes the output response to aggregate demand as follows:*

$$\Delta_\beta \frac{\partial Y_{k,t}}{\partial D_t} \equiv \frac{\partial}{\partial \beta_s}\frac{\partial Y_{k,t}}{\partial D_t} - \frac{\partial}{\partial \beta_r}\frac{\partial Y_{k,t}}{\partial D_t} = \sum_{n=0}^{\infty}\left[\tilde{A}^n_{ks} - \tilde{A}^n_{kr}\right]\left[1 + \alpha\rho\sum_{i=0}^{n}\omega^i\right]. \tag{1.9}$$

where $\tilde{A}^n_{ks}, \tilde{A}^n_{kr}$ are the elements of $\tilde{A}$ in positions k, s and k, r respectively.

c) *Lastly, a comparative static that changes the structure of the network path from industry k to final consumption, denoted by a new I-O matrix $\tilde{A}'$ leads to a change in the responsiveness of output given by*

$$\Delta_{\tilde{L}} \frac{\partial Y_{k,t}}{\partial D_t} = \sum_{n=0}^{\infty}\left[\tilde{A}'^n_k - \tilde{A}^n_k\right]\left[1 + \alpha\rho\sum_{i=0}^{n}\omega^i\right] B. \tag{1.10}$$

Proof. See Appendix 1.B. ■

The first result in Proposition 1.2 shows that the effect of a change in final demand on sectoral output can be decomposed in two distinct terms. The first one, which is the standard term in production network economies, states that the change in output is a function of the structure of the network and, in particular, of the centrality of the sector. The second term states that an additional response is driven by the behaviour of inventories. The more important inventories are in the economy and the more autocorrelated demand shocks are, the larger the additional effect of changes in demand on output.

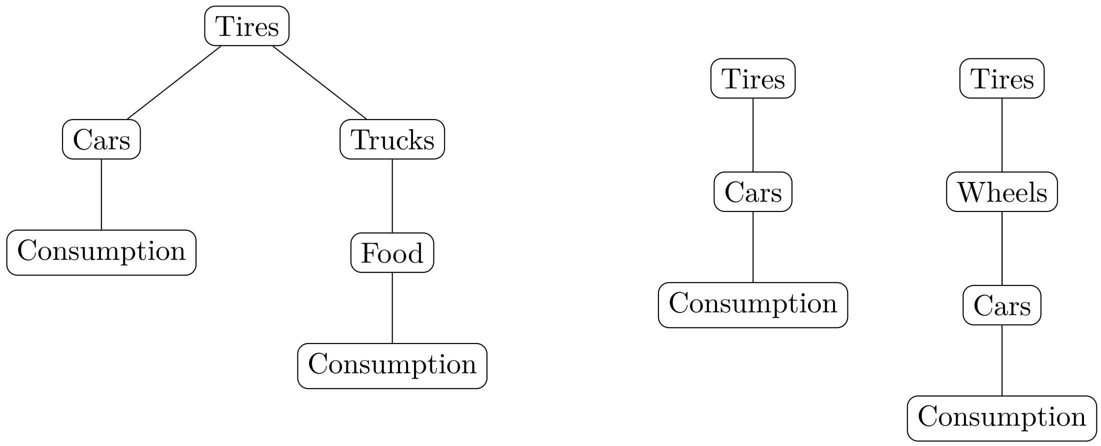
The second result in Proposition 1.2 states that the change in output response to shifts in aggregate demand, following a variation of the demand composition, depends on the relative magnitude of the appropriate elements of the augmented Leontief matrix. Effectively a change in demand composition implies a change in the position of the firm in the network. This, in turn, affects output responsiveness to changes in demand.

The second part of the Proposition characterizes the change in output response following a change in the network. This comparative static is a direct change of the firm's position in the network, hence the change in output response.

These results are best understood via two examples, one changing demand composition in a specific fashion and another changing technology.

Example 1 (Change in Demand Composition)**.** Slightly abusing notation, denote β^n_{ks} the weight of a final good s that is n stages removed from sector k. To exemplify this, think of a tires producer. The output is used to produce cars and trucks, cars are consumed while trucks

Figure 1.1: Comparative Statics



(a) Change in Demand Composition: for given total consumption, food consumption increases and car consumption decreases

(b) Technology Shift: an extra step of production is added to the existing chain

are used as input by the food industry, and food is a final good. In this example tires are at both distance 2 and 3 from consumers. This case is illustrated in Figure 1.1a.

To study how distance from consumption changes the responsiveness of output to final demand shocks, consider an infinitesimal increase in β^{n+1}_{ks} (food in the example) coupled with an equally sized decrease in β^{n}_{kr} (cars). Note that this comparative static implies a marginal increase in the industry upstreamness as more of the sector's output is now used for a longer chain than before. Applying Proposition 1.2

$$\sum_{n=0}^{\infty}\left[\tilde{A}^{n}_{ks}-\tilde{A}^{n}_{kr}\right]\left[1+\alpha\rho\sum_{i=0}^{n}\omega^{i}\right]=\tilde{A}^{n+1}_{ks}\left(1+\alpha\rho\sum_{i=0}^{n+1}\omega^{i}\right)-\tilde{A}^{n}_{kr}\left(1+\alpha\rho\sum_{i=0}^{n}\omega^{i}\right).$$

Assuming that the chains until the affected goods are identical, i.e. $A^{n}_{ks}=A^{n}_{kr}$, then

$$\operatorname{sgn}\Delta_{\beta}\frac{\partial Y_{k,t}}{\partial D_t}=\operatorname{sgn}\left[(1+\alpha\rho\sum_{i=0}^{n}\omega^{i})(\tilde{A}\mathbf{1}_k-1)+\alpha\rho\omega^{n+1}\tilde{A}\mathbf{1}_k\right]. \tag{1.11}$$

Equation (1.11) shows the effect of marginally moving more upstream on the responsiveness of output to demand shocks. The first term on the RHS states that moving more upstream implies exposure of potential dissipation by the network. The second term represents the additional inventory amplification. Depending on which of the two forces prevails, the change in demand will produce further amplification or dissipation.

This result states that if the inventory amplification effect dominates the network dissipation effect, then the change in the demand composition implies shocks will be magnified upstream. This effect is driven by the increase in the sector's distance from final consumers.

The previous example shows that a change in the demand composition affecting the industry position can lead to more or less responsiveness to aggregate demand shifts.

Similarly, an increase in the industry upstreamness generated by the introduction of an additional step in the production chain implies a change in responsiveness of output. This comparative static can be thought of as a new necessary step in the production of some final good. The next example formalises the result.

Example 2 (Technology Shift)**.** The second comparative statics example is the addition of a new step of production. In the case of the tires producer this would be equivalent to moving from a tires-cars-consumption chain to a tires-wheels-cars-consumption one. This is illustrated in Figure 1.1b. Applying Proposition 1.2

$$\sum_{n=0}^{\infty}\left[\tilde{A}'^{n}_{k}-\tilde{A}^{n}_{k}\right]\left[1+\alpha\rho\sum_{i=0}^{n}\omega^{i}\right]B=\sum_{n=0}^{\infty}\tilde{A}^{n}_{k}\left[\left(1+\alpha\rho\sum_{i=0}^{n}\omega^{i}\right)(\tilde{A}\mathbf{1}_{k}-1)+\alpha\rho\omega^{n+1}\tilde{A}\mathbf{1}_{k}\right]B.$$

As in Example 1,

$$\operatorname{sgn}\Delta_{\tilde{L}}\frac{\partial Y_{k,t}}{\partial D_{t}}=\operatorname{sgn}\left[(1+\alpha\rho\sum_{i=0}^{n}\omega^{i})(\tilde{A}\mathbf{1}_{k}-1)+\alpha\rho\omega^{n+1}\tilde{A}\mathbf{1}_{k}\right]. \tag{1.12}$$

The result in Example 2 shows that the same effect on amplification or dissipation of demand shocks can be generated by a change in the demand composition or by a change in the supply chain structure, provided that they alter the firm's position in the same way. Finally, note that the assumption that the outdegree is independent of firm's position is not required for Example 2 since the comparative static is adding a production step to an existing chain.

It is worth discussing how these two examples relate to one another and why they yield equivalent results. The two comparative statics differ in the origin of the variation: in Example 1 there is a change in composition of demand that alters the position of the industry in the production chain; in Example 2 the change is on the technology, or the network. The key assumption such that the two results coincide is the one laid out in Example 1, namely that the outdegree is independent of the stage of production. Without this assumption the two results differ since moving from the shorter path from industry k to consumption to the longer one implies comparing two different sets of outdegrees, as can be seen in equation 1.11. The assumption effectively implies looking at identical paths that only differ in one step of production, which is observationally equivalent to the comparative static in Example 2.

1.4.3 Discussion

The results in examples 1 and 2 state that increasing an industry's distance from consumption could generate amplification or dissipation depending on whether the inventory or the network effect prevails. The latter is ambiguous since the outdegree is potentially larger than 1, which would produce amplification even if firms do not hold inventories (i.e. α=0).

The object $\sum_v \tilde{a}^{kv}$ in the model can be observed in the data. As shown in Figure 1 in the Online Appendix, in the World Input Output Database (WIOD) for the year 2000, the outdegree distribution ranges between 0 and 9.3, with 87% of the sample displaying an outdegree lower than 1. [15]

This implies that most of the industries in the sample lie in the empirically interesting case in which the network can dissipate demand shocks as the distance from consumption increases.

Furthermore, in the WIOD data the correlation between industry upstreamness[16] and outdegree is .3 (see Figure 1.2 in the Appendix 1.I), suggesting that the further from consumption the higher the number of industries served by a given sector. This correlation should suggest that the higher the upstreamness, the more likely it is that the condition in examples 1 and 2 is satisfied.

It is also worth noting that the inventory effect could change the sign if α was negative. This would be evidence of the stock of inventory being used as a precautionary device by firms. Looking at the sample in the NBER CES dataset industries, the inventory-to-future-sales ratio is on average .15, ranging from .02 to .48. This can be thought of as a proxy for α if agents correctly forecast demand.[17] Further details on inventories in the NBER CES dataset are provided in Section 1.J in the Online Appendix.

Furthermore, note that the condition $\omega \in [0, 1]$ is only needed for general networks. In particular such condition implies that amplification occurs at a decaying rate as one moves away from consumption. This is necessary to bound output when chains are allowed to be of infinite length. In the case of line network economies this is only possible if there are infinitely many industries. In the general network case this is true if there are cycles. If the network is restricted to be a Directed Acyclic Graph then such assumption can be dispensed with as boundedness of output is insured by convergence of finite Neumann series.

Lastly, one additional source of heterogeneity that is not modeled in this setup is possible

[15]For the coverage of WIOD data see Tables 1.2 and 1.3 in Appendix 1.C.

[16]This measure is computed as described in the model (see Lemma 1.3). The empirical counterpart, as designed by Antràs et al. (2012), is discussed in Section 1.6.

[17]The sample only includes manufacturing industries, which implies that the estimates for the average α is presumably an upper bound for the WIOD sample, which contains service industries.

heterogeneity in inventory shares (see Online Appendix 1.G for a discussion of this case). Zavacka (2012) reports that, for the sample of manufacturing industries in the NBER CES dataset, the correlation between inventory share and upstreamness is -0.127, implying that industries further away from consumption hold a lower fraction of output in inventories. Note that as the main source of amplification is given by the inventory shares of downstream industries this is still consistent with the result of upstream amplification.

The result in the examples imply that it is empirically unclear whether one should observe output responses that increase or decrease with the distance from consumers.
The remainder of this paper uses the World Input-Output Database to empirically assess the effect of industries' distance from consumption on the responsiveness of output to final demand shocks.

1.5 Data

The main source of data in this paper is the World Input Output Database (2016 release, see Timmer et al., 2015). This contains the Input-Output structure of sector to sector flows for 44 countries from 2000 to 2014 at the yearly level. The data is available at the 2-digit ISIC revision 4 level. The total number of sectors in WIOD is 56. This amounts to 6,071,296 industry to industry flows and 108,416 industry to country flows for every year in the sample. The full coverage of the data in terms of countries and industries is shown in Table 1.2 and 1.3 in the Appendix. Additional data on macroeconomic aggregates of countries is taken from the Penn World Table 9 (see Feenstra et al., 2015).

The structure of the WIOD data is represented in Figure 1.2

Figure 1.2: World Input Output Table

			Input use & value added							**Final use**			**Total use**
			Country 1			...	Country J			Country 1	...	Country J	
			Industry 1	...	Industry S	...	Industry 1	...	Industry S				
Intermediate	Country 1	Industry 1	Z_{11}^{11}	...	Z_{11}^{1S}	...	Z_{1J}^{11}	...	Z_{1J}^{1S}	F_{11}^{1}	...	F_{1J}^{1}	Y_1^1
		...	...	Z_{11}^{rs}	...	...	...	Z_{1J}^{rs}	...	...	...	...	...
		Industry S	Z_{11}^{S1}	...	Z_{11}^{SS}	...	Z_{1J}^{S1}	...	Z_{1J}^{SS}	F_{11}^{S}	...	F_{1J}^{S}	Y_1^S
inputs	...	...	...	...	...	Z_{ij}^{rs}	...	...	...	...	F_{ij}^{r}	...	Y_i^r
	Country J	Industry 1	Z_{J1}^{11}	...	Z_{J1}^{1S}	...	Z_{JJ}^{11}	...	Z_{JJ}^{1S}	F_{J1}^{1}	...	F_{JJ}^{1}	Y_J^1
supplied		...	...	Z_{J1}^{rs}	...	...	...	Z_{JJ}^{rs}	...	...	...	...	...
		Industry S	Z_{J1}^{S1}	...	Z_{J1}^{SS}	...	Z_{JJ}^{S1}	...	Z_{JJ}^{SS}	F_{J1}^{S}	...	F_{JJ}^{S}	Y_J^S
Value added			VA_1^1	...	VA_1^S	VA_j^s	VA_J^1	...	VA_J^S				
Gross output			Y_1^1	...	Y_1^S	Y_j^s	Y_J^1	...	Y_J^S	...			

The World Input-Output Table represents a world economy with J countries and S industries per country. The $(S \times J)$ by $(S \times J)$ matrix whose entries are denoted by Z represents flows of output used by other industries as intermediate inputs. Specifically Z_{ij}^{rs} denotes the output of industry r in country i used as intermediate input by industry s in country j. In addition to the square matrix of input use the table provides the flows of output used for final consumption. These are denoted by F_{ij}^{r}, representing output of industry r in country i consumed by households, government and non-profit organizations in country j. Following the literature I denote $F_i^r = \sum_j F_{ij}^r$, namely output of sector r in country i consumed in any country in the world. By the definition of output, all rows sum to the total production of an industry. Finally the table provides a row vector of value added for every industry, this implies that columns too sum to sectoral output.

This data source is complemented with information about sectoral inventories from the NBER-CES Manufacturing Industry Database. This dataset contains information about sales and end of the period inventories for 473 6-digit 1997 NAICS manufacturing industries from 1958 to 2011.

The next Section describes how this data can be used to construct measures of distance from final consumers both globally and to specific partner countries.

1.6 Methodology

This section describes the empirical methodology used in this paper. I start by reviewing the existing measure of upstreamness as distance from final consumption proposed by Antràs et al. (2012) and then extends it to disentangle the distance from final consumption of a specific partner country. Next, I discuss the identification strategy based on the shift-share design. I show how to compute the sales share in the industry portfolio accounting for indirect linkages.This allows to evaluate the exposure of industry output to specific partner country demands fluctuations even when goods reach their final destination by passing through third countries. Finally, I discuss the fixed effect model used to extract and aggregate country and time specific demand shocks from the final consumption data.

1.6.1 Upstreamness

The measure of upstreamness of each sector counts how many stages of production there are between the industry output and final consumers proposed by Antràs et al. (2012). The index can be thought of as a duration, counting on average the number of intermediate steps between production and consumption. The measure is bounded below by 1, when the entirety of sectoral output is used directly for final consumption.

Antràs et al. (2012) provide a characterization of Upstreamness based on counting the steps between production and consumption. In particular the index is constructed by assigning value 1 to the share of output directly sold to final consumers, value 2 to the share sold to consumers after it was used as intermediate by another industry and so on.[18] Formally:

$$U_i^r = 1 \times \frac{F_i^r}{Y_i^r} + 2 \times \frac{\sum_{s=1}^{S}\sum_{j=1}^{J} a_{ij}^{rs} F_j^s}{Y_i^r} + 3 \times \frac{\sum_{s=1}^{S}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{k=1}^{K} a_{ij}^{rs} a_{jk}^{st} F_k^t}{Y_i^r} + ... \qquad (1.13)$$

where F_i^r is output of sector r in country i consumed anywhere in the world and Y_i^r is the total output of sector r in country i. a_{ij}^{rs} is dollar amount of output of sector r from country i needed to produce one unit of output of sector s in country j, defined as $a_{ij}^{rs} = Z_{ij}^{rs}/Y_j^s$. This formulation of the measure is effectively a weighted average of output, where the weights are the number of steps of production between the specific share of output and final consumption.

Provided that $\sum_i \sum_r a_{ij}^{rs} < 1$, which is a natural assumption given the definition of a_{ij}^{rs} as input requirement[19], this measure can be computed by rewriting it in matrix form:

$$U = \hat{Y}^{-1}[I - A]^{-2}F, \qquad (1.14)$$

where U is a $(J \times S)$ by 1 vector whose entries are the upstreamness measures of every industry in every country. $\hat{Y}^{-1}$ denotes the $(J \times S)$ by $(J \times S)$ diagonal matrix whose diagonal entries are the output values of all industries. The term $[I - A]^{-2}$ is the power of the Leontief inverse, in which A is the $(J \times S)$ by $(J \times S)$ matrix whose entries are all a_{ij}^{rs} and finally the vector F is an $(J \times S)$ by 1 whose entries are the values of the part of industry output that is directly consumed.

[18] This measure is shown to be equivalent to an alternative formulation proposed by Fally (2012). This characterization is based on a recursion such that upstreamness of sector r is computed as 1 plus the weighted upstreamness of industries that use the output of sector r as intermediate input. Formally, the upstreamness for sector r in country i is computed as

$$U_i^r = 1 + \sum_{s=1}^{S}\sum_{j=1}^{J} b_{ij}^{rs} U_j^s,$$

where b_{ij}^{rs} is defined as Z_{ij}^{rs}/Y_i^r. This denotes the dollar amount of sector r output from country i used by industry s output in country j.

[19] For this not to be true one would need that some industry has negative value added since $\sum_i \sum_r a_{ij}^{rs} > 1 \Leftrightarrow \sum_i \sum_r Z_{ij}^{rs}/Y_j^s > 1$, meaning that the sum of all inputs used by industry s in country j is larger than the value of its total output.

As equation 1.13 shows, the value of upstreamness of a specific industry r in country i can only be 1 if all its output is sold to final consumers directly, formally this requires that $Z_{ij}^{rs} = 0, \forall s, j$, which immediately implies that $a_{ij}^{rs} = 0, \forall s, j$. [20]

Table 1.4 provides the list of the most and least upstream industries in the WIOD sample. Predictably services are very close to consumption while raw materials tend to be very upstream.

Appendix 1.K provides additional summary statistics and stylized facts on sectors' and countries' positions in GVCs.

1.6.2 Bilateral Upstreamness

The measure outlined above describes the position of each industry in each country with respect to all countries' final consumers.
In this section I discuss how to construct a similar measure for bilateral flows.

This boils down to restricting the end point of the chains to a specific destination j while still allowing intermediate steps to go through any country in the world. This measure is then, for each industry r in country i to a specific destination country j

$$U_{ij}^{r} = \frac{1 \times F_{ij}^{r} + 2 \times \sum_s \sum_k a_{ik}^{rs} F_{kj}^{s} + 3 \times \sum_s \sum_k \sum_t \sum_m a_{ik}^{rs} a_{km}^{st} F_{mj}^{t} + ...}{F_{ij}^{r} + \sum_s \sum_k a_{ik}^{rs} F_{kj}^{s} + \sum_s \sum_k \sum_t \sum_m a_{ik}^{rs} a_{km}^{st} F_{mj}^{t} + ...} \quad . \tag{1.15}$$

This definition is the bilateral counterpart of equation 1.13. There are two key differences between the two: firstly, F is replaced by F_j, meaning that instead of accounting for global final consumption only chains whose final node is country j consumption are included; secondly, the denominator is not the total output of industry i in country r, this is replaced by the part of sectoral output that will eventually be consumed in country j. As before it is intuitive to think about this measure as a weighted average where the weights are the steps of production.

In matrix form, denote by the subscript $\cdot j$ the upstreamness of the flows from all industries to destination j. The resulting matrix form definition is

$$U_{\cdot j} = C_{\cdot j}^{-1} [I - A]^{-2} F_j. \tag{1.16}$$

Where F_j is the vector of final consumption of country j and $C_{\cdot j}$ is a diagonal square matrix whose diagonal elements are the elements of the vector $[I - A]^{-1} F_j$.

[20] To compute the measure of upstreamness I apply the inventory correction suggested by Antràs et al. (2012), the discussion of the method is left to the Appendix.

1.6.3 Identification Strategy

Given the stated goal to evaluate the responsiveness of output to demand shocks at different levels of distance from consumption, I use a shift-share instrument approach to gauge the causal effect of interest.

This approach, in this setting, boils down to generating plausibly exogenous changes in final demand for producing industries as averages of destination specific aggregate changes weighted by the appropriate measure of exposure.

The methodology, as described in Borusyak et al. (2018), Goldsmith-Pinkham et al. (2018) and Adão et al. (2019) requires exogeneity of either the shares or of the shocks.

In the case analysed in this paper it is implausible to assume the destination shares to be exogenous as firms choose the destinations the serve. Identification can be obtained by plausibly as good as randomly assigned shifters (destination specific shocks).

Define the shift-share changes in final demand for industry r in country i at time t as

$$\hat{\eta}^r_{it} = \sum_j \xi^r_{ijt-1} \hat{\eta}_{jt}. \tag{1.17}$$

Where ξ^r_{ijt-1} represent the fraction of output of industry r in country i consumed directly or indirectly in destination j at time $t-1$ and $\hat{\eta}_{jt}$ the change in final consumption of country j at time t across all products from all destinations.

As shown in Borusyak et al. (2018) the shift-share instrument estimator is consistent provided that the destination specific shocks are conditionally as good as randomly assigned and uncorrelated.

In the next two sections I describe how I compute the destination shares and the destinations shocks.

Sales Shares

The standard measure of sales composition studies the relative shares in a firm's sales represented by different partner countries, see Kramarz et al. (2016). Such a measure however may overlook indirect dependencies through third countries. To exemplify such a problem, take the manufacturing of wood in Canada, the output of this industry can be used both by final consumers and by firms as intermediate input. Assume that half of its production is sold directly to Canadian consumers and the other half to the furniture manufacturing industry in the US. The standard trade data based sales share would state that the sales composition of the industry is split halfway between Canada and the US. This, however, is not necessarily true since the US industry may sell its output back to Canadian consumers. Take the extreme example in which the whole US furniture industry output being exported back to Canada, then the only relevant demand for the Canadian wood manufacturing industry is the one from Canadian consumers.

This example illustrates that, particularly for countries that are very interconnected through trade, measuring portfolio composition only via direct flows may ignore a relevant share of final demand exposure.

Using the Input-Output structure of the data it is possible to account for these indirect links when analysing sales portfolio composition.

Define the share of output of industry r in country i that is eventually consumed by country j as

$$\xi_{ij}^r = \frac{F_{ij}^r + \sum_s \sum_k a_{ik}^{rs} F_{kj}^s + \sum_s \sum_k \sum_t \sum_m a_{ik}^{rs} a_{km}^{st} F_{mj}^t + ...}{Y_i^r}. \tag{1.18}$$

The first term in the numerator represents output from sector r in country i directly consumed by j, the second term accounts for the fraction of output of sector r in i sold to any producer in the world which is then sold to country j for consumption. The same logic applies to higher order terms. By the definition of industry output

$$\sum_j \xi_{ij}^r = 1.$$

As a final remark the next proposition formalises the link between the standard upstreamness measure and the bilateral version, through the sales shares.

Proposition 1.3 (Bilateral Upstreamness)
The upstreamness measure proposed by Antràs et al. (2012) can be obtained as a weighted average of bilateral upstreamness using as weights the bilateral sales portfolio shares.

$$U_i^r = \sum_j \xi_{ij}^r U_{ij}^r. \tag{1.19}$$

Proof. See Appendix 1.B. ■

Hence one could interpret the present discussion as a further decomposition of the standard upstreamness measure based on the portfolio composition and bilateral positioning.

Estimating Demand Shocks

To evaluate what is the total demand innovation that affects a specific industry one needs to estimate country specific demand shocks.

I do so by means of a fixed effect model applied to the change in final consumption.[21] Define the output of industry r in country i that is consumed by country j at time t as F^r_{ijt} and denote f^r_{ijt} its natural logarithm. Then the fixed effects model used to estimate demand innovations takes the following form

$$\Delta f^r_{ijt} = \hat{\eta}_{jt} + \nu^r_{ijt}. \tag{1.20}$$

Where ν^r_{ijt} is a normal distributed error term. The country and time specific demand innovations would then be the series of $\hat{\eta}_{jt}$.[22] This set of fixed effects extracts the change in consumption of destination market j at time t that is common to all sellers.

Recall that the goal is to generate shocks for a specific industry r in country i. Using 1.20 it could be that industry r chooses how much to sell to j and it is a sizeable fraction of j's consumption. Thus, one cannot claim exogeneity of η_{jt} to industry r in country i. To further insure exogeneity, I estimate a different model for every producing country i, specifically

$$\Delta f^r_{kjt} = \eta_{jt}(i) + \nu^r_{kjt} \quad k \neq i \tag{1.21}$$

For each industry r of country i, we need a shock that removes the possible choice mentioned above. Therefore, I estimate country's j fixed effect using all industries of all countries except those of country i.

These can be aggregated as described above into producing industry r effective demand shocks

$$\hat{\eta}^r_{it} = \sum_j \xi^r_{ijt-1} \hat{\eta}_{jt}(i). \tag{1.22}$$

Where the portfolio shares are lagged to eliminate the dependence of portfolio shares themselves on simultaneous demand innovations. This procedure implies that sales from i do not affect $\hat{\eta}_{jt}(i)$ and, therefore, $\hat{\eta}^r_{it}$.

The identification of demand shocks relies on the rationale that the fixed effect model in equation 1.20 captures the variation that is common to all industries when selling to a specific partner country in a given year. The estimation makes the demand shocks exogenous to the producing industry, thereby providing the grounds for causal identification of their effects on output growth.

---

[21] A similar approach is used by Kramarz et al. (2016) and Alfaro et al. (2019a).

[22] Different fixed effects model to estimate demand shocks are used as robustness checks.

1.6.4 Aggregation of Upstreamness Measures

Finally, to evaluate the position of macroeconomic flows in value chains I aggregate the industry level upstreamness using the appropriate weights.

From the industry level bilateral upstreamness it is possible to aggregate it into the following measures

$$
\begin{aligned}
\text{Output} \qquad U_i^Y &= \frac{\sum_{r=1}^S y_i^r \sum_{j=1}^J \xi_{ij}^r U_{ij}^r}{\sum_{r=1}^S y_i^r \sum_{j=1}^J \xi_{ij}^r}, \\
\text{Bilateral Exports} \qquad U_{ij}^X &= \frac{\sum_{r=1}^S \xi_{ij}^r y_i^r U_{ij}^r}{\sum_{r=1}^S \xi_{ij}^r y_i^r}, \\
\text{Total Exports} \qquad U_i^X &= \frac{\sum_{r=1}^S y_i^r \sum_{j\neq i}^J \xi_{ij}^r U_{ij}^r}{\sum_{r=1}^S y_i^r \sum_{j\neq i}^J \xi_{ij}^r}, \\
\text{Bilateral Imports} \qquad U_{ij}^M &= \frac{\sum_{r=1}^S \xi_{ji}^r y_j^r U_{ji}^r}{\sum_{r=1}^S \xi_{ji}^r y_j^r}, \\
\text{Total Imports} \qquad U_i^M &= \frac{\sum_{j\neq i}^J \sum_{r=1}^S \xi_{ji}^r y_j^r U_{ji}^r}{\sum_{j\neq i}^J \sum_{r=1}^S \xi_{ji}^r y_j^r}.
\end{aligned}
$$

Where superscripts Y, X and M denote total output, exports and imports. All these measures are computed at yearly level.

The upstreamness of output is computed by aggregating industry level upstreamness through sectoral output shares. The measures for the flows aggregate the industry upstreamness via the combination of industry output shares and sales portfolio shares. This allows to exclude the part of output that is consumed domestically. The distinction between total and bilateral upstreamness is key for the correct calculation of the trade flows measures.

Given the set of bilateral upstreamness measures it is possible to build two novel indicators for the total steps embedded in a trade balance and the degree of mismatch between what a country exports and what it imports.

The rationale for these two measures are that, given the heterogeneous amplification of shocks along production chains, the distance from consumption of trade flows has implications on the cyclical movement and the volatility of a country's trade balance.

First, I define total upstreamness, unweighted and weighted by trade flows, as

$$
\begin{aligned}
U_{i,j}^{TOT} &= U_{i,j}^X + U_{i,j}^M, \\
{U_{i,j}^{TOT}}_w &= \frac{X_{i,j}U_{i,j}^X + M_{i,j}U_{i,j}^M}{X_{i,j} + M_{i,j}}.
\end{aligned}
$$

This measure contains information about how upstream both flows are.

Second, by taking the difference one can build a measure of mismatch of the upstreamness of exports and imports for any given partner country.

$$U_{i,j}^{NX} = U_{i,j}^{X} - U_{i,j}^{M},$$
$$U_{i,j}^{NX}{}_{w} = \frac{X_{i,j}U_{i,j}^{X} - M_{i,j}U_{i,j}^{M}}{X_{i,j} + M_{i,j}}.$$

I will relate these two indicators to volatility and cyclicality of net exports.

1.7 Results

This section provides the results from the empirical analysis. These consist of a first set of findings regarding how demands shocks amplify along the value chain to industry output. Secondly, I provide evidence for the similarity in GVC positioning of countries' output being the key driver of bilateral output comovement. Lastly, I show that countries differ in their trade balance cyclical behaviour depending on the position of their production and consumption in GVCs.

1.7.1 Demand Shock Amplification and GVC Positioning

The model described in Section 3.2 provides a relationship between final demand shocks and changes in output at different stages of the supply chain. The model suggests that, in absence of network effects, amplification is exponential in distance from consumption (as in the line network). In more complex networks the responsiveness to final demand shocks might dissipate along the production chain if the network dampening effect is strong enough.

To empirically test which effect prevails I use the demand shocks extracted by the fixed effects model in equation 1.21. The estimated outcome is a vector of innovations for every destination country in every period. To aggregate these shocks at the producing industry I use the portfolio shares described above. Using equation 1.22 one has a vector of "relevant" demand shocks at the producing industry time level.

In all the analysis in the remainder of this section I drop values of industry output growth rates larger than 100%. The 98^{th} percentile of the industry growth distribution is 69%. The results are consistent with different cuts of the data and without dropping any entry.

These shocks are positively correlated with the industry output growth rates and explain 23% of their variance, as shown in Table 1.1.

Quantitatively, the estimation suggests that a 1 percentage point increase in the growth rate of final demand produces a .64 increase in the growth rate of industry output.

Table 1.1: Industry Output Growth and Demand Shocks

	(1)
	$\Delta \ln Y_{it}^r$
$\hat{\eta}_{it}^r$	0.641^{***}
	(0.00650)
Constant	0.0700^{***}
	(0.000940)
N	31921
R^2	0.234

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: the table shows the regression of the growth rate of industry output on the weighted demand shocks that the industry receives.

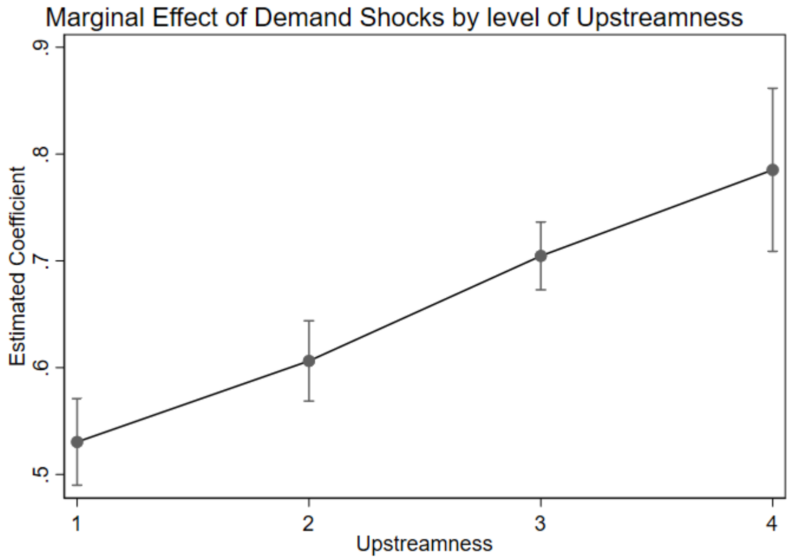
The exogeneity of the estimated demand shocks allows for a causal identification of their effect on output growth. In particular, to test the model prediction I run an econometric model in which the exogenous demand shock can be considered a treatment and the upstreamness level is a moderator of the treatment effect.

I split the upstreamness distribution into dummies taking values equal 1 if $U_i^r \in [1,2]$ and $[2,3]$ and so on. Formally, I estimate

$$\Delta \ln(Y_{it}^r) = \sum_j \beta_j \mathbb{1}\{U_{it}^r \in [j, j+1]\}\hat{\eta}_{it}^r + \nu_{it}^r, \qquad j = \{1,2,3,4\}. \tag{1.23}$$

Since only 2% of the observations are above 5, I include them in the last indicator function, $\mathbb{1}\{U_{it}^r \in [4,\infty)\}$. The resulting coefficients are plotted in Figure 1.3 while the regression output is displayed in Table 1.6 in the Appendix.

Figure 1.3: Effect of Demand Shocks on Output Growth Standard Deviation by Upstreamness Level



Note: the figure shows the marginal effect of demand shocks on industry output changes by industry upstreamness level. The vertical bands show the 95% confidence intervals around the estimates. Note that due to relatively few observations above 5, all values above have been included in the $U \in [4, 5]$ category.

The results suggest that equally sized demand growth rate shock produce largely heterogeneous responses in the growth rate of industry output. Particularly industries located between one and two steps from consumers respond approximately 40% less than industries located 4 or more steps away. This results, which is robust across different fixed effects specifications, highlights how amplification along the production chain can generate sizable heterogeneity in output responses.

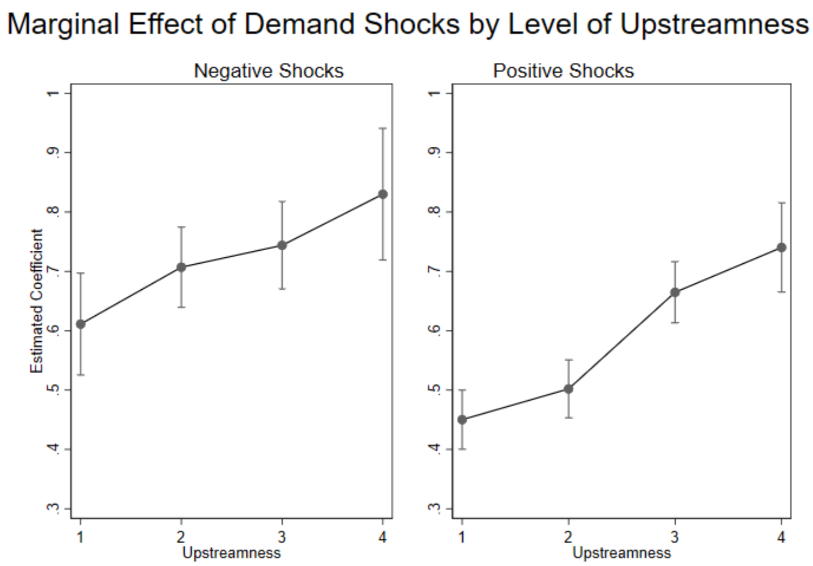
This estimation also suggests that every additional unit of distance from consumption increases the responsiveness of industry output to demand shocks by approximately .09, which represents 14% of the average response.

I further decompose this effect depending on the sign of the demand shock. This analysis aims at studying whether the amplification described above is independent of whether firms receive a positive or negative demand innovation.

Specifically I re-estimate the model by interacting the upstreamness dummies with an indicator for the sign of the shock. The result in Figure 1.4 suggests that amplification takes place in both instances. However sectoral output responds between 10 and 20% more to negative demand shocks for all levels of upstreamness, suggesting an asymmetric effect.

This asymmetry is possibly due to a differential response in terms of network formation and disruption or to heterogeneous constraints in shock absorption capacity. An example of the latter could be firms choosing capacity utilization. In the presence of negative shocks firms can reduce plants utilization, thereby amplifying shocks upstream. When faced with positive shocks firms are bounded above by the existing plants and may be unwilling to pay the fixed cost to permanently increase capacity. Such an asymmetry could produce the observed empirical result.

Figure 1.4: Effect of Demand Shocks on Output Growth Standard Deviation by Upstreamness Level



Note: the figure shows the marginal effect of demand shocks on industry output changes by industry upstreamness level, divided by the sign of the demand shock. The vertical bands show the 95% confidence intervals around the estimates. Note that due to relatively few observations above 5, all values above have been included in the $U \in [4, 5]$ category.

1.7.2 Network Importance

The theoretical model suggests that the degree of amplification or dissipation depends on a combination of industry position and importance in the network. The former, in the model, carries an effect due to inventory amplification. The available data does not allow me to directly test this mechanism. However it is possible to measure the theoretical objects defining the network in the model. In particular it is possible to compute, for every industry, the outdegree and the Leontief inverse coefficient. Defined as

$$outdegree_i^r = \sum_j \sum_s \tilde{a}_{ij}^{rs},$$
$$leontief_i^r = \sum_j \sum_s \tilde{\ell}_{ij}^{rs},$$

where $\tilde{a}_{ij}^{rs}$ is an element of $\tilde{A}$ and $\tilde{\ell}_{ij}^{rs}$ is an element of $\tilde{L}$. The outdegree summarises the intensity of outward connections for the sector, while the leontief cofficient describes the overal importance of the industry in the Input-Output matrix. These can be added to the previous regressions as controls.

The results of the estimation including these network measures is displayed in Table 1.7 in the Appendix. All the conclusions for the baseline estimation are confirmed both qualitatively and quantitatively.

As a second robustness check for the network role, I also estimate the following regression

$$\Delta \ln Y_{it}^r = \beta_1 \hat{\eta}_{it}^r + \beta_2 U_{it}^r \times \hat{\eta}_{it}^r + \beta_3 outdegree_{it}^r \times \hat{\eta}_{it}^r + \beta_4 leontief_{it}^r \times \hat{\eta}_{it}^r + \epsilon_{it}^r.$$

The coefficients of interest are β_2, β_3 and β_4 which show how a sector's position, outdegree and leontief coefficient change the effect of demand shocks on output growth. The results of this estimation are reported in Table 1.2. The results show that the marginal effect of a 1 percentage point change in final demand on the growth rate of output increases by approximately 8 percentage points for every additional upstreamness level. Hence an industry at distance 1 will respond $.40 + .08$ while an industry at distance 2 will respond $.40 + 2 \times .08$. This result is robust to the inclusion of the measures of network importance.

Table 1.2: Marginal Effect of Demand Shocks on Output Growth by Upstreamness Level

	(1) $\Delta \ln Y_{it}^r$	(2) $\Delta \ln Y_{it}^r$	(3) $\Delta \ln Y_{it}^r$	(4) $\Delta \ln Y_{it}^r$
$\hat{\eta}_{i,t}^r$	0.404***	0.404***	0.398***	0.383***
	(0.0324)	(0.0325)	(0.0329)	(0.0332)
$U \times \hat{\eta}_{i,t}^r$	0.0842***	0.0838***	0.0796***	0.0832***
	(0.0114)	(0.0113)	(0.0115)	(0.0115)
$outdegree \times \hat{\eta}_{i,t}^r$		0.00218		-0.0997**
		(0.0163)		(0.0376)
$leontief \times \hat{\eta}_{i,t}^r$			0.00884*	0.0342***
			(0.00467)	(0.0118)
Constant	0.0705***	0.0705***	0.0706***	0.0707***
	(0.00209)	(0.00209)	(0.00210)	(0.00210)
N	31921	31921	31921	31921
R^2	0.239	0.239	0.239	0.239

Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of log industry output on demand shocks both in isolation and interacted with the measure of upstreamness of the industry. Columns 2-4 include the interactions of demand shocks with network importance measures as the sectoral outdegree and the sector's cumulative Leontief inverse coefficient. Standard errors are clustered at the producing industry level.

1.7.3 Inventory Propagation

The evidence presented so far suggests that shocks to final demand amplify upstream in production chains. To test the inventory mechanism more directly I leverage the inventory data of the NBER-CES Manufacturing dataset which contains information on sales and end of the period inventories for US manufacturing industries.

I start by making the assumption that inventories policies are the same within industry across country, formally $\alpha_{it}^r = \alpha_{US,t}^r$, $\forall i$.

I then build a measure of inventory importance in production chains by constructing

$$\tilde{\alpha} = [I - A]^{-1}\alpha$$

This measure accounts for downstream inventories, such that for every industry r in country i

it states what is the inventory share in the entire production chain between the industry and final consumption.

Given this measure, and restricting the sample to manufacturing industries, I estimate the following fully saturated model

$$\Delta \ln Y_{it}^r = \beta_1 \hat{\eta}_{it}^r + \beta_2 U_{it}^r \times \hat{\eta}_{it}^r + \beta_3 U_{it}^r \times \tilde{\alpha}_{it}^r \times \hat{\eta}_{it}^r + \beta_4 U_{it}^r \times \tilde{\alpha}_{it}^r + \beta_5 \alpha_{it}^r \times \hat{\eta}_{it}^r + \beta_6 U_{it}^r + \beta_7 \alpha_{it}^r + \epsilon_{it}^r.$$

The main coefficients of interest in this model are $\beta_1, \beta_2, \beta_3$, where the theoretical model would suggest the following signs $\beta_1 > 0, \beta_2 < 0, \beta_3 > 0$.

Table 1.3 presents the results of the estimation. Columns 1-3 use the direct measure of inventories of the industry α_{it}^r, while columns 4-6 use the chain inventory $\tilde{\alpha}_{it}^r$.

The results confirm the predictions of the theoretical framework. In particular it is possible to retrieve the dissipating role of distance in the network ($\beta_2 < 0$) and the amplifying role of inventories and distance ($\beta_3 > 0$).

Given the previous aggregate result it must be that the inventory propagation mechanism outweighs the dissipating role of distance, thereby generating the patterns shown in Figure 1.3.

1.7.4 Business Cycle and Global Value Chains

In light of the evidence regarding how shocks propagate and amplify in production chains I move to the analysis of how industrial structure and sector position in GVCs can affect countries' business cycle behaviour.

A common finding in cross country studies of bilateral output comovement is that the similarity of industrial structure is a key predictor of bilateral comovement (see Clark and van Wincoop, 2001; Imbs, 2004; Ng, 2010).

The standard measure of similarity in sectoral composition is defined as

$$IS_{ij} = 1 - \sum_r |s_i^r - s_j^r|.$$

Where s_i^r is the industry output share of sector r in country i. This measure evaluates the difference in the sectoral shares of countries' output, however it does not account for within sector heterogeneity and differences in sector positions in production chains (see Figure 1.2 in the Appendix). I build a similar measure for the similarity in GVC positioning by computing

$$US_{ij} = 1 - \frac{1}{S} \sum_r^S \frac{|U_i^r - U_j^r|}{(U_i^r + U_j^r)/2}.$$

Table 1.3: Marginal Effect of Demand Shocks on Output Growth by Upstreamness Level

	(1) $\Delta \ln Y_{it}^r$	(2) $\Delta \ln Y_{it}^r$	(3) $\Delta \ln Y_{it}^r$	(4) $\Delta \ln Y_{it}^r$	(5) $\Delta \ln Y_{it}^r$	(6) $\Delta \ln Y_{it}^r$
$\hat{\eta}_{i,t}^r$	1.326***	1.039***	0.739***	1.291***	1.046***	0.731***
	(0.196)	(0.189)	(0.162)	(0.200)	(0.189)	(0.164)
$U_{i,t}^r \times \hat{\eta}_{i,t}^r$	-0.239***	-0.151**	-0.137***	-0.225***	-0.153***	-0.134***
	(0.0592)	(0.0589)	(0.0493)	(0.0600)	(0.0581)	(0.0495)
$U_{i,t}^r \times \alpha_{i,t}^r \times \hat{\eta}_{i,t}^r$	2.452***	1.923***	1.528***	2.198***	1.824***	1.423***
	(0.444)	(0.442)	(0.367)	(0.419)	(0.405)	(0.343)
$U_{i,t}^r \times \alpha_{i,t}^r$	-0.155**	-0.120	-0.179**	-0.0773	-0.0308	-0.0757
	(0.0751)	(0.0865)	(0.0815)	(0.0697)	(0.0733)	(0.0697)
$\alpha_{i,t}^r \times \hat{\eta}_{i,t}^r$	-6.425***	-4.860***	-4.122***	-5.826***	-4.651***	-3.868***
	(1.387)	(1.346)	(1.147)	(1.341)	(1.277)	(1.101)
$U_{i,t}^r$	0.0381***	0.132***	0.0841***	0.0291***	0.122***	0.0742***
	(0.0108)	(0.0153)	(0.0152)	(0.0108)	(0.0145)	(0.0147)
$\alpha_{i,t}^r$	0.364	0.0934	0.00769			
	(0.265)	(0.303)	(0.298)			
$\tilde{\alpha}_{i,t}^r$				0.158	-0.134	-0.186
				(0.252)	(0.269)	(0.274)
Constant	-0.0359	-0.332***	-0.136***	-0.0109	-0.306***	-0.116**
	(0.0389)	(0.0520)	(0.0523)	(0.0392)	(0.0501)	(0.0514)
Time FE	No	No	Yes	No	No	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes
N	8961	8961	8961	8961	8961	8961
R^2	0.300	0.454	0.513	0.299	0.453	0.513

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of log industry output on demand shocks both in isolation and interacted with the measure of upstreamness and inventories the industry. Columns 1-3 use the industry specific inventory while columns 4-6 use the production chain inventories. Columns 2 and 5 include industry fixed effects, columns 3 and 6 include industry and time fixed effects.

The difference is rescaled by the pairwise mean so that $US_{ij} \in (-1, 1]$ and high values correspond to similar positioning of sectors.

I then estimate the importance of the two measures of similarity in predicting the degree comovement in the cyclical components of output by running

$$\rho_{ij} = \beta_1 IS_{ij} + \beta_2 TI_{ij} + \beta_3 US_{ij} + \gamma_i + \gamma_j + \epsilon_{ij}.$$

Where ρ_{ij} is the correlation between the cyclical component of output of country i and country j and $TI_{ij} = \frac{X_{ij}+M_{ij}}{Y_i+Y_j}$ is a commonly used measure of bilateral trade intensity.
The results, shown in Table 1.4, suggest that the predictive power of the measure of industrial structure similarity vanishes when the regression is augmented with the index of position similarity. This evidence highlights how the position of a country's industries in production chains is a more relevant indicator of bilateral comovement. In Columns (5) and (6) I add the

Table 1.4: Comovement and Industry Structure

	(1)	(2)	(3)	(4)	(5)	(6)
	$\rho_{i,j}$	$\rho_{i,j}$	$\rho_{i,j}$	$\rho_{i,j}$	$\rho_{i,j}$	$\rho_{i,j}$
$IS_{i,j}$	0.171***	-0.0234	0.313***	0.0514	-1.621***	-0.172
	(0.0403)	(0.0401)	(0.0735)	(0.0735)	(0.306)	(0.266)
$TI_{i,j}$	15.27***	12.45***	10.96***	8.602***	10.36***	8.361***
	(2.347)	(2.174)	(2.055)	(1.947)	(2.181)	(1.967)
$US_{i,j}$		1.068***		1.554***	0.357**	1.458***
		(0.0823)		(0.147)	(0.158)	(0.184)
$US_{i,j} \cdot IS_{i,j}$					2.137***	0.297
					(0.406)	(0.340)
Constant	0.586***	-0.139**			0.380***	
	(0.0159)	(0.0578)			(0.114)	
Country FE	No	No	Yes	Yes	No	Yes
N	946	946	944	944	946	944
R^2	0.0796	0.219	0.587	0.634	0.241	0.635

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of the bilateral comovement of the cyclical component of output over measures of industry composition and upstreamness similarity between countries. Columns 3 and 4 include 2 sets of country fixed effects.

interaction term between the two measures of similarities. Such inclusion shows that, in the specification without country fixed effects, the industry composition metric turns negative, highlighting that the positive effect of similarity and comovement mostly runs through its joint effect with the measure of positioning. The result in Column (6) suggests that the industry similarity measure now captures possible substitutability between the two countries' outputs, whereas the complementarity, that drives the positive comovement, is absorbed by the interaction. Lastly, when one includes country fixed effects both the IS measure and its interaction with US are not statistically significant.

1.7.5 Global Value Chains and Trade along the Business Cycle

Building on these results I study how industry composition and GVC position can shed some light on the observed cross-country heterogeneity in trade balances behaviour over the business cycle.

Given the previous discussion on the amplification of shocks upstream in a value chain, two main facts should be found in the data:

1. Countries with higher total upstreamness should display higher volatility. Fixing the covariance between export and import, for a given demand shock, higher upstreamness implies higher response.

2. Countries with higher net upstreamness should display more procyclical trade balances. For a given global demand shock the response of the more upstream flow should be larger than the less upstream flow one. This implies that with positive net upstreamness, exports should respond more than imports, generating a more procyclical trade balance.

In this analysis I use the country aggregated indicators described in Section 1.6.4. I average them across years to study their relation with volatility and cyclicality measures.

To evaluate the first potential relationship I regress the log of the standard deviation of a country's trade balance on the log of its trade balance total upstreamness, specifically

$$\ln \sigma_i^{NX} = \beta_0 + \beta_1 \ln U_i^{TOT} + \epsilon_i. \tag{1.24}$$

The result of this estimation is displayed in Table 1.5 and in Figure 1.4 in the Appendix. The regression shows a positive correlation, with an estimated effect of 7% increased volatility for a 1% increase in the total upstreamness of the trade balance. To check that this correlation is not entirely driven by a country's development level, I add log per capita GDP and the result remains consistent. The upstreamness measure explains 25% of the observed cross-sectional variability in net exports volatility.

The results hold when using total upstreamness weighted by trade flows. It is also worth mentioning that the inclusion of the index of net upstreamness does not change the results, highlighting how it is really the total steps embedded in the trade balance that correlates with its volatility.

Table 1.5: Volatility and Total Upstreamness

	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln \sigma_{nx}$	$\ln \sigma_{nx}$	$\ln \sigma_{nx}$	$\ln \sigma_{nx}$	$\ln \sigma_{nx}$	$\ln \sigma_{nx}$
$\ln U^{TOT}$	7.661***	8.227***	7.626***			
	(2.199)	(2.604)	(2.323)			
log per capita income		0.200			0.186	
		(0.376)			(0.385)	
U^{NX}			0.0544			
			(0.624)			
$\ln U^{TOT}_w$				7.620***	8.133***	7.048***
				(2.224)	(2.539)	(2.229)
U^{NX}_w						0.575
						(0.453)
Constant	-6.733	-9.856	-6.659	-1.376	-3.898	-0.634
	(4.251)	(7.489)	(4.495)	(2.770)	(5.953)	(2.766)
N	44	44	44	44	44	44
R^2	0.244	0.251	0.244	0.243	0.249	0.273

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of the log of the standard deviation of a country's detrended trade balance on the log of total upstreamness. Column (2) adds log per capita income as a control, while Column (3) includes the net upstreamness measure. Columns 4-6 replicate the analysis with the weighted upstreamness measures.

A similar approach is taken for the second relation, regressing the country specific correlation between net exports and output, both detrended, over the measure of trade balance net upstreamness.

$$\rho_i(NX, Y) = \beta_0 + \beta_1 U_i^{NX} + \epsilon_i. \tag{1.25}$$

The results are displayed in Table 1.6 and in Figure 1.5 in the Appendix. The correlation between the two measures is positive, suggesting that indeed a higher positive mismatch between the position of exported and imported good can affect the degree of procyclicality of the trade balance. In particular the regression shows that a 1 point increase in net upstreamness of trade may increase the cyclicality of net exports between .47 and .58. These results imply that a one standard deviation increase in net upstreamness implies a 1/3 standard deviation increase in the degree of procyclicality of the trade balance. The net upstreamness measure is able to explain 10% of the variance of the trade balance and output correlations. The relation is again robust to controlling for the degree of development of the country. In this case

Table 1.6: Cyclicality and Net Upstreamness

	(1)	(2)	(3)	(4)	(5)	(6)
	$\rho_{nx,y}$	$\rho_{nx,y}$	$\rho_{nx,y}$	$\rho_{nx,y}$	$\rho_{nx,y}$	$\rho_{nx,y}$
U^{NX}	0.561***	0.469**	0.580***			
	(0.195)	(0.235)	(0.218)			
log per capita income		0.159			0.0293	
		(0.158)			(0.0956)	
U^{TOT}			-0.0547			
			(0.145)			
U_w^{NX}				0.955***	0.941***	1.002***
				(0.139)	(0.144)	(0.131)
U_w^{TOT}						-0.291
						(0.206)
Constant	0.00327	-1.619	0.388	-0.0254	-0.322	0.996
	(0.0902)	(1.612)	(1.036)	(0.0645)	(0.967)	(0.736)
N	44	44	44	44	44	44
R^2	0.0987	0.127	0.102	0.505	0.506	0.530

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of the correlation between a country's detrended trade balance and its detrended output on the measure of net upstreamness. Column (2) adds log per capita income as a control, while Column (3) includes the total upstreamness measure. Columns 4-6 replicate the analysis with the weighted upstreamness measures.

using the weighted version of the net upstramness measure changes the results quantitatively. In particular with this index the effect of a 1 point increase generates approximately a 1 point increase in the cyclicality. This can be read as a 1 standard deviation increase in the weighted net upstreamness implies a 3/4 of a standard deviation increase in procylicality. The explanatory power of this measure also increases significantly to approximately 50% of the observed cross country variation.

Finally it is worth mentioning that these results are robust to the inclusion of total upstreamness, suggesting that the mismatch dimension is the one explaining the variation in the data.

1.8 Robustness Checks

In this section I provide a set of robustness tests for the analysis of upstream amplification of shocks.

1.8.1 Ordinal Effects of Upstreamness

First I estimate a similar model to the main specification in the results section but I use ordinal measures from the upstreamness distribution. Namely I interact the industry level shocks with dummies taking value 1 if an industry belongs to an upstreamness decile. Formally the estimated model is

$$\Delta \ln(Y_{it}^r) = \sum_j \beta_j \mathbb{1}\{U_{it}^r \in D_j\}\hat{\eta}_{it}^r + \nu_{it}^r \qquad , j = \{1...10\}. \tag{1.26}$$

Where D_j denotes the mass between $j-1^{th}$ and the j^{th} deciles of the upstreamness distribution. The results are shown in Column 1 of Table 1.8 in the Appendix. The estimation suggests that moving upward in production chains increases the responsiveness of output to final demand shocks. The effect almost doubles when moving from the first to the last decile. This corresponds to moving from 1.17 to 4.37 production stages away from final demand.

As in the main specification, the results suggest that the output response to demand shocks increase with distance from consumption. Ordinally the estimation states that moving from the first to the last decile of the distribution implies an increase in the output response from .49 to .76 percentage points. Note that all the results in this section are robust to the inclusion of industry, country and upstreamness decile fixed effects.

Secondly, I run a model in which instead of using industry output growth rates I use their standard deviation over time, regressed on the standard deviation of the relevant final demand shocks. Formally

$$\sigma_{\Delta \ln(Y_{it}^r)} = \sum_j \beta_j \mathbb{1}\{U_{it}^r \in D_j\}\sigma_{\hat{\eta}_{it}^r} + \nu_i^r \qquad , j = \{1...10\}. \tag{1.27}$$

The results,displayed in Column 2 of Table 1.8 in the Appendix, suggest that relationship still holds for the standard deviations. In particular, moving from the first to the last decile of the upstreamness distribution entails a change in the effect of one point of the standard deviation of shocks on the standard deviation of output growth from .71 to 1.22. Quantitatively, the estimation suggests that the standard deviation of growth increases of 0.04 for every decile of upstreamness. The average standard deviation of output growth in the sample is .16, which implies that moving upward between any upstreamness decile produces a 25% increase in output standard deviation. This is also equivalent to half a standard deviation of the outcome.

1.8.2 Alternative Fixed Effects Models

In the previous section, I used the fixed effect model used to gauge the idiosyncratic demand shocks. Such model may be confounding other sources of variation. To inspect this possibility I use two alternative econometric models to extract the demand shocks.
In the first one I follow more closely Kramarz et al. (2016) and include producer fixed effects, γ^r_{it} is the fixed effect for the producing industry r in country i at time t, namely

$$\Delta f^r_{ijt} = \gamma^r_{it} + \eta_{jt} + \delta_t + \nu^r_{ijt}. \tag{1.28}$$

The third alternative, that is closer to the specification used in the main results, uses only partner country year fixed effect but excludes domestic industries, formally

$$\Delta f^r_{ijt} = \eta_{jt} + \delta_t + \nu^r_{ijt}, \quad \forall i \neq j. \tag{1.29}$$

The condition that $i \neq j$ ensures that domestically produced goods used for final consumption are not included in the estimation. The underlying rationale is that these industries would be the ones whose non demand related variation (think of supply shocks) may be highly correlated with demand shocks themselves.

The results of these two procedures for the cardinal effect of upstreamness (equation 1.23) are presented in Table 1.9.

The results remain consistent with the previous findings. When the producing industry variation is absorbed upon computing the demand shocks the relationship between the effect of the shocks and upstreamness flattens out at high distance from consumption. This can be seen by the relatively small difference between the effect of shocks at upstreamness between 3 and 4 and above 4. The opposite result is observed when excluding domestic industries in computing demand shocks. The relationship becomes steeper.

As a further robustness check I include a different set of fixed effects in the estimation. Namely I include year, producing country, producing industry and upstreamness level fixed effects. The results are displayed in Table 1.10 in the Appendix.

The inclusion of these additional sets of fixed effects changes the magnitude of the results, reducing the effect of a 1pp demand shock from .47pp to .25pp for the industries with upstreamness between 1 and 2 and similarly for all other levels. The qualitative result however remains robust in that industries' responsiveness to demand shocks remains ranked according to distance from consumption. All the specifications suggest that for the same shock industries very far from consumption respond between 1.5 and 2 times as much as industries close to final consumers.

I propose two additional robustness checks that use shocks estimated with different specifications. The first analysis employs shocks to log final consumption, rather than to the growth rate of final consumption and estimates the elasticity of different industries to demand shocks. Formally I estimate

$$f_{ijt}^{r} = \gamma_{it}^{r} + \eta_{jt} + \delta_t + \nu_{ijt}^{r},$$

where f_{ijt}^{r} is the log of final consumption in destination country j of output from industry r in country i. This specification implies that the estimated destination-time specific innovations are in terms of log consumption. I then aggregate these innovations at the industry level and use them to estimate

$$\ln(Y_{it}^{r}) = \sum_{j} \beta_j \mathbb{1}\{U_{it}^{r} \in [j, j+1]\} \hat{\eta}_{it}^{r} + \nu_{it}^{r}, \qquad j = \{1, 2, 3, 4\}.$$

In this context the estimated β_j represent the different elasticities of output to demand changes.

The last method I use to test the robustness of the results consists of using the methodology employed in the main result to estimate demand shocks (estimated on growth rates) and then using the base year of final consumption to determine the level of the shock. In other words I construct, for every destination country $\hat{F}_{jt} = F_{jt-1}\hat{\eta}_{jt}$, where $\hat{\eta}_{jt}$ is the exogenous component of the growth rate and F_{jt-1} is the level in the first year of the sample. One can then have level innovations of demand for every industry, once appropriately aggregated through portfolio weights. I then run the same log-log specification to estimate the elasticity of industries at different levels of production to demand shocks.

The results of these two robustness checks are displayed in Tables 1.11 and 1.12 in the Appendix, respectively. Qualitatively they confirm the increasing elasticity of output to demand shocks once one moves further away from consumers. These results are robust to the inclusion of several sets of fixed effects, thereby assessing only within variation.

As a further robustness I use the deflated version of the WIOD dataset (Los et al., 2014, see) and replicate the entire industry level empirical analysis to test whether price movements could possibly be responsible for the results discussed above. The results of this check are displayed in Table 1.13 in the Appendix. The results in the main estimation are confirmed both qualitatively and quantitatively.

Lastly, as discussed in previous work studying the effect of demand shocks and their propagation in the network (see Acemoglu et al., 2016) I include lags of the output growth rate. The results of the estimation are shown in Table 1.14 and Figure 1.6 in the Appendix. This robustness check confirms the results of the main estimation both qualitatively and quantitatively.

1.9 Conclusions

This paper starts from the premise that firms and sectors position themselves at different stages of production chains. I model this aspect, together with a flexible production network structure and procyclical inventory adjustment, to show that demand shocks can amplify or dissipate in the network. Two potentially counteracting effects are at play in this model. First, procyclical inventory adjustment can produce amplification of demand shocks along the production chain. Secondly, the structure of the network can either dissipate or amplify shocks.

In particular, if the network features small outdegrees (smaller than 1), it may be able to dissipate demand shocks travelling upstream, provided that the inventory amplification channel is relatively small. On the other hand, networks featuring nodes with high outward connections (high outdegree) may strengthen the amplification generated through inventories.

Then, I empirically test the demand shock propagation using data from the World Input-Output Database. I apply a shift-share instrumental design, using as exogenous demand shocks the destination country-time specific variation across all selling industries and aggregate them using industries' sales portfolio shares.

Regressing sectoral output growth on these industry specific shocks, I find that moving from upstreamness between 1 and 2 to upstreamness above 4, implies an output response to a 1 percentage point demand change from .53 to .78 percentage points.

Furthermore splitting the sample in upstreamness deciles and using as outcome the standard deviation of output growth, I find that moving from the first to the last decile of the distribution implies a 71% increase in the volatility of output growth.

These results provide evidence for amplification of demand shocks travelling upstream in production chains. Through the lenses of the model, this can be interpreted as either the network effect amplifying shocks or the inventory channel overturning the network dissipation effect. These results remain unchanged when controlling for measures of network importance. Hence one can conclude that the observed heterogeneity in the elasticity of output to demand shocks is driven solely by the position in the production chain. Furthermore, leveraging inventory data for US manufacturing industries I show that the empirical effect can be decomposed into a dissipating role of the network and an amplifying role of inventories.

Given these findings, I study how countries' industrial structure composition and positioning affects their business cycle behaviour.

Firstly, I show that controlling for an index of bilateral similarity in countries' GVC position eliminates the correlation between bilateral comovement in cyclical output and measures of sectoral composition similarity.

The last result of this paper relates two novel indicators of a country's trade balance with its cyclicality and volatility. In particular, using a measure of how many steps of production are embedded in a country's net exports, I show that a 1% increase in this index correlates with 8% higher trade balance volatility. This measure explains approximately 25% of the trade balance volatility. Secondly, using a measure of the mismatch between the upstreamness of exports and the one of imports, I show that a 1 standard deviation increase in such measure correlates with an increase in trade balance cyclicality between 1/3 and 3/4 of a standard deviation, explaining between 10% and 50% of its variation alone. This result follows from the intuition that, since export and imports enter the trade balance with opposite signs, and, since higher upstreamness implies higher responsiveness to shocks, a country with high net upstreamness is expected to have a more procyclical trade balance, ceteris paribus.

Bibliography

Acemoglu, Daron, Ufuk Akcigit, and William Kerr, "Networks and the Macroeconomy: An Empirical Exploration," *NBER Macroeconomics Annual*, 2016, *30* (1), 273–335.

_ , Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "The Network Origins of Aggregate Fluctuations," *Econometrica*, 2012, *80* (5), 1977–2016.

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales, "Shift-Share Designs: Theory and Inference," *The Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.

Aguiar, Mark and Gita Gopinath, "Emerging Market Business Cycles: The Cycle Is the Trend," *Journal of Political Economy*, 2007, *115*, 69–102.

Alessandria, George, Joseph P Kaboski, and Virgiliu Midrigan, "The Great Trade Collapse of 2008-09: An Inventory Adjustment?," *IMF Economic Review*, December 2010, *58* (2), 254–294.

Alfaro, Laura, Manuel García-Santana, and Enrique Moral-Benito, "On the direct and indirect real effects of credit supply shocks," Technical Report, National Bureau of Economic Research 2019.

_ , Pol Antràs, Davin Chor, and Paola Conconi, "Internalizing global value chains: A firm-level analysis," *Journal of Political Economy*, 2019, *127* (2), 508–559.

Altomonte, Carlo, Filippo Di Mauro, Gianmarco I. P. Ottaviano, Armando Rungi, and Vincent Vicard, "Global Value Chains During the Great Trade Collapse: A Bullwhip Effect?," CEP Discussion Papers dp1131, Centre for Economic Performance, LSE February 2012.

Andrews, Dan, Peter Gal, and William Witheridge, "A genie in a bottle?: Globalisation, competition and inflation," OECD Economics Department Working Papers 1462, OECD Publishing March 2018.

Antràs, Pol and Davin Chor, "Organizing the Global Value Chain," *Econometrica*, November 2013, *81* (6), 2127–2204.

__ **and** __ , "On the Measurement of Upstreamness and Downstreamness in Global Value Chains," NBER Working Papers 24185, National Bureau of Economic Research, Inc January 2018.

__ **,** __ **, Thibault Fally, and Russell Hillberry**, "Measuring the Upstreamness of Production and Trade Flows," *American Economic Review*, May 2012, *102* (3), 412–416.

Axtell, Robert L., "Zipf Distribution of U.S. Firm Sizes," *Science*, 2001, *293* (5536), 1818–1820.

Baldwin, R.E., *The Great Trade Collapse: Causes, Consequences and Prospects* A VoxEU.org publication, Centre for Economic Policy Research, 2011.

Baqaee, David Rezza and Emmanuel Farhi, "The macroeconomic impact of microeconomic shocks: beyond Hulten's Theorem," *Econometrica*, 2019, *87* (4), 1155–1203.

Barrot, Jean-Noël and Julien Sauvagnat, "Input specificity and the propagation of idiosyncratic shocks in production networks," *The Quarterly Journal of Economics*, 2016, *131* (3), 1543–1592.

Boehm, Christoph E, Aaron Flaaen, and Nitya Pandalai-Nayar, "Input linkages and the transmission of shocks: Firm-level evidence from the 2011 Tōhoku earthquake," *Review of Economics and Statistics*, 2019, *101* (1), 60–75.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-Experimental Shift-Share Research Designs," Working Paper 24997, National Bureau of Economic Research September 2018.

Carvalho, Vasco M, "Aggregate Fluctuations and the Network Structure oF Intersectoral Trade," *mimeo*, 2010.

__ **, Makoto Nirei, Yukiko Saito, and Alireza Tahbaz-Salehi**, "Supply chain disruptions: Evidence from the great east Japan earthquake," *Columbia Business School Research Paper*, 2016, (17-5).

Chen, Frank, Zvi Drezner, Jennifer Ryan, and David Simchi-levi, "Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information," *Management Science*, 07 1999, *46.*

Clark, Todd E. and Eric van Wincoop, "Borders and business cycles," *Journal of International Economics*, 2001, *55* (1), 59 – 85.

di Giovanni, Julian and Andrei A. Levchenko, "Putting the Parts Together: Trade, Vertical Linkages, and Business Cycle Comovement," *American Economic Journal: Macroeconomics*, April 2010, *2* (2), 95–124.

_ , _ , and Isabelle Mejean, "The Micro Origins of International Business-Cycle Comovement," *American Economic Review*, January 2018, *108* (1), 82–108.

_ , _ , and Romain Rancière, "Power laws in firm size and openness to trade: Measurement and implications," *Journal of International Economics*, September 2011, *85* (1), 42–52.

Fally, Thibault, "Production Staging: Measurement and Facts," *mimeo*, 01 2012.

Feenstra, Robert C, Robert Inklaar, and Marcel P Timmer, "The next generation of the Penn World Table," *American economic review*, 2015, *105* (10), 3150–82.

Foster, Lucia, John C. Haltiwanger, and C. J. Krizan, "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," in "New Developments in Productivity Analysis" NBER Chapters, National Bureau of Economic Research, Inc, August 2001, pp. 303–372.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," Working Paper 24408, National Bureau of Economic Research March 2018.

Huneeus, Federico, "Production Network Dynamics and the Propagation of Shocks," *mimeo*, February 2019.

Imbs, Jean, "Trade, Finance, Specialization, and Synchronization," *The Review of Economics and Statistics*, August 2004, *86* (3), 723–734.

Kohn, David, Fernando Leibovici, and Hakon Tretvoll, "Trade in Commodities and Emerging Market Business Cycles," 2017 Meeting Papers 743, Society for Economic Dynamics 2017.

Kramarz, Francis, Julien Martin, and Isabelle Mejean, "Volatility in the Small and in the Large: The Lack of Diversification in International Trade," CEPR Discussion Papers 11534, C.E.P.R. Discussion Papers September 2016.

Lee, L. Hau, V Padmanabhan, and Seungjin Whang, "Information Distortion in a Supply Chain: The Bullwhip Effect," *Management Science*, 12 2004, *43*, 546–558.

Los, Bart, Reitze Gouma, Marcel Timmer, and Pieter IJtsma, "Note on the construction of wiots in previous years prices," Technical Report 2014.

Metters, Richard, "Quantifying the Bullwhip Effect in Supply Chains," *Journal of Operations Management*, 05 1997, *15*, 89–100.

Miller, Ronald E. and Umed Temurshoev, "Output Upstreamness and Input Downstreamness of Industries/Countries in World Production," *International Regional Science Review*, 2017, *40* (5), 443–475.

Miranda-Pinto, Jorge, "Production Network Structure, Service Share, and Aggregate Volatility," *mimeo*, February 2019.

Ng, Eric C.Y., "Production fragmentation and business-cycle comovement," *Journal of International Economics*, 2010, *82* (1), 1 – 14.

OECD, WTO, and World Bank, "Global Value Chains: Challenges, Opportunities, and Implications for Policy," in "" 2014.

Timmer, Marcel P, Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J De Vries, "An illustrated user guide to the world input–output database: the case of global automotive production," *Review of International Economics*, 2015, *23* (3), 575–605.

Uribe, M. and S. Schmitt-Grohé, *Open Economy Macroeconomics*, Princeton University Press, 2017.

Wen, Yi, "Input and Output Inventory Dynamics," *American Economic Journal: Macroeconomics*, October 2011, *3* (4), 181–212.

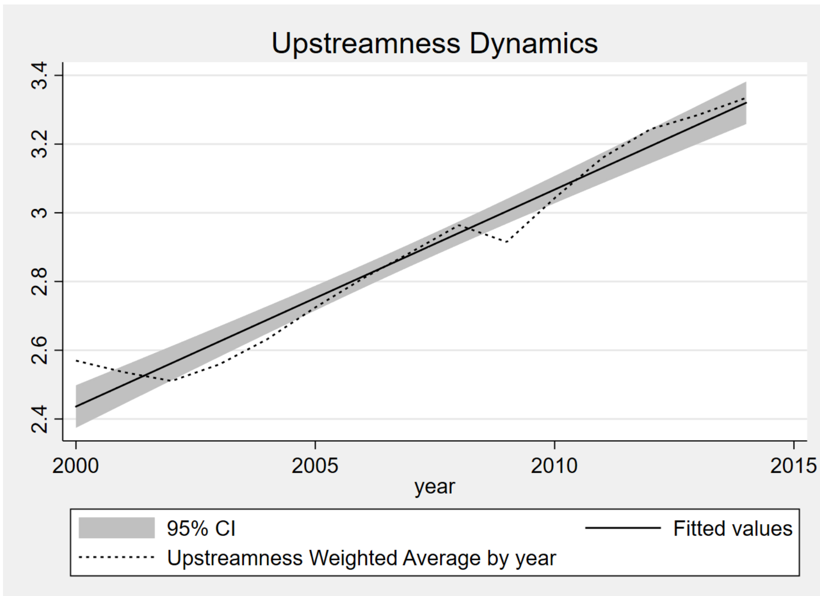
World Bank, *World Development Report 2020* number 32437. In 'World Bank Publications.', The World Bank, 12-2019 2020.

Zavacka, Veronika, "The bullwhip effect and the Great Trade Collapse," 2012.

Appendix A

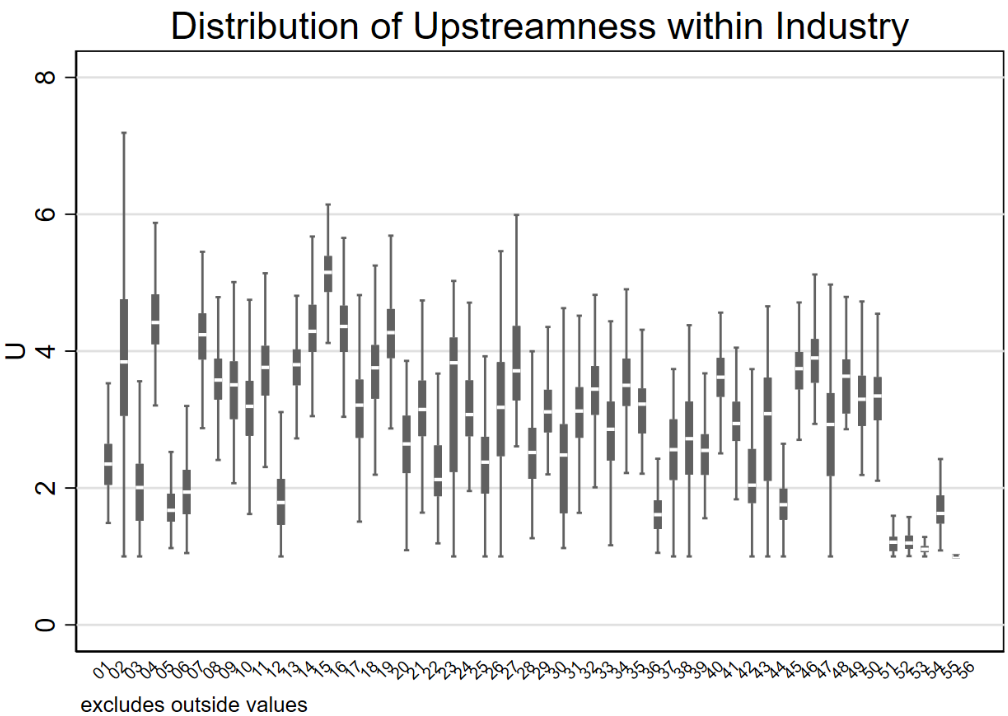
1.A Motivating Evidence

Figure 1.1: Upstreamness Dynamics



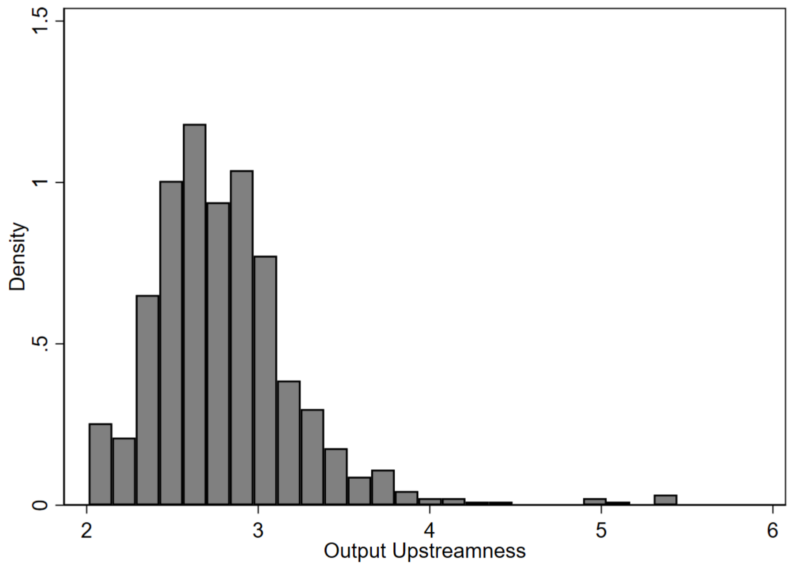
Note: the figure shows the dynamics of the weighted upstreamness measure computed as $U_t = \frac{\sum_i \sum_r y_{it}^r U_{it}^r}{\sum_i \sum_r y_{it}^r}$. It includes the estimated linear trend and the 95% confidence interval around the estimate.

Figure 1.2: Within Sector Upstreamness Distribution



Note: the graph plots the within sector box plot of upstreamness across all countries and years.

Figure 1.3: Country Output Upstreamness Distribution



Note: the figure shows distribution of country-year output upstreamness.

Table 1.1: Estimation of $I'(\cdot)$

	(1) Inventories
$\frac{\partial\, Inventories_t^i}{\partial\, Sales_t^i}$	0.0813***
	(0.00828)
Industry FE	YES
N	5668
R^2	0.633

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the non-parametric kernel estimation of $I'(\cdot)$, the derivative of the inventory function with respect to current sales. The estimation is based on the data of the NBER CES Manufacturing Industries Dataset for the years 2000-2011. Both series are HP-filtered. Standard errors are bootstrapped.

1.B Proofs

Proof of Proposition 1.1. The goal is to prove that if $0 < I'(x) < \frac{1}{1-\rho}$ then $\frac{\partial Y_t^n}{\partial D_t^0} > \frac{\partial Y_t^{n-1}}{\partial D_t^0}$, $\forall n, t$. The proof starts by characterising $\frac{\partial Y_t^n}{\partial D_t^0}$.

At stage 0, from 1.1

$$\frac{\partial Y_t^0}{\partial D_t^0} = 1 + \frac{\partial I(\mathbb{E}_t D_{t+1}^0)}{\partial D_t^0} = 1 + \rho I'$$

Similarly at stage 1

$$\begin{aligned}\frac{\partial Y_t^1}{\partial D_t^0} &= \frac{\partial Y_t^0}{\partial D_t^0} + \frac{\partial I(\mathbb{E}_t D_{t+1}^1)}{\partial D_t^0} \\ &= 1 + \rho I' + I' \left[\frac{\partial \mathbb{E}_t}{\partial D_t^0}\left[D_{t+1}^0 + I(\mathbb{E}_{t+1} D_{t+2}^0) - I(\mathbb{E}_t D_{t+1}^0)\right]\right] \\ &= 1 + \rho I' + \rho I'[1 + \rho I' - I']\end{aligned}$$

Similarly for stage 2, after some algebra

$$\begin{aligned}\frac{\partial Y_t^2}{\partial D_t^0} &= \frac{\partial Y_t^1}{\partial D_t^0} + \frac{\partial I(\mathbb{E}_t D_{t+1}^2)}{\partial D_t^0} \\ &= 1 + \rho I' + \rho I'[1 + \rho I' - I'] + \rho I'[1 + \rho I' - I']\end{aligned}$$

From the recursion

$$\frac{\partial Y_t^n}{\partial D_t^0} = \frac{\partial Y_t^{n-1}}{\partial D_t^0} + \rho I'[1 + \rho I' - I']$$

Given $\rho > 0$, if $0 < I'(x) < \frac{1}{1-\rho}$ then the last term is positive and $\frac{\partial Y_t^n}{\partial D_t^0} > \frac{\partial Y_t^{n-1}}{\partial D_t^0}$.
The statement follows. ∎

Proof of Lemma 1.1. From equation 1.1 and the assumption $I(x_t^n) = \alpha x_t^n$, $\forall n, t$

$$Y_t^0 = D_t^0 + \alpha \mathbb{E}_t D_{t+1}^0 - \alpha \mathbb{E}_{t-1} D_t^0 = D_t^0 + \alpha\rho\Delta_t$$

Using the production function identity $D_t^1 = Y_t^0$ and the definition of Y_t^1 as a function of demand at stage 0 and inventory adjustment

$$Y_t^1 = D_t^0 + \alpha\rho(2 - \alpha + \alpha\rho)\Delta_t = Y_t^0 + \alpha\rho(1 - \alpha(\rho - 1))\Delta_t.$$

Similarly, after tedious algebra,

$$\begin{aligned} Y_t^2 &= D_t^0 + \alpha\rho(3 + 3\alpha\rho - 3\alpha + \alpha^2 - 2\alpha^2\rho + \alpha^2\rho^2) \\ &= Y_t^1 + \alpha\rho(1 + \alpha(\rho - 1))\Delta_t. \end{aligned}$$

It follows from the recursion that

$$Y_t^n = Y_t^{n-1} + \alpha\rho(1 + \alpha(\rho - 1))\Delta_t,$$

or, as a fuction of final demand,

$$Y_t^n = D_t^0 + \alpha\rho \sum_{i=0}^{n} (1 + \alpha(\rho - 1))^i \Delta_t.$$

As stated in the Lemma. ■

Proof of Lemma 1.2. The first part of the Lemma follows immediately from the definition of output at a specific stage n and total sectoral output being the sum over stage specific production. The proof of the second part requires the following steps: first, using the definition of χ_k^n and denoting $\omega = 1 + \alpha(\rho - 1)$, rewrite total output as

$$\begin{aligned} Y_{k,t} &= \sum_{n=0}^{\infty} \chi_k^n \left[D_t + \alpha\rho \sum_{i=0}^{n} \omega^i \Delta_t \right] \\ &= \left[\tilde{A}^0 + \tilde{A}^1 + \dots \right]_k B D_t + \alpha\rho \left[\tilde{A}^0 \omega^0 + \tilde{A}^1(\omega^0 + \omega^1) + \dots \right]_k B\Delta_t \\ &= \tilde{L}_k B D_t + \alpha\rho \left[\sum_{n=0}^{\infty} \tilde{A}^n \sum_{i=0}^{n} \omega^i \right]_k B\Delta_t. \end{aligned}$$

The equality between the second and the third row follows from the convergence of a Neumann series of matrices satisfying the Brauer-Solow condition. To show that output exists non-negative for $\omega - 1 = \alpha(\rho - 1) \in [-1, 0]$, note that if $\omega - 1 = -1$ then $\omega = 0$, the second term vanishes and existence and non-negativity follow from $\tilde{L}$ finite and non-negative. If $\omega - 1 = 0$, then $\omega = 1$ and

$$\begin{aligned} Y_{k,t} &= \tilde{L}_k B D_t + \alpha\rho \left[\sum_{n=0}^{\infty} (n+1)\tilde{A}^n \right]_k B\Delta_t \\ &= \tilde{L}_k B D_t + \alpha\rho \left[\tilde{A}^0 + 2\tilde{A}^1 + 3\tilde{A}^2 + \dots \right]_k B\Delta_t \\ &= \tilde{L}_k B D_t + \alpha\rho \tilde{L}_k^2 B\Delta_t, \end{aligned}$$

where the last equality follows from $\sum_{i=0}^{\infty}(i+1)A^i = [I - A]^{-2}$ if A satisfies the Brauer-Solow condition. Existence and non-negativity follow from existence and non-negativity of $[I - \tilde{A}]^{-2}$.

If $\omega - 1 \in (-1, 0)$, then $\omega \in (0, 1)$. As this term is powered up in the second summation and as it is strictly smaller than 1, it is bounded above by $n + 1$. This implies that the whole second term

$$\sum_{n=0}^{\infty} \tilde{A}^n \sum_{i=0}^{n} \omega^i < \sum_{n=0}^{\infty} (n+1)\tilde{A}^n = \tilde{L}^2.$$

Alternatively, note that the second summation is strictly increasing in ω, as $\omega \leq 1$ the summation is bounded above by $n + 1$. Which completes the proof.

■

Proof of Lemma 1.3. An economy with general input-output structure can be thought of as an infinite collection of vertical production chains with length $n = 0, 1, 2, ...$ Defining upstreamness as

$$U_k = \sum_{n=0}^{\infty} (n+1) \frac{Y_k^n}{Y_k},$$

I need to prove that this metric is well defined. First, recall $Y_k = \sum_{n=0}^{\infty} Y_k^n$. Secondly, by Lemma 1.2 the following holds

$$Y_k = \tilde{L}_k B D_t + \alpha\rho \left[\sum_{n=0}^{\infty} \tilde{A}^n \sum_{i=0}^{n} \omega^i \right]_k B \Delta_t,$$

and

$$Y_k^n = \tilde{A}_k^n B D_t + \alpha\rho \tilde{A}_k^n \sum_{i=0}^{n} \omega^i B \Delta_t.$$

Then

$$U_k = \left[\tilde{L}_k B D_t + \alpha\rho \left[\sum_{n=0}^{\infty} \tilde{A}^n \sum_{i=0}^{n} \omega^i \right]_k B \Delta_t \right]^{-1} \left[\sum_{n=0}^{\infty} (n+1) \left[\tilde{A}_k^n B D_t + \alpha\rho \tilde{A}_k^n \sum_{i=0}^{n} \omega^i B \Delta_t \right] \right].$$

To show that U_k is finite, first note that $\sum_{n=0}^{\infty} (n+1) \tilde{A}_k^n B D_t = [I - \tilde{A}]_k^{-2} B D_t$ which is finite. Hence, I am left to show that the last term is finite. Following similar steps to the proof of Lemma 1.2 note that if $\omega = 0$ then the last term is 0. If $\omega = 1$ then $\sum_{n=0}^{\infty} (n+1) \alpha\rho \tilde{A}_k^n \sum_{i=0}^{n} \omega^i B \Delta_t =$

$\alpha\rho \sum_{n=0}^{\infty}(n+1)^2 \tilde{A}_k^n B\Delta_t$. Note that

$$\begin{aligned}
\sum_{n=0}^{\infty}(n+1)^2 \tilde{A}_k^n &= \sum_{n=0}^{\infty} n^2 \tilde{A}_k^n + 2\sum_{n=0}^{\infty} n\tilde{A}_k^n + \sum_{n=0}^{\infty} \tilde{A}_k^n \\
&= \sum_{n=0}^{\infty} n^2 \tilde{A}_k^n + 2\sum_{n=0}^{\infty}(n+1)\tilde{A}_k^n - \sum_{n=0}^{\infty} \tilde{A}_n^k \\
&= \sum_{n=0}^{\infty} n^2 \tilde{A}_k^n + 2[I-\tilde{A}]_k^{-2} - [I-\tilde{A}]_k^{-1}.
\end{aligned}$$

To show that the first term is bounded, totally differentiate

$$\begin{aligned}
&\frac{\partial}{\partial \tilde{A}} \sum_{n=0}^{\infty}(n+1)\tilde{A}_k^n = \frac{\partial}{\partial \tilde{A}}[I-\tilde{A}]_k^{-2} \\
&\sum_{n=0}^{\infty} n^2 \tilde{A}_k^{n-1} + \sum_{n=0}^{\infty} n\tilde{A}_k^{n-1} = 2[I-\tilde{A}]_k^{-3} \\
&\sum_{n=0}^{\infty} n^2 \tilde{A}_k^n = 2\tilde{A}[I-\tilde{A}]_k^{-3} - \tilde{A}[I-\tilde{A}]_k^{-2}.
\end{aligned}$$

As both terms of the right hand side are bounded, so is the term on the left hand side. This implies that $\sum_{n=0}^{\infty}(n+1)^2\tilde{A}_k^n$ is bounded. As the term is bounded for $\omega = 1$ and it is strictly increasing in ω, U_k is well defined for any $\omega \in [0,1]$. Finally, note that $U_k = 1$ iff $Y_k = Y_k^0$. ∎

Proof of Proposition 1.2. The result in part *a* follows from the partial derivative of output from Lemma 1.2. The statement in part *b* can be shown as follows

$$\begin{aligned}
\Delta_\beta \frac{\partial Y_{k,t}}{\partial D_t} &\equiv \frac{\partial}{\partial \beta_s}\frac{\partial Y_{k,t}}{\partial D_t} - \frac{\partial}{\partial \beta_r}\frac{\partial Y_{k,t}}{\partial D_t} = \\
&= \tilde{L}_{ks} + \alpha\rho \sum_{n=0}^{\infty} \tilde{A}_{ks}^n \sum_{i=0}^{n} \omega^i - \tilde{L}_{kr} - \alpha\rho \sum_{n=0}^{\infty} \tilde{A}_{kr}^n \sum_{i=0}^{n}\omega^i \\
&+ \sum_{n=0}^{\infty}\left[\tilde{A}_{ks}^n - \tilde{A}_{kr}^n\right]\left[1 + \alpha\rho\sum_{i=0}^{n}\omega^i\right].
\end{aligned}$$

Where the last equality follows from the definition of $\tilde{L}$.

Finally, the result in part c can be derived analogously

$$\begin{aligned}
\Delta_{\tilde{L}} \frac{\partial Y_{k,t}}{\partial D_t} &\equiv \frac{\partial Y_{k,t}|\tilde{A}'}{\partial D_t} - \frac{\partial Y_{k,t}|\tilde{A}}{\partial D_t} \\
&= \Delta \tilde{L}_k B + \alpha\rho \sum_{i=0}^{n} \omega^i \left[\tilde{A}'^n_k - \tilde{A}^n_k\right] B \\
&= \sum_{n=0}^{\infty} \left[\tilde{A}'^n_k - \tilde{A}^n_k\right] \left[1 + \alpha\rho \sum_{i=0}^{n} \omega^i\right] B.
\end{aligned}$$

Where the last equality follows from the definition of $\tilde{L}$. ■

Proof of Proposition 3.2. Denote $\tilde{U}_i^r$ the weighted average for a specific industry r in country i:

$$\begin{aligned}
\tilde{U}_i^r &= \sum_j \xi_{ij}^r U_{ij}^r = \\
&= \sum_j \frac{F_{ij}^r + \sum_s \sum_k a_{ik}^{rs} F_{kj}^s + ...}{Y_i^r} \frac{1 \times F_{ij}^r + 2 \times \sum_s \sum_k a_{ik}^{rs} F_{kj}^s + ...}{F_{ij}^r + \sum_s \sum_k a_{ik}^{rs} F_{kj}^s + ...} \\
&= \sum_j \frac{1 \times F_{ij}^r + 2 \times \sum_s \sum_k a_{ik}^{rs} F_{kj}^s + ...}{Y_i^r} \\
&= \frac{1 \times \sum_j F_{ij}^r + 2 \times \sum_s \sum_k a_{ik}^{rs} \sum_j F_{kj}^s + ...}{Y_i^r} \\
&= \frac{1 \times F_i^r + 2 \times \sum_s \sum_k a_{ik}^{rs} F_k^s + ...}{Y_i^r} = U_i^r,
\end{aligned}$$

where the equality between the fourth and the fifth line follows from $F_i^r = \sum_j F_{ij}^r$. ■

Table 1.2: Countries

Country			
Australia	Denmark	Ireland	Poland
Austria	Spain	Italy	Portugal
Belgium	Estonia	Japan	Romania
Bulgaria	Finland	Republic of Korea	Russian Federation
Brazil	France	Lithuania	Slovakia
Canada	United Kingdom	Luxembourg	Slovenia
Switzerland	Greece	Latvia	Sweden
China	Croatia	Mexico	Turkey
Cyprus	Hungary	Malta	Taiwan
Czech Republic	Indonesia	Netherlands	United States
Germany	India	Norway	Rest of the World

1.C WIOD Coverage

Table 1.3: Industries

Industry	Industry
Crop and animal production	Wholesale trade
Forestry and logging	Retail trade
Fishing and aquaculture	Land transport and transport via pipelines
Mining and quarrying	Water transport
Manufacture of food products	Air transport
Manufacture of textiles	Warehousing and support activities for transportation
Manufacture of wood and of products of wood and cork	Postal and courier activities
Manufacture of paper and paper products	Accommodation and food service activities
Printing and reproduction of recorded media	Publishing activities
Manufacture of coke and refined petroleum products	Motion picture
Manufacture of chemicals and chemical products	Telecommunications
Manufacture of basic pharmaceutical products and pharmaceutical preparations	Computer programming
Manufacture of rubber and plastic products	Financial service activities
Manufacture of other non-metallic mineral products	Insurance
Manufacture of basic metals	Activities auxiliary to financial services and insurance activities
Manufacture of fabricated metal products	Real estate activities
Manufacture of computer	Legal and accounting activities
Manufacture of electrical equipment	Architectural and engineering activities
Manufacture of machinery and equipment n.e.c.	Scientific research and development
Manufacture of motor vehicles	Advertising and market research
Manufacture of other transport equipment	Other professiona activitiesl
Manufacture of furniture	Administrative and support service activities
Repair and installation of machinery and equipment	Public administration and defence
Electricity	Education
Water collection	Human health and social work activities
Sewerage	Other service activities
Construction	Activities of households as employers
Wholesale and retail trade and repair of motor vehicles and motorcycles	Activities of extraterritorial organizations and bodies

Table 1.4: Highest and Lowest Upstreamness Industries

Industry	Upstreamness
Activities of extraterritorial organizations and bodies	1
Human health and social work activities	1.14
Activities of households as employers	1.16
Education	1.22
Public administration and defence	1.22
Accommodation and food service activities	1.66
⋮	⋮
Construction	3.96
Manufacture of wood and of products of wood and cork, except furniture	4.22
Manufacture of fabricated metal products, except machinery and equipment	4.27
Manufacture of machinery and equipment n.e.c.	4.28
Manufacture of other non-metallic mineral products	4.39
Mining and quarrying	4.52
Manufacture of basic metals	5.13

1.D Test of Uncorrelatedness of Instruments

As discussed in the main text the indentifying assumption for the validity of the shift share design is conditional independence of shocks and potential outcomes. Since this assumption is untestable a provide some evidence that the shares and the shocks are uncorrelated to reduce endogeneity concerns.

I test the conditional correlation by regressing the shares on future shocks and industry fixed effect. Formally

$$\xi^r_{ijt} = \beta\hat{\eta}_{jt+1}(i) + \gamma^r_{it} + \epsilon^r_{ijt}.$$

This estimation yields the following result

Table 1.5: Test of Exogeneity of Instruments

	ξ^r_{ijt}
$\hat{\eta}_{jt+1}(i)$	-0.0121
	(0.00762)
N	1517824
R^2	0.00284

Clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The test shows that the two are uncorrelated, suggesting that the shift share instrument is valid.

1.E Results

Table 1.6: Effect of Demand Shocks on Output Growth by Upstreamness Level

	(1) $\Delta \ln Y^r_{it}$
Uptreamness in [1,2]	0.530***
	(0.0202)
Uptreamness in [2,3]	0.606***
	(0.0188)
Uptreamness in [3,4]	0.705***
	(0.0158)
Uptreamness in [4, ∞)	0.785***
	(0.0381)
Constant	0.0705***
	(0.00208)
N	31921
R^2	0.238

Clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of industry output growth rates on demand shocks interacted with dummies taking value 1 if the Upstreamness level of the industry is in a given interval, e.g. [1,2].

Table 1.7: Effect of Demand Shocks on Output Growth by Upstreamness Level

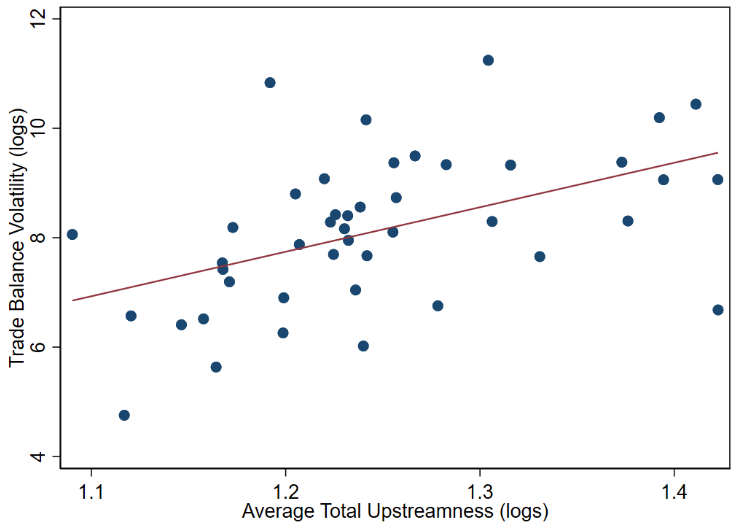
	(1)	(2)	(3)	(4)
	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$
Uptreamness in [1,2]	0.530***	0.530***	0.531***	0.530***
	(0.0202)	(0.0201)	(0.0202)	(0.0200)
Uptreamness in [2,3]	0.606***	0.606***	0.607***	0.605***
	(0.0188)	(0.0188)	(0.0188)	(0.0189)
Uptreamness in [3,4]	0.705***	0.706***	0.705***	0.705***
	(0.0158)	(0.0157)	(0.0158)	(0.0157)
Uptreamness in [4, ∞)	0.785***	0.788***	0.787***	0.785***
	(0.0381)	(0.0383)	(0.0386)	(0.0386)
Sector Outdegree		0.0110***		0.0323***
		(0.00258)		(0.00952)
Sector Leontief Coefficient			0.00129	-0.00648*
			(0.00101)	(0.00339)
Constant	0.0705***	0.0646***	0.0675***	0.0680***
	(0.00208)	(0.00259)	(0.00342)	(0.00408)
N	31921	31921	31921	31921
R^2	0.238	0.240	0.239	0.241

Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

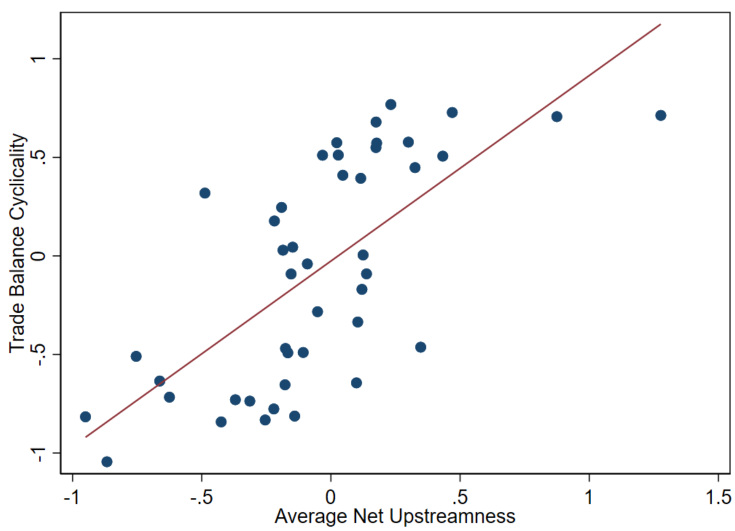
Note: this table displays the results of the regression of growth rate of industry output on demand shocks interacted with dummies taking value 1 if the Upstreamness level of the industry is in a given interval, e.g. [1,2]. Columns 2-4 include measures of network importance at the industry level. In particular the sector's outdegree and the cumulative Leontief inverse coefficient. Standard errors are clustered at the producing industry level.

Figure 1.4: Volatility and Total Upstreamness



Note: the graph displays the binscatter of the relationship between the log of the standard deviation of a country's trade balance and the log of the average embedded content (U^{TOT}). The graph is produced after controlling for log per capita GDP of the country.

Figure 1.5: Cyclicality and Net Upstreamness



Note: the graph displays the binscatter of the relationship between the cyclicality of the trade balance, measured as the correlation between the trade balance and output, and the measure of mismatch in the trade balance (U^{NX}). The graph is produced after controlling for log per capita GDP of the country.

1.F Robustness Checks

Table 1.8: Effect of Demand Shocks on Output Growth by Upstreamness Decile

	(1)	(2)
	$\Delta \ln Y_{it}^r$	$\sigma_{\Delta \ln Y_{it}^r}$
decile 1	0.493***	0.864***
	(0.0203)	(0.0561)
decile 2	0.556***	0.927***
	(0.0299)	(0.0679)
decile 3	0.580***	1.047***
	(0.0228)	(0.0583)
decile 4	0.607***	1.067***
	(0.0293)	(0.0582)
decile 5	0.611***	1.049***
	(0.0200)	(0.0531)
decile 6	0.671***	1.085***
	(0.0227)	(0.0488)
decile 7	0.682***	1.174***
	(0.0319)	(0.0470)
decile 8	0.678***	1.108***
	(0.0208)	(0.0371)
decile 9	0.666***	1.231***
	(0.0353)	(0.0425)
decile 10	0.760***	1.323***
	(0.0573)	(0.0981)
N	31921	2327
R^2	0.197	0.708

Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of industry output growth rates on demand shocks interacted with dummies taking value 1 if the upstreamness level of the industry belongs to a specific decile of the upstreamness distribution. Column (1) displays the regression of output growth rate on demand shocks, while Column (2) shows the regression of the industry specific variance of output growth rates on the variance of the demand shocks said industry faces.

Table 1.9: Effect of Demand shocks by level of Upstreamness

	(1) Supply Shocks Included $\Delta \ln Y_{it}^r$	(2) Domestic Industries Included $\Delta \ln Y_{it}^r$
Uptreamness in [1,2]	0.541***	0.358***
	(0.0206)	(0.0114)
Uptreamness in [2,3]	0.619***	0.493***
	(0.0192)	(0.0165)
Uptreamness in [3,4]	0.715***	0.579***
	(0.0162)	(0.0183)
Uptreamness in [4,∞)	0.795***	0.675***
	(0.0382)	(0.0436)
Constant	0.0712***	0.0662***
	(0.00209)	(0.00211)
N	31921	31921
R^2	0.239	0.164

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of industry output growth rates on demand shocks interacted with dummies taking value 1 if the upstreamness level of the industry is in a given interval, e.g. [1,2]. Column (1) runs the model on the demand shocks estimated absorbing producing industry-year variation. Column (2) uses the demand shocks calculated by excluding domestic industries final goods consumption.

Table 1.10: Effect of Demand shocks by level of Upstreamness

	(1)	(2)	(3)	(4)	(5)
	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$
Uptreamness in [1,2]	0.530***	0.360***	0.359***	0.225***	0.222***
	(0.0202)	(0.0750)	(0.0788)	(0.0718)	(0.0721)
Uptreamness in [2,3]	0.606***	0.415***	0.414***	0.280***	0.276***
	(0.0188)	(0.0574)	(0.0593)	(0.0714)	(0.0709)
Uptreamness in [3,4]	0.705***	0.481***	0.484***	0.338***	0.333***
	(0.0158)	(0.0777)	(0.0751)	(0.0884)	(0.0884)
Uptreamness in [4, ∞)	0.785***	0.529***	0.544***	0.385***	0.381***
	(0.0381)	(0.0761)	(0.0762)	(0.0985)	(0.0985)
Constant	0.0705***	0.0688***	0.0689***	0.0678***	0.0678***
	(0.00208)	(0.00821)	(0.00809)	(0.000222)	(0.000192)
Time FE	No	Yes	Yes	Yes	Yes
Country FE	No	No	No	Yes	Yes
Level FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes
N	31921	31921	31921	31921	31921
R^2	0.238	0.277	0.280	0.403	0.409

Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of industry output growth rates on demand shocks interacted with dummies taking value 1 if the upstreamness level of the industry is in a given interval, e.g. [1,2]. Column (1) displays the result of the simple OLS without any fixed effect. Column (2) adds year fixed effects. Column (3) includes both year and upstreamness level fixed effects. Column (4) adds producing country fixed effects and column (5) includes also producing industry fixed effects.

Table 1.11: Effect of Demand Shocks on Output by Upstreamness Level

	(1)	(2)
	$\ln Y_{it}^{r}$	$\ln Y_{it}^{r}$
Uptreamness in [1,2]	1.750***	0.140**
	(0.172)	(0.0646)
Uptreamness in [2,3]	2.992***	0.183**
	(0.160)	(0.0723)
Uptreamness in [3,4]	3.930***	0.256***
	(0.140)	(0.0570)
Uptreamness in [4,∞)	4.438***	0.399***
	(0.0988)	(0.0648)
Time FE	No	Yes
Country FE	No	Yes
Industry FE	No	Yes
N	32588	32588
R^2	0.421	0.648

Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of log industry output on demand shocks interacted with dummies taking value 1 if the Upstreamness level of the industry is in a given interval, e.g. [1,2]. Standard errors are clustered at the producing industry level in column 1 and at the producing industry, country and year level in column 2.

Table 1.12: Effect of Demand Shocks on Output by Upstreamness Level

	(1)	(2)
	$\ln Y_{it}^{r}$	$\ln Y_{it}^{r}$
Uptreamness in [1,2]	1.035***	0.578***
	(0.00407)	(0.0393)
Uptreamness in [2,3]	1.105***	0.598***
	(0.00548)	(0.0408)
Uptreamness in [3,4]	1.191***	0.635***
	(0.0155)	(0.0411)
Uptreamness in [4,∞)	1.253***	0.681***
	(0.0330)	(0.0412)
Time FE	No	Yes
Country FE	No	Yes
Industry FE	No	Yes
N	31634	31634
R^2	0.987	0.952

Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of log industry output on demand shocks interacted with dummies taking value 1 if the Upstreamness level of the industry is in a given interval, e.g. [1,2]. Standard errors are clustered at the producing industry level in column 1 and at the producing industry, country and year level in column 2.

Table 1.13: Effect of Demand Shocks on Output Growth by Upstreamness Level - Deflated Data

	(1) $\Delta \ln Y_{it}^r$	(2) $\Delta \ln Y_{it}^r$	(3) $\Delta \ln Y_{it}^r$	(4) $\Delta \ln Y_{it}^r$	(5) $\Delta \ln Y_{it}^r$
Uptreamness in [1,2]	0.615***	0.386***	0.386***	0.278***	0.279***
	(0.0236)	(0.0792)	(0.0826)	(0.0755)	(0.0761)
Uptreamness in [2,3]	0.731***	0.468***	0.469***	0.353***	0.352***
	(0.0260)	(0.0787)	(0.0809)	(0.0669)	(0.0682)
Uptreamness in [3,4]	0.900***	0.583***	0.580***	0.454***	0.454***
	(0.0275)	(0.102)	(0.0997)	(0.0833)	(0.0847)
Uptreamness in [4,∞)	1.078***	0.709***	0.702***	0.549***	0.546***
	(0.0436)	(0.150)	(0.137)	(0.128)	(0.129)
Constant	0.0756***	0.0769***	0.0769***	0.0774***	0.0774***
	(0.00214)	(0.00388)	(0.00386)	(0.000449)	(0.000377)
Time FE	No	Yes	Yes	Yes	Yes
Country FE	No	No	No	Yes	Yes
Level FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes
N	29809	29809	29809	29809	29809
R^2	0.205	0.240	0.241	0.275	0.285

Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of industry output growth rates on demand shocks interacted with dummies taking value 1 if the upstreamness level of the industry is in a given interval, e.g. [1,2]. Column (1) displays the result of the simple OLS without any fixed effect. Column (2) adds year fixed effects. Column (3) includes both year and upstreamness level fixed effects. Column (4) adds producing country fixed effects and column (5) includes also producing industry fixed effects.

Table 1.14: Effect of Demand Shocks on Output Growth by Upstreamness Level

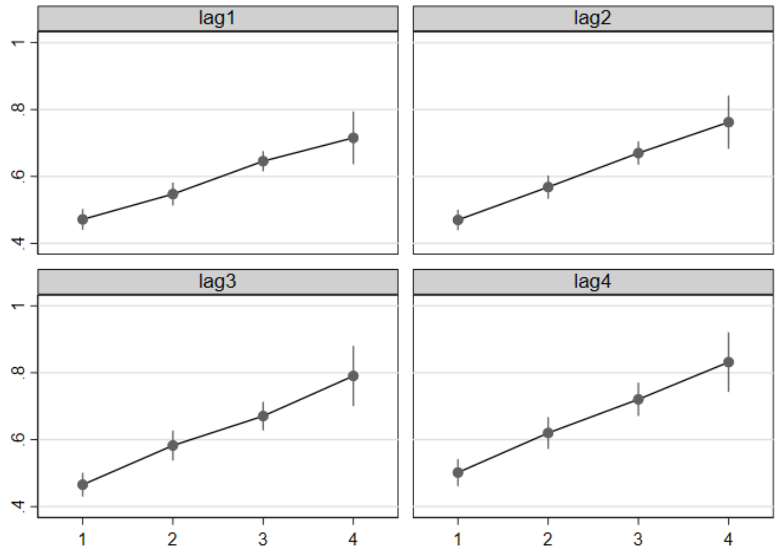
	(1)	(2)	(3)	(4)	(5)
	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$	$\Delta \ln Y_{it}^r$
Uptreamness in [1,2]	0.530***	0.472***	0.470***	0.465***	0.502***
	(0.0202)	(0.0156)	(0.0152)	(0.0178)	(0.0202)
Uptreamness in [2,3]	0.606***	0.547***	0.568***	0.583***	0.620***
	(0.0188)	(0.0172)	(0.0174)	(0.0224)	(0.0238)
Uptreamness in [3,4]	0.705***	0.646***	0.670***	0.670***	0.721***
	(0.0158)	(0.0151)	(0.0172)	(0.0213)	(0.0249)
Uptreamness in $[4,\infty)$	0.785***	0.716***	0.762***	0.790***	0.832***
	(0.0381)	(0.0393)	(0.0397)	(0.0450)	(0.0444)
L. $\Delta \ln Y_{it}^r$		0.0864***	0.0339***	0.0404***	0.0347**
		(0.0107)	(0.0115)	(0.0130)	(0.0171)
L2. $\Delta \ln Y_{it}^r$			0.0899***	0.0155	0.0214**
			(0.0122)	(0.0103)	(0.0104)
L3. $\Delta \ln Y_{it}^r$				0.0828***	0.0699***
				(0.00856)	(0.00899)
L4. $\Delta \ln Y_{it}^r$					0.0303***
					(0.00928)
Constant	0.0705***	0.0739***	0.0752***	0.0758***	0.0773***
	(0.00208)	(0.00218)	(0.00290)	(0.00334)	(0.00383)
N	31921	29077	26392	23887	21509
R^2	0.238	0.287	0.322	0.359	0.357

Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table displays the results of the regression of industry output growth rates on demand shocks interacted with dummies taking value 1 if the Upstreamness level of the industry is in a given interval, e.g. [1,2]. The first column of the table includes the first lag of the dependent variable, the other columns progressively add lags up $t - 4$.

Figure 1.6: Effect of Demand Shocks on Output Growth by Upstreamness Level - Output Growth Lags



Note: the figure shows the marginal effect of the variance of demand shocks on the variance of industry output changes by industry upstreamness level. The vertical bands show the 95% confidence intervals around the estimates. Note that due to relatively few observations above 4, all values above have been included in the U=4 category. The first panel of the figure includes the first lag of the dependent variable, the other panels progressively add lags up $t-4$.

Appendix B

Not for Publication

1.G Model Extension - Heterogeneous Storability

Assume that different sectors have different storage ability (think of some service industry being part of the production chain). Rewriting the model just described with stage specific storage, indexed by α_n implies the following amplification structure

$$Y_t^0 = D_t^0 + \alpha_0(\rho D_t^0 - \rho D_{t-1}^0).$$

From this, output at stage 1 is

$$\begin{aligned} Y_t^1 = & D_t^0 + \alpha_0 \rho D_t^0 - \alpha_0 \rho D_{t-1}^0 + \\ & + \alpha_1 \rho D_t^0 + \alpha_1 \alpha_0 \rho^2 D_t^0 - \alpha_1 \alpha_0 \rho^2 D_{t-1}^0 - \\ & - \alpha_1 \rho D_{t-1}^0 - \alpha_1 \alpha_0 \rho^2 D_{t-1}^0 + \alpha_1 \alpha_0 \rho^2 D_{t-2}^0, \end{aligned}$$

and at stage 2

$$\begin{aligned} Y_t^2 = & \left[(1 + \alpha_0 \rho + \alpha_1 \rho + \alpha_1 \alpha_0 \rho^2) D_t^0 - (\alpha_0 \rho + 2\alpha_1 \alpha_0 \rho^2 + \alpha_1 \rho) D_{t-1}^0 + \alpha_1 \alpha_0 \rho^2 D_{t-2}^0\right](1 + \alpha_2 \rho) - \\ & - \alpha_2 \rho \left[(1 + \alpha_0 \rho + \alpha_1 \rho + \alpha_1 \alpha_0 \rho^2) D_{t-1}^0 - (\alpha_0 \rho + 2\alpha_1 \alpha_0 \rho^2 + \alpha_1 \rho) D_{t-2}^0 + \alpha_1 \alpha_0 \rho^2 D_{t-3}^0\right]. \end{aligned}$$

This implies that contemporary amplification at this stage is

$$\begin{aligned} \frac{\partial Y_t^2}{\partial D_t^0} & = (1 + \alpha_0 \rho + \alpha_1 \rho + \alpha_1 \alpha_0 \rho^2)(1 + \alpha_2 \rho) \\ & = (1 + \alpha_0 \rho)(1 + \alpha_1 \rho)(1 + \alpha_2 \rho). \end{aligned}$$

Assume that stage 1 producers are in an industry whose product is not storable ($\alpha_1 = 0$), then amplification becomes

$$\frac{\partial Y_t^2}{\partial D_t^0} = (1 + \alpha_0 \rho)(1 + \alpha_2 \rho).$$

At a generic stage n, this relationship becomes

$$\frac{\partial Y_t^n}{\partial D_t^0} = \prod_{i=0}^{n} (1 + \alpha_i \rho).$$

This states that sectors whose goods are not storable do not contribute to upstream amplification but they do not erase the amplification coming from other sectors in the economy. They simply pass whatever shock the receive from customers to suppliers one-to-one.

1.H Inventory Adjustment

Antràs et al. (2012) define the measure of upstreamness based on the Input-Output tables. This measure implicitly assumes the contemporaneity between production and use of output. This is often not the case in empirical applications since firms may buy inputs and store them to use them in subsequent periods. This implies that before computing the upstreamness measure one has to correct for this possible time mismatch.

The WIOD data provides two categories of use for these instances: net changes in capital and net changes in inventories. These categories are treated like final consumption, meaning that the data reports which country but not which industry within that country absorbs this share of output.

The WIOD data reports as Z_{ijt}^{rs} the set of inputs used in t by sector s in country j from sector r in country i, independently of whether they were bought at t or in previous periods. Furthermore output in the WIOD data includes the part that is stored, namely

$$Y_{it}^{r} = \sum_{s}\sum_{j} Z_{ijt}^{rs} + \sum_{j} F_{ijt}^{r} + \sum_{j} \Delta N_{ijt}^{r}. \tag{1.30}$$

As discussed above the variables reporting net changes in inventories and capital are not broken down by using industry, i.e. the data contains ΔN_{ijt}^{r}, not ΔN_{ijt}^{rs}.

This feature of the data poses a set of problems, particularly when computing bilateral upstreamness. First and foremost including net changes in inventories into the the final consumption variables may result in negative final consumption whenever the net change is negative and large. This cannot happen since it would imply that there are negative elements of the F vector when computing

$$U = \hat{Y}^{-1}[I - A]^{-2}F.$$

However, simply removing the net changes from the F vector implies that the tables are not balanced anymore. This is also problematic since then, by the definition of output in equation 1.30 it may be the case that the sum of inputs is larger than output. When this is the case $\sum_i \sum_r a_{ij}^{rs} > 1$, which is a necessary condition for the convergence result, as discussed in the Methodology section.

To solve this set of problems I apply the the inventory adjustment suggested by Antràs et al. (2012). This boils down to reducing output by the change of inventories. This procedure however requires an assumption of inventory usage. In particular, as stated above, the data reports ΔN^{r}_{ijt} but not ΔN^{rs}_{ijt}. For this reason, the latter is imputed via a proportionality assumption. Namely if sector s in country j uses half of the output that industry r in country i sells to country j for input usages, then half of the net changes in inventories will be assumed to be used by industry s. Formally:

$$\Delta N^{rs}_{ijt} = \frac{Z^{rs}_{ijt}}{\sum_s Z^{rs}_{ijt}} \Delta N^{r}_{ijt}.$$

Given the inputed vector of ΔN^{rs}_{ijt}, output of industries is corrected as

$$\tilde{Y}^{rs}_{ijt} = Y^{rs}_{ijt} - \Delta N^{rs}_{ijt}.$$

Finally, whenever necessary, Value Added is also adjusted so that the the columns of the I-O tables still sum to the corrected gross output.

This corrections insure that the necessary conditions for the matrix convergence are always satisfied.

1.I Descriptive Statistics

This section provides additional descriptive statistics on the World Input-Output Database (WIOD) data.

1.I.1 Degree Distributions

After calculating the input requirement matrix A, whose elements are $a^{rs}_{ij} = Z^{rs}_{ij}/Y^{s}_{j}$. One can compute the industry level in and outdegree

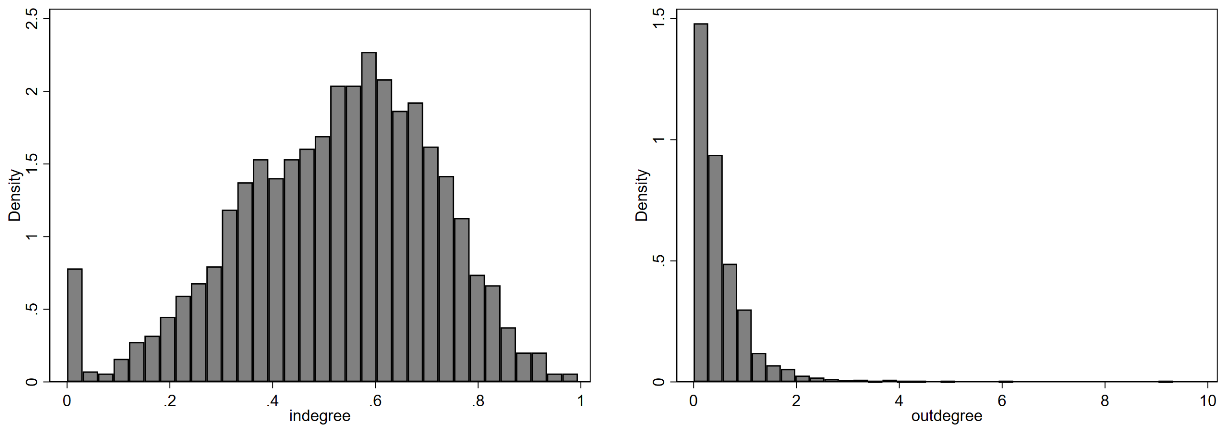
$$indegree^{r}_{i} = \sum_{i}\sum_{r} a^{rs}_{ij}, \tag{1.31}$$

$$outdegree^{r}_{i} = \sum_{j}\sum_{s} a^{rs}_{ij}. \tag{1.32}$$

The indegree measures the fraction of gross output that is attributed to inputs (note that $indegree^{r}_{i} = 1 - va^{r}_{i}$ where va^{r}_{i} is the value added share).

The weighted outdegree is defined as the sum over all using industries of the fraction of gross output of industry r in country i customers that can be attributed to industry r in country i. This measure ranges between 0, if the sector does not supply any inputs to other industries, and $S * J$, being the total number of industries in the economy, if industry r in country i is the sole supplier of all industries. In the data the average weighted outdegree is .52.

The distributions of these two measures are in Figure 1.1.



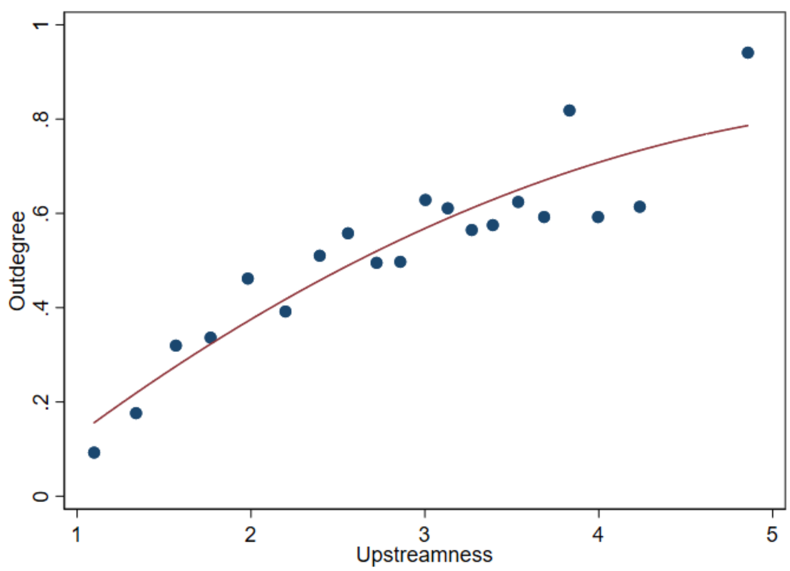
(a) Indegree

(b) Outdegree

Figure 1.1: Degree Distributions

In the WIOD sample industries' outdegree positively correlate with upstreamness, which suggests that industries higher in production chains serve a larger number (or a higher fraction) of downstream sectors. This relationship is shown in Figure 1.2.

Figure 1.2: Outdegree and Upstreamness



Note: the figure plots the binscatter of industries' outdegree and upstreamness.

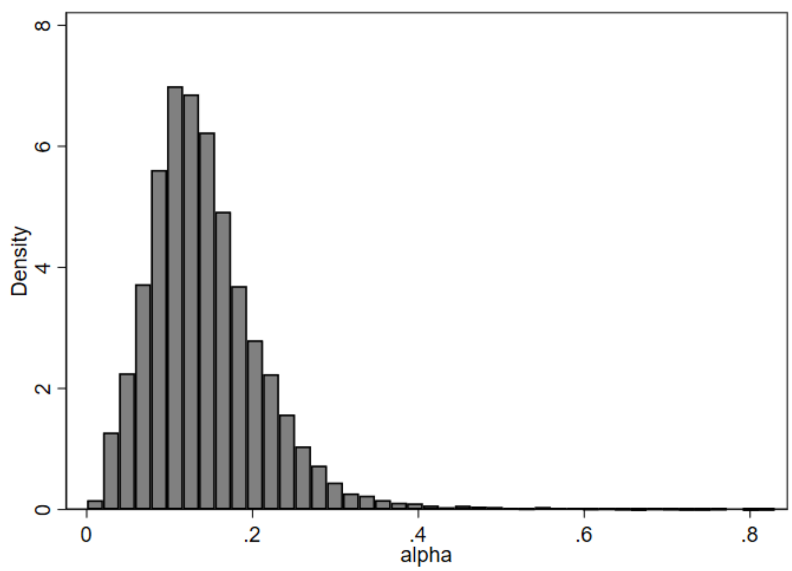
1.J Inventories

In the model presented in this paper part of the amplification is driven by procylical inventory adjustment. The WIOD data does not provide industry specific inventory stock or change, eliminating the possibility of a direct test of the mechanism.

To provide partial evidence of the behaviour of inventories I use the NBER CES Manufacturing Industry data. This publicly available dataset covers 473 US manufacturing industries at the 6-digit NAICS from 1958 to 2011. The data contains industry specific information about sales and end of the period inventories.

As mentioned in the main body of the paper, computing the parameter $\alpha \equiv I_t/\mathbb{E}_t D_{t+1}$ as $\alpha_t = I_t/D_{t+1}$ provides a set of numbers between 0 and 1, with an average of approximately 15%. Figure 1.3 shows the distribution of α across all industries and years.

Figure 1.3: Distribution of α



In the model the key assumption is that α is a constant across industries and time. This would imply that inventories are a linear function of sales. Figure 1.4 shows the scatter plot of the end of the period stock of inventories as a function of current sales (the same picture arises for next period sales). The data is first demeaned at the sectoral level to partial out industry specific differences and only exploit within sector variation. The graph includes a quadratic fit.

Figure 1.4: Inventories and Sales

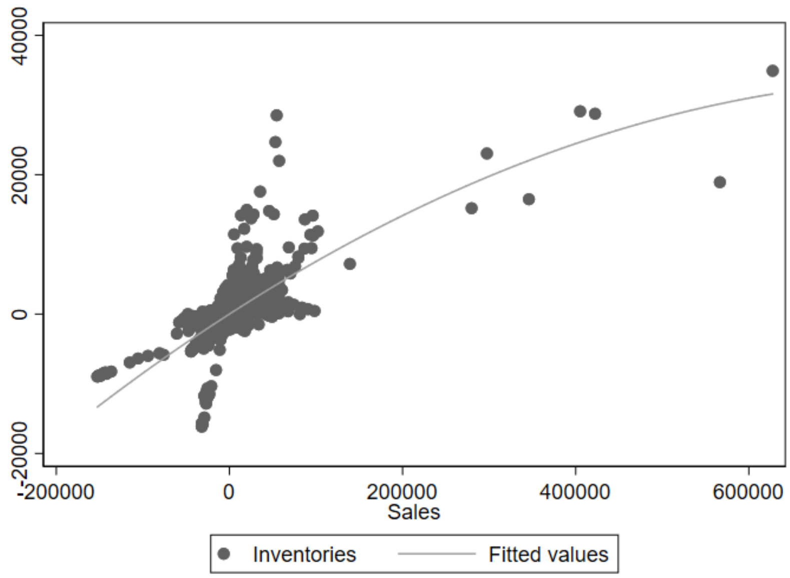
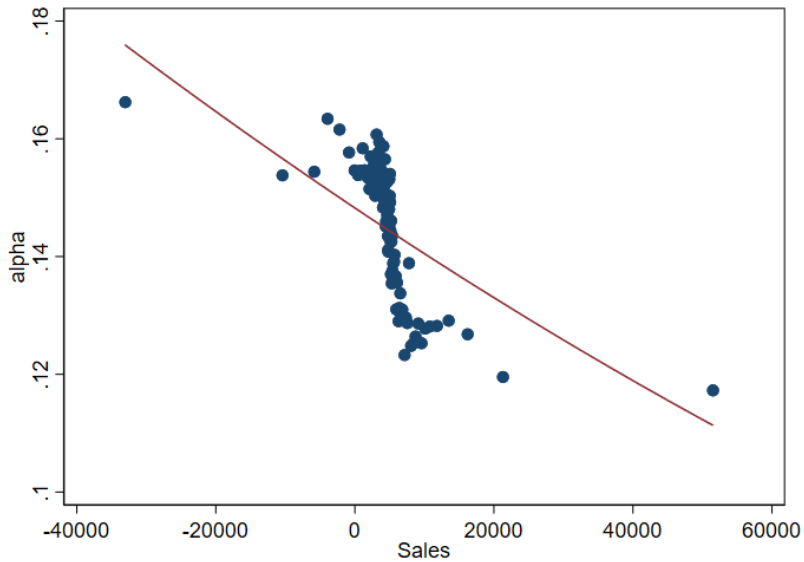


Figure 1.4 suggests that linearity assumption is relatively close to the data when sales are below the 90th percentile of their distribution. At very high sales level the function significantly deviates from linearity. As discussed in section 3.2 all the results go through, provided that the function is not "too concave", in a sense specified there. The necessary condition is expressed in terms of the elasticity of α. The assumptions made there are that α is positive and a decreasing function of demand. Furthermore, to observe amplification, one needs the function to be either strictly concave or "not too convex". Figure 1.5 shows the binscatter for the relationship between α and sales, after controlling for sector and year fixed effects. The plot includes a quadratic fit.

Figure 1.5: α and Sales



The graph shows that α is indeed decreasing, positive and slightly convex. Table 1.1 provides the results for the fixed effects regression of inventories and α over sales in columns 1 and 2. Column 3 provides the estimates of the change in inventories over the change in sales, to test procyclical adjustments.

Table 1.1: Inventories and Sales

	(1)	(2)	(3)	(4)
	I_t	α_t	α_t	ΔI_t
S_t	0.0669***	-0.000000177***		
	(0.000439)	(2.79e-08)		
$ln(S_t)$			-0.0128***	
			(0.000608)	
ΔS_t				0.0119***
				(0.000546)
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
N	24912	24912	24912	24439
R^2	0.821	0.713	0.717	0.0886

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table shows the results of the estimation of inventories response to sales. Column (1) shows the regression of end of the period inventories on current sales. Column (2) shows the regression of α on contemporaneous sales, while Column (3) uses the log of sales. Column (4) displays the results for the changed in inventories regressed on the change in sales.

As shown in Column 4 a positive change in sales correlates with a positive change in inventories, suggesting that the latter are procyclically adjusted. Furthermore, as shown in Column 3, for the class of functions $\alpha(x) = x^\beta$, the estimated functions satisfies the condition for amplification laid out in Section 3.2 of the Appendix.

1.K Stylized Facts

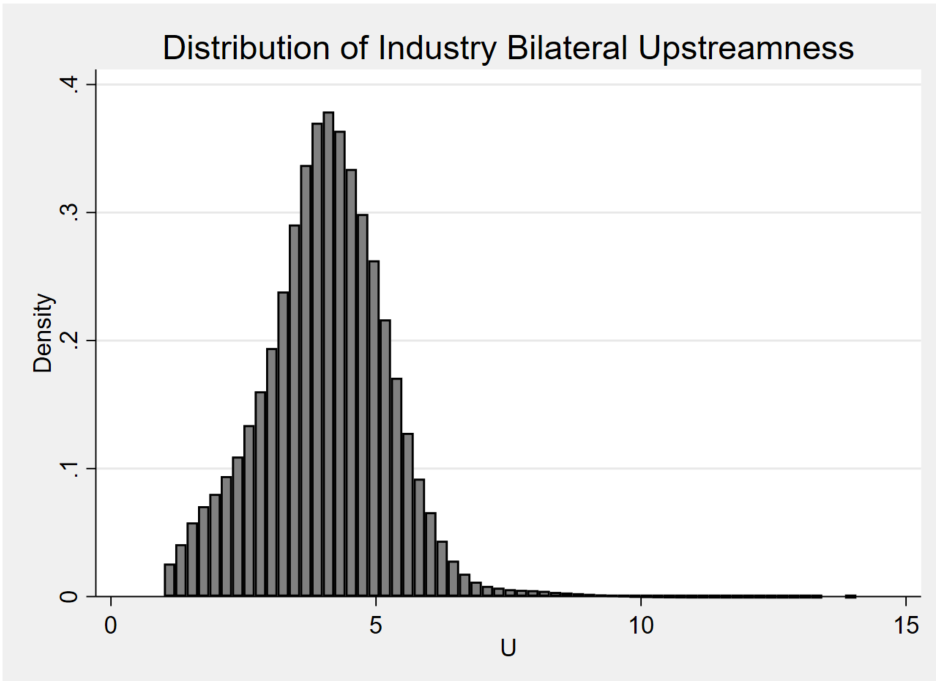
This sections provides a set of novel stylized facts regarding how countries place themselves in global value chains depending on their degree of development, the salient features of industry sales portfolios and some well known features of trade over the business cycle. These empirical regularities extend the facts discussed by Antràs and Chor (2018) and Miller and Temurshoev (2017).

1.K.1 GVC Positioning and Development

Industries place themselves at different stages of production chains depending on their country of origin and the specific partner country they are trading with. This section provides a set of descriptive statistics about the measure and GVC positioning.

First I plot the distribution of bilateral upstreamness for all industry-partner country-year combinations. This amounts to 1,626,240 different points for U^r_{ijt}.

Figure 1.6: Distribution of Industry Bilateral Upstreamness

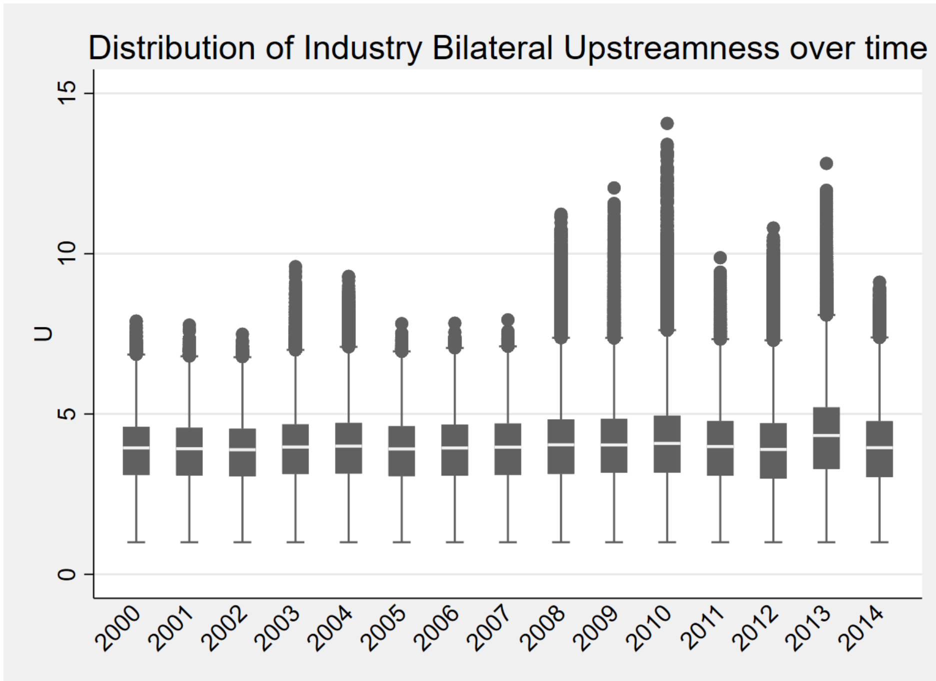


Note: the figure plots the distribution of industry specific bilateral upstreamness pooling all industries, partner countries and years.

The distribution is right skewed, with the average upstreamness being approximately 4 and a long right tail with values up to 14. The central 80% of the distribution lies between 2.5 and 5.5 production stages away from final consumption.

The evidence suggests that over the sample period (2000-2014) the bulk of the distribution did not move, as evidenced by Figure 1.7 which provides the time specific box plot of the industry bilateral upstreamness measure.

Figure 1.7: Distribution of Industry Bilateral Upstreamness over time



Note: the figure shows the box plots of the year specific distribution of bilateral industry upstreamness

The graph also suggests that over time the right tail of the distribution shifted further to the right. This may be evidence of increasing length of production processes for those products that were already complex in nature.

To assess which channel explains the observed increase in upstreamness I apply the decomposition proposed by Foster et al. (2001) to the weighted upstreamness changes , namely

$$\Delta U_t = \sum_i \sum_r \underbrace{\Delta U_{it}^r w_{it-1}^r}_{Within} + \underbrace{U_{it-1}^r \Delta w_{it}^r}_{Between} + \underbrace{\Delta U_{it}^r \Delta w_{it}^r}_{Covariance} . \tag{1.33}$$

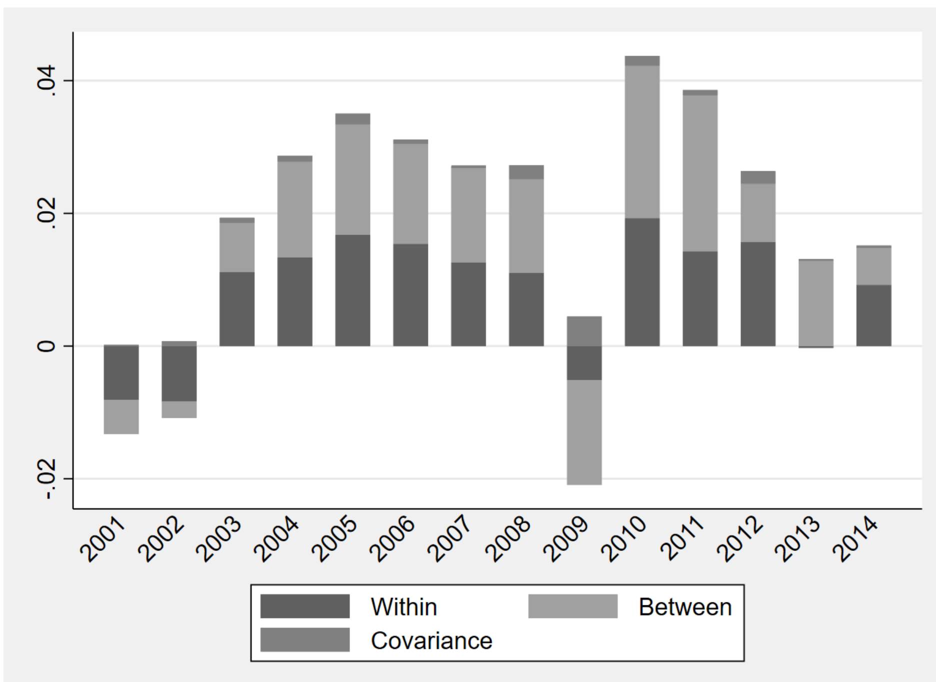
Where $w_{it}^r = \frac{y_{it}^r}{\sum_r y_{it}^r}$ is the industry r output weight within country i. This decomposition separates the contribution to the outcome changes in changes within, namely, given the weights, changes in the level of upstreamness; between, given the level of upstreamness, changes in industry weights and the covariance term.[23]

The results of the decomposition are plotted in Figure 1.8. The analysis suggests that the

[23]I dispense of the two terms for the contribution of entrants and exiters since given the aggregate nature of the data virtually no flow is zero.

observed changes in average upstreamness over time are due to the Within and Between components in equal shares, implying that most of the growth is stemming from large flows increasing the length of the production process and flows of complex goods becoming larger over time. One last interesting stylized fact stemming from the decomposition is that the covariance terms is always positive, independently of whether the average upstreamness is increasing or decreasing in a given year. This suggests that the reallocation is such that flows of products becoming more complex (or further away from consumption) are increasing in relative size or flows becoming less complex are decreasing in their relative importance. This is true even in 2009 during what the literature has labeled the *Great Trade Collapse*, see Baldwin (2011), suggesting that the effect of the crisis was heterogeneous on flows with different degrees of complexity. The specific contributions are displayed in Table 1.2.

Figure 1.8: Upstreamness Dynamics Decomposition



Note: the figure shows the stacked contributions (in levels) of the different components of the changes in the weighted upstreamness measure calculated as shown in equation 1.33.

In order to further inspect possible determinants of industry positioning I turn to the analysis of the correlations between the measure and economic development, proxied by GDP per capita. To evaluate this I construct the weighted upstreamness by origination country,

Table 1.2: Upstreamness Dynamics Decomposition Contributions

	mean	sd	min	max
within	.4744478	.1917807	-.0215916	.8264168
between	.5242279	.2169459	.2443104	1.010596
cov	.0013245	.0862674	-.270217	.0776356

Note: this tables shows the results from the decomposition of the changes of the weighted upstreamness measure. The table displays the contribution of the different components, namely the *within*, representing changes of the upstreamness level given the weights, *between*, representing changes in the weights given the level of upstreamness, and the *covariance* term, being the simultaneous changes in the level of upstreamness and the weights.

using as weights industry output shares

$$U_{it} = \frac{\sum_r y_{it}^r U_{it}^r}{\sum_r y_{it}^r}, \tag{1.34}$$

and run the following model

$$\ln U_{it} = \beta \ln y_{it} + \delta_t + \epsilon_{it}. \tag{1.35}$$

The results of this estimation are provided in Table 1.3 (and plotted in Figure 1.9). The model shows that there is a positive correlation between a country's economic development and how upstream its industries tend to be. Note that the relationship seems to be consistent only within country, meaning that the initial levels of upstreamness and development are uncorrelated, but that, given a country's baseline, the correlation turns positive and significant. Specifically, a 1% increase in per capita GDP results in a .2% increase in the measure of industry upstreamness. The relationship remains consistent when controlling for the country size, proxied by log GDP, which negatively correlates with the degree of upstreamness, suggesting that larger countries' industries tend to be closer to final consumption.

Table 1.3: Weighted Upstreamness and Economic Development

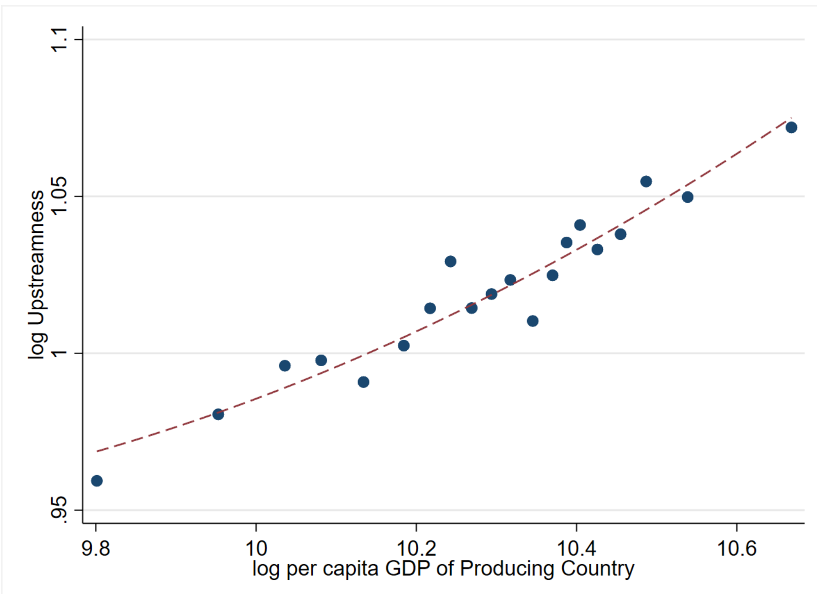
	(1) log Upstreamness	(2) log Upstreamness	(3) log Upstreamness	(4) log Upstreamness
log per capita GDP of Producing Country	-0.00591	-0.00639	0.114***	0.201***
	(-1.11)	(-1.21)	(12.02)	(4.32)
log GDP of Producing Country		0.00937***		-0.0917*
		(2.96)		(-1.90)
Time FE	Yes	Yes	Yes	Yes
Country FE	No	No	Yes	Yes
N	660	660	660	660
R^2	0.0840	0.0963	0.929	0.929

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table presents the results of the regression of the log of upstreamness on the producing country per capita GDP. Weighted upstreamness is computed as described in equation 1.34. Column (1) shows the results of the model including only time fixed effects. Column (2) adds the log of GDP of the producing country as a control. Finally Columns (3) and (4) replicate the models in (1) and (2) but include country fixed effects.

Figure 1.9: Industry Bilateral Upstreamness and Economic Development



Note: the figure shows the binscatter of the regression of log upstreamness on log per capita GDP of the producing country. The red dotted line shows the quadratic fit line.

Next I turn to how industries position themselves depending on the partner country's degree of economic development. First I construct the weighted upstreamness by partner country as

$$U_{\cdot jt} = \frac{\sum_i \sum_r y_{it}^r U_{ijt}^r}{\sum_i \sum_r y_{it}^r},$$

and estimate the following econometric model

$$\ln U_{\cdot jt} = \beta \ln y_{jt} + \delta_t + \epsilon_{jt}. \tag{1.36}$$

The results in Table 1.4 (and plotted in Figure 1.10) suggest that the correlation between industry bilateral upstreamness and partner country development is negative, with a 1% increase in the purchasing country per capita GDP implying a .5% drop in the industry bilateral upstreamness measure. The relationship turn insignificant when controlling for partner country size, suggesting that the larger the partner country the closer to consumption industries are when trading with it.

Table 1.4: Bilateral Industry Upstreamness and Partner Country Economic Development

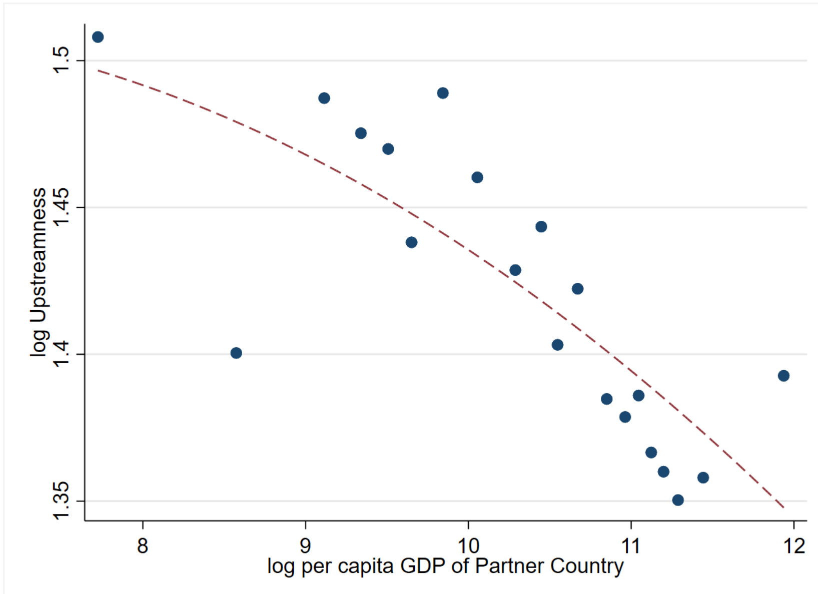
	(1) log Upstreamness	(2) log Upstreamness	(3) log Upstreamness	(4) log Upstreamness
log per capita GDP of Partner Country	-0.0363***	-0.0357***	-0.0563***	0.0813
	(-11.10)	(-11.27)	(-4.00)	(1.18)
log GDP of Partner Country		-0.0129***		-0.146**
		(-6.82)		(-2.04)
Time FE	Yes	Yes	Yes	Yes
Country FE	No	No	Yes	Yes
N	660	660	660	660
R^2	0.306	0.352	0.690	0.692

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table presents the results of the regression of the log of upstreamness on the partner country per capita GDP. Weighted upstreamness is computed as described in equation 1.36. Column (1) shows the results of the model including only time fixed effects. Column (2) adds the log of GDP of the partner country as a control. Finally Columns (3) and (4) replicate the models in (1) and (2) but include country fixed effects.

Figure 1.10: Industry Bilateral Upstreamness and and Partner Country Economic Development



Note: the figure shows the binscatter of the regression of log upstreamness on log per capita GDP of the partner country. The red dotted line shows the quadratic fit line.

A further interesting empirical regularity is that countries tend to trade among each other following a specific pattern of specialization in bilateral flows. Figure 1.11 displays the pattern of bilateral net upstreamness over the difference in log per capita income between the two countries. What emerges is a strong positive correlation. This suggests that when developed economies trade with emerging economies they sell upstream goods and buy downstream ones. Similarly when countries with comparable levels of development trade their net upstreamness is relatively more concentrated around zero, suggesting that they trade in similarly upstream or complex goods. To estimate this relationship I run the following model

$$U_{ijt}^{NX} = \beta \Delta \ln y_{ijt} + \nu_{ijt}. \tag{1.37}$$

Where $U_{ijt}^{NX} = U_{ijt}^{X} - U_{ijt}^{M}$, $\Delta \ln y_{ijt} = \ln y_{it} - \ln y_{jt}$ with $\ln y_{it}$ denoting log per capita income of country i at time t.

The estimates of this relationship are displayed in Table 1.5. Note that the regression does not include within country flows (bet upstreamness and income differences are zero by definition) and since both the measures are symmetric ($U_{ijt}^{NX} = -U_{jit}^{NX}$) it only includes pairs once, independently of the direction of the flows, i.e. it drops flows from j to i whenever flows from i to j are in the data, hence the sample size is $J \times (J-1)/2 \times T$, where J is the number of countries.

The estimation suggests that the higher the difference in per capita GDP the higher the difference in net upstreamness. In particular when developed countries trade to developing ones they export more upstream than they import and viceversa. Quantitatively the results state that increasing the difference in log per capita GDP by one point produces a .19 increase in the bilateral net upstreamness.

Table 1.5: Net Bilateral Upstreamness and per capita GDP difference

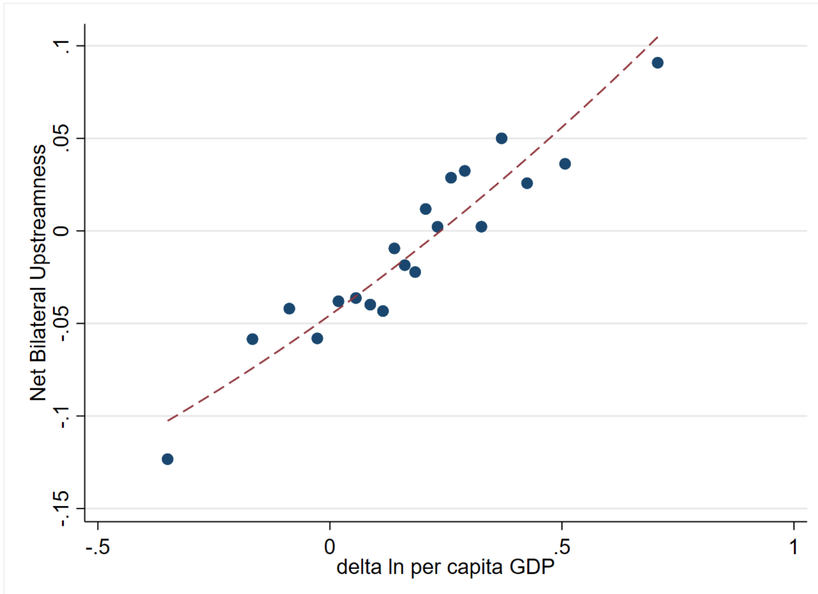
	(1) U_{ijt}^{NX}	(2) U_{ijt}^{NX}
Log per capita Income Difference	0.196***	0.196***
	(11.78)	(16.91)
Time FE	Yes	Yes
Country FE	Yes	No
Partner Country FE	Yes	No
Pair FE	No	Yes
N	14190	14190
R^2	0.529	0.785

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: this table presents the results of the regression of Net Bilateral upstreamness on the difference in log per capita GDP of the producing and partner country. Column (1) includes time and country fixed effects, while columns (2) replaces the countries fixed effects with country pair fixed effects. Standard errors are clustered accordingly.

Figure 1.11: Net Bilateral Upstreamness and per capita GDP difference



Note: the figure shows the binscatter of the regression of Net Bilateral upstreamness on the difference in log per capita GDP of the producing and the partner country. The red dotted line shows the quadratic fit line.

1.K.2 Sales Portfolio Composition

The distribution of sales portfolio shares is computed as described in the methodology. The goal of this section is to study whether there composition of sales portfolios would allow for demand shocks diversification. Table 1.6 reports the summary statistics of the portfolio shares for all industries and all periods.

Table 1.6: Portfolio Shares Summary Statistics

	count	mean	sd	min	max	p25	p50	p90	p95	p99
portfolio share	1522475	.0227	.1029	3.33e-13	.9999	.0004	.0017	.026	.0659	.7143
domestic portfolio share	34632	.6146	.2744	.0001	.9999	.4176	.6674	.9442	.9793	.9974
export portfolio share	1487843	.0089	.0273	3.33e-13	.962	.0003	.0016	.0199	.0418	.1224

Note: the table displays the summary statistics of the sales portfolio shares. Shares equal to 0 and 1 have been excluded. The latter have been excluded because they arise whenever an industry has 0 output. No industry has an actual share of 1.

The first noticeable feature of the data is that the distribution is very skewed, with the median share being equal to .01%. The skewness is largely driven by domestic sales, which mostly lie in the very right tail of the [0,1] interval. The median of domestic sales is 67%. This also points to relatively low share of trade, even when accounting for third countries linkages. The predominant relevant demand for industry is still the domestic one. The bin scatters of the two distributions are shown in Figure 1.12.

The distribution of all portfolio shares is skewed. To test the skewness of the distribution, I replicated the methods by di Giovanni et al. (2011) and Axtell (2001). These methods are used to estimate the coefficient of the power law according to which the data is thought to be distributed. Note that portfolio shares cannot be really distributed as a power law due to the inherent bounded support. This procedure is effectively just a way to assess how skewed their distribution is.
The procedure to estimate the coefficient of the power law relies on the definition of the distribution

$$P(S > s) = Cs^{-\zeta},$$

which can be estimated in log log as

$$\ln(P(S > s)) = \ln(C) - \zeta \ln(s).$$

Alternatively it can be studied by regressing the log of the (rank-0.5) on the log of the shares themselves as suggested by Gabaix and Ibragimov (2011). The estimation is

$$\ln(Rank_i - 0.5) = \beta_0 + \zeta \ln(s_i) + \epsilon_i.$$

These two procedures yield very similar results, reported in Table 1.7. The estimated power law coefficient being approximately .38, suggests that the distribution has a very fat tail. For this reason the scope for diversification is limited, particularly regarding domestic shocks.

Table 1.7: Portfolio Shares Regressions

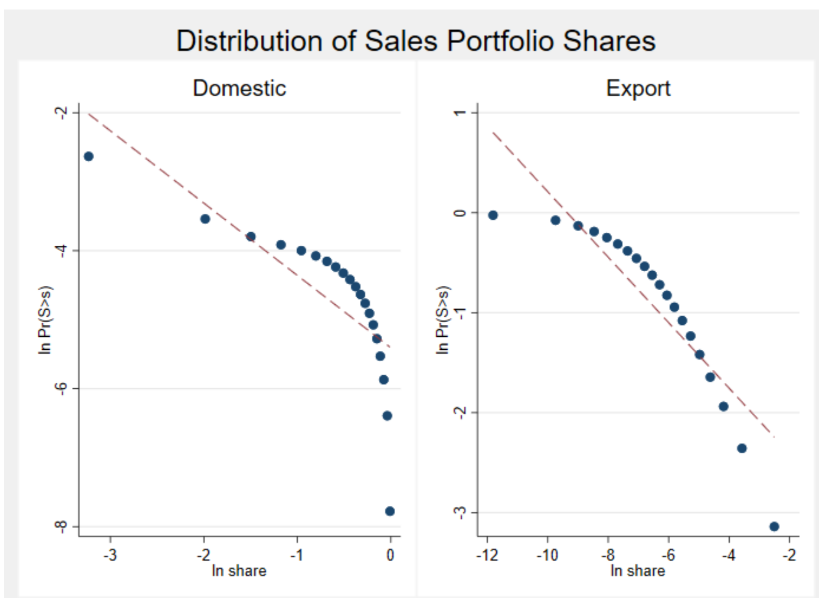
	(1) ln(Pr(S>s))	(2) ln Rank
ln s	-0.373***	0.394***
	(-45.46)	(106.69)
Constant	-3.401***	15.77***
	(-70.08)	(761.84)
N	1522474	1522475
R^2	0.770	0.857

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: shares equal to 0 and 1 have been excluded. The latter have been excluded because they arise whenever an industry has 0 output. No industry has an actual share of 1.

Figure 1.12: Portfolio Shares Distributions



Note: the figure shows the binscatter of the regressions of the log of the countercumulative frequency sales portfolio shares on the log of the of the portfolio shares. The left panel displays the relationship for domestic sales and right panel for export sales. The red dotted line represent the estimated fit of the regression.

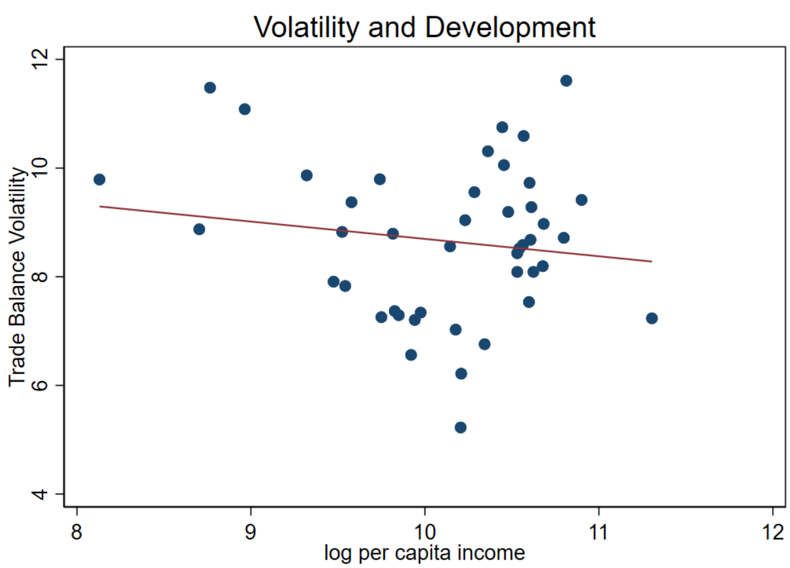
The main takeaway from this analysis is that trade is still relatively limited in the industry sales portfolio and given the high heterogeneity in the portfolios themselves demand shocks may not be diversified away.

1.K.3 Business Cycle Facts

Net Export Volatility

The first empirical regularity in international macroeconomics is that emerging economies display a larger volatility of net exports than developed countries. This fact is evident from Figure 1.13. The figure displays the log of the standard deviation of the trade balance against the log of per capita income. In order to detrend the trade balance I follow Uribe and Schmitt-Grohé (2017) and rescale the trade balance by the trend component of output before taking the quadratic trend.

Figure 1.13: Volatility and Development

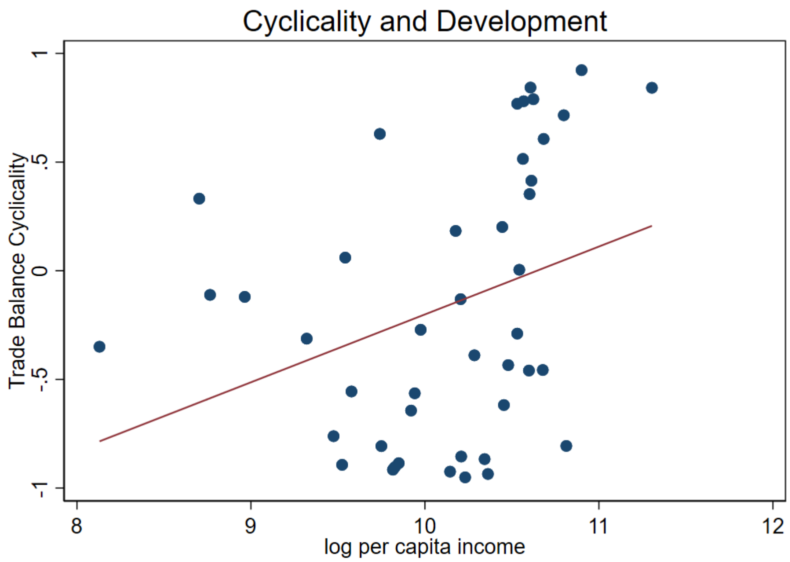


The graph displays a negative correlation between the volatility of the trade balance and the degree of development of the country, measured by per capita GDP.

Net Export Cyclicality

The second business cycle fact is that emerging economies display more countercyclical trade balances than developed countries. In Figure 1.14 I plot the correlation of the detrended trade balance (as described above) with log quadratically detrended output[24] from the World Bank data. The correlation with log per capita income is significantly positive.

Figure 1.14: Cyclicality and Development



[24]Unless otherwise specified detrending is performed by HP filtering the series. The results presented are robust to alternative methods like log quadratic detrending.

Chapter 2

Low Competition Traps

Alessandro Ferrari[1] , Francisco Queirós

2.1 Introduction

The US economy appears to have experienced a fundamental change over the past four decades. Several studies and indicators have highlighted different secular trends concerning the structure of product markets, the distribution of income across factors and the distribution of activity across firms. Some of the trends attracting prominent attention in the recent debate include[2]

1. the decline in the labor share,

2. the increase in the profit share,

3. the increase in aggregate markups,

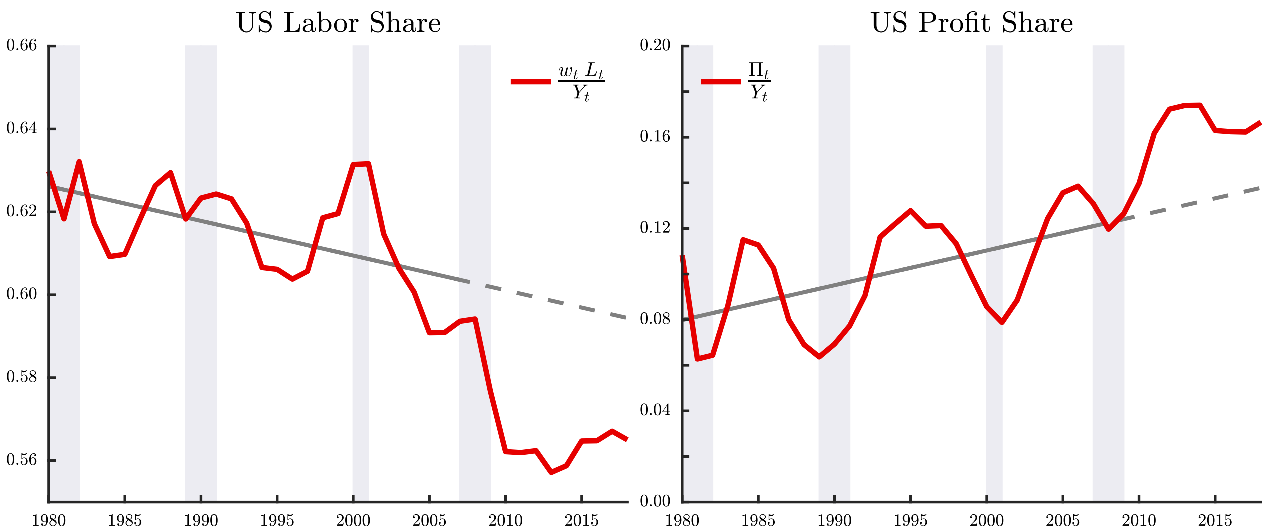
4. the decline in the firm entry rate.

Although these trends are still the object of discussion in the literature, two aspects are starting to gain consensus. First, they seem to be driven by a change in the market structure, and

[1]Ferrari: European University Institute, Via delle Fontanelle, 18, 50014 San Domenico di Fiesole, Italy (e-mail: alessandro.ferrari@eui.eu); Queirós: European University Institute, Max Weber Programme, Via dei Roccettini, 9, 50014 San Domenico di Fiesole, Italy (e-mail: francisco.queiros@eui.eu). We thank Árpád Ábrahám, Fernando Broner, Jesús Bueren, Giacomo Calzolari, Russell Cooper, Jan Eeckhout, Luca Fornaro, Manuel García-Santana, Philipp Kircher, Ramon Marimon, Giacomo Ponzetto, Edouard Schaal, Jaume Ventura and participants at the EUI Macro Working Group and the EEA 2019 Congress (Manchester). All errors are our own.

[2]See for example Eggertsson et al. (2018), Aghion et al. (2019) and Akcigit and Ates (2019a).

in particular an increasing share of large firms, with high markups and low labor shares, in production.[3] Second, they have become especially pronounced in the aftermath of the last two recessions – the 2001 crisis and, in particular, the great recession of 2008. For example, Figure 2.1 shows that both the decline in the labor share and the increase in the profit share experienced a persistent deviation from their pre-crisis trends after 2008. An identical deviation can be observed in the decline of the firm entry rate.[4] The 2008 recession had, however, a broader impact on the macroeconomy, with real GDP and most macroeconomic aggregates also deviating from their pre-crisis trends – the so called *great deviation* (Figure 2.2).



(a) **Labor Share of the US Business Sector** (b) **Profit Share of the US Business Sector**

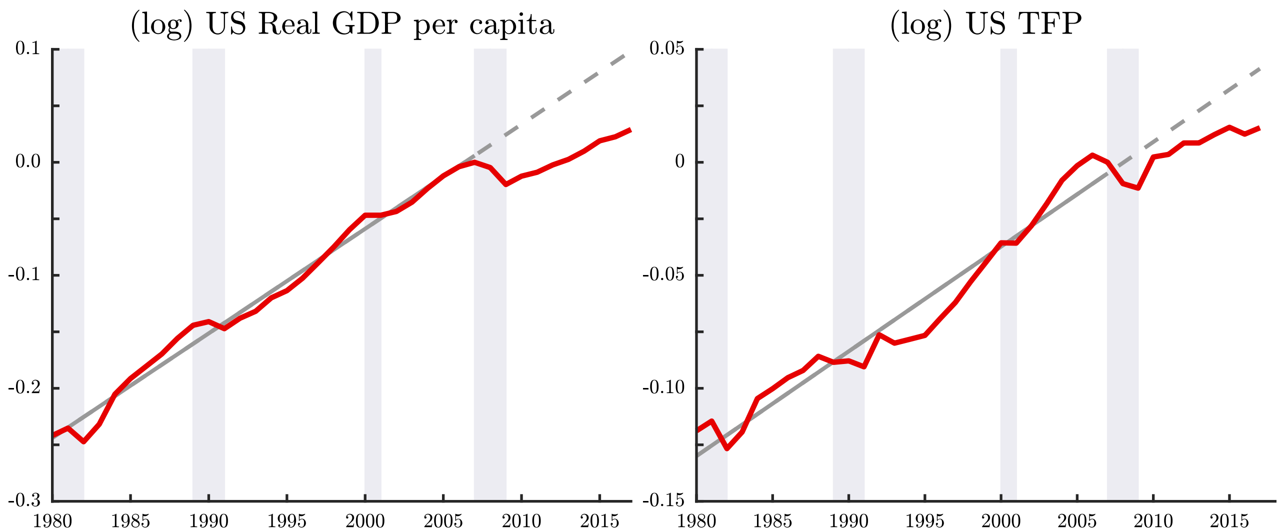
Figure 2.1: Labor and Profit Shares of the US Business Sector (1980-2018)
Note: The left panel shows the evolution of the labor share, while the right panel shows a 3-year moving average the aggregate profit share, constructed as the ratio of aggregate profits to gross value added (see Appendix 2.A.2 for details). The linear trends are computed for the pre-crisis period 1980-2007.

These observations suggest that the trends in the list above can be the consequence of forces operating at different frequencies – long-run forces that have driven a slow process of reallocation towards large firm (i.e. the trend between 1980-2007), and the persistent impact of the 2008 recession on the economy (i.e. the deviation from trend after 2008). In this paper, we investigate a natural, yet unexplored, connection between these two sets of observations. First, we ask whether a recession can have a persistent impact on the macroeconomy, capable

[3]See for example Autor et al. (2017) and Kehrig and Vincent (2018) for evidence on the labor share and De Loecker et al. (2020) for evidence on markups.

[4]See Appendix 2.A.1. This has been associated with a persistent decline in the number of active firms after 2008.

of explaining the persistent deviations from trend in both Figures 2.1 and 2.2. Second, we ask whether the long-run process of reallocation towards large firms (taking place already before the crisis) may have contributed to the severity and persistence of the crisis itself.



(a) **US Real GDP per capita** (b) **US Real TFP**

Figure 2.2: The *Great Deviation*
Note: The left panel shows real GDP per capita (from BEA). The right panel shows Fernald (2012) aggregate TFP series. Data are in logs, undetrended and centered around 2007. The linear trends are computed for the 1980-2007 period.

Our theory builds on the neoclassical growth model, with a representative household and the standard accumulation of capital. We depart, however, from the canonical model by introducing an endogenous market structure. We model a multi-industry economy, with endogenous entry and oligopolistic competition as in Atkeson and Burstein (2008). Firms face fixed production costs and make their entry decisions based on their idiosyncratic productivity draws and the state of the economy. The endogenous number of players in every market, together with the distribution of productivities, determines the overall distribution of markups and market shares. In this environment, there is a complementarity between capital accumulation and the degree of competition in product markets. On the one hand, a larger stock of capital allows more firms to break even and results in a more competitive market structure. Consequently, profit shares decline and factor shares increase. On the other hand, higher competition increases the incentives for capital accumulation. Larger factor shares result in higher factor prices (wages and rental rates) and hence a joint increase in the supply of labor and capital.

Two main insights emerge from our theory. The first is that the complementarity between capital accumulation and competition may give rise to multiple competitive regimes or stochastic steady-states. In particular, there can be regimes featuring a large stock of capital, a large number of firms and hence intense competition (low profit shares and high factor shares); and regimes featuring a low stock of capital, a small number of firms and weak competition (*low competition traps*).[5] Large (temporary) shock can trigger a transition across regimes. In particular, when an economy in a high competition regime is hit by a negative shock that significantly depresses capital accumulation, it can experience a persistent transition to a low competition regime. In such a case, the economy follows a path that, in many aspects, resembles the 2008 recession and subsequent *great deviation* – there is a persistent decline in the labor share, an increase in the profit share, as well as a persistent drop in the number of firms.

The second insight to emerge from our theory concerns the long-run process of reallocation towards large firms. We model this trend by considering a comparative static that fattens the right tail of the firm productivity distribution.[6] This shift has two main consequences. First, from a static point of view, it generates a reallocation of activity towards large, high markup firms, which is able to explain the first three facts listed above. Second, as firm heterogeneity increases, firms at the bottom of the distribution lose market share and reduce their markups. This makes them increasingly likely to exit the market upon a negative aggregate shock. In other words, a high competition regime becomes increasingly difficult to sustain, so that even a relatively small temporary shock can trigger a persistent transition to a *low competition trap*. Our model therefore suggests that larger firm differences may have increased the likelihood of a recession like the 2008 crisis, with output experiencing a persistent deviation from trend.

We calibrate our model to match first and second moments of the markup distribution of public firms in 2007. The economy features two competitive regimes. We then ask whether the model can generate a transition like the one in Figure 2.2. We feed the economy with a sequence of negative TFP shocks (to replicate the behavior of aggregate TFP in 2008-2009) and show that they can trigger a transition from the high to the low regime. Quantitatively, the model replicates the persistent drop in output, employment, investment and aggregate TFP observed in the data. It also generates a persistent drop in the labor share.

Furthermore, to evaluate the role of larger market power, we recalibrate our model to match the same moments of the markup distribution in 1985, when markups were lower and

[5]The existence of multiple regimes or stochastic steady-states does not rely on the existence of multiple equilibria. In other words, our economy can feature multiple steady-states in spite of the existence of a unique equilibrium path. The steady-state the economy will reach depends on the initial condition and the history of aggregate shocks.

[6]In our model, firm heterogeneity is driven by differences in idiosyncratic productivity. Other sources of heterogeneity (e.g. firm-specific demand shifters) would yield identical results.

less dispersed. We show that, relative to the 2007 model, the 1985 economy exhibits lower amplification and persistence. Furthermore, the size of the shock needed to trigger a transition from the high to the low regime is larger in the 1985 economy. We estimate that the 2007 economy is between 3 and 8 times more likely to experience a deep recession than the 1985 economy.[7] Overall, these results indicate that the increase in average markups and in markup dispersion may have rendered the US economy more vulnerable to aggregate shocks.

Finally, we present cross-industry evidence on our mechanism. Our model predicts that industries featuring initially a larger concentration should experience a larger contraction as the economy enters a low competition trap. We test this prediction using data from the US census and focusing again on the 2008 crisis. Consistent with the predictions of the model, we find that industries featuring a larger concentration in 2007 experienced a greater cumulative contraction over the 2007-2016 period, as well as a larger drop in the labor share.

The rest of the paper is organized as follows. Section 2.3 presents the model. Section 2.4 discusses the calibration and presents the quantitative results. Section 2.5 focuses on the US great recession and its aftermath. Section 2.7 presents the cross-industry empirical evidence. Section 2.8 concludes.

2.2 Related Literature

Our paper is related to three different strands of the literature. In the first place, there is a large literature studying the aggregate consequences of imperfect competition and variable markups. We are not the first to show how multiple equilibria and/or multiple steady-states can arise in a context of imperfect competition (see for example Cooper and John (1988), Pagano (1990), Chatterjee et al. (1993), Galí and Zilibotti (1995), Jaimovich (2007)). As in our theory, these models typically rely on a complementarity between firm entry decisions in a context of variable markups and elastic factor supply.[8] We contribute to this literature, however, by discussing the role of firm heterogeneity in generating multiplicity and shaping the response of the economy to aggregate shocks. We also provide a quantification of our mechanism and link

[7]In particular, the sequence of negative TFP shocks that we feed in the 2007 economy is not sufficient to generate a transition from the high to the low regime in 1985.

[8]Without relying on multiple equilibria or multiple steady-states, Cooper and John (2000) and Bilbiie et al. (2012) show that a combination of imperfect competition with endogenous entry can generate endogenous amplification and persistence of aggregate fluctuations.

it to the 2008 crisis.[9] Second, this paper relates to a large and growing literature documenting long-run trends in firm heterogeneity and market power. There are several signs indicating rising market power in the US and other advanced economies. For example, Autor et al. (2017) use data from the US census and document rising sales and employment concentration, while Akcigit and Ates (2019b) document a rise in patenting concentration. Other studies have documented a secular rise in price-cost markups. Using data from national accounts, Hall (2018) finds that the average sectoral markup increased from 1.12 in 1988 to 1.38 in 2015. De Loecker et al. (2020) document a steady increase in sales-weighted average markups for US public firms between 1980 and 2016.[10] This was driven by both an increasing share of large firms and by rising dispersion in the markup distribution.[11] In our model, rising dispersion in size and markups is driven by increasing productivity differences. Several studies have indeed documented a secular increase in productivity differences across firms (Andrews et al. (2015), Kehrig (2015), Decker et al. (2018)). We contribute to this literature by investigating the business cycle implications of these trends, and in particular their on the 2008 crisis and the subsequent *great deviation*.

Lastly, this paper relates to the literature focusing on the persistent impact of the 2008 crisis. Schaal and Taschereau-Dumouchel (2015) study a model with endogenous capacity utilization. Their model features a complementarity between firms' capacity utilization and aggregate output, which gives rise to multiple steady-states. Like us, they interpret the post-2008 deviation as a transition to a low steady-state. Other authors have proposed explanations based on the complementarity between firms' innovation decisions and aggregate output (Benigno and Fornaro (2017), Anzoategui et al. (2019)). Finally, Clementi et al. (2017) argue that the persistent decline in firm entry, observed after 2008, is crucial to understand the slow recovery. While we view our theory as complementary to the above-mentioned articles, we believe we are the first to link the *great deviation* to the long-run increase in firm level heterogeneity, and to propose an explanation based on the interactions between market size

[9] Our paper is also linked to the literature studying the cyclical properties of markups, which includes Rotemberg and Saloner (1986), Jaimovich and Floetotto (2008), Bils et al. (2018) and Burstein et al. (2019).

[10] Edmond et al. (2018) show that a cost-weighted average markup (as opposed to sales-weighted) displays a less pronounced upward trend. They show that a cost-weighted average markup is the one that is relevant for welfare analysis, as it accounts for the fact that high markup firms are also more productive. See also Traina (2018), Karabarbounis and Neiman (2019) and Bond et al. (2020) for a critique on the De Loecker et al. (2020) methodology.

[11] Identical findings are obtained by Díez et al. (2019) and by Calligaris et al. (2018), who use data from ORBIS (Bureau van Dijk) and include different countries in their analysis.

and market structure.

2.3 A Growth Model with Variable Markups

This section presents our theoretical framework. We start by describing the demand side and the technology structure. Then we analyze the equilibrium of a particular industry (taking aggregate variables as given). Finally, we characterize the general equilibrium.

2.3.1 Preferences

Time is discrete and indexed by $t = 0, 1, 2, \ldots$. There is a representative, infinitely-lived household with lifetime utility

$$U_t = \mathbb{E} \sum_{t=0}^{\infty} \beta^t \, U \left(C_t, L_t\right),$$

where $0 < \beta < 1$ is the discount factor, $C_t \geq 0$ is consumption of the final good and $L_t \geq 0$ is labor. We adopt a period utility function as in Greenwood et al. (1988)

$$U \left(C_t, L_t\right) = \frac{1}{1-\gamma} \left(C_t - \frac{L_t^{1+\nu}}{1+\nu}\right)^{1-\gamma}, \tag{2.1}$$

where $0 \leq \gamma \leq 1$ and $\nu > 0$.

The representative household contains many individual members, which will be denoted by j. Each individual member can run a firm in the corporate sector. We assume that if two or more individuals run a firm in the same industry, they will behave in a non-cooperative way – i.e. they will compete against each other and will not collude. Nevertheless, all individuals will pool together the profits they make. Hence there is a single dynamic budget constraint

$$K_{t+1} = \left[R_t + (1-\delta)\right] K_t + W_t L_t + \Pi_t^N - C_t, \tag{2.2}$$

where K_t is capital, R_t is the rental rate, W_t is the wage rate and $\Pi_t^N = \sum_j \Pi_j^N$ are the profits accruing from all the firms in the economy net of fixed production costs. Capital depreciates at rate $0 \leq \delta \leq 1$ and factor prices R_t and W_t are taken as given. The representative household therefore maximizes (2.1) subject to (2.2). Our choice of GHH preferences implies that the aggregate labor supply is a simple function of the wage rate

$$L_t = W_t^{1/\nu},$$

Hence, ν is the inverse of the wage elasticity of labor supply.[12]

2.3.2 Technology

There is a final good (the *numeraire*), which is a CES aggregate of I different industries

$$Y_t = \left(\sum_{i=1}^{I} y_{it}^{\rho} \right)^{\frac{1}{\rho}},$$

where y_{it} is the quantity of industry $i \in [0,1]$, $0 < \rho < 1$ and $\sigma_I = \dfrac{1}{1-\rho}$ is the elasticity of substitution across industries. I is assumed to be large, so that each individual industry has a negligible size in the economy. The output of each industry i is itself a CES composite of differentiated goods or varieties

$$y_{it} = \left(\sum_{j=1}^{n_{it}} y_{jit}^{\eta} \right)^{\frac{1}{\eta}},$$

where n_{it} is the number of active firms in industry i at time t (to be determined endogenously), $0 < \eta \leq 1$ and $\sigma_G = \dfrac{1}{1-\eta}$ is the within-industry elasticity of substitution. Following Atkeson and Burstein (2008), we assume that goods are more easily substitutable within industries than across industries.

Assumption. $0 < \rho < \eta \leq 1$

Given these assumptions, the inverse demand for each variety j in industry i is given by

$$p_{ijt} = \left(\frac{Y_t}{y_{it}} \right)^{1-\rho} \left(\frac{y_{it}}{y_{ijt}} \right)^{1-\eta}. \tag{2.3}$$

We assume that in every industry $i \in \{0, \ldots, I\}$ there is a maximum number of entrepreneurs $N \in \mathbb{N}$, so that $n_{it} \leq N$. Entrepreneur j can produce his variety by combining capital k_{ijt} and labor l_{ijt} through a Cobb-Douglas technology

$$y_{it} = \underbrace{e^{z_t} \, \pi_{ij}}_{\tau_{ijt}} \left(k_{ijt}\right)^{\alpha} \left(l_{ijt}\right)^{1-\alpha}. \tag{2.4}$$

[12] From the perspective of the household, labor and entrepreneurial decisions are separable. We abstract from an occupational decision problem (becoming an entrepreneur *versus* working) for simplicity. This assumption does not however imply that there is a fixed supply of entrepreneurs, since the equilibrium number of entrepreneurs (and hence the number of active firms) will be endogenous.

Note that the productivity of each entrepreneur τ_{ijt} is the product of two terms (i) a time-varying aggregate component e^{z_t} (common to all industries and types) and (ii) a time-invariant idiosyncratic term π_{ij}. We refer to z_t as aggregate productivity and to π_{ij} as j's idiosyncratic productivity. Aggregate productivity z_t follows an auto-regressive process

$$z_t = \phi_z z_{t-1} + \varepsilon_t, \tag{2.5}$$

with $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$.With no loss of generality, we order idiosyncratic productivities according to

$$\pi_{i1} \geq \pi_{i2} \geq \pi_{i3} > \cdots .$$

Labor is hired at the competitive wage W_t and capital at the rental rate R_t. Entrepreneur j can thus produce her variety with constant marginal cost Θ_t/τ_{ijt}, where

$$\Theta_t \equiv \left(\frac{R_t}{\alpha}\right)^{\alpha}\left(\frac{W_t}{1-\alpha}\right)^{1-\alpha}$$

is the marginal cost function for a Cobb-Douglas technology with unit productivity. We refer to Θ_t as the *factor cost index*. In addition to all variable costs, the production of each variety entails a fixed production cost $c_i \geq 0$ per period (which can be possibly different across industries). Such a cost is in units of the *numeraire*.[13]

2.3.3 Market Structure

To conclude the description of the model, we must specify the way firms interact. We assume that all firms that enter (and thus incur the fixed cost c_i) play a static Cournot game: they will simultaneously announce quantities, taking the output of the other competitors as given.[14] Therefore, each entrepreneur j solves

$$\max_{y_{ijt}} \quad \left(p_{ijt} - \frac{\Theta_t}{\tau_{ijt}}\right) y_{ijt} \qquad \text{s.t.} \quad \begin{aligned} p_{ijt} &= \left(\frac{Y_t}{y_{it}}\right)^{1-\rho}\left(\frac{y_{it}}{y_{ijt}}\right)^{1-\eta} \\ y_{it} &= \left(\sum_{k=1}^{n_{it}} y_{kit}^{\eta}\right)^{\frac{1}{\eta}} . \end{aligned} \tag{2.6}$$

---

[13]According to this formulation, fixed production costs do not change with factor prices. We introduce variable fixed costs in an extension (Appendix 2.I.3).

[14]We follow Atkeson and Burstein (2008) and assume that firms make sequential entry decisions in reverse order of productivity.

The solution to this static problem yields a system of n_{it} first order conditions (one for each firm)

$$p_{ijt}\left[\eta - (\eta - \rho)\, s_{ijt}\right] = \frac{\Theta_t}{\tau_{ijt}}, \tag{2.7}$$

where s_{ijt} is the market share of firm $j = 1, \ldots, n_{it}$.[15] From 2.7 entrepreneur j will charge a markup

$$\mu_{ijt} = \frac{1}{\eta - (\eta - \rho)\, s_{ijt}} \tag{2.8}$$

over his marginal cost Θ_t/τ_{ijt}. Note the two following implications of equation 2.8. First, equation (2.8) establishes a positive relationship between market shares and markups. To understand such a relationship, note that firms internalize the impact of their size on the price they charge p_{ijt}. Large firms end up restricting output disproportionately more (relative to productivity), thereby charging a high markup. Second, market shares are themselves a positive function of revenue TFP $p_{ijt}\,\tau_{ijt}$, as equation (2.7) also highlights. Our model thus features a positive association between revenue productivity, size and markups. Therefore, a shock that generates larger productivity differences across firms will also lead to larger markup dispersion (a point to which we will return later).

The set of first order conditions in (2.7) defines a system of n_{it} non-linear equations in the prices $\{p_{ijt}\}_{j=1}^{n_{it}}$. Such a system admits a close-form solution only in the limit case in which there is no differentiation within an industry ($\eta = 1$), as shown in Appendix 2.G.1.

To conclude the description of the industry equilibrium, we need to determine the number of active firms n_{it}. To this end, let

$$\Pi\left(\, j \,,\, n_{it} \,,\, \mathcal{F}_{it} \,,\, X_t \,\right) \coloneqq \left(p_{it} - \frac{\Theta_t}{\tau_{ijt}}\right) y_{ijt}$$

denote the equilibrium profits of firm $j \leq n_{it}$ in industry i (gross of the fixed production cost), when there are n_{it} active firms, given a productivity distribution $\mathcal{F}_{it} \coloneqq \{\, \pi_{i1} \,,\, \pi_{i2} \,,\, \ldots \}$ and a vector of aggregate variables $X_t \coloneqq [\, z_t \,,\, Y_t \,,\, \Theta_t \,]$.

The equilibrium number of firms must be such that (i) the profits of each active firm are not lower than the fixed cost c_i and (ii) if an additional firm were to enter, its profits would be lower than the fixed cost. Mathematically, an interior solution $n^*_{it} < N$ to the equilibrium number of firms must satisfy

$$\left[\Pi\left(n^*_{it}, n^*_{it}, \mathcal{F}_i, X_t\right) - c_i\right]\left[\Pi\left(n^*_{it}+1, n^*_{it}+1, \mathcal{F}_i, X_t\right) - c_i\right] \leq 0. \tag{2.9}$$

[15]It is defined as $s_{ijt} = p_{ijt}\, y_{ijt} / \sum_{k=1}^{n_{it}} (p_{ikt}\, y_{ikt})$

Proposition 2.7 in Appendix 2.G.1 provides an analytical characterization of the profit function under the special case of $\eta = 1$. In particular, we show that the profits of any firm j increase in its own idiosyncratic productivity π_{ij} and decrease in the idiosyncratic productivity of all the other firms π_{ik}. Therefore, as top firms increase their productivity advantage over small firms, small firms make lower profits become closer to their break even point (*ceteris paribus*). This is key to understand some aggregate results that we describe next.

2.3.4 General Equilibrium

Equilibrium Definition We start by defining an equilibrium for this economy. Denoting the history of aggregate productivity shocks by $Z^t = \{z_t, z_{t-1}, ...\}$ we have the following definition.

Definition 2.1
A general equilibrium consists of a sequence of household policies $\{C_t(Z^t), K_t(Z^t), L_t(Z^t)\}$, *firm policies* $\{y_{ijt}(Z^t), k_{ijt}(Z^t), l_{ijt}(Z^t)\}$, *and a set of active firms* $\{n_{it}(Z^t)\}_{i=1}^{I}$ *such that*

(i) households optimize

(ii) all active firms optimize

(iii) all active firms do not make a loss

(iv) no additional firm is willing to enter

(v) capital and labor markets clear

Static Equilibrium

We now describe the general equilibrium of this economy. We start by focusing on a static equilibrium, in which we describe production and labor supply decisions, taking the aggregate level of capital K_t as given. Later on, we describe the equilibrium dynamics.

Aggregate Production Function Given a $(I \times N)$ matrix of productivity draws $\mathbb{Z}_{\tau t}$ and when the n_{it} most productive firms of industry i are active, aggregate output can be written as

$$Y_t = \Phi\left(\mathbb{Z}_{\tau t}, \{n_{it}\}_{i=1}^{I}\right) \ L_t^{1-\alpha} K_t^{\alpha}. \tag{2.10}$$

The term $\Phi\left(\mathbb{Z}_{\tau t}, \{n_{it}\}_{i=1}^{I}\right)$ represents aggregate TFP and is a function of the number of active firms, individual firm productivities and market shares. An expression for $\Phi(\cdot)$ is provided in Appendix 2.B.1.

Factor Prices and Factor Shares We can aggregate firms' best responses, given by equation (2.7), to find an expression for the aggregate factor cost index. Given a $(I \times N)$ matrix of productivity draws $\mathbb{Z}_{\tau t}$ and when there the n_{it} most productive firms are active in industry i, the equilibrium factor cost index is equal to

$$\Theta\left(\mathbb{Z}_{\tau t}, \{n_{it}\}_{i=1}^{I}\right) = \left\{\sum_{i=1}^{I}\left[\sum_{j=1}^{n_{it}}\left(\frac{\tau_{ijt}}{\mu_{ijt}}\right)^{\frac{\eta}{1-\eta}}\right]^{\frac{1-\eta}{\eta}\frac{\rho}{1-\rho}}\right\}^{\frac{1-\rho}{\rho}}. \tag{2.11}$$

Note that factor price index will be a negative function of markups μ_{ijt}. From (2.11), we can also see how it varies with the number of firms. Suppose that the number of firms in each industry i increases from n_{it} to $n_{it}+1$. In such a case, $\Theta(\cdot)$ changes for two reasons. First, there is one additional firm in each industry, which necessarily increases factor demand even when all the remaining players do change their markups (*entry effect*). Second, the entry of one additional firm increases the level of competition in the industry, i.e. the preexisting firms respond by increasing factor demand and cutting their markups (*competition effect*).

$$\left\{\sum_{i=1}^{I}\left[\underbrace{\sum_{j=1}^{n_{it}}\left(\frac{\tau_{ijt}}{\tilde{\mu}_{ijt}}\right)^{\frac{\eta}{1-\eta}}}_{\uparrow(\text{competition effect})} + \underbrace{\left(\frac{\tau_{ikt}}{\tilde{\mu}_{ikt}}\right)^{\frac{\eta}{1-\eta}}}_{\uparrow(\text{entry effect})}\right]^{\frac{1-\eta}{\eta}\frac{\rho}{1-\rho}}\right\}^{\frac{1-\rho}{\rho}} > \left\{\sum_{i=1}^{I}\left[\sum_{j=1}^{n_{it}}\left(\frac{\tau_{ijt}}{\mu_{ijt}}\right)^{\frac{\eta}{1-\eta}}\right]^{\frac{1-\eta}{\eta}\frac{\rho}{1-\rho}}\right\}^{\frac{1-\rho}{\rho}}.$$

Factor Market Clearing Having obtained an expression for the aggregate factor cost index, we can determine the factor demand schedules for labor L_t and capital K_t

$$\begin{aligned} W_t &= (1-\alpha)\ \Theta\left(\mathbb{Z}_{\tau t}, \{n_{it}\}_{i=1}^{I}\right)\ L_t^{-\alpha}K_t^{\alpha}, \\ R_t &= \alpha\ \Theta\left(\mathbb{Z}_{\tau t}, \{n_{it}\}_{i=1}^{I}\right)\ L_t^{1-\alpha}K_t^{\alpha-1}. \end{aligned} \tag{2.12}$$

These two demand schedules can be combined with the labor and capital supply equations

$$\begin{aligned} L_t^S &= W_t^{1/\nu}, \\ K_t^S &= K_t \end{aligned} \tag{2.13}$$

to determine the factor market equilibrium. Combining equations (2.12) and (2.13) we can obtain an expression for equilibrium employment

$$L_t = \left[(1-\alpha)\,\Theta\left(\mathbb{Z}_{\tau t}, \{n_{it}\}_{i=1}^{I}\right)\right]^{\frac{1}{\nu+\alpha}} K_t^{\frac{\alpha}{\nu+\alpha}}. \tag{2.14}$$

Finally, we can combine equations (2.10) and (2.14) to write aggregate output as a function of the aggregate capital stock K_t, the productivity distribution $\mathbb{Z}_{\tau t}$ and the set of active firms

$\{n_{it}\}_{i=1}^{I}$

$$Y_t = \Phi\left(\mathbb{Z}_{\tau t}, \{n_{it}\}_{i=1}^{I}\right)\left[(1-\alpha)\,\Theta\left(\mathbb{Z}_{\tau t}, \{n_{it}\}_{i=1}^{I}\right)\right]^{\frac{1-\alpha}{\nu+\alpha}} K_t^{\alpha\frac{1+\nu}{\nu+\alpha}}. \tag{2.15}$$

To conclude the characterization of the static equilibrium, we need to determine the set of active firms $\{n_{it}\}_{i=1}^{I}$.

Equilibrium Set of Firms The number of active firms in each industry i is jointly determined by equations (2.11), (2.15) and the set of inequalities defined in (2.9). Such a joint system does not admit a general analytical characterization. We can nevertheless analyze the particular case in which all industries are identical. Proposition 2.1 states the conditions for a symmetric equilibrium with n firms per industry.

Proposition 2.1
Suppose that all industries have the same distribution of idiosyncratic productivities $\mathcal{F}_i = \mathcal{F}$. Then, there can be a symmetric equilibrium with n firms per industry if K_t is such that

$$\underline{K}(\mathcal{F}, n) \leq K_t \leq \overline{K}(\mathcal{F}, n),$$

The bounds $\overline{K}(\cdot)$ and $\underline{K}(\cdot)$ are both increasing in n.

Proof. See Appendix 2.B.3. ■

Intuitively, K_t must be (i) sufficiently large so that all existing n firms can break even, (ii) but cannot be too high, for otherwise an additional firm could profitably enter in at least one industry. When $K_t \in \left[\overline{K}(\mathcal{F}, n), \underline{K}(\mathcal{F}, n+1)\right]$, a symmetric equilibrium is not possible. In such a case, the economy will have some industries with n players, and others with $n+1$ players. The fraction of industries with $n+1$ firms will be such that the last firm exactly breaks even, i.e.[16]

$$\Pi(n+1, n+1, \mathcal{F}, \Theta_t, Y_t) = c.$$

Figure 2.3 illustrates these results. It shows aggregate output, the profit share, and the equilibrium wage and rental rate as a function of aggregate capital K_t. In the regions represented by the full line, the economy features a symmetric equilibrium (all industries have the same number of firms). In the regions represented by the dashed line, not all industries have the same number of firms. When the capital stock is such that $\underline{K}(1) \leq \overline{K} \leq (1)$, the economy can accommodate only one firm per industry – it will consist of a collection of identical monopolies. As capital increases and surpasses $\overline{K}(1)$, then some industries will have a second player. When it achieves $\underline{K}(2)$, all industries will have two players. As one can see, output Y_t

[16] A detailed characterization of this non-symmetric equilibrium is provided in online appendix 2.H.2.

is not globally concave in the capital stock K_t. This fact is explained by a complementarity between firm entry and labor supply. As the average number of firms increases (say from $n = 1$ to $n = 2$), competition becomes more intense and the profit share decreases (top right panel). This translates into a larger factor share and a disproportionately larger wage (bottom left panel) and labor supply (through equation (2.13)). The fact that factor shares increase as the economy transitions into a more competitive regime also explains why, as in the case represented in Figure 2.3, the rental rate R_t is not strictly decreasing in the aggregate capital stock K_t.[17].

To conclude, note that in the example of Figure 2.3 the equilibrium is always unique, i.e. the aggregate capital stock (K_t) pins down the number of firms per industry (n_t) and all other equilibrium variables. However, if the complementarity between firm entry and labor supply is strong enough, there can be generate multiple equilibria. See Appendix 2.B.4 for a discussion.

Equilibrium Dynamics

We are now ready to explore the dynamic properties of our economy. Even though we cannot derive a general law of motion for our economy in closed form, we can nevertheless characterize the steady-state savings rate.

Proposition 2.2
(Steady-State Savings Rate) In a steady-state with a fixed distribution of firm productivities $\mathbb{Z}_\tau$ and a fixed set of active firms $\{n_i\}_{i=1}^{I}$, the aggregate savings rate is equal to

$$s^* = \frac{\beta\delta}{1-(1-\delta)\beta}\,\alpha\,\Omega\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^{I}\right).$$

where $\Omega(\cdot)$ denotes the aggregate factor share

$$\Omega(\cdot) \coloneqq \frac{WL+RK}{Y}$$

Proof. See Appendix 2.H.3. ■

What Proposition 2.2 says is that the aggregate savings rate will increase with the level of competition in the economy. Note that $\Omega(\cdot)$ reflects the share of production accruing to labor and capital; the aggregate profit share (gross of fixed production costs) is thus equal to $1-\Omega$.

[17]Such a result helps us understand how multiple steady-states can occur. Note that the steady-state rental rate is pinned down by the household's discount factor β. If the the steady-state rental rate is such that it crosses the map represented in Figure 2.3 more than once, then steady-state multiplicity occurs as we will see in the next subsection.

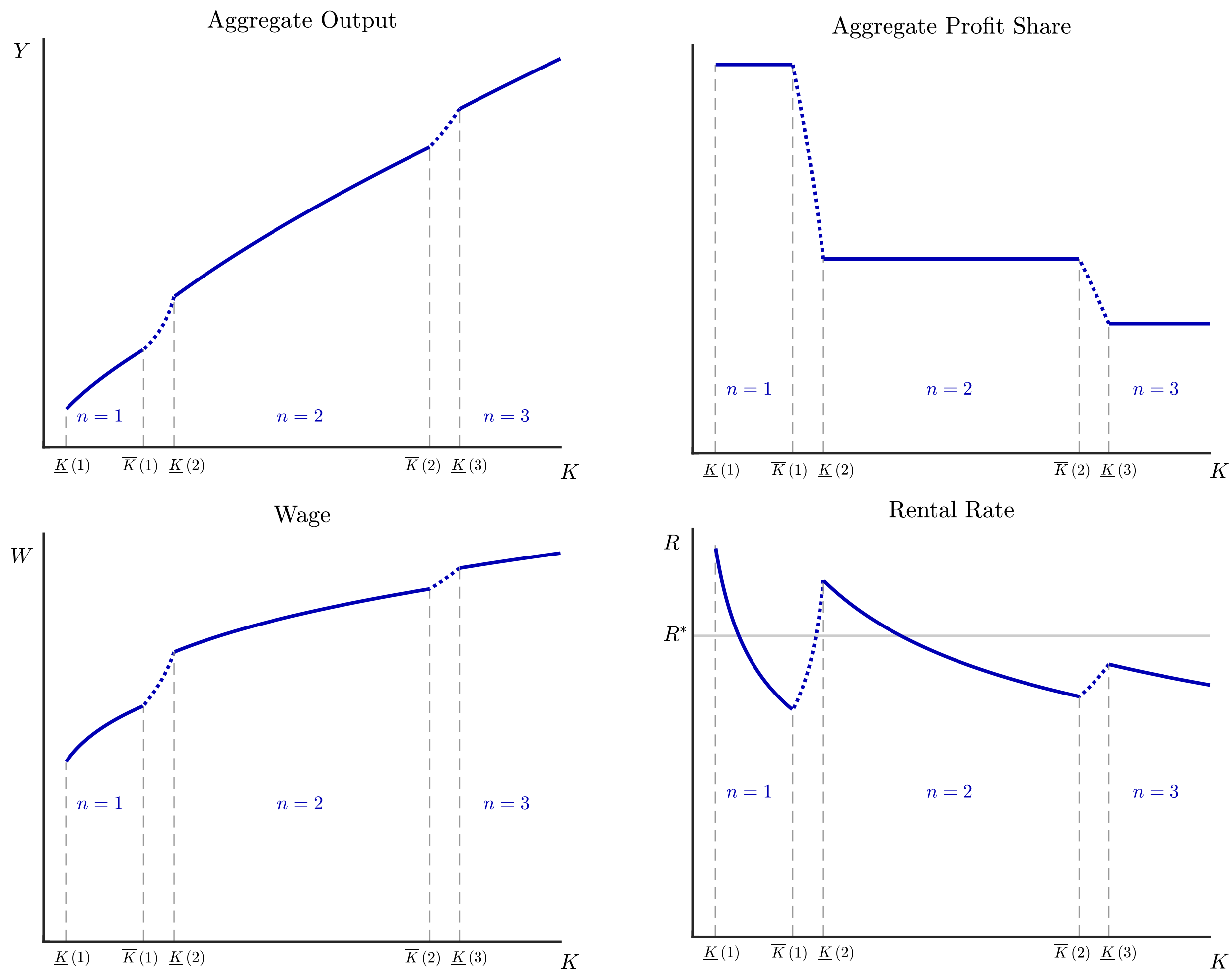


Figure 2.3: Static equilibrium
Note: the figure shows how output, the profit share and input prices move with capital. Solid segments represent economies with symmetric market structure (same number of firms), while dotted segments represent non-symmetric equilibria. We use $\alpha = 1/3$, $\rho = 0.6$, $\eta = 1$, $\nu = 0.4$, $\pi_{ij} = 1$ and $c_i = 0.015$.

Although we cannot provide a general characterization of $\Omega(\cdot)$, we can, however, characterize it in the particular case in which (i) $\eta = 1$ and (ii) all industries are identical.

Proposition 2.3
Let $\Omega(\mathcal{F}, n)$ be the aggregate factor share in a symmetric equilibrium in which all industries are identical (have the the same productivity distribution $\mathcal{F}$ and the same number of firms n). We have that

1. *$\Omega(\mathcal{F}, n)$ increases in n*

$$\Omega(\mathcal{F}, n+1) > \Omega(\mathcal{F}, n).$$

2. Let π_j be an idiosyncratic productivity type such that $\pi_j \geq \frac{1}{n}\sum_{k=1}^{n}\pi_k$. Suppose that π_j increases to $\pi_j' > \pi_j$ in all industries but all other types remain unchanged. Then, the new distribution $\mathcal{F}'$ is such that

$$\Omega\left(\mathcal{F}', n\right) < \Omega\left(\mathcal{F}, n\right).$$

Proof. See Appendix 2.H.1. ■

The previous proposition states two important results. First, the larger is the number of firms per industry, the more intense is the degree of product market competition and hence the aggregate factor share $\Omega(\cdot)$. Second, the aggregate factor share decreases when the distribution of individual productivities becomes more dispersed. The intuition is simple. In every industry, high productivity firms are larger and charge higher markups. As large firms are able to increase their markups even further, the aggregate profit share increases and the aggregate factor share decreases. Proposition 2.3 therefore says that rising productivity/size differences across firms generate lower capital and labor shares. Proposition 2.2 says that lower capital and labor shares translate into a lower steady-state savings rate.

Law of Motion Figure 2.4 below shows the law of motion of this economy. For the sake of clarity, we impose a common distribution of productivities $\mathcal{F}_i = \mathcal{F}$ so that all industries are ex-ante identical; we also fix the level of aggregate productivity $z_t = 1$. Note that the law of motion is not globally concave and exhibits a convex region for $K_t \in \left[\overline{K}(1), \underline{K}(2)\right]$. Such a convexity occurs for the mechanism highlighted earlier – as K_t increases, the economy moves towards a more competitive regime, with a larger factor share $\Omega(\mathcal{F}, 2) > \Omega(\mathcal{F}, 1)$. A larger factor share results in a higher wage and labor supply (resulting in larger Y_t for a given K_t), but also in a higher savings rate (resulting in larger K_{t+1} for a given Y_t). Because of this complementarity between capital accumulation and competition, the law of motion exhibits two steady-states: one where all industries are a monopoly (K_1^{ss}), and another where all industries are a duopoly (K_2^{ss}).[18] The steady-state to which the economy will converge depend on its initial condition K_0.[19] Note that despite the existence of two steady-states, this example features a unique equilibrium: there is a unique value of K_{t+1} for each value of K_t (the state variable).

[18] The existence of exactly two steady-states is obviously not guaranteed. Figures 2.2 and 2.3 in Appendix 2.H.3 show examples of economies with one or three steady-states.

[19] Figure 2.4 represents the law of motion for a fixed value of aggregate productivity $z_t = 1$. However, the law of motion will change with aggregate productivity z_t. Appendix 2.B.6 shows an example with stochastic productivity.

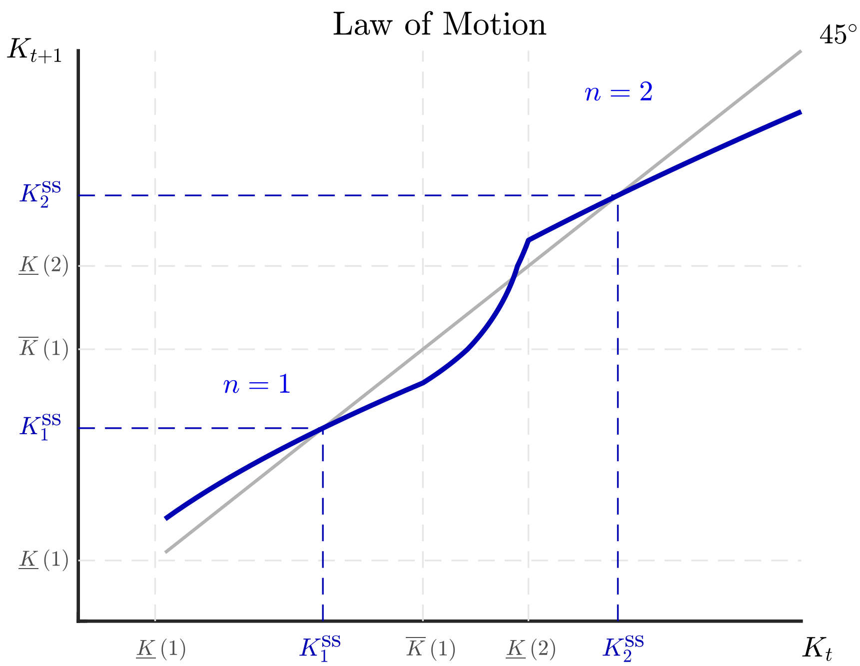
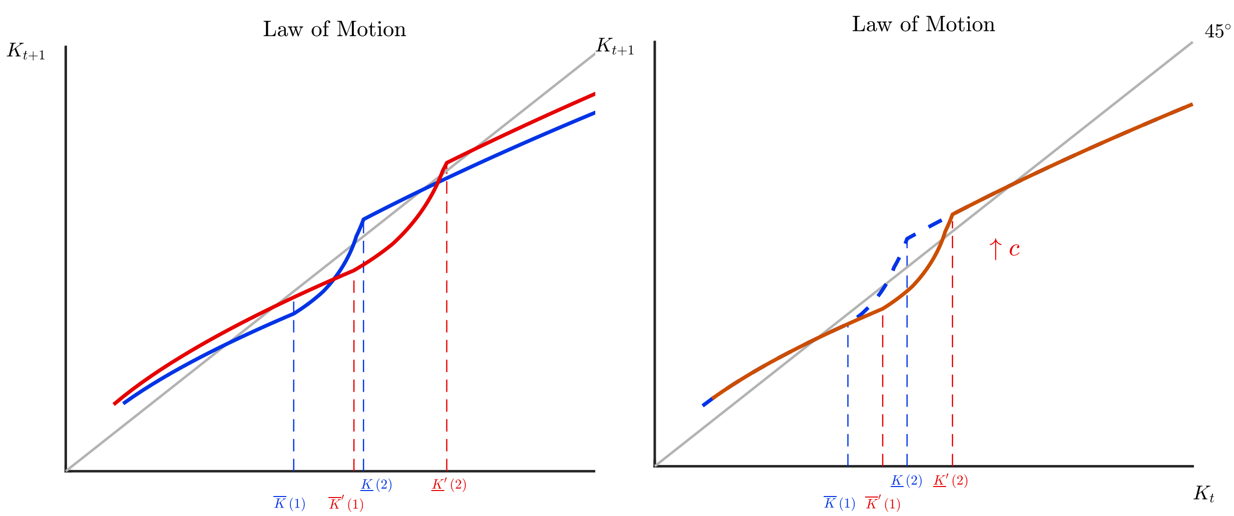


Figure 2.4: Law of motion
Note: the figure plots the law of motion of capital. This example features two stable steady states and an unstable one. We use $\gamma = 1$, $\delta = 1$, $\alpha = 1/3$, $\rho = 0.6$, $\eta = 1$, $\nu = 0.4$, $\pi_{ij} = 1$ and $c_i = 0.015$.

This simple law of motion already provides a theory for some of the facts described in the introduction. Suppose that the economy starts in a neighborhood of the high steady-state K_2^{SS} but is hit by a temporary shock that brings capital close to $\overline{K}(1)$. The economy will then enter the basin of attraction of the low steady-state K_1^{SS} and experience a persistent drop in aggregate output, a persistent decline in the labor share and an increase in the profit share. Viewed through the lens of this model, a transition to a less competitive regime can explain the deviation of output from trend after 2008.

However, as discussed in the introduction, several signs suggest a long-run increase in market power since at least the 1980. This seems to be associated with rising firm heterogeneity and a reallocation of activity towards large firms (Van Reenen (2018)). Since these trends can have a direct impact on our mechanism, we ask what are the dynamics consequences of rising firm heterogeneity. We will also discuss the consequences of rising fixed costs.

Rising Firm Differences Let us now revisit the example of Figure 2.4. Suppose that in each industry there is a productive firm with productivity $\pi_1 = \pi > 1$ (the leader), while all the other firms $j = 2, 3, \ldots$ have productivity $\pi_j = 1$ (the followers). What happens when π increases? Figure 2.5a represents the effects of an increase in π. The law of motion under the new value of π is represented in red. Two facts stand out. First, the two concave segments of the law of motion (representing the symmetric equilibria with $n = 1$ and with $n = 2$ firms) move up. This fact simply represents an expansion in the economy's production possibility frontier – because of the larger productivity advantage of the leaders, aggregate output will increase for any fixed number of active firms n. Second, part of the convex segment lying between $\overline{K}(1)$ and $\underline{K}(2)$ lies below the initial law of motion. This change reflects the fact that the leaders increase their productivity over the followers. Because of such a larger advantage, the followers can only enter at increasingly larger levels of aggregate capital, which results in a simultaneous increase in $\overline{K}(1)$ and $\underline{K}(2)$. Therefore, the increase in π has two effects on the law of motion: (i) for a fixed number of firms, it results in an expansion of the production possibility frontier, (ii) however an equilibrium with $n \geq 2$ can only be sustained with larger values of aggregate capital.



(a) Larger productivity differences

(b) Larger fixed costs

Figure 2.5: Law of motion (comparative statics)
Note: the figure shows changes in the law of motion of capital after an increase in productivity differences (Panel (a)) and in fixed costs (Panel (b)). We use $\gamma = 1$, $\delta = 1$, $\alpha = 1/3$, $\rho = 0.6$, $\eta = 1$ and $\nu = 0.4$.

Note that the increase in π results in a reduction of the basin of attraction of the high steady-state. Although we cannot provide a full analytical characterization of the impact of π on the basin of attraction (i.e. whether it always shrinks or can actually increase), we can nevertheless characterize its impact on the subset of the basin of attraction that falls under a symmetric equilibrium. Proposition 2.4 states conditions under which an increase in the productivity of the leader π reduces the basin of attraction of the high steady-state.

Proposition 2.4

Let $\eta = 1$ and suppose that all industries and firms are identical to start ($\pi_j = \pi\ \forall j$). Let $K^*(n)$ be a steady-state with n firms. Then we have that

$$\frac{\partial}{\partial \pi_k}\left[\frac{K^*(n)}{\underline{K}(n)}\right] < 0.$$

if and only if

$$\frac{1/\nu + \alpha}{1-\alpha} < \frac{(4 + 1/n)\left[n - (1-\rho)\right] - n}{1-\rho}.$$

Proof. See Appendix 2.B.5. ■

To understand Proposition 2.4, suppose that some type π_k experiences a productivity increase, while the productivity of all other types remains constant. Proposition 2.4 gives the condition under which the basin of attraction falling under $\left[\underline{K}(n), K^*(n)\right]$ shrinks. This is satisfied provided that (i) labor supply is relatively inelastic (low $1/\nu$), (ii) there are strong diminishing returns to capital (low α) and (iii) a low degree of product differentiation (high ρ). To have an intuition, note that labor supply needs to be relatively rigid for types π_n to be crowded out as types $\pi_k \geq \pi_n$ expand. The degree of product differentiation cannot be too high so that markups are low and industries are characterized by a *winner-takes-it-all* type of dynamics.[20]

Rising Fixed Costs The decline in product market competition since the 1980s may be also explained, through the lens of our model, by rising fixed costs. How can larger fixed costs affect the law of motion represented in Figure 2.4? First, we can show that if a steady-state with n firms exists, and all firms make strictly positive profits, the existence and level of such a steady-state $\left(K_n^{SS}\right)$ will not be affected by a marginal increase in c_f (see Appendix 2.B.5 for a proof). A larger fixed cost will however result in a larger level of aggregate capital consistent with n firms per industry, $\underline{K}(n)$ (see Appendix 2.B.3). Therefore, the basin of attraction of the largest steady-state shrinks, as shown in Figure 2.5b. This means that rising fixed costs also make the highest steady-state more fragile, so that now even relatively small shocks can trigger a transition to the the low competition trap. This result is stated in the next proposition.

Proposition 2.5

Let $\eta = 1$. Let $K^*(n)$ be a steady-state with n firms and suppose all firms make strictly positive

[20] When this condition is not satisfied, an increase in the productivity of the leader may increase the basin of attraction of the high steady-state. In this case, a larger production possibility frontier makes entry easier for the followers.

profits. Then we have that

$$\frac{\partial}{\partial c_i}\left[\frac{K^*(n)}{\underline{K}(n)}\right] < 0.$$

Proof. See Appendix 2.B.5. ∎

Discussion We conclude by summarizing two keys insights of our theory, which we think are relevant to understand the US growth experience after 2008. The first is that a complementarity between competition and factor supply can generate multiple competitive regimes or steady-states. A transition from a high competition to a low competition regime can in many aspects describe the 2008 recession and the subsequent great deviation. Second, changes in technology that result in larger market power (e.g. larger productivity differences across firms or larger fixed costs) make high competition regimes more difficult to sustain, and transitions to low competition traps easier to occur. Our model therefore suggests that the US economy, experiencing a long-run increase in markups and concentration since the 1980s, became increasing vulnerable to transitions like this.

We next use a calibrated version of our model to ask whether it can replicate the behaviour of the US economy in the aftermath of the 2008 crisis. We also perform some counterfactual exercises to quantify the impact of rising market power.

2.4 Quantitative Results

The goal of this section is to develop a quantitative version of the model described so far and use it to evaluate the model economy response to a recession and do policy experiments. We start by describing the calibration.

There are two important objects we need to parametrize – the distribution governing idiosyncratic productivity draws and the distribution of fixed production costs. We assume that firms draw their idiosyncratic productivities from a Pareto distribution with tail parameter λ

$$\pi_{ij} \sim \mathrm{Pa}(\lambda).$$

Recall that each industry i will be characterized by N such draws. Since N is a finite number, industries have different ex-post distributions of idiosyncratic productivities $\{\pi_{ij}\}_{j=1}^N$.

Furthermore, we assume that there are two types of industries – a fraction f_{comp} of all industries have a zero fixed cost $c_i = 0$, whereas the remaining fraction $1 - f_{\text{comp}}$ faces a positive fixed production cost $c_i = c > 0$. We hence have that

$$c_i = \begin{cases} 0 & \text{if } i \leq f_{\text{comp}} \cdot I \\ c & \text{if } i > f_{\text{comp}} \cdot I \end{cases}.$$

This is a parsimonious way of introducing heterogeneity in barriers to entry across industries. Two aspects should however be explained. First, in industries with zero fixed cost, the extensive margin will be shut down as all potential N entrants will always be active. Note however these industries will not necessarily operate close to perfect competition, as there can be large productivity differences across firms, resulting in high concentration and large markups for top players. The parameter f_{comp}, which measures the importance of this sector, will be calibrated to match the share of aggregate employment allocated to non-concentrated industries, as explained below.

Second, we assume that there is a common fixed cost $c > 0$ among all *noncompetitive* industries. Although we make this assumption mostly for simplicity, we should highlight that it is not completely innocuous. In particular, when there are differences in fixed costs within these industries, and if these differences are large, multiplicity may disappear – since for multiplicity to arise, we need a sufficiently large number of industries that move together. Recall however that, even if they share the same fixed cost $c > 0$, *noncompetitive* industries will still be heterogeneous, as they will have different ex-post distributions of idiosyncratic productivity draws $\{\pi_{ij}\}_{j=1}^{N}$. These industries may display in fact a different number of players, as we will see below.

2.4.1 Calibration

We next describe the calibration of all the parameters. The model is calibrated at a quarterly frequency. Under the parameters we use, the economy will feature two steady-states. Our calibration strategy relies on the interpretation that the economy is in the high steady-state. Some parameters are standard and taken from the literature. For the preference parameters, we work with an annualized discount factor of 0.96 and set $\gamma = 1$ to have log utility. We set $\nu = 0.4$ as in Jaimovich and Rebelo (2009), which implies a Frisch elasticity of 2.5. We set the capital elasticity to $\alpha = 0.3$ and assume a 10% depreciation rate. For the two parameters governing the elasticities of substitution, we follow Mongey (2019) and use $\sigma_I = 1.5$ and $\sigma_G = 10$. These two parameters are important for the results, as they determine the degree of complementarity between capital accumulation and competition. Edmond et al. (2015) estimate $\sigma_I = 1.24$ and $\sigma_G = 10.5$ in a static trade model with oligopolistic competition. Atkeson and Burstein (2008) use $\sigma_I \approx 1$ and $\sigma_G = 10$.[21] In general, increasing the elasticity of substitution across industries σ_I depresses markups and weakens the complementarity between capital accumulation and

[21] Several other studies also use $\sigma_I \approx 1$ so that the final good is a Cobb-Douglas aggregate of the different industries. See, for example, Hsieh and Klenow (2009) and Hottman et al. (2016).

competition, making steady-state multiplicity harder to arise.[22] Therefore, we see $\sigma_I = 1.5$ as a conservative choice. In Appendix 2.I.2, we check robustness with $\sigma_I = 1.25$ and $\sigma_I = 2$.

We set the number of industries to $I = 5,000$ and the maximum number of firms per industry to $N = 100$. The maximum number of firms per industry will play a role in the *competitive* industries (i.e. those with zero fixed cost, and where all potential firms always produce). We perform robustness exercises with $N = 50$ and $N = 200$ and obtain similar results.

Description	Parameter	Value	Source/Target
Between-Industry ES	σ_I	1.5	Mongey (2019)
Within-Industry ES	σ_G	10	Mongey (2019)
Elasticity of Labor Supply	ν	0.4	Jaimovich and Rebelo (2009)
Capital Elasticity	α	1/3	Standard value
Depreciation Rate	δ	$1 - 0.9^{1/4}$	Standard value
Discount Factor	β	$0.96^{1/4}$	Standard value
Coefficient of Risk Aversion	γ	1	log utility
Number of Industries	I	5,000	See text
Max Number of Firms (/industry)	N	100	See text
Calibrated Parameters			
Persistence of z_t	ρ_z	0.90	Autocorrelation of log TFP
Standard Deviation of ε_t	σ_ε	0.004	Standard deviation of log TFP
Fraction of Industries with $c_i = 0$	f_{85}	0.810	Emp Share Concentrated Industries
Fraction of Industries with $c_i = 0$	f_{07}	0.785	Emp Share Concentrated Industries
Pareto Tail 1985	λ_{85}	7.35	Markup Dispersion 1985
Pareto Tail 2007	λ_{07}	5.43	Markup Dispersion 2007
Fixed Cost 1985	c_{85}	4.73×10^{-3}	Average Markup 1985
Fixed Cost 2007	c_{07}	10.1×10^{-3}	Average Markup 2007

Table 2.1: Parameter Values

There are three parameters that we need to calibrate – the fraction of *competitive* industries f_{comp}, the fixed cost for the *noncompetitive* sector (c) and the Pareto shape of the productivity distribution of the pool of potential entrants (λ). These three parameters are jointly calibrated to target three data moments observed in 2007 (i.e. before the 2008 crisis). To calibrate f_{comp}, we target the fraction of aggregate employment that is allocated to highly concentrated industries. In our model, *noncompetitive* industries will be highly concentrated and will not

[22]See Jaimovich (2007) for a similar result.

have more than 4 firms. We hence define an industry as concentrated if the 4 largest firms represent at least 90% of the output of the 8 largest firms.[23] Using data from the US Census, we find that 7.62% of aggregate employment is allocated to such 6-digit industries.

We calibrate the other two parameters by targeting two moments of the markup distribution of public firms in 2007: the average sales-weighted markup (as computed by De Loecker et al. (2020)) and its standard deviation.[24] Intuitively, the average level of markups pins down the fixed cost (c) – a lower fixed cost will be associated with larger entry and hence lower average markups, for a given level of productivity dispersion. Dispersion in markups will, on the other hand, pin down the Pareto tail of entrants' productivity (λ) – given the positive link between productivity and markups, larger dispersion in productivities will be associated with larger dispersion in markups (and vice-versa) for a given number of firms. We obtain a fraction of *competitive* industries of $f_{\text{comp}} = 0.785$, a Pareto tail of $\lambda = 5.43$ and a fixed cost of $c = 0.0101$. Finally, we need to calibrate the two parameters governing the dynamics of aggregate productivity: the autocorrelation parameter ϕ_z and the standard deviation of the innovations σ_ε. We do so by targeting the first order autocorrelation and the standard deviation of aggregate TFP (between 1985 and 2018).[25]

To assess the business cycle implications of larger firm level heterogeneity, we also provide an alternative calibration of the model. In particular, we calibrate the Pareto tail λ and the fixed cost c to target the same two moments of the markup distribution in 1985. Note that both the observed sales-weighted average markup and its standard deviation are lower in 1985 than in 2007 (Table 2.2). We also recalibrate the fraction of competitive industries f_{comp} to target the same employment share in highly concentrated industries.[26] All other parameters are kept the same. In this alternative calibration, we obtain a Pareto tail of $\lambda = 7.35$, a fixed cost of $c = 0.0047$ and a share of competitive industries of $f_{\text{comp}} = 0.81$.

Table 2.2 reports our targeted moments, with their model counterparts. The model is very successful in matching the average markup in both the 1985 and the 2007 economies. We

[23] We would like to think of an industry at the highest possible level of disaggregation (e.g. 10-digit NAICS). However, the US census provides concentration metrics only at the 6-digit NAICS level. This is why we do not look directly at the share of the top 4 firms, but instead scale it by the share of the top 8. We have checked the robustness of our criterion. In particular, we considered alternative thresholds for the ratio of the top 4 to the top 8 (85% and 95%). The results were identical.

[24] See Appendix 2.C.1 for details.

[25] We use the series by Fernald (2012) and remove a linear trend, computed for the 1985-2007 period.

[26] The SUBS does not provide on employment by 6 digit industries prior to 1997. For this reason, we decide to keep the same target for the employment share of highly concentrated industries.

	1985		2007	
	Model	Data	Model	Data
Sales-weighted markup: average	1.32	1.27	1.45	1.46
Sales-weighted markup: standard deviation	1.21	1.44	1.69	1.74
Employment share in concentrated industries	9.39%	-	9.48%	7.62%
Aggregate TFP: autocorrelation	0.983	0.934*	0.936	0.934*
Aggregate TFP: standard deviation	0.027	0.025*	0.017	0.025*

*data moment computed over 1985-2018

Table 2.2: Targeted Moments and Model Counterparts

match its standard deviation reasonably well in 2007, but less so in 1985. The employment share of highly concentrated industries is slightly overestimated.

	1985	2007
Non-competitive industries: model statistics		
Number of firms per industry	1.66	1.43
Average markup (simple average)	1.82	2.24
Average markup (sales-weighted)	1.92	2.38
De Loecker et al. (2020): sales-weighted markup distribution		
90th percentile	1.66	2.25

Table 2.3: Markups (model and data moments)

Industries facing positive fixed costs will play an important role in our mechanism. Table 2.3 provides a brief characterization of these industries in the two calibrated economies. As we can see, industries with positive fixed costs will consist mostly of monopolies and duopolies – the average number of firms is 1.66 in the 1985 economy, and 1.43 in 2007. This implies a (sales-weighted) average markup of 1.92 in 1985 and of 2.38 in 2007. Note that these values are consistent with recent estimates for the US economy. For example, De Loecker et al. (2020) report that the 90th percentile for markups distribution (sales-weighted) increased from 1.66 in 1985 and of 2.25 in 2007. This means that, through the lens of our model, approximately 10% of US public charge markups consistent with the existence of a monopoly or a duopoly.

2.4.2 Quantitative Results

We start by comparing the steady-states of the 1985 and 2007 economies. We also include a parameterization for the years 1990, 1995 and 2000, assuming that average markups and markup dispersion follow a linear trend between 1985 and 2007. Figure 2.6 shows the steady-state values of output per hour, aggregate TFP, the labor share and aggregate markups for the five different parameterizations.[27] Our model predicts an overall increase in aggregate output per hour between 1985 and 2007 of roughly 30%. Aggregate TFP increases by 26%. Note that the increase in both output per hour and aggregate TFP are driven by the increase in the tail of the Pareto distribution – which results in a larger production possibility frontier. When looking at the data counterparts, we observe that real output per hour increases by 50%, while aggregate TFP increases by 26%. Therefore, through the lens of our model, the increase in the Pareto tail of the distribution of idiosyncratic draws can explain 60% of the increase in real output per hour. Our model replicates, however, the evolution of aggregate TFP, which is a non-targeted moment. Regarding the labor share, our model predicts a 3.9 percentage point decline in the aggregate labor share (from 0.564 to 0.525). In the data, it falls by only 2 percentage points (from 0.615 to 0.595). Note that the labor share in our model is about 5 to 6 percentage points lower than the one observed in the data. Therefore, our model underestimates the level of the labor share, but overestimates its decline. Such a discrepancy can be explained by the fact that we target average markups for public firms, which tend to display larger profit shares (and hence lower labor shares) than the average firm in the economy. We then compare the dynamic properties of the 1985 and the 2007 economies. We start by simulating each economy over 10,000,000 periods. Figure 2.7 shows the ergodic distribution of log output; the distributions are centered around the high steady-state, so that the horizontal axis represents output in percentage deviation from the high steady-state. The important thing to note is that, in the 2007 distribution, the two steady-states are closer to each other – a transition from the high to the low regime implies a 10%-15% reduction in output in 2007, as opposed to 35%-40% in 1985. While this fact means that transitions are less pronounced in 2007, it also implies that they are substantially more likely in 2007 than in 1985. Recall that the 1985 and the 2007 economies only differ in the Pareto tail parameter λ and the fixed production cost in the *noncompetitive* sector. In particular, the 2007 economy exhibits a more dispersed Pareto distribution and larger fixed costs. These facts mean that in 2007, small firms in the *noncompetitive* sector will have a lower share of the market and their entry/exit decisions will be more sensitive to aggregate fluctuations.

We next study the business cycle properties of our economies. Table 2.8 in Appendix 2.C.4 compares some business cycle moments with their data counterparts. To illustrate the dynamic

[27] In each of the five calibrations, the economy features two steady-states and is at the highest one.

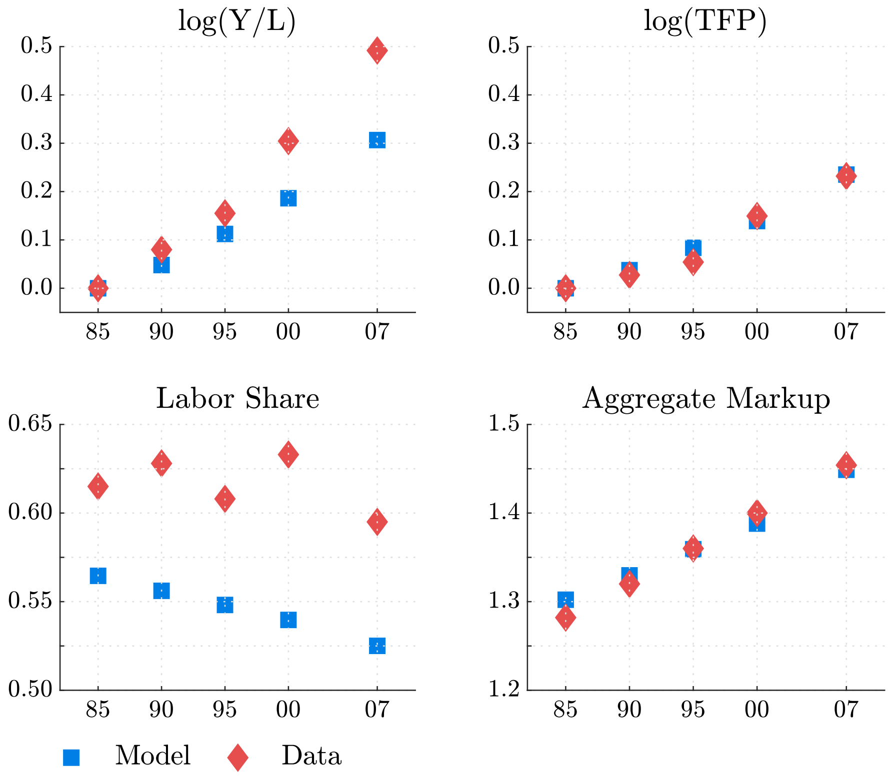
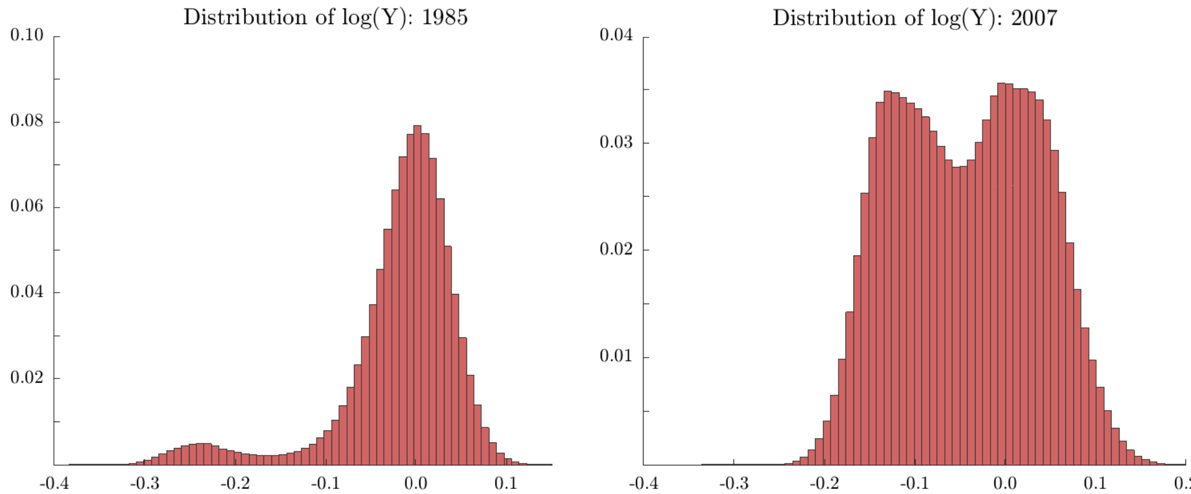


Figure 2.6: Model *versus* data: 1985-2007. The data series are respectively (i) Business Sector: Real Output Per Hour of All Persons (from BLS), (ii) Aggregate TFP from Fernald (2012), (iii) Business Sector: Labor Share (from BLS), and (iv) Aggregate Markup from De Loecker et al. (2020).

properties of our two calibrated economies we describe the response of several variables to aggregate TFP shocks.

Table 2.4 reports the probability of a deep recession for the two economies. We simulate each economy 100,000 times for 40 and 100 quarters and compute in how many simulations the economy experiences a recession where output drops by 10,15 and 20% of the high steady state level. When running the 1985 economy for 40 quarters, output drops by 10% in 2.7% of the simulations, the same number for the 2007 economy is 8.9%. This suggests that the latter is approximately 3 times more likely to experience a 10% fall in output over 10 years periods, similarly for different lengths and depth of the recession. This exercise suggests that indeed the economy is between three to eight times more likely to experience deep recessions in 2007 than the 1985 economy.

There is one aspect about our calibration strategy that should be discussed. In particular, the parameters determining the dynamics of TFP shocks (ρ_z and σ_ε) are the same in the two economies. We do not change these parameters because we want to focus on the role of the



(a) Output Distribution: 1985

(b) Output Distribution: 2007

Figure 2.7: Ergodic distribution of output
Note: the figure shows the ergodic distribution of output for the 1984 and the 2007 economies. These are given by simulating both models for 10,000,000 periods and plotting output log deviations from the high competition steady state value.

	1985 Model		2007 Model	
	T = 40	T = 100	T = 40	T = 100
$\Pr[\, y_t - y_0 < -0.10 \,]$	0.027	0.119	0.089	0.324
$\Pr[\, y_t - y_0 < -0.15 \,]$	0.002	0.019	0.015	0.141
$\Pr[\, y_t - y_0 < -0.20 \,]$	0.000	0.005	0.001	0.024

Table 2.4: Transition probabilities across regimes
Note: this table shows the transitions probabilities across regimes. Each economy starts in the high steady-state and is simulated for $T = 40$ and $T = 100$ quarters. Each simulation is then repeated 100,000 times. The probabilities $\Pr[y_t - y_0 < -\kappa]$ show the fraction of simulations in which output falls below the initial value y_0 by at least κ%.

parameters determining the competitive structure of the economy (λ, c and f).[28]

[28] There are however reasons to think that ρ_z and σ_ε changed over time. We will return to this point later.

Impulse Response Functions: Small Negative Shock We start by characterizing the reaction of the economy to a small negative shock. We consider a shock to the innovation of the exogenous TFP process that is equal to $\varepsilon_t = -\sigma_\varepsilon$ and lasts for two quarters. Such a shock will, however, have a persistent impact on exogenous TFP z_t through equation (2.5). Figure 2.8 shows the impulse responses for both the 1985 and the 2007 economy. The simulation of the transition dynamics covers 100 quarters. This shock generates different responses for the two economies, as evidenced by the middle top panel of 2.8. The 2007 economy exhibits both greater amplification and greater persistence. First, the 1985 economy experiences a 1.6% reduction in aggregate output after 5 quarters, against a 2.2% reduction in the 2007 economy. Second, the 1985 economy is back at steady state levels of output after approximately 90 quarters, while the 2007 economy has a much more prolonged downturn, being still 1% below steady-state after 100 quarters.

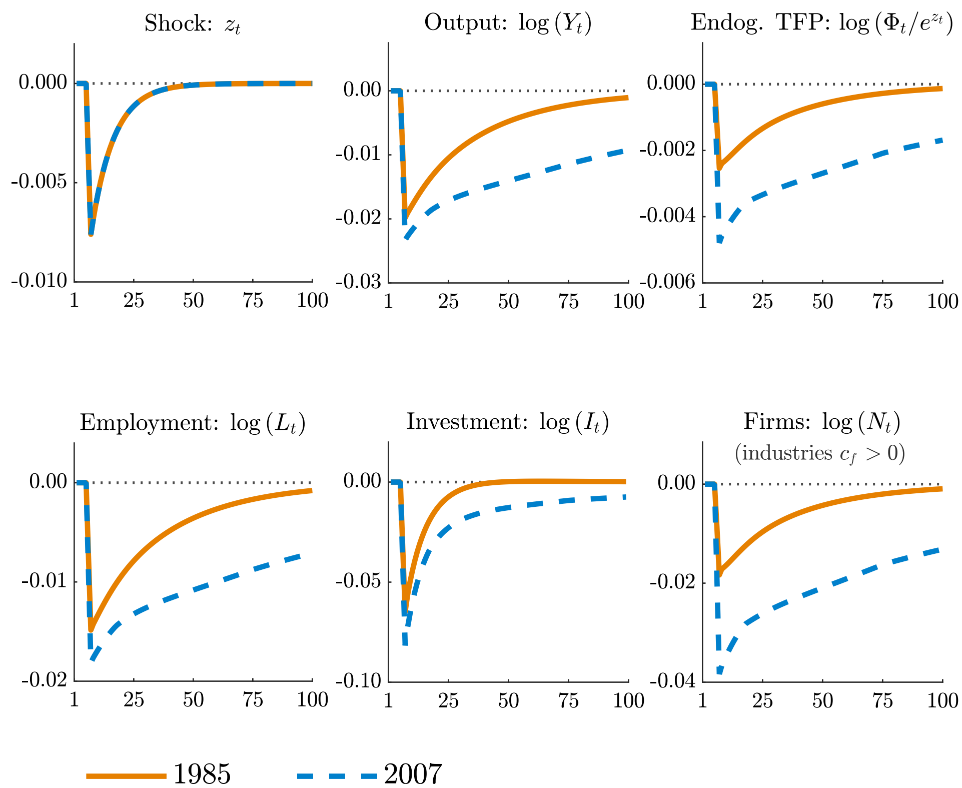


Figure 2.8: Impulse Responses: Small Shock
Note: the figure shows the impulse responses of key variables to a small aggregate TFP shock. The solid line and dashed lines represent the 1985 and 2007 economy, respectively.

The mechanism underlying such increased amplification and persistence can be better understood by looking at the right bottom panel, which plots the transition dynamics of the number of firms in the *noncompetitive* industries (which in our calibration represent 21.5% of all industries). In 2007, there is a much more significant reduction in the number of firms, due to the mechanisms outlined above: increased productivity dispersion and larger fixed costs make small, unproductive firms more sensitive to aggregate shocks. Such additional action in the extensive margins generates both additional amplification and persistence. Note that greater amplification and persistence can be observed also in employment, investment and in the endogenous component of aggregate TFP.[29]

Impulse Response Functions: Large Negative Shock The shock introduced above was small enough to make both economies transition to their initial steady-states. We now study the effect of a larger shock. To this end, we repeat the same exercise for the two economies, but now introduce a negative shock equal to $\varepsilon_t = -3\sigma_\varepsilon$, which lasts for three quarters.

The dynamics are shown in Figure 2.9. As before, there is greater amplification and persistence in the 2007 economy. However, the 2007 economy now experiences a permanent drop in aggregate output, i.e. it transitions to a lower steady-state (a low competition trap). In the example we consider, there is a permanent 11.1% loss in output. In this setup, employment drops permanently by 8.6%, while investment decreases by 72% on impact and 11.8% in the long run.

These results suggest rising firm differences and fixed costs are a source of additional amplification and persistence of shocks. This result seems however inconsistent with the idea of the *great moderation* – namely, the fact the the volatility of aggregate output declined between 1985 and 2007. Note, however, that aggregate volatility in out economy is the product of two forces – exogenous volatility (TFP shocks) and endogenous amplification and persistence. If exogenous volatility declined over time, it is possible that aggregate volatility also declined in spite of larger amplification. There are reasons to think that exogenous aggregate volatility may have decreased over time – for example, Carvalho and Gabaix (2013) argue that low-volatility sectors have gained share in production.

2.5 The 2008 Recession and Its Aftermath

In this section, we take a deeper look at the 2008 recession and its aftermath. The left panel of Figure 2.10 shows the behavior of some aggregate variables from 2006 to 2018 – real GDP, real gross private investment and total hours (all in per capita terms), as well as aggregate TFP.[30] All variables are in logs, detrended (with a linear trend computed over 1985-2007) and

[29]The decline of the endogenous component of aggregate TFP is discussed in Section 2.5.

[30]See Appendix 2.A.2 for the data sources.

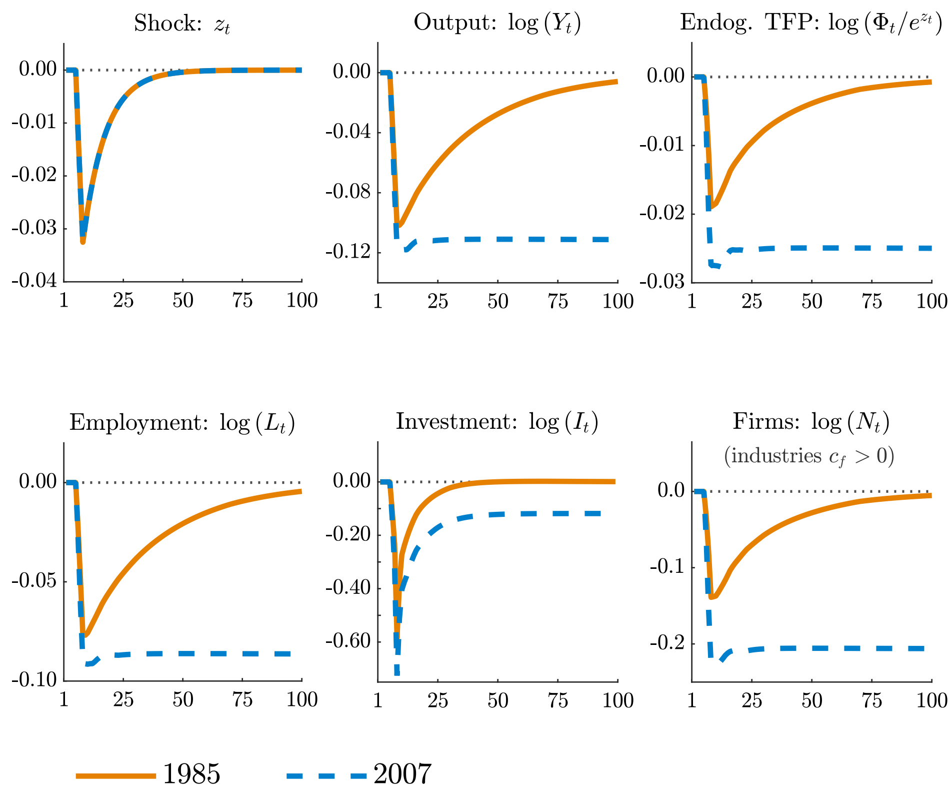


Figure 2.9: Impulse Responses: Large Shock
Note: the figure shows the impulse responses of key variables to a large aggregate TFP shock. The solid line and dashed lines represent the 1985 and 2007 economy, respectively.

centered around 2007Q4. The four variables decline on impact and do not seem to rebound to their pre-recession trends. For example, in the first quarter of 2018, real GDP per capita is 13.3% below trend (Table 2.5). Aggregate TFP has experienced a 6.8% negative deviation from trend. Investment declines by more than 40% on impact, and then seems to stabilize at approximately 20% below the pre-crisis trend.

	Data			Model		
	2009Q4	2015Q1	2018Q1	2009Q4	2015Q1	2018Q1
Output	-0.084	-0.119	-0.133	-0.109	-0.102	-0.103
Aggregate TFP	-0.037	-0.026	-0.068	-0.038	-0.016	-0.015
Hours	-0.124	-0.062	-0.037	-0.084	-0.080	-0.080
Investment	-0.352	-0.153	-0.220	-0.340	-0.133	-0.117

Table 2.5: The great recession and its aftermath

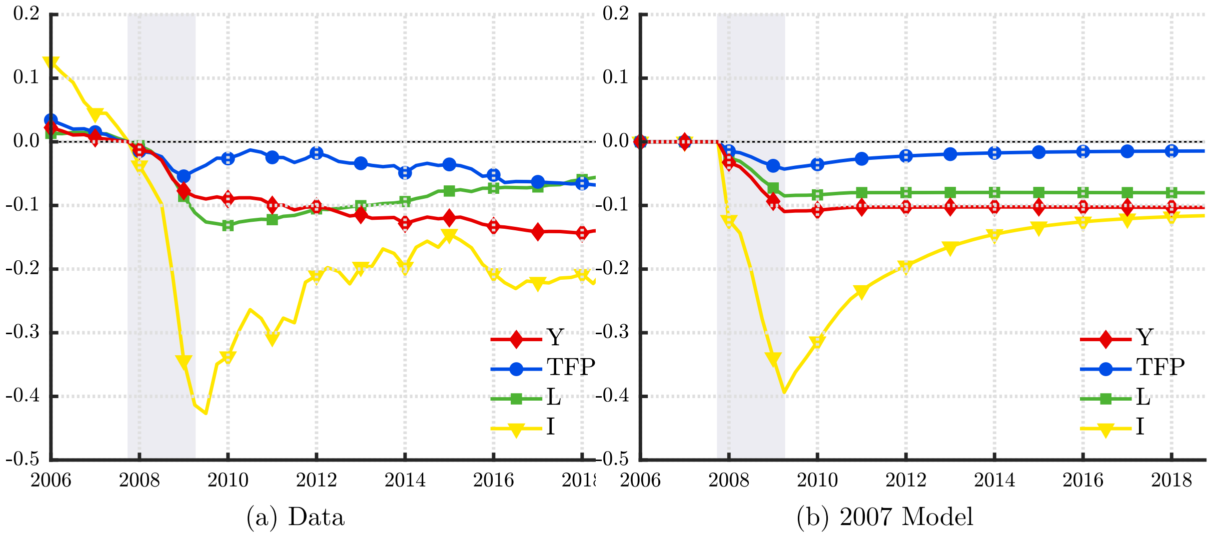


Figure 2.10: The great recession and its aftermath
Note: the figure shows the evolution of key macro aggregates in the aftermath of the 2007 recession in the data (Panel (a)) and the model (Panel (b)). The model economy is subjected to a sequence of six quarters shocks to aggregate TFP to match the change in aggregate TFP in the data.

We then ask whether our model can replicate the behavior of these four variables. We feed our model with a sequence of shocks to the innovation of TFP (ε_t) that lasts for six quarters (2008Q1:2009Q2), so that endogenous aggregate TFP in our model (Φ_t) matches the observed aggregate TFP series over the same period. The economy starts at the high steady-state (with $z_t = 0$). We set the innovations to productivity to zero after 2008Q1 and let the economy recover afterwards. The right panel of Figure 2.10 shows the implied responses of output, aggregate TFP, employment and investment, generated by our model. The series of shocks that we feed happen to be sufficient to trigger a transition to the low steady-state. Our model provides a reasonable description of the evolution of the four variables. Output experiences a 10.3% decline in the long-run, whereas employment drops by 8.0% (Table 2.5). Both reactions are of the same order of magnitude as observed in the data (with our model underpredicting the drop in output and overpredicting the drop in total hours). The same happens for investment, which declines by 34.0% on impact, and then stabilizes at 11.7% below its high steady-state value. The model also generates a 1.5% permanent drop in aggregate TFP – we can hence explain approximately 1/5 of the decline in aggregate TFP observed in the data.[31] Finally, we quantify the change in welfare. According to our model, in 2018 welfare was 4% lower (in consumption equivalent terms) relative to 2007 (2.22 in Appendix 2.C.7).

We next ask whether the sequence of aggregate TFP shocks z_t that we feed in the 2007

[31] In the next subsection, we explain the reasons for the decline in aggregate TFP.

economy can also trigger a transition to the low steady-state in the 1985 economy. Figure 2.11 shows the transition dynamics. Not only does the economy exhibit substantially less amplification, but it also reverts back to the high steady-state.

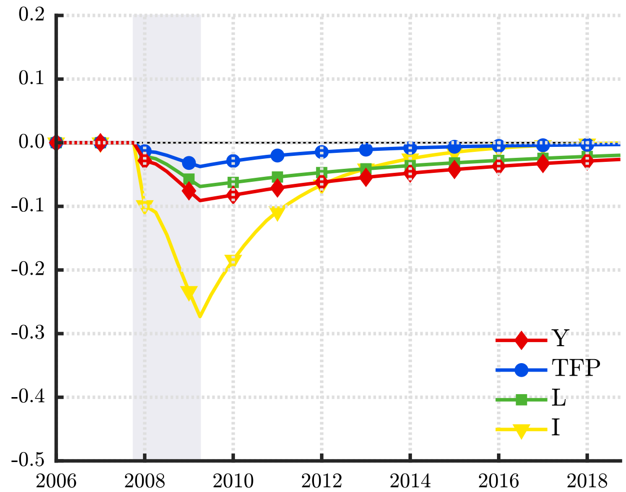


Figure 2.11: The *great recession* in the 1985 Model.
Note: This figure shows the response of the 1985 economy to the sequence of shocks used in Figure 2.10b

These results suggest that, in the 1985 economy, a downturn of the magnitude of the 2008-2009 recession would not be large enough to generate a persistent deviation from trend. The economy would have experienced a relatively fast reversal to trend, due to a lower endogenous amplification and persistence. We then conclude that the structural differences between the 1985 and the 2007 economies (namely larger productivity differences and larger fixed costs) are key to understand the 2008 crisis and the subsequent *great deviation*.

Aggregate Productivity As we have seen in Figure 2.10, the transition to the low competition regime is accompanied by a decline in aggregate TFP. This happens in spite of the exit of low productivity firms.[32] There are two reasons explaining the decline in aggregate TFP: (i) a reduction in the number of firms and (ii) an increase in cross-industry misallocation. To understand the first effect, note that our model embeds a *love for variety* effect. This can best be seen in the limit case in which there is no heterogeneity across firms or industries (all industries have n firms with identical productivity τ). In such a case, aggregate TFP is equal to

$$\Phi = I^{\frac{1-\rho}{\rho}} n^{\frac{1-\eta}{\eta}} \tau.$$

[32]Indeed, average of firm level TFP increases (see Figure 2.20 in Appendix 2.C.6).

Second, the transition to low competition trap generates an increase in cross industry misallocation. This fact happens because industries experiencing a larger contraction are the industries with positive fixed costs $c > 0$, i.e. industries whose output is already restricted. Figure 2.21 in Appendix 2.C.6 shows precisely an increase in the standard deviation of (log) industry outputs $\text{std}_i [\log (y_{it})]$.

In summary, our model provides two possible reasons why aggregate TFP may have experienced a permanent drop after 2008. Consistent with the model, such a drop in aggregate TFP may have occurred in spite of the exit of low productivity firms.[33]

Aggregate Markups and the Labor Share We now ask if our model can explain the evolution of aggregate markups and the labor share after 2008. Figure 2.12 shows the evolution of the labor share (left panel) and the De Loecker et al. (2020) aggregate markup series for publicly listed firms (right panel). The grey dashed line represents a linear trend computed for the 1985-period.

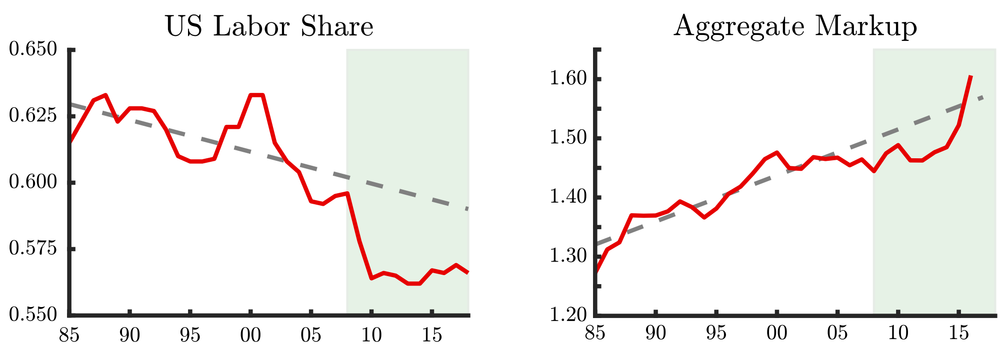


Figure 2.12: **US Labor Share and Aggregate Markup: 1985-2017**
Note: (i) Labor Share of the Corporate Business Sector (from the BLS) and (ii) De Loecker et al. (2020) aggregate markup series. For each series, the dashed grey line shows the corresponding average for the 1985-2007 period.

Table 2.6 compares the evolution of the labor share and the aggregate markup series between 2007 and 2016 observed in the data and obtained in our model. Overall, our model predicts a 0.6 pp decline in the aggregate labor share, which is approximately 20% of the observed decline between 2007 and 2016. If we account for a pre-crisis trend, we can explain approximately 38% of the deviation in 2016. Markups increase by 4.1 points in our model, which represents 29% of the observed increase (14.2 points) and 57% of the deviation from the pre-crisis trend (7.2 points).

[33]Foster et al. (2016) show that, as in previous recessions, manufacturing firms exiting during the great recession were on average less productive.

	Model	Data	
	$\Delta_{2007-2016}$	$\Delta_{2007-2016}$	$\Delta_{2007-2016} - \Delta_{\text{trend}}$
Labor Share	-0.006	-0.028	-0.016
Aggregate Markup	4.1	14.2	7.2

Table 2.6: Change in the labor share and in aggregate markups

2.5.1 The 1982 Recession

Through the lens of our model, the 2008 crisis made the US economy transition to a new steady-state. This fact has not been observed after any other postwar recession. This raises a natural question: what was special about the 2008 crisis? Was the shock hitting the economy in 2008 larger than in previous recessions? Or was the economy more fragile in 2008 and therefore more prone to experience a transition even for moderate shocks? To answer these questions, we repeat the experiment of Section 2.5 using the 1981-1982 crisis. We feed the 1985 economy with a sequence of shocks that replicate the dynamics of aggregate TFP during the 1981-1982 recession (1981Q3:1982Q4). We then take this same sequence of exogenous shocks and feed them in the 2007 economy. The results of this experiment are shown in Appendix 2.C.8. When looking at the response of the 1985 economy, we observe a temporary decline in all variables, but followed by a gradual recovery to the previous steady-state.[34] This contrasts with the response of 2007 economy, which again experiences a transition to the lower steady-state. These results suggest that, rather than the consequence to an usually large shock, the post-2008 deviation can be linked to an underlying market structure that made the economy significantly more fragile to aggregate fluctuations.

2.5.2 Robustness Checks

In the previous section we described how a reasonably calibrated version of the model admits multiple stochastic steady states and performs well when tasked with replicating the behaviour of the US economy in the aftermath of the 2008 recession. In this section we want to understand whether the existence of multiplicity, generated by the complementarity between capital accumulation and competition, is robust to alternative modelling assumptions and calibrations.

[34]The recovery seems to be slower than in the actual data. Note, however, that by design we are shutting down possible positive shocks hitting the economy in 1983 and afterwards.

Alternative Values for the Elasticity of Substitution The first robustness check we carry out is a different parametrization of the key elasticities in our model. In particular, we follow Edmond et al. (2015) and use a between industry elasticity of 1.25. Intuitively, this economy features higher market power in non competitive industries. Calibrating the Pareto parameter and the fixed cost to target the moments of the markup distribution we obtain thinner tails of the productivity distribution and higher fixed costs of production. The model retains multiplicity of steady states and, when hit with the 2007 productivity shocks, features a deeper recession than our baseline calibration. As in our benchmark model we find that the shock is such that the 1985 converges to the same steady state while the 2007 economy falls in a low competition trap.

Note that, fixing all other parameters, when we recalibrate the model using between industries elasticity above 2 multiplicity disappears.

Variable Fixed Costs We start by noting that in the model presented so far firms have to pay a fixed cost c_f every period and that such cost is paid in terms of the final good. This assumption implies that the cost of entry is independent of the state of the economy and, hence, of its competitive regime and of factor prices. If fixed costs were to change with factor prices, then entry would be cheaper (more expensive) in a low (high) competition regime, which could in principle eliminate steady-state multiplicity. To address this concern, we let firms hire labor and capital to pay the fixed cost.[35] In particular we assume production entails a fixed cost

$$c_f = k_c^{\alpha} l_c^{1-\alpha}$$

Under this assumption, firms incur an effective fixed cost $\Theta_t \cdot c_f$. Recall that the factor cost index Θ_t is increasing in the number of active firms. This implies that entry becomes ever more expensive as firms enter.

Under this assumption, the labor and capital market clearing conditions need to be updated. Appendix 2.I.3 shows the results for this version of the model. The existence of two competitive regimes is preserved under this alternative assumption. We also obtain responses for the crisis experiment performed above.

The results are qualitatively and quantitatively similar to the benchmark model. The only significant difference is that in the Cobb-Douglas fixed cost case investment drops more upon impact.

[35] We thank Greg Kaplan for suggesting a version of this robustness check.

2.6 Policy Experiment

In this section we use the benchmark model discussed above to evaluate how policy can alleviate the market power driven externality in this economy. The key intuition is that in this economy, on top of the standard inefficiency results, market power carries additional effects on the resilience to aggregate fluctuations. To study how a policy maker would reduce such effect we allow the planner to levy a tax on net profits and subsidise entry. We impose a balanced budget every period.

$$\tau_\pi \Pi_t^N = \tau_f c_f N_t.$$

where τ_π is the tax rate on net profits and τ_f is the fraction of fixed costs that is subsidized. First, note that by design the entry subsidy only affects non-competitive industries. Secondly, taxing net profits does not distort the entry choice while subsidising entry implies that the minimum productivity threshold to profitably enter decreases. By these features it is never efficient for the planner to tax entry and subsidise profits. The planner faces one key trade-off when subsidising entry: while having more firms in the economy depresses markups and directly increases welfare through love for variety in the preferences, it also implies that the average firm productivity declines. This is not necessarily true for aggregate TFP as this is increasing in N by love for variety, as discussed in Section 2.5.

We provide welfare calculations for different levels of entry subsidy in Figure 2.25 (Appendix 2.D). We simulate each economy 100,000 times for each level of the production subsidy, and calculate average welfare. The analysis suggests that the government would find a subsidy of around 80% of the fixed cost optimal. At this level however it would have to impose a tax on profit larger than 100%, implying that such level of subsidy cannot be budget balanced. The highest welfare, given the balanced budget constraint, can be achieved with a subsidy around half of the fixed cost. Note however that as this is a corner solution it implies pledging all net profits of active firms to finance entry. Interestingly, welfare becomes very steep around 1, where the government is inactive. The underlying explanation is that when the subsidy is small or negative the economy spends a large fraction of time in the low steady state, significantly decreasing average welfare. In practice, the subsidy is very effective whenever it can reduce the fraction of periods spent in the low competition regime or eliminate the low steady state.

Furthermore, note that welfare decreases for very large levels of the subsidy as the love for variety effect is dominated by new entrants becoming very unproductive.

Lastly, one can think about the optimal state dependent subsidy. If the economy is hit by a large negative shock that might trigger a steady state transition, the welfare benefit of such a subsidy is very large as it prevents the spiralling of the economy due to firm exit. On the other hand, during a recession, profits are reduced, thereby making the budget balanced constraint tighter. If the government could borrow intertemporally it would have large incentives to do so and finance entry during downturns and pay back debt during booms. This would suggest that, through the lens of our model, countercyclical subsides to firm entry (or covering the fixed cost of production) significantly alleviate recessions by preventing the economy from falling in low competition traps.

2.7 Empirical Evidence

The results presented in Section 2.3.3 offer cross-industry predictions that can be tested in the data. In particular, according to our theory, industries featuring larger concentration in 2007 should have experienced a larger contraction in 2008. This prediction follows from equation (2.7). Recall that this equation establishes a positive link between productivity, market shares and markups (for a given number of active firms). Therefore, if we take two industries with the same number of firms, the one featuring a more uneven distribution of productivities will have larger dispersion in market shares and hence larger concentration. In these industries, firms at the bottom of the distribution will be smaller and charge lower markups, and will hence be more likely to exit upon a negative shock. Note that this prediction holds for a given number of firms $n_{i,t}$; when measuring the correlation between concentration in 2007 and the size of the contraction in 2008, we must therefore control for the number of firms in the industry.

We build a dataset combining the 2002 and 2007 US Census data on industry concentration to the Statistics of US Businesses (SUSB) and the Bureau of Labor Statistics (BLS) to obtain outcomes as employment, total wage bill and the number of firms at the industry level (6-digits NAICS). The final dataset includes 791 6-digit industries. In 2016, the median industry had 1,316 firms, 36,910 workers and a total payroll of $1,880 million.

To assess whether industries with a larger concentration before the crisis experienced a larger post-crisis decline, we estimate the following model

$$\frac{\Delta y_{i,07-16}}{y_{i,07}} = \beta_0 + \beta_1 \text{ concent}_{i,07} + \beta_2 \ \log \text{firms}_{i,07} + a_s \mathbb{1}\{i \in s\} + u_i.$$

y_i is an outcome for industry i (for example total employment, total wage bill or total number of firms) and concent_i is the share of the 4 largest firms (scaled by the share of the largest 50); we also control for the number of firms before the crisis ($\text{firms}_{i,07}$). The outcomes always take the form of the annualized growth rate between 2007 and 2016 in a specific industry. In all regressions, we will also include macro sector fixed effects as a control (a_s). The unit of observation is a 6-digit industry.

We start by studying the correlation between the change in employment between 2008 and

2016 and concentration in 2007. The results, showed in Table 2.9 (Appendix 2.E), suggest that more concentrated industries experienced lower employment growth in the aftermath of the great recession. Quantitatively, the estimation suggests that a 1pp higher pre-crisis concentration correlates with a 2pp lower employment growth rate between 2007 and 2016. This pattern holds irrespective of the inclusion of the number of firms in 2007. To address the concern that industries with larger concentration in 2007 could exhibit lower growth already before the crisis, we include cumulative employment growth between 2003 and 2007 as a control (column 3); the results do not change. Finally, the results are also robust to the inclusion of sector fixed effects (column 4). While these results concern the evolution of employment growth, a similar pattern is found if we use total wage bill instead (Table 2.10). We also study the correlation between the measure of concentration and net entry after the crisis (Table 2.11). Our finding suggests that a 1 percentage point increase in the concentration measure is associated with a 2 to 3 percentage points decrease in the post crisis net entry. These results suggest that industries with larger concentration in 2007 experienced a larger contraction in activity after the crisis. They do not tell us however whether these industries experienced a larger increase in profit margins or a larger decline in labor shares. We conclude this section by providing evidence on the evolution of the labor share across industries. While the US census of firms provides data on total employment and total number of firms for all 6-digit industries, it does not contain data on the labor share. We rely on data from the BLS 'Labor Productivity and Cost' programme (see Appendix 2.A.2 for details). This database, however, only provides data on the labor share for a restricted group of industries. The results are shown in Table 2.12. Overall, there seems to be a negative relationship between the post-crisis change in the labor share and the pre-crisis level of concentration. Industries with larger concentration in 2007 experienced a larger drop in labor share between 2008 and 2016.

All in all, these results suggest that the structure of US product markets in 2007 is important to explain the consequences of the 2008 crisis. The results presented are, strictly speaking, cross-sectional – industries with larger concentration in 2007, displayed a larger post-crisis contraction. We think, however, that they can also be used to support one of the main insights of the model – namely, that rising concentration can have made the US economy more vulnerable to aggregate shocks.

2.8 Conclusion

The US economy appears to have experienced a fundamental change over the past decades, with several studies and data sources indicating a reallocation of activity towards large, high markup firms. This observation has raised concerns in academic and policy circles about increasing market power, and it has been proposed as an explanation for recent macroeconomic *puzzles* –

such as low aggregate investment, low wage growth or declining labor shares. Besides their impact on factor shares and factor prices, our model suggests that rising firm differences and greater market power can also have an impact on business cycles and provide an amplification and persistence mechanism to aggregate fluctuations. In particular, larger firm heterogeneity and greater market power may have rendered the US economy more vulnerable to aggregate shocks and more likely to experience quasi-permanent recessions. Through the lens of our theory, such increased fragility may have been difficult to identify, as it manifests itself only in reaction to large shocks.

In broader terms, our theory indicates that the firm size/markup distribution can be an important determinant of the response of the economy to aggregate shocks. This observation suggests that product market considerations should gain relevance within macroeconomic research and policy analysis. In particular, the standard toolkits used by macroeconomists should increasingly incorporate a realistic characterization of product market frictions.

We conclude by mentioning two extensions we are considering in our future work. First, we are planning to introduce endogenous growth to research the dynamic interplay between market power and innovation in a context of multiple competitive regimes. As documented in Figure 2.2, both real GDP per capita and aggregate TFP have experienced a widening deviation from trend after 2008, which indicates that growth rates have become persistently lower. We think that an extended version of our model with endogenous R&D has the potential to account for this. In a world where firms conduct R&D because of an *escape-from-competition* effect, a decrease in product market competition will likely reduce firms' incentives to innovate.

Second, we also plan to consider a setup with nominal rigidities to think about the monetary policy implications of increasing firm differences and of rising market power. Our theory suggests at least two relevant insights for the design of monetary policy. First, as industries become more concentrated, firms' pricing decisions are likely to become increasingly rigid and less sensitive to aggregate fluctuations. This suggests that the degree of price rigidity may endogenously respond to changes in the product market structure, which has obvious implications for the effects of monetary policy. Second, market power can have a negative impact on interest rates and hence be associated with the greater likelihood of a binding zero lower bound. The examination of these two hypotheses is an important avenue for future research.

Bibliography

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li, "A Theory of Falling Growth and Rising Rents," Working Paper Series, Federal Reserve Bank of San Francisco March 2019.

Akcigit, Ufuk and Sina T. Ates, "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory," NBER Working Paper 25755 2019.

_ **and** _ , "What Happened to U.S. Business Dynamism?," NBER Working Paper 25756 2019.

Andrews, Dan, Chiara Criscuolo, and Peter N. Gal, "Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries," OECD Productivity Working Papers 2015.

Anzoategui, Diego, Diego Comin, Mark Gertler, and Joseba Martinez, "Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence," *American Economic Journal: Macroeconomics*, July 2019, *11* (3), 67–110.

Atkeson, Andrew and Ariel Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, December 2008, *98* (5), 1998–2031.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," NBER Working Paper 23396 2017.

Benigno, Gianluca and Luca Fornaro, "Stagnation traps," *The Review of Economic Studies*, 2017, *85* (3), 1425–1470.

Bilbiie, Florin O., Fabio Ghironi, and Marc J. Melitz, "Endogenous Entry, Product Variety, and Business Cycles," *Journal of Political Economy*, 2012, *120* (2), 304–345.

Bils, Mark, Peter J. Klenow, and Benjamin A. Malin, "Resurrecting the Role of the Product Market Wedge in Recessions," *American Economic Review*, April 2018, *108* (4-5), 1118–1146.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch, "Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data," NBER Working Paper 27002 2020.

Burstein, Ariel, Basile Grassi, and Vasco Carvalho, "Bottom-Up Markup Fluctuations," Working Paper 2019.

Calligaris, Sara, Chiara Criscuolo, and Luca Marcolin, "Mark-ups in the digital era," OECD Science, Technology and Industry Working Papers 2018.

Carvalho, Vasco and Xavier Gabaix, "The Great Diversification and Its Undoing," *American Economic Review*, August 2013, *103* (5), 1697–1727.

Chatterjee, Satyajit, Russell Cooper, and B Ravikumar, "Strategic complementarity in business formation: Aggregate fluctuations and sunspot equilibria," *The Review of Economic Studies*, 1993, *60* (4), 795–811.

Clementi, Gian Luca, Berardino Palazzo, and Peifan Wu, "Firm Demographics and the Great Recession," Technical Report, Society for Economic Dynamics 2017.

Cooper, Russell and Andrew John, "Coordinating Coordination Failures in Keynesian Models," *The Quarterly Journal of Economics*, 1988, *103* (3), 441–463.

_ and _ , "Imperfect competition and macroeconomics : Theory and quantitative implications," *Cahiers d'Économie Politique*, 2000, *37* (1), 289–328.

Decker, Ryan A, John C Haltiwanger, Ron S Jarmin, and Javier Miranda, "Changing business dynamism and productivity: Shocks vs. responsiveness," Technical Report, National Bureau of Economic Research 2018.

Díez, Federico, Jiayue Fan, and Carolina Villegas-Sanchez, "Global Declining Competition," CEPR Discussion Papers, CEPR Discussion Papers April 2019.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, Markups, and the Gains from International Trade," *American Economic Review*, October 2015, *105* (10), 3183–3221.

_ , _ , and _ , "How costly are markups?," NBER Working Paper 24800 2018.

Eggertsson, Gauti B, Jacob A Robbins, and Ella Getz Wold, "Kaldor and Piketty's facts: The rise of monopoly power in the united states," NBER Working Paper 24287 2018.

Fernald, John G., "A quarterly, utilization-adjusted series on total factor productivity," Working Paper Series 2012-19, Federal Reserve Bank of San Francisco 2012.

Foster, Lucia, Cheryl Grim, and John Haltiwanger, "Reallocation in the Great Recession: Cleansing or Not?," *Journal of Labor Economics*, 2016, *34* (S1), 293–331.

Galí, Jordi and Fabrizio Zilibotti, "Endogenous Growth and Poverty Traps in a Cournotian Model," *Annals of Economics and Statistics*, 1995, *37-38*, 197–213.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W Huffman, "Investment, Capacity Utilization, and the Real Business Cycle," *American Economic Review*, June 1988, *78* (3), 402–417.

Hall, Robert E, "Using empirical marginal cost to measure market power in the us economy," Technical Report, National Bureau of Economic Research 2018.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein, "Quantifying the Sources of Firm Heterogeneity," *The Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Jaimovich, Nir, "Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations," *Journal of Economic Theory*, November 2007, *137* (1), 300–325.

_ **and Max Floetotto**, "Firm dynamics, markup variations, and the business cycle," *Journal of Monetary Economics*, 2008, *55* (7), 1238–1252.

_ **and Sergio Rebelo**, "Can news about the future drive the business cycle?," *American Economic Review*, 2009, *99* (4), 1097–1118.

Karabarbounis, Loukas and Brent Neiman, "Accounting for factorless income," *NBER Macroeconomics Annual*, 2019, *33* (1), 167–228.

Kehrig, Matthias, "The cyclical nature of the productivity distribution," *Earlier version: US Census Bureau Center for Economic Studies Paper No. CES-WP-11-15*, 2015.

_ **and Nicolas Vincent**, "The micro-level anatomy of the labor share decline," NBER Working Paper 25275 2018.

Loecker, Jan De and Jan Eeckhout, "Global market power," NBER Working Paper 24768 2018.

_ **,** _ **, and Gabriel Unger**, "The rise of market power and the macroeconomic implications," *The Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

Mongey, Simon, "Market Structure and Monetary Non-neutrality," *Working Paper*, 2019.

Pagano, Marco, "Imperfect Competition, Underemployment Equilibria and Fiscal Policy," *Economic Journal*, June 1990, *100* (401), 440–463.

Rotemberg, Julio and Garth Saloner, "A supergame-theoretic model of price wars during booms," *New Keynesian Economics*, 1986, *2*, 387–415.

Schaal, Edouard and Mathieu Taschereau-Dumouchel, "Coordinating business cycles," *Working Paper*, 2015.

Traina, James, "Is aggregate market power increasing? production trends using financial statements," *Production Trends Using Financial Statements (February 8, 2018)*, 2018.

Van Reenen, John, "Increasing Differences Between Firms: Market Power and the Macro-Economy," *Working Paper*, September 2018.

Appendix A

2.A Data Appendix

2.A.1 The Entry Rate and the Number of Firms

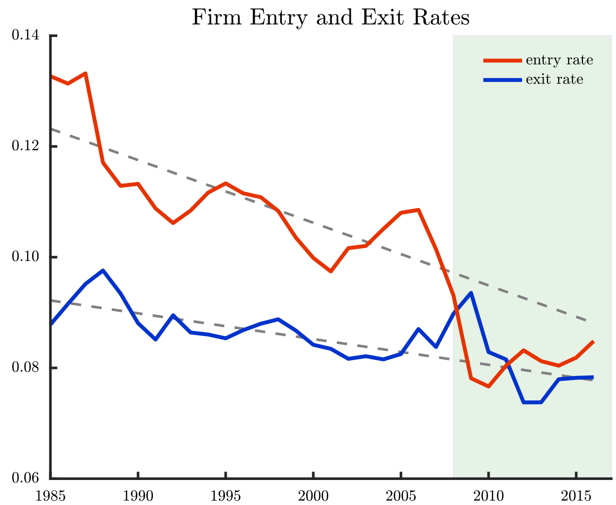


Figure 2.13: **US Firm Entry and Exit Rates: 1980-2017**
The entry (exit) rate is ratio of the number of startups (exiting firms) to the number of active firms in the previous year (data is from the US Business Dynamic Statistics). The dashed grey line shows a linear trend computed for the 1985-2007 period.

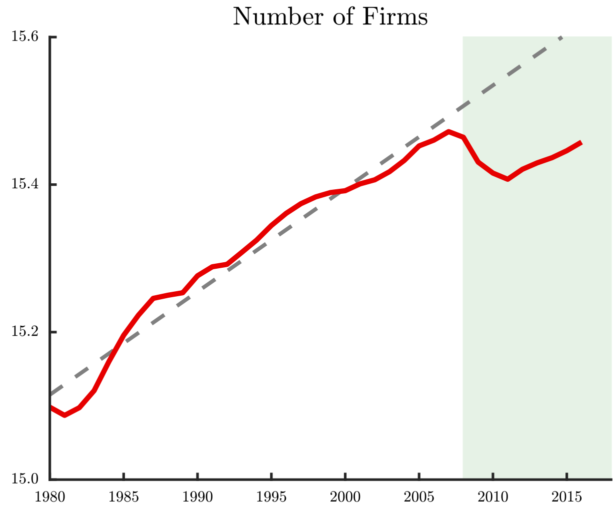


Figure 2.14: **Number of Firms per Sector: 1980-2016**
The red line shows the number of firms with at least one employee (in logs). The dashed grey line shows a linear trend computed over the 1980-2007 period. Data is from the US Business Dynamics Statistics.

2.A.2 Data Definition

Table 2.7 provides information on all the data sources used in Section 2.5.

Table 2.7: Data sources

Variable	Source
Real GDP	BEA – NIPA Table 1.1.3 (line 1)
Real Personal Consumption Expenditures	BEA – NIPA Table 1.1.3 (line 2)
Real Gross Private Domestic Investment	BEA – NIPA Table 1.1.3 (line 7)
Total Hours	BLS – Nonfarm Business sector: Hours of all persons
Aggregate TFP	Fernald (2012): Raw Business Sector TFP
Population	BEA – NIPA Table 2.1 (line 40)

Aggregate Profit Share

The aggregate profit share is computed as

$$\text{profit share}_t = \frac{\text{VA}_t - \text{W}_t\, L_t - \text{T}_t - R_t \cdot \text{K}_t - \text{DEP}_t}{\text{VA}_t}$$

VA_t is the total value added of the US business sector (NIPA Table 1.3.5, line 2), $\text{W}_t\, L_t$ is total labor compensation (NIPA Table 1.13, line 4 + line 11) and T_t is the value of taxes on production minus subsidies (NIPA Table 1.13, line 9 + line 17).

K_t is the value of nonresidential private fixed assets (including intangibles) of the US business sector (NIPA Table 4.1, line 8 - line 69 - line 73) and DEP_t is depreciation (NIPA Table 4.4, line 8 - line 69 - line 73). Finally, R_t is the required rate of return. We follow Eggertsson et al. (2018) and compute it as the difference between Moody's Seasoned BAA Corporate Bond Yield and a 5-year moving average of past CPI inflation (from BLS, used as a proxy for expected inflation).

Industry-level Labor Share

We obtain data on the labor share at the industry level from the BLS 'Labor Productivity and Costs' (LPC) database. We calculate the labor share as the ratio of 'Labor compensation' to 'Value of Production'. Note that this ratio gives the share of labor compensation in total revenues, and not in value added.[36]

2.B Derivation and Proofs: General Equilibrium

2.B.1 Aggregate TFP

Aggregate TFP is given by

$$\Phi\left(\mathbb{Z}_{\tau t},\{n_{it}\}_{i=1}^{I}\right)=\left[\sum_{i=1}^{I}\left(\sum_{j=1}^{n_{it}}\omega_{ijt}^{\eta}\right)^{\frac{\rho}{\eta}}\right]^{\frac{1}{\rho}}\left(\sum_{i=1}^{I}\sum_{j=1}^{n_{it}}\frac{\omega_{ijt}}{\tau_{ijt}}\right)^{-1}, \tag{2.16}$$

where

$$\omega_{ijt}:=\left[\sum_{k=1}^{n_{it}}\left(\frac{\mu_{ikt}}{\tau_{ikt}}\right)^{\frac{\eta}{1-\eta}}\right]^{\frac{\eta-\rho}{\eta}\frac{1}{1-\rho}}\left(\frac{\tau_{ijt}}{\mu_{ijt}}\right)^{\frac{1}{1-\eta}}.$$

2.B.2 Proof of Proposition Proposition 2.3

The aggregate factor share can be written as the ratio of the aggregate factor cost index and aggregate TFP.

$$\Omega\left(\mathcal{F},n\right)=\frac{\Theta\left(\mathcal{F},n\right)}{\Phi\left(\mathcal{F},n\right)}$$

[36] This ratio coincides with the 'Labor cost share' provided by the BLS. This variable is, however, available just for a restricted number of industries.

Assume that all industries are identical (same distribution of idiosyncratic productivities $\mathcal{F}$ and same number of active firms n) and that $\eta = 1$. As shown in online appendix 2.H.1 the aggregate factor cost index $\Theta(\mathcal{F}, n)$ is increasing in n, and aggregate TFP $\Phi(\mathcal{F}, n)$ is decreasing in n. This implies that $\Omega(\mathcal{F}, n)$ is increasing in the number of firms per industry n.

The second part of the proposition is proved in online appendix 2.H.1.

2.B.3 Number Active Firms

When there are n active firms in a given industry, the production profits of a firm with productivity π_j are equal to

$$\Pi(\pi_j, n, \mathcal{F}, \Theta, Y) = \Lambda(\pi_j, n, \mathcal{F}) \Theta^{-\frac{\rho}{1-\rho}} Y$$

where $\Lambda(\pi_j, n, \mathcal{F})$ is defined in Appendix 2.G.1. A symmetric equilibrium with n firms per industry is therefore possible provided that

$$\begin{aligned} \Lambda(\mathcal{F}, n, n) \Theta^{-\frac{\rho}{1-\rho}} Y &\geq c \\ \Lambda(\mathcal{F}, n+1, n+1) \Theta^{-\frac{\rho}{1-\rho}} Y &\leq c \end{aligned}$$

Using equation (2.15), we can write the above inequalities as

$$\underline{K}(\mathcal{F}, n) \leq K_t \leq \overline{K}(\mathcal{F}, n),$$

where

$$\begin{aligned} \underline{K}(\mathcal{F}, n) &= \left\{ \frac{c}{\Lambda(\mathcal{F}, n, n)} (1-\alpha)^{-\frac{1-\alpha}{\nu+\alpha}} [\Phi(\mathcal{F}, n)]^{-1} [\Theta(\mathcal{F}, n)]^{\frac{\rho}{1-\rho} - \frac{1-\alpha}{\nu+\alpha}} \right\}^{\frac{\nu+\alpha}{\alpha(1+\nu)}} \\ \overline{K}(\mathcal{F}, n) &= \left\{ \frac{c}{\Lambda(\mathcal{F}, n+1, n+1)} (1-\alpha)^{-\frac{1-\alpha}{\nu+\alpha}} [\Phi(\mathcal{F}, n)]^{-1} [\Theta(\mathcal{F}, n)]^{\frac{\rho}{1-\rho} - \frac{1-\alpha}{\nu+\alpha}} \right\}^{\frac{\nu+\alpha}{\alpha(1+\nu)}}. \end{aligned}$$

2.B.4 Multiple Equilibria

In the example of Figure 2.3 the equilibrium is always unique, i.e. the aggregate capital stock K_t pins down the number of firms per industry n_t and all other equilibrium variables. This happens because the bounds $\overline{K}(\cdot)$ and $\underline{K}(\cdot)$ satisfy

$$\overline{K}(\mathcal{F}, n) < \underline{K}(\mathcal{F}, n+1) \quad , \qquad \forall n$$

However, if the above condition is not satisfied, there can be multiple equilibria. Suppose that the bounds $\overline{K}(\cdot)$ and $\underline{K}(\cdot)$ satisfy

$$\underline{K}(\mathcal{F}, n+1) < \overline{K}(\mathcal{F}, n) \quad , \qquad \text{for some } n$$

When $K_t \in \left[\underline{K}(\mathcal{F}, n+1), \overline{K}(\mathcal{F}, n)\right]$, the economy features multiple equilibria: it can feature a symmetric equilibrium with n firms per industry, a symmetric equilibrium with $n+1$ per industry, and also an asymmetric equilibrium with n firms in some industries and $n+1$ in some others. Figure 2.15 shows aggregate output Y_t as a function of the aggregate capital stock K_t for an economy in which (static) multiplicity can occur.

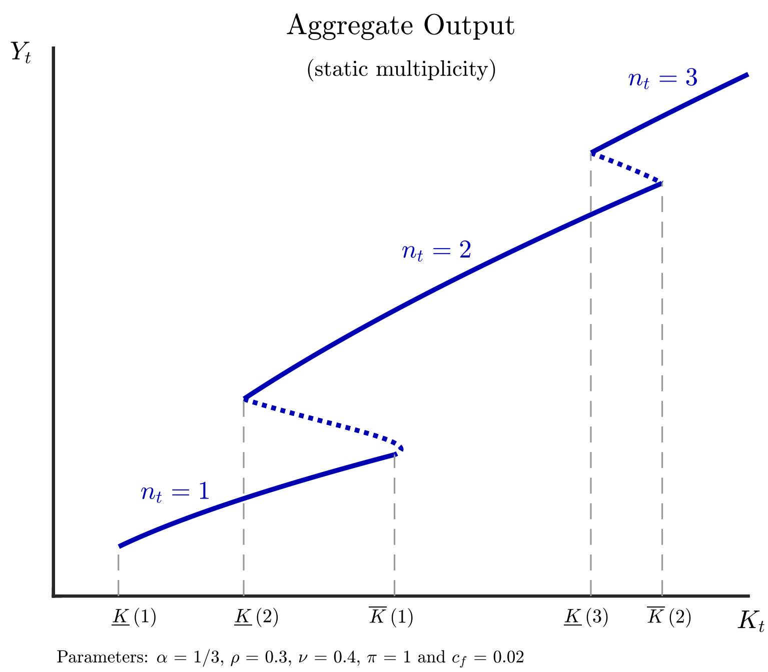


Figure 2.15: Static Multiplicity
Note: the figure represents the relation between capital and output. Solid lines depict symmetric equilibria, while dotted lines are asymmetric ones.

The full lines represent symmetric equilibria in which all industries are identical, whereas the dashed lines represent asymmetric equilibria. Note that (static) multiplicity arises because of a positive complementarity between competition and labor supply. For a given capital stock level K_t there can be an equilibrium featuring a large number of active firms, and hence high factor shares, high wages and high labor supply; and another possible equilibrium with a lower number of firms, and hence low factor shares, low wages and a depressed labor supply. The following proposition provides the conditions for (static) multiplicity.

Proposition 2.6
(Static Multiplicity) Suppose that an equilibrium with n firms per industry is possible at time t. An equilibrium with $n+1$ firms is also possible provided that

$$\frac{\Phi(\mathcal{F},n)}{\Phi(\mathcal{F},n+1)} < \left[\frac{\Theta(\mathcal{F},n)}{\Theta(\mathcal{F},n+1)}\right]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}}$$

Proof. Suppose that

$$\underline{K}(\mathcal{F},n) \leq K_t \leq \overline{K}(\mathcal{F},n)$$

so that a symmetric equilibrium with n firms in every industry is possible. A symmetric equilibrium with $n+1$ firms will also be possible provided that

$$\begin{aligned}
& \underline{K}(\mathcal{F},n+1) < \overline{K}(\mathcal{F},n) \\
\Leftrightarrow \quad & \frac{c_i}{\Lambda(n+1,\pi_{n+1})}\frac{[\Theta(\mathcal{F},n+1)]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}}}{\Phi(\mathcal{F},n+1)} < \frac{c_i}{\Lambda(n+1,\pi_{n+1})}\frac{[\Theta(\mathcal{F},n)]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}}}{\Phi(\mathcal{F},n)} \\
\Leftrightarrow \quad & \frac{\Phi(\mathcal{F},n)}{\Phi(\mathcal{F},n+1)} < \left[\frac{\Theta(\mathcal{F},n)}{\Theta(\mathcal{F},n+1)}\right]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}}
\end{aligned}$$

■

Corollary 1. *(Static Multiplicity with No Productivity Differences) When all firms are equally productive there can be equilibrium multiplicity if and only if*

$$\frac{\rho}{1-\rho} < \frac{1-\alpha}{\nu+\alpha}$$

Proof. when there are no productivity differences, the condition becomes

$$\begin{aligned}
& \left[\frac{\Theta(\mathcal{F},n)}{\Theta(\mathcal{F},n+1)}\right]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}} > 1 \\
\Leftrightarrow \quad & \frac{\rho}{1-\rho} - \frac{1-\alpha}{\nu+\alpha} < 0 \\
\Leftrightarrow \quad & \frac{\rho}{1-\rho} < \frac{1-\alpha}{\nu+\alpha} \\
\Leftrightarrow \quad & \rho < \frac{1-\alpha}{\nu+\alpha}(1-\rho) \\
\Leftrightarrow \quad & \frac{\rho}{1-\rho} < \frac{1-\alpha}{\nu+\alpha}
\end{aligned}$$

■

As the above corollary makes it clear, in the limit case in which firms are equally productive, static multiplicity depends on three main parameters: ρ, ν and α. In particular, when there are no productivity differences, static multiplicity arises whenever (i) ρ is low (so that differentiation across varieties is large and markups/factor shares display a high responsiveness to changes in the number of firms), (ii) the wage elasticity of labor supply $\frac{1}{\nu}$ is large or (iii) when the labor elasticity of output $1-\alpha$ is large. Note that in the limit case of perfect competition ($\rho = 1$, $c_i = 0$ and no productivity differences), static multiplicity can never arise.

2.B.5 Basins of Attraction

Steady-state output

In a steady-state with a constant productivity distribution $\mathbb{Z}_\tau$ and a set of active firms $\{n_i\}_{i=1}^I$, the aggregate savings rate is equal to

$$s = \frac{\beta\delta}{1-(1-\delta)\beta}\,\alpha\,\Omega\left(\mathbb{Z}_{\tau t}, \{n_i\}_{i=1}^I\right)$$

Recall that we also have

$$Y = \Phi\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\left[(1-\alpha)\,\Theta\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\frac{1-\alpha}{\nu+\alpha}} K^{\alpha\frac{1+\nu}{\nu+\alpha}}$$

We can combine the above two equations with

$$\delta K = sY$$

to write

$$\begin{aligned}
& Y = \Phi\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\left[(1-\alpha)\,\Theta\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\frac{1-\alpha}{\nu+\alpha}}\left[\frac{\beta}{1-(1-\delta)\beta}\,\alpha\,\Omega\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)Y\right]^{\alpha\frac{1+\nu}{\nu+\alpha}} \\
\Leftrightarrow\ & Y^{\frac{\nu-\alpha\nu}{\nu+\alpha}} = \Phi\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\left[(1-\alpha)\,\Theta\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\frac{1-\alpha}{\nu+\alpha}}\left[\frac{\beta}{1-(1-\delta)\beta}\,\alpha\,\Omega\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\alpha\frac{1+\nu}{\nu+\alpha}} \\
\Leftrightarrow\ & Y = \left[\Phi\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\frac{\nu+\alpha}{\nu-\alpha\nu}}\left[(1-\alpha)\,\Theta\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\frac{1-\alpha}{\nu-\alpha\nu}}\left[\frac{\beta}{1-(1-\delta)\beta}\,\alpha\,\Omega\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\alpha\frac{1+\nu}{\nu-\alpha\nu}} \\
\Leftrightarrow\ & Y = \left[\Phi\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\frac{\nu+\alpha}{\nu-\alpha\nu}-\alpha\frac{1+\nu}{\nu-\alpha\nu}}\left[\Theta\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\frac{1-\alpha}{\nu-\alpha\nu}+\alpha\frac{1+\nu}{\nu-\alpha\nu}}(1-\alpha)^{\frac{1-\alpha}{\nu-\alpha\nu}}\left[\frac{\beta\alpha}{1-(1-\delta)\beta}\right]^{\alpha\frac{1+\nu}{\nu-\alpha\nu}} \\
\Leftrightarrow\ & Y = \Phi\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\left[\Theta\left(\mathbb{Z}_\tau, \{n_i\}_{i=1}^I\right)\right]^{\frac{1+\alpha\nu}{\nu(1-\alpha)}}(1-\alpha)^{\frac{1-\alpha}{\nu-\alpha\nu}}\left[\frac{\beta\alpha}{1-(1-\delta)\beta}\right]^{\alpha\frac{1+\nu}{\nu-\alpha\nu}}
\end{aligned}$$

Proof of Proposition 2.4

A symmetric steady-state with n firms per industry is characterized by

$$Y^*(\mathcal{F},n) = \Phi(\mathcal{F},n)\left[\Theta(\mathcal{F},n)\right]^{\frac{1+\alpha\nu}{\nu(1-\alpha)}}(1-\alpha)^{\frac{1-\alpha}{\nu-\alpha\nu}}\left[\frac{\beta\alpha}{1-(1-\delta)\beta}\right]^{\alpha\frac{1+\nu}{\nu-\alpha\nu}}$$

and the minimum level of output consistent with n firms per industry is given by

$$\underline{Y}(\mathcal{F},n) = c_i(1-\rho)\left[1-\frac{\Theta(\mathcal{F},n)}{\pi_n}\right]^{-2}$$

We therefore have that

$$\frac{Y^*(\mathcal{F},n)}{\underline{Y}(\mathcal{F},n)} = \propto\ \Phi\ \left[\Theta(\mathcal{F},n)\right]^{\frac{1+\alpha\nu}{\nu(1-\alpha)}}\ \left[1-\frac{\Theta(\mathcal{F},n)}{\pi_n}\right]^2$$

For any $\pi_k > \pi_n$, we have that

$$\frac{\partial\left(\frac{Y^*}{\underline{Y}}\right)}{\partial\pi_k} < 0 = \frac{\partial\Phi}{\partial\pi_k} + \Phi\left\{\frac{1+\alpha\nu}{\nu(1-\alpha)}\Theta^{-1}\frac{\partial\Theta}{\partial\pi_k} + 2\left(-\frac{1}{\pi_j}\frac{\partial\Theta}{\partial\pi_k}\right)\left[1-\frac{\Theta}{\pi_n}\right]^{-1}\right\} < 0$$

In the special case in which $\pi_k = 1\ \forall k$

$$\frac{\partial\Phi(\mathcal{F},n)}{\partial\pi_k} = \frac{1}{1-\rho}\left[1-\frac{2n+1}{n}\Theta(\mathcal{F},n)\right]$$
$$\frac{\partial\Theta(\mathcal{F},n)}{\partial\pi_k} = \frac{\Theta(\mathcal{F},n)}{n}$$

The above condition hence becomes

$$\frac{1}{1-\rho}\left(1-\frac{2n+1}{n}\Theta\right)\left[\frac{1+\alpha\nu}{\nu(1-\alpha)}\frac{1}{n} - 2\frac{\Theta}{n}(1-\Theta)^{-1}\right] < 0$$
$$\Leftrightarrow\quad \frac{1}{1-\rho}\left[n-(2n+1)\Theta\right] + \left[\frac{1+\alpha\nu}{\nu(1-\alpha)} - 2\frac{\Theta}{1-\Theta}\right] < 0$$

Recall that $\Theta = \dfrac{n-(1-\rho)}{n}$ when $\pi_k = 1\ \forall k$, we can write

$$
\begin{aligned}
& \frac{1}{1-\rho}\left[n-(2n+1)\frac{n-(1-\rho)}{n}\right]+\left[\frac{1+\alpha\nu}{\nu(1-\alpha)}-2\frac{n-(1-\rho)}{1-\rho}\right]<0 \\
\Leftrightarrow\quad & \frac{1}{1-\rho}\left\{n-\left(2+\frac{1}{n}\right)[n-(1-\rho)]-2[n-(1-\rho)]\right\}+\frac{1+\alpha\nu}{\nu(1-\alpha)}<0 \\
\Leftrightarrow\quad & \frac{1}{1-\rho}\left\{n-\left(2+\frac{1}{n}\right)[n-(1-\rho)]-2[n-(1-\rho)]\right\}+\frac{1+\alpha\nu}{\nu(1-\alpha)}<0 \\
\Leftrightarrow\quad & \frac{1}{1-\rho}\left\{n-\left(4+\frac{1}{n}\right)[n-(1-\rho)]\right\}+\frac{1+\alpha\nu}{\nu(1-\alpha)}<0 \\
\Leftrightarrow\quad & \frac{\frac{1}{\nu}+\alpha}{1-\alpha}<\frac{\left(4+\frac{1}{n}\right)[n-(1-\rho)]-n}{1-\rho}
\end{aligned}
$$

Proof of Proposition 2.5

In a steady-state with identical industries and n firms per industry we have that

$$
\begin{aligned}
R^* = &\quad \alpha(1-\alpha)^{\frac{1-\alpha}{\nu+\alpha}}\ [\Theta(n)]^{\frac{1+\nu}{\nu+\alpha}}\ [K^*(n)]^{(1-\alpha)\frac{-\nu}{\nu+\alpha}} \\
R^* = &\quad \beta^{-1}+(1-\delta)
\end{aligned}
$$

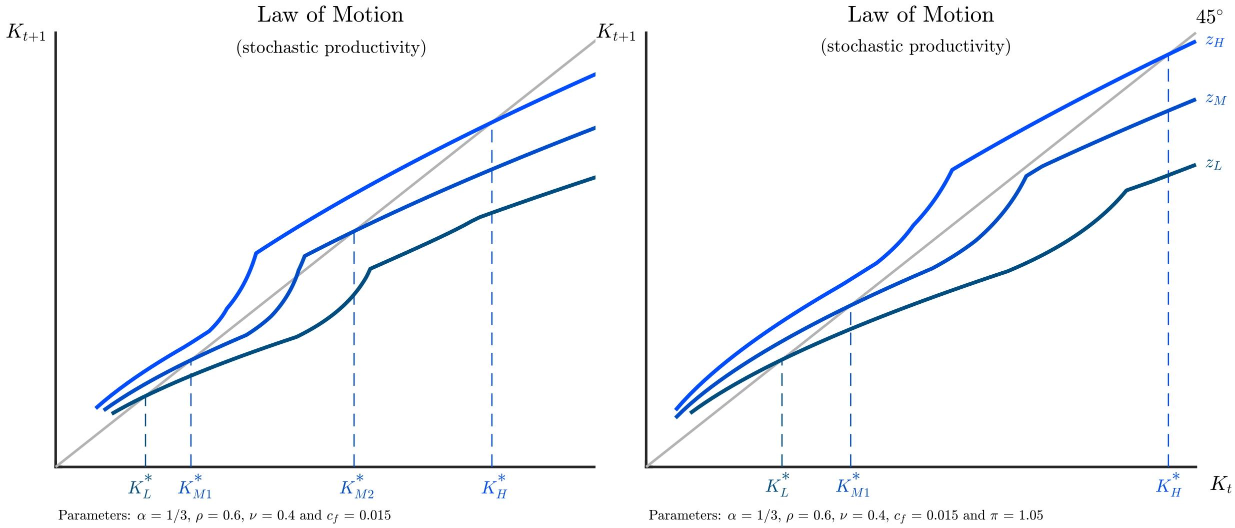
The first equation describes capital demand by fimrs, whereas the second evaluates the Euler equation (in a steady-state). $K^*(n)$ is therefore independent of c_f – provided that all n make strictly positive profits, a marginal increase in c_f will not drive any of them out of the market.

Note furthermore that $\underline{K}(n)$ is increasing in n (Appendix 2.B.3).

$$
\underline{K}(n)=\left\{\frac{c}{\Lambda(n,n)}(1-\alpha)^{-\frac{1-\alpha}{\nu+\alpha}}[\Phi(n)]^{-1}[\Theta(n)]^{\frac{\rho}{1-\rho}-\frac{1-\alpha}{\nu+\alpha}}\right\}^{\frac{\nu+\alpha}{\alpha(1+\nu)}}
$$

2.B.6 Law of Motion: Stochastic Productivity

Suppose for simplicity that z_t can take three values: a low value z_L, an intermediate value z_M and a high value z_H. Figure 2.16a represents the law of motion under each value of aggregate productivity.



(a) Baseline (b) Effect of an increase in π

Figure 2.16: Law of motion with stochastic productivity
Note: the figure shows the law of motion of capital for different levels of aggregate productivity (Panel (a)) and how it changes when the most productivity firm becomes even more productive (Panel (b)).

As one can see, when aggregate productivity is low and equal to z_L, the economy exhibits a unique steady-state where all industries are a monopoly (K^*_L). Under z_H, on the other hand, there is only a unique steady-state where all industries are a duopoly (K^*_H). Finally, when aggregate productivity takes the intermediate value z_M, the economy exhibits two steady-states: a low one where all industries are a monopoly (K^*_{M1}) and a high one where all industries are a duopoly (K^*_{M2}).

To exemplify the dynamics of the model, suppose that aggregate productivity starts at z_H and that the economy is at the steady-state K^*_H. Suppose now that there is a negative aggregate productivity shock, which reduces aggregate productivity permanently to z_M. After this shock, the economy will converge to the new steady-state K^*_{M2}. Output is lower than before, but the market structure is identical – all sectors are still a duopoly.

Now suppose that, instead of falling permanently to z_M, aggregate productivity falls first to z_L, and later increases to z_M. Suppose further that aggregate productivity remains at z_L for sufficiently large period, so that the economy approaches the low steady-state K^*_L. Then, as aggregate productivity increases to z_M, the economy will approach K^*_{M1}. Note now that in the new steady-state all sectors are a monopoly. The economy therefore experiences a persistent transition to a regime featuring a more concentrated market structure.

Figure 2.16b shows the effect of an increase in the productivity of the top firm (π). The main difference with respect to Figure 2.16a is that now, for the intermediate value of aggregate productivity z_M, there is only one steady-state – it becomes increasingly more difficult to sustain a duopoly and, as a consequence, the steady-state K^*_{M2} disappears. Note that the two steady-states K^*_L and K^*_H are larger after the increase in the leaders' productivity. The same happens with the low steady-state when aggregate productivity is equal to z_M (K^*_{M1}). This result is not surprising. Keeping the market structure constant (for example, a monopoly in every industry), the higher the productivity of the leader, the higher is aggregate output.

2.C The Quantitative Model

2.C.1 Calibration

Steady-State We perform two different calibrations of our model – to match the average level of markups and its dispersion in 1985 and in 2007. We need to calibrate three technology parameters: the Pareto tail λ, the fixed production cost c and the fraction of industries with zero fixed cost f_{comp}.

We specify a grid of possible candidates for λ, c and f_{comp}. We also specify a grid with values for the aggregate capital stock K. We then compute the aggregate equilibrium for each parameter combination (λ, c, f_{comp}) and for each value K.[37] We start by assuming that all firms are active, so that there are N firms in each of the I industries. We compute the aggregate equilibrium using equations (2.16) and (2.11). We then compute the profits net of the fixed cost that each firm makes

$$\left(p_{ijt} - \frac{\Theta_t}{\tau_{ijt}}\right) y_{ijt} - c_i$$

and identify the firm with the largest negative value. We exclude this firm and recompute the aggregate equilibrium. We repeat this iterative procedure until all firms have non-negative profits (net of the fixed production cost). For most parameter combinations, our model admits a unique equilibrium. However, if equilibrium multiplicity arises, this algorithm allows us to consistently select the equilibrium that features the largest number of firms.

For each triplet (λ, c, f_{comp}), we then have the general equilibrium computed for all possible capital values. The steady-state(s) of our economy correspond to the value(s) of K for with the rental rate R_t is equal to $\dfrac{1}{\beta} - (1-\delta)$.

Given our interpretation that the US economy was in a competition regime in both 1985 and 2007, we obtain compute model moments in the largest steady-state.

[37] Aggregate TFP e^{z_t} is assumed to be constant and equal to one.

Data Definitions For the sales weighted-average markup, we use the series computed by De Loecker et al. (2020). The authors calculate price-cost markups for the universe of public firms, using data from COMPUSTAT. The markup of a firm j in a 2-digit NAICS sector s at time t is calculated as

$$\mu_{sjt} = \xi_{st} \cdot \frac{\text{sale}_{sjt}}{\text{cogs}_{sjt}}$$

where ξ_{st} is the elasticity of sales to the total variable input bundle, sale_{sjt} is sales and cogs_{sjt} is the cost of the goods sold, which measures total variable costs.

To measure markup dispersion, we compute the standard deviation of markups within 2-digit NAICS sectors. Treating ξ_{st} as constant within a sector s and time t, we can measure markup dispersion within this sector as

$$\text{sd}_s \left[\log\left(\mu_{sjt}\right)\right] = \text{sd}_s \left[\log\left(\frac{\text{sale}_{sjt}}{\text{cogs}_{sjt}}\right)\right]$$

We calculate this measure for all 23 sectors (2-digit NAICS). We then compute an average across all such sectors, weighted by the sector sales. Figure 2.17 shows the evolution of this measure.

In our model, we compute the standard deviation of (log) markups across all firms in the economy, i.e. we do not compute it industry by industry. We think of an industry in our model as a market at the possible level of disaggregation (e.g. 10-digit NAICS). We cannot however observe data at such a fine level of disaggregation – first because most data sets only provide industry information at the 6-digit, second because many large firms are multi-product an operate in different markets. We hence think of our final good Y_t as one big-sector.

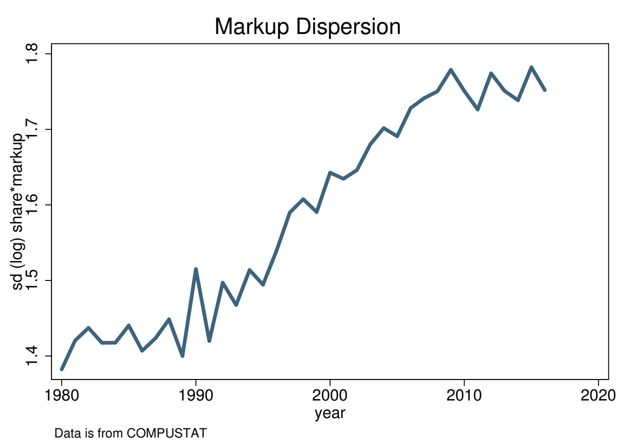


Figure 2.17: Markup Dispersion
Note: the figure shows the evolution of markup dispersion computed from COMPUSTAT data for 2-digit industries.

2.C.2 Solution Algorithm for the Dynamic Optimization Problem

We now explain the algorithm we use for the dynamic optimization problem of the representative household. We take the calibrated parameters (λ, c) and form a grid for aggregate capital with $n_K = 40$ points. This grid is centered around the highest steady-state K_H^{ss}, with a lower-bound $0.6 \times K_H^{\text{ss}}$ and upper bound $1.4 \times K_H^{\text{ss}}$.

We also form a grid for (log) aggregate TFP, z. We use Tauchen's algorithm with $n_z = 9$ points, autocorrelation parameter ϕ_Z and standard deviation for the innovations σ_ε (the last two parameters are calibrated, as explained in the main text).We compute the aggregate equilibrium for each value of K and z.

We next iterate on the policy function of the representative household. Recall that the representative household takes all aggregate variables (rental rate, wage rate and profits) as given. Specifically, he does not internalize the impact that his choice of K can have on aggregate variables. We then start with a guess for the policy function $C_t = f_C\left(K_t, z_Z\right)$. We also start with a guess for the law of motion $K_{t+1} = f_K\left(K_t, z_t\right)$. The representative household takes this law of motion as given (so that he forms expectations about the evolution of aggregate variables that are independent of his choices of capital). We iterate simultaneously on the policy function and on the law of motion.

2.C.3 Alternative Measures of Aggregate Markups

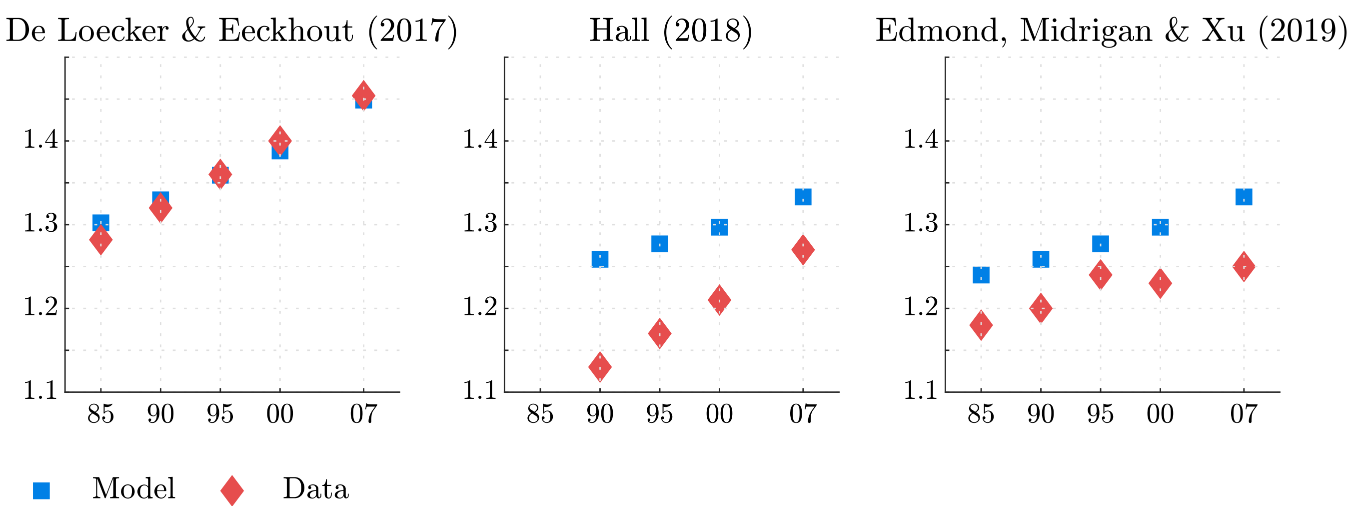


Figure 2.18: Aggregate markup: alternative measures
Note: the figure plots data and model markups computed under different definitions. The left panel shows the sales weighted markups from De Loecker and Eeckhout (2018), the central panel plots the markup from Hall (2018) and right panel the cost share weighted markup from Edmond et al. (2015). Note that the markup from De Loecker and Eeckhout (2018) is a targeted moment in our calibration.

2.C.4 Business Cycle Moments

	Output	Consumption	Investment	Hours
	Correlation with Output			
Data: 1985-2018	1.00	0.99	0.90	0.83
Model: 1985 calibration	1.00	0.99	0.97	1.00
Model: 2007 calibration	1.00	0.99	0.83	1.00
	Standard Deviation Relative to Output			
Data: 1985-2018	1.00	1.04	2.96	0.97
Model: 1985 calibration	1.00	0.98	1.12	0.76
Model: 2007 calibration	1.00	0.96	1.56	0.77

Table 2.8: Business Cycle Moments. All variables are in logs. Data variables are in per capita terms and in deviation from a linear trend computed over 1985-2007.

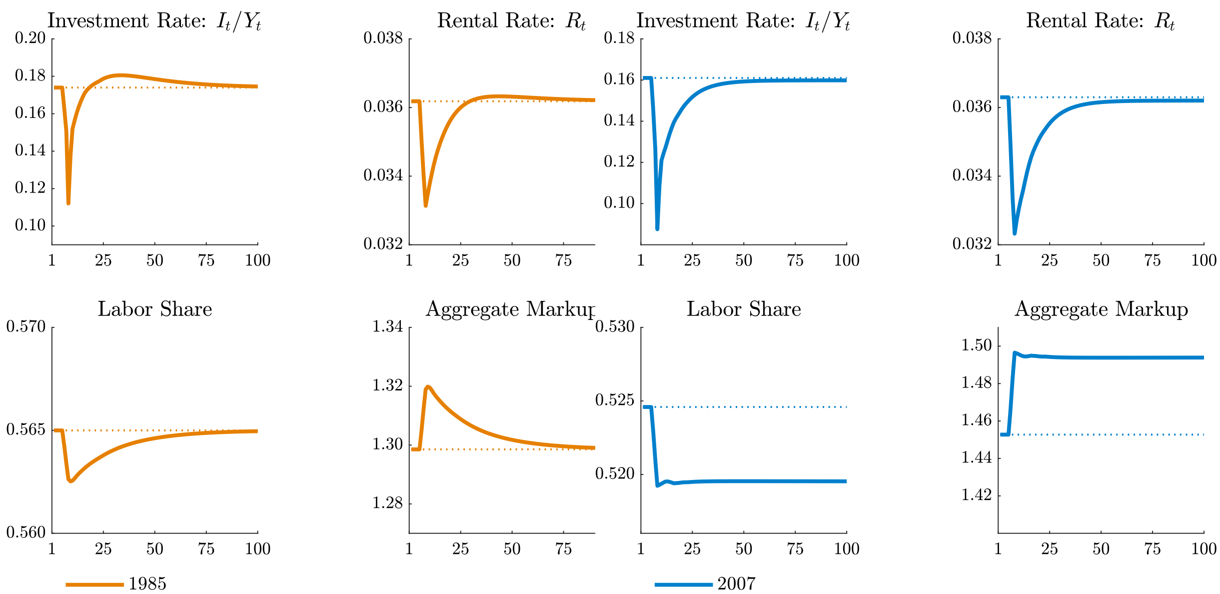
Table 2.8 shows some business cycle moments for our two calibrated economies, as well as their data counterparts. To be consistent with our interpretation that the US economy transitioned to a lower steady-state after 2008, all data variables are in deviation from a linear trend computed over 1985-2007. This fact explains the large empirical correlation between consumption and output. Comparing our two calibrated economies, we see that both economies display the same correlations of consumption and hours with output. The 2007 economy displays, however, a significantly lower correlation of investment with output. This is explained by the fact the investment appears to be more volatile in the 2007 economy.

2.C.5 Impulse Responses: Other Variables

In figures 2.19a and 2.19b we plot the responses of the gross investment rate I_t/Y_t, the rental rate, the labor share and the (sales-weighted) average markup.[38] With respect to the behavior of the investment rate, there is one difference that is worth highlighting. In the 1985 economy, the investment rate suffers a significant drop on impact, but ultimately recovers and overshoots its steady-state level – so that the capital stock can recover to its long-run value. This is not true for the 2007 economy. As the economy is converging to a lower steady state, during the transition there is “too much” capital in the system, which yields a gradual reduction of the stock through a depressed investment rate over the transition. Aggregate markups increase on

[38] In our setup, the steady-state rental rate is independent of the regime of the economy, as it is pinned down by the discount factor of the representative household.

impact as there is firm exit. In the 1985 economy, this increase is reabsorbed as the economy transitions back to its steady state and firms enter the market. In the 2007 economy, such absorption does not take place since the number of firms never goes back to the previous level. Not surprisingly, the labor share exhibits the opposite behavior. In the 1985 economy, the reduction quickly reverts, while the 2007 economy experiences a 0.5 pp permanent drop in the labor share.



(a) 1985 Economy (b) 2007 Economy

Figure 2.19: IRF (Large Shock)
Note: the figure shows the impulse responses of additional variables to a large shock to aggregate TFP for the 1985 (Panel (a)) and the 2007 (Panel (b)) economies.

2.C.6 Aggregate Productivity

Average Firm Level TFP

Figure 2.20 reports a sales-weighted average of firm level revenue TFP. A similar pattern emerges if one uses physical TFP instead.

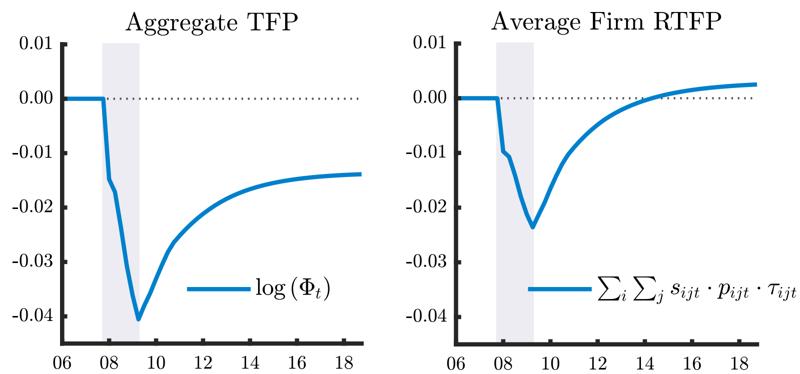


Figure 2.20: Aggregate TFP versus Average Firm Level TFP
Note: The left panel shows aggregate TFP, as defined in equation (2.16). The right panel shows a sales-weighted average of firm level revenue TFP $p_{ijt} \cdot \tau_{ijt}$.

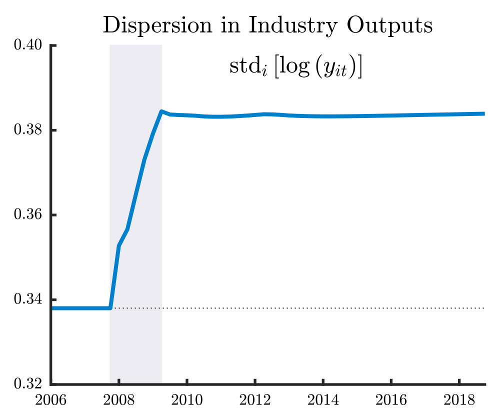


Figure 2.21: Dispersion in $\log(y_{it})$

Dispersion in Industry Output

2.C.7 Welfare

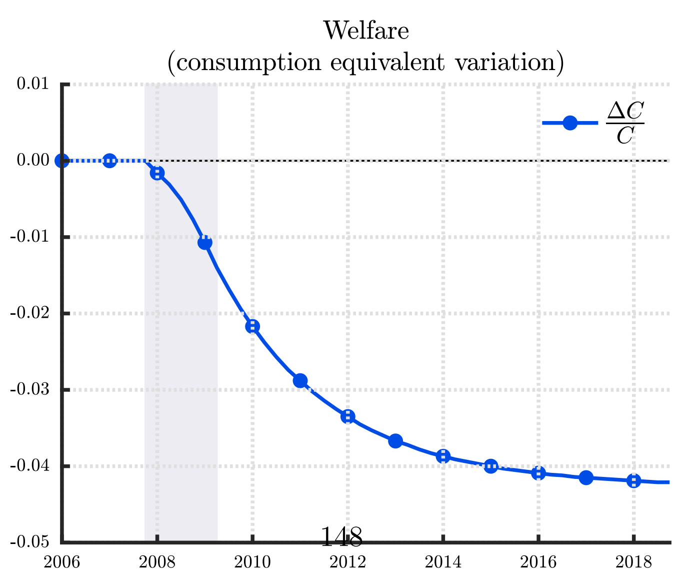
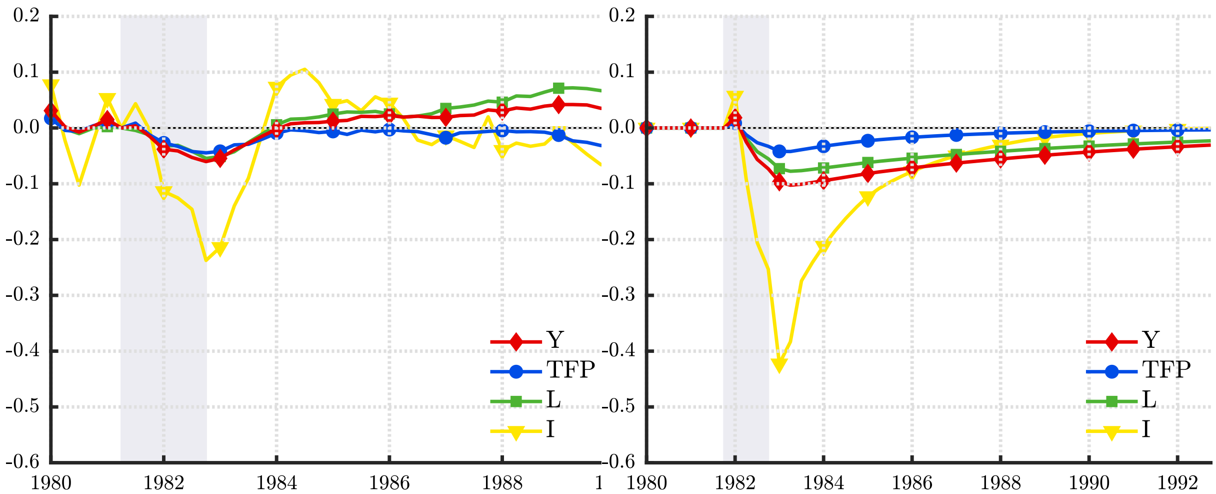


Figure 2.22: The great recession and its aftermath: welfare

2.C.8 The 1981-1982 Recession

The response in the 1985 economy



(a) 1981-1982 recession (data) (b) The 1981-1982 shocks in the 1985 model

Figure 2.23: The 1981-1982 recession

The response in the 2007 economy

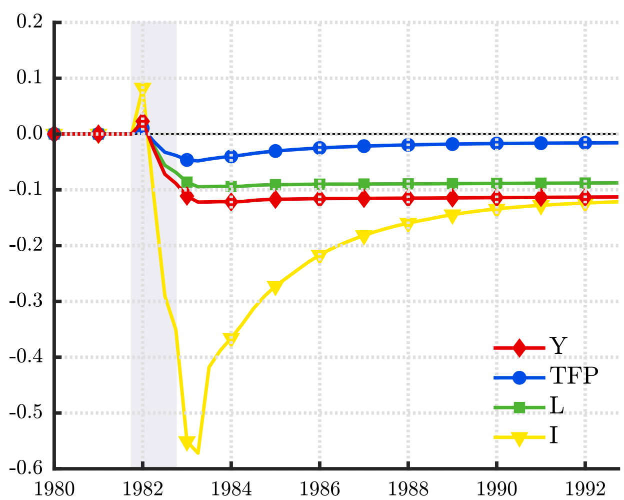


Figure 2.24: The 1981-1982 shocks in the 2007 model

2.D Policy Evaluation

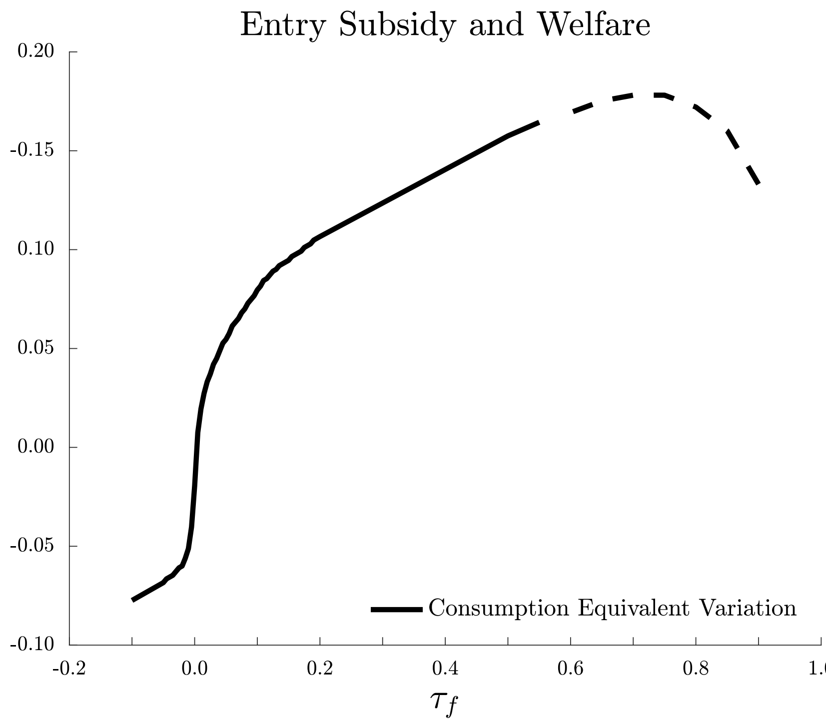


Figure 2.25: Welfare: consumption equivalent gain

Note: the figure shows the welfare effect of an entry subsidy of size τ_f of the entry cost. The dotted part of the graph represents values for which the government is unable to run a balance budget as it would require imposing a more than 100% tax rate on net profits.

2.E Regression Tables

VARIABLES	(1) $\Delta \log \text{emp}_{07-16}$	(2) $\Delta \log \text{emp}_{07-16}$	(3) $\Delta \log \text{emp}_{07-16}$	(4) $\Delta \log \text{emp}_{07-16}$
concent_{07}	-0.0223***	-0.0160**	-0.0177***	-0.0178**
	(0.00667)	(0.00688)	(0.00682)	(0.00732)
$\log \text{firms}_{07}$		0.00239***	0.00193***	0.00151
		(0.000705)	(0.000706)	(0.000983)
$\Delta \log \text{emp}_{03-07}$			0.0984***	0.0901***
			(0.0241)	(0.0247)
Observations	770	770	769	761
R-squared	0.014	0.029	0.050	0.064
Sector FE	NO	NO	NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.9: Change in Employment: 2007-2016
Note: the table shows the results of regressing the growth rate of sectoral employment between 2007 and 2016 on the measure of concentration in 2007. The table presents the results of progressively adding controls and, in the last column, sector fixed effects.

	(1)	(2)	(3)	(4)
VARIABLES	Δ log payroll$_{07-16}$	Δ log payroll$_{07-16}$	Δ log payroll$_{07-16}$	Δ log payroll$_{07-16}$
concent$_{07}$	-0.0231***	-0.0177**	-0.0189***	-0.0194***
	(0.00679)	(0.00702)	(0.00697)	(0.00749)
log firms$_{07}$		0.00203***	0.00164**	0.000991
		(0.000724)	(0.000725)	(0.00101)
Δ log payroll$_{03-07}$			0.0823***	0.0697***
			(0.0219)	(0.0225)
Observations	774	774	773	765
R-squared	0.015	0.025	0.043	0.054
Sector FE	NO	NO	NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.10: Change in Total Payroll: 2007-2016
Note: the table shows the results of regressing the growth rate of sectoral total payroll between 2007 and 2016 on the measure of concentration in 2007. The table presents the results of progressively adding controls and, in the last column, sector fixed effects.

VARIABLES	(1) $\Delta \log \text{firms}_{07-16}$	(2) $\Delta \log \text{firms}_{07-16}$	(3) $\Delta \log \text{firms}_{07-16}$	(4) $\Delta \log \text{firms}_{07-16}$
concent_{07}	-0.0432***	-0.0391***	-0.0406***	-0.0231***
	(0.00608)	(0.00637)	(0.00635)	(0.00666)
$\log \text{firms}_{07}$		0.00137**	0.00119*	0.00449***
		(0.000663)	(0.000661)	(0.000897)
$\Delta \log \text{firms}_{03-07}$			0.0881***	0.0808***
			(0.0270)	(0.0273)
Observations	791	791	791	782
R-squared	0.060	0.065	0.078	0.151
Sector FE	NO	NO	NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.11: Change in Number of Firms: 2007-2016
Note: the table shows the results of regressing the growth rate of the industry number of firms between 2007 and 2016 on the measure of concentration in 2007. The table presents the results of progressively adding controls and, in the last column, sector fixed effects.

	(1)	(2)	(3)	(4)
VARIABLES	Δlab_share$_{07-16}$	Δlab_share$_{07-16}$	Δlab_share$_{07-16}$	Δlab_share$_{07-16}$
concent$_{07}$	-0.0314*	-0.0319*	-0.0314*	-0.0301
	(0.0167)	(0.0168)	(0.0167)	(0.0196)
log firms$_{07}$		-0.00111	-0.00120	-0.00255
		(0.00240)	(0.00240)	(0.00335)
Δlab_share$_{03-07}$			0.169*	0.146*
			(0.0867)	(0.0871)
Observations	99	99	98	97
R-squared	0.035	0.037	0.075	0.111
Sector FE	NO	NO	NO	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.12: Change in Labor Share: 2007-2016
Note: the table shows the results of regressing the growth rate of sectoral labor share between 2007 and 2016 on the measure of concentration in 2007. The table presents the results of progressively adding controls and, in the last column, sector fixed effects.

Appendix B

Not for Publication

2.F Number of Firms per Sector

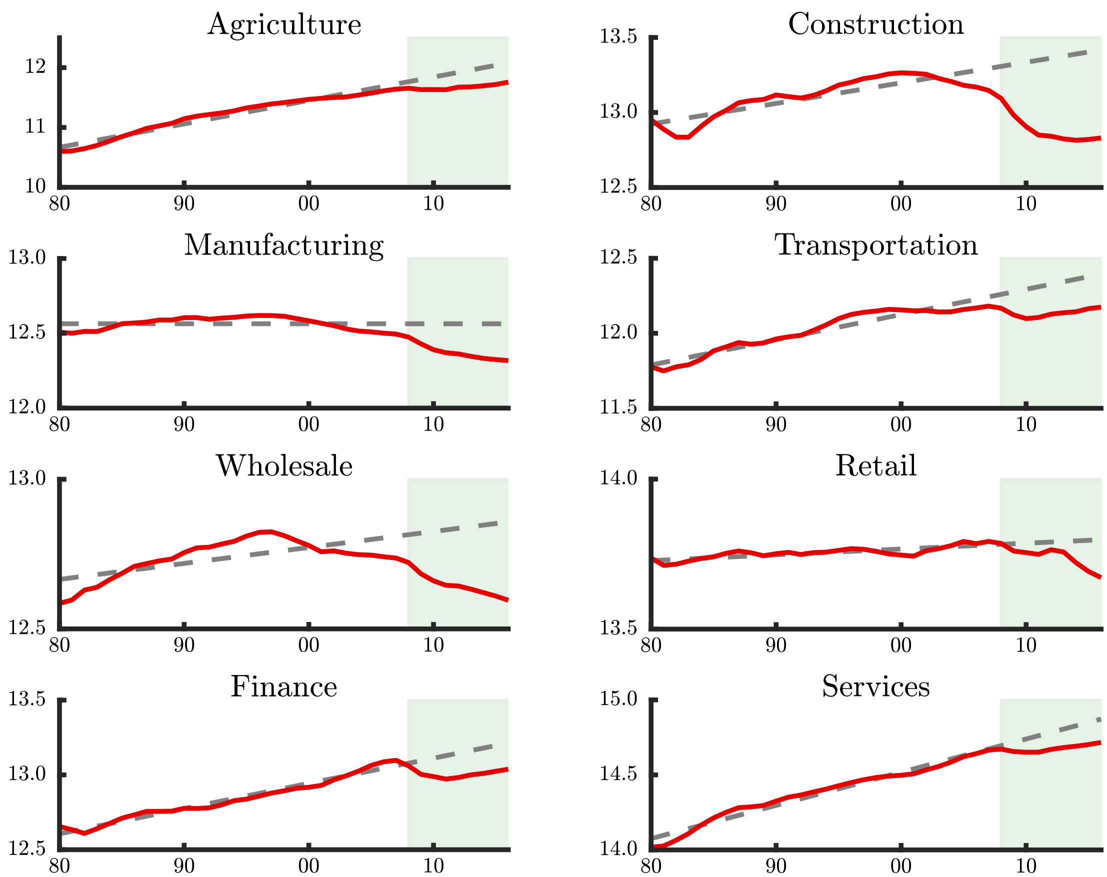


Figure 2.1: **Number of Firms per Sector: 1980-2016**
Each panel shows the number of firms with at least one employee in each sector (in logs). For each series, the dashed grey line shows a linear trend computed over the 1980-2007 period. Data is from the US Business Dynamics Statistics

2.G Derivation and Proofs: Industry Equilibrium

2.G.1 Static Cournot Game

Equilibrium Price and Output

All the results below are derived under $\eta = 1$; in this case, the model admits an analytical solution. When n firms produce, we have a system of n first order conditions

$$p\left[1-(1-\rho)\, s_j\right] = \frac{\Theta}{\pi_j}$$

Dividing the first order condition of firm j by that of firm 1 we obtain

$$\begin{aligned} & \frac{1-(1-\rho)\, s_j}{1-(1-\rho)\, s_1} = \frac{\pi_1}{\pi_j} \\ \Leftrightarrow \quad & 1-(1-\rho)\, s_j = \frac{\pi_1}{\pi_j}\left[1-(1-\rho)\, s_1\right] \\ \Leftrightarrow \quad & s_j = \frac{1}{(1-\rho)}\left\{1-\frac{\pi_1}{\pi_j}\left[1-(1-\rho)\, s_1\right]\right\} \end{aligned}$$

Note that

$$\begin{aligned} & \sum_{k=1}^{n} s_k = 1 \\ \Leftrightarrow \quad & \sum_{k=1}^{n} \frac{1}{(1-\rho)}\left\{1-\frac{\pi_1}{\pi_k}\left[1-(1-\rho)\, s_1\right]\right\} = 1 \\ \Leftrightarrow \quad & n-\pi_1\left[1-(1-\rho)\, s_1\right] \sum_{k=1}^{n} \frac{1}{\pi_k} = 1-\rho \\ \Leftrightarrow \quad & \frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} = \pi_1\left[1-(1-\rho)\, s_1\right] \end{aligned}$$

Plugging the last equation into the first order condition of firm 1 we obtain

$$\begin{aligned} & p\frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} = \Theta \\ \Leftrightarrow \quad & p = \frac{\sum_{k=1}^{n} \frac{1}{\pi_k}}{n-(1-\rho)}\Theta \end{aligned}$$

Total output is hence equal to

$$
\begin{aligned}
& y = p^{-\frac{1}{1-\rho}} Y \\
\Leftrightarrow \quad & y = \left[\frac{\sum_{k=1}^{n} \frac{1}{\pi_k}}{n - (1-\rho)} \Theta \right]^{-\frac{1}{1-\rho}} Y
\end{aligned}
$$

Market Shares

Plugging the previous equation into the first order condition of firm j we have

$$
\begin{aligned}
& 1 - (1-\rho)\, s_j = \frac{n - (1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \frac{1}{\pi_j} \\
\Leftrightarrow \quad & s_j = \frac{1}{1-\rho} \left[1 - \frac{n - (1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \frac{1}{\pi_j} \right]
\end{aligned}
$$

It is easy to verify that each firm's market share decreases in the total number of active firms. To see this, suppose that the number of firms increases from n to $n+1$. The new entrant will have a market share

$$
s_{n+1} = \frac{1}{1-\rho} \left[1 - \frac{n + 1 - (1-\rho)}{\sum_{k=1}^{n+1} \frac{1}{\pi_k}} \frac{1}{\pi_{n+1}} \right]
$$

which is non-negative provided that

$$
\pi_{n+1} \sum_{k=1}^{n+1} \frac{1}{\pi_k} > n + 1 - (1-\rho) \tag{2.17}
$$

and below one given that

$$
\pi_{n+1} \sum_{k=1}^{n+1} \frac{1}{\pi_k} < \frac{1}{\rho} \left[n + 1 - (1-\rho) \right] \tag{2.18}
$$

If we compare the market share of firm j when there n and $n+1$ firms in the market, we have

$$
\begin{aligned}
& s_j\,|n+1 < s_j\,|n \\
\Leftrightarrow \quad & \frac{1}{1-\rho}\left[1-\frac{n+1-(1-\rho)}{\sum_{k=1}^{n+1}\frac{1}{\pi_k}}\frac{1}{\pi_j}\right] < \frac{1}{1-\rho}\left[1-\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}\frac{1}{\pi_j}\right] \\
\Leftrightarrow \quad & \frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}} < \frac{n+1-(1-\rho)}{\sum_{k=1}^{n+1}\frac{1}{\pi_k}} \\
\Leftrightarrow \quad & [n-(1-\rho)]\left(\frac{1}{\pi_{n+1}}+\sum_{k=1}^{n}\frac{1}{\pi_k}\right) < [n+1-(1-\rho)]\sum_{k=1}^{n}\frac{1}{\pi_k} \\
\Leftrightarrow \quad & [n-(1-\rho)]\frac{1}{\pi_{n+1}} < \sum_{k=1}^{n}\frac{1}{\pi_k} \\
\Leftrightarrow \quad & \pi_{n+1}\sum_{k=1}^{n+1}\frac{1}{\pi_k} > n-(1-\rho)
\end{aligned}
$$

Note that the last condition is implied by (2.17).

Profits

When there are n active firms, type π_j makes production profits

$$
\begin{aligned}
\Pi\left(\pi_j, n, \mathcal{F}, \Theta, Y\right) &= \left(p-\frac{\Theta}{\pi_j}\right) s_j\, y_j \\
&= \underbrace{\frac{1}{1-\rho}\left[1-\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}\frac{1}{\pi_j}\right]^2\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}\right]^{\frac{\rho}{1-\rho}}}_{\equiv \Lambda(\pi_j, n, \mathcal{F})}\Theta^{-\frac{\rho}{1-\rho}}Y
\end{aligned}
$$

Proposition 2.7

When $\eta = 1$, the profit function $\Pi(j, n_{it}, \mathcal{F}_i, X_t)$ satisfies

$$1. \quad \frac{\partial \Pi(j, n_{it}, \mathcal{F}_i, X_t)}{\partial Y_t} > 0$$

$$2. \quad \frac{\partial \Pi(j, n_{it}, \mathcal{F}_i, X_t)}{\partial n_{it}} < 0 \qquad , n_{it} > j$$

$$3. \quad \frac{\partial \Pi(j, n_{it}, \mathcal{F}_i, X_t)}{\partial \pi_{ij}} > 0$$

$$4. \quad \frac{\partial \Pi(j, n_{it}, \mathcal{F}_i, X_t)}{\partial \pi_{ik}} < 0 \qquad , \forall k \neq j.$$

Proof. We start by showing that $\Pi(\cdot)$ increases in π_j

$$2\left[1 - \frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \frac{1}{\pi_j}\right]^{-1} \left\{ -\frac{-[n-(1-\rho)]\left[-\left(\frac{1}{\pi_j}\right)^2\right]}{\left(\sum_{k=1}^{n} \frac{1}{\pi_k}\right)^2} \frac{1}{\pi_j} + \frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \left(\frac{1}{\pi_j}\right)^2 \right\} +$$

$$\frac{\rho}{1-\rho}\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}}\right]^{-1} \frac{-[n-(1-\rho)]\left[-\left(\frac{1}{\pi_j}\right)^2\right]}{\left(\sum_{k=1}^{n} \frac{1}{\pi_k}\right)^2} > 0$$

$$\Leftrightarrow \quad 2\left[1 - \frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \frac{1}{\pi_j}\right]^{-1} \left\{ -\frac{1}{\left(\sum_{k=1}^{n} \frac{1}{\pi_k}\right)^2} \frac{1}{\pi_j} + \frac{1}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \right\} +$$

$$\frac{\rho}{1-\rho}\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}}\right]^{-1} \frac{1}{\left(\sum_{k=1}^{n} \frac{1}{\pi_k}\right)^2} > 0$$

$$\Leftrightarrow \quad 2\left[1 - \frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \frac{1}{\pi_j}\right]^{-1} \left\{ -\frac{1}{\pi_j} + \sum_{k=1}^{n} \frac{1}{\pi_k} \right\} + \frac{\rho}{1-\rho}\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}}\right]^{-1} \frac{1}{\left(\sum_{k=1}^{n} \frac{1}{\pi_k}\right)^2} > 0$$

$$\Leftrightarrow \quad 2\left[1 - \frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \frac{1}{\pi_j}\right]^{-1} \left(\sum_{k\neq j}^{n} \frac{1}{\pi_k}\right) + \frac{\rho}{1-\rho}\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}}\right]^{-1} > 0$$

To prove points (ii) and (iii) it suffices to show that $\Lambda(\cdot)$ is decreasing in $\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}$.[39] We have that

$$
\begin{aligned}
& 2\left[1-\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}\frac{1}{\pi_j}\right]^{-1}\left(-\frac{1}{\pi_j}\right)+\frac{\rho}{1-\rho}\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}\right]^{-1}<0 \\
\Leftrightarrow\quad & \frac{\rho}{1-\rho}\left[1-\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}\frac{1}{\pi_j}\right]<2\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}\right]\frac{1}{\pi_j} \\
\Leftrightarrow\quad & \frac{\rho}{1-\rho}<\left(2+\tfrac{\rho}{1-\rho}\right)\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}\right]\frac{1}{\pi_j} \\
\Leftrightarrow\quad & \rho<(2-\rho)\left[\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}\right]\frac{1}{\pi_j} \\
\Leftrightarrow\quad & \pi_j\sum_{k=1}^{n}\frac{1}{\pi_k}<\frac{2-\rho}{\rho}\left[n-(1-\rho)\right]
\end{aligned}
$$

The last condition is implied by (2.18). ■

Besides stating that profits are strictly decreasing in the number of active firms n_{it}, Proposition 2.7 says the profits of any firm j are increasing in its own idiosyncratic productivity π_{ij} and decreasing in the idiosyncratic productivity of all the other firms π_{ik} (*ceteris paribus*). The fact that the profits of j decrease in the productivity of any competitor is crucial to understand the mechanism at the heart of the model. Suppose, for example, that the most productive firm becomes even more productive, while the productivity of all the other firms remains constant. An implication of Proposition 2.7 is that the profits of the least productive firm will necessarily decrease. But if this firm experiences a sufficiently large decrease in profits, those profits may become lower than the fixed production cost c_i, and the firm may be driven out of the market.

[39] We know that $\frac{n-(1-\rho)}{\sum_{k=1}^{n}\frac{1}{\pi_k}}$ increases in n.

2.H Derivation and Proofs: General Equilibrium

2.H.1 Factor Costs and Factor Shares

Aggregate TFP

Suppose all industries are identical and that $\eta = 1$. Let s_j denote the market share of firm j and let v_j denote its input share within the industry. Note that firm j produces $\frac{s_j}{s_1}$ as much output as firm 1 and uses $\frac{s_j}{s_1}\frac{\pi_1}{\pi_j}$ as many inputs. We have that

$$
\begin{aligned}
& \sum_{k=1}^{n} v_k = 1 \\
\Leftrightarrow \quad & v_1 \frac{\pi_1}{s_1} \sum_{k=1}^{n} \frac{s_k}{\pi_k} = 1 \\
\Leftrightarrow \quad & v_1 = \frac{s_1}{\pi_1} \left(\sum_{k=1}^{n} \frac{s_k}{\pi_k} \right)^{-1}
\end{aligned}
$$

Note that we can write aggregate output as

$$
\begin{aligned}
& Y_t = \left[\sum_{i=1}^{I} \left(\sum_{k=1}^{n} y_k \right)^{\rho} \right]^{\frac{1}{\rho}} \\
\Leftrightarrow \quad & Y_t = I^{\frac{1}{\rho}} \sum_{k=1}^{n} y_k \\
\Leftrightarrow \quad & Y_t = I^{\frac{1}{\rho}} \sum_{k=1}^{n} \pi_k \left(v_k I^{-1} L \right)^{1-\alpha} \left(v_k I^{-1} K \right)^{\alpha} \\
\Leftrightarrow \quad & Y_t = I^{-\frac{1-\rho}{\rho}} \left(\sum_{k=1}^{n} \pi_k v_k \right) L^{1-\alpha} K^{\alpha} \\
\Leftrightarrow \quad & Y_t = I^{-\frac{1-\rho}{\rho}} \left(\sum_{k=1}^{n} \pi_k \frac{s_k}{s_1} \frac{\pi_1}{\pi_k} v_1 \right) L^{1-\alpha} K^{\alpha} \\
\Leftrightarrow \quad & Y_t = I^{-\frac{1-\rho}{\rho}} \left(\frac{\pi_1}{s_1} v_1 \underbrace{\sum_{k=1}^{n} s_k}_{=1} \right) L^{1-\alpha} K^{\alpha} \\
\Leftrightarrow \quad & Y_t = I^{-\frac{1-\rho}{\rho}} \left(\sum_{k=1}^{n} \frac{s_k}{\pi_k} \right)^{-1} L^{1-\alpha} K^{\alpha} \\
\Leftrightarrow \quad & Y_t = \underbrace{I^{-\frac{1-\rho}{\rho}} \left\{ \sum_{k=1}^{n} \frac{1}{\pi_k} \frac{1}{1-\rho} \left[1 - \frac{n-(1-\rho)}{\sum_{h=1}^{n} \frac{1}{\pi_h}} \frac{1}{\pi_k} \right] \right\}^{-1}}_{:=\Phi(n)} L^{1-\alpha} K^{\alpha}
\end{aligned}
$$

In this particular case, in which all industries are identical and $\eta = 1$, aggregate productivity decreases in the number of active firms. This result simply reflects the fact that firms enter in reverse order of productivity. To prove it, note the each firm's market share decreases in the number of active firms in its industry. We hence have that $s_k < \tilde{s}_k$, $\forall k$ and that $\pi_{n+1} < \pi_k$, $\forall k \leq n$. These facts imply that

$$\sum_{k=1}^{n} s_k \frac{1}{\pi_k} < \left(\sum_{k=1}^{n} \tilde{s}_k \frac{1}{\pi_k} \right) + \tilde{s}_{n+1} \frac{1}{\pi_{n+1}}$$

Aggregate Factor Cost Index

Suppose all industries are identical and that $\eta = 1$. When there are n firms in every industry we have

$$\Theta(n) = \frac{n - (1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}}$$

We can show that $\Theta(n)$ is increasing in n. To see this, note that

$$\begin{aligned}
& \frac{n + 1 - (1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k} + \frac{1}{\pi_{n+1}}} > \frac{n - (1-\rho)}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \\
\Leftrightarrow \quad & \frac{n + 1 - (1-\rho)}{n - (1-\rho)} > \frac{\sum_{k=1}^{n} \frac{1}{\pi_k} + \frac{1}{\pi_{n+1}}}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \\
\Leftrightarrow \quad & \frac{1}{n - (1-\rho)} > \frac{\frac{1}{\pi_{n+1}}}{\sum_{k=1}^{n} \frac{1}{\pi_k}} \\
\Leftrightarrow \quad & \pi_{n+1} \sum_{k=1}^{n} \frac{1}{\pi_k} > n - (1-\rho)
\end{aligned}$$

The last condition is implied by (2.17).

Proof of Proposition 2.3

Recall that the aggregate factor share is equal to

$$\Omega(\mathcal{F}, n) = \frac{\Theta(\mathcal{F}, n)}{\Phi(\mathcal{F}, n)}$$

As we have seen before, the aggregate factor cost index $\Theta(\mathcal{F}, n)$ is increasing in n, and aggregate TFP $\Phi(\mathcal{F}, n)$ is decreasing in n. To prove the second part of the proposition, note that the aggregate factor share can be written as

$$\Omega\left(\mathcal{F},n\right)=\left(\frac{n}{1-\rho}-1\right)\left\{\sum_{k=1}^{n}\frac{\frac{1}{\pi_k}}{\sum\limits_{h=1}^{n}\frac{1}{\pi_h}}\left[1-\left[n-(1-\rho)\right]\frac{\frac{1}{\pi_k}}{\sum\limits_{h=1}^{n}\frac{1}{\pi_h}}\right]\right\}$$

Take some $j<n$ such that $\pi_j\geq\frac{1}{n}\sum\limits_{h=1}^{n}\pi_h$. First note that we must have that

$$\frac{\frac{1}{\pi_j}}{\frac{1}{n}\sum\limits_{h=1}^{n}\frac{1}{\pi_h}}\leq 1$$

To see this, note that

$$\begin{aligned}
&\frac{1}{\pi_j}\leq\frac{1}{n}\sum_{h=1}^{n}\frac{1}{\pi_h}\\
\Leftrightarrow\quad & 1\leq\frac{1}{n}\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\\
\Leftrightarrow\quad & 1\leq\frac{1}{n}\sum_{h=1}^{n}\frac{\pi_j}{\frac{1}{n}\sum\limits_{k=1}^{n}\pi_k}\frac{\frac{1}{n}\sum\limits_{k=1}^{n}\pi_k}{\pi_h}\\
\Leftrightarrow\quad & 1\leq\underbrace{\frac{\pi_j}{\frac{1}{n}\sum\limits_{k=1}^{n}\pi_k}}_{\geq 1}\underbrace{\frac{1}{n}\sum_{h=1}^{n}\frac{\frac{1}{n}\sum\limits_{k=1}^{n}\pi_k}{\pi_h}}_{\geq 1}
\end{aligned}$$

Now suppose that π_j increases to $\tilde{\pi}_j>\pi_j$. We want to show that

$$
\begin{aligned}
&\Omega\left(\tilde{\mathcal{F}},n\right)<\Omega\left(\mathcal{F},n\right)\\
\Leftrightarrow\ &\sum_{k=1}^{n}\frac{\frac{1}{\tilde{\pi}_k}}{\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}}\left[1-\left[n-(1-\rho)\right]\frac{\frac{1}{\tilde{\pi}_k}}{\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}}\right]<\sum_{k=1}^{n}\frac{\frac{1}{\pi_k}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\left[1-\left[n-(1-\rho)\right]\frac{\frac{1}{\pi_k}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\\
\Leftrightarrow\ &\sum_{k=1}^{n}\frac{1}{\tilde{\pi}_k}\left[\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\frac{1}{\tilde{\pi}_k}\right]<\frac{\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\sum_{k=1}^{n}\frac{1}{\pi_k}\left[\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\frac{\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\frac{1}{\pi_k}\right]\\
\Leftrightarrow\ &\sum_{k=1}^{n}\frac{1}{\tilde{\pi}_k}\left[\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\frac{1}{\tilde{\pi}_k}\right]<\left[1+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\cdot\\
&\cdot\sum_{k=1}^{n}\frac{1}{\pi_k}\left\{\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\left[1+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\frac{1}{\pi_k}\right\}\\
\Leftrightarrow\ &\sum_{k=1}^{n}\frac{1}{\tilde{\pi}_k}\left[\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\frac{1}{\tilde{\pi}_k}\right]<\sum_{k=1}^{n}\frac{1}{\pi_k}\left[\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\frac{1}{\pi_k}\right]+\\
&+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\sum_{k=1}^{n}\frac{1}{\pi_k}\left\{\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\left[1+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\frac{1}{\pi_k}\right\}-\sum_{k=1}^{n}\frac{1}{\pi_k}\left\{\left[n-(1-\rho)\right]\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\frac{1}{\pi_k}\right\}\\
\Leftrightarrow\ &\frac{1}{\tilde{\pi}_j}\left[\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\frac{1}{\tilde{\pi}_j}\right]<\frac{1}{\pi_j}\left[\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\frac{1}{\pi_j}\right]+\\
&+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\sum_{k=1}^{n}\frac{1}{\pi_k}\left\{\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\left[1+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\frac{1}{\pi_k}\right\}-\left[n-(1-\rho)\right]\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\sum_{k=1}^{n}\left(\frac{1}{\pi_k}\right)^2\\
\Leftrightarrow\ &\left(\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}\right)\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\left[\left(\frac{1}{\tilde{\pi}_j}\right)^2-\left(\frac{1}{\pi_j}\right)^2\right]<\\
&<\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\sum_{k=1}^{n}\frac{1}{\pi_k}\left\{\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\left[1+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\frac{1}{\pi_k}\right\}-\left[n-(1-\rho)\right]\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\sum_{k=1}^{n}\left(\frac{1}{\pi_k}\right)^2\\
\Leftrightarrow\ &\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\left(\frac{1}{\pi_j}+\frac{1}{\tilde{\pi}_j}\right)>\\
&>\frac{1}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\sum_{k=1}^{n}\frac{1}{\pi_k}\left\{\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h}-\left[n-(1-\rho)\right]\left[1+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\frac{1}{\pi_k}\right\}-\left[n-(1-\rho)\right]\frac{\sum_{k=1}^{n}\left(\frac{1}{\pi_k}\right)^2}{\sum_{h=1}^{n}\frac{1}{\pi_h}}
\end{aligned}
$$

$$\Leftrightarrow \quad \sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h} - [n-(1-\rho)]\left(\frac{1}{\pi_j}+\frac{1}{\tilde{\pi}_j}\right) > \underbrace{\frac{\sum_{k=1}^{n}\frac{1}{\pi_k}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}}_{=1}\sum_{h=1}^{n}\frac{1}{\tilde{\pi}_h} - \frac{n-(1-\rho)}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\left[2+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\sum_{k=1}^{n}\left(\frac{1}{\pi_k}\right)^2$$

$$\Leftrightarrow \quad -[n-(1-\rho)]\left(\frac{1}{\pi_j}+\frac{1}{\tilde{\pi}_j}\right) > -\frac{n-(1-\rho)}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\left[2+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\sum_{k=1}^{n}\left(\frac{1}{\pi_k}\right)^2$$

$$\Leftrightarrow \quad \frac{1}{\pi_j}+\frac{1}{\tilde{\pi}_j} < \frac{1}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\left[2+\frac{\frac{1}{\tilde{\pi}_j}-\frac{1}{\pi_j}}{\sum_{h=1}^{n}\frac{1}{\pi_h}}\right]\sum_{k=1}^{n}\left(\frac{1}{\pi_k}\right)^2$$

$$\Leftrightarrow \quad 1+\frac{\pi_j}{\tilde{\pi}_j} < \frac{\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2}{\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}}\left(2+\frac{\frac{\pi_j}{\tilde{\pi}_j}-1}{\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}}\right)$$

$$\Leftrightarrow \quad \tilde{\pi}_j+\pi_j < 2\tilde{\pi}_j\frac{\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2}{\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}} - (\tilde{\pi}_j-\pi_j)\frac{\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2}{\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2}$$

$$\Leftrightarrow \quad \pi_j\left[1-\frac{\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2}{\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2}\right] < \tilde{\pi}_j\left[2\frac{\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2}{\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}} - \frac{\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2}{\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2} - 1\right]$$

$$\Leftrightarrow \quad \frac{\pi_j}{\tilde{\pi}_j} < \frac{2\frac{\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2}{\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}} - \frac{\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2}{\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2} - 1}{1-\frac{\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2}{\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2}}$$

It only suffices to show that the right hand side of the above inequality is greater than one

$$
\begin{aligned}
& 2\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)-\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2-\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2>\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2-\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2 \\
\Leftrightarrow\; & 2\sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)-\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2>\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2 \\
\Leftrightarrow\; & \sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)>\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)^2 \\
\Leftrightarrow\; & \sum_{k=1}^{n}\left(\frac{\pi_j}{\pi_k}\right)^2\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)>\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)\left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right) \\
\Leftrightarrow\; & \sum_{h=1}^{n}\left(\frac{\pi_j}{\pi_h}\right)^2>\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}
\end{aligned}
$$

It is easy to prove that the last inequality is satisfied provided that $\pi_j \geq \frac{1}{n}\sum_{k=1}^{n}\pi_k$. Note that

$$
\begin{aligned}
& \sum_{h=1}^{n}\frac{\frac{1}{n}\sum_{k=1}^{n}\pi_k}{\pi_h}\frac{\pi_j}{\pi_h}>\sum_{h=1}^{n}\frac{\pi_j}{\pi_h} \\
\Leftrightarrow\; & \left(\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}\right)+\left(\sum_{h=1}^{n}\frac{\frac{1}{n}\sum_{k\neq h}^{n}\pi_k}{\pi_h}\frac{\pi_j}{\pi_h}\right)>\sum_{h=1}^{n}\frac{\pi_j}{\pi_h}
\end{aligned}
$$

2.H.2 Asymmetric Equilibrium

When

$$\overline{K}(\mathcal{F}, n) < K < \underline{K}(\mathcal{F}, n+1)$$

there will be an asymmetric equilibrium at time $t+1$: some industries will contain n firms, whereas some industries will contain $n+1$ firms. The fraction of industries with $n+1$ will be pinned down by a zero profit condition for the marginal entrant in an industry with $n+1$ firms

$$\Lambda(\mathcal{F}, \pi_{n+1}, n+1)\,\Theta^{-\frac{\rho}{1-\rho}} Y = c_i$$

The equilibrium is characterized by 4 variables: the fraction of the industries with $n+1$ firms (η), aggregate output (Y), aggregate productivity (Φ) and the aggregate cost index (Θ). These 4 variables are pinned down by the following 4 equations

$$Y = \Phi\left[(1-\alpha)\,\Theta\right]^{\frac{1-\alpha}{\nu+\alpha}} K^{\alpha\frac{1+\nu}{\nu+\alpha}}$$

$$\Phi = \frac{\left\{(1-\eta)\left[\dfrac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\dfrac{1}{\pi_{1k}}}\right]^{\frac{\rho}{1-\rho}} + \eta\left[\dfrac{n+1-(1-\rho)}{\sum\limits_{k=1}^{n+1}\dfrac{1}{\pi_{2k}}}\right]^{\frac{\rho}{1-\rho}}\right\}^{\frac{1}{\rho}}}{(1-\eta)\left[\dfrac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\dfrac{1}{\pi_{1k}}}\right]^{\frac{1}{1-\rho}}\left(\sum\limits_{k=1}^{n}\dfrac{s_{1k}}{\pi_{1k}}\right) + \eta\left[\dfrac{n+1-(1-\rho)}{\sum\limits_{k=1}^{n+1}\dfrac{1}{\pi_{2k}}}\right]^{\frac{1}{1-\rho}}\left(\sum\limits_{k=1}^{n+1}\dfrac{s_{2k}}{\pi_{2k}}\right)}$$

$$\Theta = \left\{(1-\eta)\left[\dfrac{n-(1-\rho)}{\sum\limits_{k=1}^{n}\dfrac{1}{\pi_{k}}}\right]^{\frac{\rho}{1-\rho}} + \eta\left[\dfrac{n+1-(1-\rho)}{\sum\limits_{k=1}^{n+1}\dfrac{1}{\pi_{k}}}\right]^{\frac{\rho}{1-\rho}}\right\}^{\frac{1-\rho}{\rho}}$$

$$\Lambda(\mathcal{F}, \pi_{n+1}, n+1)\,\Theta^{-\frac{\rho}{1-\rho}} Y = c_i$$

s_{1k} is the market share of firm k in an industry with n firms, whereas s_{2k} is the market share of firm k in an industry with $n+1$ firms. They are defined in Appendix 2.G.1.

2.H.3 Steady-State

Example with Unique Steady-State

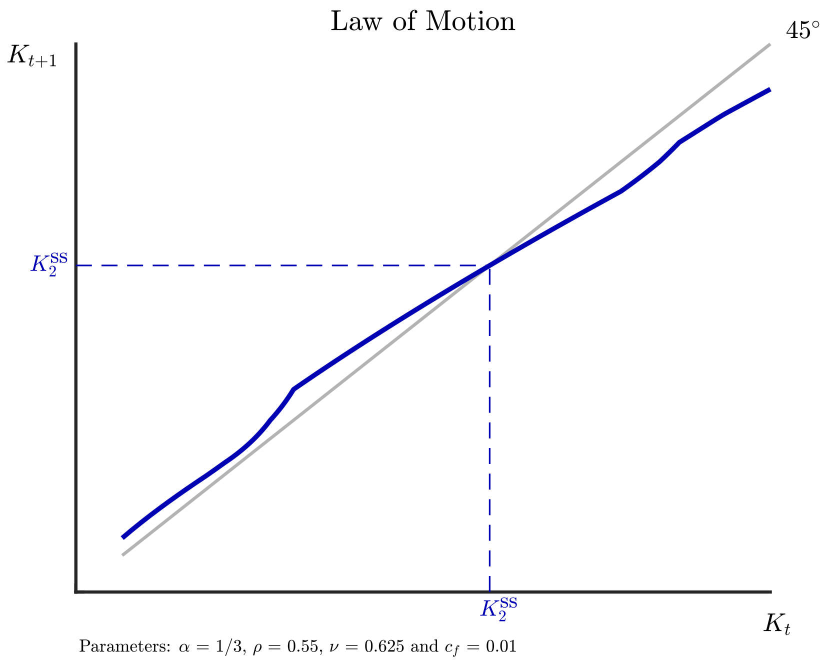


Figure 2.2: Economy with Unique Steady-State

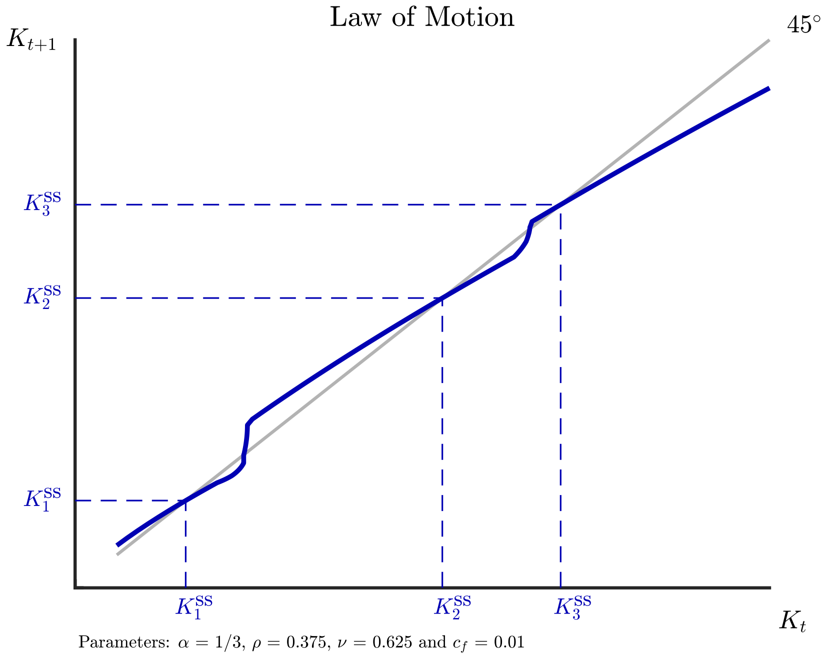


Figure 2.3: Economy with Three Steady-States

	Output	Consumption	Investment	Hours
		Correlation with Output		
Data: 1985-2007	1.00	0.87	0.75	0.52
Model: 1985 calibration	1.00	0.97	0.78	1.00
Model: 2007 calibration	1.00	0.95	0.63	1.00
		Standard Deviation Relative to Output		
Data: 1985-2007	1.00	1.20	4.84	1.72
Model: 1985 calibration	1.00	0.93	1.89	0.75
Model: 2007 calibration	1.00	0.99	2.10	0.76

Table 2.1: Business Cycle Moments. All variables are in logs. Data variables are in per capita terms and in deviation from a linear trend computed over 1985-2007.

Example with Three Steady-State

2.I Quantitative Model

2.I.1 Additional Business Cycle Moments

Business Cycle Moments: High Competition Regime

Business Cycle Moments: Model with Fixed Market Structure

2.I.2 Robustness: Different Elasticities of Substitution

$\sigma_I = 1.25$

The 2008 Crisis

	Output	Consumption	Investment	Hours
		Correlation with Output		
Data: 1985-2018	1.00	0.99	0.90	0.83
Model: 1985 calibration	1.00	0.95	0.74	1.00
Model: 2007 calibration	1.00	0.94	0.70	1.00
		Standard Deviation Relative to Output		
Data: 1985-2018	1.00	1.04	2.96	0.97
Model: 1985 calibration	1.00	0.93	2.18	0.71
Model: 2007 calibration	1.00	0.94	2.39	0.71

Table 2.2: Business Cycle Moments. All variables are in logs. Data variables are in per capita terms and in deviation from a linear trend computed over 1985-2007.

Description	Parameter	Value	Source/Target
Between-Industry ES	σ_I	1.25	Edmond et al. (2015)
Within-Industry ES	σ_G	10	Edmond et al. (2015)
Calibrated Parameters			
Persistence of z_t	ρ_z	0.90	Autocorrelation of log TFP
Standard Deviation of ε_t	σ_ε	0.004	Standard deviation of log TFP
Fraction of Industries with $c_i = 0$	f_{85}	0.870	Emp Share Concentrated Industries
Fraction of Industries with $c_i = 0$	f_{07}	0.870	Emp Share Concentrated Industries
Pareto Tail 1985	λ_{85}	7.40	Markup Dispersion 1985
Pareto Tail 2007	λ_{07}	4.76	Markup Dispersion 2007
Fixed Cost 1985	c_{85}	5.25×10^{-3}	Average Markup 1985
Fixed Cost 2007	c_{07}	17.5×10^{-3}	Average Markup 2007

Table 2.3: Parameter values in the model with $\sigma_I = 1.25$.
We only report the parameters that changed with respect to Table 2.1.

Given these assumptions, need to incur an effective fixed cost

$$\Theta_t \cdot c_f$$

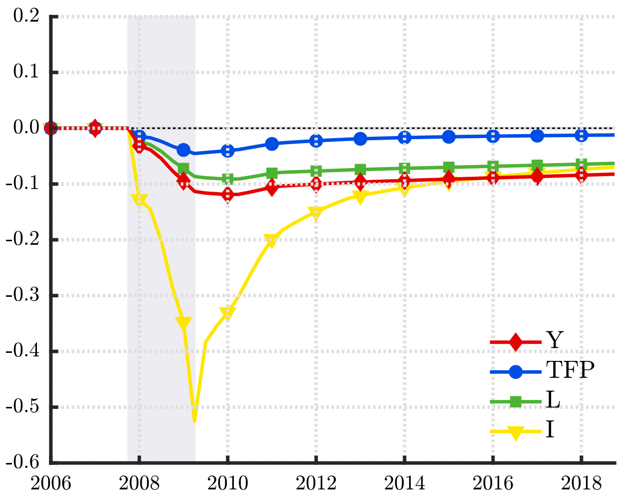
where Θ_t is the factor price index.

Denoting by L_{yt} and K_{yt} the aggregate stocks of labor and capital used in the production,

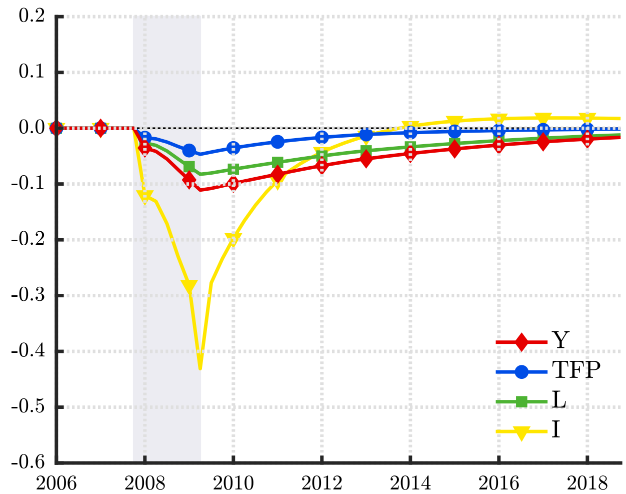
	1985		2007	
	Model	Data	Model	Data
Sales-weighted markup: average	1.33	1.27	1.55	1.46
Sales-weighted markup: standard deviation	1.54	1.44	1.88	1.74
Employment share in concentrated industries	6.23%	-	4.88%	7.62%
Aggregate TFP: autocorrelation	0.983	0.934*	0.936	0.934*
Aggregate TFP: standard deviation	0.027	0.025*	0.017	0.025*

*data moment computed over 1985-2018

Table 2.4: Targeted Moments and Model Counterparts



(a) 2007 Model
This figures replicates Figure 2.10b, under the new calibration strategy



(b) 1985 Model
This figures replicates Figure 2.11, under the new calibration strategy

Figure 2.4: The *great recession* and its aftermath

we have the following market clearing conditions for labor and capital

$$L_t = L_{yt} + N_t^c \cdot l_c$$
$$K_t = K_{yt} + N_t^c \cdot k_c$$

where N^c denotes the number of firms incurring c_f. Note that the optimal mix of l_c and k_c

chosen by each individual firm satisfies

$$\frac{k_c}{l_c} = \frac{K_{yt}}{L_{yt}}$$

Calibration

Description	Parameter	Value	Source/Target
Calibrated Parameters			
Persistence of z_t	ρ_z	0.90	Autocorrelation of log TFP
Standard Deviation of ε_t	σ_ε	0.004	Standard deviation of log TFP
Fraction of Industries with $c_i = 0$	f_{07}	0.763	Emp Share Concentrated Industries
Pareto Tail 2007	λ_{07}	5.84	Markup Dispersion 2007
Fixed Cost 2007	c_{07}	3.99×10^{-3}	Average Markup 2007

Table 2.5: Parameter values in the model with Cobb-Douglas fixed costs.
We only report the parameters that changed with respect to Table 2.1.

The 2008 Crisis

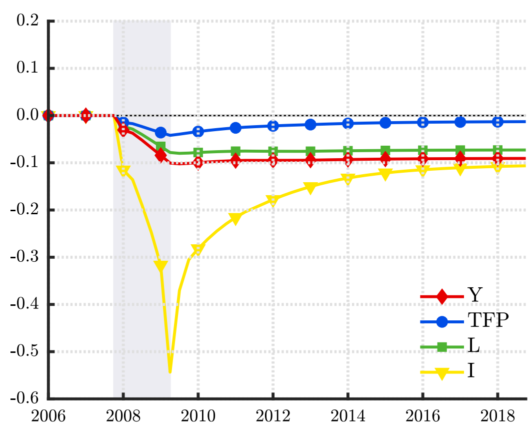


Figure 2.5: The *great recession*
This figures replicates Figure 2.11, in the model with Cobb-Douglas fixed costs.

Chapter 3

Fiscal and Currency Union with Default and Exit

Alessandro Ferrari[1] , Ramon Marimon[2] , Chima Simpson-Bell[3]

3.1 Introduction

In a federal state, with a single currency, states share risks through the federal budget (automatic stabilisers) and other risk-sharing fiscal policies. Furthermore, well-functioning, and integrated, markets – in particular, financial markets – also provide insurance against local shocks and can help to circumvent local nominal rigidities, leaving little role for an independent monetary policy. However, in a monetary union – such as the European EMU – the federal fiscal risk-sharing instruments are missing and markets may not be developed and integrated enough to provide the necessary private risk-sharing. Therefore, independent monetary policy may still have potential value. This point is made formally in Auclert and Rognlie (2014) and Farhi and Werning (2017) who derive optimal risk-sharing policies in a setting with nominal rigidities. Nevertheless, they do not account for two characteristic aspects of unions: in a union of sovereign countries there is limited enforcement. – exit is always an option even if, as in the case of Brexit, it can be a costly option –, but a union is a long-term partnership where mutually beneficial policies bind countries together, deterring them from exiting. Risk-sharing policies can play this role.

In fact, in the euro crisis the threat of exit from the Euro Area, and defaulting on payment

[1]European University Institute

[2]European University Institute, UPF - Barcelona GSE, CEPR and NBER

[3]European University Institute

obligations, has triggered sovereign debt spreads of Greece and other countries. Bayer et al. (2018) provide evidence that market participants even attached positive probability to Germany and France's exit from the common currency during the crisis. Any risk-sharing within a currency union is therefore subject to participation constraints, for both borrowing and lending countries. This point has been made by Abraham et al. (2019) who characterise constrained-efficient risk-sharing contracts as a self-enforcing mechanism within a union which is subject to limited enforcement constraints. Nevertheless, since they model a fiscal – not a monetary – union, the loss (or possible gain, if exiting) of an independent monetary policy plays no role in their analysis.

Our paper integrates these two earlier approaches by analysing long-term risk-sharing contracts as self-enforcing mechanisms in currency unions, taking into account that in monetary unions exit can take two forms. Union members can exit the union to regain control of monetary policy. For example, Sweden or Poland are full members of the European Union who persist in keeping their currencies. Alternatively, union members can exit the union to renege on their obligations; in particular, default on their debts. Default, or partial default, does not necessarily imply exit from the union (e.g. defaulting states in United States, Greece in 2012) but, as a union's 'participation constraint', the relevant case is when the possibility of default is associated with exit.

We model the union as two identical countries facing a simple nominal rigidity which creates a stabilisation role for monetary policy. There is no aggregate risk, meaning that country-risks are fully negatively correlated. We then derive the optimal history dependent transfer policies as a long-term dynamic contract subject to participation constraints. Under these constraints, the contract must improve upon an outside option in which each country has independent monetary policy, allowing it to eliminate the distortion caused by the nominal rigidity, and can borrow and lend using defaultable one-period bonds. Due to the forward looking nature of the participation constraints, we are able to characterize the constrained efficient allocation using the recursive contract solution techniques developed in Marcet and Marimon (2019). Effectively, the contract, as a social contract, gives more weight to a country whenever its participation constraint binds.

We compare the performance of the currency union against two benchmarks: an optimal fiscal union in which the nominal rigidity does is fully eliminated by independent monetary policy, and a two good version of the defaultable debt economy in Arellano (2008). We start by characterizing a number of results regarding the comparison between the fiscal union and the currency union with fiscal transfers. We show that if the fiscal transfers are able to achieve full risk-sharing then the two unions are identical. This result is a version of what the literature has labeled the "risk-sharing miracle". This result stems from the ability of a common currency to stabilize both economies when full risk-sharing is achieved on the fiscal side.

We show that when the planner cannot attain full risk-sharing the fiscal union is strictly better than the currency union. This comes from the deadweight loss that the common monetary policy entails. Such loss shrinks the production possibility frontier, thereby reducing the maximum value attainable by a planner in a currency union.

We also show that optimal common monetary policy is designed to minimize the deadweight loss and that the planner allocation in this economy is still constrained efficient.

We then simulate our economy to study three main features of our model. First we ask whether these kind of contracts are feasible. Secondly, if they are, we ask what is the optimal design of fiscal transfers in terms of size and cyclicality. Thirdly, we investigate how costly is the deadweight loss stemming from the common monetary policy.

In our simulations we find that the fiscal and the currency union are close to identical. We attribute this result to the ability of the optimal policies in the currency union to produce very small deadweight losses. Secondly, in most of our parametrizations, the steady states feature partially state dependent consumption, meaning that the limited enforcement constraints prevent the central planner from achieving full risk-sharing.

As we have mentioned, our work is close to Auclert and Rognlie (2014) and Farhi and Werning (2017) and, therefore, to Hoddenbagh and Dmitriev (2017) who takes a similar approach. We build on this work by considering the participation constraints implied by the option of unpegging from the common currency and , taking it a step forward, by deriving constrained-efficient recursive policies in a monetary union with equally patient countries. We also build on Abraham et al. (2019), although, in contrast with our work, they assume – as the sovereign literature does, to match observed levels of debt – that the 'debtor country' is impatient while the 'lender country' is risk-neutral, in fact, the latter acts as a *financial stability fund*[1]. In this respect, our work is also related to the extensive sovereign debt literature; for example, Gourinchas et al. (2019) solve for the optimal application of a no-bailout rule, finding that less than full enforcement can be sufficient to prevent a fiscally weak country from engaging in risky borrowing.

In our analysis we take the entry or formation decision for the union as given, and focus on the possibility of a breakup, but there is also a literature which considers union formation incentives. In particular, Cooper and Kempf (2003) investigate the conditions under which countries will be able to cooperate to realize the gains from entering a monetary union. Cooper et al. (2008) examines the conditions under which a central authority in a multi-region economy will find it optimal to take on the obligations of regional governments.

After solving for the constrained efficient allocation in our framework, we also propose an

[1]Their contracts also account for moral-hazard constraints: the 'debtor country' can reduce its risk profile with non-contractable effort. We abstract from this feature.

approach to implementing the net payments within the union through trading of state contingent debt contracts. For this we rely on Kehoe and Perri (2004) and Alvarez and Jermann (2000) which demonstrate how constrained efficient allocations with limited commitment can be decentralised using trading of securities. A technical contribution of our paper is that it features a unique feedback between the constrained efficient allocation and its decentralisation. The asset positions calculated in the decentralisation also determine the liabilities which each country carries into its outside option, and these outside options in turn determine the constrained-efficient allocation (recall that in our framework, exit does not imply default). This creates an interesting fixed point problem which we are able to solve numerically.

The rest of the paper is organised as follows. In Section 3.2 we describe the basic two-goods open economy with monopolistic competition in the non-tradeable goods sector and a nominal friction. We describe the contracts which make up the fiscal and currency unions, and the outside options available to each country. We then characterize the constrained efficient allocations of the union contracts and the associated implementation using state contingent debt. In this section we provide the main theoretical results. In Section 3.3 we display the policy functions in the different economies and simulate their responses to a debt crisis. We also study the behaviour of the stochastic steady states.Section 3.4 shows discusses the results under different parametrization of the model. Section 3.6 offers concluding comments.

3.2 Model

We start by modelling a fiscal union between two countries. The two countries have endowments of tradable goods and produce non-tradeables. Differently from the existing literature, we model the two countries as symmetric in terms of risk aversion and patience. Agents can partake in risk-sharing through a long term contract subject to participation constraints, where the outside option is defined by an Arellano economy. Namely, countries can opt out of the risk-sharing contract and borrow through non-state contingent bonds from a risk neutral lender. When in the Arellano economy agents can default on their debt subject to an output cost and temporary exclusion from financial markets.

A key feature of this economy is that if a country leaves the contract but does not default, it starts with a stock of liabilities equal to the present discounted value of the promised transfers in the fiscal union.

Secondly, we extend the setup to accommodate a currency union as a long term risk-sharing contract. In this context nontradables producers are subject to staggered prices nominal rigidities. In this setup the outside option allows countries to move to an Arellano economy with independent monetary policy. When countries leave the currency area they can again pay their previous obligations or default on them. Previous work focuses on optimal risk-sharing schemes within a monetary union without accounting for the incentives of the countries to leave the union.

3.2.1 Environment

Two infinitely lived countries $i \in \{1, 2\}$. Time is discrete. Each period a country receives a random endowment of an identical, freely tradeable good Y_T^i. This is the only source of uncertainty in the model. As there is no aggregate uncertainty the country specific endowments are fully negatively correlated; i.e. for all $t \geq 0$, $Y_{T,t}^1 + Y_{T,t}^2 = Y_T$. Uncertainty is described by a finite state Markov process $\{s_t\}$ with elements $s_t \in \mathcal{S}$ and transition matrix Π – in fact, $s_t = Y_{T,t}^1$. Given this Markov structure, the relevant exogenous state in this environment will be the vector $(s_{t-1}, s_t)^2$.
Preferences over consumption of tradeable goods C_T, non-tradeable goods C_{NT} and labour supply N are given by the utility function:

$$U_i = \mathbb{E}_t \sum_{k=0}^{\infty} \beta^k \left(\frac{(C_{T,t+k})^{1-\gamma}}{1-\gamma} + \alpha \left(\frac{(C_{NT,t+k})^{1-\gamma}}{1-\gamma} - \frac{N_{t+k}^{1+\phi}}{1+\phi} \right) \right) \tag{3.1}$$

All goods are perishable, non-tradeable goods must be consumed in the country in which they are produced, and labour is immobile between countries.
Non-tradeable goods are produced by each country using a technology which is linear in labour input. In each country there is a continuum of firms $j \in [0, 1]$ which produce output according to:

$$Y_{NT}^{ij} = N_{ij} \tag{3.2}$$

The consumer's utility value from consuming each of the varieties j is given by the CES aggregator:

$$C_{NT}^i = \left(\int_{j=0}^1 (C_{NT}^{ij})^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}} \tag{3.3}$$

where $\epsilon > 1$ is the elasticity of substitution between varieties. Consumption of each variety must equal output, i.e. $C_{NT}^{ij} = Y_{NT}^{ij}$, and labour market clearing implies that $N_i = \int_j N_{ij}$.

[2] With an abuse of notation, we also denote (s_{t-1}, s_t) by s, making this explicit when needed.

We assume that production of non tradeables is subsidised at $1/\epsilon$ to erase the monopoly quantity friction. Given its price P^{ij}_{NT}, the producer of variety j satisfies demand C^{ij} by hiring N^{ij} workers and earns profits

$$\Pi^{ij} = (P^{ij}_{NT} - (1+\tau^i_L)W^i)N^{ij} \tag{3.4}$$

Where τ^i_L is the government labor subsidy. Profits are distributed to households.

3.2.2 Fiscal Union

We model the fiscal union as a long term contract. In this setup countries receive state contingent net transfers of the tradeable good $\tau^i(s)$ from each other to absorb the risk associated with the realization of the tradeable goods shock. In this sense, the contract can also be interpreted as a fiscal union. Country i's consumption of the tradeable good is then

$$C^i_T(s) = Y^i_T(s) + \tau^i(s) \tag{3.5}$$

Countries remain in the contract as long as they do not choose to leave the fiscal union. Leaving the union results in the loss of the state contingent transfers.

Planner's problem

The planner arranges transfers within the union subject to each country's outside option:

$$\max_{\{C_{T,i},C_{NT,i},N_i\}} \sum_{i=1,2} \mu_i E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{(C_{T,t})^{1-\gamma}}{1-\gamma} + \alpha \left(\frac{(C_{NT,t})^{1-\gamma}}{1-\gamma} - \frac{N_t^{1+\phi}}{1+\phi} \right) \right) \tag{3.6}$$

s.t.

$$C^i_T(s) = Y^i_T(s) + \tau^i(s) \tag{3.7}$$

$$\sum_{i=1,2} \tau^i(s) = 0 \tag{3.8}$$

$$\sum_{k=0}^{\infty} \beta^k \left(\frac{(C_{T,t+k})^{1-\gamma}}{1-\gamma} + \alpha \left(\frac{(C_{NT,t+k})^{1-\gamma}}{1-\gamma} - \frac{N_{t+k}^{1+\phi}}{1+\phi} \right) \right) \geq V_i^o(s, B) \tag{3.9}$$

$$Y_{NT}^{ij} = N_{ij} \tag{3.10}$$

$$C_{NT}^i = \left(\int_{j=0}^1 (C_{NT}^{ij})^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{3.11}$$

$$N_i = \int_j N_{ij} \tag{3.12}$$

Debt B is defined as the stock of liabilities that a country would have outside the contract. We define this stock as the net present value of the promised transfers inside the contract. The decision of leaving the contract is irreversible. We assume that if a country leaves the contract it refinances its debt with a competitive outside lender borrowing at a risk free rate r. In other words, the outside option to the fiscal union is an Arellano type economy. Formally, the stock of liabilities that a country carries outside the contract is defined as

$$B_{i,t} = \mathbb{E}_t \sum_{s=t}^{\infty} q_{t,s}(Y_{i,s} - c_{i,s}) \tag{3.13}$$

Where $q_{t,s} = \prod_{s=t}^{k} q_{s,s+1}\ q_{t-1,t} \equiv \max_j \left\{ \beta \frac{U_j'^{t}}{U_j'^{t-1}} \right\}$. This assignment of liabilities will be made clearer when we outline the decentralization of the contract which describes the union. Note that the planner does not internalize the effect of within contract transfers on the value of the outside option.

Outside options

In each period, each country has the option of defaulting on its payments within the union, and choosing to leave the fiscal union. Thus, when it is inside the contract, country i faces a choice over the actions $\{SR, LR, LD\}$, i.e. stay in the fiscal union and repay transfer commitments, leave the contract and honour payments, and leave and default. We assume that defaulting on payments triggers temporary exclusion from financial markets so that the country can no longer trade bonds for a stochastic number of periods. Following Arellano (2008), a country in autarky also suffers an output cost on its endowment of the tradeable good, $\chi(Y_T^i)$; this output cost is chosen to ensure that a country is more likely to consider default when its endowment of the tradeable good is low.

We can write the decision problem of each country outside the contract in a recursive form. The value of leaving the contract takes the following form:

$$V_i^o(s,B) = \max_{LR,LD}\{V_i^{LR}(s,B), V_{LD}^i(s)\} \tag{3.14}$$

Namely, the agent can choose whether to repay the promised transfers or default on its obligations. If the country defaults it is temporarily relegated to financial autarky and faces a proportional output cost in terms of tradable endowment.

$$V_i^{LD}(s) = \max_{C_{NT,i},N_i} \frac{(Y_{T,i}-\chi(Y_{T,i}))^{1-\gamma}}{1-\gamma} + \alpha\left(\frac{C_{NT}^{1-\gamma}}{1-\gamma} - \frac{N^{1+\phi}}{1+\phi}\right) + \beta\mathbb{E}\left[\theta V_i^o(s',0) + (1-\theta)V_i^{LD}(s')\right] \tag{3.15}$$

Where θ is the probability with which the country financial markets exclusion is terminated and $\chi(E_T)$ is the output cost of financial autarky which, for a constant parameter ψ, takes the form:

$$\chi(Y_T) = \begin{cases} Y_T - \bar{Y}_T, & \text{for } Y_T \geq \bar{Y} \\ 0, & \text{for } Y_T < \bar{Y}_T \end{cases}, \text{ where } \bar{Y}_T = \psi\mathbb{E}Y_T$$

If the country regains access to financial markets, it does so with zero outstanding liabilities. If the country has left the risk-sharing agreement but opted not to default on its obligations, then the value of the problem is

$$V_i^{LR}(s,B) = \max_{C_{T,i},C_{NT,i},N_i,B_i'} \frac{C_{T,i}^{1-\gamma}}{1-\gamma} + \alpha\left(\frac{C_{NT}^{1-\gamma}}{1-\gamma} - \frac{N^{1+\phi}}{1+\phi}\right) + \beta\mathbb{E}V_i^o(s',B_i') \tag{3.16}$$

The budget constraints in each case are below. Where the price of tradables is the numeraire. If the country leaves and repays then it has access to the non contingent one period bond as saving technology:

$$C_T^i(s) + P_{NT}^i(s)C_{NT}^i(s) + B_i \leq Y_T^i(s) + W^i(s)N^i(s) + \Pi^i(s) + B_i'Q(s,B_i') \tag{3.17}$$

We omit the production subsidy since it is rebated to households and cancels out with increased profits. The subsidy is such that output reaches its efficient level. One can think of a government taxing profits and subsidising production. This will exactly cancel out in the household budget constraint as profits are rebated. Where $Q(s, B_i')$ is the bond price set between the country outside the contract and the competitive lender. The bond price is given by

$$Q(s, B_i') = \frac{1}{r}\mathbb{E}_t(1 - D(s', B_i'))$$

Where $D(s', B_i')$ is the decision to default on debt in the next period.
If the country leaves and default on past liabilities then it has no saving technology and is subject to a per period output cost:

$$C_T^i(s) + P_{NT}^i(s)C_{NT}^i(s) \leq Y_T^i(s) - \chi(Y_T^i(s)) + W^i(s)N^i(s) + \Pi^i(s) \tag{3.18}$$

3.2.3 Currency Union

Next we study a nominal version of the model to evaluate how the to design a transfer system in a currency union. In order to account for the money in this economy we assume that non-tradeable producers are subject to staggered prices.

Price setting for non-tradeables

Producers of non-tradeable goods face a rigidity in their price setting decisions. We incorporate this by assuming that at the beginning of each period firms must make their pricing decision before the realisation of the tradeable goods endowments $Y_{T,i}$, and wages cannot be conditional on this realisation[3].
Given its price P_{NT}^{ij}, the producer of variety j satisfies demand C^{ij} by hiring N^{ij} workers and earns profits

$$\Pi^{ij} = (P_{NT}^{ij} - (1 + \tau_L^i)W^i)N^{ij} \tag{3.19}$$

which are distributed to households.
For convenience, define the labour wedge $\kappa(s)$ as

$$\kappa^i(s) = 1 - \frac{U_N^i(s)}{U_{NT}^i(s)} = 1 - {C_{NT}^i}^{\gamma+\phi}(s) \tag{3.20}$$

Such wedge arises due to the lack of state-contingent non-tradeables prices.

[3]Note that in our framework having independent *state-contingent labour taxes* will play the same role as having an independent monetary policy.

The non-tradeable good price is predetermined, hence producers maximize the expected profits across states. The optimal price setting rule is characterized in the following Lemma:

Lemma 3.1 (Optimal Price Setting)
The optimal price for non-tradeables producers is:

$$P_{NT} = \frac{\sum_s \pi(s|s_{-1})\epsilon(s)^{\frac{1-\gamma}{\gamma}} C_T(s)^{1-\gamma} W(s)}{\sum_s \pi(s|s_{-1})\epsilon(s)^{\frac{1-\gamma}{\gamma}} C_T(s)^{1-\gamma}}$$

Such pricing implies zero expected labor wedge

$$\sum_s \pi(s|s_{-1})\epsilon(s)^{\frac{1-\gamma}{\gamma}} C_T(s)^{1-\gamma}\kappa(s) = 0. \tag{3.21}$$

Proof. See Appendix 3.A. ■

Lemma 3.1 shows that, in the absence of uncertainty, the optimal price is equal to the wage. Recall that this result is obtained by levying a labor subsidy that undoes the monopoly markup. Similarly, in absence of uncertainty, the labor wedge is equal to zero. This result can be understood as the lack of state-contingency being inconsequential when the state is constant.
We model the currency union as a long term contract. Countries within the currency union face the same price for tradeable goods, so that:

$$P_T^1(s) = P_T^2(s) \tag{3.22}$$

The implicit assumption is that the fixed nominal exchange rate within the currency union is 1.
Countries remain in the contract as long as they do not choose to unpeg from the common currency, or default on the net payments specified by the contract.
In this case the Ramsey planner is subject to pricing frictions and different outside options.

Outside options

In each period, each country has the option of defaulting on its payments within the union, and choosing to unpeg from the common currency and regain control of its monetary policy. We assume that either defaulting or unpegging implies abandoning the common currency. Thus, when it is inside the contract, country i faces a choice over the actions $\{PR, UR, UD\}$, i.e. maintain the peg and repay transfer commitments, unpeg and honour payments, and unpeg and default. We assume that defaulting on payments triggers temporary exclusion from financial markets so that the country can no longer trade bonds.

We also assume that the cost of unpegging from the common currency is that the country can only trade non-state contingent bonds B, limiting its consumption smoothing ability. The repayment commitments which remain are still denominated in the common currency. Unpegging is also an irreversible decision.

We can write the decision problem of each country outside the contract in a recursive form. Suppose that country i has already both defaulted on its debt and unpegged from the common currency. Its value function $V_i^{UD}(s)$ can then be written as:

$$V_i^{UD}(s) = \max_{C_{NT,i},N_i} \frac{(Y_{T,i} - \chi(Y_{T,i}))^{1-\gamma}}{1-\gamma} + \alpha\left(\frac{C_{NT}^{1-\gamma}}{1-\gamma} - \frac{N^{1+\phi}}{1+\phi}\right) + \beta\mathbb{E}\left[\theta V_U^i(s',0) + (1-\theta)V_i^{UD}(s')\right] \tag{3.23}$$

Suppose now that country i has unpegged but not defaulted yet. Its value function $V_i^U(s)$ can be written as:

$$V_i^U(s,B) = \max_{P,U}\{V_i^{UR}(s,B), V_i^{UD}(s)\} \tag{3.24}$$

where V_{UR}, the value of maintaining repaying the contractual obligations once the country has unpegged, is given by

$$V_i^{UR}(s,B) = \max_{C_{T,i},C_{NT,i},N_i,B_i'} \frac{(C_{T,i})^{1-\gamma}}{1-\gamma} + \alpha\left(\frac{(C_{NT})^{1-\gamma}}{1-\gamma} - \frac{N^{1+\phi}}{1+\phi}\right) + \beta\mathbb{E}V_U^i(s',B') \tag{3.25}$$

Finally, the outside option of a country that is still inside the currency union contract is given by the option value of, just unpegging, or defaulting and unpegging at the same time:

$$V_i^o(s,B) = \max\{V_i^{UR}(s,B), V_i^{UD}(s)\} \tag{3.26}$$

The budget constraints in each case are below. In each case the constraint is written in the common currency of the union (Euros). If the country unpegs from the common currency, $\epsilon^i(s)$ is the number of units of i's currency per Euro. Note that the price of tradables is always in Euros.

Unpeg without defaulting:

$$P_T^i(s)C_T^i(s) + \frac{P_{NT}^i(s)C_{NT}^i(s)}{\epsilon^i(s)} + B_i' \leq P_T^i(s)Y_T^i(s) + \frac{W^i(s)N^i(s)}{\epsilon^i(s)} + \frac{\Pi^i(s)}{\epsilon^i(s)} + B_i'\frac{Q(s, B_i')}{\epsilon^i(s)} \tag{3.27}$$

Unpeg and default:

$$P_T^i(s)C_T^i(s) + \frac{P_{NT}^i(s)C_{NT}^i(s)}{\epsilon^i(s)} \leq P_T^i(s)(Y_T^i(s) - \chi(Y_T^i(s))) + \frac{W^i(s)N^i(s)}{\epsilon^i(s)} + \frac{\Pi^i(s)}{\epsilon^i(s)} \tag{3.28}$$

Monetary Policy

We follow Farhi and Werning (2017) and Auclert and Rognlie (2014) in the definition of monetary policy. Within the currency union, due to the underlying price rigidity, demand externalities arise, generating a wedge between the private and social value of risk-sharing. Monetary policy optimally sets the union wide weighted wedge to zero.
Outside the monetary union, monetary policy is independent and country specific wedges are optimally equal to zero. This implies a relatively increased value of the outside option, compared to the fiscal union setup due to the independent monetary policy outside the currency area.
By the intratemporal first order condition the labour wedge is equal to zero in absence of nominal rigidities (given the subsidy to production). In presence of nominal rigidities the price of non tradeables is not equal to the wage implying that $\kappa^i(s) \neq 0$. Independent monetary policy equates the country specific labour wedge to zero.
We assume that monetary policy inside the currency union equates the weighted average labour wedge to zero. The weights are symmetric and equal to 1/2.
The following Lemma characterizes optimal monetary policy.

Lemma 3.2 (Optimal Monetary Policy)
Optimal independent monetary policy implies

$$\kappa^i(s) = 0, \quad \forall s$$

Optimal monetary policy in a currency union implies

$$\sum_{i=1,2} {C_{NT}^i}^{1-\gamma} \kappa^i(s) = 0, \quad \forall s$$

Proof. See Appendix 3.A ■

This implies that the $\kappa^i(s) \neq 0$, for $i = 1, 2$, due to asymmetry of the shock process.

3.2.4 The Union

In this section we describe how to rewrite the problem as a saddle point. We characterize the setup for the currency union since it is more general. The problem for the fiscal union with two independent monetary authorities is identical up to the presence of pricing frictions. We model the currency union with optimal transfers as a long term contract. This contract is subject to two sided limited commitment, whereby both countries can renege on the contract and switch to one of the outside options. The optimal contract is the solution to the following problem:

$$\max_{\{C_{T,i}(s^t), C_{NT,i}(s^t), N_i(s^t)\}_{i=1,2}} \sum_{i=1,2} \mu_{i,0} E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_{T,i}(s^t)^{1-\gamma}}{1-\gamma} + \alpha \left(\frac{C_{NT,i}(s^t)^{1-\gamma}}{1-\gamma} - \frac{N_i(s^t)^{1+\phi}}{1+\phi} \right) \right)$$

s.t.

$$\sum_{i=1,2} (P_T(s^t) C_T^i(s^t) + P_{NT,i} C_{NT}^i(s^t)) \leq \sum_{i=1,2} (P_T(s) Y_T^i(s) + W_i(s) N_i(s) + \Pi^i(s)) \tag{3.29}$$

$$\sum_{i=1,2} C_T^i(s^t) = \sum_{i=1,2} Y_T^i(s^t) \tag{3.30}$$

$$\mathbb{E}_t \sum_{r=t}^{\infty} \beta^r \left(\frac{C_{T,i}(s^r)^{1-\gamma}}{1-\gamma} + \alpha \left(\frac{C_{NT,i}(s^r)^{1-\gamma}}{1-\gamma} - \frac{N_i(s^r)^{1+\phi}}{1+\phi} \right) \right) \geq V_i^o(s_t, B) \tag{3.31}$$

It is known from Marcet and Marimon (2019) that this problem can be rewritten as the saddle point problem:

$$\mathcal{SP} \min_{\{\lambda_{i,t}\}_{i=1,2}} \max_{\{C_{T,i}, C_{NT,i}, N_i\}_{i=1,2}} \sum_{i=1,2} E_0 \sum_{t=0}^{\infty} \beta^t \Bigg[\mu_{i,t} \left(\frac{C_{T,i}(s^t)^{1-\gamma}}{1-\gamma} + \alpha \left(\frac{C_{NT,i}(s^t)^{1-\gamma}}{1-\gamma} - \frac{N_i(s^t)^{1+\phi}}{1+\phi} \right) \right) + \lambda_{i,t} \left(\frac{C_{T,i}(s^t)^{1-\gamma}}{1-\gamma} + \alpha \left(\frac{C_{NT,i}(s^t)^{1-\gamma}}{1-\gamma} - \frac{N_i(s^t)^{1+\phi}}{1+\phi} \right) - V_i^o(s_t, B_t) \right) \Bigg] \tag{3.32}$$

$$\mu_{i,t+1} = \mu_{i,t} + \lambda_{i,t} \tag{3.33}$$

Here $\lambda_{i,t}$ is the Lagrange multiplier of country i's participation constraint (PC). We now also have a *co-state* variable $\mu_{i,t}$ which effectively keeps track of the cost of keeping each agent inside the contract. We can then make use of a normalization which will reduce the dimension of the state space in the final problem. First we define the *relative weight* z_t of country 1 as

$$z_t = \frac{\mu_{1,t}}{\mu_{2,t}} \tag{3.34}$$

Then we rescale each country's Lagrange multiplier as follows:

$$\nu_{i,t} = \frac{\gamma_{i,t}}{\mu_{i,t}} \tag{3.35}$$

We can now derive a new equation of motion for the relative weight z_{t+1}:

$$z_{t+1} = z_t \frac{1 + \nu_{1,t}}{1 + \nu_{2,t}} \tag{3.36}$$

After this normalization, the state/co-state vector is (s, z) and the **saddle point Bellman equation** can be written as

$$\begin{aligned}\Omega(s,z) = \mathcal{SP} \min_{\{\nu_i\}_{i=1,2}} \max_{\{C_{T,i},C_{NT,i},N_i\}_{i=1,2}} z\left((1+\nu_1)(\frac{C_{T,1}(s,z)^{1-\gamma}}{1-\gamma} + \alpha\left(\frac{C_{NT,1}(s,z)^{1-\gamma}}{1-\gamma} - \frac{N_1^{1+\phi}}{1+\phi}\right) - \nu_1 V_1^o(s,B)\right) \\ +(1+\nu_2)\left(\frac{C_{T,1}(s,z)^{1-\gamma}}{1-\gamma} + \alpha(\frac{C_{NT,2}(s,z)^{1-\gamma}}{1-\gamma} - \frac{N_2(s,z)^{1+\phi}}{1+\phi}\right) - \nu_2 V_2^o(s,B) + (1+\nu_2)\beta E\Omega(s',z')\end{aligned} \tag{3.37}$$

$$z' = z\frac{1+\nu_1}{1+\nu_2} \tag{3.38}$$

$$\sum_{i=1,2}(P_T(s,z)C_T^i(s,z)+P_{NT,i}C_{NT}^i(s,z)) \leq \sum_{i=1,2}(P_T(s,z)Y_T^i(s,z)+W_i(s,z)N_i(s,z)+\Pi^i(s,z)) \tag{3.39}$$

$$\sum_{i=1,2} C_T^i(s^t) = \sum_{i=1,2} Y_T^i(s^t) \tag{3.40}$$

The policies in the union are given by the first order conditions of this problem. For tradeable goods consumption these are:

$$\frac{z(1+\nu_1)}{C_{T,1}(s,z)^\gamma} = \zeta(s,z)P_T(s,z) \tag{3.41}$$

$$\frac{1+\nu_2}{C_{T,2}(s,z)^\gamma} = \zeta(s,z)P_T(s,z) \tag{3.42}$$

Where $\zeta(s)$ is the multiplier on the resource constraint. From this we can derive the relative tradeables consumption of the two countries as:

$$\frac{C_{T,1}}{C_{T,2}} = \left(\frac{z(1+\nu_1)}{1+\nu_2}\right)^{\frac{1}{\gamma}} = (z')^{\frac{1}{\gamma}} \tag{3.43}$$

It follows that each country's consumption of the tradeable good is:

$$C_{T,1}(s,z) = \frac{(z')^{\frac{1}{\gamma}}}{1+(z')^{\frac{1}{\gamma}}} \sum_{i=1,2} Y_T^i(s^t) \tag{3.44}$$

and

$$C_{T,2}(s,z) = \frac{1}{1+(z')^{\frac{1}{\gamma}}} \sum_{i=1,2} Y_T^i(s^t) \tag{3.45}$$

The conditions for the non-tradeables consumption and labour supply of country i are then:

$$C_{NT,i}(s,z) = \left(\alpha \frac{P_T(s,z)}{P_{NT,i}(s,z)}\right)^{\frac{1}{\gamma}} C_{T,i}(s,z) \tag{3.46}$$

and

$$N_i = C_{NT,i}(s,z) \tag{3.47}$$

Furthermore the Union's value function takes the form:

$$\Omega^U(s,z) = zV_1^U(s,z) + V_2^U(s,z) \tag{3.48}$$

for $U = F, M$, depending on whether it refers to a Fiscal Union with two independent monetary authorities or to a Monetary Union[4].

Decentralization with Endogenous Borrowing Limits

We now show how the contract allocation can be decentralized as a competitive equilibrium with trading of state contingent debt contracts and borrowing constraints.
We will be interested in union allocations for which the present value, at the correctly defined prices, is finite. We say that an allocation has *high implied interest rates* if

$$E_0 \sum_{t=0}^{\infty} q(s^t, z_t \mid s_0, z_0)(Y_{1,t} + Y_{2,t}) < \infty \tag{3.49}$$

where

$$q(s_{t+1}, z_{t+1} \mid s_t, z_t) = \max_i \beta \left(\frac{C_{T,i}(s_{t+1}, z_{t+1})}{C_{T,i}(s_t, z_t)}\right)^{-\gamma} \tag{3.50}$$

and $q(s^{t+k}, z_{t+k} \mid s_t, z_t) = \prod_{n=0}^{k-1} q(s_{t+n+1}, z_{t+n+1} \mid s_{t+n}, z_{t+n})$.

Country Problem Each country i has access to a one period state contingent debt contract $B_i(s) = \{b_i(s' \mid s)\}_{s'}$, which denotes the amount of the tradeable good which country i promises to deliver in the state s'. In addition, let the price of a unit of an Arrow security which pays in state s' be $q(s' \mid s)$. Then the value of the debt contract is $\sum_{s'} q(s' \mid s) b_i(s' \mid s)$. Country i solves the following problem:

[4] In $\Omega^M(s,z)$, s denotes (s_{-1}, s).

$$\omega(b_i, s) = \max_{\{C_T, C_{NT}, N, B_i(s)\}} \frac{C_{T,i}^{1-\gamma}}{1-\gamma} + \alpha \left(\frac{C_{NT,i}^{1-\gamma}}{1-\gamma} - \frac{N_i^{1+\phi}}{1+\phi} \right) + \beta E[\omega(b_i', s') \mid s]$$

subject to

$$\begin{aligned} C_T^i(s) + P_{NT,i} C_{NT}^i(s) + b_i(s) \leq \\ Y_T^i(s) + W_i(s) N_i(s) + \Pi_i(s) + \sum_{s' \mid s} q(s' \mid s) b_i(s' \mid s)) \end{aligned} \tag{3.51}$$

and

$$b_i(s' \mid s) \leq \bar{B}_i(s') \tag{3.52}$$

Where $\bar{B}(s')$ is a state contingent **endogenous borrowing limit**.

Definition 3.1 (Equilibrium)
A competitive equilibrium with borrowing limits is a collection of borrowing limits $\{\bar{B}(s)\}$ and initial debt positions $\{b_i(s_0)\}$, together with an allocation $\{C_{T,i}(s), C_{NT,i}(s), N_i(s)\}$, state contingent debt contracts $\{B_i'(s)\}$, goods prices $\{P_T(s), P_{NT}(s), W_i(s)\}$ and asset prices $q(s' \mid s)$ such that $\{C_{T,i}(s), C_{NT,i}(s), N_i(s)\}$ solves country i's decision problem, markets clear and the resource constraint holds.

The consumption and asset choice decisions give us the Euler equation:

$$q(s' \mid s) \geq \beta \pi(s' \mid s) \left(\frac{C_T^i(s', b')}{C_T^i(s, b)} \right)^{-\gamma} \tag{3.53}$$

A competitive equilibrium with borrowing limits therefore satisfies this equation and the transversality conditions:

$$\lim_{t \to \infty} E_t \beta^t q(s^{t+1} \mid s_t) C_T^i(s_t, b_i(s_t))^{-\gamma} b_i(s_{t+1}) = 0 \tag{3.54}$$

Proposition 3.1 (Decentralized Equilibrium)
Any union allocation with high implied interest rates can be decentralized as a competitive equilibrium with endogenous borrowing limits.

Proof. See Appendix 3.A ■

With this implementation of the union allocations, we are now able to specify the liabilities generated by each country's participation in the union. In any given state, these same liability levels would also need to be financed outside of the union if one of the countries chose to exit (which, in equilibrium, never happens). Given that in the outside option, each country can decide to default completely on its debt, an obvious question is whether there is any case in which a participation constraint binds and the constrained country's preferred outside option is to continue repaying its debts. In the full currency union, this involves a complex comparison of the value of independent monetary policy with the value of enhanced risk-sharing in the contract. In the fiscal union, however, where there is no nominal friction, we are able to show that there is no case in which a country is indifferent between remaining in the union and the alternative of leaving and continuing to repay its debt.

Proposition 3.2 (Optimal Exit in Fiscal Unions)
In the fiscal union with two independent monetary authorities, whenever the participation constraint is binding for country i, $V_i^{LD}(s) > V_i^{LR}(s, B)$.

Proof. See Appendix 3.A ■

The next proposition formalises a different aspect of the two problems: if the currency union is able to achieve full risk-sharing, then it will attain the same value as the fiscal union.

Proposition 3.3 (Risk-sharing Miracle)
If in the steady state the currency union attains full risk-sharing, i.e. $(C_T^1(s)/C_T^2(s))^{-\gamma} = \bar{c}$, $\forall s$, *then the common monetary policy is able to stabilize both economies at once.*

Proof. See Appendix 3.A ■

Similarly to Auclert and Rognlie (2014) when countries attain full risk-sharing, stabilizing one economy through common monetary policy also puts the other country's labour wedge to zero. This result carries important consequences for the type of steady state that may arise in this model, and, in particular, for the comparison between fiscal and currency unions in full risk-sharing steady states.
The following definition describes the two types of steady state which can emerge in the monetary and fiscal unions.

Definition 3.2 (Steady States)

a) A steady state with perfect risk-sharing is a path in which for some k*, the relative weight* z_t *is constant for all* $t > k$*.*

b) A steady state with imperfect risk-sharing is a path in which for some k, the relative weight $z_t \in \{\underline{z}, \dots, \bar{z}\}$ (i.e. it is in the discrete support of the ergodic distribution), for all $t > k$, where $\bar{z} > \underline{z}$, $\underline{z} = \min_{s \in S \times S}\{z : V_1^U(s,z) = V_1^o(s,B)\}$ and $\bar{z} = \max_{s \in S \times S}\{z : V_2^U(s,z) = V_2^o(s,B)\}$.

Corollary 2 (Values in Full Risk-sharing Steady States)**.** *In a constant weight steady state*

$$V_i^M(s,z) = V_i^F(s,z), \;\; i = 1,2.$$

Proof. See Appendix 3.A ■

In this economy full risk-sharing always characterizes the constrained efficient allocation. The implication of Corollary 2 is that if such allocation can be attained, then the currency union delivers the same level of utility as the fiscal union. This is a direct consequence of the risk-sharing miracle.

Another equivalence result can be obtained by focusing on periods in which any participation constraint binds in an imperfect risk-sharing steady steady state.

Proposition 3.4 (Values with Binding Constraints)
If the optimal choice in the outside option economy is to leave and default on outstanding liabilities, then, whenever the participation constraint binds for country i,

$$V_i^F(s,z) = V_i^M(s,z) = V_i^o(s,B) = V_{UD}^i(s).$$

Proof. See Appendix 3.A ■

Theorem 1 (Steady States with Imperfect Risk-sharing)**.** *In steady states with imperfect risk-sharing*

$$V^F(s,z) > V^M(s,z).$$

Proof. See Appendix 3.A

■

Proposition 3.5 (Constrained Efficient Currency Unions)
The optimal allocation in a currency union is constrained efficient.

Proof. See Appendix 3.A

■

Corollary 3 (Monetary Policy with Planner Weights). *If the central bank of a currency area adopts the relative Pareto weights of the planner then risk sharing decreases. This is paired with an increased inefficiency of the non-tradeables consumption.*

Proof. See Appendix 3.A

■

In the next section we solve the model numerically to study whether the contracts are feasible (positive surplus), what is the structure of the optimal transfers and whether the optimal policies in a currency union can make up for the deadweight loss due to the lack of independent monetary policy.

3.3 Quantitative Results

In this section we describe the algorithm used to solve the model, the parameterisation and the numerical results for both the real and the nominal setup.

3.3.1 Solution Algorithm and Parameters

The solution algorithm first requires solving for the value functions and policy functions of the outside option, which is an Arellano economy with two goods. The Arellano economy is solved by value function iteration, following the algorithm in Arellano (2008), adjusted to allow updating of the bond pricing schedule in each iteration. This gives us the consumption of tradeables and the borrowing choices in terms of tradeables. Since monetary policy is independent in the outside option, so that non-tradeable production is always at the first best level, we simply set $C_{NT} = 1$ and $N = 1$ for all states; this is also true for the non-tradeable allocation in the fiscal union contract, where there is no nominal rigidity.
The contract allocations are solved for using policy function iteration. We start with an initial guess for the value functions of the contract, and the liabilities. At each iteration k, for a given assignment of liabilities $B_k(y, z)$ and a guess for the value functions $V_k(y, z)$, we find the value of relative weight z at which the participation constraint binds in each endowment state y. Using the symmetry of the environment, we can then calculate an interval $(\underline{z}(y), \bar{z}(y))$ within which the participation constraints do not bind; outside of this range, the allocations are constant due to the binding participation constraints.
Once we have updated the allocation, we can update the implied liabilities using a recursive version of the budget constraint, as in Equation 3.77. This allows us to update the assignment of the outside option values $V_i^o(y, B(y, z))$; any values of $B(y, z)$ which lie outside the grid used to solve the Arellano economy are calculated using cubic spline interpolation (or extrapolation

Description	Parameter	Value
Openness parameter	α	1
Discount Factor	β	0.95
Utility Curvature	γ	2
Labour Elasticity	ϕ	3
Risk Free Rate	r	1.02
Reinclusion Probability	θ	0.17
Default Output Cost	ψ	0.96
Endowment AR1 parameter	ρ	0.9
Endowment Shock Variance	σ_y^2	0.01

Table 3.1: Baseline parameter Values

if required). We then continue onto the next iteration $k+1$, by again finding the binding values of z. We continue iterating until the changes in the value function for the contract and the liabilities function $B(y,z)$ are sufficiently small.

This completes the solution algorithm for the fiscal union. For the currency union, there are two additional steps in each iteration, to calculate the relative prices $\frac{P_{1,NT}}{P_{2,NT}}$ and $\frac{\epsilon(s)}{P_{1,NT}}$. The exact expressions needed for calculating these relative prices can be found in Auclert and Rognlie (2014), although we have adjusted them to allow for a more flexible specification of risk aversion. Both of these prices are required to calculate the non-tradeable allocation variables C_{NT} and N.

The parameter values used for all exercises are shown in Table 3.1. These values have not been calibrated but have been chosen to lie within ranges which are common in the macroeconomic literature. An exception in this regard is the relative risk aversion parameter γ, which at 5 would be considered high. Under our current solution algorithm, it becomes extremely difficult to achieve convergence of the currency union contract solution for a lower value of γ, without introducing aggregate risk into the model. For the sake of simplicity, we have chosen to retain the higher value of γ, rather than introduce an additional state variable into the model.

The Markov process for the tradeable endowment y is produced by discretizing an AR1 process with persistence and volatility parameters ρ and σ_y^2, as given in Table 3.1. We discretize the process using the Rouwenhorst method, which achieves better performance for near unit root persistence. For all of the economies we use a 5 state Markov chain. We report the full transition matrix in equation 3.82 in the Appendix.

3.3.2 The Outside Option

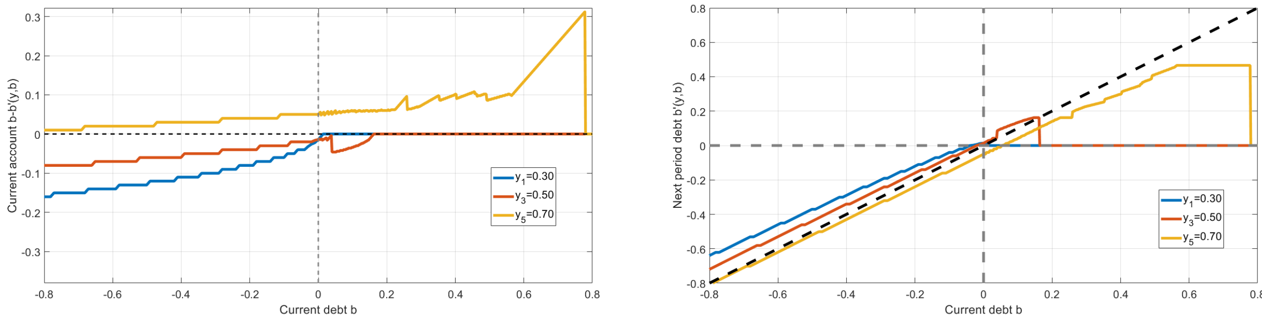
We begin our discussion of the results by first describing the behaviour of the defaultable debt economy which is the outside option to remaining in the union contract.

Figure 3.1 displays the behaviour of financial variables in the outside option economy. We plot the current account and stock of assets against different levels of current liabilities. Figure 3.1a shows the evolution of net borrowing. To the left of the zero on the horizontal axis countries have stocks of assets. The optimal policy here suggests that countries run a current account deficit when they are in the lowest endowment realization to smooth consumption. When they are in the highest endowment realization, countries run a current account surplus and increase the stock of assets. Moving rightward, countries have positive debt. Since default is costly and priced in by the lender, countries tend to run current account surpluses to deleverage and reduce the cost of borrowing. The steep declines in debt correspond to default episodes. The general deleveraging pattern is evident in Figure 3.1b since the lines slopes are less than 1, meaning that tomorrow's debt is lower than the current liability stock.

In Section 3.2.3 we outlined the full set of choices available to each country when considering whether or not to remain in the fiscal or currency union. Figure 3.1 also gives some information about the preference ordering of these outside options. Suppose the country is in the 3rd highest endowment state $y_3 = 0.5$, represented by the red lines in Figure 3.1. In Figure 3.1b, we see that for a current debt level below 0.135, the debt choice B' is non-zero (with an exception around $B = 0.02$ where the country optimally chooses to deleverage slightly to $B' = 0$), meaning that for these debt levels the country continues to participate in financial markets. However, for any current debt level above 0.135, the country's debt in the next period collapses to zero because it defaults. Using the notation of Section 3.2.3, this tells us that when $y = 0.5$ and $B \leq 0.135$, $V_i^{LR}(s, B) \geq V_i^{LD}(s, B)$, so the country chooses to continue servicing its debt; conversely, when $y = 0.5$ and $B > 0.135$, $V_i^{LR}(s, B) < V_i^{LD}(s, B)$, and so the country chooses to default on its existing liabilities.

Furthermore, these preference orderings tell us about the off-equilibrium behaviour of each country in the case that it decides to leave the union (recall that in equilibrium this will never happen because the contracts are designed so that the participation constraints are always satisfied). Consider again the case where country i's current endowment is $y_3 = 0.5$, but now assume that is inside the fiscal union. If its current liabilities inside the union are 0.1, for example, then it considers the choice between remaining in the fiscal union, and leaving the union but continuing to service the liabilities which it has accumulated. On the other hand, if it has liabilities of 0.2 (or any amount greater than 0.135), then it instead considers the choice between remaining inside the fiscal union and leaving the union and immediately defaulting on these liabilities.

In the results which follow, we will see that in the former case, where liabilities are relatively low, the country always *strictly* prefers to remain in the union. This result holds *a fortiori* for the case where the country has accumulated assets within the union. Importantly, we also find that whenever the participation constraint binds, so that the country is indifferent between



(a) Current Accounts Outside Option Economy (b) Debt Law of Motion Outside Option Economy

Figure 3.1: Outside Option Economy policis

staying in the union and leaving, its liabilities inside the union are always so large that if it were to leave the union, it would immediately default. These findings hold for both the fiscal and the currency union.

3.3.3 Fiscal Union

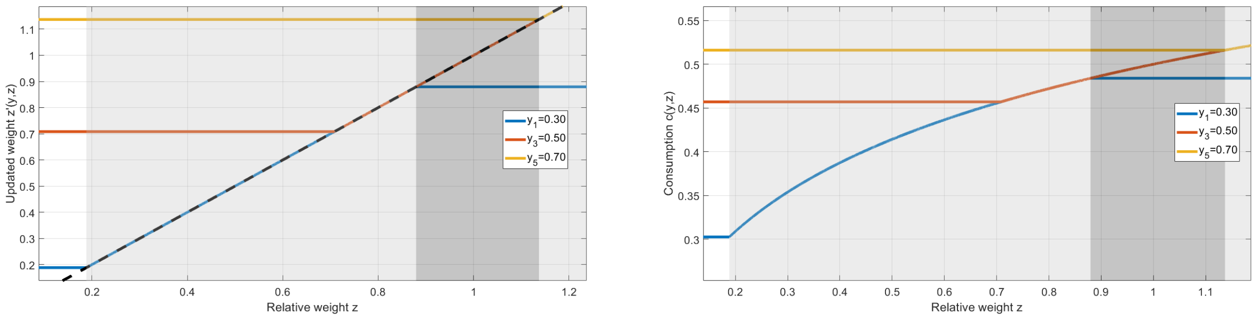
We start the description of the quantitative results of the fiscal union model by characterizing the optimal policies inside the risk-sharing contract.

In what follows we plot the policy functions inside the dynamic contract as a function of the relative weight z. We plot all policies for different endowment realizations[5].

Figure 3.2 shows the relative weights and consumption policies. The dark grey shaded area represents the set of weights that characterize the ergodic distribution of the contract. This is the set in which the weights will lie and fluctuate in the steady state. The lighter grey shaded area represents the basins of attraction of the ergodic set. All graphs contain the lowest, the median and the highest realization of the tradeable endowment.

Figure 3.2a shows the optimal relative weight policy. Every line corresponds to a specific realization of the endowment. In every line flat regions represent areas in which one of the participation constraints is binding. The flat region to the left is where the participation constraint of country 1 is binding, the flat region on the right shows where the participation constraint of country 2 binds. Since the relative weight describes the consumption allocation of country 1 relative to country 2, in the left area of the graph the relative weight is too low, meaning that country 1 is receiving too little consumption, hence the country is against its participation constraint. Conversely, as one moves rightward, there is a flat portion of the line

[5]Recall that in this setting there is no aggregate risk, meaning that when country 1 is in a high endowment state, country 2 is in a low endowment one.



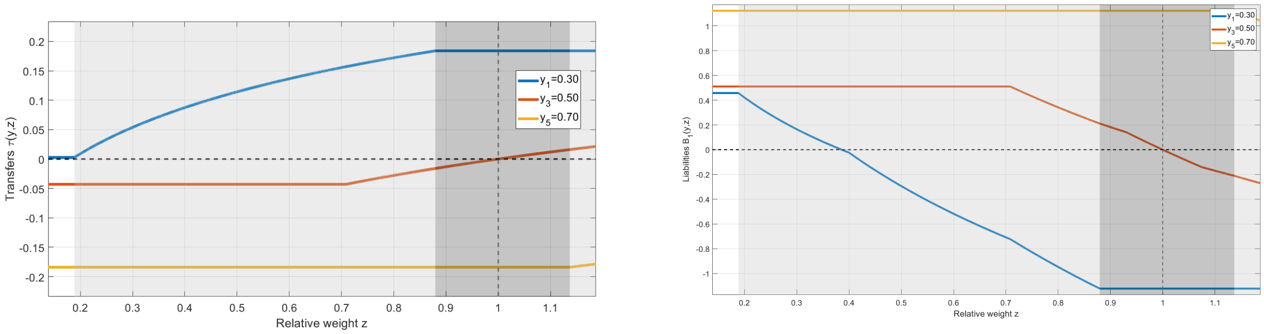
(a) z' (b) c

Figure 3.2: Outside Option Economy policies

where the relative weight is too high and country 2's participation constraint is binding. The regions in which the optimal weight coincide with the 45 degree line are areas where neither participation constraint binds, hence the weight is constant across periods.

Figure 3.2b displays the consumption allocation. When neither country wants to leave the contract the future relative weight is equal to the ratio of marginal utility of tradeable consumption between the two countries. Hence the graph shows that consumption tracks the current relative weight in the same areas where the weight is not updated. Concavity is inherited by preferences.

Figure 3.2 already shows that this contract features an imperfect risk-sharing steady state. Using Definition 3.2 it is visible that as there is no set of weights in which neither the particiation constraint of the high endowment country nor the one of the low is binding, the weights must fluctuate as the state changes. Graphically this can be seen by observing, in Figure 3.2a, that the flat region to the right on the lowest endowment relative weight lies to the left of the region where the PC stops binding for the high endowment. This steady state will then not be able to attain full risk-sharing. In particular in the steady state the following path will occur: suppose we start with both countries at the middle endowment and a relative weight of 1. At this level of weight, given the state, neither participation constrain binds. Suppose now that country 1 moves to the highest endowment level (hence country 2 moves to the lowest). At a relative weight of 1, in the highest endowment state, country 1's participation constraint binds (this can be seen on the yellow line). The planner will then increase the relative weight next period to the minimum level to make country 1 indifferent between the contract and the outside option. Such weight is rightmost point on the dark grey area, namely where the PC is barely binding in the highest endowment state. As long as the states do not change both countries' PCs are slack. Suppose now that the state changes and country 1 moves to the lowest endowment state (hence country 2 moves to the highest). At



(a) τ

(b) b

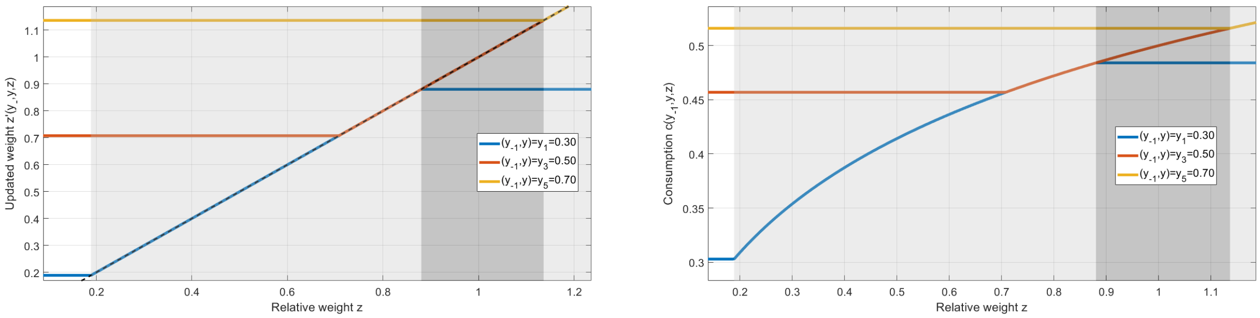
Figure 3.3: Fiscal Union policies

these relative weights country 2's participation constraint binds and the planner will increase the weight till the PC is slack again. These dynamics define the imperfect risk-sharing steady state.

The next two graphs in Figure 3.3 show the key policies inside the contract. Figure 3.3a displays the optimal transfer policy. The contract features optimally large countercyclical transfers between the countries. When countries are at symmetric endowment realizations and neither participation constraint binds, the transfers range between -18% and 18% of the total tradeable endowment. They are as large as 2/3 of the endowment for lower realizations.

Figure 3.3b shows the liabilities positions. Countries have higher stocks of debt whenever they have a low relative weight. One feature of the contract is that, since the countries are symmetric and given the persistence of the endowment, a high realization today implies future surpluses in expected terms. This, in turn, generates positive stocks of liabilities today. This feature however is not true for any level of the relative Pareto weight. This result is common to other similar models of dynamic contracts (see Abraham et al., 2019). The key difference is our setting is that the debt position can take both signs (i.e. assets or liabilities) for both countries. This feature stems from the symmetry of risk aversion and impatience in our model. This result can also be interpreted as countries with better endowment realizations being able to absorb larger stocks of debt.

Finally we discuss the steady states of this economy. The fiscal union features an imperfect risk-sharing steady states. The dark shaded area represents the ergodic set. This range of relative weights has the property that if an economy starts (say, in Period 0) with a relative weight inside this set it will always stay there. This set has a basin of attraction, both from the left and from the right, such that if an economy starts inside the basin it will eventually converge to the ergodic set. This basin of attraction is represented by the light grey areas in the graphs.



(a) z' (b) c

Figure 3.4: Monetary Union policies

3.3.4 Monetary Union

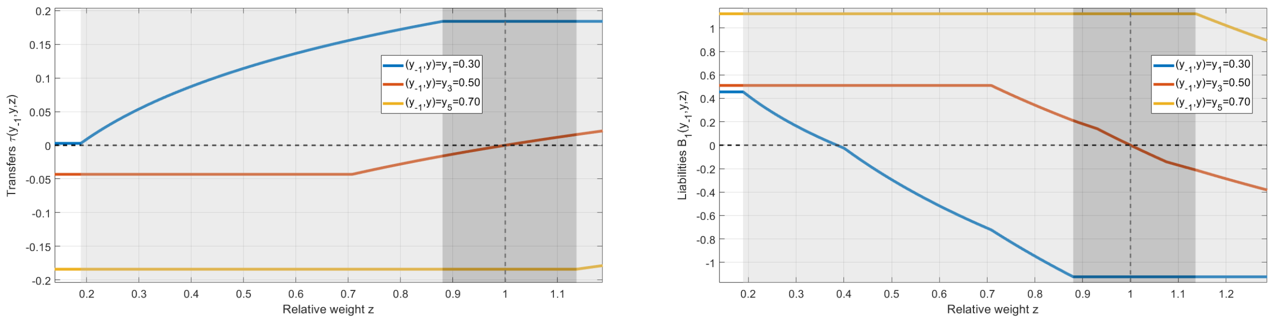
In this section we describe the results for the currency union model. In this setup non-tradeables producers face staggered prices friction. In the outside option economy countries have independent monetary policy and close the labour wedge.

One important feature of this model is that the outside option value is identical to the one in the real version of the model since monetary policy eliminates nominal rigidities entirely. However, inside the contract, countries face lower surplus since the economy is not producing at the efficient level. This is a direct consequence of Corollary 2 and Theorem 1. From Theorem 1 the currency union can never yield a higher value than the fiscal union in imperfect risk-sharing steady states. From Corollary 2 they can be at most equal in constant weights steady states. As the fiscal union features imperfect risk-sharing it is never possible for the currency union to attain the same value of the problem.

Surprisingly, the contract is qualitatively identical to the one described in the previous section. In Figure 3.4 we plot the future relative weight and consumption as a function of the current relative weight z.

Figure 3.4a displays the law of motion of the relative weight. All lines feature a flat region on the left where the country's participation constraint binds, a sloped part where it coincides with the 45 degree line and flat region on the right where the other country's participation constraint is binding. The weights are updated upward whenever the country's PC binds, downward when the other country threatens to leave the contract and they remain constant when neither is against the outside option. At first inspection, the path in the ergodic set resembles the one in the fiscal union. This feature will be extensively discussed later in the paper.

Figure 3.4b plots the consumption allocations for different levels of endowments. Consumption closely tracks the relative weight behaviour, as in the fiscal union setting.



(a) τ (b) b

Figure 3.5: Monetary Union policies

The economy features countercyclical optimal transfer of the tradeable endowment. Their size is numerical identical to the ones in the fiscal union. The same holds for the stock of liabilities, displayed in Figure 3.5b.

The outside option economy behaves identically to the real model outside the fiscal union. Hence the behaviour of current accounts and the debt law of motion can be seen in Figure 3.1. It is important to notice that while the optimal policies in the two economies representing the outside options of the contract are the same the starting levels outside are not. To see this, recall that the stock of liabilities of a country leaving the union is given by the outstanding set of promises to the other country. As the transfer policies inside the fiscal and currency unions could be different, so would be the liabilities inside the contract. Hence the starting stock of debt upon breakup could be different. In other words, conditional on a given level of b the policy for b' is the same in the two economies. However, in the same state inside the contract, upon leaving, the countries could start with different levels of b.

Exactly as in the fiscal union, this economy features an imperfect risk-sharing steady state. Hence the planner is unable to attain full risk-sharing.

3.3.5 Comparison of the Contracts

In this section we compare the optimal policies in the two contracts. From the previous discussion, we see that the two contracts seem to have very similar values and policies despite the fact that the currency union cannot achieve the optimal allocation of non-tradeable goods. In Section 2, we showed formally that in some special cases (for example, when the participation constraint is binding), the currency union attains the same value as the fiscal union; the numerical results, however, seem to suggest that the similarity is more general. How is this possible? The answer lies in a difference in the behaviour of the optimal transfers in the currency union which enables the planner to compensate partially for the labour wedge.

We start by recalling that the full state space for the economy is (y_{t-1}, y_t, z_t). We can then distinguish between two cases. In the first, between period $t-1$ and t, the endowment of tradeables goods does not change, i.e. $y_{t-1} = y_t$. In the second, the endowment does change between periods so that $y_{t-1} \neq y_t$. If the endowment is quite persistent, as we tend to assume, then in period $t-1$ agents' expectations will place a large probability mass on the first case, in which the endowment does not change. In particular, the pricing decisions of the non-tradeable good producers will put a large weight on this outcome. As a result, if the endowment does not change between periods, the labour wedge in the currency union will be relatively small, whereas if does change, it will be larger; in fact, the larger the transition $y_{t-1} \to y_t$, the large the labour wedge in period t.

This explains why in the comparisons considered so far, where the realization of y is held constant, the currency union behaves similarly to the fiscal union. If instead we consider transitions where the endowment changes between periods, we see that the transfer policy in the currency union is more complex that that of the fiscal union.

Figure 3.6 shows consumption in the two contracts when $y_t = y_1$ and the country's participation constraint is binding. We see that in the monetary union, when the constraint binds, the level of consumption depends not only on y_t but also on y_{t-1}. Moreover, when the economy arrives at y_1 from a higher previous endowment, is receives high current tradeables consumption. This higher consumption compensates for the fact that when the transition is large (say $y_5 \to y_1$), the labour wedge is also large; the additional consumption is therefore needed to keep the country in the monetary union. In contrast, in the fiscal union, monetary policy completely eliminates the wedge; as a consequence consumption does not need to be conditioned on y_{t-1} in this way.

We should reiterate at this point that the extra adjustment of transfers in the monetary union is not enough to completely undo the deadweight loss from having joint monetary policy. We showed formally that the value of the fiscal union is always weakly higher than that of the currency union. However, under certain conditions, the transfer policy in currency union can make the overall loss very small.

Steady States

Before discuss the features of the steady states in detail, it is worth describing in more depth the set of weights defining the basin of attraction of the ergodic sets of the two contracts.

In Figures 3.16 and 3.17 we plot, for every endowment realization, the set of weights in which the participation constraints do not bind. The ergodic sets are defined by the upper bound of the lowest realization of output and the lower bound of the highest realization. This area is shaded in grey.

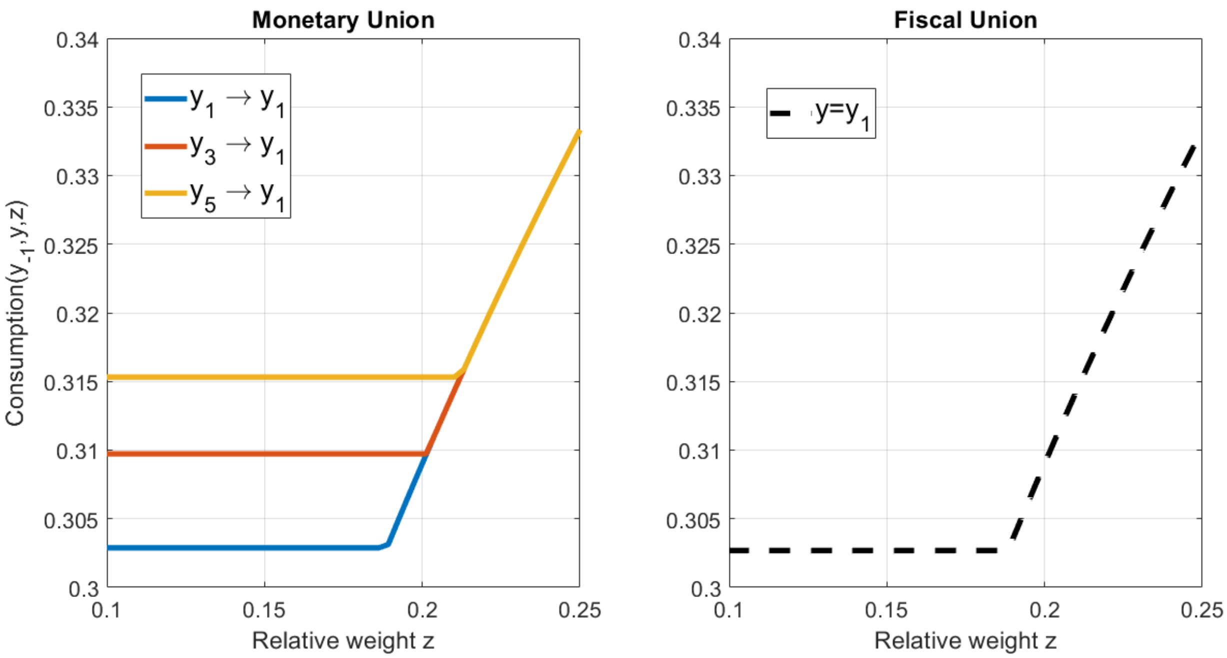
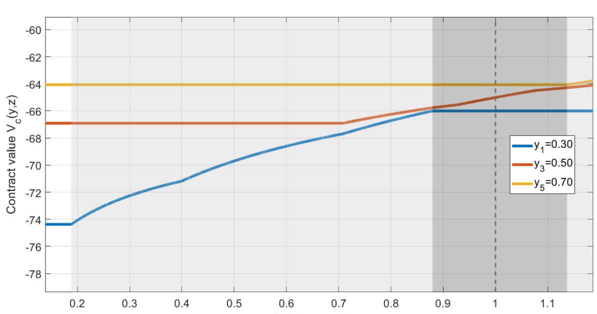
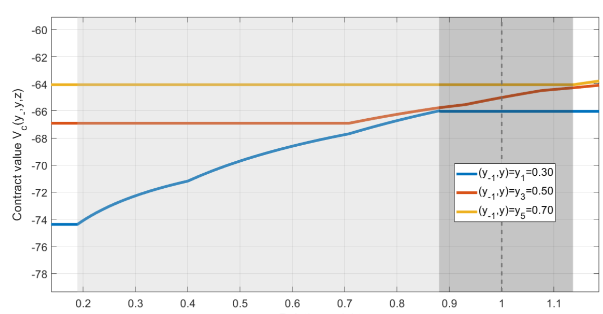


Figure 3.6: Consumption adjustment in the contracts



(a) Fiscal Union



(b) Monetary Union

Figure 3.7: Values of the Contracts

The reciprocal bounds, namely the lowerbound for the lowest endowment and the upperbound of the highest endowment, define the the basin of attraction. In order words one can read off the graphs the set of starting weights that will produce convergence to the ergodic set.

We start the discussion on the features of the stochastic steady states by providing one simulation to exemplify the dynamics. We start by simulating one history of endowments for 100 periods. We then plot the optimal policies of a country in the fiscal union, one in the currency union and one in the defaultable debt economy.

We start the contracts with a relative weight of 1 in the median endowment state. As $z = 1$ is the center of the ergodic set in both unions the economy will permanently remain in such set. We then sample 100 period and plot the simulation policies.

The path of the endowment and consumption is plotted in Figure 3.8. In the left panel, showing the endowment history, the red line shows the path for the contracts economies, while the black line for the defaultable debt one. The vertical black lines show episodes of default and financial market exclusion for the outside option economy. The endowment history is the same, though recall that when output is sufficiently large and the defaultable debt economy is in a period of exclusion from financial markets, it pays a fraction of endowment as a default cost. Hence the small deviations between the two paths when the outside option economy has default episodes. We denote periods of financial autarky, following a default, as a dot at the top of the graph, while periods of financial market access as dots at the bottom of the graph.

The left panel of Figure 3.8 shows the behaviour of the consumption of tradeables. The two contracts provide the same level of consumption (red and blue lines). The planner is able to smooth close to all of the fluctuations in the endowment. The jumps in consumption are given by updating in the relative weight, following a binding participation constraint for one of the two countries. Finally the defaultable debt economy shows high volatility in consumption as there is limited possibility to insure against the idiosyncratic risk.

Figure 3.9 shows the behaviour of the relative weight and some financial variables of these economies. The top left panel shows the behaviour of the relative weight, which is numerically identical for the two contracts. The transfer policy, in the top left panel, shows that the transfers are countercyclical and large, relatively to the endowment. Secondly it shows that, as the relative weight is stable in the first half of this history, consumption of tradeables does not change. This implies that the transfers absorb the entire difference between the constant consumption and the varying endowment. Together with the contract transfers we plot the current account balance of the outside option economy. The defaultable debt economy behaves very differently. At the beginning of this history the endowment realization is low and the economy has some assets. Once these assets are used to smooth consumption and the country starts accumulating debt, as the endowment drop, the country defaults. The country is excluded from financial markets for an extended period of time, hence the zero

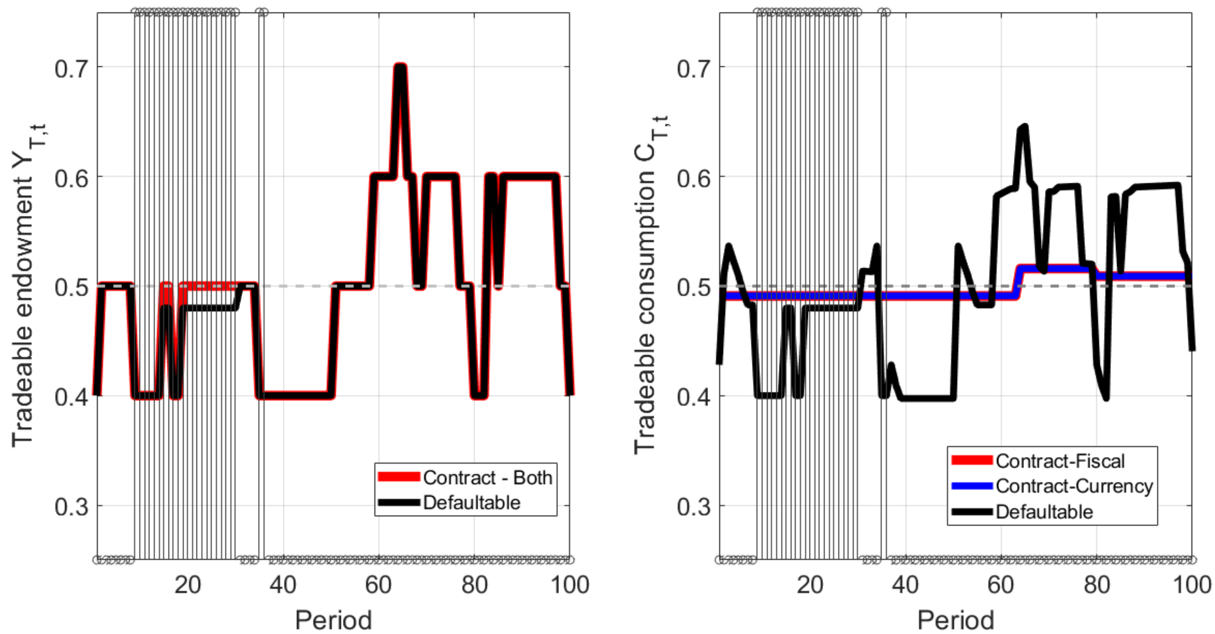


Figure 3.8: Steady State Endowment and Consumption

debt and current accounts. As the economy is reincluded in financial markets and borrows the endowment drop again and the country defaults again. Subsequently the country enjoys of sustainable borrowing and high endowment though it is still unable to absorb the large variations in the endowment and consumption is quite volatile. Lastly, the bottom right panel, shows the behaviour of the risk spread. In this graph is clear how the external lender prices in default before it happens by increasing the interest rate charge on the defaultable debt.

Table 3.2 shows some key moments of the economies in steady state. These moments are computed by averaging 50000 simulations in the steady state.

As expected the defaultable debt economy provides less consumption smoothing than the two contracts with a a consumption volatility 7 times higher. Secondly, consumption is lower in the outside option than in the contract due to default episodes in which the endowment is reduced. Thirdly, the two contracts deliver the same policies, which large (10% of GDP) countercyclical fiscal transfers and approximately the same values.

The risk-sharing value of the agreements is evidenced by the correlation between consumption and the endowment. The two contracts significantly reduce this comovement though such correlation is still positive. This stems from periods in which the weights change procyclically, for example when a country moves to high endowment and this makes the PC bind, implying an upward revision of its relative weight.

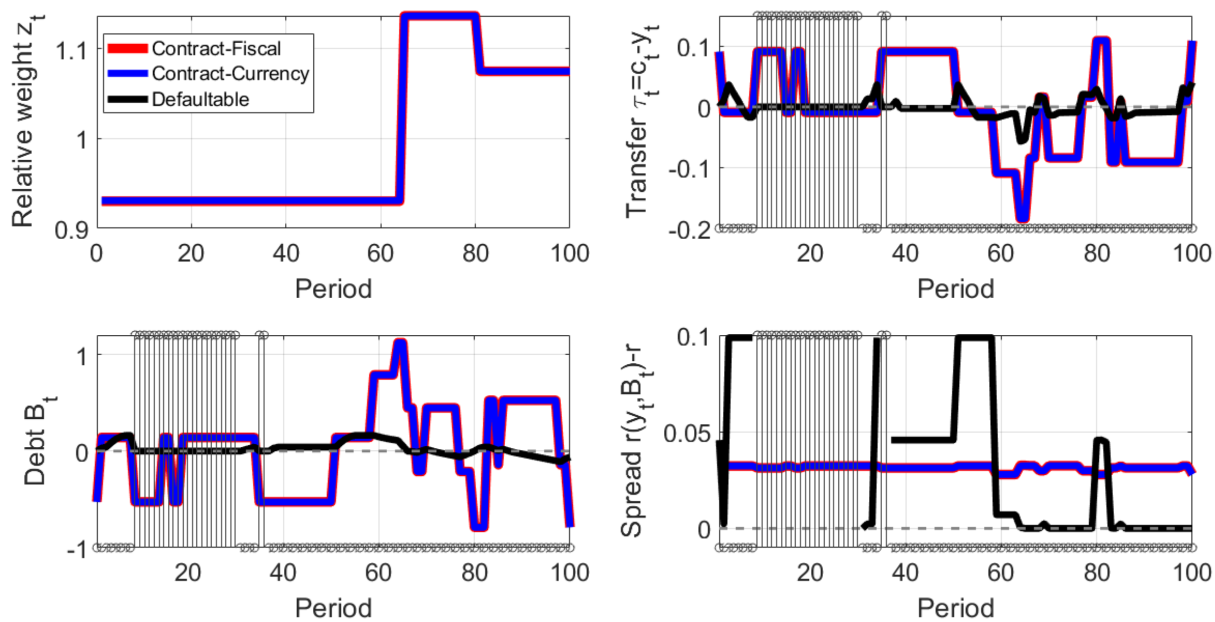


Figure 3.9: Steady State Optimal Policies

	Outside-Defaultable Debt	**Contract-Fiscal**	**Contract-Currency**
Mean			
Y_t	0.4975	0.4996	0.4996
$C_{T,t}$	0.4973	0.5	0.5
GDP_t	0.754	0.75	0.75
$\mid \tau_t \mid$	0.013	0.075	0.075
B_t	-0.001	-0.001	-0.002
z_t	-	1.005	1.005
$V(y, b/z)$	-66.832	-65.031	-65.032
$Pr(PCbinding)$	-	0.029	0.029
Standard deviation			
$\sigma(c_{T,t})$	0.094	0.013	0.013
$\sigma(Y_t)$	0.099	0.1	0.1
Correlation			
$\rho(C_{T,t}, Y_t)$	0.983	0.5	0.501
$\rho(\tau_t, Y_t)$	-0.379	-0.993	-0.993

Table 3.2: Steady State Moments

An important point of comparison between the fiscal and the currency union is in the row labelled $Pr(PCbdinding)$. This represents the fraction of periods in which any participation constraint is binding in this agreement. As discussed above, the fiscal planner in a currency union is implementing transfers that depend on y and y_{-1}. Particularly the planner is rewarding the country with the larger transition through higher tradeables consumption. This implies that the steady state path is fluctuates in narrower bands in the currency union. As such, there exists a set of pairs of endowment transitions in which a participation constraint would be binding in the fiscal union but it is not in the monetary union. This yields a lower probability of a binding PC in the currency union. In this case, however, we find that the difference is negligible numerically.

The higher risk-sharing capacity of the contracts is also visible in the maximum amount of liabilities that countries can have inside the agreements.

In the defaultable debt economy countries are unable to borrow due to the high likelihood of default. Inside the unions can accumulate liabilities. .

The graphs in Figure 3.10 carry one additional set of information. Looking at the red lines in the top graphs, the maximum amount of liabilities describe when the country would optimally default. A country leaving the risk-sharing contract with some stock of liabilities in some given endowment state would default on its obligations if debt was higher than the red line. The line depicts the maximum debt that the country would optimally repay.

In summary, the two contracts behave identically numerically. They both yield higher risk-sharing than the outside option, thereby producing higher values for the problem.

Crisis Simulation

In this section we describe the economies after a crisis event. Inside the contract we define a crisis state as a country having the lowest endowment realization and having a binding participation constraint. Outside we define the crisis as the lowest endowment realization and having a stock of debt such that the country is indifferent between repaying and defaulting.

We start by showing the result for a single simulation over 100 periods and comparing the behaviour of the fiscal union, the currency union and the defaultable debt economy. Recall that in the outside option the nominal and real economy coincide since monetary policy eliminates pricing frictions.

Figure 3.11 shows the history of endowment realizations and the consumption of tradeables in the three scenarios: in the defaultable debt economy, in the fiscal union and in the currency union.

The right panel shows the consumption of tradeables over the first 100 periods after the crisis. The black line displays the path for the economy in the outside option. Consumption

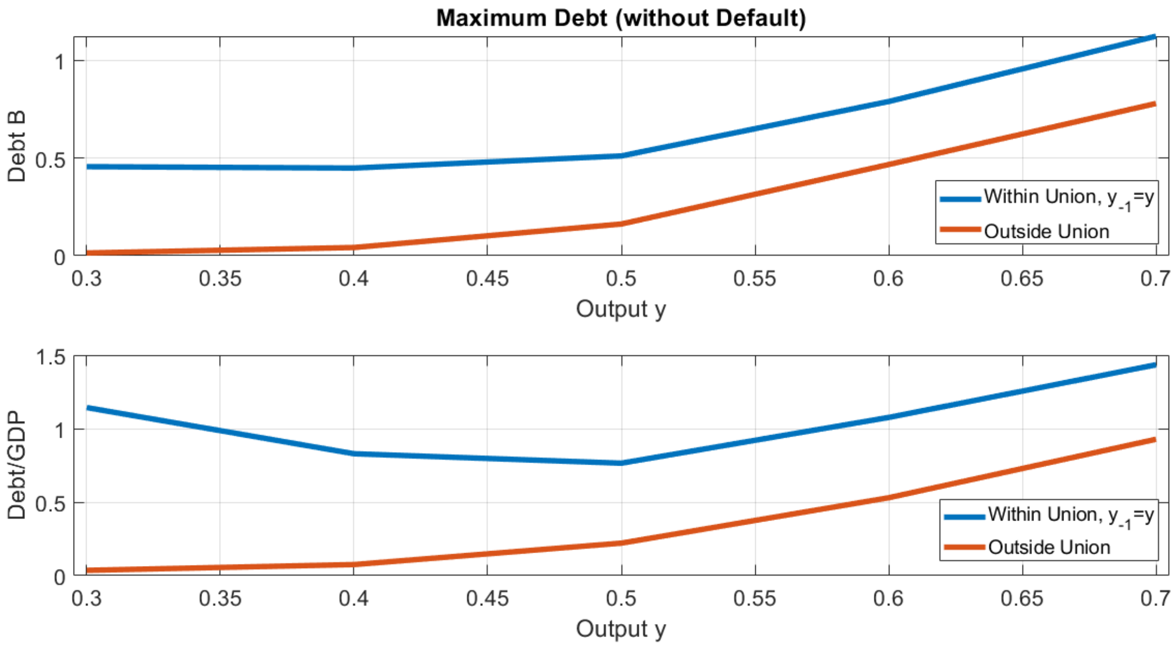


Figure 3.10: Maximum Debt

closely tracks the endowment state, showing limited scope for consumption smoothing through defaultable debt. In the outside option economy default occurs a number of times after the initial crisis before the economy manages to stabilize during a period of above average output realizations, before defaulting again towards the end of the simulation. We also see that the country pays the default cost when the endowment is high enough during exclusion, as evidenced by the difference in the endowments between the outside option and contract economies in the first panel. The two contracts behave identically, as we would expect from the policy functions: consumption increases relatively soon after the crisis and remains flat for many periods, before increasing again in response to high endowments. Since the simulation starts with the lowest endowment and relative weight, we only observe increases in consumption as the endowment reverts to its mean and the relative weight moves towards one.

In Figure 3.12 we plot the endowment, transfers, debt and the interest rate spread. The defaultable debt economy shows that when the country defaults and is temporarily excluded from financial markets, it has zero debt and zero net borrowing (which we compare with transfers in the contract). In the path of the interest rate spreads, default periods can be seen by noting that there is no spread (i.e. no debt is being traded and so there is no bond price to quote). As we would expect, spreads rise as the country approaches a default episode, reflecting the increase probability of default. After a long period of exclusion, the country

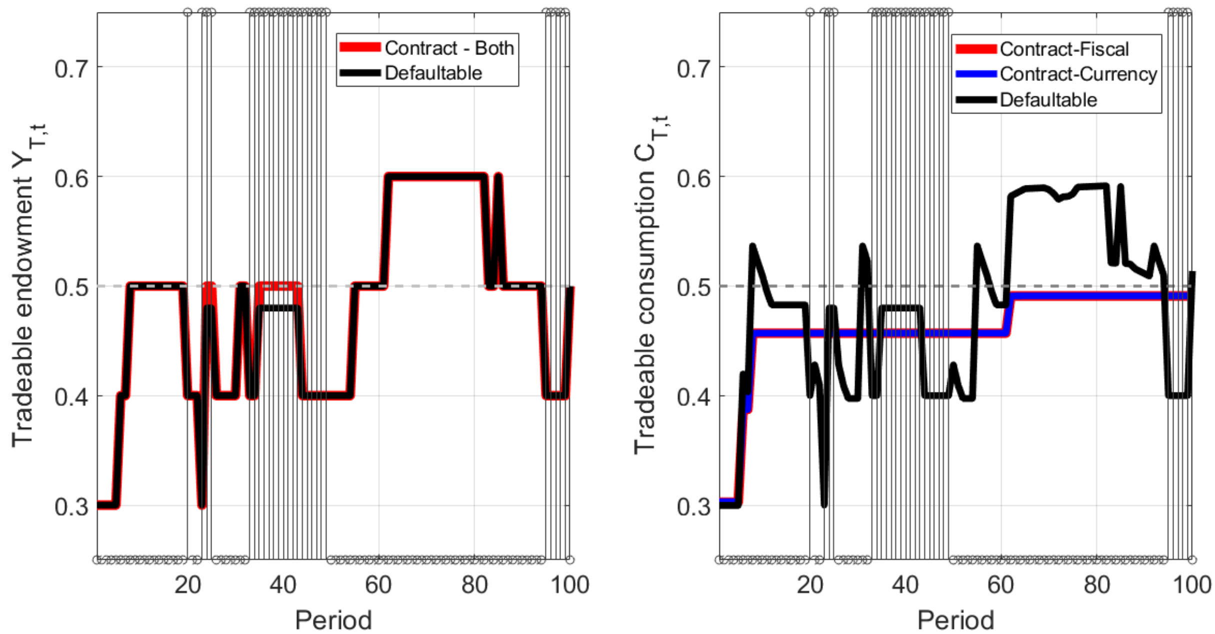


Figure 3.11: Endowment and Consumption

regains access to financial markets in the middle of the simulation, and begins to accumulate a small amount of debt before starting to save, through current account surplus, as it experiences high endowment realizations. During this saving period the interest rate spread is zero. At the end of the simulation output falls again, spreads rise as the country borrows in an attempt to smooth consumption, and eventually the economy defaults again.

Inside the risk-sharing contracts the paths of debt and transfers are the same. At the beginning of the crisis, the country is so indebted that it can no longer borrow and so it receives zero transfers; however, the stock of debt which it is able to accumulate is much higher than in the defaultable debt economy. The liabilities of the country are then reduced in response to a sharp fall in output. In this case, the fall in output it so large that the country expects to receive transfers in the near future; this corresponds to a net asset position. For the rest of the simulation, the country accumulates debt when output falls, and repays it when it rises, in order to smooth consumption. The interest spreads are lower and more stable inside the contract, and are actually negative immediately after the crisis.

Next we look at the average behaviour of the contracts in response to a crisis episode. We do so by averaging across 25000 crisis simulations. Figure 3.13 shows the average response of consumption in the fiscal and currency union. The dashed lines represent the interquartile range.

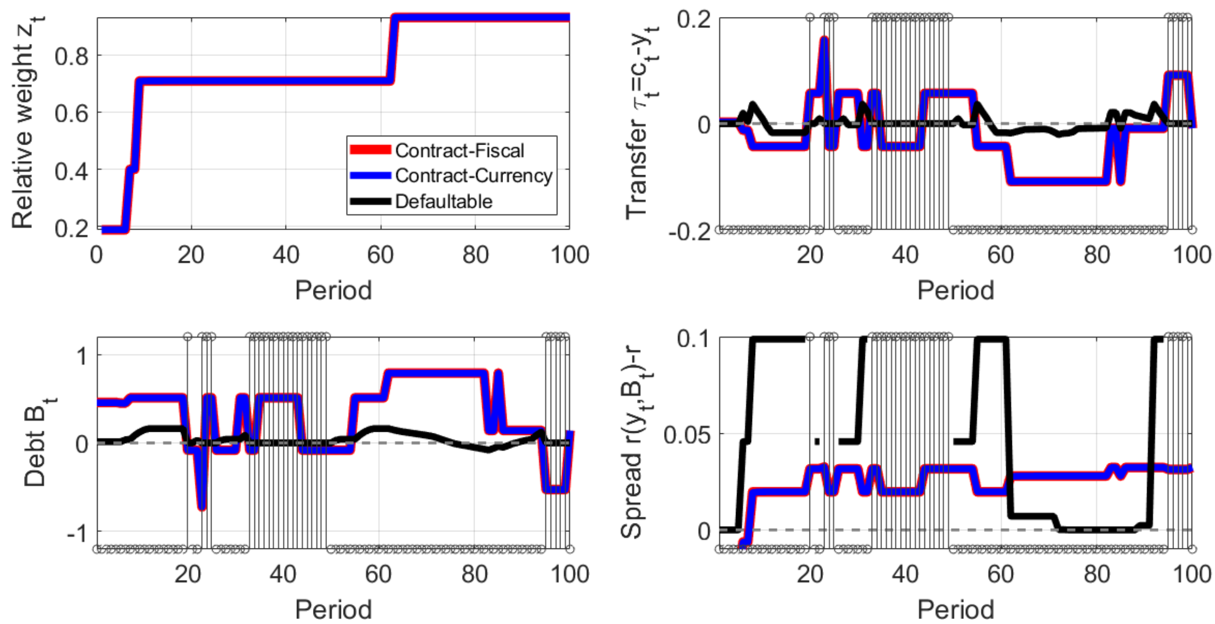


Figure 3.12: Financial Variables

Table 3.3 shows the main moments of some key outcomes of the simulations from the perspective of the country in crisis (recall that if one country is in crisis, the other must be experiencing a boom). As the economy starts in a recession and the endowment state is persistent the average endowment is .48, lower than its unconditional average of 0.5. In the economy with defaultable debt the average endowment is further decreased by the default cost. Consumption, conversely, is at its highest in the defaultable debt economy. The same ranking however holds for consumption volatility and its correlation with the endowment state. The average absolute value of transfers (current accounts) is much smaller in the defaultable debt economy compared to the unions, reflecting the reduced borrowing capacity outside the contract. The stocks of liabilities are quite different inside and outside the contract. In the outside option, the country on average has a small amount of debt, roughly one eighth of the level inside the contracts.
Finally, transfers are largely countercyclical, particularly inside the risk-sharing contracts. Countercyclicality is stronger in the union, which explains the much greater stabilization of output displayed in the simulations.
In the next two figures we plot the impulse responses of the tradeable goods in the three economies, as well as the relative weight of the crisis country in each of the contracts. The solid lines represent the average paths of the variables whereas the dashed lines represent paths one standard deviation away from the mean. In the right hand panel of Figure 3.13, we see that

	Outside-Defaultable Debt	Contract-Fiscal	Contract-Currency
Mean			
Y_t	0.478	0.480	0.480
$C_{T,t}$	0.478	0.471	0.471
GDP	0.716	0.696	0.696
$\mid \tau_t \mid$	0.011	0.0652	0.0652
B_t	0.023	0.177	0.176
z_t	-	0.839	0.840
$V(y, b/z)$	-67.541	-66.392	-66.391
$Pr(PCbinding)$	-	0.057	0.066
Standard deviation			
$\sigma(c_{T,t})$	0.1	0.052	0.052
$\sigma(Y_t)$	0.105	0.0105	0.105
Correlation			
$\rho(C_{T,t}, Y_t)$	0.987	0.617	0.617
$\rho(\tau_t, Y_t)$	-0.351	-0.872	-0.872

Table 3.3: Crisis Moments

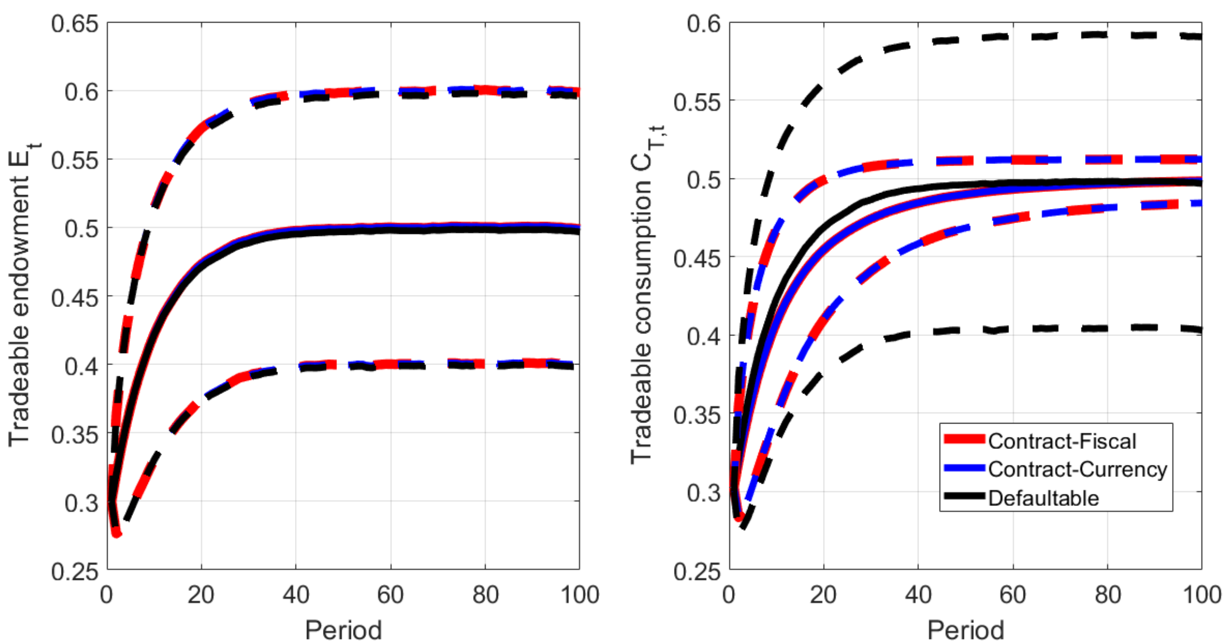


Figure 3.13: Tradeables Impulse Response After Crisis

after 100 periods, the average level of consumption is roughly the same in all three economies, and close to the mean level of the tradeable goods endowment. After the crisis, consumption also tends to recover faster in the defaultable debt economy. However, the dashed lines tell us that consumption is much more volatile outside the union than it is inside.

Figure 3.14 shows the average path of the transfers, the stock of debt and interest rate spreads in the fiscal and currency unions compared to the defaultable debt economy outside the contract. We see that on average transfers are close to zero in the defaultable debt economy, reflecting an inability to borrow, whereas in the union the crisis country initially makes net payments to the other country. The fact that the country in the union makes net payments in the aftermath of the crisis may be counterintuitive. However, as we can see in the top left panel of 3.14, the country begins the crisis with a very low relative weight, which corresponds to low consumption. Along any history which leads to this crisis state, the country will have been able to borrow large amounts to smooth consumption, an option which would not have been available outside the union.

The paths of liabilities are also very different for in the contracts, compared to the outside option. In the contracts, the economy begins the crisis with a large stock of debt, which it gradually repays over the course of the simulation. The defaultable debt economy, on the other hand, tends to spend the periods after crisis with zero net liabilities because it frequently defaults when it enters a crisis and subsequently spends some periods in financial autarky.

In the bottom right panel of Figure 3.14 we see the average path of the interest rate spreads which correspond to these movements in liabilities. The bold black line, which shows the median spreads for the defaultable debt economy [6], is calculated only for those states in which the economy does not default, since if it does default no debt is traded and there is no interest rate. We therefore see that if the economy does not default immediately during the crisis, it faces elevated interest spreads due to the high probability of default in the future. After this, the tradeables endowment reverts to its mean level, where default is less likely, and we see that spreads are volatile but typically close to zero. In contrast, in the union contracts we see *negative* spreads before the country gradually recovers from the crisis and its outstanding liabilities trade with a stable positive spread. The initial negative spreads are an artefact of the lack of aggregate risk in the union. While one country is in crisis, the other is experiencing a boom, and is therefore extremely willing to hold assets against the likelihood

[6] For the paths of interest rate spreads, we face the problem that in some states spreads in the defaultable debt economy jump to extremely high levels, which inhibits the convergence of the standard deviation and the average paths across simulation. We therefore plot the median and interquartile range for the defaultable debt economy, since these statistics are more robust to outliers.

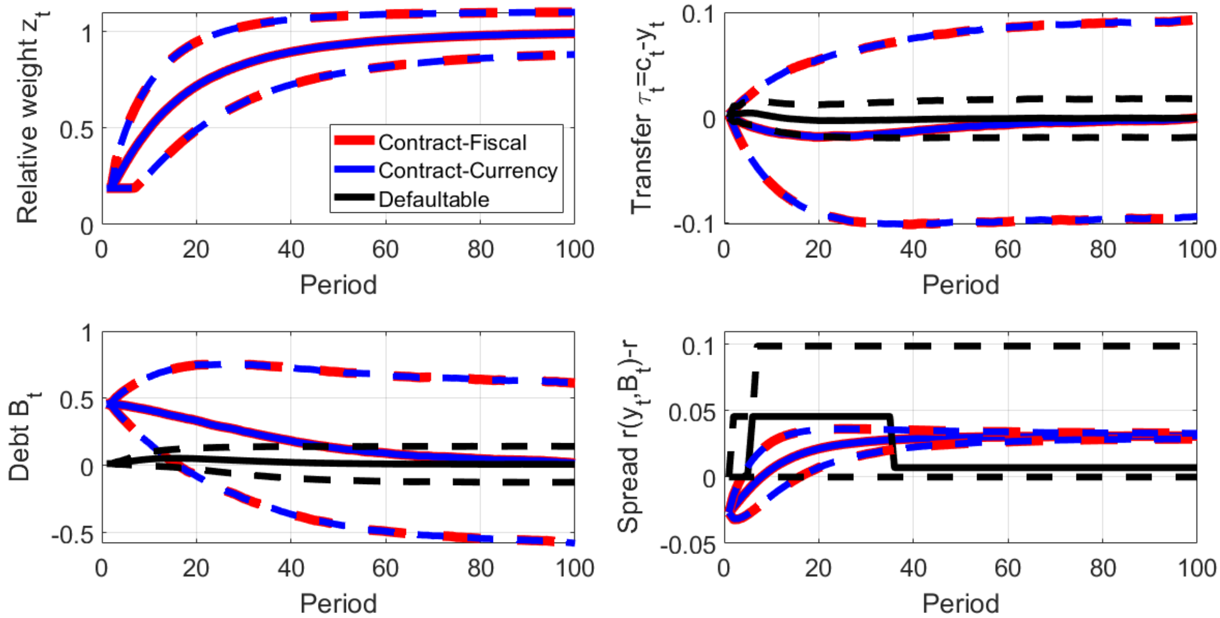


Figure 3.14: Financial Variables

that its endowment (and tradeable consumption) will fall in the near future [7].
Finally, in the top left panel of Figures 3.14 we see the impulse responses of the relative weight of the crisis country in the two contracts. During the crisis, the country receives the lowest level of tradeables endowment, and is therefore willing to accept a very low relative weight because its outside option is also very unattractive. As the country's endowment reverts to its mean however, the initial level of z is too low to satisfy the country's participation constraint, and so the relative weight is driven upwards to keep the country inside the contract, until the weight reaches one. We should recall that, due to the imperfect risk-sharing in the steady state, while the impulse response for z exhibits a smooth path, actual changes in the relative weight take place through discrete jumps (as seen, for example in Figure 3.12), due to the discrete set of values at which the participation constraint binds for different realizations of the endowment.

[7] See Appendix 3.A for the relationship between the (implicit) interest rates on the liabilities with the contracts and the marginal rates of substitutions for tradebles consumption in the two countries

	Outside-Defaultable Debt	Contract-Fiscal	Contract-Currency
Mean			
Y_t	0.498	0.5	0.5
$C_{T,t}$	0.502	0.5	0.5
GDP_t	0.637	0.625	0.625
$\mid \tau_t \mid$	0.021	0.076	0.076
B_t	-0.201	-0.002	-0.002
z_t	-	1.001	1.001
$V(y, b/z)$	-59.303	-55.018	-55.021
$Pr(PCbinding)$	-	0.0093	0.0093
Standard deviation			
$\sigma(c_{T,t})$	0.088	0.005	0.005
$\sigma(Y_t)$	0.1	0.1	0.1
Correlation			
$\rho(C_{T,t}, Y_t)$	0.954	0.364	0.364
$\rho(\tau_t, Y_t)$	-0.508	-0.999	-0.999

Table 3.4: Moments: $\gamma = 3$

3.4 Robustness Checks

In this section we provide the steady state moments for three alternative calibrations of our model. In the first two we change the parameter governing the risk aversion, which allows us to alter how agents value risk-sharing. In the final one we revert to the risk aversion of the baseline parameter set ($\gamma = 2$), and instead reduce the persistence of the endowment process. Tables 3.4 and 3.5 show the steady state moments for the same model discussed above but with the risk aversion parameter equal to 3 and 4, respectively.

Starting from Table 3.4, in the outside option economy agents show a higher level of steady state assets. This driven by the higher precautionary motif, which also, through positive interest rate on the assets, allows the country to consume more than the endowment on average.

The contracts show again similar values, with a marginally bigger difference in values between the fiscal and the currency union. As agents value smooth consumption more than in the previous simulation, the planner optimally reduces the variance of consumption by increasing the countercyclicality of transfers and increasing their average size by about 1.5%.

Table 3.5 provides a very different picture. The outside option economy increases the steady state level of assets compared to the previous economies, which significantly increases the average consumption of tradeables due to returns on the stock of assets.

The contracts are now very different from before. The risk aversion is large enough that the limited enforcement friction has little bite, allowing the planner to achieve full risk-sharing. In

	Outside-Defaultable Debt	Contract-Fiscal	Contract-Currency
Mean			
Y_t	0.5	0.5	0.5
$C_{T,t}$	0.51	0.5	0.5
GDP_t	0.577	0.563	0.563
$\mid \tau_t \mid$	0.031	0.075	0.075
B_t	-0.509	-0.002	-0.002
z_t	-	1	1
$V(y, b/z)$	-73.654	-65	-65
$Pr(PCbinding)$	-	0	0
Standard deviation			
$\sigma(c_{T,t})$	0.083	0	0
$\sigma(Y_t)$	0.1	0.1	0.1
Correlation			
$\rho(C_{T,t}, Y_t)$	0.916	0	0.
$\rho(\tau_t, Y_t)$	-0.583	-1	-1

Table 3.5: Moments: $\gamma = 4$

these economies the steady states feature a constant relative weight. As full risk-sharing is achieved this economy falls into the case described in Proposition 3.3. We therefore observe that, as there is no deadweight loss, the fiscal and currency union can attain exactly the same allocation.

Finally we consider economies with much lower persistence in the endowment process, where the AR1 parameter ρ is reduced from 0.9 to 0.5. As shown in Table 3.6, this parameter choice also delivers a constant weight steady state. However, the mechanism is slightly different. Since output reverts to the mean more quickly with lower persistence, a country currently receiving a high endowment faces a more similar future endowment stream to a country with a low endowment. The sets of relative weights which will satisfy both countries is therefore more similar, and actually overlaps. Compared to the baseline parameter set, the experience of the defaultable debt economy is much improved when the persistence of output is lower. In particular, since periods of low output are shorter on average, the economy is more able to borrow against higher future income, and the consumption smoothing which it can achieve is higher; the volatility of consumption is reduced by about one third compared to the baseline.

	Outside-Defaultable Debt	Contract-Fiscal	Contract-Currency
Mean			
Y_t	0.5	0.5	0.5
$C_{T,t}$	0.502	0.5	0.5
GDP_t	0.755	0.75	0.75
$\mid \tau_t \mid$	0.048	0.077	0.077
B_t	0	-0.001	-0.001
z_t	-	1	1
$V(y, b/z)$	-65.594	-65	-65
$Pr(PCbinding)$	-	0	0
Standard deviation			
$\sigma(c_{T,t})$	0.061	0	0
$\sigma(Y_t)$	0.1	0.1	0.1
Correlation			
$\rho(C_{T,t}, Y_t)$	0.807	0	0
$\rho(\tau_t, Y_t)$	-0.813	-1	-1

Table 3.6: Moments: $\rho = 0.5$

3.5 Productivity shocks

In this section we extend the model to include productivity shocks in the non-tradeable sector. The combination of endowment shocks and separable homothetic preferences lies behind the equivalence result presented in the previous section. The intuition is that the cost of losing independent monetary policy is proportional to the variance of consumption and is zero when consumption is perfectly smoothed across states. As fiscal policy is able to fully or almost fully smooth consumption we find that a common currency carries zero to little cost relative to independent monetary policy.

In this section we propose an extension in which non-tradeable production is subject to stochastic productivity. The goal is to check how large the welfare losses from the common currency are in presence of non-insurable variations in consumption.

Formally equation 3.2 becomes

$$Y_{NT}^{ij}(s) = A_i(s)N_{ij} \tag{3.55}$$

Equation 3.20, which defines the labour wedge, now becomes

$$\kappa^i(s) = 1 - \frac{1}{A_i(s)} \frac{U_N^i(s)}{U_{NT}^i(s)} = 1 - \frac{1}{A_i(s)} {C_{NT}^i}^{\gamma+\phi}(s) \tag{3.56}$$

and Equation 3.47 which gives the labour supply is now

$$N_i = \frac{1}{A_i(s)} C_{NT,i}(s,z) \tag{3.57}$$

Where s now denotes a two variable state vector which includes the endowment and the productivity realization. The rest of the model can be read from the previous section where everything is now contingent of both state variables.

Tables 3.7 provide the key steady state moments for the economies parametrized with risk aversion coefficient of $\gamma = 2$. Recall that in the benchmark economy, with $\gamma = 2$, we found that both the fiscal and the currency union had a steady state cycle and that their values were almost identical quantitatively. The addition of stochastic productivity breaks this negligible difference in steady state values. Starting from the currency union, we find that the contract has no surplus, hence a common currency joint with common fiscal policy cannot be sustained. As discussed above, a single monetary authority is ill-equipped to smooth the variations coming from both the endowment and the productivity shocks for both countries. At the same time in the outside option economy the independent central bank can maintain the non-tradeable side of the economy at first best levels. The fiscal union, similarly to the benchmark case, displays a steady state cycle. Consumption behaves as in the benchmark economy, even displaying the same level of volatility. The lower value of the contract is given by the higher volatility of the non-tradeable side of the economy, which now responds to the fluctuations in productivity.

To further analyse the properties of this economy we report the same steady state moments when agents are more risk-averse ($\gamma = 4$). Recall that under this parameterization the benchmark model achieved full risk sharing in both the fiscal and the currency union contracts. In the presence of productivity shocks we find that the fiscal union still achieves full risk sharing. The currency union, however, retains steady state fluctuations in tradeable consumption. This can be explained by the smaller surplus of this contract, as the outside option provides further smoothing possibilities via independent monetary policy.

Lastly, we study how these economies behave for different levels of variance of productivity shocks. When the variance $\sigma_p = 0$, we obtain our benchmark model. Figure 3.15 shows the values of the problem against the productivity shocks variance. Recall that the currency union with $\gamma = 2$ has no surplus. A number of features are worth discussion. First, when full risk sharing is achieved (fiscal union with $\gamma = 4$) the value of the problem is monotonically decreasing in the variance of productivity. By full risk sharing the economy is at first best, this

	Outside-Defaultable Debt	**Contract-Fiscal**	**Contract-Currency**
Mean			
Y_t	0.5	0.5	-
A_t	1	1	-
$C_{T,t}$	0.47	0.5	-
$\mid \tau_t \mid$	0.014	0.075	-
B_t	-0.03	-0.003	-
z_t	-	1	-
$V(y, b(z))$	-71.08	-65.46	-
$Pr(PCbinding)$	-	0.039	-
Standard deviation			
$\sigma(c_{T,t})$	0.089	0.013	-
$\sigma(Y_t)$	0.1	0.1	-
$\sigma(A_t)$	0.1	0.1	-
Correlation			
$\rho(C_{T,t}, Y_t)$	0.99	0.49	-
$\rho(\tau_t, Y_t)$	0.685	-0.01	-
$\rho(C_{T,t}, A_t)$	-0.069	0.022	-
$\rho(\tau_t, A_t)$	-0.056	-0.007	-

Table 3.7: Moments: $\gamma = 2$

	Outside-Defaultable Debt	**Contract-Fiscal**	**Contract-Currency**
Mean			
Y_t	0.5	0.5	0.5
A_t	1	1	1
$C_{T,t}$	0.5	0.5	0.4996
$\mid \tau_t \mid$	0.014	0.075	0.069
B_t	-0.025	-0.006	0
z_t	-	1	1
$V(y, b(z))$	-80.65	-65.42	-66.72
$Pr(PCbinding)$	-	0	0.035
Standard deviation			
$\sigma(c_{T,t})$	0.094	0	0.031
$\sigma(Y_t)$	0.1	0.1	0.1
$\sigma(A_t)$	0.1	0.1	0.1
Correlation			
$\rho(C_{T,t}, Y_t)$	0.983	0	0.629
$\rho(\tau_t, Y_t)$	0.821	-0.012	-0.008
$\rho(C_{T,t}, A_t)$	0.005	0	0.128
$\rho(\tau_t, A_t)$	-0.015	-0.003	0.006

Table 3.8: Moments: $\gamma = 4$

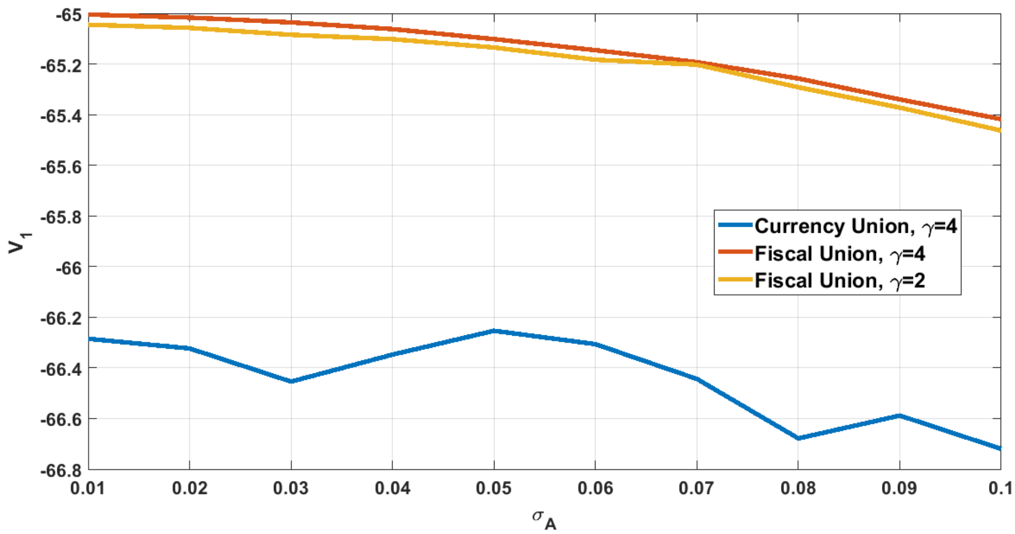


Figure 3.15: Steady State Values

implies that the only residual variation is given by the optimal non-tradeable consumption plan inheriting volatility from the productivity process. As this volatility increases with the variance of the shock, by risk aversion, the value decreases. Secondly, we cannot state a similar result for the cases in which full risk sharing is not achieved. When the economy still has non zero variance in tradeable consumption we see that the value of the contract can be locally increasing in the variance of the productivity shocks. This result is likely to be due to the behaviour of the outside option economy. We observe that increasing the variance of productivity has non-monotonic effects on the likelihood of default in the defaultable debt economy, due to the opposing effects of an increased precautionary savings motive and a higher probability of low productivity realizations. This in turn causes non-monotonic changes in the value of the outside option and the value of the contract, since the movements in steady state consumption are driven by the outside option.

3.6 Conclusions

In this paper we develop a model of fiscal and currency unions as recursive contracts. We lay down a framework in which two symmetric, equally patient, risk-averse countries face idiosyncratic risk on their tradeables endowment. There is no aggregate risk since the risks are fully negatively correlated. They partake in a risk-sharing agreement subject to a participation constraint. In this constraint the outside option is defined by an Arellano (2008) type economy, in which countries can borrow and default on a risk-neutral lender. Inside the agreement they are able to set up state contingent transfers to reduce consumption volatility. We show that a fiscal union with two independent monetary authorities manages to achieve considerable consumption smoothing.

The fiscal union with two independent monetary authorities has one more policy instrument than the currency union and, therefore, it achieves a higher value. The role of independent monetary policy is to close the labor wedge resulting from the pricing rigidities faced by non-tradeables producers. In a currency union the lack of independent monetary policy implies that the economy is producing at a suboptimal level since a single monetary policy cannot simultaneously close the wedges of both countries. Therefore, the possibility of having an independent monetary policy outside the union makes this institutional design relatively more attractive. We show that, indeed, at the steady state, the monetary union cannot do better than the fiscal union with independent monetary policies. Nevertheless, we quantitatively find that an optimal design of state-dependent transfers, taking as given the optimal monetary policy of the currency union, can compensate almost all the losses of loosing monetary independence.

We provide a characterization of the optimal cross country transfers. We show that the optimal policy requires large countercyclical transfers as a device to smooth consumption. In addition, since in the currency union larger changes in the endowment of tradeables result in large labour wedges, the optimal transfers should be higher after large transitions. It is this extra adjustment in transfers which partially closes the gap between the currency union and the fiscal union. In our simulations, where idiosyncratic risk is significant, the monetary union risk-sharing agreement also allows a significantly higher debt capacity than the defaultable debt economy. Neither the fiscal union nor the monetary union achieves full risk-sharing but they are both able to reduce the volatility of consumption by about 4/5, compared to the defaultable debt economy. However, significant cyclicality of consumption remains.

A number of extensions of this paper would be of interest. An extension of this model in which countries are not symmetric, particularly with respect to the average size of their output, would allow us to analyse situations closer to real world experiences such as the Euro Area. In the same spirit, an extension allowing for aggregate uncertainty would be of interest. Lastly, a further interesting addition would be that of a fiscal externality. This would allow the analysis of cases like the Greek debt crisis, where Greek debt was being held by German banks.

Bibliography

Abraham, Arpad, Eva Carceles-Poveda, Yan Liu, and Ramon Marimon, "On the optimal design of a Financial Stability Fund," *Link to latest version*, 2019.

Alvarez, Fernando and Urban J Jermann, "Efficiency, equilibrium, and asset pricing with risk of default," *Econometrica*, 2000, *68* (4), 775–797.

Arellano, Cristina, "Default risk and income fluctuations in emerging economies," *American Economic Review*, 2008, *98* (3), 690–712.

Auclert, Adrien and Matthew Rognlie, "Monetary union begets fiscal union," 2014.

Bayer, Christian, Chi Kim, and Alexander Kriwoluzky, "The term structure of redenomination risk," 2018.

Cooper, Russell and Hubert Kempf, "Commitment and the Adoption of a Common Currency," *International Economic Review*, February 2003, *44* (1), 119–142.

_ , _ , and Dan Peled, "Is It Is Or Is It Ain'T My Obligation? Regional Debt In A Fiscal Federation," *International Economic Review*, November 2008, *49* (4), 1469–1504.

Farhi, Emmanuel and Iván Werning, "Fiscal unions," *American Economic Review*, 2017, *107* (12), 3788–3834.

Gourinchas, Pierre Olivier, Philippe Martin, and Todd Messer, "The Economics of Sovereign Debt, Bailouts and the Eurozone," 2019.

Hoddenbagh, Jonathan and Mikhail Dmitriev, "The Financial Accelerator and the Optimal State-Dependent Contract," *Review of Economic Dynamics*, March 2017, *24*, 43–65.

Kehoe, Patrick J and Fabrizio Perri, "Competitive equilibria with limited enforcement," *Journal of Economic Theory*, 2004, *119* (1), 184–206.

Marcet, Albert and Ramon Marimon, "Recursive Contracts," *Econometrica*, 2019, *87* (5), 1589–1631.

3.A Proofs

Proof of Lemma 3.1. Non-tradeable goods producers maximize the expected profits across states, inheriting the households' nominal discount factor $1/\epsilon(s)C_T(s)^{-\gamma}$. Firms maximize

$$\Pi(p) = \sum_s \pi(s|s_{-1}) \frac{1}{\epsilon(s)C_T(s)^{-\gamma}} \left[(p - (1+\tau_L)W(s)) \left(\frac{p}{P_{NT}(s)} \right)^{-\frac{\varepsilon}{\gamma}} \left(\frac{\alpha\epsilon(s)}{P_{NT}(s)} \right)^{\frac{1}{\gamma}} C_T(s) \right] \tag{3.58}$$

The first order condition with respect to the price p is

$$\frac{\partial \Pi(p)}{\partial p} : \alpha^{\frac{1}{\gamma}} \sum_s \pi(s|s_{-1}) p^{-\frac{\varepsilon}{\gamma}} \epsilon(s)^{\frac{1-\gamma}{\gamma}} P_{NT}(s)^{\frac{\varepsilon-1}{\gamma}} C_T(s)^{1-\gamma} \left[1 - \frac{\varepsilon}{\gamma}(p - (1+\tau_L)W(s))p^{-1} \right] = 0 \tag{3.59}$$

Using $p = P_{NT}(s) = P_{NT}$, $\forall s$, this condition becomes

$$\sum_s \pi(s|s_{-1}) \epsilon(s)^{\frac{1-\gamma}{\gamma}} P_{NT}^{-\frac{1}{\gamma}} C_T(s)^{1-\gamma} \left[1 - \frac{\varepsilon}{\gamma} \left(1 - (1+\tau_L)\frac{W(s)}{P_{NT}} \right) \right] = 0 \tag{3.60}$$

Which yields

$$P_{NT} = \frac{\varepsilon}{\varepsilon - \gamma}(1+\tau_L) \frac{\sum_s \pi(s|s_{-1})\epsilon(s)^{\frac{1-\gamma}{\gamma}} C_T(s)^{1-\gamma} W(s)}{\sum_s \pi(s|s_{-1})\epsilon(s)^{\frac{1-\gamma}{\gamma}} C_T(s)^{1-\gamma}} \tag{3.61}$$

Using the labor subsidy $(1+\tau_L) = \frac{\varepsilon-\gamma}{\varepsilon}$, it simplifies to the first statement in Lemma 3.1

$$P_{NT} = \frac{\sum_s \pi(s|s_{-1})\epsilon(s)^{\frac{1-\gamma}{\gamma}} C_T(s)^{1-\gamma} W(s)}{\sum_s \pi(s|s_{-1})\epsilon(s)^{\frac{1-\gamma}{\gamma}} C_T(s)^{1-\gamma}} \tag{3.62}$$

To obtain the second statement, notice that, using the definition of the labor wedge and the household first order condition, one has

$$\frac{W(s)}{P_{NT}} = 1 - \kappa(s) \tag{3.63}$$

Hence, the optimal non-tradeable good price implies

$$\sum_s \pi(s|s_{-1})\epsilon(s)^{\frac{1-\gamma}{\gamma}} C_T(s)^{1-\gamma}\kappa(s) = 0 \tag{3.64}$$

Which completes the proof.

■

Proof of Lemma 3.2. The goal of the central bank is to maximize agents' welfare by means of the exchange rate ϵ. The exchange rate in this setting is equivalent to the price of the tradeable good P_T . Recall the following relationships from the household's first order conditions: $C_T^{-\gamma} = \frac{\alpha P_T}{P_{NT}} C_{NT}^{-\gamma}$. Using $\epsilon = P_T$ and inverting the previous relationship one gets $\frac{\partial C_{NT}}{\partial \epsilon} = \frac{1}{\gamma}\frac{C_{NT}}{\epsilon}$. Finally, recall that by labor market clearing $C_{NT} = N$. As monetary policy does not carry intertemporal effects, the central banks maximizes the contemporenous stream of utility:

$$v(\epsilon) = \frac{C_T^{1-\gamma}}{1-\gamma} + \frac{C_{NT}^{1-\gamma}}{1-\gamma} - \frac{N^{1+\phi}}{1+\phi}. \tag{3.65}$$

Maximizing with respect to the exchange rate

$$\frac{\partial v(\epsilon)}{\partial \epsilon} : \frac{C_{NT}}{\gamma\epsilon}[C_{NT}^{-\gamma} - C_{NT}^{\phi}] = 0 \tag{3.66}$$

Recalling the definition of the labor wedge

$$\kappa^i(s) = 1 - \frac{U_N^i(s)}{U_{NT}^i(s)} = 1 - C_{NT}^{\gamma+\phi}(s) \tag{3.67}$$

Then optimal monetary policy implies setting

$$k^i(s) = 0 \tag{3.68}$$

Which proves the first part of the lemma.

In a currency union, the monetary authority maximizes the weighted sum of the welfare of member states. Assuming equal weighing implies maximizing

$$v(\epsilon) = \frac{1}{2}v^1(\epsilon) + \frac{1}{2}v^2(\epsilon) \tag{3.69}$$

Maximizing with respect to the exchange rates yields

$$\frac{\partial v(\epsilon)}{\partial \epsilon} : C^1_{NT}[{C^1_{NT}}^{-\gamma} - {C^1_{NT}}^{\phi}] + C^2_{NT}[{C^2_{NT}}^{-\gamma} - {C^2_{NT}}^{\phi}] = 0 \tag{3.70}$$

Using the definition of the labor wedge, optimal monetary policy implies

$$\sum_{i=1,2} {C^i_{NT}}^{1-\gamma} \kappa^i(s) = 0, \quad \forall s$$

■

Proof of Proposition 3.1. We prove the proposition by constructing the competitive equilibrium which corresponds to the union allocation.

It will be convenient to have the following notation for the marginal rates of substitution of tradeable goods:

$$q(s', z' \mid s, z) = \max_i \beta \left(\frac{C_{T,i}(s', z')}{C_{T,i}(s, z)} \right)^{-\gamma} \tag{3.71}$$

We can now set the price of an Arrow security in this economy as

$$Q(s' \mid s) = \pi(s' \mid s) q(s', z' \mid s, z) \tag{3.72}$$

These Arrow prices clearly satisfy the Euler equation, with equality for the country which has the highest marginal rate of substitution. The value of the state contingent debt contract in state s is then

$$\sum_{s' \mid s} Q(s' \mid s) d(s' \mid s) = Eq(s', z' \mid s, z) d(s' \mid s) \tag{3.73}$$

We can derive from the equation of motion for z' that

$$\frac{z''}{z'} = \frac{1 + \nu_1(s', z')}{1 + \nu_2(s', z')} \tag{3.74}$$

And from the solution to the union contract we know that

$$\left(\frac{C_{T,2}(s,z)}{C_{T,1}(s,z)}\right)^{-\gamma} = z' \tag{3.75}$$

Thus we can write

$$\begin{aligned}\frac{z''}{z'} &= \frac{1+\nu_1(s',z')}{1+\nu_2(s',z')} \\ &= \left(\frac{C_{T,2}(s',z')}{C_{T,2}(s,z)}\right)^{-\gamma} \Bigg/ \left(\frac{C_{T,1}(s',z')}{C_{T,1}(s,z)}\right)^{-\gamma}\end{aligned} \tag{3.76}$$

From this expression we can see that the maximum marginal rate of substitution will be attained by the country which is unconstrained ($\nu_i = 0$) in state (s', z').
For the current debt position of each country, we write the budget constraint of country i as

$$b_i(s) = Y_T^i(s) - C_T^i(s,b) + \sum_{s'|s} Q(s' \mid s) b_i(s' \mid s) \tag{3.77}$$

and iterate forward on this equation and apply the transversality condition to obtain

$$b_{i,t} = \mathbb{E}_t \sum_{k=0}^{\infty} q(s^{t+k} \mid s_t)(Y_{i,t+k} - c_{i,t+k}) \tag{3.78}$$

where

$$q(s^{t+k} \mid s_t) = \prod_{n=0}^{k-1} q(s_{t+n+1} \mid s_{t+n}) \tag{3.79}$$

It should be clear from this definition of the debt position and the resource constraint that

$$B_1(s) = -B_2(s) \tag{3.80}$$

so that asset markets clear in every state. We set the initial debt positions as $b_{i,0} = \mathbb{E}_0 \sum_t^{\infty} q_{0,t}(Y_{i,t} - c_{i,t})$. We then choose borrowing constraints which are *not too tight* in the sense of Alvarez and Jermann (2000) so that

$$\omega(s, \bar{B}_i(s)) = V_o^i(s, \bar{B}_i(s)) \tag{3.81}$$

By definition, we will then have $b_i(s) = \bar{B}_i(s)$ whenever country i's participation constraint is binding.
To complete the proof we must show that an allocation which has a high implied interest rate also satisfies the transversality condition:

$$\begin{aligned}
&\lim_{t\to\infty} \mathbb{E}_t \beta^t q(s^{t+1} \mid s_t) C_T^i(s_t, b_i(s_t))^{-\gamma} b_i(s_{t+1}) \\
&= \lim_{t\to\infty} \mathbb{E}_t \left[\beta^t C_T^i(s_t, b_i(s_t))^{-\gamma} \max_i \beta \left(\frac{C_{T,i}(s_{t+1}, b_i(s_{t+1}))}{C_{T,i}(s_t, b_i(s_t))} \right)^{-\gamma} \right. \\
&\left. \times \mathbb{E}_{t+1} \sum_{k=0}^{\infty} q(s^{t+k+1} \mid s_{t+1})(Y_{i,t+k+1} - C_{T,i}(s_{t+k+1}, b_i(s_{t+k+1}))) \right] \\
&= \lim_{t\to\infty} \sum_{s_{t+1}|s_t} \pi(s_{t+1} \mid s_t) \left[\beta^t C_T^i(s_t, b_i(s_t))^{-\gamma} \max_i \beta \left(\frac{C_{T,i}(s_{t+1}, b_i(s_{t+1}))}{C_{T,i}(s_t, b_i(s_t))} \right)^{-\gamma} \right. \\
&\left. \times \mathbb{E}_{t+1} \sum_{k=0}^{\infty} q(s^{t+k+1} \mid s_{t+1})(Y_{i,t+k+1} - C_{T,i}(s_{t+k+1}, b_i(s_{t+k+1}))) \right] \\
&= \lim_{t\to\infty} \sum_{s_{t+1}|s_t} \beta^t C_T^i(s_t, b_i(s_t))^{-\gamma} \mathbb{E}_{t+1} \sum_{k=0}^{\infty} q(s^{t+k+1} \mid s_t)(Y_{i,t+k+1} - C_{T,i}(s_{t+k+1}, b_i(s_{t+k+1}))) \\
&= \lim_{t\to\infty} \sum_{s_{t+1}|s_t} \beta^t C_T^i(s_0, b_i(s_0))^{-\gamma} \frac{C_T^i(s_t, b_i(s_t))^{-\gamma}}{C_T^i(s_0, b_i(s_0))^{-\gamma}} \mathbb{E}_{t+1} \sum_{k=0}^{\infty} q(s^{t+k+1} \mid s_t)(Y_{i,t+k+1} - C_{T,i}(s_{t+k+1}, b_i(s_{t+k+1}))) \\
&\leq C_T^i(s_0, b_i(s_0))^{-\gamma} \lim_{t\to\infty} \mathbb{E}_{t+1} \sum_{k=0}^{\infty} q(s^{t+k+1} \mid s_t)(Y_{i,t+k+1} - C_{T,i}(s_{t+k+1}, b_i(s_{t+k+1}))) \\
&\leq C_T^i(s_0, b_i(s_0))^{-\gamma} \lim_{t\to\infty} \mathbb{E}_{t+1} \sum_{k=0}^{\infty} q(s^{t+k+1} \mid s_0)(Y_{1,t+k+1} + Y_{2,t+k+1}) \\
&= 0
\end{aligned}$$

Where the last equality follows from the high implied interest rate condition in Equation 3.49. ■

Proof of Proposition 3.2. We prove this by contradiction. Assume that country i has a binding participation constraint, so that $\lambda_i > 0$.
Recall that

$$V_i^o(s,B) = \max_{LR,LD}\{V_i^{LR}(s,B), V_{LD}^i(s)\}$$

We have shown in Proposition 3.1 that the union allocation can be decentralized as a competitive equilibrium with state contingent debt and endogenous borrowing constraints. Recall that $\omega(b_i, s)$ is the value of the problem in the decentralized equilibrium. If the participation constraint binds ($\lambda_i > 0$), it must be that

$$V_i^o(s,B) = \omega(b_i, s)$$

Recall that

$$B_{it} = \mathbb{E}_t \sum_{s=t}^{\infty} q_{t,s}(Y_{i,s} - c_{i,s}) = \mathbb{E}_t \sum_{k=0}^{\infty} q(s^{t+k} \mid s_t)(Y_{i,t+k} - c_{i,t+k}) = b_{i,t}$$

i.e. the face value of the debt in the outside option is the appropriately discounted value of the net payments in the decentralized economy.

In the outside option and in the decentralized economy, the agents maximize the same objective function under different constraints. The budget constraint in the case of exiting and repaying the liabilities is

$$C_T^i(s) + P_{NT}^i(s)C_{NT}^i(s) + B_i \le Y_T^i(s) + W^i(s)N^i(s) + \Pi^i(s) + B_i'Q(s, B_i')$$

whereas in the decentralization of the union allocation it is

$$C_T^i(s) + P_{NT,i}C_{NT}^i(s) + b_i(s) \le Y_T^i(s) + W_i(s)N_i(s) + \Pi_i(s) + \sum_{s'|s} q(s' \mid s)b_i(s' \mid s)$$

In the latter, the country is also subject to an endogenous borrowing limit, which we have specified in such a way that it is never binding if the participation constraint is slack. In addition, when the participation constraint binds, the country's liabilities are exactly equal to the borrowing limit. The borrowing limit therefore does not change the allocation.

Comparing the two budget constraints above, it is clear that the allocation in the outside option in case of repayment can always be exactly replicated in the decentralized fiscal union, since the state contingent debt can replicate any payments delivered by non state contingent bonds.

Hence, by optimality, it can never be that the value of the problem is higher in the case of leaving and repaying than in the decentralized fiscal union. Formally, $V_i^{LR}(s,B) < \omega(b_i, s)$ $\forall s, B$.

This implies that if the participation constraint binds, it must be that $V_i^o(s,B) = V_{UD}^i(s) >$

$\omega(b_i, s) \geq V_i^{LR}(s, B)$. In other words, it can never be that the participation constraint binds and the country would like to exit and *not* default. ■

Proof of Proposition 3.3. Using the definition of optimal non-tradeable prices, imposing full risk-sharing and taking the ratio of the non-tradeable prices in the two countries, we obtain

$$\frac{P_{NT}^1}{P_{NT}^2} = \left(\frac{C_T^1}{C_T^2}\right)^{\gamma} = \bar{c}^{-\gamma}$$

We show that imposing a zero wedge condition in one country immediately implies a zero wedge in the other. If country 1 has no labor wedge, $\kappa^1(s) = 0$, then

$$1 = \left(\frac{\alpha\epsilon}{P_{NT}^1}\right)^{\frac{1}{\gamma}} C_T^1$$

Substituting in the relative prices and the relative consumption as a function of the constant $\bar{c}$

$$1 = (\alpha\epsilon)^{\frac{1}{\gamma}} \, \bar{c}^{\frac{1}{\gamma}} {P_{NT}^2}^{-\frac{1}{\gamma}} \bar{c}^{-\frac{1}{\gamma}} C_T^2 = (\alpha\epsilon)^{\frac{1}{P_{NT}^2}^{-\frac{1}{\gamma}\gamma}} C_T^2$$

Which implies $\kappa^2(s) = 0$ and completes the proof ■

Proof of Corollary 2. A full risk-sharing steady state is a steady state in which relative weights are constant. As a consequence tradeable consumption is constant and marginal utilities are equal to some constant number. In this case Proposition 3.3 applies and the common monetary policy has no cost as countries attain the optimum on the non-tradeables side of the economy. Conditioning on the current states s, z the economy has the same level of tradeable consumption and the (optimal) non-tradeable and labour supply. As this is a steady state the continuation values are also identical, which proves that the value of the two programs coincide. ■

Proof of Proposition 3.4. Conditional on the optimal choice in the outside option being default, the value of the outside option is independent of the current relative weight as it is independent of the stock of liabilities.

Furthermore, whenever a participation constraint binds, the country's relative weight is increased exactly of the amount that makes it indifference between the contract and the outside option. This implies $V_i^F(s, z) = V_i^M(s, z) = V_i^o(s, B)$. As the outside option value $V_i^o(s, B) = V_{UD}^i(s)$ is independent of the relative weight it must be the same for the fiscal and the currency union. Hence the statement of the proposition

$$V_i^F(s, z) = V_i^M(s, z) = V_i^o(s, B) = V_{UD}^i(s).$$

■

Proof of Theorem 1. We start by showing that in a monetary union consumption fluctuates in narrower bands whenever the steady state features non constant consumption. By Lemma 3.2 in a fiscal union the wedge is zero for both countries in every period, while it is non-zero for both countries in a currency union. We also know that, other things equal, the value of the problem decreases as the wedge moves away from zero

$$\frac{\partial \Omega^M(s,z)}{\partial |\kappa|} < 0, \ \frac{\partial \Omega^F(s,z)}{\partial \kappa}|_{\kappa=0} = 0,$$

in fact, when $\kappa \neq 0$, $\frac{\partial V_i^M(s,z)}{\partial |\kappa|} < 0$ for $i = 1$ and 2, since domestic labor and consumption of non-tradeables is distorted in both economies. However, regarding tradeables, what is important is how the wedge affects limited enforcement constraints. Recall that if $\nu_i > 0$ then $\nu_j = 0$, $j \neq i$. Without loss of generality assume that $\nu_2 = 0$ and $\nu_1 > 0$. The value of the Lagrange multiplier ν_1 is given by

$$\frac{\partial \Omega(s,z)}{\partial V_1^M}|_{V_1^M = V_1^0} = \nu(s,z)$$

By concavity of $\Omega(s,z)$ it must be that

$$\nu_1(s,z)|_{\kappa(s)\neq 0} > \nu_1(s,z)|_{\kappa(s)=0}$$

Recall that $z' = z\frac{1+\nu_1}{1+\nu_2}$, therefore, it must be that

$$z'(s,z)|_{\kappa(s)\neq 0} > z'(s,z)|_{\kappa(s)=0}$$

This implies that for any given z, if a PC binds, next period z' will be larger in currency unions than in fiscal unions. By the definition of steady states with imperfect risk-sharing it must be that consumption fluctuates in *broader bands* in currency unions than in fiscal unions. Per se such higher volatility of consumption decreases the value of the problem. Furthermore, in currency unions, this is always paired with suboptimal non-tradeables. Hence in a steady state (s,z) the value of the currency union is lower than the value of the fiscal union with independent monetary policies.

■

Proof of Proposition 3.5. The common currency monetary policy objective function is such that it minimizes the deadweight loss. Such minimized deadweight loss defines the set of feasible allocations in a monetary union. Conditioning on this restricted feasible allocation set the transfer policy solves the planner problem, thereby picking the efficient allocation in the constrained set. ■

Proof of Corollary 3. A central bank using the planner's relative Pareto weight maximizes

$$v(\epsilon) = zv^1(\epsilon) + v^2(\epsilon)$$

This results in the following first order condition:

$$z{C_{NT}^1}^{1-\gamma}\kappa^1(s) + {C_{NT}^2}^{1-\gamma}\kappa^2(s) = 0, \quad \forall s$$

Without loss of generality, assume that $z < 1$. This also implies that $C_{NT}^1 < C_{NT}^2$. Comparing this monetary policy rule with the one of a central banks that weighs equally the two countries:

$${C_{NT}^1}^{1-\gamma}\kappa^1(s) + {C_{NT}^2}^{1-\gamma}\kappa^2(s) = 0, \quad \forall s,$$

country 1 will have a larger wedge as it carries less weight in the first order condition.
Following similar lines as the proof of the previous theorem, as country 1 has a larger wedge, if there is surplus in the contract, it will be rewarded with a larger z' for all current z in which the PC binds.
As in the theorem this implies a higher level of consumption fluctuations and a higher wedge, particularly so for the agent with high marginal utility.

■

3.B Quantitative Model

In section 3.3, we produce a 5 state Markov process for the stochastic endowment y of the tradeable good in each country. We do this by discretizing an AR1 process with persistence parameter $\rho = 0.9$ and shock variance $\sigma_y^2 = 0.01$, using the Rouwenhorst method. The transition matrix for this Markov process is:

$$\pi = \begin{pmatrix} 0.8145 & 0.1715 & 0.0135 & 0.0005 & 0 \\ 0.0429 & 0.8213 & 0.1290 & 0.0068 & 0.0001 \\ 0.0023 & 0.0860 & 0.8235 & 0.0860 & 0.0023 \\ 0.0001 & 0.0068 & 0.1290 & 0.8213 & 0.0429 \\ 0 & 0.0005 & 0.0135 & 0.1715 & 0.8145 \end{pmatrix} \tag{3.82}$$

The following graphs show, for each level of the tradeable endowment y, the interval $[\underline{z}(y), \bar{z}(y)]$

within which the participation constraints are satisfied. They therefore accompany the discussions in Section 3.3 on the ergodic sets for z in each contract and the basins of attraction for these ergodic sets.

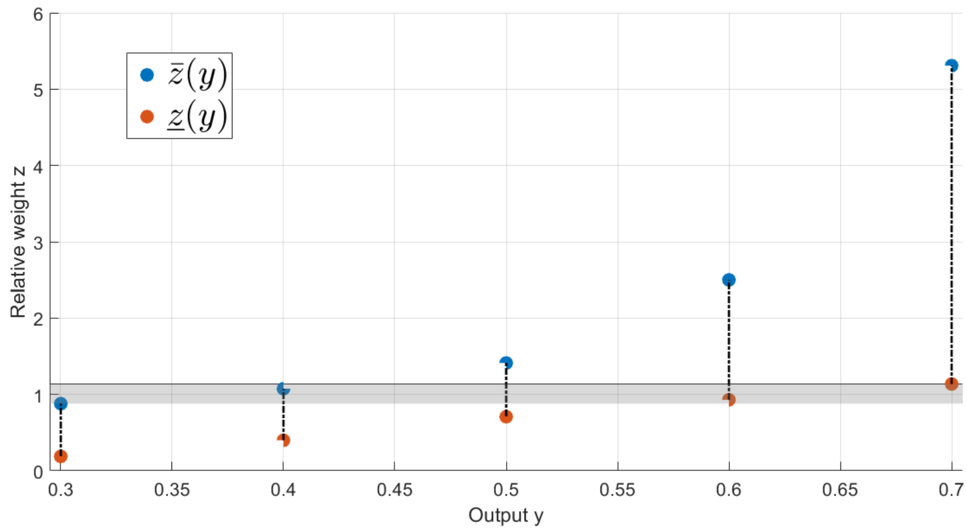


Figure 3.16: Relative Weights Bounds in Fiscal Unions

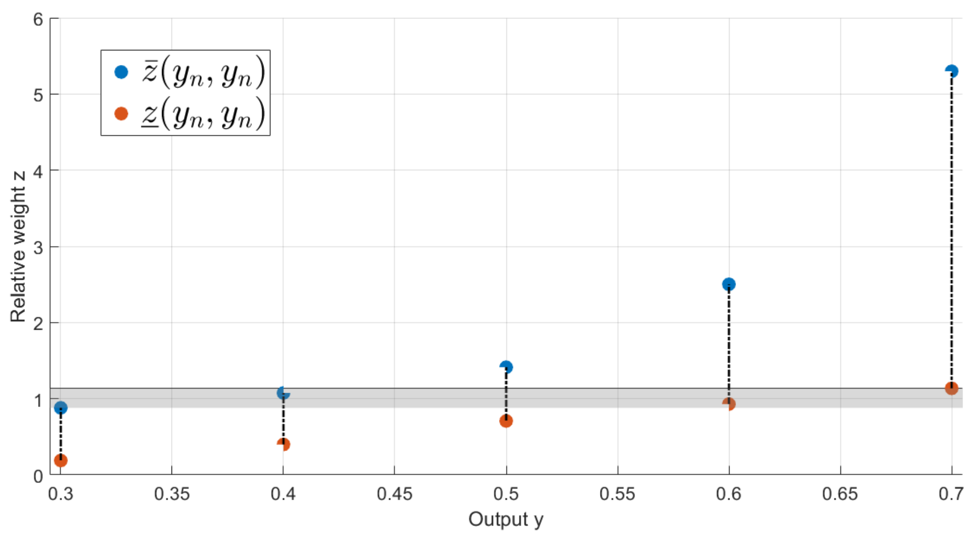


Figure 3.17: Relative Weights Bounds in Currency Unions