Essays on Macroeconomics and International Trade

by

Masao Fukui

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2021

Author . . . .
Department of Economics
May 19, 2021

Certified by . . . .
Iván Werning
Robert M. Solow Professor of Economics
Thesis Supervisor

Certified by . . . .
Arnaud Costinot
Professor of Economics
Thesis Supervisor

Accepted by . . . .
Amy N. Finkelstein
John & Jennie S. MacDonald Professor of Economics
Chairman, Department Committee on Graduate Theses

Essays on Macroeconomics and International Trade

by

Masao Fukui

Submitted to the Department of Economics
on May 19, 2021, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis consists of three essays. In the first essay, I develop a new theory of wage rigidity and unemployment fluctuations. The starting point of my analysis is a generalized version of Burdett and Mortensen's (1998) job ladder model featuring risk-neutral firms, risk-averse workers, and aggregate risk. Because of on-the-job search, my model generates wage rigidity both for incumbent workers, through standard insurance motives, and for new hires, through novel strategic complementarities in wage setting between firms. In contrast to the conventional wisdom in the macro literature, the introduction of on-the-job search implies that: (i) the wage rigidity of incumbent workers, rather than new hires, is the critical determinant of unemployment fluctuations; (ii) fairness considerations in wage setting dampen, rather than amplify, unemployment fluctuations; and (iii) new hire wages are too flexible, rather than too rigid, in the decentralized equilibrium. Quantitatively, the wage rigidity of incumbent workers caused by the insurance motive alone accounts for about one fifth of the unemployment fluctuations observed in the data.

In the second essay (joint with Arnaud Costinot and David Atkin), we study the relationship between international trade and development in a model where countries differ in their capability, goods differ in their complexity, and capability growth is a function of a country's pattern of specialization. Theoretically, we show that it is possible for international trade to increase capability growth in all countries and, in turn, to push all countries up the development ladder. This occurs because: (i) the average complexity of a country's industry mix raises its capability growth, and (ii) foreign competition is tougher in less complex sectors for all countries. Empirically, we provide causal evidence consistent with (i) using the entry of countries into the World Trade Organization as an instrumental variable for other countries' patterns of specialization. The opposite of (ii), however, appears to hold in the data. Through the lens of our model, these two empirical observations imply dynamic welfare losses from trade that are small for the median country, but pervasive and large among a few developing countries.

In the third essay, I build a model of endogenous capital flow reversal. In the data, Capital tends to flow from fast-growing countries to slow-growing countries, contrary to the prediction of neoclassical models. I propose a parsimonious theory in which slower growth *causes* capital inflow. The theory builds on the idea that financial development is demand driven. In the model, a relatively larger demand for store of value in slow-growing countries stimulates domestic financial innovation. The endogenous response of financial development can be strong enough to attract capital inflow. This contrasts with the existing theories in which slow-growing countries happen to have relatively well-developed financial markets.

Thesis Supervisor: Iván Werning
Title: Robert M. Solow Professor of Economics

Thesis Supervisor: Arnaud Costinot
Title: Professor of Economics

Acknowledgments

I am deeply indebted to many people who supported my journey of Ph.D.

First and foremost, I am truly grateful to my advisors, Iván Werning, Arnaud Costinot, Robert Townsend, Emi Nakamura, and Jón Steinsson. I cannot imagine how I would have completed my journey of Ph.D. without their constant and enormous help, advice, and support throughout my Ph.D.

Iván has been always extremely friendly and encouraging since the start of my Ph.D., although I had him as a formal main advisor toward the late in my Ph.D. When I was contemplating between the several fields for specialization in my earlier years, he always kept telling me that I should do macro. When I was struggling with my thesis, he always welcomed me to drop by his office for a chat. When I was on the job market, he supported me in every single dimension. He helped me understand my model, practiced interviews and presentations for many hours, and encouraged me with so many thoughtful words when I was not confident. Iván is also a macroeconomist that I always look up to. The moment I talked to him was the moment I learned the most about macro.

Arnaud has been always playing a central role for me throughout my Ph.D. life. I took international trade sequence in my first year, not expecting this is going to be one of my specializations. I was fascinated by Arnaud's class and his research agenda. I remember I was extremely excited when he asked me whether I would like to do RA for him at the end of my first year. Ever since then, I kept learning from him about trade as an advisee, an RA, a TA, and a coauthor. He always has been very generous in his time. I always felt I am taking too much of his time. He spent so much time in listening my half-baked ideas, giving me detailed feedbacks, and reading my papers, but he always welcomed me to do so. Even after I started to write my job market paper on macro, he kept giving me his generous support in carefully reading my job market paper and slides with extremely detailed feedbacks. One day in the future, I hope to come back to his office with a great trade paper.

Rob has been always available encouraging and friendly, and he has been a great mentor. If I was not in getting in touch with him for two weeks, he always sent me an email to set up a meeting. Every conversation in his office has been always a relaxed and fun moment. When I was struggling with my research, he always reminded me of the spirit of enjoying the research. Indeed, I was motivated whenever I see him talking about economics with full of smile and excitement on his face. He was always open to listen to any kind of my research, and he always gave me strong encouragement to pursue what I think is interesting.

I was extremely fortunate to meet with Emi and Jón in my first year at MIT when they were visiting. I presented some research ideas in their international macro class, and ever since then, Emi and Jón have been provided me an incredible amount of help and knowledge about economics. Through the countless interaction with them, I learned everything about empirical macro, the actual process of doing research, and how to write good papers. Most importantly for me, I always enjoyed working with them together. Emi always sent me endless emails with so many questions, and Jón's always gave sharp comments, both of which always deepened my understanding about macroeconomics. They also have been incredibly supportive mentors. I will forever keep their extremely warm and long emails, which they sent me after I confessed that I was struggling with my thesis. I am especially grateful that they interacted beyond academics by inviting me to New York and their house in Berkeley and by having countless daily conversations about life in general.

I am also grateful to Marios Angeletos, Martin Beraja, Ricardo Caballero, Alp Simsek, David Atkin, Dave Donaldson, and Patrick Kehoe for numerous encouragement and helpful conversations.

My Ph.D. life would not have been as enjoyable as it was without my best friend, Mikel Petri. I was fortunate to be a best friend with him since my first year, share the office with him for our entire Ph.D., and spent 90% of the time in the department together. I always miss the days that we kept talking about economics and life in our office all day. He has been always cheerful, kind, friendly, and smart. I also benefited from numerous interactions with Michele Fornino, Chen Lian, Joe Hazell, Mayara Felix, Bumsoo Kim, Marc de la Barrera, John Sturm, Ali Kakhbod, and especially, Jeremy Yang, Sophie Sun, and Nathan Zorzi.

I owe everything to my wife, Mutsumi. I cannot thank enough to Mutsumi for allowing me to pursue this career, taking care of our son, Yuto, and always being supportive of me. I hope to repay this huge debt to her through being a good husband and a good dad. I dedicate my thesis to Mutsumi and Yuto.

Contents

Chapter 1

A Theory of Wage Rigidity and Unemployment Fluctuations with On-the-Job Search

1.1 Introduction

Does wage rigidity matter for unemployment fluctuations? There is little debate about the fact that the wages of *incumbent* workers are rigid. The conventional view, however, is that this empirically well-documented source of wage rigidity in itself is inconsequential for unemployment fluctuations (Barro, 1977; Pissarides, 2009).[1] The core of the theoretical argument behind this skepticism is that the wages of *new hires*, rather than incumbent workers, are what determine a firm's marginal cost, and in turn, its hiring incentives.

The starting point of this paper is that the previous argument, as intuitive as it may sound, is at odds with one key feature of labor markets: job-to-job transitions. As Figure 1-1 shows, such transitions are a pervasive feature of the US labor market, making up more than 40% of new hires.[2] For firms hiring from a pool of unemployed *and* employed workers, the incentive to create jobs cannot be independent from prevailing incumbent wages. If, in recessions, the wages of incumbent workers do not fall, new jobs have a hard time attracting workers, which in turn discourages job creation.

Motivated by the previous fact, I propose a new theory of wage rigidity and unemployment fluctuations with on-the-job search. Among other things, it implies that: (i)

[1]See also Haefke, Sonntag, and van Rens (2013) and Rudanko (2009).

[2]The US is not an outlier. Engbom (2020) shows that although the US features higher job-to-job transition rates than most European countries, the magnitudes are comparable. Donovan, Lu, and Schoellman (2018) find that developing countries tend to have higher job-to-job transition rates than the US.

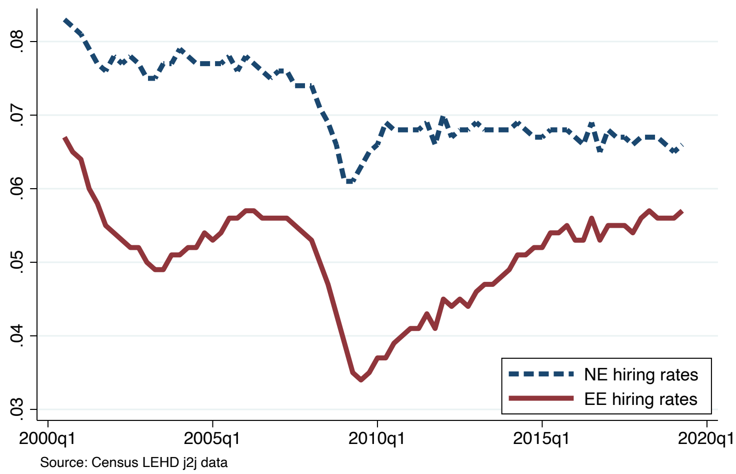


Figure 1-1: NE and EE hiring rates

Note: Figure 1-1 shows the NE (non-employment to employment) and EE (employment to employment) hiring rates from 2000-2019. The NE and EE hiring rates refer to the flow of workers from non-employment to employment, and from one employer to another as a fraction of total employment, respectively. Data are from the Census LEHD j2j database.

wages of both incumbent workers and new hires are endogenously rigid; (ii) the wage rigidity of incumbent workers, rather than new hires, is the critical determinant of unemployment fluctuations; (iii) fairness considerations in wage setting dampen, rather than amplify, unemployment fluctuations; and (iv) new hire wages are too flexible, rather than too rigid, in the decentralized equilibrium.

Section 1.2 develops a generalized version of Burdett and Mortensen's (1998) job ladder model with risk-neutral firms of heterogeneous productivity, risk-averse workers, and aggregate risk. I start with a two-period model to derive a number of sharp qualitative insights. In the first period, firms write state-contingent wage contracts with an exogenous number of incumbent workers to insure against aggregate risk.[3] In the second period, aggregate productivity shocks are realized, and firms post vacancies and wages to hire new workers. Without aggregate risk, the model is in the spirit of Burdett and Mortensen (1998). Incumbent firms and poaching firms compete for workers strategically along the job ladders subject to search frictions. While firms can commit to the wage contract, workers cannot: workers search on the job and are free to take an outside offer

[3] The insurance motive is the most common explanation for incumbent wage rigidity, which goes back at least to Azariadis (1975) and Baily (1974). Therefore, in my model, wage rigidity is an outcome of optimal contracts and does not stem from unexplained inefficiencies.

from other firms. In the presence of aggregate shocks arriving in the second period, the incumbent wage contract plays the role of insurance. Firms need to balance the provision of insurance and incentivizing the workers to stay with the firm. At the same time, firms also create new jobs to attract workers from a pool of unemployed and employed workers. The wage distributions of incumbent workers and new hires, as well as distribution of vacancy creation, endogenously respond to aggregate shocks as an equilibrium outcome.

Section 1.3 then characterizes the decentralized equilibrium. Up to a first-order approximation, I show that the equilibrium can be characterized as the solution to a system of ordinary differential equations (ODEs), in which each firm on the job ladder only cares about about the wages and hiring decisions of their neighboring competitors, not the entire distribution. This allows me to derive two main results on wage rigidity and unemployment fluctuations.

The first main result is that wages are endogenously rigid (i.e., they respond less than the aggregate productivity) not only for incumbent workers but also for new hires. The fact that incumbent wages are rigid is intuitive: firms optimally provide some insurance to workers. The fact that new hire wages are also rigid, at least for some firms, is more subtle. Using my ODE characterization of the equilibrium, I show using simple phase diagrams that new hire wages must always feature rigidity at the top of the job ladder. This comes from the fact that at the very top of the job ladder, potential new employers have no incentive to increase wages above what the incumbent firms offer because there would be no additional workers to poach. This extremely strong strategic complementarity spills over toward lower job ladder rungs, and the wages are asymptotically rigid regardless of functional forms or parameter values. This result provides an explanation for the recent evidence on new hire wage rigidity.[4]

My second main result is that the wage rigidity of incumbent workers, rather than new hires, is the critical determinant of fluctuations in job creation. In fact, in this two-period model, the aggregate response of vacancy creation *only* depends on incumbent wage responses. This implies that despite the fact that my model delivers the endogenous wage rigidity of new hires, it has no consequence on unemployment fluctuations. Moreover, to a first-order approximation, introducing exogenous rigidity in the wages of new hires has no effect, either. In this sense, incumbent wage rigidity is a sufficient statistic for unemployment fluctuations regardless of whether or why wages of new hires are rigid.

[4] Gertler et al. (2020), Hazell and Taska (2019), and Grigsby et al. (2019) show the rigidity in wages of new hires is comparable to that of incumbent wages.

This result is in contrast to the conventional view that in the textbook Diamond-Mortensen-Pissarides (DMP) models, wage rigidity of new hires is the only source of unemployment volatility. Why are the conclusions strikingly different? My result is the consequence of a combination of two assumptions: on-the-job search, as emphasized earlier, but also wage posting. The presence of on-the-job search implies that the incumbent wage rigidity does affect job creation because it affects the prospective for poaching. Wage posting further implies that any rigidity in the wages of new hires has no first order effect on the profitability of vacancy posting because of the envelope theorem: since firms set the posted wage optimally as a trade-off between hiring more workers and higher costs, any (non-)movement in posted wages has no first order effect on the incentive to create jobs.[5]

As noted earlier, incumbent wage rigidity in my model is not exogenously imposed, but, rather, is derived from a firm's motive to insure workers. This implies that privately optimal risk-sharing contracts between firms and workers drive the unemployment fluctuations. When workers are more risk-averse, the unemployment rate becomes more volatile because firms provide more insurance. This result challenges the consensus in the literature that wage rigidity derived from long-term contracting should not drive unemployment fluctuations in the canonical models of labor markets (Barro, 1977; Rudanko, 2009). Accounting for on-the-job search is crucial for reaching a starkly different conclusion. In fact, I show that in a version of my model without on-the-job search, unemployment volatility is invariant to the workers' risk aversion.

In Section 1.4, I build on the above insights to consider two extensions of the model. The first one focuses on the introduction of fairness constraints that tie wages of incumbent workers and new hires within a firm.[6] With fairness constraints, firms have to use the same wage to provide insurance for incumbent workers and to attract new hires. As a result, new hire wages become more rigid, but incumbent wages become more flexible relative to the case without such constraints. The more flexible incumbent wages, in turn, reduces unemployment volatility because wage rigidity of incumbent workers, rather than new hires, are what matters for job creation in my model. This implication is

[5]The importance of new-hire wage rigidity is claimed mainly in the context of the Diamond (1982); Mortensen (1982); Pissarides (1985) models, which not only abstract from on-the-job search but also assume wage bargaining. If wages are bargained, firms would prefer to pay wages as low as possible so long as workers accept the job. Since profits are strictly decreasing in wages, (non-)movements in wages have a first order effect on profits. This brings the issue of whether wage posting or wage-bargaining is a more realistic assumption. Existing survey evidence (Hall and Krueger, 2012; Faberman et al., 2020) suggests that wage posting is more prevalent, which is consistent with my model.

[6]Such constraints arise from social norms that workers who perform the same job should be paid the same. The presence of such social norms are documented empirically (Card et al., 2012; Breza et al., 2018; Dube et al., 2019).

the opposite of the conventional view in the previous literature that fairness constraints increase the volatility of unemployment.[7] The contrast comes from the fact that, in many existing models, more rigidity in new hire wages increases the unemployment volatility, while more flexibility in incumbent wages has no consequence.

The second extension considers the introduction of government-provided insurance. The government makes a transfer to workers during recessions and taxes workers during booms. I show that such public insurance reduces unemployment fluctuations by crowding out firm insurance. Because now that the government provides insurance, incumbent firms need to provide less of it. Consequently, incumbent wages become more flexible, which in turn reduces unemployment volatility. This exercise also clarifies the source of unemployment volatility in my model: it comes from the fact that only incumbent firms can provide insurance to workers — workers cannot write contracts with potential new employers. If workers could write contracts with potential new employers, which is in principle what the government is doing here, the unemployment volatility would disappear.

Section 1.5 turns to the efficiency of the decentralized equilibrium. As in Burdett and Mortensen (1998), I assume that firms can commit to the wage contract, but workers cannot. Therefore, when the potential new employers post vacancies, they do not internalize how their offers affect the outside option of incumbent workers, and in turn, the contracts of incumbent jobs.

First, I show that firms tend to make too aggressive wage offers as long as workers are strictly risk-averse. The planner improves welfare by forcing all firms to offer lower wages. This intervention reduces the consumption dispersion of all workers by reducing its upward potential. As workers prefer smooth consumption profiles, this makes it cheaper for incumbent firms to deliver the same utility to workers, leading to Pareto improvement. Moreover, the externality is larger for more productive firms because their high wage offers contribute most to enlarging workers' consumption dispersion. I next show that, through the same externality, the number of vacancy postings is excessive. Productive firms especially tend to over-create jobs because their vacancies distort incumbent wage contracts the most.

Next, I discuss the efficiency in the presence of aggregate risk. An important implication of my framework is that wage rigidity is not necessarily inefficient because it insulates workers from aggregate risk. In fact, I show that new hire wages are always

[7] Such views are informally described by Bewley (1999). Gertler and Trigari (2009); Snell and Thomas (2010); Gertler et al. (2020); Rudanko (2019) formalize such views. It has also been common to impose fairness constraints in wage posting models since the seminal work of Burdett and Mortensen (1998). My result clarifies the role played by such constraints within this class of models.

too flexible relative to the social optimum. This is the case for two reasons. First, competition to attract workers excessively increases the workers' consumption fluctuations. Second, flexibility in new hire wages exacerbates cyclical misallocation. As the wages of incumbent workers respond less than the wages of new hires, workers can flow from more productive firms to less productive firms in booms and reject the offers from more productive firms in recession, manifesting here as misallocation of labor. Forcing new hire wages to respond less improves the allocative efficiency. This is in contrast to the wage rigidity studied in the canonical models of labor markets (e.g., **?**Hall and Milgrom, 2008). There, wages are too rigid, and welfare can be improved by making wages more flexible.

Section 1.6 concludes by exploring the quantitative importance of the mechanisms described above in a generalized version of the baseline model with continuous time and infinite horizon. Methodologically, I propose a new computational algorithm that starts from the same ODE representation of the decentralized equilibrium as in the baseline two-period model. This allows me to construct equilibria by starting with a guess of the wage that the least productive firms offer, which is the reservation wage, and then to compute recursively the wages along the entire distribution by computationally climbing up the job ladder. Instead of having to solve infinite dimensional fixed point problems, I only need to solve a fixed-point in terms of the sequence of market tightness and reservation wages, which are low dimensional problems. Building on a recent contribution by Auclert, Bardóczy, Rognlie, and Straub (2019), I exploit sequence-space Jacobians to solve this fixed-point, which typically takes less than a few seconds to compute the transition dynamics.

Quantitatively, I find that the wage rigidity of incumbent workers caused by the insurance motive alone generates a 20% dampening of wage responses of new hires and accounts for 20% of the unemployment volatility observed in the data. Contrary to the conventional wisdom, imposing fairness constraints dampens the volatility of unemployment by 70%. Different to the two-period model, new hire wage rigidity plays a role in unemployment fluctuations, but I find that incumbent wage rigidity remains the dominant source of the fluctuations. Finally, I show that the type of wage rigidity that matters for unemployment fluctuations in my model is very different from that in the textbook Diamond-Mortensen-Pissarides model. This comes from the fact that the Burdett and Mortensen (1998) model features dynamic competition in the labor market, while such competition is absent in the DMP model.

Related Literature

This paper relates to six strands of the literature. First, it relates to the literature that puts emphasis on the new hire wage rigidity while (implicitly or explicitly) de-emphasizing the role of incumbent wage rigidity; this includes Barro (1977), Pissarides (2009), Haefke et al. (2013), and Rudanko (2009). The latter three papers make a specific point that in the textbook Diamond-Mortensen-Pissarides models, what matters for the incentive to create jobs is the presented discounted value of wage payments to new hires; thus, the response of incumbent wages to aggregate shocks themselves are irrelevant for fluctuations in vacancy creation. These papers abstract from on-the-job search, and hence they mechanically shut down any meaningful interaction between incumbent wages and labor market dynamics. Among them, perhaps the most closely related paper is Rudanko (2009). Like my paper, she micro-founds the incumbent wage rigidity as risk-neutral firms providing insurance to risk-averse workers. She demonstrates that it barely affects unemployment fluctuations compared with a model with risk-neutral workers. Contrary to Rudanko's (2009) findings, I show that the insurance motive does drive unemployment fluctuations once on-the-job search is taken into account.

Since the emergence of the above papers, subsequent literature has measured and modeled new hire wage rigidity. While Haefke et al. (2013), Kudlyak (2014), and Basu and House (2016) document strong pro-cyclicality of new hire wages, more recent papers, Gertler, Huckfeldt, and Trigari (2020), Hazell and Taska (2019), and Grigsby et al. (2019), have found weak cyclicality. The controversy comes from the difficulty in adjusting worker and job compositions that change over business cycles. In contrast, measuring incumbent wage cyclicality does not suffer from such problems, and there is a widely held consensus that incumbent wages are fairly rigid over business cycles (see Grigsby, Hurst, and Yildirmaz (2019) for the most recent evidence). The implication of my theory is that what is less controversial is what matters the most.

Theoretically, several papers have proposed mechanisms that generate endogenous new hire wage rigidity. In Menzio and Moen (2010), firms can commit to wage contracts to insure incumbent workers, but cannot commit not to fire them. This asymmetric commitment technology implies that firms have an incentive not to lower the wages of new hires to avoid replacing incumbent workers with new hires. Although my model is close in sprit in deriving incumbent wage rigidity from firm insurance, the underlying mechanisms are entirely different. For example, in Menzio and Moen (2010), it is important that a firm that posts a vacancy and a firm with incumbent workers are the same firm, but it is not in my framework. In Kennan (2010), workers do not ask for higher wages in expansions because they do not know whether the firm's productivity increased or not.

I provide another mechanism that relies on strategic complementarity in wage setting. This is a natural mechanism to explore because search friction with on-the-job search implies that firms compete for a worker in an imperfectly competitive labor market. In this sense, my paper also relates to the recent papers on strategic complementarity in price settings in oligopolistic product markets (Mongey, 2017; Wang and Werning, 2020).

A more popular and simpler way to generate new hire wage rigidity is to impose fairness constraints, together with other assumptions that generate incumbent wage rigidity. Among others, Menzio (2004), Gertler and Trigari (2009), Snell and Thomas (2010), and Rudanko (2019) pursue this approach. They all conclude that such a constraint amplifies unemployment fluctuations because in those models, new hire wage rigidity, rather than incumbent wage rigidity, is the key source of fluctuations. By contrast, I show that with on-the-job search, such a constraint dampens unemployment fluctuations.

There are models in which incumbent wage rigidity matters for unemployment fluctuations. Schoefer (2016) adds financial frictions into Diamond-Mortensen-Pissarides models and shows that incumbent wage rigidity can tighten financial constraints in recessions. Bils, Chang, and Kim (2016) add endogenous effort choice by workers. In their model, if incumbent wages are too high in recessions, incumbent workers provide too much effort, which reduces the value of the additional workforce. Eliaz and Spiegler (2014) and Carlsson and Westermark (2016) studies the role of incumbent wage rigidity in job destruction. In contrast to these papers, I provide a simple and empirically well grounded channel that operates through job creation.

The second strand of literature to which this paper relates emphasizes the role of on-the-job search in business cycle dynamics. There are three approaches in this line of research. The first approach adopts directed search and wage posting (competitive search) (Menzio and Shi, 2011; Schaal, 2017; Baley, Figueiredo, and Ulbricht, 2019). The second approach is to assume a random search together with wage bargaining (Lise and Robin, 2017; Moscarini and Postel-Vinay, 2018, 2019; Bilal et al., 2019). A third approach, which I pursue in this paper, is to assume random search and wage posting in the tradition of Burdett and Mortensen (1998) and Burdett and Coles (2003) (**?**Moscarini and Postel-Vinay, 2016b; **?**).[8] All these papers feature risk-neutral workers, thereby flexible wages, so they do not speak to the issues studied here. Relative to this strand of literature, I intro-

[8]While papers using this framework to study long-run wage and employment distribution is abundant (van den Berg and Ridder, 1998; Bontemps et al., 1999; Engbom and Moser, 2017; Heise and Porzio, 2019, just to name a few), the literature on transition dynamics is far more scarce. See also Yamaguchi (2010); Bagger et al. (2014); Jarosch (2015); Caldwell and Harmon (2019); Karahan et al. (2019); Engbom (2019), which study the role of on-the-job search in the long-run wage and firm dynamics using the bargaining framework.

duce risk-averse workers and aggregate risk into Burdett and Mortensen's (1998) models to study the nature and the consequence of wage rigidity.[9] Burdett-Mortensen model is particularly well-suited to study these issues because the model has a well-defined notion of wages.

Several other papers explore alternative mechanisms whereby the presence of on-the-job search amplifies the business cycle through the changes in aggregate search efficiency. Eeckhout and Lindenlaub (2019) study a model in which the pro-cyclical job search effort by employed workers leads to self-fulfilling fluctuations in the presence of worker sorting. Engbom (2020) shows that cyclical changes in the composition of employed and unemployed job searchers amplify separation shocks due to greater applications from the latter. I add to this literature by providing a novel mechanism through which the presence of on-the-job search amplifies business cycles.

Third, I build on the long tradition of the literature that micro-founds incumbent wage rigidity as insurance provided by firms. Azariadis (1975) and Baily (1974) are early contributions on this. Harris and Holmstrom (1982) add limited commitment to the workers' side, and this mechanism leads to downward wage rigidity.[10] Beaudry and DiNardo (1991) test its prediction in the data. Rudanko (2009) and Lamadon (2016) embed the mechanism into a search-and-matching labor market. My paper contributes to this literature by demonstrating that such insurance not only explains the wage dynamics, but also increases the volatility of unemployment. Since I focus on the first order approximation around the steady-state, I do not study the non-linear effect such as downward nominal wage rigidity that Harris and Holmstrom (1982) emphasize. However, I conjecture that the non-linear dynamics of my model features downward wage rigidity, which I leave for future work.

Fourth, my paper relates to a series of papers on Shimer (2005) puzzle: i.e., the textbook Diamond-Mortensen-Pissarides models cannot generate unemployment volatility comparable to the data. As mentioned before, many papers rely on new hire wage rigidity (e.g., **?**Dupraz, Nakamura, and Steinsson, 2019). As summarized by Ljungqvist and Sargent (2017), many other solutions rely on increasing the sensitivity of profits to labor productivity by making profits small (e.g., Hagedorn and Manovskii, 2008; Hall and Milgrom, 2008; Pissarides, 2009). However, Chodorow-Reich and Karabarbounis (2016) criticize these approaches by showing that much of the amplifications disappear if one assumes that the outside options for unemployed workers are equally as cyclical as labor

[9] **?** has an extension with exogenous wage rigidity in the form wage adjustment costs à la Rotemberg (1983) but does not separate the role played by rigidity of incumbent workers and new hires.

[10] Thomas and Worrall (1988) further extend this to an environment with two-sided limited commitments.

productivity, and they provide evidence for this. In keeping with this evidence, I adopt the assumption that outside options of unemployed workers scale with labor productivity. Hall (2017), Borovička and Borovičková (2018), Kehoe, Lopez, Midrigan, and Pastorino (2019), and Martellini, Menzio, and Visschers (2020) explore whether movements in discount rates or risk premium explain unemployment volatility.[11] My contribution to this strand of the literature is that incumbent wage rigidity and on-the-job search, which are uncontroversial features of the data, can help resolve the Shimer puzzle.

Fifth, I build on the recent developments on the computation of the transition dynamics of heterogenous agent models. It is widely believed that solving the transition dynamics of the Burdett-Mortensen model is challenging because the endogenous distribution enters as a state variable. **?** propose one methodology under a set of restrictive assumptions, and **?** applies Reiter (2009) approach to approximate the distribution as a low dimensional object. I add to this literature by providing a fast approach to compute the transition dynamics that is accurate to a first order in the size of aggregate shocks, without the need to approximate the distribution. The key idea is that firms do not care about the entire distribution when solving their decision problems. Other than a small number of aggregate variables, such as the market tightness and reservation wages, firms only care about their neighboring competitors. This implies that the equilibrium solution boils down to solving a system of ODEs, rather than infinite dimensional fixed point problems. Therefore, I only need to solve for a sequence of aggregate market tightness and reservation wages that is consistent with equilibrium. I extend the sequence-space Jacobian approach by Auclert et al. (2019) in solving the sequence of aggregates to efficiently compute the equilibrium.

Sixth, my paper touches on a growing strand of the literature on theoretical models of monopsony in the labor market (Manning 2003; Berger et al. 2019; Jarosch et al. 2019; Lamadon et al. 2019; Gouin-Bonenfant 2020). In my model, firms exercise monopsony power because of search frictions. The presence of market power is necessary to study wage rigidity because under perfect competition, wages always respond one-for-one with aggregate productivity. While the literature typically focuses on how the monopsony power shapes the level of wages, I shed light on how the monopsony power shapes the pass-through of aggregate productivity shock through endogenous changes in wage markdowns.

Layout. The rest of the paper is organized as follows. Section 1.2 describes the basic two-period model. Section 1.3 provides qualitative insights on why wages are rigid and

[11] See also Yashiv (2000) and Mukoyama (2009).

what this implies for unemployment fluctuations. Section 1.4 considers two extensions to study the implications of fairness consideration in wage setting and public insurance. Section 1.5 highlights the inefficiency of the model. Section 1.6 quantitatively explores the mechanisms by extending the basic model to an infinite horizon and continuous time setup. Section 1.7 concludes.

1.2 A Job Ladder Model with Risk Averse Workers and Aggregate Shocks

I start from a two-period model to derive a number of sharp qualitative insights. Later in Section 1.6, I will turn to the quantification of these results in a continuous time and infinite horizon version of the model. In this section, I describe the model environment and define equilibrium.

1.2.1 Preferences and Technology

Consider an economy with two dates, $t = 0, 1$. In the initial period, $t = 0$, the firms and workers write contracts (to be described later). At $t = 1$, consumption and production take place. There are two states, $s \in \{h, l\}$ at $t = 1$, with different aggregate productivity, A_s, with $A_h \geq A_l$. The aggregate productivity is revealed at the beginning of $t = 1$. The probability for each state is given by $\pi_s = 1/2$ for $s \in \{h, l\}$. In words, there will be either a boom or recession at $t = 1$ with equal probability.

The economy is populated by two types of agents: a unit mass of workers and a unit mass of firms (or entrepreneurs). Workers consume only at $t = 1$, and their preferences are given by

$$\mathbb{E}u(c_1) \quad \text{with} \quad u(c) = \frac{c^{1-\gamma}}{1-\gamma},$$

where $\gamma \geq 0$ corresponds to the relative risk aversion. At $t = 0$, workers are initially divided into two groups: a fraction $1 - \mu$ of incumbent workers and a fraction μ of unemployed. Both types of workers search for a job at the beginning of $t = 1$. Unemployed workers meet with a firm with probability λ_s^U, and incumbent workers meet with probability $\lambda_s^E \equiv \zeta \lambda_s^U$, where $\zeta > 0$ is the relative search efficiency of the employed. A worker faces no search cost. When workers end up being unemployed at $t = 1$, they enjoy home production, which produces $A_s b$ amount of consumption goods, where $b > 0$ is a parameter. Here, the outside option of unemployed scales with the aggregate productivity shock, A_s.[12]

Firms consume and produce only at $t = 1$, and they are risk-neutral,

$$\mathbb{E}c_1^e,$$

where c_1^e is the consumption of entrepreneurs. Each firm has access to production technology that is linear in labor,

$$A_s z l,$$

where l is labor and z is the idiosyncratic productivity. The idiosyncratic productivity is a fixed characteristics of a firm. The cross-sectional distribution is continuous and has a bounded support $[\underline{z}, \bar{z}]$ with $\underline{z} \geq b$. Let $G(z)$ and $g(z)$ denote the cumulative and the probability density function, respectively. Each firm z is exogenously endowed with $\ell_0(z)$ amount of employed workers at $t = 0$.

At $t = 1$, firms choose how much vacancy to post, $v_s(z)$, to attract new workers. The vacancy posting is subject to convex cost, $c_s(v; z)$. I assume the cost of vacancy creation, $c_s(v; z)$ takes the form

$$c_s(v; z) = A_s \bar{c}(z) \frac{v^{1+1/\iota}}{1 + 1/\iota} \tag{1.1}$$

where $\iota > 0$ corresponds to the elasticity of vacancy creation. The assumption that the cost function scales with the aggregate productivity follows Blanchard and Galí (2010).[13] This captures the idea that to recruit workers, existing workers must reduce their time devoted to production, which costs a firm lost output. This assumption ensures that the fluctuations in job creation are not driven by differential productivity growth between the output production and recruitment activity.

Each vacancy will meet a worker with probability λ_s^F. Although I have not described whether a firm that posts a vacancy and a firm with incumbent workers are the same firm or not, the distinction is not important. This is because, in the baseline model, firms decisions are separable between the two because of the constant-returns-to-scale technology.[14] I will refer to a firm with incumbent workers as a incumbent firm and a firm that posts a vacancy as a poaching firm, a potential new employer, or a new hire firm.

Finally, the total number of meetings between firms and workers is given by a constant-returns-to-scale matching technology $\mathcal{M}(\tilde{\mu}, V_s)$. The first input to the matching function is the total efficiency unit of search by workers, $\tilde{\mu} \equiv \mu + \zeta(1 - \mu)$. The second input is

[12]This is consistent with the evidence documented in Chodorow-Reich and Karabarbounis (2016).

[13]See also Shimer (2010), or more recently Kehoe et al. (2019).

[14]It will be important when I later introduce the fairness constraint that exogenously tie incumbent and new hire wages.

the total amount of vacancy postings, $V_s \equiv \int v_s(z)dG(z)$. Search is random. When firms meet with a worker, the worker is an unemployed with probability $\chi \equiv \frac{\mu}{\tilde{\mu}}$ and is employed with probability $1-\chi$. Likewise, when workers meet a firm, the probability that the firm has productivity z is given by $\frac{v_s(z)}{V}g(z)$.

1.2.2 Contracts and Markets

Firms that have incumbent workers at $t=0$ write state-contingent wage contracts with workers at $t=0$. A worker employed by a firm with productivity z is endowed with promised utility $\bar{W}_0(z)$: the firm has to deliver expected utility at least $\bar{W}_0(z)$ through the contract. Although this is an exogenous parameter, one can think of this as an object that is determined in the past when a worker is hired. In fact, this will be the case in the infinite horizon version of my model studied later.

The contract specifies the wage payments in each state $\{w_{0h}(z), w_{0l}(z)\}$, which are to be paid at $t=1$. There are two assumptions in the contract. First, workers cannot commit to the contract, so they are free to leave firms when receiving a better offer. Workers are also free to quit and become unemployed. In contrast, I assume firms have full commitment. Second, the contract cannot depend on the outside offers that workers received. A justification for this assumption is that outside offers are not verifiable. These two assumptions are common in the wage posting literature (e.g., Burdett and Coles, 2003).

At $t=1$, when firms post a vacancy after the realization of the aggregate productivity shock, they also post wage, $w_{1s}(z)$. Firms commit to the wage contract; therefore, the offer is a take-it-or-leave-it offer.

Timing. The timing of the model is described in Figure 1-2. First, incumbent workers and firms write contracts before the realization of aggregate productivity shock. Then, after the observing the aggregate productivity, firms post a vacancy. Next, a matching market opens, and firms and workers meet with each other. Workers either accept or reject the offer, and production and consumption take place.

1.2.3 Equilibrium Definition

In equilibrium, incumbent firms at date 0 maximize expected profits taking the wage distribution induced by wage offers by potential new employers at date 1 and the reservation wage of their workers as given, whereas new employers at date 1 maximize expected profits taking the distribution of wage contracts by incumbent firms at date 0 and the reservation wages of workers as given.

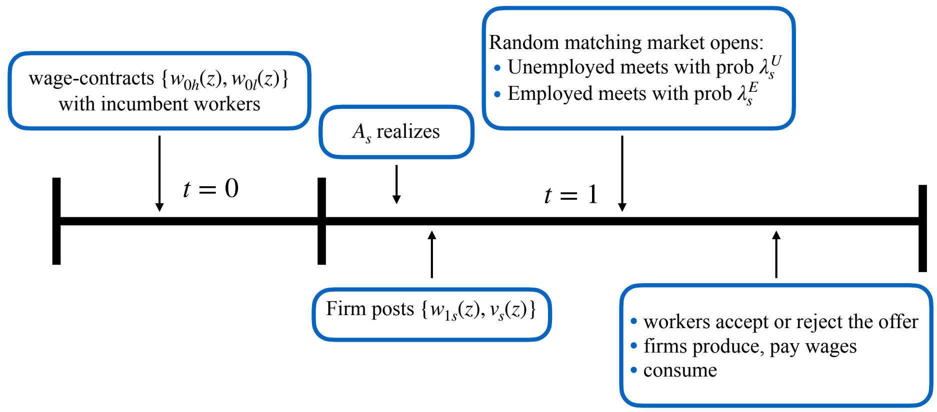


Figure 1-2: Timing

Note: Figure 1-2 describes the timing assumption of the model.

Incumbent firms' optimal contracting problem. Incumbent firms take the new hire wage distribution, which I denote as $F_{1s}(w)$, and the meeting probability λ_s^E as given. The incumbent wage contracts of firms with productivity z solves the following problem:

$$\begin{aligned} &\max_{\{w_{0h}, w_{0l}\}} \sum_{s\in\{h,l\}} \pi_s (A_s z - w_{0s})(1 - \lambda_s^E + \lambda_s^E F_{1s}(w_{0s})) \\ \text{s.t.} \quad &\sum_{s\in\{h,l\}} \pi_s \left[(1-\lambda_s^E) u(w_{0s}) + \lambda_s^E \int \max\{u(w_{0s}), u(\tilde{w})\} dF_{1s}(\tilde{w}) \right] \geq \bar{W}_0(z) \\ &w_{0h} \geq A_h b, \quad w_{0l} \geq A_l b. \end{aligned} \tag{1.2}$$

where $\bar{W}_0(z)$ is the promised utility of firm z. The objective function is the expected profits, taking into account the probability of workers being poached. With probability $1 - \lambda_s^E$, a worker does not receive an outside offer, and with probability λ_s^E, s/he receives an offer. If the offer is lower than the current wage, w_{0s}, which happens with probability $F_{1s}(w_{0s})$, s/he find it optimal to stay with the current firm. Otherwise, the worker leaves for the new firm. The constraint guarantees the worker's expected utility, which takes into account that the wage payments in the new firm, is greater than the predetermined promised utility. The constraint $w_s \geq A_s b$ captures the fact that workers can always quit and engage in home production, which firms never find it optimal to let happen.

The key trade-off in this contracting problem is insurance versus incentive. Firms would like to insure workers as the worker's utility is concave, but if firms do too much insurance, firms will not be able to keep workers during good times, while tending to keep workers during bad times. The optimal wage contract strikes a balance between the two.

New hire firms' profit maximization. The new hire firms take the incumbent wage distribution, which I denote as $F_{0s}(w)$, and the meeting probability λ_s^F as given. A firms with productivity z solves the following profit maximization problem:

$$\begin{aligned} \max_{v_s, w_{1s}} & (A_s z - w_{1s}) \lambda_s^F \left(\chi + (1-\chi) F_{0s}(w_{1s}) \right) v_s - c_s(v_s; z), \\ \text{s.t.} \quad & w_{1s} \geq A_s b \end{aligned} \tag{1.3}$$

where $\chi \equiv \mu/(\mu + \zeta(1-\mu)$ is the share of unemployed. Since firms always find it optimal to offer at least $A_s b$ because $z \geq b$ for all z, the unemployed always accept an offer. The remaining fraction $1-\chi$ of workers are already employed, and they accept the offer with probability $F_{0s}(w_{1s})$. Again, the constraint $w_{1s} \geq A_s b$ captures the fact that the firms always find it optimal to offer wages that at least attract the unemployed. The firm chooses the wage offers and the amount of vacancy to maximize expected profits after observing the aggregate productivity shock, A_s.

The equilibrium definition is as follows:

Definition 1. *Equilibrium consists of incumbent firms' wage contracts, $\{w_{0s}(z)\}$, and new hire firms' wage offers and vacancy postings, $\{w_{1s}(z), v_s(z)\}$, associated wage distribution $\{F_{0s}(w), F_{1s}(w)\}$ and meeting probabilities λ_s^E and λ_s^F such that (i) given the entrants' wage distribution F_{1s} and λ_s^E, incumbent wages $\{w_{0s}(z)\}$ solve (1.2), (ii) given the incumbents' wage distribution F_{0s} and λ_s^F, $\{w_{1s}(z), v_s(z)\}$ solve (1.3), and (iii) the wage distribution is consistent with the equilibrium wage strategies: $F_{0s}(w) = \frac{1}{1-\mu} \int_{z: w \geq w_{0s}(z)} \ell_0(z) dG(z)$, $F_{1s}(w) = \int_{z: w \geq w_{1s}(z)} (v_s(z)/V_s) dG(z)$; (iv) the meeting probabilities are given by the matching function, $\lambda_s^E = \zeta \frac{\mathcal{M}(\tilde{\mu}, V_s)}{\tilde{\mu}}$ and $\lambda_s^F = \frac{\mathcal{M}(\tilde{\mu}, V_s)}{V_s}$.*

Equilibrium is a fixed point in terms of the wage distribution (and matching probabilities). Each individual firm is infinitesimal and takes the wage distribution of competitors as an input to their decision problems. The optimization problems give the wage distribution as an outcome, which has to be consistent with the distribution that firms took as an input.

1.2.4 Discussion of the main assumptions

The assumption that workers have limited commitment and firms have full commitment is standard in the literature, which at least goes back to Harris and Holmstrom (1982) or more recently Lamadon (2016). The justification for this assumption is that firms have reputation costs of reneging the contract, while workers arguably have much less costs in doing so.

The assumption on wage posting also deserves some discussion, since it plays important roles in many of my analyses. Another common approach is to use a sequential auction protocol as in **?**. Perhaps, both wage setting protocols co-exist in a real world, but existing empirical evidence suggests wage posting is more prevalent. Survey evidence shows two-thirds of workers do not bargain over wages (Hall and Krueger, 2012; Faberman et al., 2020). Faberman et al. (2020) also document that counter-offers are rather rare: only 12% of offers that workers receive are countered by their employers. Moreover, recent evidence by Addario et al. (2020) shows that workers' wages display little dependence on past jobs, contrary to the prediction of sequential auction protocol models, but this fact is consistent with wage posting models.

Another assumption that I impose is random search, as opposed to directed search (Moen, 1997; Acemoglu and Shimer, 1999; Menzio and Shi, 2011). Both assumptions are equally common in the literature, and the reality should lie somewhere in between. It is thus an important open question to study wage rigidity in an environment with directed search, which I leave for future work.[15]

1.3 Wage Rigidity and Unemployment Fluctuations

This section studies equilibrium wage rigidity and its consequences for unemployment fluctuations. Section 1.3.1 describes the solution approach. Section 1.3.2 presents the main result on wage rigidity, and Sections 1.3.3-1.3.4 study what type of wage rigidity matters for unemployment fluctuations.

1.3.1 Solution Approach

I first characterize the equilibrium by considering the optimality conditions of firms. The first order necessary condition associated with incumbent firm's optimization problem

[15] Bilal et al. (2019) argue that the recent evidence on worker and firm flows by Bagger et al. (2020) is consistent with random search but not necessarily with directed search.

(1.2) is

$$\begin{aligned}&-(1-\lambda_s^E+\lambda_s^E F_{1s}(w_{0s}(z)))+(A_s z-w_{0s}(z))\lambda_s^E F_{1s}'(w_{0s}(z))\\&\quad+\eta(z)\left[(1-\lambda_s^E)u'(w_{0s}(z))+\lambda_s^E F_{1s}(w_{0s}(z))u'(w_{0s}(z))\right]=0\end{aligned} \tag{1.4}$$

where $\eta(z)$ is the Lagrangian multiplier on the promise-keeping constraint. The first order conditions associated with new hire firm's optimization problem (1.3) is

$$(1-\chi)F_{0s}'(w_{1s}(z))(A_s z-w_{1s}(z))-(\chi+(1-\chi)F_{0s}(w_{1s}(z)))=0 \tag{1.5}$$

$$(A_s z-w_{1s}(z))\lambda_s^F(\chi+(1-\chi)F_{0s}(w_{1s}(z)))-c_s'(v_s(z);z)=0. \tag{1.6}$$

In deriving these conditions, I have assumed that the wage distributions, F_{0s} and F_{1s}, are differentiable. I will later confirm that that they are as such in my analysis.

Symmetric equilibrium without aggregate risk. Although I have already simplified the model by focusing two-period model, analyzing the model still poses a challenge. As is clear from the equilibrium definition, the problem involves multiple infinite dimensional objects (incumbent and new hire wage distribution, as well as vacancy distribution in each state). It is intractable not only analytically but even computationally. To the best of my knowledge, there is no efficient algorithm to solve the model non-linearly because the equilibrium does not have a convenient property such as contraction mapping.

To overcome the difficulty, I propose a tractable solution approach. I consider a perturbation of a particular equilibrium with respect to the aggregate risk. I first focus on a particular parametrization that features the following properties: (i) zero aggregate risk, $A_h = A_l \equiv A$, and (ii) symmetry between incumbent and new hire wages, $w_0(z) = w_1(z) \equiv w(z)$, where I dropped the s subscript as two states are the same. These properties will naturally arise in the steady-state of an infinite horizon setup that I will study later. For this reason, I call this equilibrium as the steady-state equilibrium.

After imposing (i) and (ii), the new hire firms first-order condition (1.5) becomes

$$(1-\chi)F_0'(w(z))(Az-w(z))-(\chi+(1-\chi)F_0(w(z)))=0. \tag{1.7}$$

Here F_0', is well-defined as there cannot be a mass point in the incumbent wage distribution. If there was a mass point, then one of the incumbent or the new hire firms at the mass point can raise wages by a small amount and discontinuously increase the profits, which contradicts with the optimality of wage setting. Moreover, $w(z)$ is strictly increasing because the objective function (1.3) is strictly log-supermodular in (z, w). Because the wages are monotone, it follows that $\hat{F}_0(z) \equiv F_0(w(z)) = \frac{1}{1-\mu}\int^z \ell_0(z)dG(z)$, so

$\hat{F}_0'(z) = F_0'(w(z))w'(z)$, where $\hat{F}_0(z)$ is the employment-weighted productivity distribution (i.e., the share of workers employed in firms with productivity below z). Using these expressions, we can rewrite (1.7) as a single ODE:

$$(1-\chi)\hat{F}_0'(z)(Az - w(z)) - \left(\chi + (1-\chi)\hat{F}_0(z)\right) w'(z) = 0 \tag{1.8}$$

with the boundary condition $w(\underline{z}) = Ab$ because the least productive firms can only hire from a pool of unemployed. The solution is

$$w(z) = \frac{\chi Ab + (1-\chi)\int_b^z A\tilde{z}d\hat{F}_0(\tilde{z})}{\left(\chi + (1-\chi)\hat{F}_0(z)\right)}, \tag{1.9}$$

which corresponds to the employment weighted average productivity level conditional on productivity below z. As is standard in Burdett and Mortensen (1998) models, firms exercise monopsony power, $w(z) < Az$, because of search frictions. Appendix 1.8.1 also shows that the second-order condition is satisfied.

Given (1.9), the optimal vacancy solves

$$(Az - w(z))\lambda^F\left(\chi + (1-\chi)F(w(z))\right) = c'(v(z); z). \tag{1.10}$$

The meeting probabilities are given by

$$\lambda^F = \frac{1}{V}\mathcal{M}(\tilde{\mu}, V) \quad \lambda^E = \zeta\frac{\mathcal{M}(\tilde{\mu}, V)}{\tilde{\mu}} \quad \text{with } V = \int v(z)dG(z). \tag{1.11}$$

Finally, I have to guarantee that the incumbent firms find it optimal to offer $w(z)$. I can always guarantee this if the promised utility is appropriately chosen and if the promise-keeping constraint is binding. The promise-keeping constraint is always binding as long as λ^E is small enough, which is the case for sufficiently small ζ. Intuitively speaking, if incumbent firms do not face too tough competition from being poached, they would like to exercise monopsony power to lower wages as much as they can. Then, one can appropriately choose $\bar{W}_0(z)$ so that the incumbent firms need to offer $w_0(z) = w(z)$. I summarize the discussion as follows:

Lemma 1. *Suppose $A_h = A_l$ and the relative search efficiency of the employed, ζ, is sufficiently small. Then, there exists $\{\bar{W}_0(z)\}$ under which the equilibrium wage strategy is symmetric between incumbent and new hire firms, $w_0(z) = w_1(z) \equiv w(z)$. In such an equilibrium,*

$\{w(z), v(z), \lambda^F, \lambda^E\}$ *are given by (1.9), (1.10) and (1.11), and*

$$\bar{W}_0(z) = (1-\lambda^E)u(w(z)) + \lambda^E \int \max\{u(w(z)), u(w(\tilde{z}))\}(v(\tilde{z})/V)dG(\tilde{z}).$$

All the proofs are collected in Appendix 1.8. I next turn to the analysis with aggregate risk by taking a first order perturbation around the above symmetric equilibrium.

1.3.2 Wage Rigidity

I introduce aggregate risk into the economy by assuming $A_h > A_l$. I consider a first order perturbation that is a mean preserving spread around $A_h = A_l \equiv A$, $\ln A_h = \ln A + d\ln A$ and $\ln A_l = A - d\ln A$. I let variables with hat denote the log deviation from the steady-state equilibrium, $\hat{x} \equiv d\ln x$.

Characterization

The following lemma shows that the responses are symmetric between two states:

Lemma 2. *In the presence of small aggregate risk, $\hat{A} > 0$, to a first order, the equilibrium is symmetric between two states: $\hat{w}_{1h}(z) = -\hat{w}_{1l}(z) \equiv \hat{w}_1(z)$, $\hat{w}_{0h}(z) = -\hat{w}_{0l}(z) \equiv \hat{w}_0(z)$, $\hat{v}_h(z) = -\hat{v}_l(z) \equiv \hat{v}(z)$, $\hat{V}_h = -\hat{V}_l \equiv \hat{V}$, $\hat{\lambda}^E_h = -\hat{\lambda}^E_l \equiv \hat{\lambda}^E$, $\hat{\lambda}^F_h = -\hat{\lambda}^F_l \equiv \hat{\lambda}^F$.*

Since the equilibrium conditions are smooth with respect to endogenous variables, the symmetric aggregate productivity shocks induces the symmetric responses. This is useful because we can reduce the number of unknowns by half.

I turn to characterizing the equilibrium responses to the aggregate shock. I first concentrate on the wage responses by assuming vacancies are inelastic. Even in this case, the equilibrium is potentially very complicated because it is an infinite dimensional fixed point problem. New hire firms need to form expectation over the entire incumbent wage distribution and decide where to position their wage rank. Conversely, incumbent firms need to form expectation about entire new hire wage distribution and decide which wage offers they would like to block. These expectations need to be consistent with optimization behaviors. However, it turns out that the equilibrium solution takes a very simple form, as the following lemma shows:

Lemma 3. *Assume vacancy creation is inelastic, $\iota = 0$. In the presence of small aggregate risk, $\hat{A} > 0$, to a first order, the equilibrium incumbent wage responses, $\hat{w}_0(z)$, and new hire wage* *responses, $\hat{w}_1(z)$, solve the following two ODEs:*

$$\text{(new hire)} \quad \hat{w}_1(z) = \theta_{1a}(z)\hat{A} + \underbrace{\theta_{1w}(z)\hat{w}_0(z)}_{\text{competition within a job-ladder rung}} - \underbrace{\theta_{1a}(z)\alpha(z)\frac{w(z)}{w'(z)}\hat{w}_0'(z)}_{\text{competition between job-ladder rungs}} \tag{1.12}$$

$$\text{(incumbent)} \quad \hat{w}_0(z) = \theta_{0a}(z)\hat{A} + \underbrace{\theta_{0w}(z)\hat{w}_1(z)}_{\text{competition within a job-ladder rung}} - \underbrace{\theta_{0a}(z)\alpha(z)\frac{w(z)}{w'(z)}\hat{w}_1'(z)}_{\text{competition between job-ladders rungs}}, \tag{1.13}$$

with the boundary conditions, $\hat{w}_1(\underline{z}) = \hat{A}$ and $\hat{w}_0(\bar{z}) = \hat{w}_1(\bar{z})$. The coefficient $\alpha(z) \equiv (Az - w(z))/Az$ is the wage markdown and the other coefficients are such that $\theta_{1a}(z) > 0$, $\theta_{0a}(z) > 0$, $\theta_{1a}(z) + \theta_{1w}(z) = 1$ and $\theta_{0a}(z) + \theta_{0w}(z) \leq 1$, with equality if workers are risk-neutral, $\gamma = 0$, as shown in Appendix 1.8.3.

The two ODEs come from the log linearization of the first order conditions (1.4) and (1.5) and are the best response functions of the firms wage settings. Note that the original best response function of incumbent firm of productivity z depends on F_{1s}, F_{1s}', which in turn depends on the entire functions of $\{w_{1s}(\tilde{z})\}_{\tilde{z}}$. The key observation of Lemma 14 is that to a first-order approximation, the best response of incumbent firm of productivity z only depends on, $w_{1s}(z)$ and $w_{1s}'(z)$, not on the entire function $\{w_{1s}(\tilde{z})\}_{\tilde{z}}$, substantially reducing the dimensionality. To see this, the first order change in cumulative distribution function $F_1(w_0(z))$ is given by

$$dF_1(w_0(z)) = F_1'(w(z))w(z)\left(\hat{w}_0(z) - \hat{w}_1(z)\right)$$

and the first order change in the density function $F_1'(w_0(z))$ is given by

$$dF_1'(w_{0s}(z)) = F_1''(w(z))w(z)\hat{w}_0(z) - \left(F_1''(w(z))w(z) + F_1'(w(z))\right)\hat{w}_1(z) - F_1'(w(z))\frac{w(z)}{w'(z)}\hat{w}_1'(z).$$

That is, the competition remains always local in response to small shocks. Firms do not need to form expectations about the wage offers of firms in significantly different job ladder ranks because they won't affect the labor supply curve. Firms need to only care about how their local competitors will behave.

The term "competition within a job ladder rung" captures how the competitors with exactly the same productivity level affect the labor supply. For example, if the new hire firm with productivity z increases its wages, the incumbent firm with the same productivity z is more likely to be poached. The term "competition between job ladder rungs" captures how the neighboring competitors' wage setting affects the labor supply. For example, if the new hire firm with productivity $z - dz$ increases its wages more than those with productivity z, the incumbent firm with productivity z faces more elastic labor supply because there would now be a greater mass of marginal competitors.

The coefficients on (1.12) and (1.13) cannot be arbitrary and have theoretical restrictions. First, $\theta_{1a}(z) > 0$ and $\theta_{1a}(z) + \theta_{1w}(z) = 1$.[16] That is, the new hire firms' problem is homogenous: if the aggregate productivity increases by 1% and the incumbent firms increase wages by 1%, then it is optimal for them to increase wages just by 1%. In contrast, $\theta_{0a}(z) > 0$ and $\theta_{0a}(z) + \theta_{0w}(z) \leq 1$ with strict inequality if and only if $\gamma > 0$. That is, incumbents' overall wage responses are dampened as long as workers are risk-averse. This is intuitive. Because incumbent firms have incentives to insure workers, they do not want to fluctuate wages too much with the aggregate shocks. Moreover, these coefficients can be expressed as a function of steady-state moments, which have a clear data counterpart. The new hire firms' response $\theta_{1a}(z)$ depends only on wage markdown, $\alpha(z)$, and the elasticity of new hire wage density function, $\eta_{F_0}(z)$. These expressions have a natural counterpart in pass-through literature in the context of product price-settings (see Burstein and Gopinath (2014) for a survey). In the context of product price settings, it is well-known that pass-through of costs to product prices mainly depends on the (i) elasticity of demand function and (ii) super-elasticity of the demand function. Here, $\alpha(z)$ captures the former, and $\eta_{F_0}(z)$ captures the latter. In addition to wage markdown and the elasticity of the density function of incumbent firms, the incumbent firms' response depends also on the elasticity of workers' staying probability and the relative risk aversion.

Since the system consists of two ODEs with two unknowns, we need two boundary conditions. The first boundary condition describes what happens at the bottom of the job ladder, $\hat{w}_1(\underline{z}) = \hat{A}$. Because unemployed workers' outside options scale one for one with the aggregate productivity, the least productive firm, which hires only from a pool of unemployed, also needs to move wages one for one with the aggregate productivity. One may wonder why the same boundary condition does not apply for the incumbent firms at the bottom, $z = \underline{z}$. The reason is that the constraint $w_{0s} \geq A_s b$ strictly binds because the firm would like to insure workers as much as possible, and hence they are no longer in the interior solution. Therefore, the bottom boundary of the incumbent firms is at $\hat{w}_0(\underline{z}^+) \equiv \lim_{z \downarrow \underline{z}} \hat{w}_0(z)$, which is free. The second boundary condition describes the behavior at the top of the job ladder, $\hat{w}_0(\bar{z}) = \hat{w}_1(\bar{z})$. It says that both incumbent and new hire firms must find it optimal to offer exactly the same wages at the top of the job ladder.

[16]Second-order condition implies $\theta_{1a}(z) > 0$ for all z.

If one firm at the top offers strictly higher wages then the other at the top, it can strictly increase profits by slightly lowering wages because it does not affect the labor supply but yet reduces costs. This extremely strong form of strategic complementarity at the top of the job ladder is at the heart of the analysis that comes next.

Equilibrium Wage Rigidity

Having characterized the equilibrium wage responses as a system of ODEs, I am ready to study their properties.

Proposition 1. *Assume the elasticity of vacancy creation, ι, is sufficiently small. If workers are risk-neutral, $\gamma = 0$, then all wages are flexible, $\hat{w}_1(z) = \hat{w}_0(z) = \hat{A}$ for all z. If workers are risk-averse, $\gamma > 0$, then all incumbent wages are rigid, $\hat{w}_0(z) < \min\{\hat{A}, \hat{w}_1(z)\}$ for all z, and new hire wages are rigid at the top of the job ladder, $\hat{w}_1(z) < \hat{A}$ for z close enough to $\bar{z}$.*

The proposition states that with risk-neutral workers, wages for both incumbents and new hires are fully flexible. With risk-averse workers, incumbent wages are rigid. Perhaps more surprisingly, new hire wages are also rigid, at least toward the higher end of the job ladder. Let me turn to an explanation for each of the result, assuming $\iota = 0$. By continuity, the results hold if ι is small enough. Throughout, I often impose the elasticity of vacancy creation, ι, is sufficiently small. This is largely a technical assumption. I have not encountered any counter-example even with large ι. Moreover, it has been common to assume relatively small ι in the literature that builds on Burdett and Mortensen (1998) because as $\iota \to \infty$, the firm size distribution becomes too concentrated to the most productive firm.

As (1.12) and (1.13) consist a system of two ODEs, one can draw a phase diagram to explain the proposition, as I do in the left panels of Figure 1-3-1-5. Let us start with a case of risk-neutral workers, $\gamma = 0$, in the left panel of Figure 1-3. I plot $\hat{w}_0(z)$ on a vertical axis and $\hat{w}_1(z)$ on a horizontal axis.[17] A particular point $(\hat{w}_0(z), \hat{w}_1(z))$ in the phase diagram corresponds to a pair of incumbent and new hire wage responses at job ladder (productivity) z. If the point lies inside the gray square, it means that both incumbent and new hire wages respond less than the aggregate productivity, $\hat{w}_0(z) < \hat{A}$ and $\hat{w}_1(z) < \hat{A}$, or in other words, wages are rigid (sticky). In the figure, I draw two lines, each corresponding to the $\hat{w}_0'(z) = 0$ locus and the $\hat{w}_1'(z) = 0$ locus.

When $\gamma = 0$, the two lines have to go through a point $(\hat{A}, \hat{A})$. Moreover, the $\hat{w}_0'(z) = 0$ locus needs to have slope greater than one because $\theta_{1w}(z) < 1$, and the $w_1'(z) = 0$ locus

[17] Although the two lines are generically moving around depending on z, I study the phase diagram as if the two loci are unchanged for all z. Qualitative properties are unaffected by this consideration, as long as the coefficients are continuous in z, which is the case here.

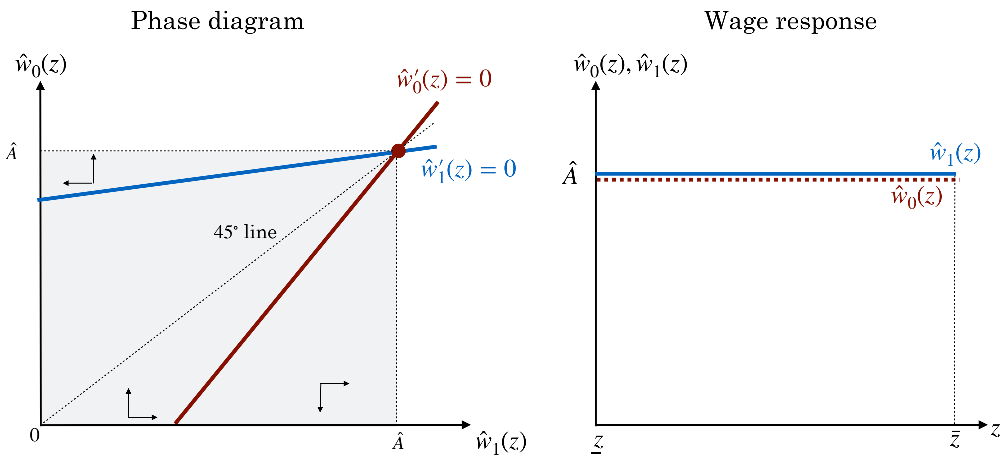


Figure 1-3: $\gamma = 0$

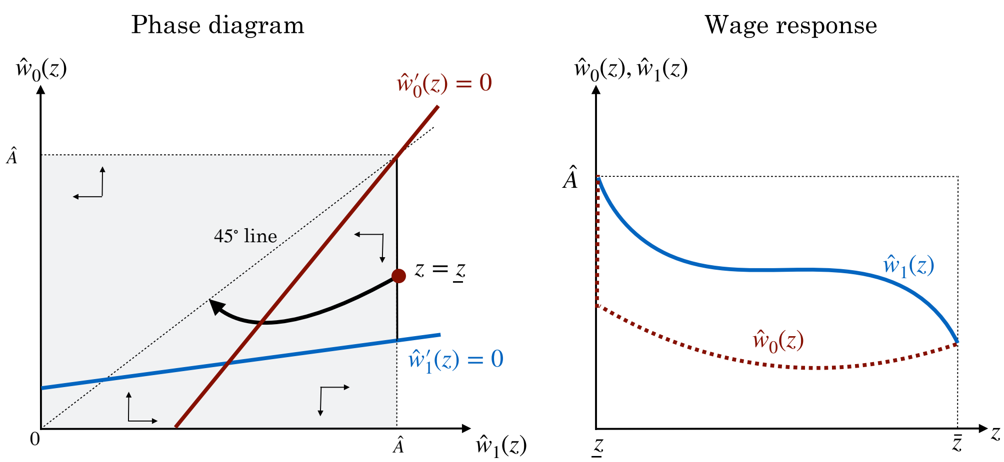


Figure 1-4: $\gamma > 0$ and $\theta_{1w} > 0$

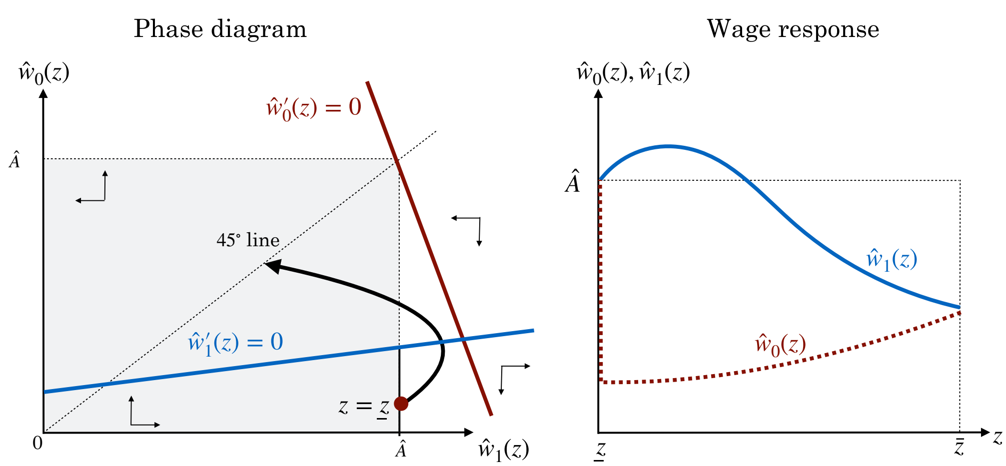


Figure 1-5: $\gamma > 0$ and $\theta_{1w} < 0$

needs to have a slope less than one because $\theta_{0w}(z) < 1$. One of the boundary conditions states that $\hat{w}_1(\underline{z}) = \hat{A}$, so starting from $z = \underline{z}$, it has to originate from somewhere in the vertical line that goes through $(\hat{A}, \hat{A})$. The other boundary condition states that $\hat{w}_1(\bar{z}) = \hat{w}_0(\bar{z})$, so the path needs to end up somewhere in the 45° line. Then, it is immediately clear that the only path that satisfies the two boundary conditions is the one that starts from $(\hat{A}, \hat{A})$ at $z = \underline{z}$ and stays there until it reaches $z = \bar{z}$. That is, both incumbent and new hire wages are fully flexible. The wage response for each z is depicted in the right panel of Figure 1-3. This result is intuitive. With risk-neutral workers, incumbent firms have no incentive to insure workers, so both incumbent and new hire firms' problems are homogenous in aggregate productivity and competitors' wages. Wages just scales up and down with the aggregate productivity.

More interesting cases arise when workers are risk averse, $\gamma > 0$. The left panel of Figure 1-4 shows the phase diagram with $\gamma > 0$ and typical parametrization $\theta_{1w} > 0$. With $\gamma > 0$, the $\hat{w}_1'(z) = 0$ locus uniformly shifts downward compared with $\gamma = 0$. With $\theta_{1w} > 0$, the $\hat{w}_1'(z) = 0$ locus is upward sloping. A path that satisfy the boundary conditions are drawn as a black line: it needs to start with $\hat{w}_1(\underline{z}) = \hat{A}$ and end up on the 45° line. In this case, since the entire path lies inside the gray square, both incumbent and new hire wages are rigid. The wage response at each job ladder is drawn in the right panel of Figure 1-4. Incumbent wages are unresponsive throughout the job ladder. In contrast, new hire wages become less and less responsive as we look at a higher job ladder rank, eventually reaching the same rigidity at the top of the job ladder.

However, Figure 1-4 is not the only possibility. Suppose $\theta_{1w} < 0$. Then the $\hat{w}_0'(z) = 0$ locus is negatively sloped, as shown in the left panel of Figure 1-5. In this case, the path that satisfies boundary condition could be the one depicted as a black line. The right panel of Figure 1-5 shows the corresponding wage response at each job ladder rank. In this case, new hire wages respond more than the aggregate productivity at the lower end of the job ladder, but incumbent wages respond less for the entire job ladder.

The result that incumbent wages are always rigid for $\gamma > 0$ is not surprising. As firms have an incentive to insure workers, they respond less than the aggregate productivity. The reason why new hire wages are also always rigid at the higher end of the job ladder comes from the strategic complementarity in wage setting. job ladder models feature a very extreme form of strategic complementarity at the top, $z = \bar{z}$: no firm wants to set wages strictly above the competitor's one. This strategic complementarity spills over from the top to lower job ladder ranks. If incumbent and new hire firms at the very top, $z = \bar{z}$, set exactly the same wages, firms at a slightly lower-rank, $z = \bar{z} - dz$, also set the similar wages. The reason why the new hire wages can overshoot at the lower-end of the job ladder is as follows. The case $\theta_{1w} < 0$ happens when

$$\eta_{F_0}(z) \geq \frac{1 - 2\alpha(z)}{\alpha(z)},$$

$\eta_{F_0}(z) = \frac{d \ln F_0'(w(z))}{d \ln w}$ is the elasticity of density of wage distributions. This says the elasticity of incumbent wage density function is large enough (but not too large as $\theta_{1a} > 0$ requires $\eta_{F_0}(z) < \frac{2-2\alpha(z)}{\alpha(z)}$). Intuitively speaking, when this is the case, the new hire firms can poach a lot more workers if they increase wages slightly more than the incumbent firms. Therefore they have an incentive to become aggressive in making high wage offers if incumbent wages are not responsive — that is, new hire firms' wage setting is strategic substitutes with respect to incumbents'.

It is worth noting that on-the-job search was the key for new hire wages to feature any kind of rigidity. Without on-the-job search, new hire firms find it optimal to offer outside options of the unemployed, $w_{1s}(z) = A_s b$, so new hire wages are fully flexible, $\hat{w}_{1s}(z) = \hat{A}$. It is precisely the competition for employed workers through which incumbent wage rigidity spills over to new hire wage rigidity.

Relationship to recent evidence on new hire wage rigidity. My result above shows that the two empirically well grounded assumptions, (i) incumbent wage rigidity and (ii) on-the-job search, naturally leads to endogenous new hire wage rigidity, especially at the top of the job ladder. This result provides an explanation for the recently documented empirical evidence. While Haefke et al. (2013) or Kudlyak (2014) originally documented that new hire wages are substantially more cyclical than incumbent wages, more recent evidence that carefully adjusts for the job compositions (Gertler et al., 2020; Hazell and Taska, 2019; Grigsby et al., 2019) shows that new hire wages are much less cyclical than previously thought. However, we tend to lack a theoretical understanding of the underlying mechanisms without imposing ad-hoc constraints on wage setting. My model provides a natural explanation for this. Although there are some other theories of endogenous new hire wage rigidity (Menzio and Moen, 2010; Kennan, 2010), a distinguishing feature of my theory is that it predicts that new hire wages should feature more rigidity at the higher job ladder rank. Consistently with this prediction, Bloesch and Taska (2019) use the data from online vacancies to document that posted wages are much less cyclical for high-wage jobs than low-wage jobs.

Non-strategic incumbent firms. The mechanism that generates endogenous new hire wage rigidity comes from strategic complementarity in wage setting. It relies on the fact *both* incumbent firms and new hire firms are acting strategically what wages to offer workers. If one has a view that the reason why incumbent wages are rigid is because of the cost of changing wages or other institutional constraints, then it might not be realistic to think incumbent firms are acting strategically. Here, I will argue that new hire wages are (asymptotically) rigid even if incumbent firms are non-strategic.

Suppose incumbent firms mechanically fix wages due to some costs of changing wages or other constraints, $\hat{w}_0(z) = 0$ for all z.[18] Then, from (1.12), new hire wage responses are given by

$$\hat{w}_1(z) = \theta_{1a}(z)\hat{A}.$$

The key question here is whether $\theta_{1a}(z) < 1$ or not. If $\theta_{1a}(z) < 1$, then new hire wages become rigid whenever incumbent firms cannot adjust wages. The following proposition shows that this is indeed always the case toward the higher end of the job ladder:

Proposition 1′ (Endogenous wage rigidity with non-strategic incumbent firms). *Assume the distribution of z is such that $\bar{z} \to \infty$ with finite variance. If incumbent firms have exogenously fixed wages, $\hat{w}_0(z) = 0$, then new hire wages are rigid at the top of the job ladder, $\hat{w}_1(z) < \hat{A}$ for z close enough to $\bar{z}$.*

Therefore, regardless of incumbent firms being strategic or not, the job-ladder model robustly predict that there should be endogenous new hire wage rigidity toward the top of the job-ladder. However, the underlying mechanism here is distinct from the one with the strategic incumbent firms. Proposition 1′ comes from the fact that very productive firms are shielded from competition in the labor market. The degree of competition in this class of model is determined by the number of neighboring competitors. Since very productive firms have fewer of them, their monopsony power is high. As firms become more monopsonistic, their wage offers are increasingly tied to the workers outside options, $w_0(z)$, which is fixed here. That is, their wage offers become disconnected from the marginal product of labor, and wages are not responsive to aggregate productivity changes. In fact, I can show

$$\lim_{z \to \infty} \theta_{1a}(z) = 0,$$

which means new hire wages become completely rigid for very productive firms when incumbent firms fix wages.

The fact that very productive firms are insulated from competition in the labor market is a common feature of Burdett and Mortensen (1998) models. Recently, Gouin-Bonenfant (2020) exploits this insight to study the implications for labor shares. I exploit the same

[18] The argument goes through for any constant $C < \hat{A}$ with $\hat{w}_0(z) = C$.

insight but shed light on the implications for wage rigidity.

1.3.3 Incumbent Wage Rigidity Drives Unemployment Fluctuations

In Section 1.3.2, I have shown the model generates both incumbent and new hire wage rigidity, but which wage rigidity is important for unemployment fluctuations? Pissarides (2009) makes a strong argument that *only* new hire wage rigidity matters for job creation. In what follows, I challenge his conclusion.

I present two results in sequence:

Proposition 2. *Aggregate vacancy, V_s, is a function only of incumbent wage distribution, $\{w_{0s}(z)\}$. To a first order approximation, firm-level and aggregate vacancy responses are given by*

$$\hat{v}(z) = \iota \left[\frac{1-\alpha(z)}{\alpha(z)} (\hat{A} - \hat{w}_0(z)) + \hat{\lambda}^F \right], \tag{1.14}$$

$$\hat{V} = \frac{\iota}{1+\iota(1-\kappa)} \mathbb{E}_v \left[\frac{1-\alpha(z)}{\alpha(z)} (\hat{A} - \hat{w}_0(z)) \right], \tag{1.15}$$

where $\kappa \equiv \frac{d \ln \mathcal{M}(\tilde{\mu}, V)}{d \ln V}$ is the elasticity of the matching function with respect to vacancy, and $\mathbb{E}_v[x(z)] \equiv \int x(z)(v(z)/V) dG(z)$ denotes the vacancy-weighted average of a given variable.

The above result shows incumbent wages are sufficient statistics for unemployment fluctuations. It comes from the fact that since wages of new hires, $\{w_1(z)\}$, are optimally chosen, the profit from vacancy positing is not a function of $\{w_1(z)\}$. Consequently, while my model delivers new hire wage rigidity endogenously, such rigidity in itself has no consequence on unemployment fluctuations. The following result shows imposing further rigidity in new hire wages has no consequence either:

Proposition 2' (Incumbent wage rigidity as sufficient statistic with constrained new hire wages). *Assume new hires wage changes are exogenously given by $\hat{w}_1(z) = \hat{w}_1^{exo}(z)$ for some $\hat{w}_1^{exo}(z)$. To a first order approximation, the firm-level and the aggregate level vacancy responses are still given by (1.14) and (1.15).*

While the result comes from the linearization of (1.6), the proposition is a striking result. It says that incumbent wage rigidity is the *only* source of fluctuations in job creation. Any form of new hire wage rigidity, no matter whether the rigidity is endogenously derived or exogenously imposed, has no consequence on unemployment fluctuations. The result is precisely the opposite from what the conventional wisdom would suggest.

First, why does incumbent wage rigidity matter for job creation? It is because incumbent wage rigidity affects the prospect of poaching. If incumbent wages do not fall when the aggregate productivity falls, then incumbent workers are better paid relative to the overall economic condition. Under this situation, potential new employers have a hard time attracting incumbent workers. This reduces the return from the posting vacancy, in turn reducing job creation. Second, why does wage rigidity of new hires not matter for job creation? It is because of envelope theorem. Without shocks to the aggregate productivity, new hire firms were optimally setting wages to maximize profits, facing trade-off between paying higher labor costs and attracting more workers. Therefore, any first order (non-)response of their wages has no effect on profits from posting a vacancy, and in turn, on job creation.

The fact that rigidity of incumbents matters is very robust; the fact that it is the only rigidity that matters, so that new wage rigidity does not matter, is less robust to extensions of the model. It is always the case that a firm is not affected by its own rigidity of the wage, but there are potentially general equilibrium effects from the wage of others. The result here is stark because of the two-period assumption. I explore the robustness of the result in the context of quantitative infinite-horizon model in Section 1.6. Although new hire wage rigidity matters there, the quantitative magnitudes are small, and I find the incumbent wage rigidity still remains the dominant source of unemployment fluctuations.

Relationship to Pissarides (2009). The fact that incumbent wage rigidity does matter for job creation comes from the presence of on-the-job search. The fact that new hire wage rigidity does not matter for job creation comes from the assumption of wage posting. These two assumptions shape the backbone of Burdett and Mortensen's (1998) model. Pissarides (2009) and many others obtained the opposite conclusion because the argument is based on the DMP model. The DMP model has been popularly used to study the business cycle dynamics of unemployment due to its tractable nature, but this class of model assumes wage-bargaining and no on-the-job search.

To clarify the difference, consider an alternative version of my model with two modifications. First, let us assume there is no on-the-job search, $\zeta = 0$. Second, assume wages are bargained for instead of posted. Since any wage $w_1(z) \in [Ab, Az]$ generates positive gains from trade between unemployed and firms with productivity z, wages can be anywhere in the bargaining set, $[Ab, Az]$, as in **?**. Starting from steady-state value of $w_1(z)$, suppose the wage responses are given by $\hat{w}_1(z)$. The rest of the models are unchanged.

Appendix 1.8.7 shows that with these assumptions, to a first order, the firm-level vacancy response is given by

$$\hat{v}(z) = \iota \left[\frac{1-\alpha(z)}{\alpha(z)} (\hat{A} - \hat{w}_1(z)) + \hat{\lambda}^F \right], \tag{1.16}$$

and the aggregate level response is

$$\hat{V} = \frac{\iota}{1+\iota(1-\kappa)}\mathbb{E}_v\left[\frac{1-\alpha(z)}{\alpha(z)}(\hat{A}-\hat{w}_1(z))\right]. \tag{1.17}$$

These expressions echo Pissarides (2009) that new hire wages are the only source of fluctuations in job creation. Equation (1.17) is also a version of Ljungqvist and Sargent's (2017) formula incorporating firm heterogeneity and the finite elasticity of vacancy creation. By comparing (1.17) with (1.15), one again sees the striking contrast between the two. The two expressions only differ in terms of whether it is new hire or incumbent wages that enter the job creation equation. Expression (1.17) does not depend on incumbent wages because by abstracting from on-the-job search, it mechanically shuts down any interaction between incumbent workers and labor market dynamics. Expression (1.17) does depend on new hire wages because firms are not optimizing what wages to offer. With wage-bargaining, firms would prefer to pay as low of wages as possible so long as workers accept the job. This implies that new hire wage rigidity does have a first order effect.

Given that a different set of assumptions deliver strikingly different implications of wage rigidity, the natural question to ask is which assumptions are empirically relevant. The prevalence of on-the-job search is hard to deny. As mentioned in the introduction, 40-50% of new hires are employer-to-employer transitions. The assumption of wage posting is more controversial, but as discussed in Section 1.2.2, the available evidence is more supportive of wage posting than wage bargaining.

Employer-to-Employer (EE) transition rates. Equation (1.15) immediately implies that the UE (unemployment to employment) transition rate is unaffected by new hire wage rigidity because log-deviation in the UE rate is simply

$$\widehat{UE} \equiv \hat{\lambda}^U = \kappa\hat{V}.$$

In contrast, the EE transition rate is affected by the new hire wage rigidity. The EE rate is defined as

$$EE_s = \lambda_s^E \int (1 - F_{1s}(w_{0s}(\tilde{z}))\ell_0(\tilde{z})dG(\tilde{z})$$

because workers employed in firm z move to new employers whenever they receive better wage offers, which happens with probability $1 - F_{1s}(w_{0s}(z))$. The log deviation in the

EE rate is

$$\widehat{EE} = \frac{\lambda^E}{EE} \int F_1'(w_0(\tilde{z}))w(\tilde{z}) \left[\hat{w}_1(\tilde{z}) - \hat{w}_0(\tilde{z})\right] \ell_0(\tilde{z})dG(\tilde{z}) + (\text{terms unrelated to } \hat{w}_1(\tilde{z})) \tag{1.18}$$

This expression implies that relative rigidity in incumbent and new hire wages matter for the EE rate, and the EE rate responds more when new hire wages are more flexible relative to incumbent wages. Intuitively speaking, when new hire wages respond more than the incumbent wages, new hire firms can poach more workers. In fact, a firm poaches workers from other firms with higher productivity, causing a misallocation of workers. Therefore, although new hire wage rigidity is irrelevant to the UE rate, it matters a lot for the EE rate.

Since $\hat{w}_1(z) \geq \hat{w}_0(z)$ in equilibrium as we saw in Proposition 1, equation (1.18) implies that EE rate is strongly amplified relative to the case with flexible wages, $\hat{w}_1(z) = \hat{w}_0(z) = \hat{A}$. This is consistent with the evidence documented in Haltiwanger et al. (2018). They show that the firm wage ladder is strongly procyclical, meaning the number of workers who climb up the job-ladder collapses in recessions.[19] My theory provides a natural explanation of this.

Beyond First Order Approximation

Non-optimizing new hire wages in the steady-state. The reason why new hire wage rigidity does not affect job creation is because of the envelope theorem. Then, it is natural to think that if new hire wages are not optimized in the steady-state, they start to matter for unemployment fluctuations. I will show that although new hire wages matter, the way it matters is more subtle than one may think. Holding the incumbent wage distribution the same as (1.9), suppose the new hire wage distribution in the steady-state equilibrium is given by $w_1^n(z) \neq w(z)$. Let

$$\tau_1(z) \equiv \frac{(1-\chi)F'(w_1^n(z))w_1^n(z)}{(\chi + (1-\chi)F(w_1^n(z)))} - \frac{w_1^n(z)}{Az - w_1^n(z)}$$

denote the wedge of the optimality condition for the wage setting of new hire firms. I consider a small deviation of $w_1^n(z)$ from $w(z)$ so that $\tau_1(z) < 0$ if $w_1^n(z) > w(z)$, $\tau_1(z) = 0$ if $w_1^n(z) = w(z)$, and $\tau_1(z) > 0$ if $w_1^n(z) < w(z)$. Let $\hat{w}_1^n(z) \equiv \ln w_{1s}(z) - \ln w_1^n(z)$ denote the arbitrary constraint on new hire wage setting expressed as a log-deviation from the

[19] See also Barlevy (2002), Mukoyama (2014), and Nakamura et al. (2019) for related evidence.

non-optimized ones. Linearizing the optimality condition for vacancy creation gives

$$\hat{v}(z) = \iota \left[\hat{\lambda}^F + \frac{1-\alpha(z)}{\alpha(z)} \left(\hat{A} - \hat{w}_0(z) \right) + \tau_1(z) \hat{w}_1^n \right].$$

This expression immediately tells us new hire wage rigidity now matters for unemployment fluctuations. Suppose $\hat{A} > 0$. The more rigidity in new hire wages (lower $\hat{w}_1^n$) implies amplification of job creation if $\tau_1(z) < 0$, but it implies dampening of job creation if $\tau_1(z) > 0$. Intuitively speaking, when the initial wage is located in the increasing part of the profit function, the fact that wages cannot increase will decrease the profit from the vacancy posting. This dampens job creation in response to the positive shock to the aggregate productivity. In contrast, when the initial wage is located in a decreasing part of the profit function, job creation is amplified. Therefore, whether new hire wage rigidity amplifies unemployment fluctuation or not is generally ambiguous.

Second-order approximation. Another consideration is to study higher-order effects. Applying the second-order approximation to the optimality condition for vacancy creation at the firm-level, one can write

$$\frac{d^2 \log v(z)}{d \log A^2} = \nu_1 \frac{d^2 \log w_1(z)}{d \log A^2} + \nu_2 \frac{d \log w_0(z)}{d \log A} \frac{d \log w_1(z)}{d \log A} + \left(\text{terms unrelated to } \frac{d \log w_1(z)}{d \log A} \right)$$

where $\nu_1 \equiv \iota \frac{(1-\alpha(z))}{\alpha(z)^2} \left[2(1-\alpha(z)) - \alpha(z) \eta_{F_0}(z) \right] < 0$ and $\nu_2 \equiv -\iota \frac{(1-\alpha(z))}{\alpha(z)} \left\{ \eta_{F_0}(z) - \frac{(1-\alpha(z))}{\alpha(z)} \right\}$. Not surprisingly, new hire wage rigidity has a second-order effect on job creation, but does new hire wage rigidity amplify job creation? Not necessarily. The first term implies that relative to the case without rigidity, $\frac{d \log w_1(z)}{d \log A} > 0$, if we impose rigid new hire wages, $\frac{d \log w_1(z)}{d \log A} = 0$, fluctuation in job creation is amplified in response to a negative shock, but it is dampened in response to a positive shock. The sign of the second term is generally ambiguous. Therefore, incorporating new hire wage rigidity in this environment does not necessarily amplify job creation, even to a second-order.

1.3.4 Firm Insurance Drives Unemployment Fluctuations

In Section 1.3.3, I studied the implications of an arbitrary form of wage rigidity on vacancy creation, but I also showed that in Section 1.3.2, my model endogenously generates wage rigidity as an equilibrium outcome. Now, I connect the two to study the unemployment fluctuations arising from equilibrium wage rigidity. Extending Lemma 14, the first order equilibrium responses with endogenous vacancy creation $\{\hat{w}_1(z), \hat{w}_0(z), \hat{v}(z), \hat{V}, \hat{\lambda}^E, \hat{\lambda}^F\}$

solve

$$
\begin{aligned}
\hat{w}_1(z) &= \theta_{1a}(z)\hat{A} + \theta_{1w}(z)\hat{w}_0(z) - \theta_{1a}(z)\alpha(z)\frac{w(z)}{w'(z)}\hat{w}_0'(z) \\
\hat{w}_0(z) &= \theta_{0a}(z)\hat{A} + \theta_{0w}(z)\hat{w}_1(z) - \theta_{0a}(z)\alpha(z)\frac{w(z)}{w'(z)}\hat{w}_1'(z) + \theta_{0a}(z)\alpha(z)\left\{1-\theta_{\lambda,p}(z)\right\}\hat{\lambda}^E \\
&\quad + \theta_{0a}(z)\alpha(z)\theta_{\lambda,r}(z)\left(\hat{V}(z)-\hat{V}\right) + \theta_{0a}(z)\alpha(z)\left(\hat{v}(z)-\hat{V}\right) \\
\hat{v}(z) &= \iota\left[\frac{1-\alpha(z)}{\alpha(z)}(\hat{A}-\hat{w}_0(z)) + \hat{\lambda}^F\right],
\end{aligned}
\tag{1.19}
$$

with $\hat{\lambda}^F = (\kappa-1)\hat{V}$, $\hat{\lambda}^E = \kappa\hat{V}$, and $\hat{V} = \frac{1}{V}\int v(z)\hat{v}(z)dG(z)$, where $\hat{V}(z) \equiv \frac{\int^z v(\tilde{z})\hat{v}(\tilde{z})dG(\tilde{z})}{\int^z v(\tilde{z})dG(\tilde{z})}$ is the log change of the cumulative amount of vacancies, $\theta_{\lambda,p}(z) \equiv \frac{\lambda^E(1-F(w(z)))}{1-\lambda^E+\lambda^E F(w(z))}$ is the share of workers who meet with other firms and are poached, and similarly $\theta_{\lambda,r}(z) \equiv \frac{\lambda^E F(w)}{1-\lambda^E+\lambda^E F(w)}$ is the share of workers who meet with other firms but reject the offer. The boundary conditions remain the same: $\hat{w}_1(\underline{z}) = \hat{A}$ and $\hat{w}_0(\bar{z}) = \hat{w}_1(\bar{z})$.

Compared with Lemma 14, the vacancy responses enter into the best response function of incumbent wage settings. For example, if there is a positive response of aggregate vacancy, $\hat{\lambda}^E > 0$, all else equal, incumbent firms have the incentive to raise wages to prevent workers from leaving.

The following proposition provides a useful starting point in studying the role of firm insurance in unemployment fluctuations:

Proposition 3. *If workers are risk-neutral, $\gamma = 0$, there is no unemployment fluctuation.*

As we move to risk-averse workers, $\gamma > 0$, the economy exhibits unemployment fluctuation, $\hat{V} > 0$. Furthermore, one can show that the maximum unemployment fluctuation occurs in the limit where workers are infinitely risk-averse, $\gamma \to \infty$:

$$
\lim_{\gamma\to\infty}\hat{V} \to \frac{\iota}{1+\iota(1-\kappa)}\mathbb{E}_v\left[\frac{1-\alpha(z)}{\alpha(z)}\right]\hat{A} > 0.
$$

.

Figure 1-6 illustrates the result by plotting wages (left-top), vacancy (right-top), the UE rate (left-bottom) and the EE rates (right-bottom) against the relative risk aversion, γ. First, the model generates no unemployment fluctuations with risk-neutral workers, $\gamma = 0$. As we have already seen, with risk-neutrality, incumbent wages move one for one with the aggregate productivity. Since the cost of vacancy also scales with the aggregate productivity, the profitability from a vacancy posting is unchanged.[20] This serves as a

[20]The same benchmark case appears in Blanchard and Galí (2010) and Kehoe et al. (2019).

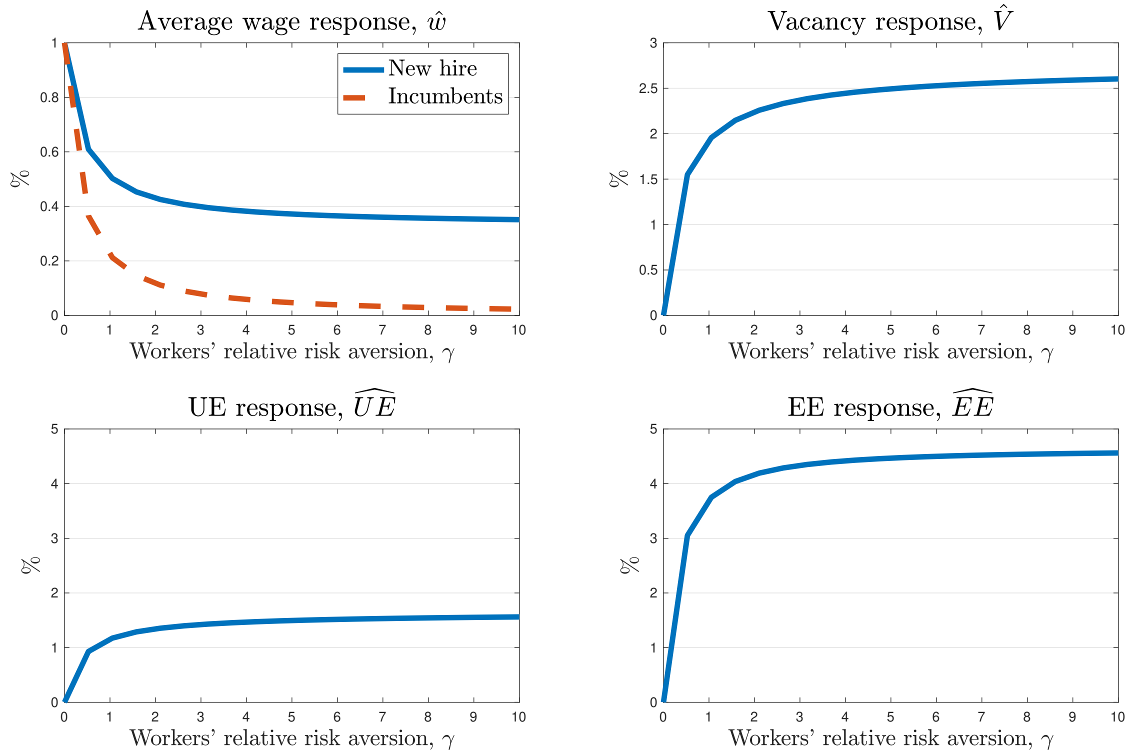


Figure 1-6: Workers' risk aversion and wage and unemployment fluctuations

Note: The figure shows a numerical example of the responses to a 1% aggregate productivity shock for each value of γ. All the reported values are log deviations from the steady-state. The parameter values are $\mu = 0.06, \zeta = 0.2, \mathcal{M}(\tilde{\mu}, V) = \tilde{\mu}^{1-\kappa}V^{\kappa}, \kappa = 0.6, \iota = 1, b = 1, \bar{c}(z) = z^{\bar{c}}, \bar{c} = 10, G(z) = 1 - z^{-\alpha}, \alpha = 5$.

useful benchmark. As soon as we move away from risk-neutral workers, incumbent firms start to insure workers, which generate incumbent wage rigidity. This rigidity tends to spill over to new hire wages, as depicted in the left-top panel of Figure 1-6. Because we have already seen that the incumbent wage rigidity drives the fluctuation in vacancy creation, the response of vacancy increases with γ (the right-top panel of Figure 1-6), reaching the limit described in the proposition as $\gamma \to \infty$. Moreover, the model predicts the EE rate to substantially respond more than the UE rate because the fact that new hire wages respond more than the incumbent wages make poaching easier, as we have already discussed.

The fact that firm insurance solely drives unemployment fluctuation is in stark contrast to the arguments made in Barro (1977) and Rudanko (2009). Both papers point out that long-term contracts between firms and workers do not contribute to unemployment fluctuations. Here, on-the-job search was crucial to reach the opposite conclusion. The following proposition formally illustrates the importance of on-the-job search:

Proposition 4. *If there is no on-the-job search, $\zeta = 0$, there is no unemployment fluctuation.*

Abstracting from on-the-job search shuts down any interaction between incumbent wages and labor market dynamics. Since the hiring pool only consists of unemployed and their outside option scales with the aggregate productivity, the return from vacancy posting is invariant to the aggregate productivity. In this case, as explained in Rudanko (2009) and Pissarides (2009), the incentive to create jobs is disconnected from the incumbent wage rigidity, no matter how rigid they are.

It is worth emphasizing wage rigidity and unemployment fluctuation in my model solely come from optimal contracting problems between firms and workers. The theory differs from existing models of wage rigidly and unemployment fluctuations in the following two senses: First, it does not rely on any unexplained inefficiencies such as the ad-hoc cost of changing wages, and, thus, immune to Barro's (1977) critique that wage rigidity should not interfere with mutually beneficial contracts. Second, it does not rely on an arbitrary choice of the wage setting rule in the bargaining set (**?**). The degree of wage rigidity and unemployment fluctuations in my model are disciplined by the structural parameters, such as workers' risk aversion. In contrast, models of wage rigidity in DMP tradition lack such discipline, so they cannot speak to the questions of how wage rigidity changes with counterfactual policies, for example.

1.4 Extensions: Internal Firm Fairness and Public Insurance

Building on the insights that I derived in Section 1.3, I consider two comparative statics in the model. The first exercise studies the effect of imposing fairness constraint within a firm. which prevents firms to discriminate incumbent workers and new hires. The second exercise considers the effect of public insurance.

1.4.1 Fairness Constraints Dampen Unemployment Fluctuations

In the baseline model, I have assumed that incumbent wages and new hire wages can be set separately in an unconstrained manner. However, in practice, if incumbent workers and new hires belong to the same firm, it might be difficult to discriminate wages due to fairness concerns.[21] It is often argued that such a constraint amplifies unemployment

[21] The presence of such social norms are documented empirically (Card et al., 2012; Breza et al., 2018; Dube et al., 2019).

fluctuations by making new hire wages more rigid. This idea at least goes back to Bewley (1999), and has been formalized later in several papers (Snell and Thomas, 2010; Rudanko, 2019; Menzio, 2004; Gertler and Trigari, 2009).[22] Because conventional wisdom says that new hire wage rigidity is the source of amplification, it is natural to expect that any constraint that prevents the flexible adjustment of new hire wages would amplify unemployment fluctuations. However, I will argue that these implications are reversed in my model. The key idea is that fairness constraints make new hire wages more rigid, but incumbent wages more flexible. As incumbent wage rigidity is the source of amplification in my model, the constraint dampens unemployment fluctuations.

I assume that that the boundary of firms are such that each productivity z corresponds to a single firm.[23] Then, I impose that the firms cannot discriminate new hire wages and incumbent wages due to fairness concerns or other social norms, $w_{0s}(z) = w_{1s}(z)$. Each firm z solves the following problem:

$$\begin{aligned}
\max_{w_{0s}, w_{1s}, v_s} \sum_{s \in \{h,l\}} \pi_s \Bigg[& (A_s z - w_{0s})(1 - \lambda_s^E + \lambda_s^E F_{1s}(w_{0s}))\ell_0(z) \\
& + \lambda_s^F v_s(\chi + (1-\chi)F_{0s}(w_{1s}))(A_s z - w_{1s}) - c_s(v_s; z) \Bigg] \\
\text{s.t.} \quad & \sum_{ss \in \{h,l\}} \pi_s \left[(1-\lambda_s^E) u(w_{0s}) + \lambda_s^E \int \max\{u(w_{0s}), u(\tilde{w})\} dF_{1s}(\tilde{w}) \right] \geq \bar{W}_0(z), \\
& w_{0s} \geq A_s b, \quad w_{1s} \geq A_s b, \\
& w_{0s} = w_{1s}.
\end{aligned}$$

Therefore, firms maximize the weighted average of profits from new hires (the first term) and incumbents (the second term) while delivering the promised utility to incumbent workers. I again consider a perturbation around the same symmetric steady-state equilibrium, as in Lemma 1.In this steady-state, the fairness constraint, $w_{0s}(z) = w_{1s}(z)$, is not binding because incumbent workers and new hires are offered the same wages anyway. The steady state wage distribution is given by (1.9), which is monotone in z.

With shocks, the constraint is binding because incumbent and new hire firms now face different incentives to set wages. Let $w_s(z) \equiv w_{0s}(z) = w_{1s}(z)$ denote the firm-level wages, and let $\hat{w}(z)$ denote their log-deviation from the steady-state.. The first-order

[22]Besides the business cycle literature, it has been common to impose fairness (equal treatment) constraints in wage posting models since Burdett and Mortensen (1998).

[23]Or one can think that all firms with the same productivity are symmetric.

condition with respect to $w_s(z)$ under the binding fairness constraint is given by

$$-(1-\lambda_s^E+\lambda_s^E F_{1s}(w_s(z)))\ell_0(z)+(A_s z-w_s(z))\lambda_s^E F_{1s}'(w_s(z))\ell_0(z)$$
$$+\eta(z)\left[(1-\lambda_s^E)u'(w_s(z))+\lambda_s^E F_{1s}(w_s(z))u'(w_s(z))\right]\ell_0(z)$$
$$+(1-\chi)F_{0s}'(w_s(z))(A_s z-w_s(z))\lambda_s^F v_s(z)-\left(\chi+(1-\chi)F_{0s}(w_s(z))\right)\lambda_s^F v_s(z)=0.$$

The first order condition with respect to vacancy creation, $v_s(z)$, remains the same (1.6). Log-linearizing these first-order conditions, first order equilibrium responses, $\{\hat{w}(z),\hat{v}(z),\hat{V},\hat{\lambda}^E,\hat{\lambda}^F\}$ solve

$$\hat{w}(z)=\theta_a^{eq}(z)d\ln A_s-\theta_a^{eq}(z)\alpha(z)\frac{w(z)}{w'(z)}\hat{w}'(z)+\theta_a^{eq}(z)\varphi(z)\alpha(z)\left\{1-\theta_{\lambda,p}(z)\right\}\hat{\lambda}^E$$
$$+\theta_a^{eq}(z)\varphi(z)\alpha(z)\theta_{\lambda,r}(z)\left(\hat{V}(z)-\hat{V}\right)+\theta_a^{eq}(z)\varphi(z)\alpha(z)\left(\hat{v}(z)-\hat{V}\right),$$
$$\hat{v}(z)=\iota\left[\frac{1-\alpha(z)}{\alpha(z)}(\hat{A}-\hat{w}(z))+\hat{\lambda}^F\right],$$

$\hat{\lambda}^F=(\kappa-1)\hat{V}$, $\hat{\lambda}^E=\kappa\hat{V}$, and $\hat{V}=\frac{1}{V}\int v(z)\hat{v}(z)dG(z)$ with the boundary condition $\hat{w}(\underline{z})=\hat{A}$, where $\theta_a^{eq}(z)\equiv\frac{1}{1+\gamma\varphi(z)\theta_\lambda(z)}$ and $\varphi(z)\equiv\frac{\lambda^E F_1'(w(z))\ell_0(z)}{\lambda^F v(z)(1-\chi)F_0'(w(z))+\lambda^E F_1'(w(z)\ell_0(z)}$.

As one might expect, with fairness constraints, the best responses are the weighted average of the best responses of (1.12) and (1.13) after imposing $\hat{w}_0(z)=\hat{w}_1(z)$. The following result is immediate:

Proposition 5. *Assume the elasticity of vacancy creation, ι, is sufficiently small. Fairness constraints raise the flexibility of incumbent wages at the bottom of the job-ladder, $\hat{w}(z)>\hat{w}_0(z)$ for z close enough to $\underline{z}$.*

The result says near the bottom of the job ladder, incumbent wages become more flexible, which comes from the boundary condition at the bottom. This is intuitive, as incumbent wages not only serve as insurance but also need to attract new workers. While I cannot prove that this holds globally, the wage rigidity at the bottom of the job ladder plays a dominant role for unemployment fluctuations. This is because low-productivity firms have a low surplus (low $\alpha(z)$), so their vacancies are particularly more sensitive to wage rigidity (see equation (1.16)).

Figure 1-7 shows a numerical example of how incorporating fairness constraints affects labor market fluctuations. The left top panel shows the responses of wages. As one would expect, the fairness constraint increases the flexibility of incumbent wages, and reduces new hire wage flexibility for most of the range of γ. Because incumbent wage rigidity is the sole driver of vacancy fluctuations, the fairness constraints dampen the va-

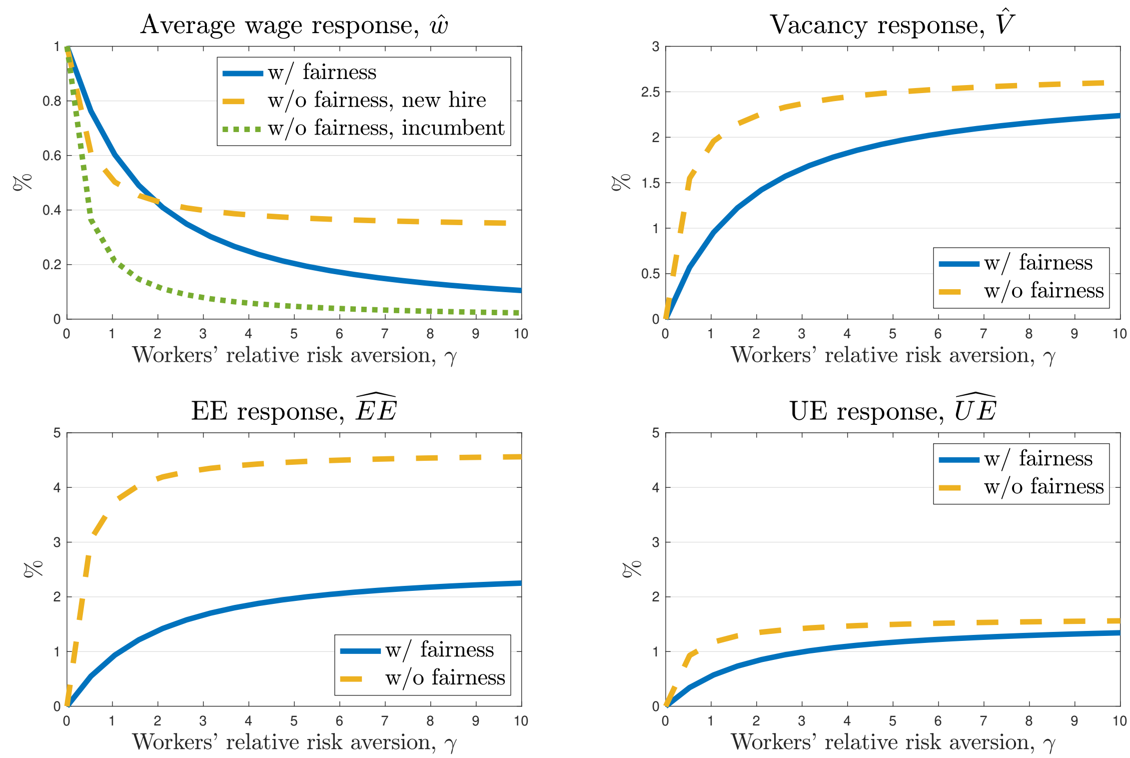


Figure 1-7: The impact of fairness constraints on labor market fluctuations

Note: The figure shows a numerical example of the responses to a 1% aggregate productivity shock for each value of γ. All the reported values are log deviations from the steady-state. The parameter values are the same as Figure 1-6.

cancy response, as the right top panel shows. This is in stark contrast to conventional wisdom. The bottom two panels show the effect on the EE and the UE rates. Notably, fairness constraint dampens the EE responses much more than the UE response. This comes from the fact that with fairness constraints, the term highlighted in (1.18) is zero. Because wages are strictly increasing in productivity z, workers always flow from less productive firms to more productive firms. Therefore, in contrast to the case without fairness constraints, there is no cyclical misallocation.

1.4.2 Government-provided Insurance Dampens Unemployment Fluctuations

The source of unemployment fluctuations in my model comes from firm insurance. What if the government could also provide insurance to workers? Suppose the government imposes lump-sum taxes/transfers, T_s, in state s to all workers (including the unemployed), financed via a lump-sum tax from entrepreneurs. Under this assumption, the only modification is the promise-keeping constraint of incumbent firms:

$$\sum_{ss\in\{h,l\}} \pi_s \left[(1-\lambda_s^E) u(w_{0s}+T_s) + \lambda_s^E \int \max\{u(w_{0s}+T_s), u(\tilde{w}+T_s)\} dF_{1s}(\tilde{w}) \right] \geq \bar{W}_0(z).$$

Because income taxes are unconditional, workers' job mobility decisions are not affected. The transfers are revenue-neutral in expectation, $\sum_s \pi_s T_s = 0$. I assume that in a steady-state, $T = 0$, and $dT_l = -dT_h \equiv dT > 0$ under small aggregate risk. This means the government transfers money in recessions and taxes in booms. The linearized best response of incumbent firms now become

$$\hat{w}_0(z) = \theta_{0a}(z)\hat{A} + \overbrace{\gamma\omega_2(z)\theta_{0a}(z)\frac{1}{w(z)}dT}^{\text{public insurance effect } (\geq 0)} + \theta_{0w}(z)\hat{w}_1(z) - \theta_{0a}(z)\alpha(z)\frac{w(z)}{w'(z)}\hat{w}_1'(z) \tag{1.20}$$
$$+ \theta_{0a}(z)\alpha(z)\left\{1-\theta_{\lambda,p}(z)\right\}\hat{\lambda}^E + \theta_{0a}(z)\alpha(z)\theta_{\lambda,r}(z)\left(\hat{V}(z)-\hat{V}\right) + \theta_{0a}(z)\alpha(z)\left(\hat{v}(z)-\hat{V}\right),$$

where $\omega_2(z) > 0$ is defined in Appendix 1.8.3. All the other equilibrium conditions are unchanged.

The following result is immediate:

Proposition 6. *If workers are risk-neutral, $\gamma = 0$, public insurance has no effect on equilibrium.*

Moving to the case wth risk-averse workers, $\gamma > 0$, we can see from equation (1.20) that, holding everything else constant, public insurance makes incumbent wages more flexible. The intuition is that firms now do not need to provide insurance as much as before because the government partially substitutes for it. If incumbent wages become more flexible, this dampens unemployment fluctuations. Figure 1-8 shows a numerical example by making a comparison with and without government insurance. I set $dT = 0.2$. The left panel confirms that incumbent wages become more flexible with government insurance. The middle panel shows that through strategic complementarity, the public insurance also increases the flexibility of new hire wages. Finally, the right panel shows that the fluctuations in vacancy creation are dampened.

While it is an incentive for firms to provide insurance to incumbent workers which drives unemployment fluctuations, providing more insurance to *all* workers reduces unemployment fluctuations. The crucial market failure in my model is that workers are allowed to write contracts *only* with their current employers. If they could write contracts with potential new hire firms, which is in principle what the government is trying do here, the unemployment fluctuations would disappear. Note that if private agents do not

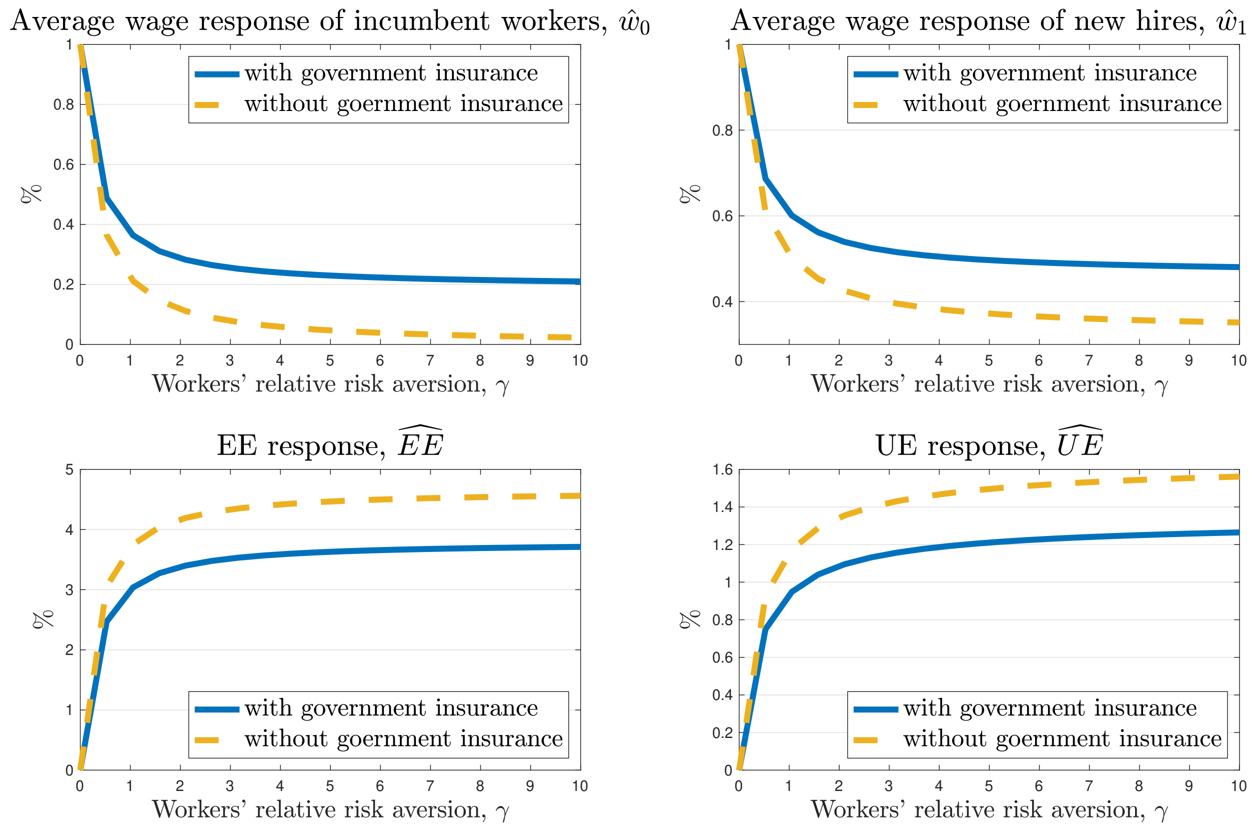


Figure 1-8: Labor market response with and without government insurance

Note: Figure 1-8 shows a numerical example of the responses to a 1% aggregate productivity shock for each value of γ. All the reported values are log-deviation from the steady-state. The cyclicality of the transfer is set to $dT = 0.2$. The remaining parameter values are the same as Figure 1-6.

anticipate that the government-provided insurance, there is no effect on unemployment fluctuations. Any ex-post tax on labor income that is imposed after the realization of aggregate productivity does not affect the firm's nor the worker's behavior. Therefore, it is precisely the ex-ante role of public insurance that crowds out firm insurance, which in turn reduce unemployment volatility.

In an extreme case where the transfer is allowed to depend on employer's identity, it is possible for the government to completely eliminate unemployment fluctuations. Letting $dT(z)$ denote the transfer to workers employed in a firm with productivity z, suppose $dT(z) = \frac{1}{\gamma\omega_2(z)\theta_{0a}(z)\frac{1}{w(z)}}(1 - \theta_{0a}(z) - \theta_{0w}(z))\hat{A}$. Then, it is straightforward to see the equilibrium features complete wage flexibility and no unemployment fluctuations. While it is not realistic to implement such an intervention, the result undermines the role of public insurance in stabilizing the labor market fluctuations.

1.5 Inefficiencies: Steady States and Response to Shocks

So far, we have focused on the positive implications of the model. What are the normative implications? I first ask whether the steady-state equilibrium is efficient or not. Then, I study whether equilibrium responses to the aggregate shock is efficient. The spirit of exercise is to consider whether a small perturbation of the equilibrium achieves Pareto improvement or not. If it does, it implies the equilibrium is constrained inefficient. Although a more satisfactory and interesting analysis is to derive the optimal policy, I leave this for the future followup work.

1.5.1 Steady-state Inefficiencies

Let us start from the welfare properties of the steady-state equilibrium. While Gautier et al. (2010) and Cai (2020) study the efficiency property of the Burnett-Mortensen model with risk-neutral workers, to the best of my knowledge, the efficiency property with risk-averse workers has not been studied before. I consider a planner who can directly intervene to perturb (i) the wage offers; (ii) the amount of vacancies; and (iii) unemployment benefit, $\{\delta w_0(z), \delta w_1(z), \delta v(z), \delta b\}$, where δx denotes the marginal changes in x. I assume the unemployment benefit is financed via a lump-sum tax on firms.

Inefficient wage offers. First, fixing $\delta v(z) = 0$, consider a small reduction in new hire firm's wage, $\delta w_1(z) < 0$, for some z, combined with $\{\delta w_0(z), \delta b\}$, which would leave workers indifferent. This exercise is meant to isolate whether the wage offers in equilibrium is efficient or not by shutting down the vacancy margin. Can such an intervention improve welfare? The following proposition shows the answer is yes if workers are risk-averse:

Proposition 7. *Consider the steady-state equilibrium. For any z, there exists a feasible perturbation featuring the same amount of vacancies at all firms, but strictly lower wages for new hires at firm z, $\delta w_1(z) < 0$, that yield a Pareto improvement if and only if workers are strictly risk-averse, $\gamma > 0$.*

Therefore, potential new employers make too aggressive wage offers, which is more true for more productive firms. In what follows, I sketch the proof. As workers have to be indifferent, the perturbation must satisfy

$$\Big((1-\lambda^E) + \lambda^E F_1(w(\tilde{z}))\Big) u'(w(\tilde{z})) dw_0(\tilde{z}) + \mathbb{I}(z > \tilde{z}) \lambda^E u'(w(z))(v(z)/V) g(z) \delta w_1(z) = 0$$

$$(1-\lambda^U) u'(Ab) A \delta b + \lambda^U u'(w(z))(v(z)/V) g(z) \delta w_1(z) = 0$$

for all $\tilde{z} \in [\underline{z}, \bar{z}]$. The question is whether such perturbation can raise the net total surplus (total firms' profits), which is given by

$$
\begin{aligned}
\mathcal{F} = & \int \Big[(Az - w_0(\tilde{z}))(1 - \lambda^E + \lambda^E F_1(w_0(\tilde{z})))\ell_0(\tilde{z}) \Big] \, dG(\tilde{z}) \\
& + \int \Big[v(\tilde{z})\lambda^F(\chi + (1-\chi)F_0(w_1(\tilde{z})))(Az - w_1(\tilde{z})) - c(v(\tilde{z}); \tilde{z}) \Big] \, dG(\tilde{z}) + \mu(1 - \lambda^U)Ab
\end{aligned}
$$

where $\tilde{F}(w) \equiv \chi + (1-\chi)F_0(w)$. As long as workers are strictly risk-averse, $\gamma > 0$, the answer is yes:

$$
\begin{aligned}
d\mathcal{F} = & -v(z)\lambda^F \left[\chi \left(1 - \frac{u'(w(z))}{u'(Ab)} \right) + (1-\chi) \int^z \left(1 - \frac{u'(w(z))}{u'(w(\tilde{z}))} \right) \frac{\ell_0(\tilde{z})}{1-\mu} dG(\tilde{z}) \right] g(z)\delta w_1(z) \\
& > 0
\end{aligned}
$$

where the last inequality follows from $\delta w_1(z) > 0$ and $u'' < 0$. This implies that we can Pareto improve welfare by forcing a new hire firm to slightly lower the wage offer. The reason is that potential new employers do not internalize their contribution to the idiosyncratic income risks. If potential new employers lower wage offers, workers' consumption dispersion goes down, and it becomes cheaper for incumbent firms to deliver the promised utility if the utility function is concave. Of course, such an intervention potentially creates a form of misallocation because workers now accept an offer from less productive firms. However, as there is no misallocation in the steady-state, such consideration has no first order effect on welfare and only has a second-order effect. Moreover, we see that the term inside parenthesis is strictly increasing in z, which means that the externality is greater for more productive firms. The reason is that, because workers only accept a better wage offer, the contribution to the income risk is greater for high-paying productive firms.

Inefficient vacancy creation. Next, I ask whether the vacancy creation is efficient or not. To focus on vacancy margin, I fix $\delta w_1(z) = 0$. Then, I consider a perturbation $\delta v(z)$, $\{\delta w_0(\tilde{z})\}$ and δb that leave workers indifferent and see whether such a perturbation can raise the total net surplus. The following expression characterizes the effect of such perturbation on the total net surplus:

$$
\begin{aligned}
d\mathcal{F} = & \overbrace{\lambda^F \int^z \left(\frac{u(w(z)) - u(w(\tilde{z}))}{u'(w(\tilde{z}))} - [w(z) - w(\tilde{z})] \right) dP(\tilde{z})}^{\text{idiosyncratic income risk externality } (\leq 0)} \delta v(z) \\
& + \underbrace{(\kappa - 1)\lambda^F \int \int^{\tilde{z}} \left[(A\tilde{z} - w(\tilde{z})) - (w(\tilde{z}) - w(\zeta)) + \frac{u(w(\tilde{z})) - u(w(\zeta))}{u'(\tilde{z})} \right] dP(\zeta) \frac{v(\tilde{z})}{V} dG(\tilde{z}) g(z)\, \delta v(z)}_{\text{congestion externality } (\leq 0)},
\end{aligned}
\tag{1.21}
$$

where $\kappa \equiv \frac{d \ln \mathcal{M}}{d \ln V}$ is the elasticity of matching function with respect to vacancy, and $P(z) \equiv \chi \mathbb{I}(z > \underline{z}) + (1-\chi) \int^z \frac{1}{1-\mu} \ell_0(\tilde{z}) dG(\tilde{z})$ is the search-efficiency weighted cumulative employment distribution (including unemployed).

The expression shows that the welfare effect of reducing the vacancy of a particular firm z can be decomposed into two (generically) non-zero terms. The first term, which I label as idiosyncratic income risk externality, captures that firms do not internalize their contribution to the worker's income risk when they create jobs. One can immediately see that if workers are risk-neutral, this term is zero. With risk-averse workers, this term is negative because more job creation increases the workers upward income risk, which makes it costlier for incumbent firms to deliver the promised utility. Welfare can be improved if the planner forces firms to reduce job creation. Moreover, the absolute size of this term is increasing in z because they contributed the most to enlarging workers' consumption dispersion. This implies productive firms tend to be too large relative to the social optimum. In contrast to this, many theories predict productive firms are too small compared with the social optimum. Under search and matching frictions, Acemoglu (2001) shows that unproductive firms create too many jobs relative to productive ones because they do not internalize that they crowd out more productive matches. Golosov, Maziero, and Menzio (2013) show also in the context of a frictional labor market that too few workers seek jobs in productive firms because of search risk. Under an oligopsonistic labor market, productive firms hire too few workers (e.g., Berger et al., 2019) due to labor market power.

The second term, which I label the congestion externality, is relatively more standard (Gautier et al., 2010; Cai, 2020). This term is zero when the elasticity of the matching function with respect to vacancy is one, $\kappa = 1$. Since the wage posting model can be interpreted as firms having all the bargaining power, this condition is analogous to Hosios'(1990) condition. As we move toward $\kappa < 1$, this term becomes negative. Firms do not internalize that their job creation will congest the market and lower meeting probabilities of other firms. Welfare can be improved by reducing job creation. Notably, this term does not depend on z, so the externality is the same for all firms.

I summarize the above discussion as follows:

Proposition 8. *Consider the steady-state equilibrium. For any z, there exists a feasible perturbation featuring the same new hire wages at all firms, but strictly lower vacancies at firm z, $\delta v(z) < 0$, that yield a Pareto improvement if and only if workers are risk-averse, $\gamma > 0$, or Hosios condition fails, $\kappa < 1$. Furthermore, for a given change in vacancies δv, the change in social surplus is increasing with the productivity of firm z.*

1.5.2 Inefficient Wage Flexibility under Aggregate Risk

So far, I have focused on the efficiency in the steady-state. The natural next question is whether the equilibrium response to aggregate shocks is efficient or not. In particular, are wages too flexible or too rigid in equilibrium? Because wage rigidity in my model is fully micro-founded as an optimal contracting problems, these questions are well defined. This is in contrast to many existing models of wage rigidity, in which rigidity is imposed exogenously.

Inefficient new hire wage flexibility. I consider a planner who can directly intervene to perturb new hire wages in each state, but cannot control vacancies. Let me first concentrate on the efficiency of new hire wage settings by imposing $\delta w_{0h}(z) = \delta w_{0l}(z) = 0$ for all z. Take a particular firm z_s in each state s, and consider a perturbation, $\delta w_{1h}(z_s), \delta w_{1l}(z_s)$. For such a perturbation to leave workers indifferent, they must satisfy

$$\begin{aligned} &\mathbb{I}(w_{0h}(\tilde{z}) \leq w_{1h}(z_h))\lambda_h^E u'(w_{1h}(z_h))\delta w_{1h}(z_h)\frac{v_s(z_h)}{V_h}g(z_h) \\ &+\mathbb{I}(w_{0l}(\tilde{z}) \leq w_{1l}(z_l))\lambda_l^E u'(w_{1l}(z_l))\delta w_{1l}(z_l)\frac{v_s(z_l)}{V_l}g(z_l) = 0 \end{aligned} \tag{1.22}$$

for all $\tilde{z}$. Note that (1.22) implies that unemployed workers are also left indifferent, because $\lambda_s^E = \zeta\lambda_s^U$. I focus on the non-trivial case with $w_{0h}(\tilde{z}) \leq w_{1h}(z_h)$ and $w_{0l}(\tilde{z}) \leq w_{1l}(z_l)$ for all $\tilde{z}$. This implies that there exists $\breve{z}$ such that $w_{0h}(\breve{z}) = w_{1h}(z_h)$ and $w_{0l}(\breve{z}) = w_{1l}(z_l)$. Since we have already learned that the vacancies are unaffected by any small movement in new hire wages, so a term that involve $\delta v_s(z)$ does not show up in the above expression. The question is whether such perturbation can raise the expected net total surplus, which is given by

$$\mathcal{F} = \sum_{s=l,h} \pi_s \Big\{ \int \Big[(A_s\tilde{z} - w_{0s}(\tilde{z}))(1-\lambda_s^E + \lambda_s^E F_{1s}(w_{0s}(\tilde{z})))\ell_0(\tilde{z}) \Big] dG(\tilde{z}) \\ + \int \Big[v_s(\tilde{z})\lambda_s^F(\chi + (1-\chi)F_0(w_{1s}(\tilde{z})))(A_s\tilde{z} - w_{1s}(\tilde{z})) - c_s(v(\tilde{z});\tilde{z}) \Big] dG(\tilde{z}) + \mu(1-\lambda_s^U)A_s b \Big\}.$$

The changes in net surplus can be computed as

$$d\mathcal{F} = \overbrace{\sum_{s=l,h} \Big[A_s\left(z_s - \check{z}\right)\lambda_s^E F'_{1s}(w_{1s}(z_s))\ell_0(\check{z}) \Big] \delta w_{1s}(z_s)}^{\text{misallocation}} \\ \underbrace{+ \Big[\mu\lambda_s^U + \lambda_s^E \int^{\check{z}} \ell_0(\tilde{z})dG(\tilde{z}) \Big] \left(\frac{u'(w_{0l}(\check{z}))}{u'(w_{0h}(\check{z}))} - 1 \right) \frac{v_l(z_l)}{V_l} g(z_l)\delta w_{1l}(z_l)}_{\text{aggregate risk-sharing}}.$$

The first term, which I labeled the misallocation, comes from the fact that there is cyclical misallocation in my model. As we have seen in Proposition 14, if workers are risk-averse, incumbent firms' wages respond less than the potential new employers with the same productivity. Therefore, workers can flow toward a less productive firm in booms, and reject the offer from a more productive firms in recessions. That is, $z_s - \check{z}$ is negative for $s = h$ and positive for $s = l$. If the planner forces potential new employers to respond less to the aggregate productivity, this will alleviate the misallocation. The second term, which I labeled the aggregate income risk-sharing, comes from the fact that potential new employers do not internalize their contribution to the aggregate income risk. As is the case with idiosyncratic income risk externality, if the planner could force potential new employers to be less aggressive, this would alleviate the limited commitment friction of the contract between incumbent firms and workers. As long as u is strictly concave, this term is positive for $\delta w_{1h}(z_h) < 0$ and $\delta w_{1l}(z_l) > 0$. This leads me to conclude:

Proposition 9. *Consider the equilibrium with aggregate risk. Assume the elasticity of vacancy creation, ι, is sufficiently small. There exists a perturbation featuring lower new hire wages in some firms in the high state, $\delta w_{1h}(z_h) < 0$, and higher new higher wages in the low state, $\delta w_{1l}(z_l) > 0$, that yields a Pareto improvement if and only if workers are risk averse, $\gamma > 0$.*

That is, the new hire wages are too flexible in equilibrium. Despite my model generates endogenous new hire wage rigidity, the planner improves the welfare by making new hire wages even more rigid. We have seen that more new hire wage rigidity has no consequence for unemployment fluctuations, but it still improves welfare.

Incumbent wage flexibility. Although it is of great interest to understand whether incumbent wages are too rigid or too flexible in equilibrium, I can only analytically study this for a special case. The complication arises in taking into account endogenous responses of job creation associated with the changes in incumbent wages. When vacancies are inelastic, $\iota \to 0$, I can shut off this channel. Appendix 1.9 shows that incumbent wages are also too flexible in equilibrium. Intuitively speaking, incumbent firms move around wages with aggregate productivity in order to block poaching from potential new employers. However, such competition is business stealing. If firms could collectively focus on insuring workers, this would raise the welfare.

1.6 Quantitative Exploration

To quantity the mechanisms, I extend the previous two-period model to a infinite horizon model in continuous time. I first describe the environment where there is no aggregate shock and consider the one-time unanticipated aggregate shock.

1.6.1 From Two-period to Infinite Horizon

Preferences and technology. The economy is populated by a mass of workers and a mass of entrepreneurs. Workers are risk-averse with preferences

$$W_0 = \int_0^\infty e^{-\rho t} u(c_{wt}) dt,$$

where $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ as before, and entrepreneurs are risk neutral,

$$\int_0^\infty e^{-\rho t} c_{kt} dt.$$

The flow value of the unemployed is Ab, where A is the aggregate productivity. Firms operate the linear production technology, $y = Azl$, and z is the firm's permanent productivity distributed according to the cumulative density function $G(z)$.

Unemployed and employed workers meet with a firm with Poisson intensity λ_t^U and $\lambda_t^E \equiv \zeta \lambda_t^U$, respectively. Employed workers exogenously separate with firms with Poisson intensity δ. Similarly to Moscarini and Postel-Vinay (2016a) and Gertler et al. (2020), I also introduce reallocation shock with intensity $\varkappa$. When hit by the reallocation shock, the worker is forced to move to another firm with the same productivity and inherits the same wage contracts. This is meant to capture the fact that not all job-to-job transitions are for climbing up the job ladders and appear for various other reasons (e.g. spousal moving). This assumption is not solely for realism. Without reallocation shock, the employer-to-employer transition rate observed in the data implies that firms face strong competition with each other. This makes it difficult to match the observed degree of wage cyclicality in the data.

Firms post a vacancy with convex cost, $Ac(v;z)$, which scales with the aggregate labor productivity, as before. This assumption ensures balanced growth, in which the permanent changes in the aggregate productivity, A, leave the long-run unemployment rate unchanged. Each vacancy meets with a worker with intensity λ_t^F. As before, among the workers firms meet with, fraction $\chi_t = \frac{\mu_t}{\mu_t + \zeta(1-\mu_t)}$ is unemployed and the remaining fraction $1-\chi_t$ is employed. The total number of meetings between firms and workers is given by a CRS matching technology $\mathcal{M}(\tilde{\mu}_t, V_t)$.

Contracts and markets. Firms compete for workers by posting wage contracts, w. I assume a wage contract can only depend on labor productivity, Az, and thus excluding the possibility of wage-tenure contracts studied in Burdett and Coles (2003). In principle, firms would like to make wages contingent on tenure to backload the incentives. Although studying such full dynamic contracts would be interesting, I believe such consideration is largely orthogonal to my focus: aggregate risk sharing.[24] Moreover, Burdett and Coles (2010) show that wage-tenure contracts with heterogenous firms massively complicates the analysis. For example, the equilibrium wage distribution need not necessarily be smooth.

The contract specifics the utility that firms deliver to workers at each point in time rather than the path of the wages. While this assumption is innocuous under perfect foresight equilibrium, it matters when hit by an unanticipated shock. Under the assumption that utility is specified in the contract, there is room for rewriting the contracts to adjust wages. Finally, as before, workers cannot commit to the contracts, and the possibility of counter-offers are excluded. With constant productivity and constant wage contracts, workers accept the arriving wage offers if and only if the offer is higher than the current wage.

Equilibrium objects. The unemployment rate, μ_t, evolves according to

$$\partial_t \mu_t = \delta(1-\mu_t) - \mu_t \lambda_t^U, \tag{1.23}$$

[24] One can justify my assumption if I let workers' elasticity of intertemporal substitution to be zero. Then, as Burdett and Coles (2003) showed, the optimal wage-tenure contract features a constant wage throughout the tenure.

where $\chi_t \equiv \frac{\mu_t}{\mu_t+\zeta(1-\mu_t)}$ and $\partial_t y_t \equiv \frac{\partial y_t}{\partial t}$ are the short-hand notation for the time derivative for any y_t. Let $P_t(w)$ denote the employment weighted wage distribution function. It follows the following law of motion:

$$\partial_t P_t(w) = -\delta P_t(w) - \lambda_t^E(1 - F_t(w))P_t(w) + \frac{1}{1-\mu_t}\mu_t \lambda_t^U F_t(w). \tag{1.24}$$

The value of a firm with productivity z per unit of employee that offers wage w satisfies[25]

$$\rho J_t(w, z) = A_t z - w - (\delta + \boldsymbol{\varkappa} + \lambda_t^E(1 - F_t(w)))J_t(w, z) + \partial_t J_t(w, z).$$

The firm choose what wages to offer and how much vacancies to post at time t by solving

$$\{w_t(z), v_t(z)\} \in \arg\max_{w,v} v\lambda_t^F Q_t(w) J_t(w, z) - A_t c(v; z), \tag{1.25}$$

where $Q(w) \equiv \chi_t \mathbb{I}(w \geq \underline{w}_t) + (1 - \chi_t)P_t(w)$ and $\underline{w}$ is the reservation wage for unemployed. The a worker's value function with wage w, $W(w)$, satisfies

$$\rho W_t(w) = u(w_t) + \delta \{U_t - W_t(w)\} + \lambda_t^E \int \max\{0, W(\tilde{w}) - W(w_t)\} dF_t(\tilde{w}) + \partial_t W_t(w), \tag{1.26}$$

where the value of unemployment, U_t, is given by

$$\rho U_t = u(A_t b) + \lambda_t^U \int \max\{0, W_t(\tilde{w}) - U_t\} dF_t(\tilde{w}) + \partial_t U_t. \tag{1.27}$$

The reservation wage for the unemployed, $\underline{w}_t$, must be such that workers are indifferent between being employed and unemployed $W_t(\underline{w}_t) = U_t$.

Appendix 1.12.1 defines the perfect-foresight equilibrium and characterizes the steady state of this economy with $A_t = A$.

Transition Dynamics in Response to Aggregate Shocks. As in the two-period model, I consider the following experiment. Before $t \leq 0$, the economy is in its steady-state. At $t = 0$, the economy experiences an unanticipated one-time increase in the variance of the

[25] I assumed away the possibility of endogenous separation (or exits). In principle, firms would like to fire workers if $J(w, z) < 0$. However, since firms at the exist threshold, $z = \underline{z}$, employ zero workers, even if I allow for the possibility of endogenous separation, there is no first order effect on the equilibrium outcomes.

aggregate productivity. The aggregate productivity for $t > 0$ is given by

$$\ln A_t = \begin{cases} \ln A_h = \ln A + d \ln A & \text{with probability } \pi_h \equiv 1/2 \\ \ln A_l = \ln A - d \ln A & \text{with probability } \pi_l \equiv 1/2 \end{cases}.$$

That is, the aggregate productivity is either permanently high or low for $t > 0$. The focus on permanent productivity shocks has been common in the search and matching literature (e.g., Ljungqvist and Sargent, 2017; Kehoe et al., 2019), and it is empirically reasonable given the high persistence on the productivity process.

Once firms and workers anticipate the aggregate risk, firms that already hire incumbent workers (re)write a state-contingent wage contracts at $t = 0$ that solve

$$\max_{\{w_{0s}^{inc}\}} \sum_{s \in \{h,l\}} \pi_s J_{0s}(w_{0s}, z)$$
$$\text{s.t.} \quad \sum_{s \in \{h,l\}} \pi_s W_{0s}(w_{0s}^{inc}) \geq W(w(z)),$$

where $s = h$ and $s = l$ denote the state with high and low productivity, respectively. In other words, firms offer a state-contingent wage that promise at least the same expected utility to a worker as before to maximize its expected profits. This is because I made an assumption that contracts are written in terms of the utility to be delivered. The optimal incumbent wage responses $w_{0s}^{inc}(z)$ satisfy the following first order condition:

$$\partial_w J_{0s}(w_{0s}^{inc}(z), z) + \eta(z) W'_{0s}(w_{0s}^{inc}(z)) = 0, \tag{1.28}$$

where $\eta(z)$ is the Lagrangian-multiplier on the promise-keeping constraint. After $t > 0$ onwards, given the incumbent wages as initial conditions, the equilibrium follows the perfect foresight path described above.

It is again worth noting that without aggregate risk-sharing (i.e., risk-neutral workers) or there is no on-the-job search, the economy exhibits no fluctuation in unemployment. Formally, Propositions 3 and 4 continue to hold, as established in Appendix 1.8.14. This ensures that it is precisely the complementarity between risk-sharing and on-the-job search that drives nontrivial labor market dynamics in my model. In what follows, I explore this complementarity by assuming $\gamma > 0$ and $\zeta > 0$.

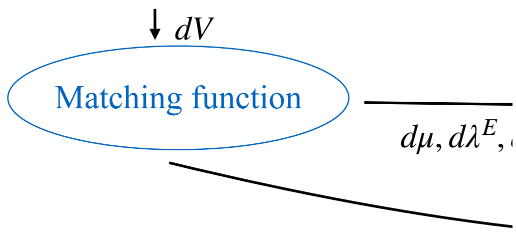


Figure 1-9: DAG representation of first order responses of the economy

Note: Figure 1-9 shows a directed acyclical graph (DAG) representation of the first order equilibrium responses, following Auclert et al. (2019). The economy takes the productivity shock dA as an exogenous input, and two endogenous variables, aggregate vacancy dV and the reservation wage $d\underline{w}$, as endogenous inputs. Given dV, one can compute the sequence of unemployment rates, $d\mu$, and meeting probabilities, $d\lambda^E, d\lambda^U, d\lambda^F$. Given $d\underline{w}$, $d\lambda^E, d\lambda^U, d\lambda^F$, and dA, one can compute the path of the distribution of wage and vacancies, $\{dw(z), dv(z)\}$, through a system of ODEs. Then, we can compute the implied aggregate vacancy and reservation wages to check the consistency. The key observation is that we do not need take the entire wage and vacancy distribution $\{dw(z_i), dv(z_i)\}_i$ as inputs (as indicated by a red diagonal line). In the figure, $d\bar{V}(z) \equiv \int^z d(v(\tilde{z})/V)dG(\tilde{z})$ is the cumulative vacancy distribution.

1.6.2 New Solution Method

As is well-known, solving the transition dynamics of the wage posting model has been considered as a challenge because of the need to keep track of the endogenous evolution of distribution. I develop a general and efficient computational approach to solve the transition dynamics of a wide class of wage posting job ladder models. Throughout, I focus on first order responses, which is crucial for my approach.

The key idea behind the computational algorithm is the same as how I solved the two-period model. Although it is widely believed that one needs to keep track of the path of wage and employment distribution to compute the equilibrium, I argue this is not the case. As long as we focus on the first order response, no firm cares about the entire distribution per se. Firms only care about the wages and vacancies of their neighbors. That is, best responses can be still represented as a system of ODEs in the infinite horizon model.

Figure 1-9 shows the DAG (directed acyclical graph) representation of the equilibrium, following Auclert et al. (2019). Instead of solving the fixed points of the entire distribution wages and vacancy, $\{dw(z), dv(z)\}$, if I have a guess of the reservation wage, $d\underline{w}$, I can compute what the least productive firms would offer, $dw(z_1) = d\underline{w}$. This, in turn, allows me to compute the wages and the vacancy of second least productive firms through the linearized best response, which is a system of ODEs. By continuing this logic, I can compute the entire path of the wage and vacancy distribution just by computationally climbing up the job ladder. In this process, I also need a guess of the path of aggregate vacancy, dV, to compute the path of matching probabilities. Therefore, I only need to iterate over a sequence of two endogenous variables, $\{d\underline{w}_t, dV_t\}$, instead of infinitely many endogenous variables. In solving for a fixed point of $\{dw_t(\underline{w}), dV_t\}$, I build on Auclert et al. (2019) to use the sequence-space Jacobian. Auclert et al. (2019) note that not covering wage posting models is a limitation of their methodology.[26] My contribution is to show that it is entirely possible for their methodology to cover wage posting models.

Among others, two key advantages of my approach are worth emphasizing. First, the computation is extremely efficient. It typically takes less than a second to compute the transition dynamics with a small number of grid points. Even with a large number of grid points, it only takes 1-5 seconds. While I do not pursue here, the efficiency of computation enables one to fully estimate the model using business cycle moments. Second, my approach does not require approximation of the distribution. An alternative approach to solve the first order transition dynamics of Burdett and Mortensen (1998) is to use Reiter's (2009) method to consider the first-order approximation in terms of state-space, as done by **?**. However, as emphasized by Auclert et al. (2019), such an approach is infeasible or requires approximating distribution with large state-space. Approximating distribution is especially not ideal in the context of wage-posting models because the best responses of firms depend on the entire shape of the wage distribution, as I showed in Lemma 14.[27]

There is also an alternative approach by Moscarini and Postel-Vinay (2016b) that solves the non-linear dynamics. Appendix 1.10 describes the solution method in more details.

1.6.3 Calibration

Functional form assumptions. The vacancy cost function is parametrized as the iso-elastic function, $c(v;z) \equiv \bar{c}z^{\iota_z}\frac{v^{1+1/\iota}}{1+1/\iota}$. The matching function is assumed to be a Cobb-Douglas form, $\mathcal{M}(\tilde{\mu}, V) = \bar{m}\tilde{\mu}^{1-\kappa}V^{\kappa}$. The firm's productivity distribution is parametrized as a Pareto distribution, $G(z) = 1 - (b/z)^{\Lambda}$ with $\Lambda > 1$ being the tail parameter.

Parameter values. Table 1.1 describes the calibration. The time frequency is monthly. I first set the elasticity of the matching function to $\kappa = 0.6$ following Blanchard and Di-

[26]In fact, they write "For instance, in the model of on-the-job search in Burdett and Mortensen (1998), agents take the full distribution of wages as an input to their decision problem, and it is impossible to represent this via a DAG of feasible dimension." To the contrary, I show it is possible, as I do in Figure 1-9.

[27]In contrast, in Bewly-Hugget-Aiyaragari models, only the mean of the (asset) distribution matters for interest rates and wages.

Parameter	Descripton	Value	Source/Target
Panel A. Externally assigned			
κ	Elasticity of matching function	0.6	Blanchard and Diamond (1989)
ρ	Dicount rate	0.004	5% annual interest rate
δ	Separation rate	0.016	1.6% EU rate
ι	Elasticity of vacancy cost function	1	Kaas and Kircher (2015)
Λ	Pareto tail of productivity distribution	5	S.d. of log productivity 0.2
$\varkappa$	Reallocation shock	0.0075	Share of EE with wage increase 1/2
$\bar{m}$	Matching efficiency	0.1	Normalization
b	Outside option of unemployed	1	Normalization
Panel B. Internally calibrated parameters			
$\bar{c}$	Vacancy cost parameter	0.035	Unemployment rate 6%
ι_z	Vacancy cost parameter	8	Aggregate profit share 14%
ζ	Relative efficiency of on-the-job search	0.08	1.5% EE rate
γ	Relative risk aversion	15	Wage volatility relative to output 38%

Table 1.1: Parametrization

Note: Table 1.1 describes the choice of the parameter values and their sources or targeted moments. Panel A shows the parameters exogenously assigned. Panel B shows the parameters that are internally calibrated to match the data moments.

amond (1989). The discount rate is set to match the 5% annual interest rate, $\rho = 0.004$. The separation rate is set to 1.6%, $\delta = 0.016$, corresponding to the average of the BLS labor status flow from employed to unemployed over the period of 1990-2019. I also set the matching efficiency parameter, $\bar{m} = 0.1$, which is a normalization. The elasticity of vacancy creation is set to $\iota = 1$, following Kaas and Kircher (2015) and Gertler and Trigari (2009), and I normalize $b \equiv 1$. The tail parameter of productivity is set so that the standard deviation of log productivity is 0.2, which corresponds to the lower end of the value reported in Foster et al. (2016). I set $\varkappa$ so that half of job changers experience wage increases, which follows Gertler et al. (2020) and Moscarini and Postel-Vinay (2016a).

I choose $\{\bar{c}, \iota_z, \zeta, \gamma\}$ to match the (i) steady-state unemployment rate of 6%; (ii) monthly job-to-job transition rate of 1.5%; (iii) the aggregate profit share of 14% reported by Gutiérrez and Philippon (2018) for the US; and (iv) the relative standard deviation of average real wage growth to the real output growth of 0.38. I use the real output in the non-farm business sector as a measure of the real output, and average hourly earnings of production and nonsupervisory employees deflated by PCE as a measure of the real wage. Both series are obtained from BLS. I explicitly target the profit share because this is the key determinant of fluctuations in job creation for the given level of wage rigidity, as can

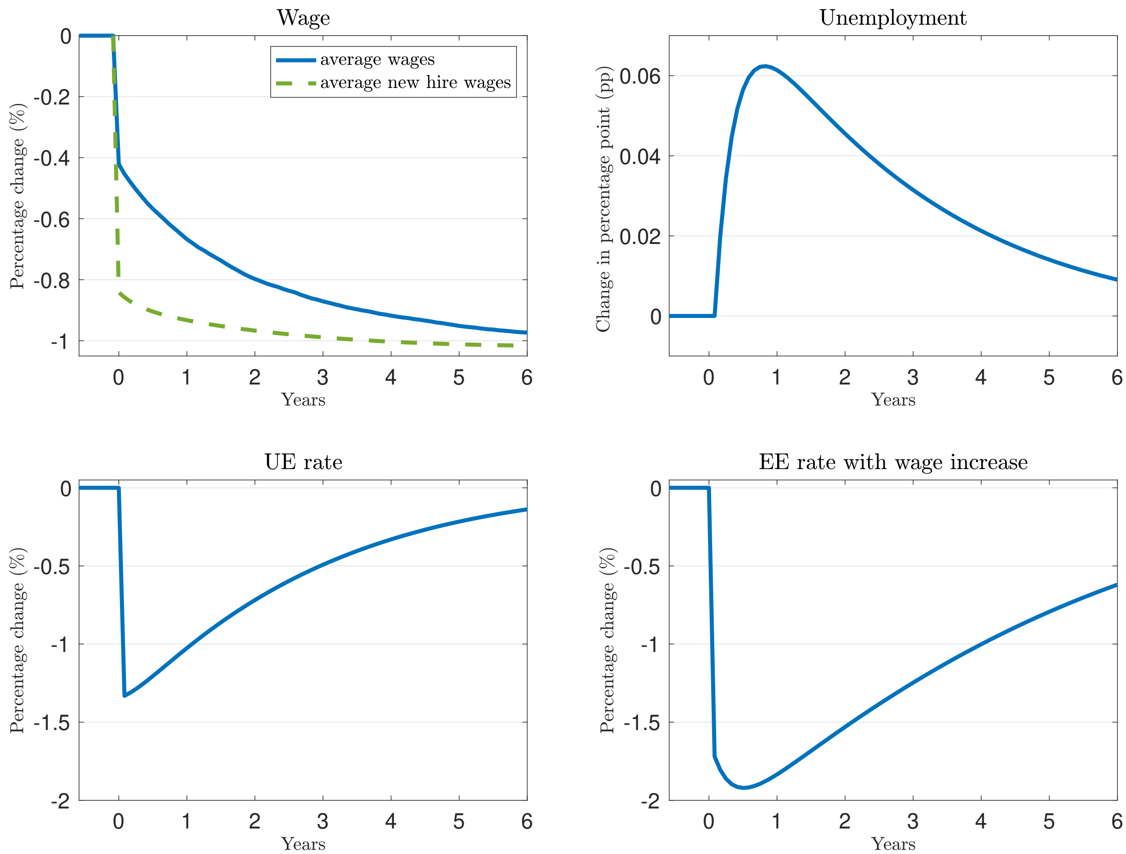


Figure 1-10: Impulse response to the realization of a 1% negative productivity shock

Note: Figure 1-10 shows the impulse response of the economy to the realization of a 1% negative productivity shock.

been from equation (1.15) (see also Ljungqvist and Sargent (2017)). To match the relative volatility of real wages, the model requires a relative risk aversion of 15. This value is higher than the most macro models, but is fairly consistent with the value used in the finance literature. The reason why I need a relatively high risk aversion is that the presence of on-the-job search implies that between firm competition acts as a strong force in preventing effective risk-sharing. This is in contrast to Rudanko (2009). She uses a model with risk-sharing but without on-the-job search and shows that the model tends to deliver too rigid wages compared with the data.

1.6.4 Results

Figure 1-10 shows the impulse response function for the realization of a 1% negative productivity shock. The left-top panel shows the dynamics of average wages in the blue solid line and the average new hire wages in the dashed green line. The average wage is substantially sluggish mainly because firm insurance induces a muted incumbent wage response. The new hire wage response is also dampened through the strategic complementarity highlighted in the two-period model. Compared with the fully flexible case, the initial new hire wage response is dampened by 15-20%. The magnitude is smaller than the two-period model because when potential new employers decide on their wage offers, this not only takes into account the competition with incumbent firms but also with the subsequent potential new employers. Since wages need to eventually adjust fully in the long-run, subsequent potential new employers offer more flexible wages. To prevent being poached by those firms, current firms have an incentive to offer more flexible wages than the two-period model.

The right-top panel shows the response of unemployment rate. Note that with risk-neutral workers, $\gamma = 0$, there should not be any response of unemployment rate. As in two-period model, as soon as we move away toward risk-averse workers, the model does generate unemployment fluctuations, which is in stark contrast to the conventional wisdom that long-term contracts should not contribute to unemployment fluctuations. The bottom left and the bottom right panels show the UE rate and the EE rate with a wage increase, respectively. EE rate with a wage increase declines more sharply and recovers more slowly than the UE rate. This collapse in the number of workers who climb up the job ladder is consistent with the fact documented in Haltiwanger et al. (2018). They show that the firm wage ladder is strongly procyclical, and my theory provides a natural explanation of this.[28]

Decomposing unemployment response. The model generates a sluggish adjustment in both incumbent and new hire wages as well as volatility in unemployment. It is then natural to ask what drives unemployment fluctuation: is it incumbent wage rigidity or new hire wage rigidity? The answer to this question was stark in the two-period model, but it is not in an infinite horizon model. In the infinite horizon model, sluggish adjustments in future new hire wages lowers the probability of being poached in the future for the current firms, which raises the incentive to create jobs.

To shed light on this issue, I exogenously change each of the incumbent wage and the path of new hire wages separately and simulate the model. In the first experiment, I force all the incumbent firms to adjust wage one for one with productivity holding the path of new hire wages fixed. In the second experiment, I force all the new hire wages to adjust one for one with productivity, holding incumbent wages fixed. Figure 1-11 shows

[28] See also Barlevy (2002), Mukoyama (2014), and Nakamura et al. (2019) for related evidence.

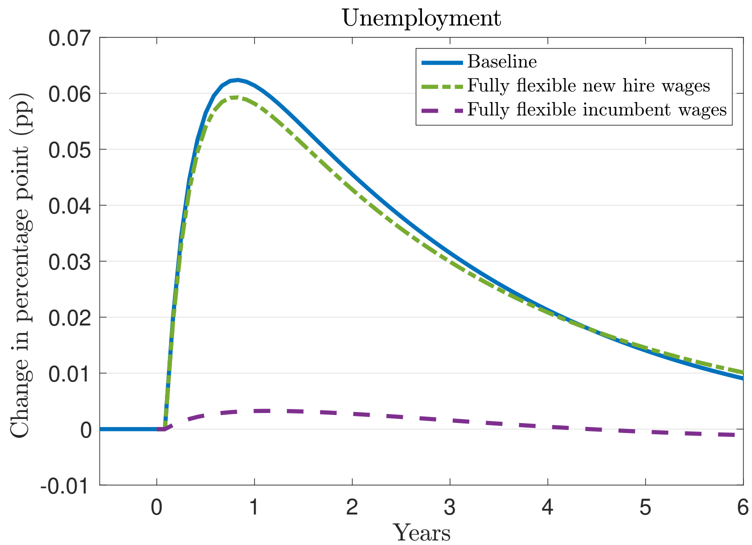


Figure 1-11: Decomposition of the unemployment response

Note: Figure 1-11 shows the decomposition of the impulse response of the unemployment rate. The green dash-dotted line assumes new hire wages respond one for one with the aggregate productivity, holding the incumbent wages the same as the basline response. The purple dashed line assumes the incumbent wages respond one for one with the aggregate productivity, holding the new hire wages the same as the basline response.

the response of unemployment under each counterfactual scenario. One can immediately see that most of the unemployment response disappears if incumbent wages are flexible. In contrast, the response is barely affected even if new hire wages are fully flexible. This decomposition shows that my results are indeed driven by incumbent wage rigidity.

Business cycle moments. Table 1.2 compares the business cycle moment of the model to the data. By design, the model matches the standard deviation of real wage growth. The model generates roughly 20% of the volatility in the UE rate and vacancy. Since the volatility in the unemployment rate not only comes from the UE rate but also fluctuations in separation rate, which I abstract from, my model generates volatility in unemployment rate smaller than 20% of the data. To make a fair comparison, I construct a time-series of the unemployment rate that assumes constant separation rate, following Shimer (2012). Specifically, the adjusted unemployment rate is given by $u_{t+1}^{adj} = \delta(1-u_t^{adj}) + (1-UE_t)u_t^{adj}$, where $\delta = 1.6\%$ and UE_t is the UE rate taken from the data. The model explains roughly 20% of volatility of this variable.

The magnitude of volatility is relatively small compared with the data. This comes from two reasons. The first reason is relatively standard. I have chosen parameters so

Moments	**Model**	**Data**
Panel A. Relative s.d. to real output growth		
average real wage growth	0.38	0.38
UE rate	1.36	6.97
log vacancy	7.36	37.9
unmployment rate	0.30	2.23
constant seperation unmployment rate	0.30	1.61
Panel B. Autocorrelation		
real wage growth	0.10	0.18
UE rate	0.97	0.96
unemployment rate	0.99	0.99
log vacancy	0.95	0.98
Panel C. Correlation with unemployment		
real wage growth	-0.20	-0.13
UE rate	-0.96	-0.83
log vacancy	-0.93	-0.52

Table 1.2: Business cycle moments

Note: Table 1.2 shows the business cycle moments in the model and in the data. The real output measure is the real output in nonfarm business sector from BLS. The real wage is average hourly earnings of production and nonsupervisory employees deflated by PCE, also from BLS. The constant separation unemployment rate assumes the EU rate is constant at $\delta = 1.6\%$. Vacancy data comes from the composite Help-Wanted index by Barnichon (2010).

that the aggregate profit share is 15%, which is consistent with the data. As is well-known, search and matching models require low surplus (profit share) to generate amplifications (Ljungqvist and Sargent, 2017). This is true in my model, as equation (1.15) crucially depends on $\alpha(z)$, the profit share. Using a standard DMP model with a representative firm, **?**, Hagedorn and Manovskii (2008), and others have been able to generate realistic volatility in labor market because they have assumed that the profit share is less than 5%. It is difficult for the wage posting job ladder model to deliver such low profit share with reasonable heterogeneity in firm productivity. Highly productive firms are profitable, so they become large in size, raising the aggregate profit share. Since these channels are well understood and not my focus, I do not pursue an approach to engineer my model to generate a low profit share.

Second reason comes from the type of wage rigidity that matters for unemployment fluctuations in my model. As I will shortly explain in detail in Section 1.6.5, my model gives an important role not only to incumbent wage rigidity but also to a full dynamic response of wages: how much the wages will adjust over the very long horizon. Since

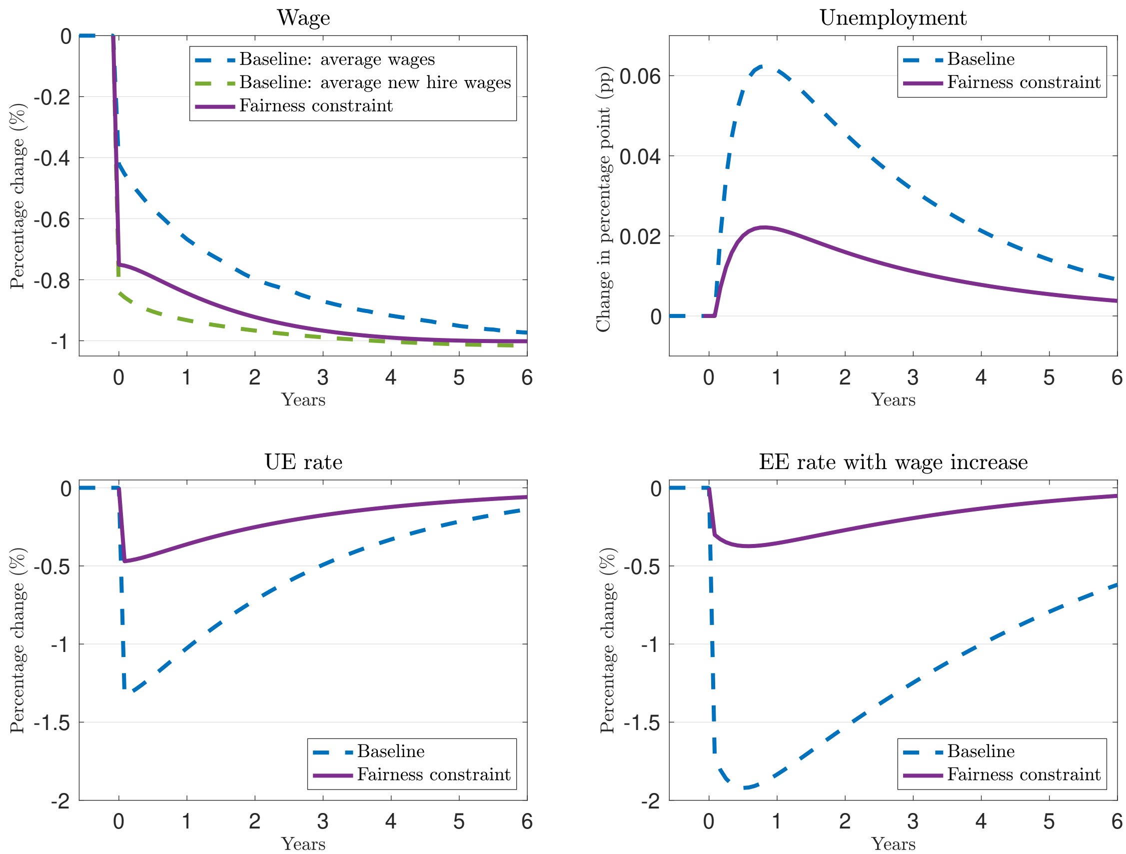


Figure 1-12: Impulse response with fairness constraints

Note: Figure 1-12 shows the impulse response of the economy to the realization of a 1% negative productivity shock with and without fairness constraints.

wages will be fully flexible in the long-run (after 4-5 years in my model), this tends to diminish the amplification of the model.

Fairness constraints. I revisit the question of whether the fairness constraint amplifies or dampens unemployment fluctuations by using this quantitative model. This is interesting also from the perspective of the literature. The literature that uses the wage posting job ladder model to study business cycle almost always imposed fairness constraints, following the tradition of Burdett and Mortensen (1998) (e.g., **??**). It is worth clarifying the role that such a constraint was playing in these papers.

I impose a restriction that firms cannot discriminate wages across employees. Firms commit to a sequence of wage payments $\{w_s\}$ that delivers W_t of the expected lifetime utility to the workers employed at the firm. Workers accept the job that offers a higher value. I delegate the detail description of the environment to Appendix 1.12.2.

Figure 1-12 shows the impulse response with fairness constraints. The top left panel shows the wage response, and as one would expect, the wage response lies in the middle of new hire and average wages in the baseline model. The right-top panel shows that fairness constraints indeed dampens unemployment response by around 70%. This has two implications. First, the common practice of imposing fairness constraints in the wage posting job ladder model tends to worsen Shimer puzzle. Second, while some papers argue that fairness constraints amplify unemployment fluctuations, the implications are reversed once we take into account on-the-job search. The bottom two panels show the response of the UE and the EE rate. As in the two-period model, fairness constraints dampen the EE response much more than the UE response.

1.6.5 Which Wage Rigidity Matters?

In this infinite horizon model, what type of wage rigidity is relevant for the incentive to create jobs? The answer to this question helps us understand the above simulation results. I explain it using the notion of how much job values are sensitive to wage changes at each point in time.

Let us focus on the baseline infinite horizon model without fairness constraints. I consider its discrete time approximation where a period corresponds to a month. The value of job with productivity z and a wage contract w is

$$J_t(w,z) = Az - w + (1 - (\hat{\rho} + \lambda^E_{t+1}(1 - F_{t+1}(w)))J_{t+1},$$

where $\hat{\rho} \equiv \rho + \delta + \boldsymbol{\varkappa}$. The optimality condition for vacancy creation at $t = 1$ is

$$\lambda^F_1 Q_1(w) J_1(w,z) = Ac'(v).$$

I consider the following thought experiment: Before $t = 1$, the economy is in a steady-state. Suppose at $t = 1$, there is exogenous increase in A combined with the arbitrary changes in wage distribution, $\{w_t(z)\}_{t,z}$, (including its own wage), with $w_0(z)$ being the incumbent wages of firm z. How do the changes in the wage distribution affect the value of job creation? The exercise is the partial equilibrium, so I fix all other variables (λ^F_t, λ^E_t, and vacancies) fixed. Therefore, we are interested in

$$\mathcal{E}_{1,t}(z,\tilde{z}) \equiv \frac{\partial \ln \left(Q_1(w(z))J_1(w(z),z)\right)}{\partial \ln w_t(\tilde{z})}.$$

First, it is straightforward to see $\mathcal{E}_{1,t}(z,\tilde{z}) = 0$ for $z \neq \tilde{z}$: small wage changes of infra-

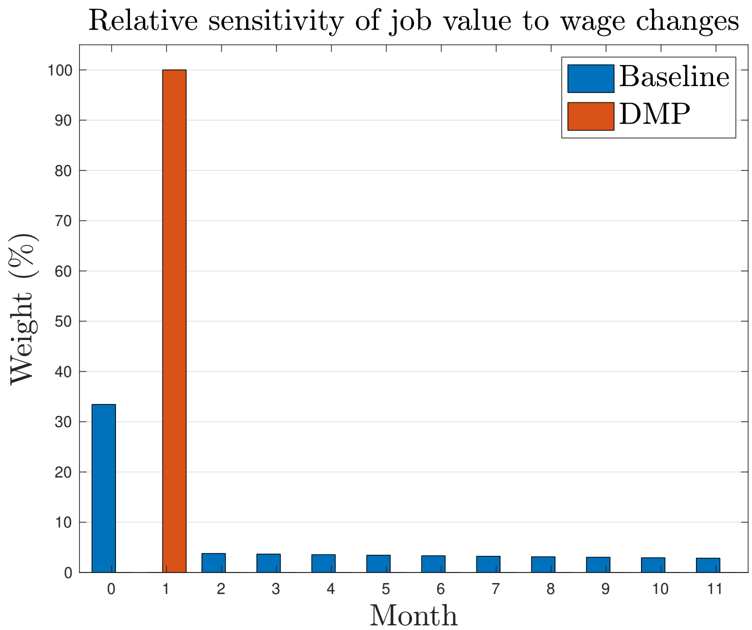


Figure 1-13: Relative sensitivity of job value to wage changes in each point in time

Note: Figure 1-13 plots $weight_t$, defined in (1.29), which captures the relative sensitivity of job value to wages in each point in time. Month 1 corresponds to the sensitivity to the new wages when the job is created. Month 0 corresponds to the sensistivity to incumbent wages. Month $t > 1$ corresponds to the sensitivity to wage offer at month t.

marginal competitors do not affect the job value. Second,

$$\sum_{s=0}^{\infty} \mathcal{E}_{1,s}(z,z) = -\frac{1-\alpha(z)}{\alpha(z)},$$

where $\alpha(z) \equiv (Az - w(z))/Az$ is the profit share. I define

$$weight_t \equiv \int \frac{\mathcal{E}_{1,t}(z,z)}{\sum_{s=0}^{\infty} \mathcal{E}_{1,s}(z,z)} (v(z)/V) dG(z), \tag{1.29}$$

which captures how much the value of job is sensitive to the wage changes at each point in time after integrated using vacancy as density.

The blue bar in Figure 1-13 shows the weight in the baseline model. First, it places 35% of the weight on incumbent wages ($t = 0$). Second, it places 0% of the weight on contemporaneous wage changes. Third, the weight is spread over the entire period with each having 3-4%. The first result indicates that the incumbent wage rigidity is the most important determinant of job creation. The second result comes from the envelope theorem as in the two-period model. The third result comes from the fact that firms face a constant threat of being poached and that higher wage offers in the future makes this possibility more likely. All these results are consequence of dynamic competition in the labor market. Firms that post a job today not only compete with incumbent firms to poach workers, but also with future poaching firm.

The red bar in Figure 1-13 shows the same object in DMP model. In stark contrast, DMP models put 100% weight on contemporaneous wage. The reason is that the labor market competition is completely absent in this class of models. The value of a job in DMP model is

$$J_t^{DMP}(w,z) = Az - w + (1-(\rho+\delta))J_{t+1}^{DMP},$$

and the optimal vacancy creation is

$$\lambda_1^F J_1^{DMP}(w,z) = Ac'(v),$$

which does not depend on any other wages than its own wage, $\mathcal{E}_{1,s}(z,z) = 0$ for all $s \neq 1$. One can also compute that $\mathcal{E}_{1,s}(z,z) = -\frac{1-\alpha(z)}{\alpha(z)}$, so that $\sum_{s=0}^{\infty} \mathcal{E}_{1,s}(z,z) = -\frac{1-\alpha(z)}{\alpha(z)}$, which implies that the total response is the same with the baseline model, but the importance of the wage rigidity in each point in time completely differ.

This has an implication for measuring the theoretically relevant notion of wage rigidity. Since Bils (1985), it has been common to estimate *contemporaneous* wage rigidity, which is the contemporaneous correlation between wage changes and unemployment rate. While this is theoretically well grounded from the viewpoint of DMP model, it is not if one believes in the wage posting model with on-the-job search. As Figure 1-13 shows, the theory implies that we need to measure *intertemporal* wage rigidity, which consists of (i) incumbent wage rigidity, and (ii) how the wages at time s respond to the aggregate shock at time $t < s$. This comes from the fact that in this class of the job ladder model, labor market competition is inherently dynamic. Firms that intend to create jobs today need to compete with incumbent firms and future jobs. Consequently, those competitors' wage responses become the important determinant of job creation.

1.7 Conclusion

Is incumbent wage rigidity important for unemployment fluctuations? Conventional wisdom says no. My paper says yes by arguing that the key missing piece in the conventional view is on-the-job search. Models of wage rigidity have been abstracting from on-the-job search, thereby mechanically shutting down any meaningful interaction between incumbent wages and labor market dynamics. I showed that once we take into account on-the-job search in an environment where firms insure incumbent workers, (i) both new hire and incumbent wages are endogenously rigid; (ii) but only the latter form of wage rigidity is the key determinant for unemployment fluctuations.

I operationalize the idea using a generalized version of Burdett and Mortensen's (1998) job ladder model featuring risk-neutral firms, risk-averse workers, and aggregate risk. Besides the main messages, I showed a number of other results such as the fact that fairness constraints and public insurance dampen unemployment fluctuations, and the novel source of inefficiency makes wages too flexible in equilibrium. Overall, I believe my theory provides a useful starting point in rethinking the nature and the consequence of wage rigidity in an arguably more realistic labor market model than the canonical DMP model.

I conclude by discussing several avenues for future research. First, my model features wage rigidity that is symmetric between booms and recessions, because of the first order approximation. I conjecture that my model will feature downward wage rigidity with a higher order approximation, through the mechanism of Harris and Holmstrom (1982). Since downward wage rigidity is the pervasive feature of the data, it would be promising to study its interaction with the labor market dynamics in a micro-founded manner. Second, while it has been common to assume an exogenously incomplete market in the heterogenous household literature, my model features an endogenously incomplete market through firm insurance subject to limited commitment frictions. It would be interesting to add consumption and saving decisions in my model to study the interaction between aggregate demand, equilibrium wage rigidity, and labor market dynamics. Third, while I mostly focused on theoretical aspects, my theory provides a new angle for looking at the data. For example, it would be fruitful to look into the relationship between the prevalence of on-the-job search, wage rigidity, and employment fluctuations at various levels of disaggregation.

Appendix

1.8 Proofs

1.8.1 Proof of Lemma 1

The derivation of (1.9) and (1.10) are provided in the main text. We need to show that the second-order condition for the potential new employers

$$F_0''(w(z))(Az - w(z)) - 2F_0'(w(z)) < 0$$

is satisfied. Totally differentiating (1.7) gives

$$F_0''(w(z))(Az - w(z)) = -F_0'(w(z))A\frac{1}{w'(z)} + 2F_0'(w(z)).$$

Therefore the second-order condition is

$$\begin{aligned} F_0''(w(z))(Az - w(z)) - 2F_0'(w(z)) &= -F_0'(w(z))A\frac{1}{w'(z)} \\ &< 0 \end{aligned}$$

since $F'(w(z)) > 0$ and $w'(z) > 0$.

Now consider incumbent firms. We have to guarantee that the promise-keeping constraint is binding. It is enough to impose the following assumption.

Assumption 1. *Parameters are such that $\Pi_0(w; z) \equiv (Az - w)(1 - \lambda^E + \lambda^E F_1(w))$ is decreasing in w for all z, where $F_1(w) \equiv \int_{w \geq w(z)} v(\tilde{z})/V dG(\tilde{z})$, and $w(z), v(z), \lambda^E, V$ are given by (1.9), (1.10) and (1.11).*

The assumption is always satisfied as long as λ^E is small enough. In fact, if the cost of vacancy is such that the vacancy is constantly proportional to employment, $v(z) = \bar{v}\ell_0(z)$,[29] it is sufficient to have $\lambda^E < 1 - \chi$. Empirically, the share of employer-to-employer transitions among new hires is 40%, which implies $1 - \chi = 0.4$, while employer-to-employer transition rate at the quarterly frequency is around 5%, which implies $\lambda^E \approx 0.1$.[30] Therefore the assumption is arguably natural to impose. Under Assumption 1,

[29] This is empirically reasonable. Davis et al. (2013) document that vacancy rate ($v(z)/\ell_0(z)$) is uncorrelated with firm-size measures.

[30] Under the constant vacancy rate, $v(z)/\ell_0(z)$, the half of workers transition to new employer conditional on meetings.

equation (1.4) implies that the constraint must be binding, $\eta(z) > 0$. As the worker's utility is strictly increasing in $w_0(z)$, if

$$\bar{W}_0(z) = (1-\lambda^E)u(w(z)) + \lambda^E \int \max\{u(w(z)), u(w(\tilde{z}))\}(v(\tilde{z})/V)dG(\tilde{z}),$$

the incumbent firms have to set $w_0(z) = w(z)$. These complete that the equilibrium has the the properties claimed.

1.8.2 Proof of Lemma 2

The set of equilibrium conditions are

$$\begin{aligned}
(1-\chi)F'_{0s}(w_{1s}(z))(A_s z - w_s(z)) - (\chi + (1-\chi)F_{0s}(w_{1s}(z))) = 0 \\
-(1-\lambda^E_s + \lambda^E_s F_{1s}(w_{0s}(z))) + (A_s z - w_s(z))\lambda^E_s F'_{1s}(w_{0s}(z)) \\
+\eta(z)\left[(1-\lambda^E_s)u'(w_{0s}(z)) + \lambda^E_s F_{1s}(w_{0s}(z))u'(w_{0s}(z))\right] = 0 \\
\sum_{s\in\{h,l\}} \pi_s \left[(1-\lambda^E_s)u(w_{0s}(z)) + \lambda^E_s \int \max\{u(w_{0s}(z)), u(\tilde{w})\}dF_{1s}(\tilde{w})\right] = \bar{W}_0(z) \\
(A_s z - w_{1s}(z))\lambda^F_s(\chi + (1-\chi)F_{0s}(w_{1s}(z))) = c'_s(v_s(z); z) \\
\lambda^F_s = \frac{1}{V}\mathcal{M}(\tilde{\mu}, V), \quad \lambda^E = \zeta\frac{\mathcal{M}(\tilde{\mu}, V)}{\tilde{\mu}} \quad \text{with } V = \int v(z)dG(z)
\end{aligned}$$

Applying the generalized implicit function theorem (Luenberger, 1969) jointly to $\{w_{0s}(z), w_{1s}(z), v_s(z), V$
with respect to $d \ln A$, we have

$$\begin{aligned}
\frac{d\ln w_{1s}(z)}{d\ln A} &= \theta_{1a}(z)(\mathbb{I}(s=h)-\mathbb{I}(s=l)) + \theta_{1w}(z)\frac{d\ln w_{0s}(z)}{d\ln A} - \theta_{1a}(z)\alpha(z)\frac{w(z)}{w'(z)}\frac{d}{dz}\frac{d\ln w_{0s}(z)}{d\ln A} \\
\frac{d\ln w_{0s}(z)}{d\ln A} &= \theta_{0a}(z)(\mathbb{I}(s=h)-\mathbb{I}(s=l)) + \theta_{0w}(z)\frac{d\ln w_{1s}(z)}{d\ln A} - \theta_{0a}(z)\alpha(z)\frac{w(z)}{w'(z)}\frac{d}{dz}\frac{d\ln w_{1s}(z)}{d\ln A} \\
&\quad + \theta_{0a}(z)\alpha(z)\left\{1-\theta_{\lambda,p}(z)\right\}\frac{d\ln\lambda_h^E}{d\ln A} \\
&\quad + \theta_{0a}(z)\alpha(z)\theta_{\lambda,r}(z)\left(\frac{1}{\int^z v(\tilde{z})dG(\tilde{z})}\int\frac{dv_s(\tilde{z})}{d\ln A}dG(\tilde{z}) - \frac{1}{V}\frac{dV_h}{d\ln A}\right) \\
&\quad + \theta_{0a}(z)\alpha(z)\left(\frac{1}{v(z)}dv_h(z) - \frac{1}{V}dV_h\right) \\
&\quad + \theta_{0a}(z)\left[(1-\lambda^E)u'(w_0(z)) + \lambda^E F_1(w_0(z))u'(w_0(z))\right]\frac{d\eta(z)}{d\ln A} \\
0 &= (1-\lambda^E+\lambda^E F_1(w(z)))u'(w(z))w(z)\sum_s\frac{d\ln w_{0s}(z)}{d\ln A} \\
&\quad + \int\sum_s\frac{d\ln w_{1s}(\tilde{z})}{d\ln A}(v(\tilde{z})/V)dG(\tilde{z}) \\
&\quad + \lambda^E\left(\int\max\{u(w(z)),u(\tilde{w})\}dF_1(\tilde{w}) - u(w(z))\right)\frac{d\ln\lambda_s^E}{d\ln A} \\
&\quad + \lambda^E\int_z u(w(\tilde{z}))\sum_s\left(\frac{1}{V}\frac{dv_s(\tilde{z})}{d\ln A} - \frac{v(\tilde{z})}{V^2}\frac{dV_s}{d\ln A}\right)dG(\tilde{z}) \\
\frac{1}{v(z)}\frac{dv_s(z)}{d\ln A} &= \iota\left[\frac{d\ln\lambda_s^F}{d\ln A} + \frac{1-\alpha(z)}{\alpha(z)}\left(\mathbb{I}(s=h)-\mathbb{I}(s=l)-\frac{d\ln w_{0s}(z)}{d\ln A}\right)\right] \\
\frac{d\ln\lambda_s^F}{d\ln A} &= (\kappa-1)\frac{1}{V}\frac{dV_s}{d\ln A} \\
\frac{d\ln\lambda_s^E}{d\ln A} &= \kappa\frac{1}{V}\frac{dV_s}{d\ln A} \\
\frac{dV}{d\ln A} &= \int\frac{dv_s(z)}{d\ln A}dG(z),
\end{aligned}$$

where I used the assumption that $\pi_s = 1/2$. One can see that all the endogenous variables enter symmetrically between two states. We also know that at the top, wage offers must be symmetric

Lemma 4. *To a first order,* $\frac{d\ln w_{1h}(\bar{z})}{d\ln A} = \frac{d\ln w_{0h}(\bar{z})}{d\ln A} = -\frac{d\ln w_{1l}(\bar{z})}{d\ln A} = -\frac{d\ln w_{0l}(\bar{z})}{d\ln A}$

Proof. I proceed in four steps.

Step 1: $w_{1s}(\bar{z}) \leq w_{0s}(\bar{z})$ for $s \in \{h, l\}$ in equilibrium. Suppose not: $w_{1s}(\bar{z}) > w_{0s}(\bar{z})$

holds in equilibrium. Then the new hire firms can strictly increase profits by slightly lowering the wage (no change in labor supply, but lower costs). This is a contradiction.

Step 2: to a first order, $\frac{d\ln w_{0h}(\bar{z})}{d\ln A} = -\frac{d\ln w_{0l}(\bar{z})}{d\ln A}$. This is implied by the promise-keeping constraint at the top:

$$\sum_{s\in\{h,l\}} \pi_s u(w_{0s}(\bar{z})) = \bar{W}_0(\bar{z}) \tag{1.30}$$

because the first step implies that there cannot be higher wage offers than $w_{0s}(\bar{z})$ in equilibrium.

Step 3: $w_{0h}(\bar{z}) = w_{0l}(\bar{z})$. Suppose to the contrary $w_{0h}(\bar{z}) > w_{1h}(\bar{z})$. Then by slightly reducing $w_{0h}(\bar{z})$ and raising $w_{0h}(\bar{z})$ by the same amount will (i) weakly increases the labor supply, and (ii) relaxes the constraint $\sum_{s\in\{h,l\}} \pi_s u(w_{0s}(\bar{z})) \geq \bar{W}_0(\bar{z})$. This is a contradiction that $w_{0h}(\bar{z})$ was optimally set. Combined with Claim 1, $w_{1h}(\bar{z}) = w_{0h}(\bar{z})$.

Step 4: $\frac{d\ln w_{0l}(\bar{z})}{d\ln A} = \frac{d\ln w_{1l}(\bar{z})}{d\ln A}$. Suppose to the contrary that $\frac{d\ln w_{0l}(\bar{z})}{d\ln A} > \frac{d\ln w_{1l}(\bar{z})}{d\ln A}$. Then consider a perturbation of incumbent firms strategy that changes $\frac{d\ln w_{0l}(\bar{z})}{d\ln A}$ by $\Delta w_{0l}(\bar{z}) < 0$ and changes $\frac{d\ln w_{1l}(\bar{z})}{d\ln A}$ by $\Delta w_{0h}(\bar{z}) > 0$, with $\Delta w_{0l}(\bar{z}) = -\Delta w_{0h}(\bar{z})$. Around a symmetric steady-state, this does not impact worker's welfare to a first order, and therefore does not affect the constraint (1.30):

$$\begin{aligned}\sum_{s\in\{h,l\}} \pi_s u(w_{0s}(\bar{z})) &= \pi_s u'(w(\bar{z}))w(\bar{z})\left(\Delta w_{0l}(\bar{z}) + \Delta w_{0h}(\bar{z})\right) \\ &= 0.\end{aligned}$$

However, this has the first order increase in labor supply:

$$\Delta(\text{labor supply}) = \underbrace{0 \times \Delta\hat{w}_{0l}(\bar{z})}_{\text{because there was no mass in the neighborhood}} + \underbrace{F'(w(\bar{z})) \times \Delta\hat{w}_{0h}(\bar{z})}_{\text{because there is a mass of competittors (from Step 3)}},$$

which in turn implies that this has a first order increase in profits. This is a contradiction that $\frac{d\ln w_{0l}(\bar{z})}{d\ln A}$ was optimum.

From Step 2, 3 and 4, we confirm

$$\frac{d\ln w_{1h}(\bar{z})}{d\ln A} = \frac{d\ln w_{0h}(\bar{z})}{d\ln A} = -\frac{d\ln w_{1l}(\bar{z})}{d\ln A} = -\frac{d\ln w_{0l}(\bar{z})}{d\ln A}.$$

□

Therefore Lemma 4 implies that at the boundary, $z = \bar{z}$, wage responses must be symmetric between two states. Then given all the coefficients in the system of ODEs enter symmetrically between two states, any solution has to be symmetric as well.

1.8.3 Proof of Lemma 14

Linearizing the new hire firms' FOC (1.5) gives

$$(1-\chi)F_{0s}'(w_{1s}(z))(A_s z - w_s(z)) - (\chi + (1-\chi)F_{0s}(w_{1s}(z))) = 0$$

$$w(z)\left[-F_0''(w(z))(Az-w(z)) + 2F_0'(w(z))\right]\hat{w}_{1s} \tag{1.31}$$
$$= F_0'(w(z))Az\hat{A} + (Az-w(z))\partial_{\mathbf{w}_0}F_{0s}(w(z)) - \partial_{\mathbf{w}_0}F_{0s}(w(z)), \tag{1.32}$$

where $\partial_{\mathbf{w}_0}$ denote the partial derivative with respect to entire distribution of $\{w_0(z)\}$. Using the fact that

$$\begin{aligned}
\partial_{\mathbf{w}_0}F_{0s}(w_{0s}(z)) &= -F_0'(w(z))w(z)\hat{w}_{0s}(z)\\
\partial_{\mathbf{w}_0}F_{0s}'(w_{0s}(z)) &= \partial_{\mathbf{w}_0}\left(\frac{\ell_0(z)g(z)}{w_{0s}'(z)}\right)\\
&= \partial_{\mathbf{w}_0}\left(\frac{\ell_0(z)g(z)}{w_{0s}(z)\frac{d\ln w_{0s}(z)}{dz}\Big|_{z=\zeta_{0s}(w)}}\right) \quad \text{(where } \zeta_{0s} \text{ is the inverse function of } w_{0s}\text{)}\\
&= -\frac{\ell_0(z)g(z)}{w(z)\frac{d\ln w(z)}{dz}}d\ln w_{0s}(z) - \frac{\partial}{\partial z}\left(\frac{\ell_0(z)g(z)}{w(z)\frac{d\ln w(z)}{dz}}\right)\zeta_0'(w_0(z))w_0(z)d\ln w_{0s}(z)\\
&\quad - \frac{\ell_0(z)g(z)}{w(z)\frac{d\ln w(q)}{dz}}\frac{1}{\frac{d\ln w(z)}{dz}}d\frac{d\ln w_{0s}(q)}{dz}\\
&= -F_0'(w(z))d\ln w_{0s}(z) - \frac{\partial}{\partial z}\left(\frac{\ell_0(z)g(z)}{w(z)\frac{d\ln w(z)}{dz}}\right)\frac{1}{w'(z)}w(z)d\ln w_0(z)\\
&\quad - F_0'(w(z))\frac{w(z)}{w'(z)}d\frac{d\ln w_{0s}(z)}{dz}\\
&= -F_0'(w(z))d\ln w_{0s}(z) - F_0''(w(z))w(z)d\ln w_0(z) - F_0'(w(z))\frac{w(z)}{w'(z)}\hat{w}_0'(z),
\end{aligned}$$

on can rewrite (1.31) as

$$[2(1-\alpha(z)) - \alpha(z)\eta_{F_0}(z)]\,\hat{w}_1(z) = \hat{A} + [2(1-\alpha(z)) - \alpha(z)\eta_{F_0}(z) - 1]\,\hat{w}_0(z) - \alpha(z)\frac{w(z)}{w'(z)}\hat{w}_0'(z).$$

Rearranging, one obtains (1.12).

Similarly, linearizing incumbent firms' FOC (1.4),

$$
\begin{aligned}
-\lambda^E F'(w(z))w(z)\hat{w}_{0s}(z) - \lambda^E \partial_{\mathbf{w}_1} F_{1s}(w) + Az\lambda^E F'(w)\hat{A}_s \\
+(A_s z - w_s)\lambda^E F''(w) w\hat{w}_{0s}(z) + (A_s z - w_s)\lambda^E \partial_{\mathbf{w}_1} F'_{1s}(w_{0s}) \\
+\eta(1-\lambda^E)u''(w)w\hat{w}_{0s}(z) + \eta\lambda^E F(w)u''(w)w\hat{w}_{0s}(z) + \eta\lambda^E f(w)u'(w)\hat{w}_{0s}(z) \\
+\eta\lambda^E u'(w_s)\partial_{\mathbf{w}_1} F_{1s}(w_{0s}) + d\eta(z)\left[(1-\lambda^E)u'(w_s) + \lambda^E F(w_s)u'(w_s)\right] = 0.
\end{aligned}
$$

One can use symmetry from Lemma 2 to eliminate $d\eta(z)$:

$$
\begin{aligned}
-\lambda^E F'(w(z))w(z)\hat{w}_0(z) - \lambda^E \partial_{\mathbf{w}_1} F_{1s}(w) + Az\lambda^E F'(w)\hat{A} \\
+(Az - w(z))\lambda^E F''(w) w\hat{w}_0(z) + (Az - w(z))\lambda^E \partial_{\mathbf{w}_1} F'_{1s}(w_{0s}) \\
+\eta(z)(1-\lambda^E)u''(w(z))w\hat{w}_0(z) + \eta(z)\lambda^E F(w)u''(w(z))w(z)\hat{w}_0(z) + \eta(z)\lambda^E F'_1(w(z))u'(w(z))\hat{w}_0(z) \\
+\eta(z)\lambda^E u'(w(z))\partial_{\mathbf{w}_1} F_{1s}(w_{0s}(z)) = 0
\end{aligned}
$$

Using

$$
\begin{aligned}
\partial_{\mathbf{w}_1} F_{1s}(w_{0s}(z)) &= -F'_1(w(z))w(z)\hat{w}_{1s}(z) \\
\partial_{\mathbf{w}_1} F'_{1s}(w_{0s}(z)) &= -F'_1(w(z))d\ln w_{1s}(z) - F''_1(w(z))w(z)d\ln w_1(z) - F'_1(w(z))\frac{w(z)}{w'(z)}\hat{w}'_1(z)
\end{aligned}
$$

and the Lagrangin multipliers at the steady-state, $\eta(z) = \frac{(1-\lambda^E+\lambda^E F(w(z)))+(Az-w(z))\lambda^E F'(w(z))}{u'(w(z))[(1-\lambda^E)+\lambda^E F(w(z))]}$ and rearranging, one obtains

$$
\begin{aligned}
&\left[2(1-\alpha(z)) - \alpha(z)\eta_{F_1}(z) - \{(1-\alpha(z)) + \alpha(z)\eta_\lambda(z)\} + \frac{\gamma}{\eta_\lambda(z)}\{(1-\alpha(z)) + \alpha(z)\eta_\lambda(z)\}\right]\hat{w}_0(z) \\
=&\hat{A} + \left[2(1-\alpha(z)) - \alpha(z)\eta_{F_1}(z) - 1 - \{(1-\alpha(z)) + \alpha(z)\eta_\lambda(z)\}\right]\hat{w}_1(z) - \alpha(z)\frac{w(z)}{w'(z)}\frac{d}{dz}\hat{w}'_1(z).
\end{aligned}
$$

Define

$$
\begin{aligned}
\omega_1(z) &\equiv 2(1-\alpha(z)) - \alpha(z)\eta_{F_1}(z) - \{(1-\alpha(z)) + \alpha(z)\eta_\lambda(z)\} \\
\omega_2(z) &\equiv \frac{1}{\eta_\lambda(z)}\{(1-\alpha(z)) + \alpha(z)\eta_\lambda(z)\},
\end{aligned}
$$

where $\eta_{F_0}(z) = \frac{d\ln F'_0(w(z))}{d\ln w}$ and $\eta_{F_1}(z) = \frac{d\ln F'_1(w(z))}{d\ln w}$ are the elasticity of density of wage distributions, $\eta_\lambda(z) \equiv \frac{d\ln\left(1-\lambda^E+\lambda^E F_1(w)\right)}{d\ln w}$ is the elasticity of worker's staying probability, and γ is the relative risk aversion of workers. Then we obtain (1.13).

Now, I turn to the boundary conditions. Lemma 4 shows the boundary condition at the top is $\hat{w}_0(\bar{z}) = \hat{w}_1(\bar{z})$. Regarding the bottom, it must be the case that either $\hat{w}_1(\underline{z}) = \hat{A}$ or $\hat{w}_0(\underline{z}) = \hat{A}$ as an interior solution. Because of the constraint that wages must be higher than the outside option of being unemployed, $\hat{w}_1(\underline{z}) \geq \hat{A}$ and $\hat{w}_0(\underline{z}) \geq \hat{A}$. Suppose that $\hat{w}_1(\underline{z}) > \hat{A}$ and $\hat{w}_0(\underline{z}) > \hat{A}$. Then one of the firms offering lower wages can lower wages without affecting the labor supply, contradicting to the optimality. If $\hat{w}_1(\underline{z}) = \hat{A}$ at an interior, then it must be the constraint $\hat{w}_0(\underline{z}) \geq \hat{A}$ must be (weakly) binding (not at an interior solution) because if $\hat{w}_0(\underline{z}) > \hat{A}$, then it would contradict the presumption that $\hat{w}_1(\underline{z}) = \hat{A}$ was an interior solution. in this case, the boundary of the incumbent firms is $w_0(\underline{z} + dz)$ and by continuity of wage strategy, it must be $w_0(z + dz) \leq \hat{A}$. Similarly, $\hat{w}_0(\underline{z}) = \hat{A}$ at an interior solution, then $\hat{w}_1(\underline{z} + dz) \leq \hat{A}$. Finally, I claim that $\hat{w}_1(\underline{z}) = \hat{A}$ is the appropriate boundary condition, and $\hat{w}_0(\underline{z} + dz)$ is a free jump variable for the bottom of incumbent firms. Suppose to the contrary that $\hat{w}_0(\underline{z}) = \hat{A}$ is the boundary condition, then the system of ODEs imply that for any $\gamma > 0$, $\hat{w}_1(z) < \hat{A}$ and $\hat{w}_0(z) > \hat{A}$ for all z. To prove this, starting from $\hat{w}_0(\underline{z}) = \hat{A}$ and $\hat{w}_1(\underline{z} + dz) \in [0, \hat{A}]$, $\hat{w}_0'(z) > 0$ at $\hat{w}(z) = \hat{A}$ for any $\hat{w}_1(z) \in [0, \hat{A}]$ and $\hat{w}_1'(z) < 0$ at $\hat{w}_1(z) = \hat{A}$ for all $\hat{w}_0 \geq \hat{A}$. Therefore the path needs to feature $\hat{w}_1(z) < \hat{A}$ and $\hat{w}_0(z) > \hat{A}$ for all z, but then it would never be able to satisfy the boundary condition at the top, $\hat{w}_1(\bar{z}) = \hat{w}_0(\bar{z})$, a contradiction. Therefore $\hat{w}_1(\underline{z}) = \hat{A}$ is an appropriate boundary condition.

1.8.4 Proof of Proposition 1

Since all the coefficients on linear ODEs are continuous in z, there must exist a unique solution. Part (i) follows from the fact $\theta_{1a}(z) + \theta_{1w}(z) = 1$ and $\theta_{0a}(z) + \theta_{0w}(z) = 1$. Then one can easily verify $\hat{w}_1(z) = \hat{w}_0(z) = \hat{A}$ is a unique solution that satisfy boundary conditions.

In order to prove part (ii), consider whether there exists ζ such that $\hat{w}_0(\zeta) > \hat{A}$. There can potentially be two such cases. First case is $(\hat{w}_0(\zeta), \hat{w}_1(\zeta))$ with $\hat{w}_0(\zeta) > \hat{A}$ and $\hat{w}_1(\zeta) \leq \hat{A}$. Then starting from such a point, it is not possible to satisfy the boundary condition at the top. This is because $\hat{w}_0'(z) > 0$ at $\hat{w}(z) = \hat{A}$ for any $\hat{w}_1(z) \in [0, \hat{A}]$ and $\hat{w}_1'(z) < 0$ at $\hat{w}_1(z) = \hat{A}$ for all $\hat{w}_0 \geq \hat{A}$, and therefore the path features $\hat{w}_1(z) < \hat{A}$ and $\hat{w}_0(z) > \hat{A}$ for all $z > \zeta$.

Second case is $(\hat{w}_0(\zeta), \hat{w}_1(\zeta))$ with $\hat{w}_0(\zeta) > \hat{A}$ and $\hat{w}_1(\zeta) > \hat{A}$, but such a point is never reached. Starting from $w_1(\underline{z}) = \hat{A}$ and $\hat{w}_0(\underline{z} + dz) \in [0, \hat{A}]$, in order to reach such a point, it needs to go through either (i) $\hat{w}_0(z) = \hat{A}$ and $\hat{w}_1(z) \in [0, \hat{A}]$ or (ii) $\hat{w}_0(z) = \hat{A}$ and $\hat{w}_1(z) > \hat{A}$. Case (i) is already excluded from the previous paragraph. Case (ii) is also not possible because $\hat{w}_0'(z) < 0$ and $\hat{w}_1(z) < 0$ at $\hat{w}_0(z) = \hat{A}$ and $\hat{w}_1(z) > \hat{A}$. These arguments complete the proof that $\hat{w}_0(z) < \hat{A}$.

In order to prove $\hat{w}_0(z) < \hat{w}_1(z)$, suppose to the contrary that there exists ζ such that $\hat{w}_0(\zeta) > \hat{w}_1(\zeta)$. It is always true that in such a region, $\hat{w}_0'(\zeta) > 0$. Then there can be potentially two cases: (i) $\hat{w}_1'(\zeta) < 0$ or (ii) $\hat{w}_1'(\zeta) > 0$. In the former case, it would never be able to satisfy the boundary condition at the top. The latter case is never reached.

Lastly, since the path needs to end up with $\hat{w}_0(\bar{z}) < \hat{A}$ and $\hat{w}_1(\bar{z}) = \hat{w}_0(\bar{z})$, and the path is continuous, it is immediate to see that there must exist $\breve{z}$ such that $\hat{w}_1(z) < \hat{A}$ for $z > \breve{z}$.

1.8.5 Proof of Proposition 1′

Note that

$$\theta_{1a}(z) = 2(1 - \alpha(z)) - \alpha(z)\eta_{F_0}(w(z)).$$

Totally differentiating (1.7) with respect to z gives

$$2(1 - \alpha(z)) - \alpha(z)\eta_{F_0}(w(z)) = \frac{w(z)}{w'(z)z}.$$

As $z \to \infty$, $w(z) \to \chi Ab + (1-\chi)\int_b^\infty A\tilde{z}d\hat{F}(\tilde{z})$. From (1.8), we have

$$\begin{aligned} w'(z)z &= \frac{1}{(\chi + (1-\chi)\hat{F}_0(z))}(1-\chi)\hat{F}_0'(z)(Az - w(z))z \\ &\leq \frac{1}{(\chi + (1-\chi)\hat{F}_0(z))}(1-\chi)\hat{F}_0'(z)Az^2 \end{aligned}$$

Taking the limit, $z \to \infty$,

$$\begin{aligned} \lim_{z\to\infty} w'(z)z &\leq \lim_{z\to\infty} \frac{1}{(\chi + (1-\chi)\hat{F}_0(z))}(1-\chi)\hat{F}_0'(z)Az^2 \\ &= 0 \end{aligned}$$

where the last inequality follows from the assumption of finite variance. Therefore

$$\begin{aligned} \lim_{z\to\infty} \theta_{1a}(z) &= \frac{w'(z)z}{w(z)} \\ &= 0, \end{aligned}$$

which completes the proof.

1.8.6 Proof of Proposition 2

The optimality condition for vacancy creation is

$$(A_s z - w_{1s}(z))\lambda_s^F \left(\chi + (1-\chi)F_{0s}(w_{1s}(z))\right) = A_s \bar{c}(z)(v_s(z))^{1/\iota}.$$

Taking log derivative,

$$\begin{aligned}\hat{\lambda}^F + \frac{1}{\alpha(z)}\hat{A} + \underbrace{\left(\frac{(1-\chi)F_0'(w(z))w(z)}{(\chi + (1-\chi)F_0(w(z)))} - \frac{w(z)}{Az - w(z)}\right)}_{=0 \text{ (from FOC of wages)}} \hat{w}_1^{exo}(z) \\ -\frac{(1-\chi)F'(w(z))w(z)}{(\chi + (1-\chi)F(w(z)))}\hat{w}_0^{exo}(z)) = \hat{A} + \frac{1}{\iota}\hat{v}(z) \\ \Leftrightarrow \quad \hat{\lambda}^F + \frac{1-\alpha(z)}{\alpha(z)}\left(\hat{A} - \hat{w}_0^{exo}(z))\right) = \frac{1}{\iota}\hat{v}(z),\end{aligned}$$

which is the firm-level vacancy response. To derive the aggregate response, note $\hat{\lambda}^F = (\kappa - 1)\hat{V}$. After multiplying both sides by $v(z)/V$ and adding up for all z, we obtain the aggregate response.

1.8.7 Vacancy response in DMP models

Without on-the-job search and with wage bargaining, the optimality condition for vacancy creation is

$$(A_s z - w_{1s}(z))\lambda_s^F = A_s \bar{c}(z)(v_s(z))^{1/\iota}.$$

Taking log-derivative,

$$\frac{1-\alpha(z)}{\alpha(z)}(\hat{A} - \hat{w}_1(z)) + \hat{\lambda}_s^F = \frac{1}{\iota}\hat{v}(z).$$

To derive the aggregate response, note $\hat{\lambda}^F = (\kappa - 1)\hat{V}$. After multiplying both sides by $v(z)/V$ and adding up for all z, we obtain the aggregate response.

1.8.8 Derivations of equilibrium conditions with endogenous vacancy

The incumbent firm's FOC is

$$-(1 - \lambda_s^E + \lambda_s^E F_{1s}(w_{0s})) + (A_s z - w_s)\lambda_s^E F_{1s}'(w_{0s}) + \eta\left[(1-\lambda_s^E)u'(w_{0s}) + \lambda_s^E F_{1s}(w_{0s})u'(w_{0s})\right] = 0.$$

Linearizing using symmetry,

$$\left[2(1-\alpha(z))-\alpha(z)\eta_{F_1}(z)-\{(1-\alpha(z))+\alpha(z)\eta_\lambda(z)\}+\frac{\gamma}{\eta_\lambda(z)}\{(1-\alpha(z))+\alpha(z)\eta_\lambda(z)\}\right]\hat{w}_0(z)$$
$$=\hat{A}+[2(1-\alpha(z))-\alpha(z)\eta_{F_1}(z)-1-\{(1-\alpha(z))+\alpha(z)\eta_\lambda(z)\}]\hat{w}_1(z)-\alpha(z)\frac{w(z)}{w'(z)}\frac{d}{dz}\hat{w}_1'(z)$$
$$\frac{1}{AzF'(w)}\left\{(1-F_1(w(z)))+(Az-w(z))F_1'(w(z))-\left(1+(Az-w(z))\frac{1}{w(z)}\eta_\lambda(z)\right)(1-F_1(w(z)))\right.$$
$$\left\{\alpha(z)\frac{1}{w(z)}\eta_\lambda(z)\right\}d\left(\frac{\bar{V}(z)}{V}\right)$$
$$\alpha(z)\frac{1}{v(z)/V}d\left(\frac{v(z)}{V}\right),$$

which I can rewrite further to obtain

$$\left[2(1-\alpha(z))-\alpha(z)\eta_{F_1}(z)-\{(1-\alpha(z))+\alpha(z)\eta_\lambda(z)\}+\frac{\gamma}{\eta_\lambda(z)}\{(1-\alpha(z))+\alpha(z)\eta_\lambda(z)\}\right]\hat{w}_0(z)$$
$$=\hat{A}+[2(1-\alpha(z))-\alpha(z)\eta_{F_1}(z)-1-\{(1-\alpha(z))+\alpha(z)\eta_\lambda(z)\}]\hat{w}_1(z)-\alpha(z)\frac{w(z)}{w'(z)}\frac{d}{dz}\hat{w}_1'(z)$$
$$+\alpha(z)\left\{1-\frac{\lambda^E(1-F(w))}{1-\lambda^E+\lambda^E F(w)}\right\}\hat{\lambda}^E$$
$$\left\{\alpha(z)\frac{\lambda^E F(w)}{1-\lambda^E+\lambda^E F(w)}\right\}d\left(\frac{\bar{V}(z)}{V}\right)$$
$$\alpha(z)(\hat{v}(z)-\hat{V}),$$

which is the one in the lemma. To complete the proof that boundary conditions are unchanged, note that if

$$\theta_{0a}(z)\hat{A}+\theta_{0w}(z)\hat{A}$$
$$+\underbrace{\theta_{0a}(z)\alpha(z)\left\{1-\theta_{\lambda,p}(z)\right\}\hat{\lambda}^E+\theta_{0a}(z)\alpha(z)\theta_{\lambda,r}(z)\left(\hat{V}(z)-\hat{V}\right)+\theta_{0a}(z)\alpha(z)\left(\hat{v}(z)-\hat{V}\right)}_{\equiv B}<\hat{A},$$

then the same argument as in Lemma 14 applies because it only relied on the fact that $\hat{w}_1'(z)=0$ locus in the phase diagram shifts downward. I can always guarantee this if ι is small enough, as $\lim_{\iota\to 0}B=0$.

1.8.9 Proof of Proposition 3

We can verify that $\hat{w}_0(z)=\hat{A}$, $\hat{w}_1(z)=\hat{A}$, $\hat{v}(z)=0$, and $\hat{V}=0$ are the solutions the ODEs with $\gamma=0$. The proposition follows because there is a unique solution. Next, as $\gamma\to\infty$,

$\hat{w}_0(z) \to 0$ almost everywhere. Then combined with Proposition 2, we have the claim.

1.8.10 Proof of Proposition 4

As the new hire firms only hire from unemployed, the wage offer to unemployed is $w_{1s}(z) = A_s b$. Then the vacancy creation condition is

$$\lambda_s^F (A_s z - A_s b) = A_s \bar{c}(v)^{1/\iota},$$

in which A_s cancel out. Therefore vacancy is a constant, which implies the unemployment rate is a constant in response to the shock to aggregate productivity.

1.8.11 Proof of Proposition ??

The first order condition with binding fairness constraint is

$$\begin{aligned}
&\lambda_s^F v(z) \left[(1-\chi)F_{0s}'(w_s(z))(A_s z - w_s(z)) - (\chi + (1-\chi)F_{0s}(w_s(z)))\right] \\
&\quad -\ell_0(z)\Bigg[(1-\lambda_s^E + \lambda_s^E F_{1s}(w_{0s}(z))) + (A_s z - w_s(z))\lambda_s^E F_{1s}'(w_s(z)) \\
&\qquad +\eta(z)\left[(1-\lambda_s^E)u'(w_s(z)) + \lambda_s^E F_{1s}(w_s(z))u'(w_s(z))\right]\Bigg] = 0
\end{aligned}$$

Taking the first order approximation, we have

$$\begin{aligned}
&\lambda^F v(1-\chi)F_0'(w)\left[\hat{A} - \hat{w}(z) - \alpha(z)\frac{w(z)}{w'(z)}\hat{w}'(z))\right] \\
&+\lambda^E F'(w)\ell(z)\Bigg[\hat{A} - \hat{w}(z) - \alpha(z)\frac{w(z)}{w'(z)}\hat{w}'(z) - \gamma\theta_\lambda(z)\hat{w}(z) \\
&\alpha(z)\left\{1-\theta_{\lambda,p}(z)\right\}\hat{\lambda}^E + \alpha(z)\theta_{\lambda,r}(z)\left(\hat{V}(z) - \hat{V}\right) + \alpha(z)\left(\hat{v}(z) - \hat{V}\right)\Bigg] = 0.
\end{aligned}$$

Rearranging, we obtain the expression in the proposition.

1.8.12 Proof of Proposition 8

In order to leave workers indifferent, the set of perturbation must satisfy

$$
\begin{aligned}
&\Big((1-\lambda^E)+\lambda^E F_1(w_0(\tilde{z}))\Big)\, u'(w_0(\tilde{z}))dw_0(\tilde{z}) \\
&+\mathbb{I}(z<\tilde{z})u(w_1(\tilde{z}))g(z)(\lambda^E/V)dv(z)+\mathbb{I}(z>\tilde{z})u(w_1(z))g(z)(\lambda^E/V)dv(z) \\
&-u(w_0(\tilde{z}))\lambda^E\int_z v(\tilde{z})dG(\tilde{z})\frac{\partial(\lambda^E/V)}{\partial v(z)}dv(z)+\int_{\tilde{z}} u(w_1(\zeta))v(\zeta)g(\zeta)d\zeta\frac{\partial(\lambda^E/V)}{\partial v(z)}dv(z)=0 \\
&(1-\lambda^U)u'(Ab)Ad\flat+u(w_1(z))g(z)(\lambda^E/V)dv(z)+\int_{\tilde{z}} u(w_1(\zeta))(v(\zeta)/V)g(\zeta)d\zeta\frac{\partial(\lambda^U/V)}{\partial v(z)}dv(z)=0
\end{aligned}
$$

The effect on net total surplus is

$$
\begin{aligned}
&-\int (A\tilde{z}-w(\tilde{z}))\left[\lambda^E\int_{\tilde{z}} v(\zeta)dG(\zeta)\right]dG(\tilde{z}) \\
&-\int (A\tilde{z}-w(\tilde{z}))\left[\lambda^E\mathbb{I}(z>\tilde{z})\right]dG(\tilde{z})
\end{aligned}
$$

$$
\begin{aligned}
d\mathcal{F} &= -c'(v(z))g(z) \\
&\quad + \int \left\{ -(A\tilde{z} - w(\tilde{z})) \int_{\tilde{z}} v(\zeta) dG(\zeta) \frac{\partial \lambda^E / V}{\partial v(z)} + (A\tilde{z} - w(\tilde{z}))\lambda^E \mathbb{I}(z < \tilde{z}) \right\} \ell_0(\tilde{z}) dG(\tilde{z}) \\
&\quad + \lambda^F (\chi + (1-\chi) F_0(w_1(z)))(Az - w_1(z)) \\
&\quad + \int \left\{ v(\tilde{z})(\chi + (1-\chi)F_0(w_1(\tilde{z})))(Az - w_1(\tilde{z})) \frac{\partial \lambda^F}{\partial v(z)} \right\} dG(\tilde{z}) \\
&\quad - \mu(Ab - A\flat) \int_{\underline{z}} v(\zeta) dG(\zeta) \frac{\partial \lambda^U / V}{\partial v(z)} \\
&\quad - \lambda^E \int^z (A\tilde{z} - w(\tilde{z})) \ell_0(\tilde{z}) dG(\tilde{z}) dv(z) - \mu \lambda^U (Ab - A\flat) dv(z) \\
&\quad - \int \left\{ (1 - \lambda^E + \lambda^E F_1(w_0(\tilde{z}))) dw_0(\tilde{z}) \right\} \ell_0(\tilde{z}) dG(\tilde{z}) - \mu(1-\lambda^U) d\flat \\
&= \underbrace{- \int \left\{ (A\tilde{z} - w(\tilde{z})) \int_{\tilde{z}} v(\zeta) dG(\zeta) \frac{\partial \lambda^E / V}{\partial v(z)} + \mu A b \int_{\underline{z}} v(\zeta) dG(\zeta) \frac{\partial \lambda^U / V}{\partial v(z)} \right\} \ell_0(\tilde{z}) dG(\tilde{z})}_{(1)} \\
&\quad \underbrace{+ \int \left\{ v(\tilde{z})(\chi + (1-\chi)F_0(w_1(\tilde{z})))(Az - w_1(\tilde{z})) \frac{\partial \lambda^F}{\partial v(z)} \right\} dG(\tilde{z})}_{(2)} \\
&\quad \underbrace{-\lambda^E \int^z (A\tilde{z} - w(\tilde{z})) \ell_0(\tilde{z}) dG(\tilde{z}) dv(z) - \mu \lambda^U (Ab - A\flat) dv(z)}_{(3)} \\
&\quad \underbrace{- \int \left\{ (1 - \lambda^E + \lambda^E F_1(w_0(\tilde{z}))) dw_0(\tilde{z}) \right\} \ell_0(\tilde{z}) dG(\tilde{z}) - \mu(1-\lambda^U) d\flat}_{(4)},
\end{aligned}
$$

where I used FOC of vacancy creation in the last equality. The first line (1) is

$$
\begin{aligned}
&- \int \left\{ (A\tilde{z} - w(\tilde{z})) \int_{\tilde{z}} v(\zeta) dG(\zeta) \frac{\partial \lambda^E / V}{\partial v(z)} + \mu A b \int_{\underline{z}} v(\zeta) dG(\zeta) \frac{\partial \lambda^U / V}{\partial v(z)} \right\} \ell_0(\tilde{z}) dG(\tilde{z}) \\
&= - \left(\int \left[w(\tilde{z}) P_0(\tilde{z}) - \int^{\tilde{z}} w(\zeta) dP_0(\zeta) \right] v(\tilde{z}) dG(\tilde{z}) \right) \frac{\partial (\mathcal{M}/V)}{\partial v(z)}
\end{aligned}
$$

where $P_0(z) \equiv \chi \mathbb{I}(z > \underline{z}) + (1-\chi)\frac{1}{1-\mu} \int^z \ell_0(\tilde{z}) dG(\tilde{z})$ is the cumulative employment distribution, and I used the expression of steady-state $w(z)$. The second line (2) is

$$
\begin{aligned}
&\int \left\{ v(\tilde{z})(\chi + (1-\chi)F_0(w_1(\tilde{z})))(Az - w_1(\tilde{z})) \frac{\partial \lambda^F}{\partial v(z)} \right\} dG(\tilde{z}) \\
&= \int \left\{ v(\zeta)(A\zeta - w(\zeta)) P(\zeta) \right\} dG(\zeta) \frac{\partial \mathcal{M}/V}{\partial v(z)}
\end{aligned}
$$

The third line is

$$
\begin{aligned}
&- \lambda^E \int^z (A\tilde{z} - w(\tilde{z}))\ell_0(\tilde{z}) dG(\tilde{z}) dv(z) - \mu\lambda^U Abdv(z) \\
=& - \lambda^F \int^z \left[w(z) - w(\tilde{z})\right] dP_0(\tilde{z}) dv(z)
\end{aligned}
$$

The forth line is

$$
\begin{aligned}
&- \int \left\{ (1 - \lambda^E + \lambda^E F_1(w_0(\tilde{z}))) dw_0(\tilde{z}) \right\} \ell_0(\tilde{z}) dG(\tilde{z}) - \mu(1 - \lambda^U) d\flat \\
=& \left(- \int \int^{\tilde{z}} \frac{u(w(\zeta))}{u'(w(\tilde{z}))} dP(\zeta) v(\tilde{z}) dG(\tilde{z}) + \int \left\{ v(\zeta) \frac{u'(w(\zeta))}{u'(w(\tilde{z}))} P(\zeta) \right\} dG(\zeta) \right) \frac{\partial \mathcal{M}/V}{\partial v(z)} dv(z) \\
&+ \lambda^F \left(\int^z \frac{1}{u'(w(\tilde{z}))} u(w(z)) - \frac{1}{u'(w(\tilde{z}))} u(w(\tilde{z})) dP(\tilde{z}) \right) dv(z)
\end{aligned}
$$

Combining, we obtain the desired expression.

$$
\begin{aligned}
d\mathcal{F} =& \int \left((A\tilde{z} - w(\tilde{z})) P(\tilde{z}) - \int^{\tilde{z}} \left[(w(\tilde{z}) - w(\zeta)) - \frac{1}{u'(\tilde{z})} \left(u(w(\tilde{z})) - u(w(\zeta)) \right) \right] dP(\zeta) \right) v(\tilde{z}) dG(\tilde{z}) \frac{\partial(\mathcal{M}}{\partial v} \\
&+ \lambda^F \int^z \left(\frac{1}{u'(w(\tilde{z}))} \left[u(w(z)) - u(w(\tilde{z})) \right] - \left[w(z) - w(\tilde{z}) \right] \right) dP(\tilde{z})
\end{aligned}
$$

1.8.13 Proof of Proposition 9

Provided in the main text.

1.8.14 Neutrality Result in Infinite Horizon setup

Assume workers are risk-neutral. Let $w(z) = A\bar{w}(z)$. Then we can guess and verify that in the steady-state, all the value functions are homogenous in A:

$$
W(w) = AW(\bar{w}), \quad U = A\bar{U}, \quad J(w, z) = AJ(\bar{w}, z).
$$

Now consider aggregate risk with $\{A_h, A_l\}$ Since the cost of vacancy scales with A, if $\{w(z), v(z)\}$ is a steady-state equilibrium with A, $\{A_s/Aw(z), v(z)\}$ for $s = h, l$ is also an equilibrium with no transition dynamics. Finally, (1.28) and the promise keeping constraint are satisfied as value functions are homogenous in A. Therefore we have that wages scale with aggregate productivity and no changes in employment and vacancy distribution. Without on-the-job search, wages are concentrated at $w(z) = Ab$ for all z. Therefore again, the value functions scale with A.

1.9 Inefficiently flexible incumbent wages

I assume $\iota = 0$ and consider a small perturbation of incumbent wages of a particular firm z, $dw_{0h}(z)$ and $dw_{0l}(z)$. In order to leave workers indifferent, such perturbation must satisfyFor incumbents,

$$\sum_s \pi_s \left(1 - \lambda^E + \lambda^E F_{1s}(w_{0s}(z))\right) u'(w_{0s}(z)) dw_{0s}(z) = 0 \quad .$$

Then changes in net total surplus can be computed as

$$\begin{aligned} d\mathcal{F} = & - v(z)\lambda^F (1-\chi) F'_{0h}(w_{0h}(z)))(A_h w_{1h}^{-1}(w_{0h}(z)) - w_{0h}(z)) dw_{0h}(z) \\ & - v(z)\lambda^F (1-\chi) F'_{0l}(w_{0l}(z)))(A_l w_{1l}^{-1}(w_{0l}(z)) - w_{0l}(z)) dw_{0l}(z) \\ = & \underbrace{-v(\tilde{z})\lambda^F_h \left(\chi + (1-\chi) F_{0h}(w_{0h}(z))\right) \left(1 - \frac{u'(w_{0h}(z))\left(1 - \lambda^E_h + \lambda^E_h F_{1h}(w_{0h}(z))\right)}{u'(w_{0l}(z))\left(1 - \lambda^E_l + \lambda^E_l F_{1l}(w_{0l}(z))\right)}\right)}_{<0} d\, w_{0h}(z), \end{aligned}$$

where I used new hire firm's FOC in the last equality. Therefore $dw_{0h}(z) < 0$ and $dw_{0l}(z) > 0$ improves welfare. We thus conclude

Proposition 10. *Assume $\iota \to 0$. Consider the equilibrium with aggregate risk. There exists a small perturbation of new hire wages $dw_{0h}(z), dw_{0l}(z)$ with $dw_{0h}(\tilde{z}) < 0$ and $dw_{0l}(\tilde{z}) > 0$ that yield Pareto improvement.*

1.10 Details of solution method in infinite horizon model

I first log-linearize all the optimality condition. The first order approximation of the firm's value function that hired a worker at time τ is given by

$$\begin{aligned} \rho_f dJ_t(w,z) = & \, Azd\ln A_t - wd\ln w - \lambda^E(1 - F(w))J(w,z)d\ln \lambda^E_t \\ & + \lambda^E J(w,z) dF_t(w) - (\delta + \lambda^E_t(1 - F_t(w)))dJ_t(w,z) + \partial_t dJ_t(w,z), \end{aligned}$$

where

$$dF_t(w) = -F'(w)wd\ln w_t(z) + \frac{1}{V}d\bar{V}_t(z) - \frac{1}{V^2}dV_t,$$

and $d\bar{V}_t(z) \equiv \int^z dv(\tilde{z})dG(\tilde{z})$. The response of $\partial_w J_t(w,z)$ is

$$\begin{aligned}\rho_f d\partial_w J_t(w,z) = &-\lambda^E F'(w)\partial_w J(w,z)d\ln w_t(z)\\ &+\left\{\lambda_t^E(1-F(w)))\partial_w J(w,z)+\lambda^E F'(w)J(w,z)\right\}d\ln\lambda_t^E\\ &+\lambda^E J(w,z)dF_t(w)+\lambda^E J(w,z)dF_t'(w)-(\delta+\lambda_t^E(1-F_t(w)))d\partial_w J(w,z)\\ &+\lambda_t^E F_t'(w)dJ(w,z)+\partial_t d\partial_w J_t(w,z),\end{aligned}$$

where

$$\begin{aligned}dF_t'(w) = &-F'(w)d\ln w_t(z)-F''(w)wd\ln w_t(z)-F'(w)\frac{w(z)}{w'(z)}\frac{d}{dz}(d\ln w_t(z))\\ &+F'(w)\left\{\frac{1}{v(z)g(z)}\frac{d}{dz}(d\bar{V}_t(z))-\frac{1}{V}dV_t\right\}\end{aligned}$$

The response of employment distribution is

$$\begin{aligned}\partial_t dP_t(w) = &-\left(\delta+\lambda_t^E(1-F_t(w))\right)dP_t(w)-\lambda^E(1-F(w))P_t(w)d\ln\lambda_t^E\\ &+\left(\lambda^E P(w)+\frac{1}{1-\mu}\mu\lambda^U\right)dF_t(w)+\frac{1}{(1-\mu)^2}\lambda^U F(w)d\mu_t+\delta F(w)d\ln\lambda_t^U\end{aligned}$$

and the response of $P_t'(w)$ is

$$\begin{aligned}\partial_t dP_t'(w) = &-\left(\delta+\lambda_t^E(1-F_t(w))\right)dP_t'(w)+\left(F'(w)P_t(w)-\lambda^E(1-F(w))P_t'(w)\right)d\ln\lambda_t^E\\ &+\lambda^E F'(w)dP_t(w)+\lambda^E P'(w)dF_t(w)+\left(\lambda^E P(w)+\delta\right)dF_t'(w)\\ &+\frac{1}{(1-\mu)^2}\lambda^U F'(w)d\mu_t+\delta F'(w)d\ln\lambda_t^U.\end{aligned}$$

The unemployment rate follows

$$\partial_t d\mu_t = -(\lambda^U+\delta)d\mu_t-\lambda^U\mu d\ln\lambda_t^U,$$

and the matching function implies

$$d\ln\lambda_t^U = d\ln\lambda_t^E = -\kappa d\ln\tilde{\mu}_t+\kappa d\ln V_t.$$

Therefore the first order response of new hire wages solve

$$\begin{aligned} Q(w)d\partial_w J_t(w,z) + (1-\chi)\partial_w J(w,z)dP_t(w) + (1-\chi)J(w,z)dP_t'(w) \\ +(1-\chi)P'(w)dJ_t(w,z) + \big((1-P(w))\partial_w J(w,z) - P'(w)J(w,z)\big)\, d\chi_t = 0, \end{aligned} \tag{1.33}$$

where $d\chi_t \equiv \left\{\frac{\zeta}{(\zeta(1-\mu_t)+\mu_t)^2}\right\} d\mu_t$. As in the two-period model, the above expression only depends on a few number of variables, $\{d\ln w_t(z), dV_t, d\bar{V}_t(z), \frac{d}{dz}d\ln w_t(z), \frac{d}{dz}d\bar{V}_t(z), d\mu_t, d\ln\lambda_t\}$. Crucially, it does not depend on the wage distribution, and only depends on the wages of the neighboring competitors. The first order approximation of the optimality condition for vacancy creation is

$$\lambda^F \hat{P}_t(w)J(w,z)d\ln\lambda_t^F + \lambda^F J(w,z)dQ_t(w) + \lambda^F Q(w)dJ_t(w,z) = Ac''(v(z))\frac{1}{g(z)}\frac{d}{dz}(d\bar{V}_t(z)). \tag{1.34}$$

The incumbent's wage response is

$$d\partial_w J_0(w,z) - \frac{\partial_w J(w,z)}{W'(w)}dW_0'(w) = 0, \tag{1.35}$$

where $dW_t(w)$ is given by

$$\rho_w dW_t'(w) = u'(w_t)d\ln w - \delta dW_t'(w) - \lambda_t^E(1-F_t(w))dW_t'(w) + \lambda^E W'(w)dF_t(w) + \partial_t dW_t'(w).$$

The boundary conditions are

$$\begin{aligned} d\ln w_t(z) &= d\ln \underline{w}_t \\ d\ln w_0^{inc}(\bar{z}) &= d\ln w_0(\bar{z}). \end{aligned}$$

I solve the transition dynamics using the following algorithm. First, I guess the sequence of two aggregates: $\{d\ln\underline{w}_t,\ dV_t\}$. Given these two aggregates, one can immediately compute $\{d\mu_t, d\chi_t, d\ln\lambda_t^E, d\ln\lambda_t^U, d\ln\lambda_t^F\}$ using the matching function. Then I solve a system of linear ODEs, (1.33), (1.34) and (1.35) to obtain $\{d\ln w_t(z), d\bar{V}_t(z)\}$. I iterate over the guess of $\{d\ln\underline{w}_t,\ dV_t\}$ until I have $W_t(\underline{w}_t) = U_t$ and $d\bar{V}_t(\bar{z}) = dV$. In practice, I simply invert the matrix to find equilibrium $\{d\ln\underline{w}_t, dV_t\}$. This takes less than a second to compute the transition dynamics. Following Auclert et al. (2019), figure 1-9 shows the directed acyclical graph (DAG) representation of the first order responses of the economy.

With fariness constraint. Let ω_{1t} and ω_{2t} denote the Lagrangian multipliers on the constraint (1.38) and (1.39), respectively. The optimality conditions of firm's problem are

$$\partial_t \ell_t = -(\delta + \lambda_t^E(1 - F_t(W_t))\ell_t + v\lambda_t^F \left(\chi_t + (1-\chi_t)P_t^{eq}(W_t)\right)$$

$$\rho_w W_t = u(w_t) + \delta\left\{U_t - W_t\right\} + \lambda_t^E \int \max\left\{0, \tilde{W} - W_t\right\} dF_t(\tilde{W}) + \partial_t W_t,$$

$$-\ell - \omega_2 u'(w) = 0$$

$$(Az - w) - \omega_1\left(\delta + \lambda_t^E(1 - F_t(W_t))\right) = \rho_f \omega_1 - \dot{\omega}_1$$

$$\omega_1\left(\lambda_t^E F'(W)\ell + v\lambda^F(1-\chi_t)P'(W)\right) + \omega_2\left\{\rho_w + \delta + \lambda^E(1 - F_t(W))\right\} = \rho_f\omega_2 - \dot{\omega}_2$$

$$\lambda_t^F\left(\chi_t + (1-\chi_t)P_t^{eq}(W_t)\right)\omega_{1t} = c'(v).$$

The initial condition W_0 is pinned down by the risk-sharing condition:

$$\omega_1\left(\lambda_t^E F'(W_0)\ell + v\lambda^F(1-\chi_t)P'(W_0)\right) + \omega_2 + \eta = 0,$$

where η is the Lagrangian multiplier constraint on the promise-keeping constraint.

As before, linearizing the equilibrium conditions to obtain the system of linear ordinary differential equations.

$$\begin{aligned}\left(\lambda_t^E F'(W)\ell + v\lambda^F(1-\chi_t)P'(W)\right)u'(w)d\omega_1 + \omega_1\left(\lambda_t^E F'(W)\ell + v\lambda^F(1-\chi_t)P'(W)\right)u''(w)wd\ln w \\ + u'(w)\omega_1\lambda_t^E\ell dF'(W) + u'(w)\omega_1\lambda_t^E\ell F'(W)d\ln\lambda_t^E + u'(w)\omega_1\lambda_t^E F'(W)d\ell \\ + \lambda^F(1-\chi_t)P'(W)u'(w)\omega_1 dv + v\lambda^F(1-\chi_t)P'(W)u'(w)\omega_1 d\ln\lambda^F \\ + v\lambda^F(1-\chi_t)u'(w)dP'(W) - v\lambda^F\omega_1 u'(w)P'(W)d\chi \\ - \left\{\rho_w + \delta + \lambda^E(1-F_t(W))\right\}d\ell - \left\{\lambda^E(1-F_t(W))\right\}\ell d\ln\lambda^E - \gamma\partial_t d\ln w_t = 0\end{aligned}$$

$$\begin{aligned}\partial_t d\ell_t = &-(\delta + \lambda_t^E(1-F_t(W_t))d\ell_t + v\lambda_t^F(1-P_t(W_t))d\chi_t + \lambda_t^E\ell_t d\frac{J(z)}{V} \\ &+ \left(\chi_t + (1-\chi_t)P_t(W_t)\right)\lambda_t^F dv - \lambda_t^E(1-F_t(W_t))\ell_t d\ln\lambda_t^E + \left(\chi_t + (1-\chi_t)P_t(W_t)\right)\lambda_t^F v d\ln\lambda_t^F \\ &+ v\lambda_t^F(1-\chi_t)dP_t(W_t)\end{aligned}$$

$$\partial_t dW = u'(w) d\ln w - (\rho + \lambda^E + \delta) dW_{t+\Delta} + \delta d\bar{U}_{t+\Delta} + \lambda^E \int \max\{d\tilde{W}_t, dW_t\} dF^{eq}(\tilde{W}_t)$$
$$+ \lambda^E \left[-W(w_s) + \int \max\{\tilde{W}, W\} dF^{eq}(\tilde{W}) \right] d\ln \lambda^E + \lambda^E \int \max\{W(\tilde{z}), W\} d\hat{F}'(\tilde{z}) d\tilde{z}$$

$$d\omega_{1,t} = Azd\ln A_t - wd\ln w - (\rho + \delta + \lambda^E(1 - F^{eq}(W)) d\omega_{1,t} - \omega_1 \lambda_t^E (1 - F_t(W_t)) d\ln \lambda_t^E + \omega_1 \lambda_t^E d\hat{F}(z)$$
$$\left(\lambda_t^E F'(W_0) \ell + v\lambda^F (1 - \chi_t) P'(W_0) \right) d\omega_1 + \omega_1 d \left(\lambda_t^E \ell d + v\lambda^F (1 - \chi_t) P'(W_0) \right) + d\omega_2 + d\eta = 0,$$

and the rest of the equilibrium conditions are unchanged from the one without fairness constraint.

1.11 Importance of New hire wage rigidity in other environments

1.11.1 Competitive search

Consider the following model with competitive search (Moen, 1997; Acemoglu and Shimer, 1999) without on-the-job search. The model is static that follows Wright et al. (2019). Each firm with productivity z posts wage w, and workers see all the wages and direct their search. Let $q(w)$ denote the unemployment-to-vacancy ratio in the sub-market with wage w. Then $\lambda^F(w) \equiv \mathcal{M}(q(w), 1)$ denote the meeting probability of a firm when the firm posts w, and $\lambda^U(w) \equiv \frac{1}{q(w)} \mathcal{M}(q(w), 1)$ is the meeting probability of unemployed. Unemployed workers earn Ab. Workers must be indifferent across sub-markets:

$$\begin{aligned} \bar{U} &= \lambda^U(w) w + (1 - \lambda^U(w)) Ab \\ &= \frac{1}{q(w)} \mathcal{M}(q(w), 1) w + \frac{1}{q(w)} (1 - \mathcal{M}(q(w), 1)) Ab, \end{aligned}$$

which defines $q(w)$ implicitly.

If they can, then firms set wages so as to

$$w^* = \arg\max_w \Pi(w; A, z) = \mathcal{M}(q(w), 1)(Az - w).$$

The optimal amount of vacancy creation for a given wage w is that

$$v^*(w; A, z) = \arg\max_v \Pi(w; A, z) v - c(v; A, z).$$

Then by envelope theorem, if the firm was setting the wage optimally, there is no first order effect of wages on profits:

$$\frac{\partial \Pi(w^*, A)}{\partial w} = 0,$$

which implies that vacancy is unaffected by wage as well

$$\frac{\partial v^*(w^*; A, z)}{\partial w} = 0.$$

In this environment, what matters for the vacancy creation of a particular firm is not whether that firm can adjust wages or not, but that whether other firms can adjust wages. How the wages of other firms determined? In the baseline competitive search environment, there is a perfectly elastic free-entry, $c'' = 0$. If those entrants can freely choose wages, then there cannot be any equilibrium new hire wage rigidity. Similar results hold in the context of price-setting, as studied by Bilbiie (2020). One can work with new hire wage rigidity with inelastic entry, but this also kills the tractability of competitive search.

1.12 Quantitative Infinite-Horizon Setup

1.12.1 Perfect-foresight Equilibrium and Steady-state Characterization without Fairness Constraints

The wage offer distribution is

$$F_t(w) = \frac{1}{V_t} \int_{z:w(z) \leq w} v_t(z) dG(z). \tag{1.36}$$

The meeting probabilities are

$$\lambda_t^U = \frac{1}{\tilde{\mu}_t} \mathcal{M}(\tilde{\mu}_t, V_t), \quad \lambda_t^E = \zeta \lambda_t^U, \quad \lambda_t^F = \frac{1}{V_t} \mathcal{M}(\tilde{\mu}_t, V_t), \quad \text{where } V_t \equiv \int v_t(z) dG(z). \tag{1.37}$$

Equilibrium definition is as follows:

Definition 2. *Equilibrium with constant aggregate productivity consists of a sequence of* $\{w_t(z), v_t(z)\}$, $\{P_t(w), F_t(w), \underline{w}_t, \mu_t\}$, $\{\lambda_t^U, \lambda_t^E, \lambda_t^F\}$ *such that (i) given* $\{P_t(w), F_t(w), \lambda_t^U, \lambda_t^E, \lambda_t^F, \underline{w}_t\}$, $\{w_t(z), v_t(z)\}$ *solve (1.25); (ii) the reservation wages satisfy* $W_t(\underline{w}_t) = U_t$, *where* U_t *and* W_t *are given by (1.26) and (1.27), respectively; (iii) the unemployment, the wage employment distribution,* $P_t(w)$, *and the wage offer distribution,* $F_t(w)$, *satisfy (1.23), (1.24), and (1.36), respectively; and (iv) meeting probabilities are given by (1.37).*

The steady-state unemployment rate is given by $\mu = \frac{\delta}{\delta + \lambda^U}$. The steady-state employment weighted wage distributions are

$$P(w) = \frac{\delta F(w)}{\delta + \lambda^E(1 - F(w))}, \quad Q(w) = \frac{\delta}{\delta + \lambda_t^E(1 - F_t(w))}.$$

The firm's Bellman equation in the steady-state is

$$J(w, z) = \frac{Az - w}{\rho_f + \delta + \boldsymbol{\varkappa} + \lambda^E(1 - F(w))}.$$

Using the above expressions, one can rewrite firms' FOCs as

$$\lambda^E F'(w) \frac{Az - w(z)}{\delta + \lambda^E(1 - F(w(z)))} + \lambda^E F'(w) \frac{Az - w(z)}{\rho + \delta + \boldsymbol{\varkappa} + \lambda^E(1 - F(w(z)))} = 1$$

and

$$\lambda^F Q(w(z)) J(w(z), z) = Ac'(v(z)).$$

Because firms' profits are log-supermodular in (w, z), wages are increasing in firm's productivity. Therefore, as in Burdett and Mortensen (1998), the steady-state equilibrium is rank-preserving in a sense that workers always move toward more productive firms. Defining $\hat{F}(z) \equiv F(w(z)) = \frac{1}{V} \int^z v(\tilde{z}) dG(\tilde{z})$ and $\hat{Q}(z) \equiv Q(w(z))$, we can write the steady-state equilibrium wage and vacancy distribution $\{w(z), \hat{F}(z)\}$ as the solution to the following system of ODEs:

$$\lambda^E \hat{F}'(z) \frac{Az - w(z)}{\delta + \lambda_t^E(1 - \hat{F}(z))} + \lambda^E \hat{F}'(z) \frac{Az - w(z)}{\rho + \delta + \boldsymbol{\varkappa} + \lambda_t^E(1 - \hat{F}(z))} = w'(z)$$

$$\lambda^F \hat{Q}(z) J(w(z), z) = Ac' \left(V\hat{F}'(z)/g(z) \right)$$

with the boundary conditions $w(\underline{z}) = \underline{w}$, where $\underline{w}$ satisfies $W(\underline{w}) = U$, and $\hat{F}(\underline{z}) = 0$. One still needs to solve fixed point in terms of aggregate vacancy, V, because meeting probabilities, λ^F and λ^U, and unemployment rate μ, using (1.37). Note that in the steady-state, workers' risk aversion, γ, plays no role.

Given the wage and vacancy distributions, workers value function in the steady-state are given by

$$\rho W(w) = u(w) + \delta \{U - W(w)\} + \lambda^E \int \max\{0, W(\tilde{w}) - W(w)\} dF(\tilde{w})$$
$$\rho U = u(Ab) + \lambda^U \int \max\{0, W(\tilde{w}) - U\} dF(\tilde{w}).$$

1.12.2 Environment with Fairness Constraints

I impose a restriction that firms cannot discriminate wages across employees. Firms commit to a sequence of wage payments $\{w_s\}$ that delivers W_t of the expected lifetime utility to the workers employed at the firm. Workers accept the job that offers a higher value. Let $F_t^{eq}(W) \equiv \frac{1}{V_t}\int_{z:W_t(w_t(z))\leq W} v_t(z)dG(z)$ denote the cumulative distribution function of the offer distribution. The employment distribution of worker value P_t^{eq} evolves in an analogous manner as in (1.24):

$$\partial_t P_t^{eq}(W) = -\delta P_t^{eq}(W) - \lambda_t^E(1-F_t^{eq}(W))P_t^{eq}(W) + \frac{1}{1-\mu_t}\mu_t\lambda_t^U F_t^{eq}(W).$$

The employment in a particular firm z evolves according to

$$\partial_t \ell_t = -(\delta + \lambda_t^E(1-F_t^{eq}(W_t))\ell_t + v\lambda_t^F\left(\chi_t + (1-\chi_t)P_t^{eq}(W_t)\right) \tag{1.38}$$

For a given W_0, a firm chooses its wage policy and vacancies to maximize profits

$$\Pi(W_0;z) \equiv \max_{\{W_t,w_t,v_t\}} \int e^{-\rho t}\left\{(Az_t - w_t)\ell_t - c(v_t)\right\}dt$$

subject to (1.38) and the worker's Bellman equation:

$$\rho_w W_t = u(w_t) + \delta\left\{U_t - W_t\right\} + \lambda_t^E\int \max\left\{0, \tilde{W} - W_t\right\}dF_t^{eq}(\tilde{W}) + \partial_t W_t. \tag{1.39}$$

The rest of the models are unchanged from before. Appendix 1.12.2 characterizes the steady-state of this economy.

Again, I consider the following dynamics. At $t = 0$, the economy is at the steady-state. Then there is a news that the aggregate productivity could be permanently high or low in the following periods. Firms insure workers by writing state contingent wage contracts that deliver the expected utility that is at least as large as promised in the steady-state:

$$\max_{\{W_t^s\}} \sum_{s\in\{h,l\}} \pi_s \Pi(W_0^s;z)$$
$$\text{s.t.} \quad \sum_s \pi_s W_t^s \geq W(z).$$

Given the initial W_0^s, the economy follows the perfect foresight equilibrium described above. Differently from before, wages are not fixed during the tenure period. Rather, firms offer the same time-varying wages to both incumbent workers and new hires.

As before, the steady-state is rank-preserving: more productive firms offer higher wage (values) to workers. The steady-state wages $w(z)$ and wage offer distribution $\hat{F}(z) \equiv \int^{z} v(\tilde{z})/V dG(\tilde{z})$ solve the following system of ODEs:

$$
\begin{aligned}
2\lambda^{E}\hat{F}'(z)\frac{Az - w(z)}{\delta + \lambda^{E}(1-\hat{F}'(z))} &= 1 \\
\lambda^{F}\hat{Q}(z)J(w(z), z) &= Ac'(V\hat{F}'(z)/g(z)).
\end{aligned}
$$

Compared with the model without equal treatments, we can immediately see that as $\rho \to 0$, the steady-state equilibrium coincide. I focus on a symmetric steady-state, in which firms with the same productivity employ the same number of workers. Let $W(z)$ denote the utility that a firm with productivity z promises to workers in the steady-state.

Chapter 2

Globalization and the Ladder of Development: Pushed to the Top or Held at the Bottom?

2.1 Introduction

A popular metaphor about development is that countries sit at different rungs of a ladder, each associated with a different set of economic activities. As countries develop, they become more capable, move up the ladder, and start to produce and export more complex goods. In this paper, we propose to take this metaphor seriously and use it as a starting point to study the relationship between international trade and development.

As simple as it is, the previous metaphor points towards two distinct mechanisms through which international trade and development may be related. On the one hand, countries that develop—because of technological innovations, the adoption of better domestic policies, or any other channel unrelated to trade—may acquire a comparative advantage in more complex goods and, in turn, tilt their exports towards these goods. On the other hand, countries that specialize in more complex goods—because of the removal of trade barriers or technological innovations in the rest of the world—may start growing faster, as a result of greater opportunities for knowledge accumulation and technological spillovers in those sectors.

The distinction between the two mechanisms has potentially important implications, both from a normative and a positive perspective. The first mechanism corresponds to the static channel between productivity and trade at the core of any Ricardian model. In such a model, changes in trade patterns are a by-product of technological progress, specialization according to comparative advantage is Pareto efficient, and laissez-faire policy is optimal. The second mechanism corresponds to the dynamic effects of trade more often emphasized by models with external economies of scale. It suggests, in contrast, that industrial policies subsidizing more complex sectors at the expense of others could be welfare improving. It also opens up the possibility that the emergence of large countries like China in the world economy may push some countries to the top of the ladder, while holding others at the bottom.

Our paper offers a formalization of the previous ideas and an exploration of their empirical validity. As a theoretical matter, we show that if two key features of the ladder metaphor are satisfied, namely that specialization in more complex goods generates positive spillovers and that fewer countries at the top of the ladder produce more complex goods, then it is possible for international trade to raise capability in all countries. As an empirical matter, however, we only find support for the first of these two qualitative features. In the data, more complex goods tend to be produced by more countries. Through the lens of our model, this implies dynamic welfare losses from trade that are small for the median country, but pervasive and large among a few developing countries.

Section 2.2 develops a Ricardian model of international trade with nested CES preferences and static and dynamic effects. There are many countries and many sectors. Within each sector, goods produced by different countries are imperfect substitutes. In line with the previous metaphor, we assume that countries can be ranked in terms of their capability, while goods can be ranked in terms of their complexity. In a given period, capability and complexity determine the distribution of productivity across countries and sectors. Over time, capability may increase in all countries, but technological progress is unequal and depends, in part, on what countries specialize in, which atomistic firms do not internalize. Specifically, we assume that more complex goods generate more opportunities for learning. So, when the distribution of employment is tilted towards those goods, capability growth increases.

Beside the fact that goods and countries may each be ranked along a single dimension, the ladder metaphor also points towards productivity differences manifesting themselves at the extensive margin. Capable countries sitting at the top of the ladder can produce the most complex goods, whereas countries at lower rungs cannot. To shed light on the implications of these extensive margin considerations, we first focus on a special case of our general Ricardian environment in which the only difference across goods is that some goods, the most complex ones, are produced by fewer countries, the most capable ones, as in Krugman (1979). We refer to this benchmark environment as a pure ladder economy.

Without international trade, all countries in that economy would produce all the goods that they know how to produce. With international trade, they can source some of those goods from the rest of the world. From the point of view of any individual country, among the goods that it knows how to produce, the rest of the world tends to have a comparative advantage in its less complex goods, since a greater number of foreign competitors knows how to produce those goods. More competition at the bottom of the ladder tends to push all countries to specialize in their most complex sectors and, in turn, to raise capability and real income around the world.

A recurrent theme of the earlier literature on the dynamic effects of trade, as reviewed for instance by Grossman and Helpman (1995), is that there are good sectors, with opportunities for learning, and bad sectors, without them. For countries with a static comparative advantage in the former sectors, free trade therefore slows down productivity growth, opening up the possibility of welfare losses from trade liberalization. Our simple ladder economy maintains a similar good-sector-bad-sector dichotomy, but by moving the focus to extensive margin considerations, it clarifies that dynamic gains from trade, like static ones, do not have to be zero sum. In the pure ladder economy, all countries that are not at the bottom of the ladder experience strictly positive dynamic gains (since they face strictly more competition for their least complex goods), whereas the poorest country sitting at the bottom experiences neither dynamic losses nor gains (since it faces the same competition from the rest of the world in all sectors in which it is able to produce).

To explore the empirical relevance of the pervasive dynamic gains predicted by our pure ladder economy, we propose to proceed in two ways. In our baseline analysis, we start with measures of complexity and capability that, in light of earlier empirical work by Hausman et al. (2007) and Hausman et al. (2013), are likely to generate positive spillovers, we will then estimate the magnitude of those spillovers (if any), and finally we will assess the extent to which opening up to trade indeed shifts most countries towards their more complex sectors. In our sensitivity analysis, we will then proceed in reverse by defining the more complex goods as those that fewer countries produce and then testing whether or not they generate positive spillovers.

Section 2.3 presents of baseline measures of complexity and capability. Keeping the focus of our analysis on extensive margin considerations and taking inspiration from the work of Hausman et al. (2013), we propose using disaggregated trade data from the United Nations Comtrade Database to measure the complexity of hundreds of manufacturing goods, defined as an SITC 4-digit product, and the capability of 146 countries from 1962 to 2014. We then propose to infer complexity and capability from the assumption that more capable countries are more likely to export more complex goods. Accordingly, if a country is known to be more capable than another, say the United States versus

Bangladesh, then one can identify more complex goods as those that are relatively more likely to be exported by the United States. Conversely, if a good is known to be more complex than another, say medicines versus underwear, then one can identify more capable countries as those that are relatively more likely to export medicines. Our revealed measures of complexity and capability should then be consistent with both types of observations.

Overall, measures of complexity and capability reveal reasonable patterns. Throughout this period, rich countries, like the United States and Western Europe, are revealed to be among the most capable in the world, whereas poor countries, like much of Africa, remain at the bottom. East Asian countries like Korea and China experience rapid increases in capability growth while much of Latin America sees relative declines. Across goods, Medicaments, Cars, and Medical Instruments are consistently revealed to be among the most complex, whereas Wood Panels, Hand Woven Rugs, and Men's Underwear are among the least complex.

Section 2.4 focuses on the estimation of dynamic spillovers. We specify the law of motion for capability as an auto-regressive process of order 1. In every period, shocks are drawn from a distribution whose mean linearly depends on the average complexity of a country's output mix. Dynamic spillovers are positive if the mean of a country's capability shocks is increasing with average complexity.

The key empirical challenge to estimate the previous spillovers is reverse causality, running from country capability to sectoral employment through changes in a country's comparative advantage over time. To deal with these issues, we need instrument variables correlated with a country's sectoral employment but uncorrelated with the unobserved determinants of its capability. We propose to use the entry of countries into the World Trade Organization (WTO) to construct time- and country-varying shifters of average complexity in other countries that rely on first-order approximations to the changes in sectoral employment caused by lower trade costs in our model. The reduced form of our IV also provides a comparative static of independent interest: does a country like China's entry into the WTO push more capable countries up the ladder and less capable countries down?

Our baseline IV estimates point to dynamic economies of scale in more complex sectors that are positive and statistically significant. Exogenous employment shifts towards more complex sectors tends to raise capability. This conclusion is robust to a range of alternative specifications and robustness checks, including alternative data samples and lag structures. The same exogenous shift in sectoral employment is also associated with significant increases in real GDP per capita.

Section 2.5 returns to our Ricardian model, in its most general form, allowing patterns of international specialization to be shaped both by intensive and extensive margin considerations. In order to quantify the static and dynamic effects of trade, we ask the following counterfactual question. Suppose that a country were to move to autarky in 1962, the first year of our sample, while still being subject to the same domestic technological shocks, what would happen to the path of its capability and real consumption? Combining our estimates of dynamic spillovers with a non-parametric specification of productivity differences across origin, destination, and sectors, we conclude that about XXX% of countries in our sample would experience *higher* capability under autarky. For the median country, these dynamic considerations lower the welfare gains from trade by XXX%, though a few developing countries experience much larger welfare losses. The reason behind these pervasive losses is that in sharp contrast to the benchmark predictions of our pure ladder economy, sectors that we have identified as more complex in Section 2.3 tend to face more rather than less foreign competition.

Section 2.6 explores the robustness of the previous conclusions to alternative measures of complexity and capability that are based instead on the assumption that more capable countries are those that tend to produce more goods, whereas more complex goods are those that tend to be produced by fewer countries. While these new measures of capability remain positively correlated with our earlier measures, with richer countries being revealed as more capable on average, the correlation between the two measures of complexity is negative. As a result, when using the same IV strategy, we conclude that there are negative dynamic spillovers in more complex sectors, now defined as those where more capable countries are more likely to export.

Our bottom line about the dynamic consequences of international trade, however, remains unchanged. The reason is that for dynamic gains to arise, two conditions need to be simultaneously satisfied. First, more complex sectors need to be associated with dynamic positive spillovers (so that their expansion creates capability growth); and second, they need to face less foreign competition (so that they expand under free trade). In our baseline analysis, the first condition holds, but not the second. In our sensitivity analysis, the second condition holds, by construction, but not the first. In both cases, we therefore end up concluding that there are small, but pervasive dynamic losses from trade.

Related Literature

On the theory side, the static part of our model, with its emphasis on the interaction between a single country characteristic, capability, and a single good characteristic, complexity, is reminiscent of Krugman's (1986) technology gap model, Ricardian models of trade and institutions, like Matsuyama (2005), Levchenko (2007), Costinot (2009), and Melitz and Cunat (2012), and the recent work on quality and capability by Sutton and Trefler (2016) and Schetter (2020).[1] The special case of a pure ladder economy, which we study analytically, is a strict generalization of Krugman (1979). Like Krugman (1979), our model emphasizes differences in comparative advantage across countries that take place at the extensive margin, a key feature of the ladder metaphor motivating our analysis. But unlike Krugman (1979), our model allows for more than two countries and imperfect substitutability between goods from different countries. The first generalization allows us to distinguish what happens at the top and the bottom of the ladder from what happens in most countries in the middle. The second generalization makes foreign costs a strictly decreasing function of the number of foreign countries that can produce a good, which gives all countries a comparative advantage in more complex goods relative to the rest of the world.

The dynamic part of our model, with its emphasis on external economies of scale, is related to earlier work by Krugman (1987), Boldrin and Scheinkman (1988), as well as Grossman and Helpman (1990), Young (1991) and Stokey (1991) who also allow inter-industry spillovers. As mentioned earlier, we share with these papers an emphasis on the dichotomy between good sectors, that are more conducive to learning and growth, and bad sectors, that are not. Motivated by the ladder metaphor, however, we turn our attention away from intensive margin considerations (in a two-country world) towards extensive margin considerations (in a many-country world). This seemingly small change of perspective has important welfare implications. In the pure ladder economy, dynamic gains from trade, like the static ones, do not have to be zero-sum, as all countries may simultaneously specialize in the good sectors.

The previous feature is related to recent work by Perla, Tonetti, and Waugh (2015), Sampson (2016), and Buera and Oberfield (2017). They focus on economies where firms of heterogeneous productivity can learn from each other. Since opening up to trade reallocates production towards larger, more productive firms, from which other firms have more learn, it also raises aggregate productivity. Hence, we share the same general feature that trade may lead to a reallocation of economic activities that is potentially growth-enhancing in all countries, though the empirical content and policy implications are very different. In the previous papers, large firms should be subsidized; in our paper, if there

[1] A similar focus on a ladder of countries can be found in Matsuyama (2004) and Matsuyama (2013) where productivity differences between countries arise endogenously through symmetry breaking under free trade.

are positive dynamic spillovers, the most complex sectors should be subsidized.

On the empirical side, we view our revealed measures of complexity and capability as a bridge between the original, descriptive work of Hidalgo and Hausman (2009) and Hausman et al. (2013) and recent, structural work on comparative advantage by Costinot, Donaldson, and Komunjer (2012), Levchenko and Zhang (2016), and Hanson, Lind, and Muendler (2016). In the spirit of Hausman et al. (2013), we focus on the extensive margin of trade, that is, whether or not a country exports a particular good, as a way to reveal capability and complexity. But like Costinot, Donaldson, and Komunjer (2012), Levchenko and Zhang (2016), and Hanson, Lind, and Muendler (2016), we use a difference-in-difference strategy that controls for exporter-importer and importer-industry fixed effects. This allows us to separate capability and complexity from bilateral trading frictions and demand differences across countries.

Our estimation of dynamic spillovers is related to the influential work of Hausman et al. (2007) and the general debate about whether what countries export matters, as discussed, for instance, in Lederman and Maloney (2012). Our instrumental variable strategy, based on the differential effects of new WTO members on countries with different industry mixes, aims to provide credible causal evidence that trade indeed matters for the pattern of development, rather than development mattering for the pattern of trade. Our evidence complements the recent work of Bartelme et al. (2019b) who study the heterogeneous impact of sectoral foreign demand shocks on real income as well as recent papers such as Bloom et al. (2016) and Autor et al. (2017) that focus on the differential impact of Chinese imports, caused by the removal of trade barriers or productivity growth in China, on direct measures of innovation, like patents, across sectors.

2.2 Theory

2.2.1 Environment

We consider an economy with many countries, indexed by i, and a continuum of goods, indexed by k. The total measure of goods is one. Time is continuous and indexed by $t \geq 0$. Labor is the only factor of production, with $L_{i,t}$ the labor supply in country i at date t.

Preferences. In each country, there is a representative agent who derives utility from an infinite stream of consumption,

$$U_i = \int_0^\infty e^{-\rho_i t} u_i(C_{i,t}) dt,$$

where $\rho_i > 0$ is the discount factor and u_i is strictly increasing, strictly concave, and twice differentiable. Aggregate consumption $C_{i,t}$ itself derives from consuming varieties from different countries in different sectors,

$$C_{i,t} = (\int (C_{i,t}^k)^{(\epsilon-1)/\epsilon} dk)^{\epsilon/(\epsilon-1)}, \tag{2.1}$$

$$C_{i,t}^k = (\sum_j (c_{ji,t}^k)^{(\sigma-1)/\sigma})^{\sigma/(\sigma-1)}, \tag{2.2}$$

where $\epsilon > 1$ is the elasticity of substitution between goods from different sectors, $\sigma > 1$ is the elasticity of substitution between varieties from different countries within a given sector, and $\sigma > \epsilon$ so that there is more substitutability within than between sectors. This implies that if a country i faces more foreign competition in a sector, that is lower foreign prices, then total expenditure on country i's variety in that sector decreases.

Technology. Goods differ in their complexity, $n_t^k \geq 0$, whereas countries differ in their capability, $N_{i,t} \geq 0$. We let $N_t = \{N_{i,t}\}$ denote the state of the world technology at date t and F_t denote the cumulative distribution of complexity across goods,

$$F_t(n) = \int_{0 \leq n_t^k \leq n} dk, \text{ for all } n \geq 0.$$

For all goods, production functions are linear,

$$q_{ij,t}^k = A_{ij,t}^k \ell_{ij,t}^k, \tag{2.3}$$

where $A_{ij,t}^k \geq 0$ denotes the productivity of firms producing good k for country j in country i at date t, inclusive of any transport cost, and $\ell_{ij,t}^k \geq 0$ denote their employment. Conditional on a good's complexity and the state of the world technology, we assume for now that the vector of productivity, $A_t^k = \{A_{ij,t}^k\}$, is drawn independently across all k from a general multivariate distribution,

$$\text{Prob}(A_t^k \leq a) = G_t(a | n_t^k = n, N_t).$$

Over time, changes in a country's capability are determined by its present capability and its endogenous pattern of specialization,

$$\dot{N}_{i,t} = H_{i,t}(N_{i,t}, F_{i,t}^{\ell}), \tag{2.4}$$

with $H_{i,t}$ a monotonic function of the cumulative distribution of employment across sectors of different complexity,

$$F^{\ell}_{i,t}(n) = \frac{\sum_j \int_{0\leq n^k \leq n} \ell^k_{ij,t} dk}{\sum_j \int \ell^k_{ij,t} dk} \text{ for all } n \geq 0. \tag{2.5}$$

Below, we assume that $H_{i,t}$ is increasing in $F^{\ell}_{i,t}$ in the sense that if $F^{\ell'}_{i,t}$ stochastically dominates $F^{\ell}_{i,t}$ in terms of the Monotone Likelihood Ratio Property (MLRP), then $H_{i,t}(N_{i,t}, F^{\ell'}_{i,t}) > H_{i,t}(N_{i,t}, F^{\ell}_{i,t})$. In words, complex sectors are "good" sectors in the sense that employment in more complex sectors, perhaps due to international trade, causes higher capability growth.[2] The estimation of such spillover effects will be the main focus of our empirical analysis.

2.2.2 Competitive Equilibrium

We focus on a competitive equilibrium with free trade in goods and financial autarky. At each date t, firms maximize profits, consumers maximize their utility, and goods and labor markets clear. Conditional on the vector of countries' capabilities $N_t = \{N_{i,t}\}$, these static equilibrium conditions determine wages, good prices, consumption, and employment. Employment shares across countries and sectors then determine countries' future capabilities, whereas the path of aggregate consumption determines the interest rate in each country, without any further consequences for our analysis.

Static Equilibrium Conditions. Profit maximization by perfectly competitive firms requires the price of a variety of good k produced in country i and sold in country j to be equal to its unit cost,

$$p^k_{ij,t} = w_{i,t}/A^k_{ij,t} \tag{2.6}$$

with $w_{i,t}$ the wage in country i at date t. If country i cannot produce good k at date t, then $A^k_{ij,t} = 0$ and $p^k_{ij,t} = \infty$. Utility maximization requires

$$c^k_{ij,t} = \frac{(p^k_{ij,t})^{-\sigma}}{(P^k_{j,t})^{1-\sigma}} \frac{(P^k_{j,t})^{1-\epsilon} w_{j,t} L_{j,t}}{(P_{j,t})^{1-\epsilon}}, \tag{2.7}$$

[2] At this point, it is worth noting that this restriction is no less general than assuming that $H_{i,t}$ is monotonic in $F^{\ell}_{i,t}$. Indeed, if $H_{i,t}$ is decreasing in $F^{\ell}_{i,t}$, then one can always reindex goods by a new complexity index $\tilde{n}^k \equiv -n^k$, such that $H_{i,t}$ is increasing in $\tilde{F}^{\ell}_{i,t}$ with $\tilde{F}^{\ell}_{i,t}(n) \equiv \int_{0\leq \tilde{n}^k \leq n} \sum_j \ell^k_{ij,t} dk / \int \sum_j \ell^k_{ij,t} dk$.

where the sector-level price index, $P^k_{j,t}$, and the aggregate price index, $P_{j,t}$, are given by

$$P^k_{j,t} = [\sum_i (p^k_{ij,t})^{1-\sigma}]^{1/(1-\sigma)}, \tag{2.8}$$

$$P_{j,t} = [\int (P^k_{j,t})^{1-\epsilon} dk]^{1/(1-\epsilon)}. \tag{2.9}$$

Good market clearing requires

$$c^k_{ij,t} = A^k_{ij,t} \ell^k_{ij,t}, \tag{2.10}$$

whereas labor market clearing requires

$$\sum_j \int \ell^k_{ij,t} dk = L_{i,t}. \tag{2.11}$$

Dynamic Equilibrium Conditions. For given employment levels $\{\ell^k_{ij,t}\}$, the evolution of capabilities across countries is described by equations (2.4) and (2.5). Finally, the consumer's Euler equation pins down the interest rate in each country,

$$\frac{\dot{C}_{i,t}}{C_{i,t}} = \frac{1}{\nu_i(C_{i,t})}(r_{i,t} - \frac{\dot{P}_{i,t}}{P_{i,t}} - \rho_i). \tag{2.12}$$

where $\nu_i(C) \equiv -d \ln u'_i / d \ln C$ is the elasticity of the consumer's marginal utility.

Definition of a Competitive Equilibrium. A competitive equilibrium corresponds to capabilities, $\{N_{i,t}\}$, wages, $\{w_{i,t}\}$, good prices, $\{p^k_{ij,t}, P^k_{j,t}, P_{j,t}\}$, interest rates, $\{r_{i,t}\}$, consumption levels, $\{c^k_{ij,t}, C^k_{j,t}, C_{j,t}\}$, employment levels, $\{\ell^k_{ij,t}\}$, and employment distributions, $\{F^\ell_{i,t}\}$, such that equations (2.1)-(2.12) hold. Provided that F_t, G_t, and H_t are smooth enough, such a competitive equilibrium exists and is unique. We maintain this assumption throughout. For the interested reader, Appendix 2.8.1 offers a formal discussion.

2.2.3 Pushed to the Top or Held at the Bottom?

Beside the fact that goods and countries may each be ranked along a single dimension, a distinctive feature of the ladder metaphor is that productivity differences across countries and sectors manifest themselves at the extensive margin: capable countries sitting at the top of the ladder can produce the most complex goods, whereas countries at lower rungs cannot. Before turning to our empirical and quantitative analysis, we propose to zoom in on those extensive margin considerations and explore their implications for the relationship between trade, technological capability, and welfare.

The Pure Ladder Economy. Consider an economy, which we refer to as the pure ladder economy, where the only difference across goods is that some goods, the most complex ones, are produced by fewer countries, the most capable ones sitting at the top of the ladder. Formally, we assume that the distribution of productivity G_t is such that

$$A^k_{ij,t} = \begin{cases} A_{ij,t} & \text{if } n^k_t \leq N_{i,t}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

Equation (2.13) allows for arbitrary trading frictions: $A^k_{ij,t}$ may vary across origin and destination countries and over time. The critical restriction that we impose is that $A^k_{ij,t}$ is independent of k for all goods below a country's capability. Hence, comparative advantage is a purely extensive-margin affair.

***All* Pushed to the Top.** To evaluate the consequences of globalization, we compare the time paths of capabilities $\{N_{i,t}\}$ and aggregate consumption $\{C_{i,t}\}$ in the original equilibrium with productivity levels $\{A^k_{ij,t}\}$ to their time paths in a counterfactual autarky equilibrium with productivity levels $\{(A^k_{ij,t})'\}$ such that

$$(A^k_{ij,t})' = \begin{cases} A_{ij,t} & \text{if } n^k_t \leq N_{i,t} \text{ and } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{2.14}$$

All other structural parameters, including the function $H_{i,t}(\cdot,\cdot)$ that determines the law of motion of a country's capability, are held fixed in the two equilibria.

In the autarky equilibrium, all goods produced in a given country i have the same prices, $w_i/A_{ii,t}$; consumers there demand them in the same proportions; and employment shares are equal across sectors. As a result, the autarky employment distribution $F^{\ell,A}_{i,t}$ is equal to F_t in all countries. In the trade equilibrium, this is not the case. By equations (2.6)-(2.10), country i's employment in a sector k with complexity $n^k_t \leq N_{i,t}$ at date t is given by

$$\ell^k_{i,t} = \sum_j \frac{(A_{ij,t})^{\sigma-1}(w_{i,t})^{-\sigma}}{\left(\sum_{l:N_{l,t}\geq n^k_t}(w_{l,t}/A_{lj,t})^{1-\sigma}\right)^{\frac{\epsilon-\sigma}{1-\sigma}}} \frac{w_{j,t}L_{j,t}}{(P_{j,t})^{1-\epsilon}}.$$

As the complexity of goods increases, fewer and fewer countries are able to produce, i.e., fewer countries l satisfy $N_{l,t} > n^t_k$. Under the assumption that $\sigma > \epsilon > 1$, this increases the price index in that sector, $P^k_{j,t} = [\sum_{l:N_{l,t}\geq n^t_k}(w_{l,t}/A_{lj,t})^{1-\sigma}]^{1/(1-\sigma)}$, which further raises sales and employment in country i. For a given level of capability, the distribution of employment $F^{\ell}_{i,t}$ therefore shifts up in terms of MLRP, as illustrated in Figure 2-1. Going from trade to autarky therefore causes a decrease in growth capability, at impact, and a decrease in the level of capability, at all subsequent dates. From a welfare standpoint, these dynamic considerations always strengthen the static case for the gains from trade.

We summarize this discussion in the next proposition. The formal proof can be found in Appendix 2.8.2.

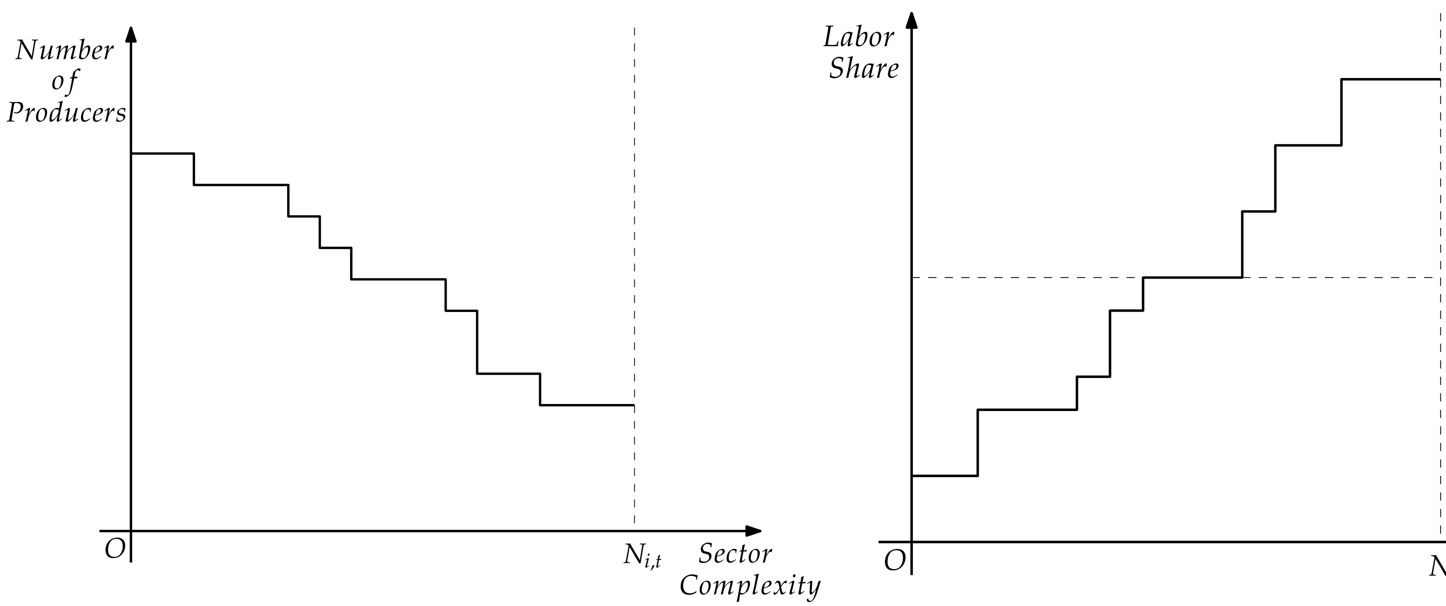
Proposition 11. *In the pure ladder economy, openness to trade raises capability and aggregate consumption at all dates in all countries.*

In the pure ladder economy, all countries gain from trade both because of static and dynamic considerations. The static considerations are standard. For given levels of capability, countries must achieve higher aggregate consumption under trade than under autarky because the autarkic consumption bundle remains achievable under free trade, as in Samuelson (1939). The dynamic considerations pertain to the endogenous evolution of countries' capability under trade and autarky.

At arbitrary points in time, changes in capability may well be lower under trade than what it would have been under autarky. Nevertheless, whenever capability levels coincide in the trade and autarky equilibria, differences in the distribution of employment between the two equilibria imply higher capability growth in the former. This is sufficient to guarantee higher levels of capability under trade at all dates.Dynamic gains from trade are not zero-sum, in the sense that some countries experience dynamic gains at the expense of others by specializing in the good sectors, whereas other countries do not. Here, globalization pushes all countries up the ladder. At worst, capability remains the same in the two equilibria, which is what happens in the country with the lowest capability. For this country, since competition from the rest of the world is exactly the same in all sectors at the bottom of the ladder, the distribution of employment across sectors remains given by F_t after opening up to trade.

Although the pure ladder economy imposes strong restrictions on the distribution of productivity across countries and sectors, those that we think capture well the original ladder metaphor, it allows for a general law of motion for capability and, in turn, rich dynamics for the distribution of productivity across countries and sectors. Proposition 11, for instance, can accommodate scale effects, such that variation in the size of sectors, rather than their shares of total employment, matters for capability. This simply corresponds to the special case where $H_{i,t}$ is a function of $L_{i,t}$. We can also generalize Proposition 11 in a straightforward manner to environments where productivity differences across countries and sectors take the form $A^k_{ij,t} = A_{ij,t} B^k_{j,t} C_{i,t}$.[3]

[3]The critical feature of a pure ladder economy, already emphasized above, is that comparative advan-



(a) Foreign Competition

(b) Employment Distribution

Figure 2-1: Changes in employment distribution after opening up to trade

Valuation of the Gains from Trade. A natural way to measure the extent to which dynamic economies of scale affect the gains from trade is to compare gains from trade in the present environment to those predicted by the formula in Arkolakis, Costinot, and Rodríguez-Clare (2012) (ACR). In line with the static analysis in ACR, we assume that the world economy is initially at a steady state, with no technological shocks ($F_t = F$, $G_t = G$, and $H_{i,t} = H_i$) and no population growth ($L_{i,t} = L_i$), and that after moving to autarky, the economy converges to a new steady state, with the transitional dynamics to this new steady state determined by H_i. We can then define the gains from trade as the (permanent) difference between the income level required to achieve the (lifetime) utility under free trade and the income level required to achieve the (lifetime) utility under autarky, both evaluated at the free trade prices and expressed as a fraction of a country's income level under free trade.

Let $\lambda_{ii}(n)$ denote country i's share of expenditure on domestic goods with complexity n in steady state under free trade, let $e_i(n)$ denote country i's share of total expenditure on goods with complexity n in that same steady state, and let $H_i^{-1}(0, \tilde{F})$ denote the capability level $\tilde{N}$ that solves $0 = H_i(\tilde{N}, \tilde{F})$. As shown in Appendix 2.8.2, the previous statistics can be used to value the gains from trade (GT).

Proposition 12. *In the pure ladder economy, gains from trade in any country i are bounded from below by $\underline{GT}_i$ and above by $\bar{GT}_i$ such that*

$$\underline{GT}_i = 1 - \underbrace{\left[\int e_i(n)(\lambda_{ii}(n))^{\frac{\epsilon-1}{\sigma-1}} dF(n) \right]^{\frac{1}{\epsilon-1}}}_{\text{Static Gains}},$$

$$\bar{GT}_i = 1 - \underbrace{\left[\int e_i(n)(\lambda_{ii}(n))^{\frac{\epsilon-1}{\sigma-1}} dF(n) \right]^{\frac{1}{\epsilon-1}}}_{\text{Static Gains}} \cdot \underbrace{\left[H_i^{-1}(0, F_i^{\ell}) / H_i^{-1}(0, F) \right]^{\frac{1}{(1-\epsilon)}}}_{\text{Dynamic Gains}}.$$

The first term, $\left[\int e_i(n)(\lambda_{ii}(n))^{\frac{\epsilon-1}{\sigma-1}} dF(n) \right]^{\frac{1}{\epsilon-1}}$, is standard in the quantitative trade literature. It corresponds to the static gains from trade, as previously described by Costinot and Rodríguez-Clare (2013) in the case of a similar multi-sector Armington model with

tage only expresses itself at the extensive margin. More generally, if we were to allow more capable countries to have a comparative in more complex goods—in the sense that $A_{ij,t}^k$ is log-supermodular in $(n_t^k, N_{i,t})$, but does not necessarily satisfy condition (2.13)—then there would be dynamic gains for the most capable country at the top of the ladder, dynamic losses for the least capable one at the bottom, and either dynamic losses or gains for any country in between. The basic logic is unchanged. The most capable country still faces tougher foreign competition for its least complex products, now both through intensive and extensive margin considerations. In contrast, the least capable country now faces tougher competition for its most complex products, exclusively through intensive margin considerations.

nested CES utility. Like the original one-sector ACR formula, this expression uses expenditure shares in the current trade equilibrium, $\{\lambda_{ii}(n)\}$ and $\{e_i(n)\}$, to infer by how much country i's terms-of-trade would worsen as the economy goes back to autarky, a terms-of-trade adjustment that also depends on the elasticities of substitution across and within sectors, ϵ and σ. The lower those two elasticities are, the more a country's terms-of-trade worsen, and the larger the gains from trade are.

The second term, $\left[H_i^{-1}(0, F_i^{\ell})/H_i^{-1}(0, F)\right]^{\frac{1}{(1-\epsilon)}}$, is the main focus of our analysis. It captures the fact that, in addition to the previous welfare losses, country i's production possibility frontier would also be affected by moving back to autarky. In the trade steady state, country i's capability is given by $N_i = H_i^{-1}(0, F_i^{\ell})$. In the autarky steady state, in contrast, uniform employment across sectors implies $(N_i)' = H_i^{-1}(0, F)$. The (inverse of the) function H_i determines by how much changes in the pattern of sectoral employment, from F_i^{ℓ} to F, affects the growth of capability along the transition path and, in turn, the level of capability in steady state. With F_i^{ℓ} stochastically dominating F, moving back to autarky also reduces aggregate productivity in country i, with the mapping between the change in the number of goods that country can produce, $N_i/(N_i)'$, and aggregate productivity given by $\frac{1}{1-\epsilon}$. This captures the standard "love of variety" associated with CES utility.

Our two bounds on the gains from trade derive from the observation that any point along the transition path from the trade steady state to the autarky steady state, country i's capability must always lie between N_i and $(N_i)'$. As a result, the welfare losses from autarky, and hence the gains from trade, must be bounded from below by the welfare loss, $\underline{GT}_i$, that would occur absent any change in capability. Likewise, the gains from trade must be bounded from above by the welfare loss, $\overline{GT}_i$, that would occur if capability jumped immediately and permanently to its lower steady state level under autarky.

2.3 Measuring Capability and Complexity

The end goal of our paper is to test empirically whether, as illustrated by Proposition 11, opening up to trade may be a force that tends to push all countries up the capability ladder by allowing them to specialize in their more complex sectors. To confront that hypothesis with data, we first need measures of capability and complexity.

2.3.1 Empirical Strategy

Our general empirical strategy is to use zero and non-zero trade flows in order to partially identify the distribution of productivity $G_t(\cdot|n, N)$ and, in turn, the capability and complexity indices that shape that distribution at any given point in time. For our baseline analysis, we restrict the productivity distribution G_t to be such that

$$\text{Prob}(A^k_{ij,t} > 0) = \delta_{ij,t} + \gamma^k_{j,t} + N_{i,t} n^k_t, \tag{2.15}$$

with independence across origins and sectors, but not necessarily across destinations within the same origin and sector. Consistent with the original ladder metaphor discussed in the introduction, equation 2.15 captures well the notion that "complex goods are what capable countries do" in the sense that more complex goods are more likely to be exported by more capable countries. This is also similar in spirit to the existing empirical literature, e.g. Hausman et al. (2007) and Hausman et al. (2013), that extracts measures of country and product sophistication from export patterns.

Note that equation 2.15 includes exporter-importer-year and importer-good-year specific terms, $\delta_{ij,t}$ and $\gamma^k_{j,t}$, respectively. In line with recent work on revealed comparative advantage (e.g. Costinot, Donaldson, and Komunjer 2012, Levchenko and Zhang, 2016, and Hanson, Lind, and Muendler 2016), this allows us to separate the determinants of comparative advantage, capability and complexity, from bilateral trading frictions and demand differences across countries.

In our model, trade flows $x^k_{ij,t} \equiv p^k_{ij,t} c^k_{ij,t}$ by country i to country $j \neq i$ in sector k at date t are strictly positive if and only if productivity $A^k_{ij,t}$ is strictly positive. Thus, under the assumption above, we can estimate $N_{i,t}$ and n^k_t as interacted country-year and sector-year fixed effects in a linear probability model of the form,

$$\pi^k_{ij,t} = \delta_{ij,t} + \gamma^k_{j,t} + N_{i,t} n^k_t + \epsilon^k_{ij,t}, \tag{2.16}$$

where $\pi^k_{ij,t}$ is the dummy variable for whether or not $x^k_{ij,t} > 0$ and the error term $\epsilon^k_{ij,t} \equiv \pi^k_{ij,t} - E[\pi^k_{ij,t}]$. Intuitively, if a country is known to be more capable than another, say the United States (US) versus Bangladesh (BG), then one can identify any good k as more complex than another reference good k_0 if, relative to the reference good, it is more likely to be exported by the United States than Bangladesh. Indeed, if there are no error terms, then equation (2.16) directly implies

$$n^k_t - n^{k_0}_t = [(\pi^k_{USj,t} - \pi^{k_0}_{USj,t}) - (\pi^k_{BGj,t} - \pi^{k_0}_{BGj,t})]/(N_{US,t} - N_{BG,t}).$$

Conversely, if a good is known to be more complex than another, say medicines (ME) versus men's underwear (UW), then one can identify any country i_1 as more capable than another reference country i_0 if, relative to the reference country, it is more likely to export medicines than underwear. Again, in the absence of error terms, equation (2.16) implies

$$N_{i,t} - N_{i_0,t} = [(\pi_{ij,t}^{ME} - \pi_{ij,t}^{UW}) - (\pi_{i_0j,t}^{ME} - \pi_{i_0j,t}^{UW})]/(n_t^{ME} - n_t^{UW}).$$

Our estimators of capability and complexity generalizes the previous idea to the case where there are error terms, but those are mean zero.[4] Our procedure requires an initial guess of either which countries are capable or which goods are complex. We follow the first path and assert that the original members of the G-10 are capable which provides complexity measures for all goods which can then be used to recover capabilities for all countries. The details of that estimation procedure can be found in Appendix 2.9.2.

Our procedure identifies capability and complexity, up to affine transformation. For the purposes of identifying dynamic spillovers in Section 2.4.3, we will further assume that the lowest and highest complexity levels are time-invariant and always equal to 0 in the least complex sector, i.e. there is no spillover from producing the least complex product, and moving from specializing in the least to the most complex product generates the same-sized spillover in any period. Given these two assumptions, we can normalize complexity in the most complex sector to 1 without further loss of generality. In our empirical exercises below, year fixed effects will sweep out average capability in any period. To ease exposition, we normalize capability such that the average capability of the US is equal to 1.

2.3.2 Data

Our baseline empirical analysis uses trade data from the UN Comtrade database for 146 countries and 715 4-digit SITC Rev. 2 manufacturing products from 1962 to 2014.

The UN Comtrade database contains more than 3 billion records on annual imports and exports by detailed product code going back as far as 1962. We start by extracting all trade transactions between 1962 and 2014 with transactions in at least one year for 233 countries.[5] Transactions are concorded to the 4-digit SITC rev 2 level by Comtrade

[4] Specifically, we assume $\epsilon_{ij,t}^k = \xi_{i,t}^k + u_{ij,t}^k$, where $\xi_{i,t}^k$ is i.i.d and mean zero across both products and origins and $u_{ij,t}^k$ is i.i.d and mean zero across products, origins and destinations.

[5] We combine East and West Germany in the years prior to reunification. Several countries report jointly for subsets of years in the database. For this reason, we combine: Belgium and Luxembourg; the islands that formed the Netherlands Antilles; North and South Yemen; and Sudan and South Sudan.

and all trade flows are converted into real 2010 US dollars using the US CPI. We then perform a number of data cleaning steps that closely follow the cleaning exercise outlined in Feenstra et al. (2005) (e.g. giving primacy to importer's reports where available, correcting values where UN values are known to be inaccurate, and accounting for re-exports of Chinese goods through Hong Kong).[6] This procedure gives us the value of trade flows $x^k_{ij,t}$ by country i to country $j \neq i$ in sector k at date $t = 1962, ..., 2014$. To ensure that estimates of the linear probability model (2.16) are picking up genuine exporting relationships as opposed to sending samples or small quantities of re-exports, we set the dummy variable $\pi^k_{ij,t}$ for a strictly positive trade flow equal to 1 if the value of exports is equal or greater than \$100,000 in 2010 US dollars and zero otherwise.

We restrict our attention to manufacturing sectors. These are the sectors where we expect technological spillovers emphasized in our theory to be relevant. Out of 1067 4-digit SITC rev 2 products in the full dataset, this leaves us with 715 sectors.

Our baseline sample of countries satisfies two restrictions. First, as we will ultimately be running panel regressions exploring how capability growth responds to the complexity of goods being produced, we eliminate countries with fewer than 40 years of data. This restriction eliminates 84 countries that are either newly formed, no longer exist, or infrequently report. Second, to ensure that results are not driven by the world's smallest countries, we eliminate 33 countries whose exports, averaged over any 5 year period, never rise above \$100 million in 2010 prices. As there is substantial overlap in the countries eliminated by these two restriction, the final sample contains 146 countries that we list in Appendix Table 2.9.1. For robustness we explore additional samples that remove the panel requirement, remove the size threshold, or expand the size threshold to US \$1 billion.

2.3.3 Baseline Estimates of Capability

Before using our capability and complexity estimates to uncover the sign and strength of dynamic spillovers in Section 2.4, we start by describing how capability varies across countries and over time. Figure 2-2 plots the evolution of the recovered capability estimates, $N_{i,t}$, for a range of similarly-sized countries spanning different level of incomes both today and in the 1960s.

[6]The dataset produced by Feenstra et al. (2005) has two shortcomings for our purposes. First, it only covers the years 1962-1999. Second, purchasing restrictions meant that for the years 1984-1999 they were only able to use trade flows that exceeded \$100,000 per year and only for 72 reporter countries. Thus we use the Feenstra et al. (2005) dataset from 1962-1983 but construct our own dataset for the years 1984-2014 using the full set of trade flows and reporter countries. We perform robustness exercises replacing the 1984-1999 entries in our dataset with the entries from Feenstra et al. (2005).

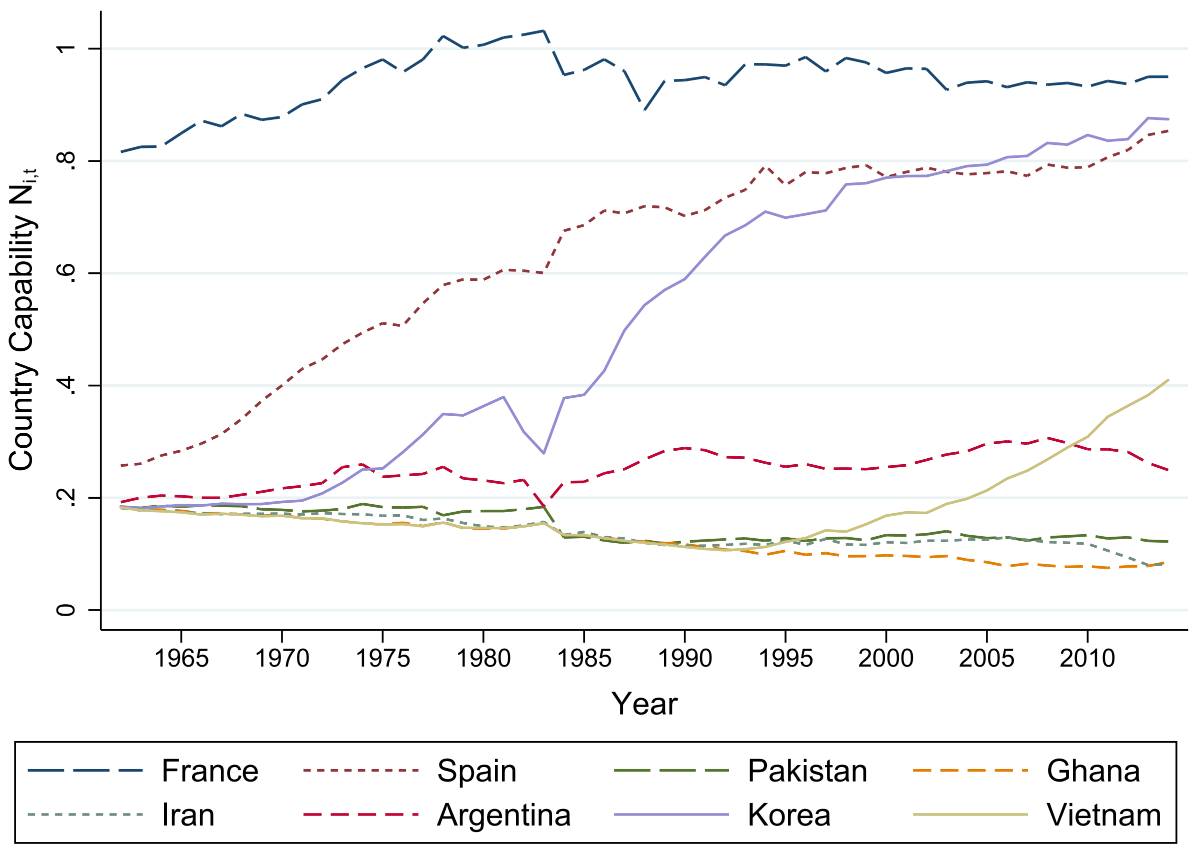


Figure 2-2: Baseline Capability ($N_{i,t}$)

Notes: Figure 2-2 reports the country capability measure $N_{i,t}$ from the linear probability model estimation of equation (2.16) in a given year t. Capability is normalized so that the average capability of the US is 1.

The estimates of $N_{i,t}$ resonate well with widespread priors about levels of economic development across countries and over time. Recall that our normalization sets capability in the US to 1 in every period. Western European countries (e.g. France) have consistently maintained high levels of capability only a little below the US. Starting from somewhat lower initial positions, Southern European countries such as Spain saw their capabilities converge with Western Europe, with particularly rapid convergence in the first 20 years of our sample. Poor African countries such as Ghana had massively lower capability at the start of the period and have seen limited catch up since. A similar lack of catch up is evident for South Asian and Middle-Eastern countries such as Pakistan and Iran. While starting at somewhat higher levels, middle-income South American countries such as Argentina display a similarly-flat trajectory. Finally, the rapid ascent of the East Asian Tigers (e.g. Korea) in the 1960's through 1990s and the more recent South-East Asian growth miracles such as Vietnam show up clearly.

We can also study the relationship between our capability estimates and levels of development more formally by exploring the association with real GDP per capita from the Penn World Tables. To examine the variation across countries within each year, we run a panel regression of log real GDP per capita on both capabilities and year fixed effects. We find a very strong positive relationship—a coefficient of 2.9 with a standard error of 0.05—and a high within R-squared of 0.3 (i.e. we explain about a third of the variation within year). If we additionally include country fixed effects, and so are also exploiting variation across time within countries, we still find a strong relationship although explain less of the variation (a coefficient of 1.7 with a standard error of 0.06 and a within R-squared of 0.11).[7]

2.3.4 Baseline Estimates of Complexity

We now turn to our baseline estimates of product complexity. Table 2.1 reports the goods with the 10 highest and 10 lowest average complexity across all years from 1962 to 2014. The ranking of those products also fits well priors about technological sophistication across sectors, and hence the potential for knowledge spillovers. Medicaments, chemicals and cars, for instance, are among the most complex products throughout our sample, whereas men's underwear, wood panels and plastic ornaments are among the least complex ones.

[7] Figures 2.9.1 and 2.9.2 in the appendix present these relationships visually via binned scatterplots.

Table 2.1: Baseline Complexity (n_t^k)

	Sectors with highest n^k (Average Value, 1962-2014)	
1	Medicaments	0.964
2	Miscellaneous Non-Electrical Machinery Parts	0.878
3	Chemical Products	0.872
4	Cars	0.861
5	Miscellaneous Non-Electrical Machines	0.857
6	Miscellaneous Electrical Machinery	0.831
7	Miscellaneous Hand Tools	0.808
8	Medical Instruments	0.805
9	Electric Wire	0.768
10	Fasteners	0.759
	Sectors with lowest n^k (Average Value, 1962-2014)	
1	Wool Undergarments	0.067
2	Undergarments of Other Fibres	0.083
3	Men's Underwear	0.100
4	Wood Panels	0.096
5	Aircraft Tires	0.089
6	Rotary Converters	0.081
7	Sheep and Lamb Leather	0.110
8	Retail Yarn of More Than 85% Synthetic Fiber	0.091
9	Women's Underwear	0.115
10	Plastic Ornaments	0.137

Notes: Table 2.1 reports the sectors with the 10 highest and 10 lowest average values of n_t^k from 1962 to 2014 for products with at least 40 years of data. Complexity n_t^k is estimated year-by-year using the linear probability model described in equation (2.16). Complexity is set to 0 and 1 for the least and most complex good, respectively, at all dates t.

2.3.5 Comparison to Earlier Work

To conclude, we compare our baseline measures of capability and complexity to the original work of Hausman et al. (2007) and Hausman et al. (2013) who also use trade data to construct technological indices across products and countries.

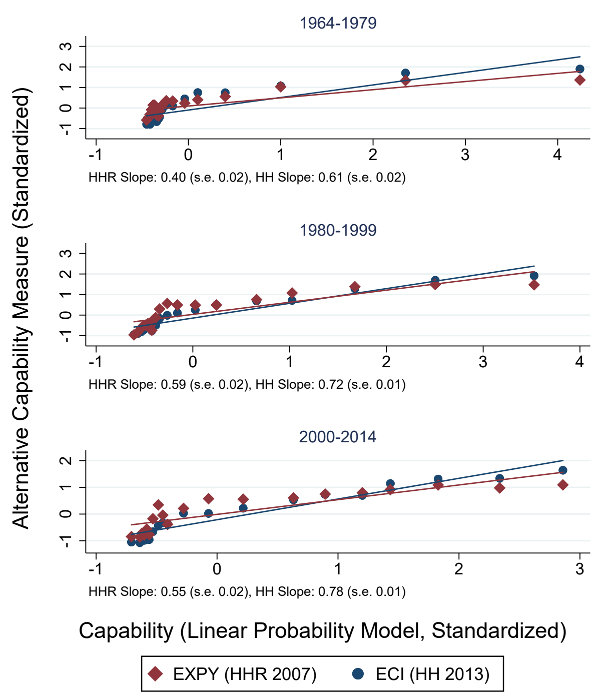
In Hausman et al. (2007), the counterpart of product's complexity, $PRODY^k$, is defined as the weighted-sum of GDP per capita, Y_i, with weights equal to Balassa's (1965) measure of revealed comparative of comparative advantage in country i and sector k, whereas the counterpart of a country's capability, $EXPY_i$, is equal to the weighted sum of $PRODY^k$, with weights equal to the share country i's exports in sector k.

In Hausman et al. (2013), the counterparts of capability and complexity, ECI_i and PCI^k, also focus on the extensive margin of trade. In practice, Hausman et al. (2013) go first from the raw matrix of zero trade flows to a matrix whose entries take a value of one if Balassa's (1965) revealed measure of comparative advantage is greater than one, and zero otherwise; they then compute normalized versions of the product of that rectangular matrix with its transpose as well as the product of the transpose with the matrix; and finally, they define the vectors of complexity and capability as the eigenvectors associated with the second-largest eigenvalues of these two matrices, normalized by the mean and standard deviation of each eigenvector.[8]

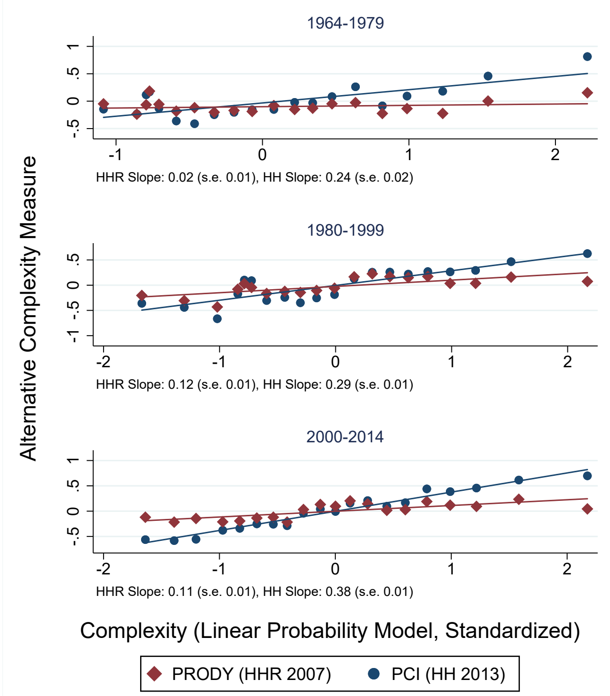
Figure 2-3 reports how our baseline measures (on the x-axis) correlate with the measures of complexity and capability in Hausman et al. (2007) and Hausman et al. (2013) (red diamonds and blue circles) in different decades of our sample.[9] As can be clearly seen from Figures 2-3a and 2-3b, the three empirical measures are strongly and positively correlated. This derives from the fact that all three are designed to capture the same general idea that complex goods are what capable countries exports, and vice versa. A benefit of our linear probability model, and the reason why we use it instead of those existing measures, is that it directly maps into primitive assumptions about technology. We will use this feature to conduct counterfactual and welfare analysis in Section 2.5.

[8] Schetter (2019) uses a similar approach, but starts from structural estimates of productivity in a multi-sector model à la Costinot et al. (2012) rather than Balassa's (1965) measure.

[9] Figures are binscatters from regressing each alternative measure on our baseline, controlling for year fixed effects. We start in 1964 rather than 1962 as the Hausman et al. (2007) measures are only available from that year forward.



(a) Capability Measures (N_{it})



(b) Complexity Measures (n_t^k)

Figure 2-3: Alternative Measures of Capability and Complexity

Notes: Figure 2-3 compares our baseline measures of capability $N_{i,t}$ and complexity n_t^k from the linear probability model estimation of equation (2.16), as described in Section 2.3.1, to the capability and complexity measures in Hausman et al. (2007) (labeled EXPY and PRODY) and Hausman et al. (2013) (labeled ECI and PCI). Figure plots binscatters of regressions of our measures on the alternative measures, absorbing year fixed effects and pooling observations by time period. Regression slope and standard error shown under each figure. All measures standardized mean 0 standard deviation 1 in each year.

2.4 Estimating Dynamic Spillovers

2.4.1 Baseline Specification

The theoretical framework of Section 2.2 focuses on an environment with continuous time and a general law of motion for capability. To estimate dynamic spillovers, and later to quantify their implications, we assume instead that time is discrete,

$$N_{i,t+\Delta} - N_{i,t} = H_{i,t}(N_{i,t}, F^{\ell}_{i,t})$$

with Δ corresponding to a 5-year period in our baseline analysis, and we impose the following parametric restrictions on the law of motion for capability,

$$H_{i,t}(N_{i,t}, F^{\ell}_{i,t}) = \beta \int n dF^{\ell}_{i,t}(n) + (\phi - 1) N_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t}.$$

Finally, given the lack of comparable production data at the product level across many countries and time periods, we use trade data to proxy country i's distribution of employment, $F^{\ell}_{i,t}(n)$, by its distribution of exports, $F^{x}_{i,t}(n) = \sum_j \int_{0 \leq n^k \leq n} x^k_{ij,t} dk / \sum_j \int x^k_{ij,t} dk$.[10] For future reference, note that this implies that country i is able to produce good k for the domestic market, $A^k_{ii,t} > 0$, if and only if it is able to export it to at least one of its 145 trading partners. We will maintain this assumption in our quantitative analysis.

Combining the two previous equations and letting $S_{i,t} = \int n dF^{x}_{i,t}(n)$ denote the average complexity of country i's industry mix at date t, we obtain the following baseline specification,

$$N_{i,t+\Delta} = \beta S_{i,t} + \phi N_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t}. \tag{2.17}$$

The first parameter, β, is the main coefficient of interest. It measures the magnitude of the dynamic spillovers. If $\beta > 0$, then a shift in the distribution of employment that increases the average complexity of country i's industry mix—for example, opening up to international trade that by construction exposes the least complex goods to the largest increases in foreign competition—also increases capability growth at impact. If $\beta < 0$, the converse is true and increases in average complexity reduce capability growth.

The second parameter, ϕ, determines the persistence of shocks. If $\beta > 0$ and $\phi < 1$, then positive and permanent shocks to the average complexity of a country's industry

[10] This is equivalent to assuming that the unobserved sector-level domestic sales, $x^k_{ii,t}$, are proportional to total exports in each sector, $x^k_{ii,t} = \zeta_{i,t}(\sum_{j \neq i} x^k_{ij,t})$, for some time-and-country specific shifter $\zeta_{i,t}$. In Section 2.5, we will use data on total gross output in manufacturing to pin down $\zeta_{i,t}$ so that total domestic sales are consistent with both aggregate trade and production data.

mix leads to an increase in capability changes, in the short-run, and convergence to a new steady state with higher capability level in the long-run. In the knife-edge case $\phi = 1$, permanent shocks to average complexity have permanent effects on capability changes.[11]

The third parameter, γ_i, captures all country-specific determinants of capability growth that are constant over the 50-year horizon that we consider, such as geography or the origin of country i's legal system. The fourth parameter, δ_t, captures time-specific determinants due to global innovation such as the introduction of the internet. The final term, $\varepsilon_{i,t}$, captures all other idiosyncratic sources of technological innovations and domestic policies that may affect capability growth.[12]

2.4.2 Construction of Instrumental Variables

The main endogeneity concern is that shocks to country i's capability, $\varepsilon_{i,t}$, may be correlated with shocks to the average complexity of its industry mix in period t, $S_{i,t}$. For example, this may occur because "good" policies implemented in period t, like investment in R&D and education, simultaneously promote specialization in complex sectors and capability growth, leading to upward bias in β. Or, this may occur because "bad" policies, like subsidies to more complex sectors associated with rent-seeking, expand more complex sectors, but reduce capability growth, leading to downward bias in β. We now describe how we construct instrumental variables to deal with this issue.[13]

We consider two distinct instrumental variables (IV). For both of them, the general idea is to use the entry of countries into the WTO as an exogenous shifter of other countries' distribution of employment, $F^{\ell}_{i,t}$, and, in turn, the average complexity of its exports, $S_{i,t}$. As country c enters the WTO at date t, it faces lower tariffs from current WTO mem-

[11]Mathematically, ϕ plays a similar role as the returns to scale for ideas in endogenous growth models, for which $\phi = 1$, and semi-endogenous growth models, for which $\phi < 1$. See Jones Jones (1999) and Burstein and Atkeson (2019) for general discussions. Quantitatively, the magnitude of the dynamic gains from trade depends both on β and ϕ. For instance, in the case of the pure ladder economy described in Section (2.2.3), if $H_i(N_i, F^{\ell}_i) = \gamma_i + \beta \int n dF^{\ell}_i(n) + (\phi - 1)N_i$, then dynamic gains are given by

$$\left[H_i^{-1}(0, F^{\ell}_i)/H_i^{-1}(0, F)\right]^{\frac{1}{(1-\epsilon)}} = \left[\frac{\beta[\int n dF^{\ell}_i(n) - \int n dF(n)]}{(1-\phi)}\right]^{\frac{1}{(1-\epsilon)}}.$$

[12]Although $L_{i,t}$ does not appear on the right-hand side of the previous specification, it is worth noting that it implicitly allows for some scale effects. Here, both systematic differences in country size, absorbed in γ_i, and uniform changes in the world population, absorbed in δ_t, may affect capability growth.

[13]Another endogeneity concern is standard in panel models with fixed effects. The lagged dependent variable, $N_{i,t}$, is mechanically correlated with the demeaned error term that accounts for the country fixed effect, $\varepsilon_{i,t} - \sum_{s=1}^{T} \frac{\varepsilon_{is}}{T}$, the so-called Nickell (1981) bias. We briefly discuss how we deal with this issue as well at the end of this section.

bers. This tends to increase the demand for labor from country c and lowers the demand for labor from other countries, but differentially so across sectors depending on their exposure to exports from country c. This, in turn, leads to differential effects of the entry of country c on another country i's average complexity depending on whether country i's more complex sectors are those that are more or less exposed to country c—either because c sells a similar set of products, sells to a similar set of countries, or both. Our two IVs builds on the previous observation and the assumption that the changes in trade costs associated with country c's accession to the WTO, and hence any function of them, are orthogonal to shocks to capacity, $\{\varepsilon_{i,t}\}$, in other countries.

More specifically, as described in Appendix 2.9.4, we model the entry of any given country c into the WTO at some date t_c as a uniform and permanent trade cost shock. We then compute, up to a first-order approximation, the counterfactual change in the average complexity of other countries that that would have been observed at dates $t \geq t_c$, assuming the entry of country c was the only shock occurring from period t_c onward and ignoring general equilibrium adjustments in wages.[14] Finally, we sum the previous changes across all WTO entry events prior to date t.

Given the nested structure of preferences in our model, this approach delivers two IVs. The first one corresponds to the changes in average complexity caused by the changes in sector-level price indices associated with the trade cost shock,

$$Z^{I}_{i,t} = \sum_{c \neq i} \mathbb{1}_{\{t \geq t_c\}} \sum_{k} n^{k}_{t_c-1} \times \underbrace{\omega^{k}_{i,t_c-1} \big(\sum_{j \neq c} \rho^{k}_{ij,t_c-1} \lambda^{k}_{cj,t_c-1} - \sum_{k'} \omega^{k'}_{i,t} \sum_{j \neq c} \rho^{k'}_{ij,t_c-1} \lambda^{k'}_{cj,t_c-1}\big)}_{\text{shift in } k\text{'s employment share predicted by sector-level price changes}} ,$$

where $\omega^{k}_{i,t} \equiv \ell^{k}_{i,t}/L_{i,t}$ is the employment share of sector k in country i at date t; $\rho^{k}_{ij,t} \equiv \ell^{k}_{ij,t}/\ell^{k}_{i,t}$ is the share of employment in country i and sector k associated with destination j at that date, and is $\lambda^{k}_{cj,t}$ the share of expenditures on goods from country c in sector k and destination j. The second IV corresponds to the changes in average complexity caused by the changes in aggregate-level price indices,

$$Z^{II}_{i,t} = \sum_{c \neq i} \mathbb{1}_{\{t \geq t_c\}} \sum_{k} n^{k}_{t_c-1} \times \underbrace{\omega^{k}_{i,t_c-1} \big(\sum_{j \neq c} \rho^{k}_{ij,t_c-1} \lambda_{cj,t_c-1} - \sum_{k'} \omega^{k'}_{i,t} \sum_{j \neq c} \rho^{k'}_{ij,t_c-1} \lambda_{cj,t_c-1}\big)}_{\text{shift in } k\text{'s employment share predicted by aggregate-level price changes}} ,$$

[14] Taking into account those adjustments would require us to already take a stand on the structural parameters of the model. For the purposes of constructing IV, ignoring those adjustments may weaken our first stage, but it does not affect the validity of our exclusion restriction. As mentioned above, the assumption that we impose is that any function of changes in trade costs associated with country c's accession to the WTO is orthogonal to capacity shocks in other countries.

where $\lambda_{cj,t}$ denotes the share of expenditure on goods from country c (across all sectors) in country j at date t.

Intuitively, $Z^{I}_{i,t}$ captures similarity in both the products i and the WTO entrant sells to different countries as well as which countries they sell to, while $Z^{II}_{i,t}$ focuses on the overlap in i's export mix and the countries the WTO entrant sells to. While the destination-level variation in λ_{cj,t_c-1} is subsumed in the destination-product level variation $\lambda^{k}_{cj,t}$, the first order approximation requires both $Z^{I}_{i,t}$ and $Z^{II}_{i,t}$ are included when the elasticity of substitution between sectors, ϵ, is not unity. Of course, if there is no variation in expenditure shares across sectors, so that $\lambda^{k}_{cj,t} = \lambda_{cj,t}$, then changes in sector-level price indices are perfectly collinear with changes in aggregate-level prices across destinations. In this case, $Z^{I}_{i,t}$ and $Z^{II}_{i,t}$ are perfectly collinear as well. Finally, note that since equation (2.17) features country and year fixed effects, the identifying variation comes from new entrants in a given time period that disproportionately affect country i' mix of more and less complex products based on the entrants export patterns prior to their entry.

Figure (2-4) illustrates the time path of our first instrumental variable for the same subset of countries as in Figure 2-2. To illustrate our identifying variation, consider the dramatic rise in $Z^{I}_{i,t}$ in 2001 for Vietnam associated with the entry of China into the WTO. That is, according to our first-order approximation, competition from new WTO entrants in 2001 affected products that were relatively complex compared to Vietnam's product mix, potentially shifting them towards less complex products (a relationship we will document in our first stage regressions). In contrast, France experienced a drop in $Z^{I}_{i,t}$ with China's entry as its relatively less complex sectors experienced greater competition, potentially tilting it towards producing more complex products. To identify whether those sectors are good or bad for capability growth, we can therefore verify whether Vietnam experienced a slowdown relative to other countries post 2000 and whether France experienced an acceleration.[15]

2.4.3 Estimates of Dynamic Spillovers

Before presenting our estimates of the sign and size of dynamic spillovers, we first show our first stage regressions for the IV strategy. In Table 2.9.3 we regress the average complexity of country i's industry mix, $S_{i,t}$, on the two instruments described above,

$$S_{i,t} = \alpha_1 Z^{I}_{i,t} + \alpha_2 Z^{II}_{i,t} + \gamma_i + \delta_t + u_{i,t} \tag{2.18}$$

[15] Appendix Figure 2.9.3 reports a very similar timepath for our second IV.

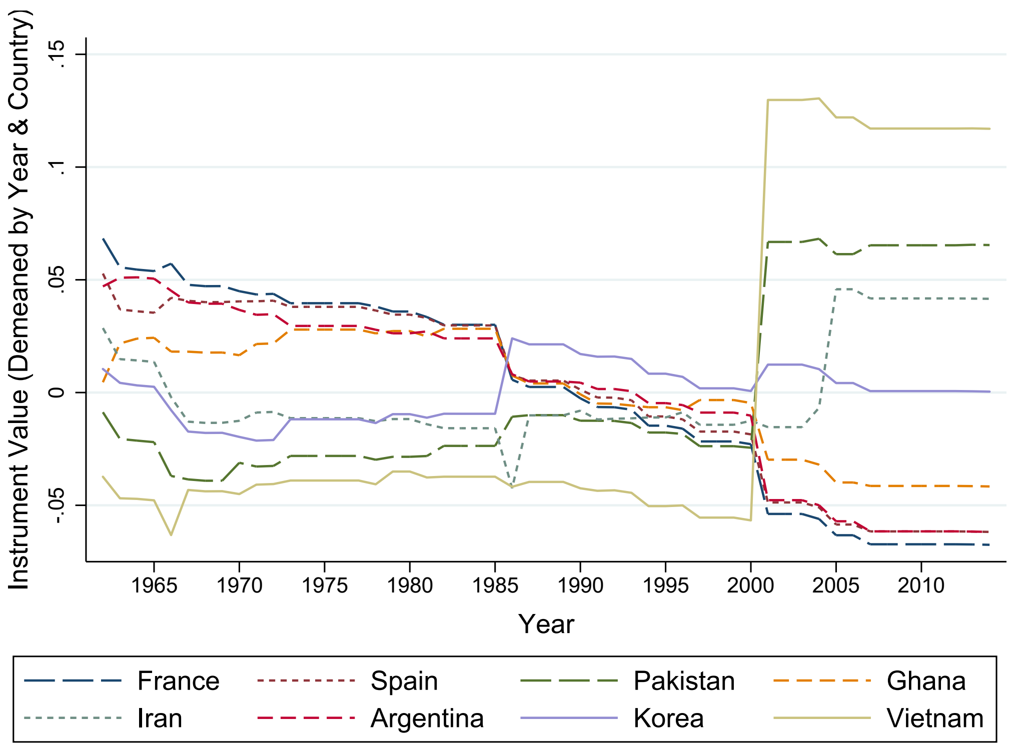


Figure 2-4: Time Path of $Z^I_{i,t}$

Notes: Figure 2-4 plots the value of the instrument $Z^I_{i,t}$ over time for a selection of similarly-sized countries in our sample. The instrument captures the change in complexity-weighted competition due to sector-level price index changes induced by other countries' entry into the WTO and derives from a first-order approximation of the change in average complexity due to trade cost shocks to WTO entrants (see Appendix 2.9.4).

Table 2.2: First Stage Regressions: Responses of Industry Structure to WTO Entrants

	Average Complexity $S_{i,t}$	
	(1)	(2)
WTO Entrant Shock $Z^{I}_{i,t}$	-0.674***	-0.186
(Product-Destination Level)	(0.212)	(0.223)
WTO Entrant Shock $Z^{II}_{i,t}$		-4.017***
(Destination Level)		(0.793)
Country and year FEs	Yes	Yes
Observations	7,617	7,617
R-squared	0.586	0.592
Clusters	1588	1588

Notes: Table 2.2 reports estimates of α_1 and α_2 in equation (2.18) using the baseline measure of complexity n^k_t from the linear probability model estimation of equation (2.16). Standard errors clustered at the 5-year-period-country level.

As in the second stage regressions, we include year and country fixed effects, γ_i and δ_t respectively. Column 1 presents the first stage regression using only a single instrument $Z^{I}_{i,t}$, while column 2 presents both instruments.We find very strong negative relationships, either when focusing on sector-level price shifts or when adding aggregate-level price effects. When both instruments are included, it is the second instrument that focuses on overlap in i's export mix and the countries the WTO entrant sells to that dominates. Interpreted through the lens of our first-order approximation, the negative sign of α_1 points towards a lower-level elasticity of substitution (between countries within a sector) that is greater than the upper-level elasticity of substitution (between sectors), whereas the negative sign of α_2 suggests an upper-level elasticity of substitution that is strictly greater than one; see Appendix 2.9.4.

We now turn to estimating dynamic spillovers, β in equation (2.17) above. Table 2.9.4 presents the main regressions of the country capability on the average complexity of the product mix in the previous period.[16] Column 1 shows the ordinary least squares regressions while columns 2 and 3 present the IV regressions using the WTO-entry instruments.

[16]Recall that we chose a period length of 5 years as is common in growth-type regressions of this sort. Thus, when indexed in years instead, equation (2.17) corresponds to a regression of $N_{i,year+5}$ on $S_{i,year}$ and control for both the initial level of country capability, $N_{i,year}$, as well as country and year fixed effects. Given the 5-year lead on the dependent variable, observations within 5-year periods are not independent and so we cluster standard errors at the 5-year-period-country level.

Table 2.3: Changes in Capability and Industrial Structure

	Country Capability $N_{i,t+\Delta}$		
	(1)	(2)	(3)
	OLS	IV ($Z^I_{i,t}$)	IV ($Z^I_{i,t}$ and $Z^{II}_{i,t}$)
Average Complexity $S_{i,t}$	0.00840**	0.368***	0.288***
	(0.00390)	(0.141)	(0.0902)
Initial Capability $N_{i,t}$	0.936***	0.831***	0.855***
	(0.0211)	(0.0468)	(0.0364)
Country and year FEs	Yes	Yes	Yes
Observations	6,872	6,872	6,872
R-squared	0.988	0.619	0.701
Clusters	1438	1438	1438
CD F-Stat		32.66	36.03
KP F-Stat		9.330	8.445

Notes: Table 2.3 reports estimates of β and ϕ in equation (2.17) using the baseline measures of complexity n^k_t and capability $N_{i,t}$ from the linear probability model estimation of equation (2.9.1). Columns 2 and 3 instrument average complexity $S_{i,t}$ by the WTO shocks $Z^I_{i,t}$ and $Z^{II}_{i,t}$ (in both cases using n^k_t calculated using the linear probability model). Standard errors clustered at the 5-year-period-country level.

Following the first stage discussion above, column 2 reports results using only the single instrument $Z^{I}_{i,t}$ that captures both country and product similarity in the WTO entrant's export mix, while column 3 allows these two dimensions to have different effects on the complexity of i's product mix by using both $Z^{I}_{i,t}$ and $Z^{II}_{i,t}$ as instruments.

Both of our IV specifications show positive and significant coefficient estimates on the complexity-weighted product mix, i.e. $\beta > 0$. Producing more complex goods raises a country's capability growth. The fact that the OLS estimate is much closer to zero is consistent with endogeneity concerns that bias β downwards, for example "good policies" such as investments in R&D and education that both raise capability growth (i.e. a country's propensity to export) and lead to greater specialization in more complex sectors (i.e. those that are less likely to be exported).

2.4.4 Sensitivity

Before using our estimates of dynamic spillovers to calculate the impact of trade on capability growth, we explore a range of alternative specifications and robustness checks. For reference purposes, column 1 of Table 2.9.5 and 2.9.6 repeat our baseline IV specification, instrumenting with the complexity of the industry mix with both WTO-entry instruments (column 3 of Table 2.9.4). We use the two-instrument IV strategy throughout.

Table 2.9.5 focuses on the sensitivity of our estimates to alternative data samples. Recall our baseline utilizes the raw Comtrade data for consistency and applies the basic cleaning procedures outlined in Feenstra et al. (2005) but to the full 1962-2014 timespan of our data. Column 2 reproduces our results using the actual Feenstra et al. (2005) dataset for years where available, and our more recent data with restrictions designed to mimic Feenstra et al. (2005). Unsurprisingly results are similar. Columns 3 to 5 consider alternative samples of countries. Column 3 expands our baseline sample of 146 countries to 200 countries by removing the restriction that we need to have observed a country for at least 40 years(and soincludes countries such as those created with the fall of the Soviet Union and those with spotty reporting). Columns 4 and 5 alter the restriction that a country must export a total of 100 million USD or more at some point in our sample (using 2010 US dollars and averaging annual exports over 5 year periods). Column 4 removes this restriction leaving us with 149 countries, while column 5 enlarges this threshold to 1 billion USD reducing the sample to countries. The coefficient on average complexity remains highly significant in all these cases, and rise substantially when restricting our sample to larger countries in the last column.

Table 2.9.6 first explores the sensitivity of our results to alternative lag structures. Col-

Table 2.4: Changes in Capability and Industrial Structure: Sensitivity (I)

	Country Capability $N_{i,t+\Delta}$				
	(1)	(2)	(3)	(4)	(5)
	Baseline	Feenstra Dataset	All Length Panels	No Size Threshold	High Size Threshold
Average Complexity $S_{i,t}$	0.288***	0.298**	0.223***	0.291***	0.414***
	(0.0902)	(0.127)	(0.0732)	(0.0901)	(0.149)
Initial Capability $N_{i,t}$	0.855***	0.929***	0.868***	0.857***	0.805***
	(0.0364)	(0.0416)	(0.0359)	(0.0354)	(0.0532)
Country and year FEs	Yes	Yes	Yes	Yes	Yes
Observations	6,872	6,864	7,905	6,995	5,986
R-squared	0.701	0.721	0.711	0.689	0.648
Clusters	1438	1438	1673	1466	1249
CD F-Stat	36.03	17.52	37.97	34.09	27.05
KP F-Stat	8.445	4.145	9.282	8.475	5.551

Notes: Table 2.4 reports estimates of β and ϕ in equation (2.17) using the baseline measures of complexity n_t^k and capability $N_{i,t}$ from the linear probability model estimation of equation (2.9.1). All columns use the two-instrument IV strategy. Column 1 reports our baseline estimates (column 3 of Table 2.9.4). Column 2 uses data from Feenstra et al. (2005) whenever possible. Column 3 expands our sample to include countries with fewer than 40 years of data. Column 4 removes the threshold value of total exports required to be included in our sample. Column 5 raises the threshold value of total exports required to be included in our sample from 100 million to 1 billion USD (at 2010 prices). Standard errors clustered at the 5-year-period-country level.

Table 2.5: Changes in Capability and Industrial Structure: Sensitivity (II)

	Country Capability $N_{i,t+\Delta}$				$GNI_{i,t+1}$
	(1)	(2)	(3)	(4)	(5)
	Baseline	10-year Lag	1 Obs. per 5-year Cluster	IV $N_{i,t}$	
Average Complexity $S_{i,t}$	0.288***	0.405***	0.205**	0.275***	0.906**
	(0.0902)	(0.144)	(0.0877)	(0.0955)	(0.417)
Initial Capability $N_{i,t}$	0.855***	0.690***	0.876***	0.721***	
	(0.0364)	(0.0651)	(0.0381)	(0.0981)	
GNI per capita $GNI_{i,t}$					0.758***
					(0.0330)
Country and Year FEs	Yes	Yes	Yes	Yes	Yes
Observations	6,872	6,151	1,295	6,195	6,107
R-squared	0.701	0.308	0.751	0.669	0.588
Clusters	1438	723	1295	1303	1269
CD F-Stat	36.03	35.85	7.177	12.98	63.55
KP F-Stat	8.445	8.733	5.094	3.674	16.70

Notes: Table 2.5 reports estimates of β and ϕ in equation (2.17) using the baseline measures of complexity n_t^k and capability $N_{i,t}$ from the linear probability model estimation of equation (2.9.1). All columns use the two-instrument IV strategy. Column 1 reports our baseline estimates (column 3 of Table 2.9.4). Column 2 reports the same estimates using 10-year lags. Column 3 uses one observation per 5-year cluster. Column 4 instruments initial capability using lagged-values of the WTO shocks $Z_{i,t}^{I}$ and $Z_{i,t}^{II}$. Column 5 uses GNI per capita instead of capability. Standard errors clustered at the 5-year-period-country level.

umn 2 considers a 10-year rather than 5-year lag (and the instruments use the export structure of the future entrant 10 years before entry). The dynamic spillovers become approximately one third larger over this extended time period. As an alternative to including all years of data and clustering standard errors at the country 5-year period level, column 3 only includes one observation from each cluster (observations from years ending in five or zero). The magnitude of the coefficient falls slightly but remains significant.

Although our 1962-2014 panel is relatively long, there may still be Nickel-bias concerns for the inclusion of a lagged dependent variable, as discussed in footnote 13. To address this issue, we use lags of our instruments as additional IVs. Through the lens of our model, those lagged variables are correlated with initial capability and are orthogonal to capability shocks under the same conditions as our non-lagged IVs. Column 4 reports results treating the initial level of capability as endogenous and additionally including 5-year lags of our two instruments in our instrument set. Reassuringly, the β coefficient on the complexity of the industry mix changes little, and the coefficient on the initial level of capability only falls by a small amount, suggesting Nickell-bias worries are limited.

Finally, column 5 replaces the capability of country i with its GDP per capita. As with capability, we find that the changes in the complexity of the industry mix (instrumented by shocks coming from WTO entrants) reduces future GDP per capita (conditioning on initial GDP per capita and both country and year fixed effects). This result is of independent interest, with the specification having close similarities to the growth regressions in Hausman et al. (2007) among others. We come back to this issue in Section 2.6.

2.5 Does Trade Push All Countries to the Top?

In Section 2.2.3, we have provided sufficient conditions under which international trade may raise capability in all countries. In Section 2.4, we have estimated positive dynamic positive spillovers in more complex sectors. In a pure ladder economy, this would lead to pervasive dynamic gains from trade. We now explore whether the same result holds in a less stylized environment that is flexible enough to match the pattern of trade flows observed in the data.

2.5.1 Baseline Economy

Throughout this section, we maintain the functional form assumptions imposed on preferences in Section 2.2 as well as those imposed on technology in Sections 2.3 and 2.4. Time is discrete, with each period t corresponding to a year. In addition to the manufacturing sectors analyzed in our empirical analysis, we add a non-tradable sector that enter preferences in a Cobb-Douglas fashion and allow for trade imbalances in the form of exogenous lump-sum transfers across countries.

On the demand side, preferences are nested CES with elasticity of substitution, ϵ and σ, as described in equations (2.1) and (2.2). We set $\epsilon = 1.36$ in line with the elasticity of substitution between 4-digit sectors in Redding and Weinstein (2018) and $\sigma = 2.7$ in line with the median elasticity of substitution between foreign varieties in Broda and Weinstein (2006), also estimated at the 4-digit level.[17]

On the supply side, we need to specify the labor endowments, L_{it}, as well as the productivity draws, $A^k_{ij,t}$. For each country i and each year t, we choose units so that wages per efficiency units are equal to one, $w_{i,t} = 1$. Under this normalization, L_{it} is equal to the total sales, across all destinations and sectors, from country i at date t. We then set the realization of the productivity draws $A^k_{ij,t}$ so that the baseline economy matches trade data. As demonstrated in Appendix 2.8.3, given estimates of ϵ and σ as well as data on bilateral trade flows, $x^k_{ij,t}$, productivity levels $A^k_{ij,t}$ are exactly identified, up to a time-and-destination productivity shifter,

$$\frac{A^k_{ij,t}}{A^1_{jj,t}} = \left(\frac{x^k_{ij,t}}{x^1_{jj,t}}\right)^{\frac{1}{\sigma-1}} \left[\frac{\sum_i x^k_{ij,t}}{\sum_i x^1_{ij,t}}\right]^{\frac{(\epsilon-\sigma)}{(\sigma-1)(1-\epsilon)}} \equiv \hat{A}^k_{ij,t}. \tag{2.19}$$

In what follows, we set $A^1_{jj,t}$ to one, both in the autarkic and trade equilibria, for all j and t. This affects the level of real consumption $C_{j,t}$ in both the autarkic and trade equilibria, but not the proportional changes between the two, which is what we are interested in.

We follow a similar approach for dynamic considerations. We assume that the law of motion for capability is an AR1, with persistence ϕ, that depends on the average complexity of a country's output mix, with β controlling the magnitude of dynamic spillovers, as described in equation (2.17). We use $\beta = 0.739$, as reported in column 3 of Table 2.3, and $\phi = (0.575)^{1/5} = 0.838$, which is the one-year counterpart of the 5-year coefficient reported in column 3 of Table 2.3. We then set the capability shocks $\varepsilon_{i,t}$ so that conditional on the measure of complexity estimated in Section 2.3 as well as the country and time fixed effects, γ_i and δ_t, estimated in Section 2.4.3, the baseline economy perfectly matches

[17] We are not aware of other estimates of the elasticity of substitution between sectors at the 4-digit level. At the 2-digit level, Oberfield and Raval (2014) report estimates of the elasticity of substitution between sectors centered around one, whereas the preferred estimate of Bartelme et al. (2019a) is 1.47. For higher levels of aggregation, for instance agriculture, manufacturing, and services, Herrendorf et al. (2013) estimate a lower elasticity, as one would expect, around 0.9. As already mentioned in Section 2.4.3, our first stage coefficients point towards an elasticity of substitution between sectors ϵ that is greater than one.

Parameter	Value	Choice Calibration
Panel A: Nested CES Preferences		
σ	2.7	Broda and Weinstein (2006)
ϵ	1.36	Redding and Weinstein (2018)
Panel B: Dynamic Spillovers		
β	0.288	Baseline estimate (Table 2.3, Column 3)
ϕ	0.855	Baseline estimate (Table 2.3, Column 3, yearly adjusted)

Table 2.6: Baseline Economy

the path of capability $N_{i,t}$ estimated in Section 2.3,

$$\varepsilon_{i,t} = N_{i,t+\Delta} - \beta S_{i,t} - \phi N_{i,t} - \gamma_i - \delta_t.$$

Table 2.6 reports the values of the main structural parameters used in our baseline economy.

2.5.2 Construction of the Counterfactual Autarkic Equilibrium

To quantify the static and dynamic effects of trade for development, we return to the counterfactual question of Section 2.2.3: If a country were to go back to autarky from 1962 onwards, what would be the consequences for its capability and welfare?

For each country i and each year t from 1962 to 2014, we construct the counterfactual autarkic equilibrium as follows. In 1962, we start by setting the counterfactual autarkic capability to the value observed in the initial equilibrium, $(N_{i,1962})' = N_{i,1962}$. We then proceed iteratively. In any year $t \geq 1962$, given the counterfactual autarkic capability $(N_{i,t})'$ and the observed measures of complexity n_t^k, we first determine the set of goods with strictly positive probability under autarky, $(A_{ii,t}^k)' > 0$.

Consistent with the empirical analysis of Sections 2.3 and 2.4, we assume that a country i is able to produce good k for its domestic market at date t (under autarky) if and only if it would have been able to export it to at least one foreign market (under trade), with the probability of exporting to any individual foreign market given by equation (2.15), XXX TO FIX XXX

$$\text{Prob}((A_{ii,t}^k)' > 0) = 1 - \left[1 - \frac{e^{((N_{i,t})' - n_t^k)}}{1 + e^{((N_{i,t})' - n_t^k)}}\right]^{M-1},$$

where $M = 96$ is the total number of countries used in our counterfactual exercise.[18]

If good k has strictly positive productivity in both the trade and autarkic equilibria, we set $(A^k_{ii,t})' = A^k_{ii,t}$. If good k has strictly positive productivity only in the autarkic equilibrium, then we randomly draw $(A^k_{ii,t})'$ from a log-normal distribution whose mean is equal to the sum of the country-time and sector-time fixed effects, $A_{i,t}$ and A^k_t, estimated from the following log-linear regression

$$\ln A^k_{ii,t} = A_{i,t} + A^k_t + \alpha^k_{i,t},$$

and whose standard deviation is equal to the standard deviation of the estimated residuals. For all destinations $j \neq i$, we then set $(A^k_{ij,t})' = 0$.

Given the set of autarky productivity draws $(A^k_{ij,t})'$ and using country i's labor as our numeraire, $(w_{i,t})' = 1$, we can then use equations (2.6), (2.7), (2.8), (2.9), and (2.10) to solve for all autarky prices and quantities at date t in country i: $(p^k_{ii,t})'$, $(c^k_{ii,t})'$, and $(l^k_{ii,t})'$.[19] Once the counterfactual employment distribution $(F^{\ell}_{i,t})'$ is known, the counterfactual autarkic capability at date $t+1$ can be computed using equation (2.17),

$$(N_{i,t+1})' = \gamma_i + \delta_t + \beta \int n d(F^{\ell}_{i,t})'(n) + \phi(N_{i,t})' + \varepsilon_{i,t},$$

where γ_i, δ_t, and $\varepsilon_{i,t}$ are at the same value as in the baseline economy.

2.5.3 Static and Dynamic Consequences of International Trade

For each year $t \geq 1962$, we define the gains from trade for country i at that date as

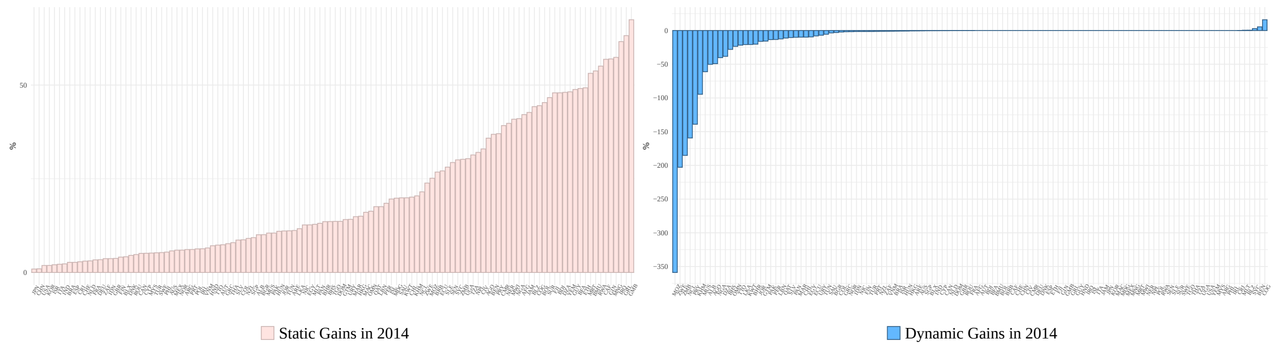
$$GT_{i,t} = 1 - \frac{(C_{i,t})'}{C_{i,t}}, \tag{2.20}$$

where $C_{i,t}$ and $(C_{i,t})'$ are the aggregate consumption levels in the original trade equilibrium and the counterfactual autarkic equilibrium.[20] The values of the structural parameters in the trade equilibrium are those described in Section 2.5.1, with productivity levels set to $\hat{A}^k_{ij,t}$ for all goods with positive productivity. By construction, the original trade

[18] A number of small countries do not report trade data in the final years of our sample, hence the downward adjustment from 138 to 96 countries.

[19] Given good market clearing, the labor market clearing condition (2.11) in country i necessarily holds, an expression of Walras' Law.

[20] Like in Section 2.2.3, this corresponds to the difference between the income level required to achieve the utility under free trade (at date t) and the income level required to achieve the utility under autarky (at that same date t), both evaluated at the free trade prices and expressed as a fraction of a country i's income level under free trade.



(a) Static Gains from Trade **(b)** Dynamic Gains from Trade

Figure 2-5: Welfare Consequences of International Trade

Notes: Figure 2-5a reports the static gains from trade, $GT_{i,t}$, as described in equation (2.21), for all countries in our sample in 2014. Figure 2-5b reports the dynamic gains from trade, $GT^D_{i,t}$, as described in equation (2.22), for the same countries and year.

equilibrium matches bilateral trade flows for all countries and sectors at date t. Aggregate consumption under trade can be computed using $c^k_{ji} = x^k_{ji,t}/(w_{j,t}/\hat{A}^k_{ji,t})$ and substituting into equations (2.1) and (2.2). Aggregate consumption in the autarkic equilibrium can be computed in a similar manner using the counterfactual consumption levels $(c^k_{ii,t})'$ from Section 2.5.2.

To decompose the gains from trade into a static and dynamic component, we consider a second counterfactual autarkic equilibrium in which we also set $(A^k_{ij,t})'' = 0$ for all destinations $j \neq i$, but we keep capability at the same level as in the original trade equilibrium, $(N_{i,t})'' = N_{i,t}$ for all t, and all goods produced in the original trade equilibrium remain produced with the same productivity, $(A^k_{ii,t})'' = \hat{A}^k_{ii,t}$. The static gains from trade at date t then correspond to

$$GT^S_{i,t} = 1 - \frac{(C_{i,t})''}{C_{i,t}}, \tag{2.21}$$

where $(C_{i,t})''$ denotes the aggregate consumption level associated with that second autarkic equilibrium. The dynamic gains from trade, in turn, are defined as the difference between the total gains from trade and the static component,

$$GT^D_{i,t} = GT_{i,t} - GT^S_{i,t}. \tag{2.22}$$

For expositional purposes, we focus on 2014, the last year of our sample. Figure 2-5 reports the static and dynamic gains from trade across countries. In Figure 2-5a, we see that the static gains from trade are positive for all countries, as our perfectly competi-

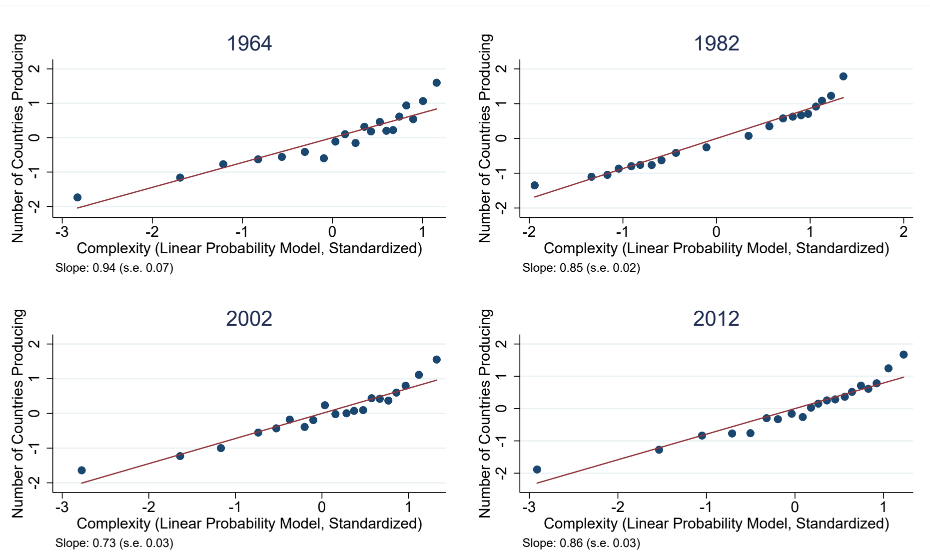


Figure 2-6: Complexity vs. Foreign Competition

tive model necessarily predicts, and large. This derives both from the low elasticity of substitutions used in our calibration—which tends to make domestic and foreign labor services very imperfect substitutes—and from the fact that many countries in our sample have very little domestic production in manufacturing sectors. These were already the two considerations shaping up the magnitude of the static gains from trade in the pure ladder economy, as established in Proposition 12.In contrast, we see in Figure 2-5b that the dynamic gains from trade are either negative or zero for most of the countries in our sample. Despite our estimates of positive dynamic spillovers in more complex sectors, we are very far from the qualitative predictions derived in the case of the pure ladder economy.

Figure 2-6 explains why. In sharp contrast to the assumptions imposed in the pure ladder economy, more complex goods tend to be produced by *more* countries. Since more complex sectors face more foreign competition, they shrink relative to other sectors in the trade equilibrium. And since we have identified those sectors as the source of dynamic spillovers, opening up to trade tends to lower capability around the world.

2.6 Robustness

2.6.1 Alternative Measures of Complexity and Capability

A natural concern with the previous analysis is that there is some arbitrariness in how we have defined our measures of complexity and capability. What if we had made different functional-form assumptions about how complexity and capability map into zero trade flows, could we have reached very different conclusions?

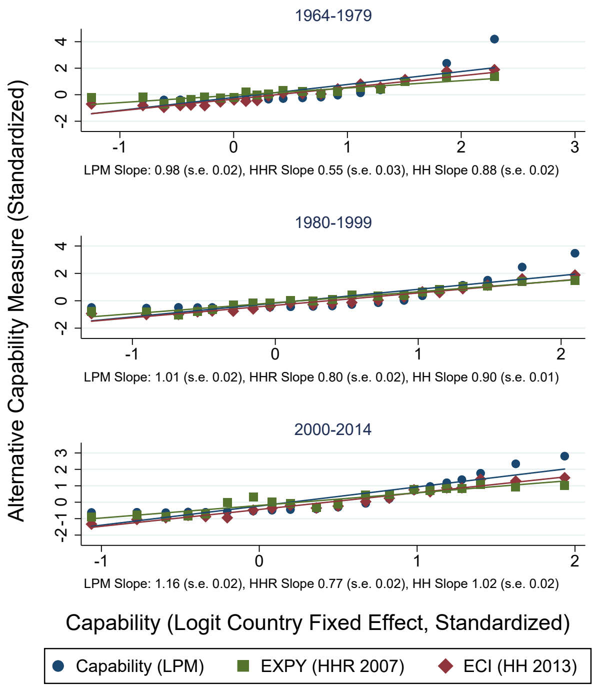
In the pure ladder economy, we have shown that dynamic gains from trade arise because countries face less foreign competition in the more complex sectors that generate positive dynamic positive spillovers. A more direct test of this hypothesis would start by identifying complex sectors as those that generate more spillovers and then check whether or not they face less foreign competition or, conversely, by identifying complex sectors as those that face less foreign competition and then check whether or they generate more spillovers. We now follow this second route.

Alternative Measures. Instead of equation 2.15, we now assume that the productivity distribution G_t is such that,

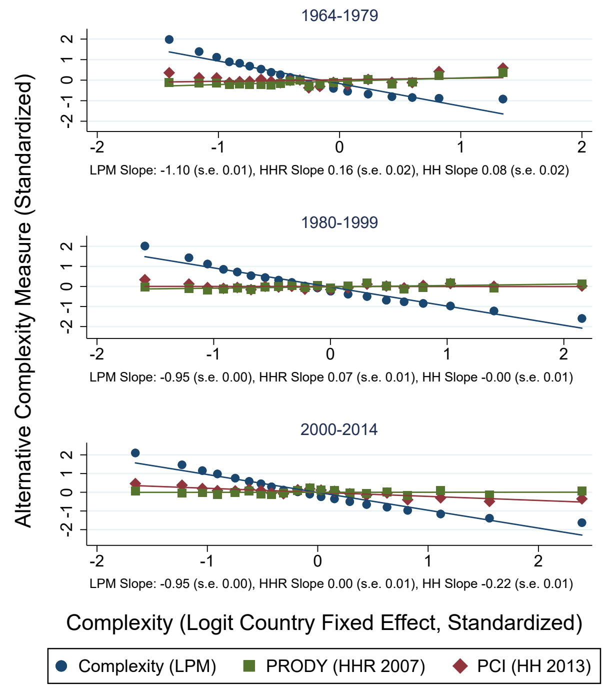
$$\text{Prob}(A^k_{ij,t} > 0) = \frac{e^{(N_{i,t} - n^k_t)}}{1 + e^{(N_{i,t} - n^k_t)}}, \text{ for all } i \neq j, k, \text{ and } t, \tag{2.23}$$

with independence across origins, destinations, and sectors. Consistent with the pure ladder economy presented in Section 2.2.3, this standard logistic function implies that the probability of a strictly positive productivity draw and hence a non-zero trade flow is: (i) increasing with capability, (ii) decreasing in complexity, and (iii) log-supermodular in both. By construction, more complex goods are therefore those that tend to face less foreign competition, since less countries, on average, are able to export them.

Separately for each year t in our sample, we can recover estimates of $N_{i,t}$ and n^k_t by fitting a logit model via maximum likelihood where the binary response is whether variable country i exports good k to country j, and the explanatory variables are sets of country and good fixed effects. Since sets of fixed effects are only identified up to a constant, we normalize US capability to 0 in every period (i.e. it the US fixed effect is omitted in the logit specification). Figure 2.9.5 plots these alternative measures of capability and complexity against those already presented in Section 2.3.5. While all four capability measures are positively correlated, this is not the case for the complexity measure. As can be seen from Figure 2-7b, our new logit estimate, which infers complexity from the number of



(a) Capability Measures (N_{it})



(b) Complexity Measures (n_t^k)

Figure 2-7: Alternative Measures of Capability and Complexity

countries that export in a particular sectors, is strongly negatively correlated with our baseline measure and, at least in recent decades, with the PCI index of Hausman et al. (2013).

Pushed to the Top or Held at the Bottom? For our purposes, the interesting question, however, is whether using these alternative measures of complexity and capability would affect our conclusions about the consequences of international trade. To address it, we reestimate dynamic spillovers using equation 2.17 and our two WTO instruments (recalculated using our new measures of $N_{i,t}$ and n_t^k). Our results are reported in Appendix 2.9.5. Not surprisingly, since baseline and alternative capability measures are positively correlated whereas baseline and alternative complexity measures are negatively correlated, we now find evidence of negative dynamic spillovers, with $\beta = 0.739$ in our preferred two-instrument specification (Table 2.9.4, Column 3).[21]

Since more complex sectors that face less foreign now generate negative dynamic

[21] Appendix 2.9.5 also repeats the same set of robustness exercises examined in Section 2.4.4 for the logit specification (equation (2.23)).

spillovers, we end up with the same general conclusion as in Section 2.5: pervasive dynamic losses, as illustrated in Figure 2-8.

2.6.2 Complexity and Foreign Competition

We conclude our sensitivity analysis by exploring the extent to which other assumptions about the nature of foreign competition and dynamic spillovers might have affected our conclusions about the consequences of international trade. Unless stated otherwise, the calibrated parameters are those described in Section 2.5.1.

Complexity and Foreign Competition. Our baseline analysis rules out any heterogeneity across sectors in the lower-level elasticity of substitution: $\sigma^k = \sigma$. This implies that how many countries are able to produce a good determines the extent of foreign competition in that sectors. In practice, variations in σ^k may also affect the extent to which foreign competition shifts labor demand across sectors. If sectors that are less complex tend to be those with lower elasticities, then opening up to trade may not move countries away from those sectors, potentially reverting our welfare conclusions.

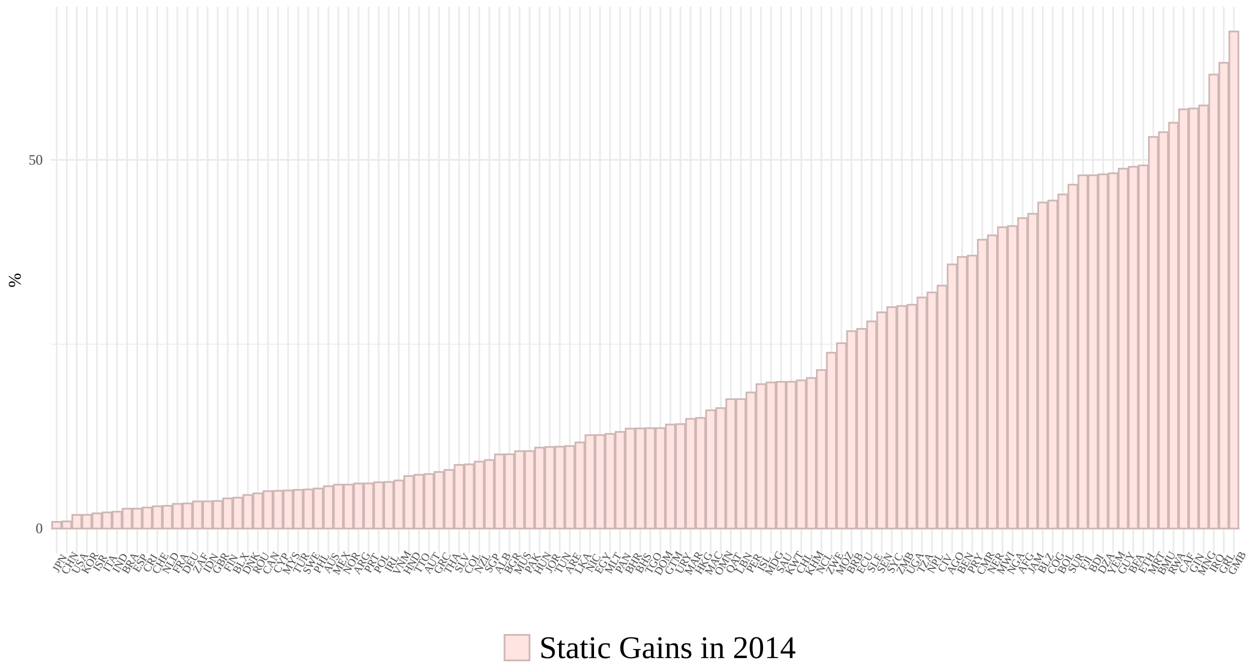
A simple way to assess whether this channel might be important is to plot the estimates of σ^k in Broda and Weinstein (2006) against our baseline estimates of complexity n_t^k. TBD.

2.7 Concluding Remarks

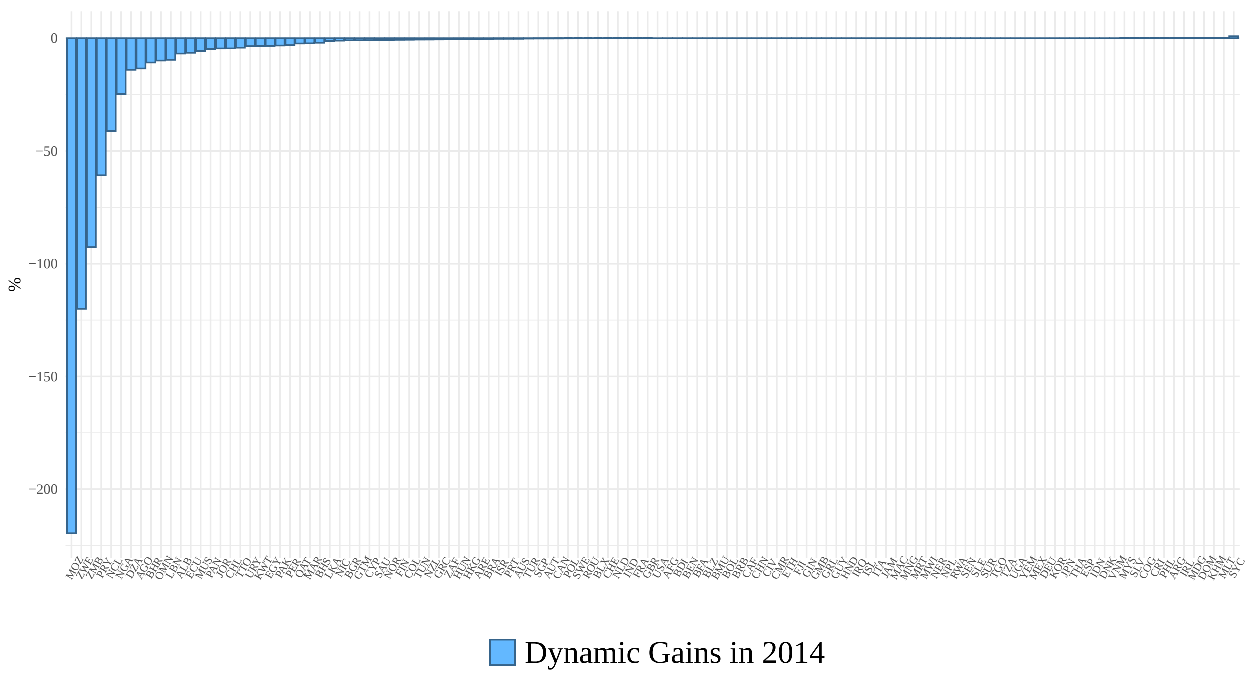
Motivated by the metaphor of a development ladder, we have developed a dynamic trade model in which countries differ in their capability, goods differ in their complexity, and capability growth is a function of the average complexity of the goods that each country produces. Two insights have emerged from our analysis.

First, on the theory side, we have demonstrated that the dynamic gains from trade do not have to be zero sum, with some countries specializing in the "good" sectors that are conducive to growth and others specializing in the "bad" sectors that are not. Instead, upon opening up to trade, all countries may move towards more complex sectors that face less foreign competition. And if those sectors create positive dynamic spillovers, all countries may gain.

Second, on the empirical side, we have demonstrated that the conditions required for pervasive dynamic gains do not appear to be satisfied. Using the entry of other countries into the WTO as exogenous shifter of countries' industry mix, we have shown that



(a) Static Gains from Trade



(b) Dynamic Gains from Trade

Figure 2-8: Welfare Consequences of International Trade (Alternative complexity and capability measures)

Notes: Figure 2-8 is the counterpart of Figure 2-5 when capability and complexity are estimated using equation (2.23).

more complex sectors that face less foreign competition tend to create negative dynamic spillovers. As a result, rather than pushing countries up the development ladder, opening up to international trade tends to hold most of them back.

2.8 Theoretical Appendix

2.8.1 Existence and Uniqueness of a Competitive Equilibrium

By equations (2.6)-(2.11), the equilibrium wages $\{w_{i,t}\}$ solve

$$L_{i,t} = \int_n \int_a \sum_j \frac{(a_{ij})^{\sigma-1}(w_{i,t})^{-\sigma}}{(\sum_l (w_{l,t}/a_{lj})^{1-\sigma})^{\frac{\epsilon-\sigma}{1-\sigma}}} \frac{w_{j,t}L_{j,t}dG_t(a|n,N_t)dF_t(n)}{\int_{n'}\int_{a'}(\sum_l (w_{l,t}/a'_{lj})^{1-\sigma})^{\frac{1-\epsilon}{1-\sigma}}dG_t(a'|n',N_t)dF_t(n')}. \tag{2.8.1}$$

for each i and t. Below we first establish the existence and uniqueness of $\{w_{i,t}\}$ that solve (2.8.1). The existence and uniqueness of $\{p^k_{ij,t}\}$, $\{c^k_{ij,t}\}$, and $\{\ell^k_{ij,t}\}$ directly follow from equations 2.6-(2.10). We conclude by characterizing the smoothness conditions on F_t, G_t, and H_t required for the existence and uniqueness of $\{N_{i,t}\}$ that solve (2.4).

Existence and uniqueness of $\{w_{i,t}\}$. Define the excess labor demand function,

$$z_{i,t}(\mathbf{w}_t) \equiv \int_n \int_a \sum_j \frac{(a_{ij})^{\sigma-1}(w_{i,t})^{-\sigma}}{(\sum_l (w_{l,t}/a_{lj})^{1-\sigma})^{\frac{\epsilon-\sigma}{1-\sigma}}} \frac{w_{j,t}L_{j,t}dG_t(a|n,N_t)dF_t(n)}{\int_{n'}\int_{a'}(\sum_l (w_{l,t}/a'_{lj})^{1-\sigma})^{\frac{1-\epsilon}{1-\sigma}}dG_t(a'|n',N_t)dF_t(n')} - L_{i,t},$$

where $\mathbf{w}_t \equiv [w_{i,t}]_i$, and $\mathbf{z}_t \equiv [z_{i,t}]_i$. Then $z_{i,t}$ is continuous and homogenous of degree zero, $\mathbf{w} \cdot \mathbf{z}(\mathbf{w}) = 0$ for all $\mathbf{w}$, and $\max_i\{z_{i,t}\} \to \infty$ as $w_{l,t} \to 0$ for some l. Therefore assumptions for Proposition 17.C.1 in Mas-Colell et al. (1995) are satisfied. This establishes the existence of $\{w_{i,t}\}$ that solve (2.8.1).

Next, let us show that $z_{i,t}(\mathbf{w}_t)$ satisfies gross-substitute properties, $\frac{\partial z_{i,t}(\mathbf{w}_t)}{\partial w_{l,t}} > 0$ for $i \neq l$. Let

$$\begin{aligned} L_{i,t}(\mathbf{w}_t) &\equiv \frac{1}{w_{i,t}} \int_n \int_a \sum_j \frac{(w_{i,t}/a_{ij})^{1-\sigma}}{(\sum_l (w_{l,t}/a_{lj})^{1-\sigma})^{\frac{\epsilon-\sigma}{1-\sigma}}} \frac{w_{j,t}L_{j,t}dG_t(a|n,N_t)dF_t(n)}{\int_{n'}\int_{a'}(\sum_l (w_{l,t}/a'_{lj})^{1-\sigma})^{\frac{1-\epsilon}{1-\sigma}}dG_t(a'|n',N_t)dF_t(n')} \\ &= \frac{1}{w_{i,t}} \int_n \int_a \sum_j \lambda^n_{ij,t}(\mathbf{w}_t)\Lambda^n_{j,t}(\mathbf{w}_t) w_{j,t}L_{j,t}dG_t(a|n,N_t)dF_t(n), \end{aligned}$$

with

$$\lambda_{ij,t}(a;\mathbf{w}_t) \equiv \frac{(w_{i,t}/a_{ij})^{1-\sigma}}{\sum_l (w_{l,t}/a_{lj})^{1-\sigma}},$$

$$\Lambda_{j,t}(a;\mathbf{w}_t) \equiv \frac{(\sum_l (w_{l,t}/a_{lj})^{1-\sigma})^{\frac{1-\epsilon}{1-\sigma}}}{\int_{n'}\int_{a'}(\sum_l (w_{l,t}/a'_{lj})^{1-\sigma})^{\frac{1-\epsilon}{1-\sigma}}dG_t(a'|n',N_t)dF_t(n')}.$$

Taking log-derivative for $l \neq i$,

$$
\begin{aligned}
\frac{\partial \ln L_{i,t}(\mathbf{w}_t)}{\partial \ln w_{l,t}} = \frac{1}{w_{i,t} L_{i,t}} \int_n \int_a \Bigg(& \lambda_{il,t}(a;\mathbf{w}_t)\Lambda_{l,t}(a;\mathbf{w}_t) w_{l,t} L_{l,t} \\
& + \sum_j \lambda_{ij,t}(a;\mathbf{w}_t)\Lambda_{j,t}(a;\mathbf{w}_t) w_{j,t} L_{j,t} \left[(\sigma - \epsilon)\lambda_{lj,t}(a;\mathbf{w}_t) \right. \\
& \left. + (\epsilon - 1) \int_{n'} \int_{a'} \lambda_{lj,t}(a';\mathbf{w}_t)\Lambda_j(a';\mathbf{w}_t) dG_t(a'|n', N_t) dF_t(n') \right] \Bigg) dG_t(a|n, N_t) dF_t(n),
\end{aligned}
$$

which is strictly positive under the assumptions that $\sigma > \epsilon > 1$. This shows that $L_{i,t}(\mathbf{w}_t)$ is strictly increasing in $w_{k,t}$ for $k \neq i$, implying that $z_{i,t}(\mathbf{w}_t)$ satisfies gross-substitute property. Applying Proposition 17.F.3 in Mas-Colell et al. (1995), the equilibrium wages $\{w_{i,t}\}$ are unique.

Existence and Uniqueness of $\{N_{i,t}\}$. Let $\mathbf{w}_t(N_t)$ denote the unique equilibrium vector of wages in period t as a function of the capability vector $N_t \equiv \{N_{i,t}\}$ and let $F^{\ell}_{i,t}(\cdot; N_t)$ denote the associated equilibrium distribution of employment across sectors of different complexity,

$$
F^{\ell}_{i,t}(n; N_t) = \frac{\int_{n' \leq n} \int_a \sum_j \frac{(w_{i,t}(N_t)/a_{ij})^{1-\sigma}}{(\sum_l (w_{l,t}(N_t)/a_{lj})^{1-\sigma})^{\frac{\epsilon-\sigma}{1-\sigma}}} \frac{w_{j,t}(N_t) L_{j,t} dG_t(a|n', N_t) dF_t(n')}{\int_{n''} \int_{a'} (\sum_l (w_{l,t}(N_t)/a'_{lj})^{1-\sigma})^{\frac{1-\epsilon}{1-\sigma}} dG_t(a'|n'', N_t) dF_t(n'')}}{\int_{n'} \int_a \sum_j \frac{(w_{i,t}(N_t)/a_{ij})^{1-\sigma}}{(\sum_l (w_{l,t}(N_t)/a_{lj})^{1-\sigma})^{\frac{\epsilon-\sigma}{1-\sigma}}} \frac{w_{j,t}(N_t) L_{j,t} dG_t(a|n', N_t) dF_t(n')}{\int_{n''} \int_{a'} (\sum_l (w_{l,t}(N_t)/a'_{lj})^{1-\sigma})^{\frac{1-\epsilon}{1-\sigma}} dG_t(a'|n'', N_t) dF_t(n'')}} \text{ for all } n. \tag{2.8.2}
$$

By equations (2.4) and (2.5), the equilibrium capability vector N_t solves the following ODE,

$$
\dot{N} = V(N, t), \tag{2.8.3}
$$

where $V : \mathbb{R}^I \times \mathbb{R} \to \mathbb{R}^I$ is such that for any $i = 1, ..., I$,

$$
V_i(N, t) \equiv H_{i,t}(N_i, F^{\ell}_{i,t}(\cdot; N)). \tag{2.8.4}
$$

Existence and uniqueness of $\{N_t\}$ follow from the conditions of Picard Theorem being satisfied. That is, for any $N_0 = \{N_{i,0}\}$ and any finite time horizon T, there exists a unique solution $\{N_t\}$ to (2.8.3) for $t \in [0, T]$ with initial value N_0 provided that F_t, G_t and $H_{i,t}$ are such that V is Lipschitz-continuous with respect to N and continuous with respect to t.

2.8.2 Comparative Statics

Proof of Proposition 11

We consider a country i that moves from trade to autarky at date 0. We first demonstrate that any date $t \geq 0$, country i's capability must be lower in the autarky equilibrium than what it would have been in the trade equilibrium if there are positive dynamic spillovers, whereas country i's capability must be higher in the autarky equilibrium if there are negative dynamic spillovers. We then conclude that at any date $t \geq 0$, country i's aggregate consumption in the autarky equilibrium must be lower if there are positive dynamic spillovers and may be higher if there are negative dynamic spillovers.

Change in Capability. Suppose first that there are positive dynamic spillovers. Let $N_{i,t}^A$ and $N_{i,t}$ denote the capability of country i at date t in the autarky and trade equilibrium, respectively. At date 0, we know that $N_{i,0}^A = N_{i,0}$. To show that $N_{i,t}^A \leq N_{i,t}$ for all $t \geq 0$, it is therefore sufficient to show that if $N_{i,t_0}^A = N_{i,t_0}$ at any date $t_0 \geq 0$, then $\dot{N}_{i,t_0} \geq \dot{N}_{i,t_0}^A$. By equation (2.4), under the assumption that $H_{i,t}$ is increasing in $F_{i,t}^\ell$, this is equivalent to show that if $N_{i,t_0}^A = N_{i,t_0}$, then F_{i,t_0}^ℓ first-order stochastically dominates $F_{i,t_0}^{\ell,A}$.

Take a date t_0 such that $N_{i,t_0} = N_{i,t_0}^A = N_i$. The density of employment in sectors of complexity n in country i at date t_0 in the autarky equilibrium is

$$f_{i,t_0}^{\ell,A}(n) = \begin{cases} f_{t_0}(n) & \text{, for all } n \leq N_i, \\ 0 & \text{otherwise.} \end{cases}$$

The same density in the trade equilibrium is

$$f_{i,t_0}^{\ell}(n) = \begin{cases} \dfrac{\sum_j \frac{(A_{ij,t_0})^{\sigma-1}(w_{i,t_0})^{-\sigma}}{(\Sigma_{l:N_{l,t_0}\geq n}(w_{l,t_0}/A_{lj,t_0})^{1-\sigma})^{\frac{\epsilon-\sigma}{1-\sigma}}} \frac{w_{j,t_0}L_{j,t_0}}{\int (P_{j,t_0}(m))^{1-\epsilon} dF_{t_0}(m)} f_{t_0}(n)}{\int \sum_j \frac{(A_{ij,t_0})^{\sigma-1}(w_{i,t_0})^{-\sigma}}{(\Sigma_{l:N_{l,t_0}\geq n'}(w_{l,t_0}/A_{lj,t_0})^{1-\sigma})^{\frac{\epsilon-\sigma}{1-\sigma}}} \frac{w_{j,t}L_{j,t_0}}{\int (P_{j,t_0}(m))^{1-\epsilon} dF_{t_0}(m)} dF_{t_0}(n')} & \text{, for all } n \leq N_i \\ 0 & \text{, otherwise.} \end{cases}$$

Now take $n_1 \leq n_2 \leq N_i$. Since $\sigma > \epsilon > 1$, we must have

$$(\sum_{l:N_{l,t_0}\geq n_2} (w_{l,t_0}/A_{lj,t_0})^{1-\sigma})^{\frac{\epsilon-\sigma}{1-\sigma}} \leq (\sum_{l:N_{l,t_0}\geq n_1} (w_{l,t_0}/A_{lj,t_0})^{1-\sigma})^{\frac{\epsilon-\sigma}{1-\sigma}} \text{ for all } j.$$

This implies

$$\frac{f_{i,t_0}^\ell(n_2)}{f_{i,t_0}^\ell(n_1)} \geq \frac{f_{t_0}(n_2)}{f_{t_0}(n_1)} = \frac{f_{i,t_0}^{\ell,A}(n_2)}{f_{i,t_0}^{\ell,A}(n_1)}.$$

Hence, F_{i,t_0}^ℓ dominates the distribution of employment in country i under autarky, $F_{i,t_0}^{\ell,A}$, in terms of the Monotone Likelihood Ratio Property (MLRP). It follows that $\dot{N}_{i,t_0}^A \leq \dot{N}_{i,t_0}$ and, in turn, that $N_{i,t}^A \leq N_{i,t}$ for all $t \geq 0$.

Suppose instead that there are negative dynamic spillovers. Under the assumption that $H_{i,t}$ is decreasing in $F_{i,t}^\ell$, the previous argument implies that if $N_{i,t_0}^A = N_{i,t_0}$ at any date $t_0 \geq 0$, then $\dot{N}_{i,t_0} \leq \dot{N}_{i,t_0}^A$ and, in turn, that $N_{i,t}^A \geq N_{i,t}$ for all $t \geq 0$.

Change in Aggregate Consumption. Suppose first that there are positive dynamic spillovers. Let $C_{i,t}^A$ and $C_{i,t}$ denote aggregate consumption in country i at date t in the autarky and trade equilibrium, respectively, and let $\bar{C}_{i,t}^A$ denote aggregate consumption in country i at date t in a hypothetical autarky equilibrium where capability levels remain fixed at their trade equilibrium values, $N_{i,t}$, at all dates. For fixed capability levels $N_{i,t}$, our economy features a representative agent, perfect competition, and no distortion. Hence, standard arguments (e.g. Samuelson, 1939) imply $\bar{C}_{i,t}^A \le C_{i,t}$. Since $N_{i,t}^A \le N_{i,t}$ for all t, we must also have $C_{i,t}^A \le \bar{C}_{i,t}^A$. Combining the two previous observations, we get $C_{i,t}^A \le C_{i,t}$ for all $t \ge 0$.

Suppose instead that there are negative dynamic spillovers. In this case, the previous arguments imply $\bar{C}_{i,t}^A \le C_{i,t}$, but $C_{i,t}^A \ge \bar{C}_{i,t}^A$. Lower aggregate consumption arises if dynamic considerations dominate static ones (e.g. if there are no static gains from trade, which arise for a country with the highest capability at all dates if $\sigma \to \infty$).

Proof of Proposition 12

Suppose first that there are positive dynamic spillovers. Let N_i^A and N_i denote the capability of country i in the autarky and trade steady state, respectively. From the proof of Proposition 11, we already know that $N_{i,t}^A \le N_i$ for all t. This implies $C_{i,t}^A \le \bar{C}_i^A$, where $\bar{C}_i^A$ denotes aggregate consumption under autarky if country i's capability had remained at its trade steady state value, N_i. We can therefore compute a lower-bound on the cost of autarky, and hence the gains from trade, as

$$\underline{GT}_i = 1 - \frac{\bar{C}_i^A}{C_i}.$$

Since $N_{i,0}^A = N_i \ge N_i^A$, we must also have $N_{i,t}^A \ge N_i^A$ for all t. This implies $C_{i,t}^A \ge \underline{C}_i^A$, where $\underline{C}_i^A$ denotes aggregate consumption under autarky if country i's capability had jumped immediately to its autarky steady state value, N_i^A. We can therefore compute an upper-bound as

$$\bar{GT}_i = 1 - \frac{\underline{C}_i^A}{C_i}.$$

We now describe how to compute $\underline{GT}_i$ and $\bar{GT}_i$ using the same general strategy as in Costinot and Rodríguez-Clare (2013). Consider $\underline{GT}_i$ first. In the trade and autarky equilibria with identical capability N_i, budget balance in every period implies

$$C_i = w_i L_i / P_i, \tag{2.8.5}$$

$$\bar{C}_i^A = \bar{w}_i^A L_i / \bar{P}_i^A. \tag{2.8.6}$$

By equation (2.9), we also have

$$\frac{\bar{P}_i^A}{P_i} = \left[\int_{n \le N_i} \left(\frac{P_i(n)}{P_i}\right)^{1-\epsilon} \left(\frac{\bar{P}_i^A(n)}{P_i(n)}\right)^{1-\epsilon} dF(n)\right]^{\frac{1}{1-\epsilon}}.$$

Using the fact that $e_i(n) = (P_i(n)/P_i)^{1-\epsilon}$, $\bar{P}_i^A(n) = \bar{w}_i^A / A_{ii}$ for all $n \le N_i$, and $\lambda_{ii}(n) = (w_i/(A_{ii}P_i(n)))^{1-\sigma}$ for all $n \le N_i$ and zero otherwise, this can be rearranged as

$$\frac{\bar{P}_i^A}{P_i} = \frac{\bar{w}_i^A}{w_i}\left[\int e_i(n)(\lambda_{ii}(n))^{\frac{\epsilon-1}{\sigma-1}} dF(n)\right]^{\frac{1}{1-\epsilon}}.$$

Combining this expression with equations (2.8.5) and (2.8.6), we obtain

$$\underline{GT}_i = 1 - \left[\int_0^{N_i} e_i(n)(\lambda_{ii}(n))^{\frac{\epsilon-1}{\sigma-1}} dF(n)\right]^{\frac{1}{\epsilon-1}}.$$

Next, consider $\bar{GT}_i = 1 - \frac{\underline{C}_i^A}{C_i}$. As before, budget balance in every period implies

$$\underline{C}_i^A = \underline{w}_i^A L_i / \underline{P}_i^A,$$

whereas equations (2.6) and (2.9) imply

$$\frac{\bar{P}_i^A}{\underline{P}_i^A} = \frac{\bar{w}_i^A}{\underline{w}_i^A}\left(\frac{N_i}{N_i^A}\right)^{\frac{1}{1-\epsilon}}.$$

Noting that $\underline{C}_i^A / C_i = (\underline{C}_i^A / \bar{C}_i^A)(\bar{C}_i^A / C_i)$, we get

$$\bar{GT}_i = 1 - \left[\int_0^{N_i} e_i(n)(\lambda_{ii}(n))^{\frac{\epsilon-1}{\sigma-1}} dF(n)\right]^{\frac{1}{\epsilon-1}} \left[N_i / N_i^A\right]^{\frac{1}{1-\epsilon}}.$$

In the trade steady state, we know that $0 = H(N_i, F_i^\ell)$, so that $N_i = H_i^{-1}(0, F_i^\ell)$. In the autarky steady state, we also know from the proof of Proposition 11 that the employment distribution is equal to F, so that $N_i^A = H_i^{-1}(0, F)$. Substituting for N_i and N_i^As, we finally obtain

$$\bar{GT}_i = 1 - \left[\int e_i(n)(\lambda_{ii}(n))^{\frac{\epsilon-1}{\sigma-1}} dF(n)\right]^{\frac{1}{\epsilon-1}} \left[H_i^{-1}(0, F_i^\ell) / H_i^{-1}(0, F)\right]^{\frac{1}{(1-\epsilon)}}.$$

Suppose instead that there are negative dynamic spillovers. In this case, Proposition 11 implies $N_{i,t}^A \ge N_i$ for all t and, in turn, $\bar{C}_i^A \le C_{i,t}^A$. Thus, $\underline{GT}_i = 1 - \frac{\bar{C}_i^A}{C_i}$ provides an upper-bound on the gains from trade. Likewise, since $N_{i,0}^A = N_i \le N_i^A$, we must also have $N_{i,t}^A \le N_i^A$ for all t and, in turn, $C_{i,t}^A \le \underline{C}_i^A$. Thus, $\bar{GT}_i = 1 - \frac{\underline{C}_i^A}{C_i}$ provides a lower-bound on the gains from trade.

2.8.3 Identification of Productivity Draws

Let $x^k_{ij,t}$ denote the value of sales by country i to country j in sector k at date t. Equations (2.6) and (2.7) imply

$$\frac{x^k_{ij,t}}{x^1_{jj,t}} = \frac{(w_{i,t}/A^k_{ij,t})^{1-\sigma}}{(w_{j,t}/A^1_{jj,t})^{1-\sigma}} \frac{(P^k_{j,t})^{\sigma-\epsilon}}{(P^1_{j,t})^{\sigma-\epsilon}}. \tag{2.8.7}$$

Combined with equation (2.8), equations (2.6) and (2.7) further imply

$$\frac{\sum_i x^k_{ij,t}}{\sum_i x^1_{ij,t}} = \frac{(P^k_{j,t})^{1-\epsilon}}{(P^1_{j,t})^{1-\epsilon}}. \tag{2.8.8}$$

Using equation (2.8.8) to substitute for $P^k_{j,t}/P^1_{j,t}$ in equation (2.8.7), we obtain, after rearrangements,

$$\frac{A^k_{ij,t}}{A^1_{jj,t}} = \left(\frac{w_{i,t}}{w_{j,t}}\right) \left(\frac{x^k_{ij,t}}{x^1_{jj,t}}\right)^{\frac{1}{\sigma-1}} \left[\frac{\sum_l x^k_{lj,t}}{\sum_l x^1_{lj,t}}\right]^{\frac{(\sigma-\epsilon)}{(\sigma-1)(\epsilon-1)}}.$$

Under our choice of units of account, $w_{i,t} = 1$ for all i and t. Equation (2.19) follows.

2.9 Empirical Appendix

2.9.1 Sample of Countries

Table 2.9.1: Sample countries

Country Name	Years in Sample	Max Exports 5-yr Avg ($B)	Country Name	Years in Sample	Max Exports 5-yr Avg ($B)	Country Name	Years in Sample	Max Exports 5-yr Avg ($B)
Afghanistan	54	0.79	Ghana	54	9.67	Nigeria	54	92.07
Albania	54	1.89	Gibraltar	52	0.19	North Korea	54	2.82
Algeria	54	52.35	Greece	54	20.23	Norway	54	120.40
Angola	54	58.92	Greenland	48	0.80	Oman	54	39.25
Argentina	54	71.70	Guatemala	54	9.70	Pakistan	54	22.92
Australia	54	239.80	Guinea	54	2.14	Panama	54	6.09
Austria	54	143.00	Guinea-Bissau	54	0.25	Papua New Guinea	54	7.19
Bahamas	54	5.18	Guyana	54	1.36	Paraguay	54	5.95
Bahrain	54	6.03	Haiti	54	0.94	Peru	54	37.78
Bangladesh	44	26.46	Honduras	54	7.75	Philippines	54	69.80
Barbados	54	0.66	Hong Kong	54	77.39	Poland	54	160.40
Belgium-Luxembourg	54	316.10	Hungary	54	89.43	Portugal	54	51.90
Belize	51	1.13	Iceland	54	4.69	Qatar	54	94.11
Benin	54	1.03	India	54	215.60	Republic of the Congo	54	10.13
Bermuda	54	0.80	Indonesia	54	184.50	Romania	54	53.81
Bolivia	54	9.16	Iran	54	93.76	Rwanda	53	0.35
Brazil	54	229.60	Iraq	54	72.40	Saint Kitts and Nevis	52	0.40
Bulgaria	54	21.89	Ireland	54	155.50	Saudi Arabia	54	291.40
Burkina Faso	54	1.42	Israel	54	56.84	Senegal	54	1.69
Burma	54	10.98	Italy	54	438.40	Seychelles	47	0.46
Burundi	54	0.26	Jamaica	54	2.63	Sierra Leone	54	1.11
Cambodia	54	8.93	Japan	54	723.90	Singapore	54	169.20
Cameroon	54	4.72	Jordan	54	5.67	Somalia	54	0.49
Canada	54	413.40	Kenya	54	4.73	South Africa	54	124.10
Central African Repul	54	0.35	Kiribati	54	0.48	South Korea	54	465.40
Chad	54	2.90	Kuwait	54	68.84	Spain	54	246.40
Chile	54	73.64	Laos	54	3.14	Sri Lanka	54	9.47
China	54	2054.00	Lebanon	54	3.18	Sudan	54	9.91
Colombia	54	51.28	Liberia	54	2.90	Suriname	54	1.97
Costa Rica	54	31.16	Libya	54	41.53	Sweden	54	145.50
Cote d'Ivoire	54	9.32	Macau	53	3.81	Switzerland	54	293.00
Cuba	54	5.03	Madagascar	54	1.81	Syria	54	6.53
Cyprus	54	3.29	Malawi	53	1.05	Tanzania	54	3.98
Democratic Republic	54	6.92	Malaysia	54	236.40	Thailand	54	207.70
Denmark	54	87.86	Mali	54	1.95	Togo	54	1.70
Djibouti	54	0.18	Malta	54	4.05	Trinidad and Tobago	54	14.37
Dominican Republic	54	7.24	Mauritania	54	2.95	Tunisia	54	15.38
Ecuador	54	23.62	Mauritius	54	2.35	Turkey	54	119.60
Egypt	54	27.31	Mexico	54	345.90	Uganda	54	1.39
El Salvador	54	4.76	Mongolia	54	4.14	United Arab Emirates	50	178.60
Equatorial Guinea	54	10.99	Morocco	54	20.97	United Kingdom	54	399.00
Ethiopia	54	2.15	Mozambique	54	4.08	United States	54	1260.00
Falkland Islands	46	0.19	Nepal	54	0.79	Uruguay	54	9.19
Fiji	54	0.79	Netherland Antilles a	54	9.03	Venezuela	54	59.20
Finland	54	75.61	Netherlands	54	393.20	Vietnam	54	116.60
France	54	519.90	New Caledonia	54	1.44	Yemen	54	8.17
Gabon	54	9.73	New Zealand	54	35.41	Zambia	54	6.85
Gambia	54	0.25	Nicaragua	54	4.60	Zimbabwe	53	2.47
Germany	54	1227.00	Niger	54	0.72			

Notes: Table reports the 146 countries in our sample alongside the number of years of data and the maximum value of exports over any 5 year period 1962-2014 (in billions of 2010 US dollars).

2.9.2 Baseline Measures of Capability and Complexity: Construction

This appendix describes how we construct our baseline measures of capability $N_{i,t}$ and complexity n_t^k from the assumption that more capable countries are more likely to export more complex goods. As described in Section 2.3.1, we posit the following linear probability model:

$$\pi_{ij,t}^k = \delta_{ij,t} + \gamma_{j,t}^k + N_{i,t} n_t^k + \epsilon_{ij,t}^k \tag{2.9.1}$$

where $\pi_{ij,t}^k$ is a dummy variable that takes the value 1 if positive exports of good k are observed between i and j in period t and $\epsilon_{ij,t}^k$ is a mean-zero error term, independently drawn across origins and sectors, but not necessarily across destinations within the same origin and sector,

$$\epsilon_{ij,t}^k = \xi_{i,t}^k + u_{ij,t}^k,$$

where $\xi_{i,t}^k$ is i.i.d and mean zero across both products and origins and $u_{ij,t}^k$ is i.i.d and mean zero across products, origins and destinations.

To estimate both capability $N_{i,t}$ and complexity n_t^k in any year, we start by taking a double difference, $DD_{ii_0,t}^{kk_0}$, of equation 2.9.1 with respect to a base good k_0 and a base exporter i_0, and average across all j destinations:

$$DD_{ii_0,t}^{kk_0} \equiv \sum_j \frac{1}{J}\left[(\pi_{ij,t}^k - \pi_{i_0 j,t}^k) - (\pi_{ij,t}^{k_0} - \pi_{i_0 j,t}^{k_0})\right] \rightarrow_{J\rightarrow\infty} (N_{i,t} - N_{i_0,t})(n_t^k - n_t^{k_0}) + (\xi_{i,t}^k - \xi_{i_0,t}^k) - (\xi_{i,t}^{k_0} - \xi_{i_0,t}^{k_0}), \tag{2.9.2}$$

where we have applied the law of large numbers across J destination countries to eliminate the u_{ij}^k shocks.

Capability Estimator. In order to estimate $N_{i,t}$, up to affine transformation, we first average this difference-in-difference over goods k to get

$$\sum_k \frac{1}{K} DD_{ii_0,t}^{kk_0} \equiv \sum_k \frac{1}{K} \sum_j \frac{1}{J}\left[(\pi_{ij,t}^k - \pi_{i_0 j,t}^k) - (\pi_{ij,t}^{k_0} - \pi_{i_0 j,t}^{k_0})\right] \rightarrow_{J,K\rightarrow\infty} (N_{i,t} - N_{i_0,t})(\sum_k \frac{1}{K} n_t^k - n_t^{k_0}) - (\xi_{i,t}^{k_0} - \xi_{i_0,t}^{k_0}),$$

where we have applied the law of large numbers across K sectors to eliminate the $(\xi_{i,t}^k - \xi_{i_0,t}^k)$ shocks. This deals with any potential bias due fact that to the fact that country i may be unusually prone to export any particular good k relative to the benchmark country i_0. To address the bias that i may be unusually productive in making the benchmark good relative to the benchmark country (the $\xi_{i,t}^{k_0} - \xi_{i_0,t}^{k_0}$ term), we then take a second weighted average over benchmark goods, with the weights $\omega_t^{k_0} \neq \frac{1}{K}$ chosen such that $(\sum_k \frac{1}{K} n_t^k - \sum_{k_0} \omega_t^{k_0} n_t^{k_0}) \neq 0$ and for which the law of large numbers still applies to the weighted average, i.e. $\sum_{k_0} \omega_t^{k_0}(\xi_{i,t}^{k_0} - \xi_{i_0,t}^{k_0}) \to_{K\to\infty} 0$. This implies

$$\hat{N_{i,t}} \equiv \sum_{k_0} \omega_t^{k_0} \sum_k \frac{1}{K} DD_{i,i_0}^{k,k_0} \to_{J,K\to\infty} (N_{i,t} - N_{i_0,t})(\sum_k \frac{1}{K} n_t^k - \sum_{k_0} \omega_t^{k_0} n_t^{k_0}). \tag{2.9.3}$$

We can therefore use $\hat{N_{i,t}}$ as an estimator of $N_{i,t}$, up to affine transformation,

$$N_{i,t} = a_t \hat{N_{i,t}} + b_t,$$

with $a_t \equiv 1/(\sum_k \frac{1}{K} n_t^k - \sum_{k_0} \omega_t^{k_0} n_t^{k_0})$ and $b_t \equiv N_{i_0,t}$. We discuss the choice of weights $\omega_t^{k_0}$ as well as how we deal with a_t and b_t after introducing our estimator of product complexity.

Complexity Estimator.

We can follow the same steps to obtain an estimator of complexity n_t^k, up to affine transformation. Starting from equation (2.9.2) and averaging across origin countries implies

$$\sum_i \frac{1}{I} DD_{ii_0,t}^{kk_0} \equiv \sum_i \frac{1}{I} \sum_j \frac{1}{J} \left[(\pi_{ij,t}^k - \pi_{i_0j,t}^k) - (\pi_{ij,t}^{k_0} - \pi_{i_0j,t}^{k_0})\right] \to_{J,I\to\infty} (\sum_i \frac{1}{I} N_{i,t} - N_{i_0,t})(n_t^k - n_t^{k_0}) - (\xi_{i_0,t}^k - \xi_{i_0,t}^{k_0}),$$

where we have applied the law of large numbers across I origin countries to eliminate the $(\xi_{i,t}^k - \xi_{i,t}^{k_0})$ term. Averaging again over benchmark countries using weights $\omega_{i_0,t}$ such that $(\sum_i \frac{1}{I} N_{i,t} - \sum_{i_0} \omega_{i_0,t} N_{i_0,t}) \neq 0$ and $\sum_{i_0} \omega_{i_0,t}(\xi_{i_0,t}^k - \xi_{i_0,t}^{k_0}) \to_{I\to\infty} 0$ implies

$$\hat{n}_t^k \equiv \sum_{i_0} \omega_{i_0,t} \sum_i \frac{1}{I} DD_{ii_0,t}^{kk_0} \to_{J,I\to\infty} (\sum_i \frac{1}{I} N_{i,t} - \sum_{i_0} \omega_{i_0,t} N_{i_0,t})(n_t^k - n_t^{k_0}). \tag{2.9.4}$$

We can therefore use $\hat{n}_t^k$ as an estimator of n_t^k, up to affine transformation,

$$n_t^k = c_t \hat{n}_t^k + d_t,$$

with $c_t \equiv 1/(\sum_i \frac{1}{I} N_{i,t} - \sum_{i_0} \omega_{i_0,t} N_{i_0,t})$ and $d_t \equiv n_t^{k_0}$.

Choosing weights.

Our estimators of capability and complexity each require weights, $\omega_t^{k_0}$ and $\omega_{i_0,t}$, respectively. Provided that $\omega_t^{k_0}$ is such that $\sum_k \frac{1}{K} n_t^k - \sum_{k_0} \omega_t^{k_0} n_t^{k_0} \neq 0$ and $\sum_{k_0} \omega_t^{k_0}(\xi_{i,t}^{k_0} - \xi_{i_0,t}^{k_0}) \to_{K\to\infty} 0$ and $\omega_{i_0,t}$ is such that $\frac{1}{I} N_{i,t} - \sum_{i_0} \omega_{i_0,t} N_{i_0,t} \neq 0$ and $\sum_{i_0} \omega_{i_0,t}(\xi_{i_0,t}^k - \xi_{i_0,t}^{k_0}) \to_{I\to\infty} 0$, the previous discussion establishes that $\hat{N_{i,t}}$ and $\hat{n}_t^k$ are consistent estimators of $N_{i,t}$ and n_t^k, up to affine transformation. In small samples, though, the choice of $\omega_t^{k_0}$ and $\omega_{i_0,t}$ may matter for our estimates of $N_{i,t}$ and n_t^k. We now describe how we choose those weights through an iterative procedure.

We start with initial weights $\omega_{i_0,t}^{(0)}$ that focuses on whether country i_0 is a G-10 country in 1962-

1964,

$$\omega_{i_0,t}^{(0)} = \begin{cases} \frac{1}{11} & \text{if } i_0 \in \text{G-10}, \\ 0 & \text{if } i_0 \notin \text{G-10}. \end{cases}$$

By including 11 countries rather than a single one in our reference group, we expect $\sum_{i_0} \omega_{i_0,t}^{(0)}(\xi_{i_0,t}^{k} - \xi_{i_0,t}^{k_0})$ to be close to zero. By only including countries that we expect to be more capable, we also expect $\sum_i \frac{1}{I} N_{i,t} - \sum_{i_0} \omega_{i_0,t}^{(0)} N_{i_0,t} \neq 0$ to hold. These weights give us an initial set of estimates of complexity, $\hat{n}_t^{k,(0)}$.

In any step $s \geq 1$, given estimates of complexity $\hat{n}_t^{k,(s-1)}$ obtained in step $s-1$, we then set

$$\omega_t^{k_0,(s)} = \frac{\max_k \hat{n}_t^{k,(s-1)} - \hat{n}_t^{k_0,(s-1)}}{\sum_l [\max_k \hat{n}_t^{k,(s-1)} - \hat{n}_t^{l,(s-1)}]},$$

which is such that $\omega_t^{k_0,(s)} \in [0,1]$, $\sum_{k_0} \omega_t^{k_0,(s)} = 1$, and $\omega_t^{k_0,(s)}$ is decreasing in $\hat{n}_t^{k_0,(s-1)}$, but increasing in $n_t^{k_0}$ if $\sum_i \frac{1}{I} N_{i,t} - \sum_{i_0} \omega_{i_0,t}^{(s-1)} N_{i_0,t} < 0$. These weights give us a new set of estimates of capability, $\hat{N}_{i,t}^{(s)}$, and a new set of country weights,

$$\omega_{i_0,t}^{(s)} = \frac{\max_i \hat{N}_{i,t}^{(s)} - \hat{N}_{i_0,t}^{(s)}}{\sum_l [\max_i \hat{N}_{i,t}^{(s)} - \hat{N}_{l,t}^{(s)}]},$$

which is also such that $\omega_{i_0,t}^{(s)} \in [0,1]$, $\sum_{i_0} \omega_{i_0,t}^{(s)} = 1$, and $\omega_{i_0,t}^{(s)}$ is decreasing in $\hat{N}_{i_0,t}^{(s)}$, but increasing in $N_{i_0,t}$ if $\sum_k \frac{1}{K} n_t^k - \sum_{k_0} \omega_t^{k_0,(s)} n_t^{k_0} < 0$. These weights give us a new set of estimates of complexity, $\hat{n}_t^{k,(s)}$.

We iterate until convergence of weights $\omega_t^{k_0,(s)}$ and $\omega_{i_0,t}^{(s)}$ to $\omega_t^{k_0}$ and $\omega_{i_0,t}$.

Measuring capability and complexity across time

Our procedure only identifies capability and complexity up to affine transformation. Specifically, recall that we recover $\hat{N_{i,t}}$ and $\hat{n}_t^k$ such that

$$N_{i,t} = a_t \hat{N}_{i,t} + b_t,$$

$$n_t^k = c_t \hat{n}_t^k + d_t,$$

with $a_t \equiv 1/(\sum_k \frac{1}{K} n_t^k - \sum_{k_0} \omega_t^{k_0} n_t^{k_0})$, $b_t \equiv N_{i_0,t}$, $c_t \equiv 1/(\sum_i \frac{1}{I} N_{i,t} - \sum_{i_0} \omega_{i_0,t} N_{i_0,t})$ and $d_t \equiv n_t^{k_0}$. Note that we can rearrange

$$a_t = \frac{1}{c_t(\sum_k \frac{1}{K} \hat{n}_t^k - \sum_{k_0} \omega_t^{k_0} \hat{n}_t^{k_0})}$$

which implies

$$c_t = \frac{1}{a_t(\sum_k \frac{1}{K}\hat{n}_t^k - \sum_{k_0}\omega_t^{k_0}\hat{n}_t^{k_0})}$$

In sum, we have

$$\begin{aligned} N_{i,t} &= \frac{a_t}{c_t}\frac{\hat{N}_{i,t}}{\sum_k \frac{1}{K}\hat{n}_t^k - \sum_{k_0}\omega_t^{k_0}\hat{n}_t^{k_0}} + N_{i_0,t} \\ n_t^k &= \frac{c_t}{a_t}\frac{\hat{n}_t^k}{\sum_i \frac{1}{I}\hat{N}_{i,t} - \sum_{i_0}\omega_{i_0,t}\hat{N}_{i_0,t}} + n_t^{k_0} \end{aligned}$$

So we can measure $N_{i,t}$ and n_t^k up to three constants: $\frac{a_t}{c_t}$, $N_{i_0,t}$, and $n_t^{k_0}$.

To pin down the level and scale of our measures across periods we further assume that the lowest and highest complexity levels are time-invariant and always equal to 0 in the least complex sector, i.e. there is no spillover from producing the least complex product, and moving from specializing in the least to the most complex product generates the same-sized spillover in any period:

$$\begin{aligned} min(n^k) &= 0 \\ max(n^k) &= 1. \end{aligned}$$

Then we know $\frac{c_t}{a_t}$ and $n_t^{k_0}$:

$$\begin{aligned} n_t^k &= \frac{c_t}{a_t}\frac{\hat{n}_t^k}{\sum_i \frac{1}{I}\hat{N}_{i,t} - \sum_{i_0}\omega_{i_0,t}\hat{N}_{i_0,t}} + n_t^{k_0}, \\ n_t^{k_0} &= -\frac{c_t}{a_t}\frac{max(\hat{n}_t^k)}{\sum_i \frac{1}{I}\hat{N}_{i,t} - \sum_{i_0}\omega_{i_0,t}\hat{N}_{i_0,t}}, \\ n_t^{k_0} &= 1 - \frac{c_t}{a_t}\frac{min(\hat{n}_t^k)}{\sum_i \frac{1}{I}\hat{N}_{i,t} - \sum_{i_0}\omega_{i_0,t}\hat{N}_{i_0,t}}, \\ \frac{a_t}{c_t} &= -\frac{max(\hat{n}_t^k) - min(\hat{n}_t^k)}{\sum_i \frac{1}{I}\hat{N}_{i,t} - \sum_{i_0}\omega_{i_0,t}\hat{N}_{i_0,t}}. \end{aligned}$$

Using

$$N_{i,t} = \frac{a_t}{c_t}\frac{\hat{N}_{i,t}}{\sum_k \frac{1}{K}\hat{n}_t^k - \sum_{k_0}\omega_t^{k_0}\hat{n}_t^{k_0}} + N_{i_0,t}$$

this implies that we also know $N_{i,t}$ up to a constant $N_{i_0,t}$. As our dynamic spillovers specification is linear in $N_{i,t}$, the constant $N_{i_0,t}$ is absorbed by the fixed effects when we estimate the size of the dynamic spillovers. (i.e. the regression specification in equation 2.17) and so it does not need to be pinned down. For expositional purposes, we first demean our recovered capability measure, $\frac{a_t}{c_t}\frac{\hat{N}_{i,t}}{\sum_k \frac{1}{K}\hat{n}_t^k - \sum_{k_0}\omega_t^{k_0}\hat{n}_t^{k_0}}$, by year and then chose $N_{i_0,t}$ to ensure that the US has an average capability of 1 across periods.[22]

2.9.3 Baseline Measures of Complexity and Capability: Additional Figures and Tables

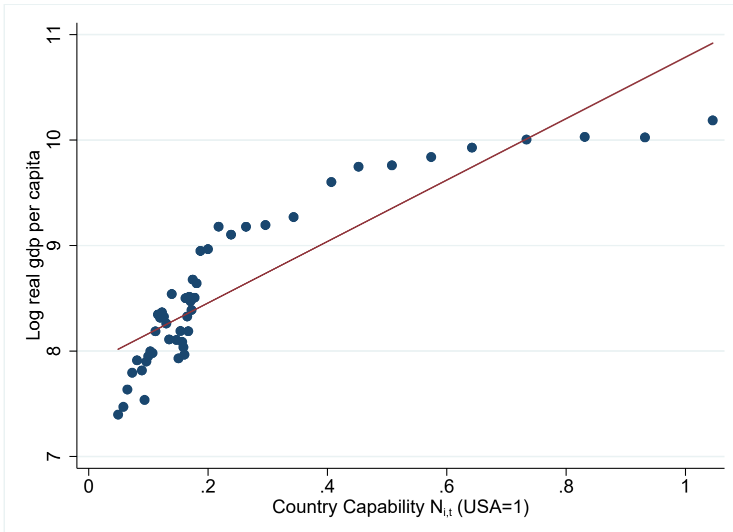


Figure 2.9.1: Real GDP per capita vs. capability (within years)

Notes: Figure 2.9.1 is the binned scatter plot associated with a regression of log real GDP per capita on both capabilities and year fixed effects.

[22]I.e. $N_{i,t} = \hat{Z}_{i,t} - \sum_i \frac{\hat{Z}_{i,t}}{N} + (1 - \sum_t \frac{\hat{Z}_{i,t}}{T} + \sum_t \sum_i \frac{\hat{Z}_{i,t}}{NT})$ where $\hat{Z}_{i,t} = \frac{a_t}{c_t} \frac{\hat{N}_{i,t}}{\sum_k \frac{1}{K} \hat{n}_t^k - \sum_{k_0} \omega_t^{k_0} \hat{n}_t^{k_0}}$.

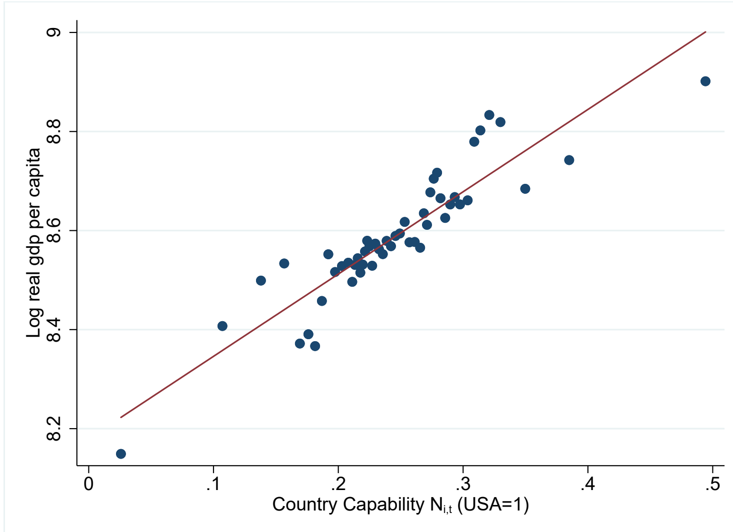


Figure 2.9.2: Real GDP per capita vs. capability (within years and countries)

Notes: Figure 2.9.2 is the binned scatter plot associated with a regression of log real GDP per capita on both capabilities, year fixed effects and country fixed effects.

2.9.4 Construction of Instrumental Variables

Our goal is to construct instrumental variables that predict average complexity, $S_{i,t}$, in a country i as a function of the entry of other countries c into the WTO at dates $t_c \leq t$. To do so, we model the entry of any country c into the WTO as a uniform trade cost shock such that for all $t \geq t_c$,

$$(A^k_{ij,t})'_c = \begin{cases} e^{\alpha} A^k_{ij,t_c-1} & \text{if } i = c \text{ and } j \neq c, \\ A^k_{ij,t_c-1} & \text{otherwise.} \end{cases}$$

with $\alpha > 0$. We then compute, up to a first-order approximation, the counterfactual change in country i's average complexity, $(\Delta S_i)_c$, that would have been observed in any period $t \geq t_c$ if the entry of country c was the only shock occurring from period t_c onward and all wages were to remain fixed. We finally sum the previous changes across all WTO entry events that are prior to date t to construct predictors of $S_{i,t}$.

Formally, let $\omega^k_{i,t_c-1} \equiv \ell^k_{i,t_c-1}/L_{i,t_c-1}$ denote the share of employment in sector k and country i at date $t_{c,WTO} - 1$ and $(\omega^k_{i,t})'_c$ denote the counterfactual share associated with the entry of country c in the WTO if it were the only shock occurring up to date $t > t_c - 1$. The counterfactual value of $S_{i,t}$ is given by $(S_{i,t})'_c = \sum_k n^k_{t_c-1}(\omega^k_{i,t})'_c$. We can therefore express the associated change $(\Delta S_{i,t})_c \equiv (S_{i,t})'_c - S_{i,t_c-1}$ as

$$(\Delta S_i)_c = \sum_k n^k_{t_c-1}(\Delta\omega^k_{i,t})_c,$$

with $(\Delta\omega^k_{i,t})_c \equiv (\omega^k_{i,t})'_c - \omega^k_{i,t_c-1}$. Up to a first-order approximation, we also have

$$
\begin{aligned}
(\Delta\omega^k_{i,t})_c/\omega^k_{i,t_c-1} = & \\
& \sum_{j\neq c}\rho^k_{ij,t_c-1}[(\sigma-1)\alpha + (\sigma-\epsilon)\Delta\ln(P^k_{j,t})_c + (\epsilon-1)\Delta\ln(P_{j,t})_c] \\
- \sum_{k'}\omega^{k'}_{i,t_c,-1} & \sum_{j\neq c}\rho^{k'}_{ij,t_c-1}[(\sigma-1)\alpha + (\sigma-\epsilon)\Delta\ln(P^k_{j,t})_c + (\epsilon-1)\Delta\ln(P_{j,t})_c],
\end{aligned}
$$

with $\rho^k_{ij,t_c-1} = \ell^k_{ij,t_c-1}/[\sum_{j'}\ell^k_{ij',t_c-1}]$ the share of employment in country i and sector k associated with destination j and the log-changes in prices $\Delta\ln(P^k_{j,t})_c \equiv \ln(P^k_{j,t})'_c - \ln P^k_{j,t_c-1}$ and $\Delta\ln(P_{j,t})_c \equiv \ln(P_{j,t_c})' - \ln P_{j,t_c-1}$ given by

$$
\begin{aligned}
\Delta\ln(P^k_{j,t})_c &= -\alpha\lambda^k_{cj,t_c-1}, \text{ for all } j\neq c, \\
\Delta\ln(P_{j,t})_c &= -\alpha\lambda_{cj,t_c-1}, \text{ for all } j\neq c,
\end{aligned}
$$

with $\lambda^k_{cj,t}$ the share of country j's expenditure on good k allocated to country c at date t, $e^k_{j,t}$ the share of expenditure of country j on sector k, and $\lambda_{cj,t} = \sum_k e^k_{j,t}\lambda^k_{cj,t}$ the total share of expenditure on goods from country c in destination j. Regrouping terms, this leads to

$$
\begin{aligned}
(\Delta S_{i,t})_c = & -\alpha(\sigma-\epsilon)\{\sum_k n^k_{t_c-1}\omega^k_{i,t_c-1}[\sum_{j\neq c}\rho^k_{ij,t_c-1}\lambda^k_{cj,t_c-1} - \sum_{k'}\omega^{k'}_{i,t_c-1}\sum_{j\neq c}\rho^{k'}_{ij,t_c-1}\lambda^{k'}_{cj,t_c-1}]\} \\
& -\alpha(\epsilon-1)(\sigma-\epsilon)\{\sum_k n^k_{t_c-1}\omega^k_{i,t_c-1}[\sum_{j\neq c}\rho^k_{ij,t_c-1}\lambda_{cj,t_c-1} - \sum_{k'}\omega^{k'}_{i,t_c-1}\sum_{j\neq c}\rho^{k'}_{ij,t_c-1}\lambda_{cj,t_c-1}]\}.
\end{aligned}
$$

Summing across all WTO entry events by a country $c\neq i$ that have occurred before a given date t, we obtain the following predictor $\hat{S}_{i,t}$ of average complexity in country i at date t,

$$
\hat{S}_{i,t} \equiv \sum_{c\neq i}\mathbb{1}_{\{t\geq t_{c,WTO}\}}(\Delta S_{i,t})_c = -\alpha(\sigma-\epsilon)Z^I_{i,t} - \alpha(\epsilon-1)(\sigma-\epsilon)Z^{II}_{i,t},
$$

where $Z^I_{i,t}$ and $Z^{II}_{i,t}$ are such that

$$
\begin{aligned}
Z^I_{i,t} &= \sum_{c\neq i}\mathbb{1}_{\{t\geq t_c\}}\sum_k n^k_{t_c-1}\times\omega^k_{i,t_c-1}(\sum_{j\neq c}\rho^k_{ij,t_c-1}\lambda^k_{cj,t_c-1} - \sum_{k'}\omega^{k'}_{i,t_c-1}\sum_{j\neq c}\rho^{k'}_{ij,t_c-1}\lambda^{k'}_{cj,t_c-1}), \\
Z^{II}_{i,t} &= \sum_{c\neq i}\mathbb{1}_{\{t\geq t_c\}}\sum_k n^k_{t_c-1}\times\omega^k_{i,t_c-1}(\sum_{j\neq c}\rho^k_{ij,t_c-1}\lambda_{cj,t_c-1} - \sum_{k'}\omega^{k'}_{i,t_c-1}\sum_{j\neq c}\rho^{k'}_{ij,t_c-1}\lambda_{cj,t_c-1}).
\end{aligned}
$$

These are the two instrumental variables used to estimate equation (2.17).

2.9.5 Alternative Measures of Capability and Complexity

Alternative Measures.

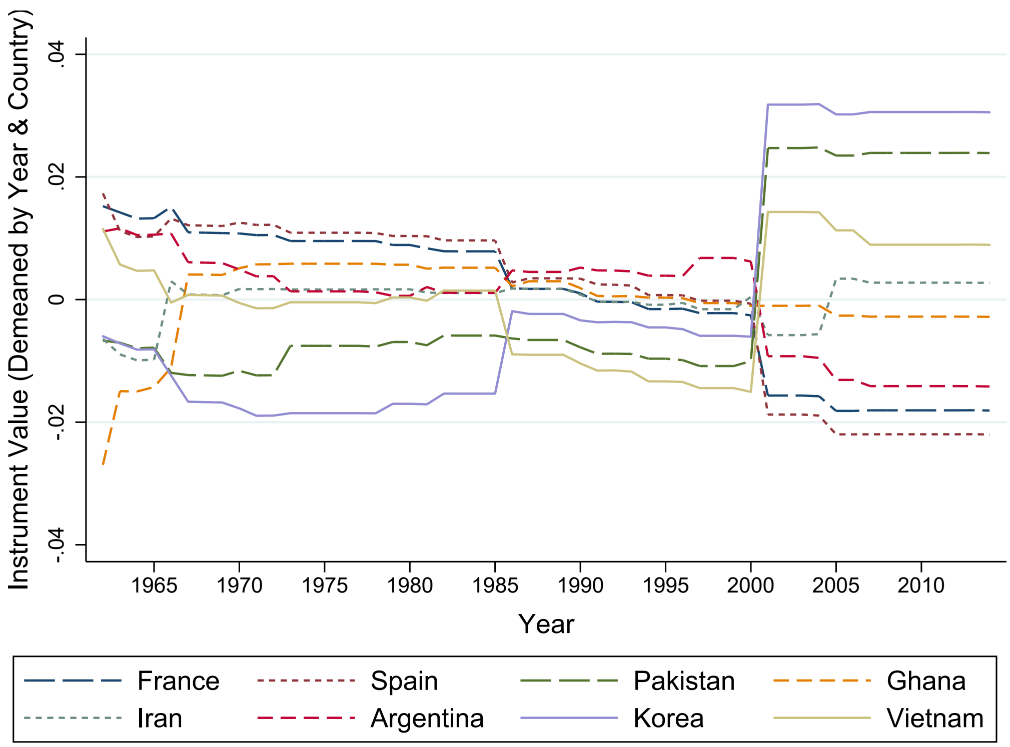


Figure 2.9.3: Time Path of $Z_{i,t}^{II}$

Notes: Figure 2.9.3 plots the value of the instrument $Z_{i,t}^{II}$ over time for a selection of similarly-sized countries in our sample. The instrument captures the change in complexity-weighted competition due to aggregate-level price index changes induced by other countries' entry into the WTO and derives from a first-order approximation of the change in average complexity due to trade cost shocks to WTO entrants (see Appendix 2.9.4).

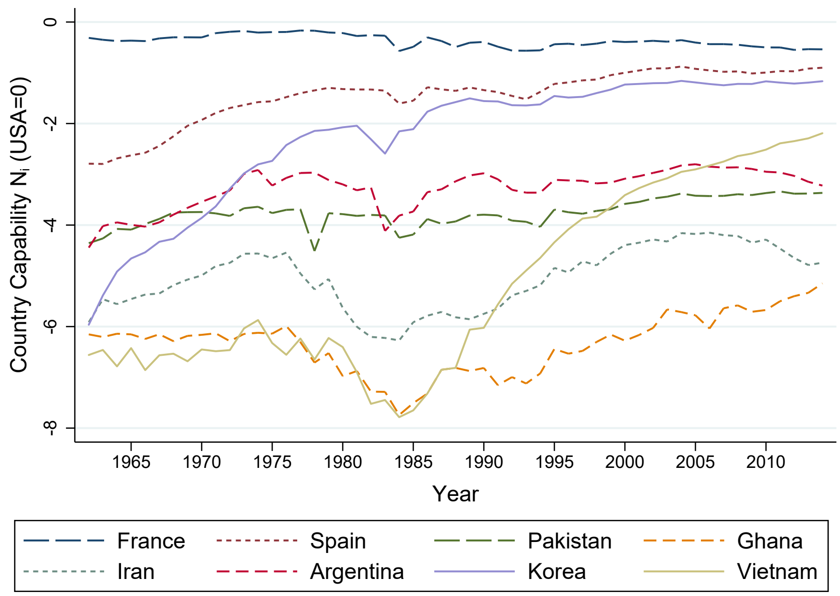


Figure 2.9.4: Alternative Capability ($\tilde{N}_{i,t}$)

Notes: Figure 2.9.4 reports the country fixed effects $\tilde{N}_{i,t}$ in the maximum likelihood estimation of equation (2.23) in a given year t, as described in Section 2.3.1. Fixed effects are normalized so that $N_{USA,t} = 0$ for all t.

Table 2.9.2: Alternative Complexity ($\tilde{n}_t^k$)

	Sectors with highest n^k (Average Value, 1962-2014)	
1	Railway Passenger Cars	3.233
2	Electric Trains	3.230
3	Warships	3.193
4	Mechanically Propelled Railway	2.894
5	High-pressure hydro-electric conduits of steel	2.690
6	Leather Articles Used in Machinery	2.665
7	Rotary Converters	2.557
8	Hats	2.533
9	Aircraft Tires	2.526
10	Nuclear Reactors	2.526
	Sectors with lowest n^k (Average Value, 1962-2014)	
1	Medicaments	-1.626
2	Chemical Products	-1.237
3	Miscellaneous Non-Electrical Machinery Parts	-1.157
4	Miscellaneous Electrical Machinery	-1.128
5	Miscellaneous Non-Electrical Machines	-1.067
6	Finished Cotton Fabrics	-1.007
7	Footwear	-1.001
8	Medical Instruments	-0.985
9	Electric Wire	-0.969
10	Miscellaneous Hand Tools	-0.969

Notes: Table 2.9.2 reports the 10 highest and 10 lowest values of the sector fixed effects $\tilde{n}_t^k$ in the maximum likelihood estimation of equation (2.23) averaged across all years from 1962 to 2014 for products with at least 40 years of data.

Table 2.9.3: First Stage Regressions: Responses of Industry Structure to WTO Entrants

	Average Complexity $S_{i,t}$	
	(1)	(2)
WTO Entrant Shock $Z^{I}_{i,t}$	-2.945***	-1.064
(Product-Destination Level)	(0.541)	(0.660)
WTO Entrant Shock $Z^{II}_{i,t}$		-12.03***
(Destination Level)		(2.057)
Country and year FEs	Yes	Yes
Observations	7,617	7,617
R-squared	0.723	0.729
Clusters	1588	1588

Notes: Table 2.9.3 reports estimates of α_1 and α_2 in equation (2.18). Standard errors clustered at the 5-year-period-country level.

Alternative Estimates of Dynamic Spillovers.

Robustness

Table 2.9.5: Changes in Capability and Industrial Structure: Robustness (I)

	Country Capability $N_{i,t+1}$				
	(1)	(2)	(3)	(4)	(5)
	Baseline	Feenstra et al. Dataset	All Length Panels	No Size Threshold	Higher Size Threshold
Average Complexity $S_{i,t}$	-0.390**	-0.223	-0.488**	-0.398**	-0.458**
	(0.196)	(0.227)	(0.205)	(0.201)	(0.192)
Initial Capability $N_{i,t}$	0.549***	0.567***	0.491***	0.543***	0.546***
	(0.0296)	(0.0291)	(0.0296)	(0.0298)	(0.0345)
Country and year FEs	Yes	Yes	Yes	Yes	Yes
Observations	6,872	6,864	7,905	6,995	5,986
R-squared	0.348	0.383	0.261	0.333	0.371
Clusters	1438	1438	1673	1466	1249
CD F-Stat	119.7	96.88	122.9	113.3	97.36
KP F-Stat	23.43	18.75	24.17	22.98	21.37

Notes: Table 2.9.5 reports estimates of β and ϕ in equation (2.17). All columns use the two-instrument IV

Table 2.9.4: Changes in Capability and Industrial Structure

	Country Capability $N_{i,t+1}$		
	(1)	(2)	(3)
	OLS	IV ($Z^I_{i,t}$)	IV ($Z^I_{i,t}$ and $Z^{II}_{i,t}$)
Average Complexity $S_{i,t}$	0.0412	-0.0474	-0.390**
	(0.0302)	(0.249)	(0.196)
Initial Capability $N_{i,t}$	0.595***	0.586***	0.549***
	(0.0210)	(0.0320)	(0.0296)
Country and year FEs	Yes	Yes	Yes
Observations	6,872	6,872	6,872
R-squared	0.970	0.405	0.348
Clusters	1438	1438	1438
CD F-Stat		107.5	119.7
KP F-Stat		21.65	23.43

Notes: Table 2.9.4 reports estimates of β and ϕ in equation (2.17). Columns 2 and 3 instrument average complexity $S_{i,t}$ by the WTO shocks $Z^I_{i,t}$ and $Z^{II}_{i,t}$. Standard errors clustered at the 5-year-period-country level.

Table 2.9.6: Changes in Capability and Industrial Structure: Robustness (II)

	Country Capability $N_{i,t+1}$				$GNI_{i,t+1}$
	(1)	(2)	(3)	(4)	(5)
	Baseline	10-year Lag	1 Obs. per 5-year Cluster	IV $N_{i,t}$	
Average Complexity $S_{i,t}$	-0.390**	-0.447	-0.511*	0.213	-0.212***
	(0.196)	(0.280)	(0.302)	(0.215)	(0.0802)
Initial Capability $N_{i,t}$	0.549***	0.283***	0.547***	1.039***	
	(0.0296)	(0.0422)	(0.0469)	(0.202)	
GNI per capita $GNI_{i,t}$					0.766***
					(0.0308)
Country and Year FEs	Yes	Yes	Yes	Yes	Yes
Observations	6,872	6,151	1,295	6,195	6,107
R-squared	0.348	0.057	0.284	0.164	0.605
Clusters	1438	723	1295	1303	1269
CD F-Stat	119.7	107.5	20.18	8.123	128.4
KP F-Stat	23.43	21.83	11.73	2.315	26.94

Notes: Table 2.9.6 reports estimates of β and ϕ in equation (2.17). All columns use the two-instrument IV strategy. Column 1 reports our baseline estimates (column 3 of Table 2.9.4). Column 2 reports the same estimates using 10-year lags. Column 3 uses one observation per 5-year cluster. Column 4 instruments initial capability using lagged-values of the WTO shocks $Z^{I}_{i,t}$ and $Z^{II}_{i,t}$. Column 5 uses GNI per capita instead of capability. Standard errors clustered at the 5-year-period-country level.

Chapter 3

Home-Market Effect in Asset Production

3.1 Introduction

Capital tends to flow from fast-growing countries to slow-growing countries. This pattern holds both between developing countries and developed countries as well as within developing countries (Gourinchas and Jeanne, 2013). This pattern is puzzling from the viewpoint of neoclassical models, as these models strongly predict the opposite. While the dominant explanations attribute this puzzle to the differences in financial development, virtually all theories assume that slow-growing countries happen to have relatively well-developed financial sectors. Without an understanding of the determinants of the financial development, policy implications drawn from these theories can be misleading.

This paper proposes a parsimonious theory in which slower growth causes financial development and capital inflows through a home-market effect. The theory puts forward the idea that financial development is demand driven. A slow growing country has a relatively larger amount of wealth that needs to be stored. The larger domestic demand for store of value, in turn, stimulates domestic financial innovation. This endogenous response of financial development can be strong enough to attract capital *inflows.*

I operationalize the idea in the context of a model in the spirit of Caballero, Farhi, and Gourinchas (2008, hereafter CFG) by endogeneizing financial development. In their framework, the share of future output that is accrued by the asset holders differs exogenously across countries. They interpreted this parameter as capturing the difference in financial development. Under the assumption that slow-growing countries have high enough financial development, their model generates capital flow reversals. In my model, this financial develoment is endogenously determined through the entry of banks in the monopolistically competitive financial sector á la Krugman (1980). Slow-growing countries endogenously achieve higher financial development because their relatively larger local demand for store of value stimulates the entry of banks. When the returns to scale in the financial sector are sufficiently high, they become net exporter of assets.

The key assumption of the model is that the financial sector features increasing returns to scale. Both theoretical and empirical works are supportive of the view that this is a realistic feature. On the theoretical front, a larger operating scale enables banks to diversify risks both on the liability side and the asset side. A traditional theory of banking emphasizes the key role of banks as diversification on the liability side (Diamond and Dybvig, 1983). Diversification on the asset side enables securitization, which was the main driver of the popularity of U.S. financial assets in the 2000s. A more recent theory by Phelan (2017) argues that the efficiency of banking come from the fact that the borrowers' information acquisition features increasing returns to scale when the relevant information is correlated across borrowers.

On the empirical front, the majority of available bank-level estimates of returns to scale suggest the presence of the increasing returns to scale among U.S. banks (Wheelock and Wilson, 2012; Hughes and Mester, 2013; Anderson and Joeveer, 2012). Although these estimates do not necessarily imply the existence of the increasing returns to scale at the sectoral level, they are suggestive considering that the financial sector is typically dominated by a small number of banks.

Related Literature. The insights of this paper build on home-market effect in the international trade literature (Linder, 1961; Krugman, 1980). This literature is motivated by the fact that high-income countries tend to be net exporters of income elastic goods, which is exactly the opposite of what happens in the neoclassical trade models. While Krugman (1980) has formally shown increasing returns and trade costs can reverse these predictions by assuming exogenous taste differences, Fajgelbaum, Grossman, and Helpman (2011) and Matsuyama (2018) revisited by endogeneizing demand differences through non-homothetic preferences. Costinot, Donaldson, Kyle, and Williams (2016) provide empirical evidence in support of these theories in the context of pharmaceutical industries.

This paper is close in the spirit to Martin and Rey (2004), who also linked the insights from the new trade theory to international capital flows. They emphasized the country size as the determinant of capital flows in the context of the two-period model. My work differs from the literature in that I study intertemporal trade in an infinite horizon model incorporating the endogenous demand differences for store of value through the differences in growth rates.

There have been many studies explaining capital flow reversals by assuming slow-growing countries happen to have a feature to attract capital inflows, as surveyed in Gourinchas and Rey (2014). Many papers (Song, Storesletten, and Zilibotti, 2011; Buera and Shin, 2017; Coeurdacier, Guibaud, and Jin, 2015) share the insights with CFG that fast-growing countries suffer more from financial friction so that it lacks the ability to generate store of value. Other explanations include the lack of ability of fast-growing countries to generate insurance assets (Mendoza, Quadrini, and Rios-Rull, 2009; Angeletos and Panousi, 2011) and its less capital intensive industrial structure (Jin, 2012). While I build on the framework of CFG, the general insights in endogenizing the comparative advantage in asset production may well apply to all the other specific mechanisms.

Layout. Section 3.2 describes the model setup. Section 3.3 shows analytically that the capital flows from fast-growing to slow-growing countries around a symmetric balanced growth path equilibrium. Section 3.4 complements the analytical results with the numerical simulation, which allows me to study transition dynamics. Section 3.5 concludes.

3.2 Model

This section describes the model setup. The model extends CFG and its generalization by Gourinchas and Rey (2014). The key ingredient of the model is that the demand for store of values determines the incentive for financial innovation, which, in turn, determines the direction of capital flows.

There are two countries, $i = 1, 2$. Each country is populated by overlapping generations of perpetual youth á la Yaari (1965) and Blanchard (1985). Each household faces an i.i.d. instantaneous probability of dying θ, and a fraction θ of the population is born every instant. As is well known, a competitive life-insurance market will offer a rate of return θ per unit of wealth conditional on surviving. I assume that households have log-utility with discount rate ρ and receive all non-financial income at birth.

Let $c_i(s,t), w_i(s,t), z_i(s,t)$ denote the consumption, wealth, and non-financial income at time t of a household born at time s in country i, respectively. Each household solves

$$\max \int_t^\infty e^{-(\rho+\theta)(u-s)} \log c_i(s,u) du$$

subject to

$$\frac{dw_i(s,t)}{dt} = (\tilde{r}_{it} + \theta)\, w_i(s,t) - c_i(s,t) + z_i(s,t), \tag{3.2.1}$$

where $z(s,t) = Z_t$ for $s = t$ and zero otherwise, Z_t denotes the aggregate non-financial income at time t, and $\tilde{r}_{it}$ is the individual return on their wealth, which will be described in detail below. With log-utility, it is straightforward to derive that the consumption function takes the form

$$c_i(s,t) = (\rho + \theta) w_i(s,t) \tag{3.2.2}$$

for $t > s$ and $c(t,t) = (\rho + \theta) z(t,t)$. Define the aggregate wealth as $W_{it} \equiv \int w_i(s,t) \theta e^{-\theta(t-s)} ds$, where $\theta e^{-\theta(t-s)}$ is the size of the cohort born at time $s < t$. Aggregating the budget constraint (3.2.1) and the consumption function, we have the following law of motion for the aggregate wealth, $W_{it} \equiv \int w_i(s,t) \theta e^{-\theta(t-s)} ds$, where $\theta e^{-\theta(t-s)}$ is the size of the cohort born at time $s < t$:

$$\frac{dW_{it}}{dt} = (R_{it} - \rho - \theta) W_{it} + Z_{it},$$

where R_{it} is the aggregate return on wealth in country i.

Households can deposit their wealth into either of the countries. At each point in time, I assume that the households draw idiosyncratic shocks that generate heterogenous beliefs about the asset returns. They know each others' belief, and they agree to disagree. I also assume investing abroad is costly, which erodes the rate of return. In particular, household h in country i believes the return from financial assets from country j is $\epsilon_{jth} r_{jt} / \chi_{ji}$, where ϵ_{ith} follows the Fréchet distribution with shape parameter $\kappa \geq 1$, and $\chi_{ji} \geq 1$ is the international transaction costs in assets with $\chi_{ii} = 1$ being the normalization. The well-known property of the Fréchet distribution gives the aggregate portfolio share of country j invested in country i as

$$\lambda_{jit} = \frac{(r_{jt}/\chi_{ji})^{\kappa}}{\sum_l (r_{lt}/\chi_{li})^{\kappa}}. \tag{3.2.3}$$

Given this, the aggregate return on wealth in each country is

$$R_{it} = \sum_j \lambda_{jit} r_{jt} / \chi_{ji},$$

which is the weighted average of the asset return from the two countries.

Equation (3.2.3) gives an iso-elastic demand curve for assets from different origins, also known as a gravity equation. In the standard international macro models, assets from different origins were typically assumed to be perfect substitutes so that $\kappa = \infty$. I introduce imperfect substitutability for two reasons. The first reason is technical. I need some form of imperfect substitutability in order to accommodate increasing returns to scale in the financial sector to avoid pathological features such as multiple equilibria (see Kucheryavyy, Lyn, and Rodríguez-Clare, 2016). The second reason is empirical. Broner, Didier, Erce, and Schmukler (2013) document that the average gross capital flow in the 2000s is more than 30 times larger than the gross capital flow. If assets were perfect substitutes, there should be no reason to have such large discrepancies between gross and net capital flow. The belief heterogeneity gives a parsimonious micro-foundation for why assets from different origins could be imperfect substitutes.[1]

Now I turn to the description of the supply side of the economy. Each country i is endowed with output Y_{it} in each period, and the output grows at rate $g_{it} \equiv \dot{Y}_{it}/Y_{it} > 0$. Here, I assume the output is exogenous for simplicity, but in Section 3.4, I endogenize. The only asset available in the economy is capital, which returns a fraction of output every period. I assume there is no depreciation of capital.

Households are not able to directly hold capital. Instead, households deposit their wealth to banks, and banks manage capital. The amount of output that is accrued as the claims to capital is endogenously determined by the aggregate amount of financial services, S_{it}. The remaining output, $Y_{it} - S_{it}$, is captured by newborns as non-financial income, so that $Z_{it} = Y_{it} - S_{it}$. The aggregate financial service in the country is a CES composite of financial services provided by the banks in that country:

$$S_{it} \equiv \left(N_{it}^{-\varphi} \int_0^{N_{it}} (s_{it}(\omega))^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}, \tag{3.2.4}$$

where $s_{it}(\omega)$ is the financial services provided by a bank ω, N_{it} is the measure of banks, $\sigma > 1$ is the elasticity of substitution across financial services, and φ controls returns to the increase in the number of banks. Setting $\varphi = 0$ recovers the Dixit and Stiglitz (1977) demand system, whereas setting $\varphi > 0$ weakens the returns from love of variety.

In order for each bank ω to produce a strictly positive amount of financial services, it requires $\psi_i s(\omega) + fY_{it}$ units of deposit (wealth), where ψ_i is the marginal cost and fY_{it} is the entry cost of banks. The entry cost scales with the output so as to ensure balanced growth. Such an assumption is commonly used in the growth literature (e.g., Romer, 1990; Aghion and Howitt, 1992) and is consistent with the facts documented in Bollard, Klenow, and Li (2016). Since banks are monopolistically competitive, they charge a constant markup, $\mu \equiv \sigma/(\sigma-1)$. The free-entry condition is $(\mu-1)\psi_i s = fY_{it}$, where I dropped the dependence on ω because all the banks are symmetric. The total amount of

[1] I am very grateful to Arnaud Costinot for suggesting this micro-foundation. An alternative approach is to assume assets from different countries differ in their risk characteristics (Martin and Rey, 2004; Okawa and Van Wincoop, 2012).

wealth deposited to country i, K_{it}, has to satisfy $K_{it} = N_{it}(\psi_i s + fY_{it})$. Combining these equations, we have $N_{it} = K_{it}/(fY_{it}\sigma)$ and $s = fY_{it}/((\mu-1)\psi_i)$. Substituting these back into (3.2.4), we have

$$S_{it} = \bar{\delta}_i \left(\frac{K_{it}}{Y_{it}}\right)^{\phi} Y_{it},$$

where $\phi \equiv \frac{\sigma}{\sigma-1}(1-\varphi)$ and $\bar{\delta}_i \equiv (f\sigma)^{\frac{\sigma}{\sigma-1}(\varphi-1)}(\sigma-1)f/\psi_i$. Therefore, the fraction

$$\delta_i = \bar{\delta}_i \left(K_{it}/Y_{it}\right)^{\phi}$$

of output can be pledged as the claims to capital, which I refer to as financial development. The return on deposit satisfies $r_{it}K_{it} = \bar{\delta}_i(K_{it}/Y_{it})^{\phi}Y_{it}$, which in turn can be rewritten as

$$r_{it} = \bar{\delta}_i(K_{it}/Y_{it})^{\phi-1}. \tag{3.2.5}$$

I define the financial development of country i as the share of output that accrues to the owner of capital, $\delta_i \equiv \bar{\delta}_i(K_{it}/Y_{it})^{\phi}$. CFG correspond to the case with $\phi = 0$ (or $\varphi = 1$), which implies the exogenous financial development $\delta_i = \bar{\delta}_i$. For $\phi \in [0,1)$, we have decreasing returns to scale in financial sector, as a higher capital-to-GDP (K_{it}/Y_{it}) ratio lowers the rate of return on capital. For $\phi > 1$ (or $\varphi < 1/\sigma$), we have increasing returns, as the larger demand for store of values increases the rate of return.

I emphasize here that what is important is the increasing returns to scale in asset production, captured through equation (3.2.5), while the underlying specific micro-foundation is not important. Although I have borrowed a particular structure from Krugman (1980) to micro-found the equation equation (3.2.5), this is not the only micro-foundation that leads to the equation (3.2.5). Other possible micro-foundations are Marshallian external economies of scale, Melitz (2003)-style model with Pareto productivity distribution in the financial sector, or endogenous innovation by a monopolist (see Kucheryavyy, Lyn, and Rodríguez-Clare (2016) and Costinot, Donaldson, Kyle, and Williams (2016)).

3.3 Analytical Results

I describe the analytical results by focusing on small changes around a symmetric balanced growth path equilibrium. First and foremost, the capital flows into slow-growing countries when the returns to scale are sufficiently high. As a corollary, I then show that promoting domestic savings can increase the current account deficits. Second, financial globalization magnifies the cross-country differences in financial development. This implies that increase in the global imbalances is amplified through the endogenous

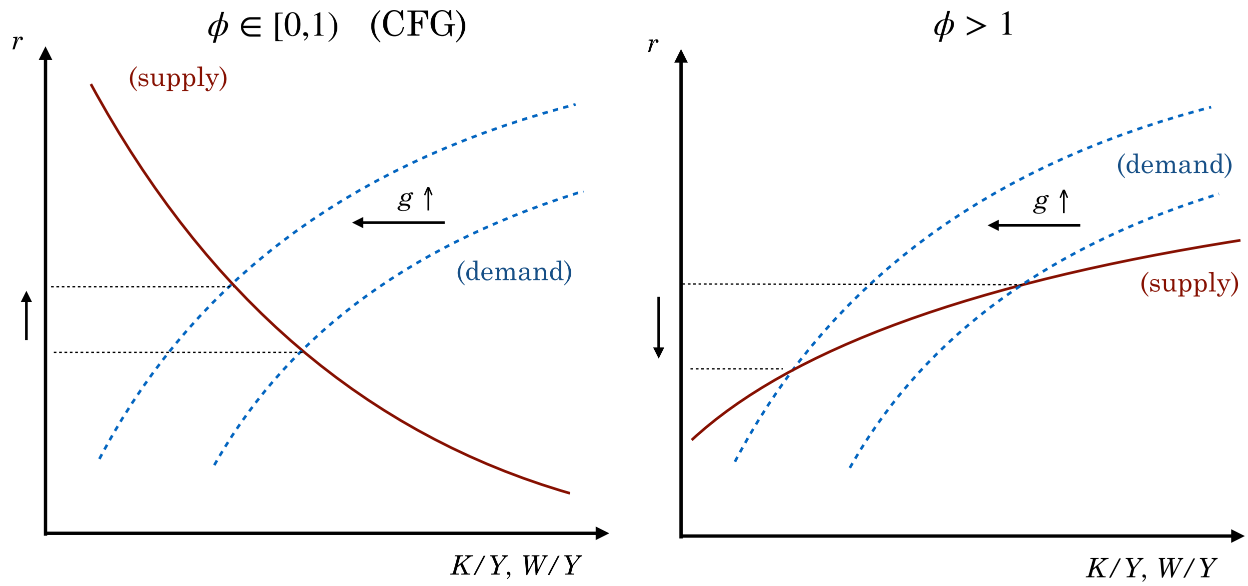


Figure 3.3.1: Metzler diagram

responses of the financial sector.

3.3.1 Financial Autarky

Let me start by describing financial autarky, $\chi_{ij} = \infty$ for $i \neq j$. In financial autarky, $K_{it} = W_{it}$. Using the fact that $Z_{it} = (1 - \bar{\delta}(K_{it}/Y_{it})^{\phi})Y_{it}$, along the balanced growth path, we can express the demand for store of value as

$$\frac{W_{it}}{Y_{it}} = \frac{(1 - \bar{\delta}_i(W_{it}/Y_{it})^{\phi})}{g_i + \rho + \theta - r_i}. \tag{3.3.1}$$

The expression provides positive relationship between W_{it}/Y_{it} and r_i. The supply of store of value is

$$\frac{K_{it}}{Y_{it}} = \left(\frac{\bar{\delta}_i}{r_i}\right)^{\frac{1}{1-\phi}}. \tag{3.3.2}$$

If $\phi \in [0, 1)$, it gives the negative relationship between K_{it}/Y_{it} and r_i. In contrast, if $\phi > 1$, it gives the positive relationship. Figure 3.3.1 shows the familiar Metzler diagram, where supply and demand for assets are plotted. For $\phi < 1$, we recover CFG, where the supply curve of assets is downward sloping. With increasing returns to scale in the financial sector, $\phi > 1$, the supply curve is upward sloping.

Solving (3.3.1) and (3.3.2) immediately gives the financial development in the autarky

$$\delta_i = \bar{\delta}_i \left(\frac{1}{g_i + \rho + \theta} \right)^{\phi}$$

and the autarky interest rate

$$r_i^a = \bar{\delta}_i \left(\frac{1}{g_i + \rho + \theta} \right)^{\phi - 1}.$$

This leads to the following proposition:

Proposition 13. *In the steady state autarky equilibrium, an increase in the growth rate decreases the autarky interest rate if and only if $\phi > 1$.*

In the CFG case ($\phi < 1$), higher growth will, ceteris paribus, make the interest rate higher. Higher growth implies that the country has less wealth that needs to be stored relative to the current output. If the supply of the store of value is exogenous ($\phi = 0$) or features decreasing returns ($\phi \in (0, 1)$), the interest rate needs to rise in order for the asset market to clear. Under the increasing returns ($\phi > 1$), a higher demand for assets will enable the country's financial sector to operate at a larger scale, which raises efficiency and leads to a higher interest rate.

3.3.2 Direction of Capital Flow in an Open Economy

Now I move on to an open economy case with $\chi_{ij} < \infty$. The balanced growth equilibrium is $\{K_i, r_i, W_i\}_{j=1,2}$ such that the market for capital clears for each country

$$K_{jt} = \sum_{i=1,2} \lambda_{jit} W_{it},$$

where λ_{ijt} is given by (3.2.3). Along the balanced growth path, a country's wealth is given by

$$\frac{W_{jt}}{Y_{jt}} = \frac{(1 - \bar{\delta}_j (K_{jt}/Y_{jt})^{\phi})}{g_i + \rho + \theta - \sum_l \lambda_{ljt} r_{lt} / \chi_{lj}},$$

where $r_{jt} = \bar{\delta}_i (K_{jt}/Y_{jt})^{\phi - 1}$ is the return on deposit in each country.

Note that CFG correspond to a special case with $\kappa = \infty$ (the assets are perfect substitutes across countries), $\chi_{ij} = 1$ for all i and j, and $\phi = 0$. Country i's asset position on country j is

$$A_{ji} \equiv \lambda_{ji} W_i.$$

The net foreign asset position of country i is its asset minus liability:

$$NFA_i \equiv A_{ji} - A_{ij}.$$

Accordingly, the current account is defined as the changes in the net foreign asset positions: $CA_i \equiv N\dot{F}A_i$. These definitions are standard except that my model takes into account the presence of gross capital flow due to belief heterogeneity.

Growth and Capital Flow. I consider a first-order perturbation with respect to the past growth rate, g_i, around the symmetric equilibrium, which imposes $g_1 = g_2 \equiv g$, $\chi_{12} = \chi_{21} \equiv \chi > 1$, and $Y_{1t} = Y_{2t} \equiv Y_t$. This not only simplifies the algebra but also ensures that the growth rate is the sole cross-country difference. The symmetric equilibrium is given by $W/Y = K/Y = 1/(g+\rho+\theta)$, and $r = \bar{\delta}(g+\rho+\theta)^{1-\phi}$, which are the same as the autarky equilibrium. It is illustrative to first treat the wealth to GDP ratio as exogenous. The log-difference between asset and liability can be expressed as

$$d\log(A_{21}/A_{12}) = \frac{1+\lambda\kappa(\phi-1)(2\lambda-3)}{(1-4\lambda(1-\lambda)\kappa(\phi-1))}\left(d\log(W_1/Y_1) - d\log(W_2/Y_2)\right),$$

where $\lambda \equiv \lambda_{11} = \lambda_{22} > 1/2$ is the domestic portfolio share. Under the condition that $\kappa(\phi-1) \leq 1$, the denominator is always positive. If we further impose $\phi < \bar{\phi} \equiv 1 + \frac{1}{\lambda\kappa(2\lambda-3)}$, the numerator is also positive. This implies that a higher wealth-to-GDP ratio is associated with positive net-foreign asset positions. This is what happens in CFG and virtually any existing model of capital flow. When a country needs more store of value, it becomes a net importer of assets or equivalently runs a positive net foreign asset position.

In contrast, when $\phi \geq \bar{\phi}$, the higher wealth-to-GDP ratio leads to negative net-foreign asset positions. This is exactly the home-market effect studied in the international trade literature (Linder, 1961; Krugman, 1980). The higher local demand for store of values generates efficiency gains in the financial sector, which can be strong enough that they become net exporters of assets (i.e. run negative net-foreign asset positions).

However, the wealth-to-GDP ratio is not exogenous in my model. It not only depends on the growth rate, but also on its financial development and interest rates in both countries. Despite these considerations, the capital flow reversal occurs under the same condition:

Proposition 14. *Under the stability condition stated in the Appendix, around a symmetric equilibrium, an increase in growth rate in country i causes capital outflows in country i if and only if*

$\phi > \bar{\phi}$:

$$\frac{d}{dg_i}\left(\frac{NFA_i}{Y_i}\right) > 0 \quad and \quad \frac{d}{dg_i}\left(\frac{CA_i}{Y_i}\right) > 0.$$

The fact that the country has been growing slower implies that it requires more store of value relative to current output. This will induce more banks to enter the local financial sector, which in turn improves the country's efficiency in generating store of values. With returns to scale high enough, the capital flows into the slow growing country in net. In contrast, banks find it less profitable to locate in the fast-growing country because its local demand for store of value is low. The stability condition is required to ensure the supply curve is not steeper than the demand curve.

The result is in sharp contrast to existing theories of capital flow reversal, in which faster growth will, ceteris paribus, cause capital outflows. These theories explain the capital flow reversal by assuming the fast-growing country happens to have a sufficiently less developed financial sector that off sets the growth effect. My result demonstrates that the correlation between slow growth and financial development is not a coincidence but rather is causally linked.

Promoting Domestic Savings. When faced with a surge in the current account deficit, a country may have reasons to employ policies that promote domestic savings if it is concerned with the risk that foreign lending dries up. For example, Obstfeld (2018) writes "countries with lower-than-warranted external current account balances should reduce fiscal deficits and encourage household saving." A corollary of Proposition 14 is that such a policy may backfire. Suppose a government imposes a tax rate, τ_i, on the gross return on savings. For simplicity, assume that the government spends all the collected revenue. The policy will result in one simple modification to the wealth accumulation equation:

$$\frac{dW_{it}}{dt} = (R_{it} - \rho - \theta - \tau_i)W_{it} + Z_{it}.$$

Corollary 1. *Subsidies on domestic savings can increase the current account deficit if $\phi > \bar{\phi}$.*

The underlying mechanisms are the same as Proposition 14. Subsidies on domestic savings increases the demand for the store of value, which in turn will result in more capital inflows if the returns to scale are sufficiently high.

Financial Globalization. Now I consider the effect of financial globalization, which I capture through a reduction in χ. The following proposition examines how financial globalization affects the cross-country differences in financial development and contrasts the effect on global imbalances with the model of exogenous financial development.

Proposition 15. *Consider a symmetric equilibrium with sufficiently small real interest rate, r, and suppose $\phi > \bar{\phi}$. A reduction in the financial transaction costs, χ, magnifies the difference in the financial development between the slow-growing country and the fast-growing country. As a result, the absolute size of global imbalances increases more relative to the model with exogenous financial development.*

Financial globalization magnifies the cross-country differences in financial development, because it becomes less costly to concentrate asset holdings in the relatively more financially developed country, which in turn further increases its efficiency. Although financial globalization increases global imbalances, even in the model of exogenous financial development, the increase is amplified in the model with endogenous financial development.

The above result implies that the interest rate increases in the slow-growing country and decreases in the fast-growing country in response to financial globalization. One might argue that this is inconsistent with the observation that safe interest rates have been declining in most of countries over the past 30 years, in which the financial globalization took place. However, the interest rate in my model should not necessarily be interpreted as the safe interest rate. It captures the investors' perceived return on various financial assets. Iachan, Nenov, and Simsek (2015) argue that when investors' have speculative motives, financial innovation, which increases the number of available risky assets, increases their perceived rate of return while decreasing the safe interest rate.

3.4 Numerical Illustrations

I complement the analytical results with simulation results, which allows me to focus away from the symmetric equilibrium. The goal is not to provide a fully-fledged calibration but to give a rough sense of the quantitative magnitude.

3.4.1 Endogenous Production

I extend the previous model by endogenizing the production side. I assume the production function is Cobb-Douglas in capital and labor:

$$Y_{it} = (K_{it})^{\alpha}(A_{it}L)^{1-\alpha},$$

where A_{it} is labor productivity, which grows at rate $g_{it} \equiv \dot{A}_{it}/A_{it}$. The production function suffers from financial friction where an amount S_{it} of output can be pledged as claims to capital. The firm and labor are owned by newborns. The firm solves

$$\max_{K_{it},L_{it}} (K_{it})^{\alpha}(A_{it}L)^{1-\alpha} - r_{it}K_{it} - w_{it}L$$
$$\text{s.t.} \quad r_{it}K_{it} \le S_{it}.$$

The constraint is binding as long as $S_{it} < \alpha Y_{it}$, which I assume throughout. As in the previous section, capital is intermediated by the financial sector, which determines the financial development, S_{it}. Therefore $S_{it} = \bar{\delta}_i \left(K_{it}/Y_{it}\right)^{\phi} Y_{it}$.

An alternative interpretation of the model is that it features endogenous distortion. Defining the capital wedge, τ_k, as the difference between real interest and the marginal product: $r_{it} = (1-\tau_k)MPK_k$, where $MPK_k \equiv \alpha(K_{it})^{\alpha-1}(A_{it}L)^{1-\alpha}$ is the marginal product of capital, then

$$\tau_k = 1 - \frac{\delta}{\alpha}\left(\frac{K_{it}}{Y_{it}}\right)^{\phi} > 0,$$

which is decreasing in K_{it}/Y_{it}. Therefore, the model can also be interpreted as having endogenous distortion, where distortion is decreasing in the market size. Note that CFG correspond to a fixed distortion.

3.4.2 Parameter Values

Time frequency is annual. I consider a situation where two countries are symmetric and growing at a common rate of $g_1 = g_2 = 2\%$ until $t = 0$. At $t = 0$, country 2 experiences growth acceleration, where it grows at a rate of $g_2 = 4\%$. I se Cobb-Douglas share of capital in the production function to $\alpha = 0.4$, which is standard. Following CFG, I set $\rho + \theta$ to match the wealth to output ratio of 4 in the initial steady state, $\rho + \theta = 23\%$. There is no existing estimate that corresponds to κ in my model, so I simply set $\kappa = 10$. The financial transaction cost, χ, is set to match the average degree of home bias surveyed in Coeurdacier and Rey (2013), $\lambda = 77\%$. This leads to $\chi = 1.13$. I consider two values of ϕ, $\phi = 1.1$ (so that $\kappa(\phi - 1) = 1$) and $\phi = 0$ (CFG case). I chose $\bar{\delta}$ so that the real interest rate in the initial balanced growth path equilibrium is 3%.

3.4.3 Simulation Results

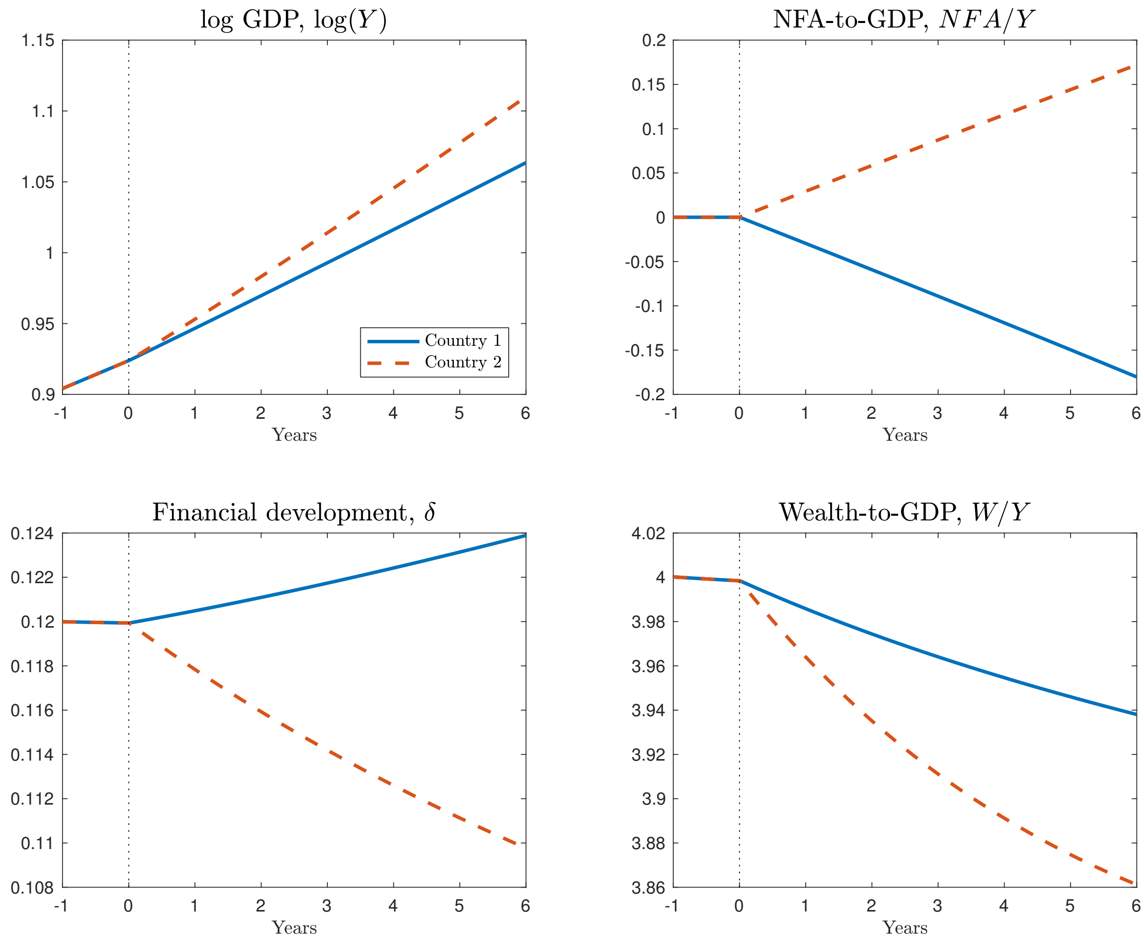


Figure 3.4.1: Growth acceleration: increasing returns in the financial sector ($\phi = 1.1$)

Notes: The underlying scenario is the growth acceleration, where country 2 starts to grow faster at $t = 0$.

Growth Acceleration. Figure 3.4.1 shows the case with increasing returns, $\phi = 1.1$. The right-top panel shows that in response to the growth acceleration of country 2 at $t = 0$, country 2 keeps increasing its net foreign asset positions, while country 1 keeps running negative net foreign asset positions. In short, the faster growth in country 2 causes net capital outflow. The bottom two panels clarify that this result is driven by the endogenous responses of financial developments. The bottom-left panel shows that the financial development (or the fraction of pledgeable output) improves in the slow-growing country 1, but it deteriorates in the fast-growing country 2. The right-bottom panel explains that this is because country 2's wealth-to-GDP ratio decreases relative to country 1. Since the local demand for the store of value decreases, the financial sector finds more profitable to locate in country 1 rather than in country 2, which in turn drives the cross-country differences in financial development.

Figure 3.4.2 shows the result with CFG case ($\phi = 0$). The right-top panel shows that the

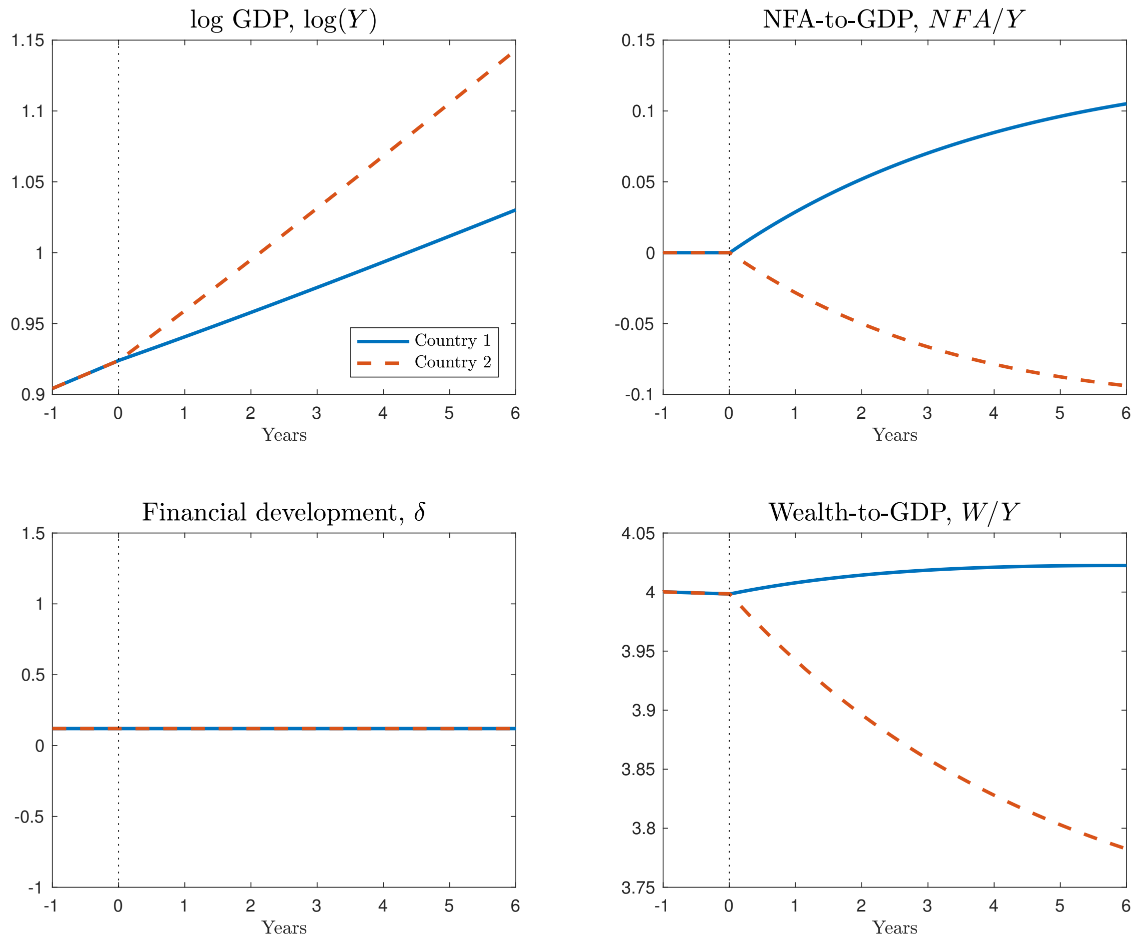


Figure 3.4.2: Growth acceleration: exogenous financial development ($\phi = 0$)

Notes: The underlying scenario is growth acceleration, where country 2 starts to grow faster at $t = 0$. The financial development is constant in both countries by design.

fast-growing country experiences net capital inflow, which is what happens in neoclassical models and is exactly the opposite from the above result. This model has constant financial development by design (the left-bottom panel), and as a result, the relatively lower wealth-to-GDP ratio in fast-growing country implies that it will be a net exporter of assets (negative net foreign asset position). One may notice that the gap in GDP is greater in this case than before (the left-top panel). This is because of the difference in capital flow. Since the fast-growing country attracts more capital when $\phi = 0$, it will grow faster. In contrast, when $\phi = 1.1$, the fast-growing country experiences capital outflow. Therefore, the difference in growth rate is dampened in this case.

Financial Globalization. Figure 3.4.3 illustrates the effect of financial globalization, which is captured through a reduction in χ. The figure considers the growth acceleration episode

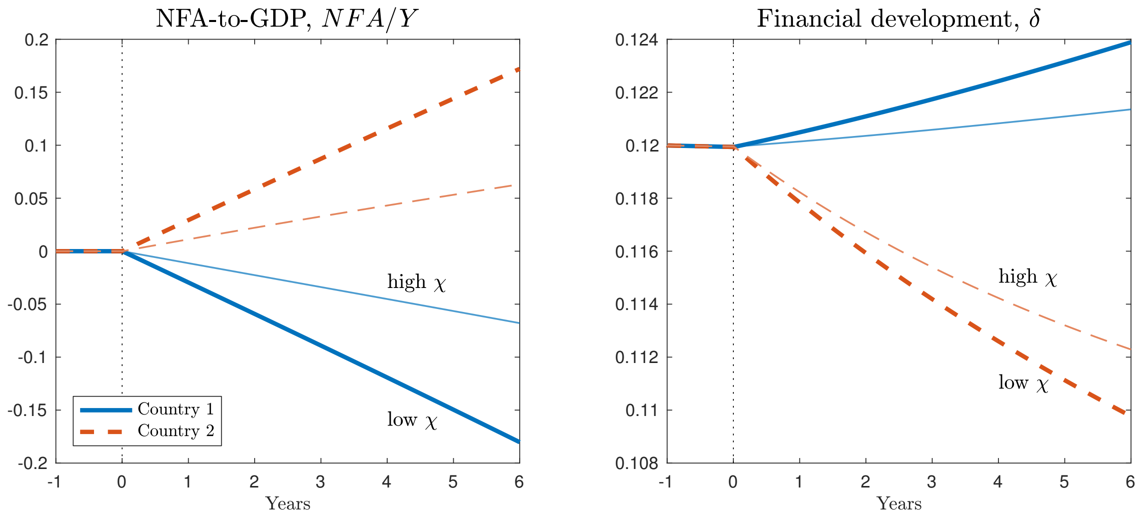


Figure 3.4.3: Effect of financial globalization (lower χ): $\phi = 1.1$

Notes: The underlying scenario is the growth acceleration, where country 2 starts to grow faster at $t = 0$. The dark lines show case with benchmark values of χ, while lighter lines show the case with 10% higher values of χ.

with $\phi = 1.1$ for two different values of χ: the baseline value and 10% higher. When financial transaction cost, χ, is lower, the growth difference has a greater effect on global imbalances (left panel). This is in part because lower financial transaction costs widen the cross-country differences in financial development. Thus, financial globalization makes the cross-country differences in local demand for store of value even more important for local financial development.

3.5 Conclusion

This paper proposed a simple theory in which slower growth causes capital inflows. In explaining why capital tends to flow from fast-growing countries to slow-growing countries, virtually all existing theories assumed that slow-growing countries happen to have relatively well-developed financial markets. I put forward the idea that financial development is demand driven. Without the need to store wealth, no banks would exist. The slow-growing country has relatively more wealth that needs to be stored, which drives demand-induced financial innovation. When the returns to scale are sufficiently high, capital will flow into the slow-growing country. Thus the model provides a simple explanation for the puzzling pattern of capital flow.

Appendix

.1 Proofs

.1.1 Proof of Proposition 14

The equilibrium conditions are

$$\begin{aligned}
k_1 = {} & \frac{(\bar{\delta}(k_1)^{\phi-1})^{\kappa}}{(\bar{\delta}(k_1)^{\phi-1})^{\kappa} + (\bar{\delta}(k_2)^{\phi-1}/\chi)^{\kappa}} \frac{(1-\bar{\delta}(k_1)^{\phi})}{g_1+\rho+\theta-\lambda_{11}r_1-\lambda_{21}r_2/\chi} \\
& + \frac{(\bar{\delta}(k_1)^{\phi-1}/\chi)^{\kappa}}{(\bar{\delta}(k_1)^{\phi-1}/\chi)^{\kappa} + (\bar{\delta}(k_2)^{\phi-1})^{\kappa}} \frac{(1-\bar{\delta}(k_2)^{\phi})}{g_2+\rho+\theta-\lambda_{11}r_1/\chi-\lambda_{22}r_2} \\
k_2 = {} & \frac{(\bar{\delta}(k_2)^{\phi-1}/\chi)^{\kappa}}{(\bar{\delta}(k_1)^{\phi-1})^{\kappa} + (\bar{\delta}(k_2)^{\phi-1}/\chi)^{\kappa}} \frac{(1-\bar{\delta}(k_1)^{\phi})}{g_1+\rho+\theta-\lambda_{11}r_1-\lambda_{21}r_2/\chi}, \\
& + \frac{(\bar{\delta}(k_2)^{\phi-1})^{\kappa}}{(\bar{\delta}(k_1)^{\phi-1}/\chi)^{\kappa} + (\bar{\delta}(k_2)^{\phi-1})^{\kappa}} \frac{(1-\bar{\delta}(k_2)^{\phi})}{g_2+\rho+\theta-\lambda_{11}r_1/\chi-\lambda_{22}r_2}
\end{aligned}$$

where $k_i \equiv K_i/Y$. Taking log-derivatives, we have

$$\begin{aligned}
d\ln k_1 &= a d\ln k_1 + b d\ln k_2 - \lambda c dg_1 - (1-\lambda) c dg_2 \\
d\ln k_2 &= b d\ln k_1 + a d\ln k_2 - (1-\lambda) c dg_1 - \lambda c dg_2,
\end{aligned}$$

where

$$\begin{aligned}
a &\equiv 2\lambda(1-\lambda)\kappa(\phi-1) + \lambda f_1 + (1-\lambda) f_2 \\
b &\equiv -2\lambda(1-\lambda)\kappa(\phi-1) + (1-\lambda) f_1 + \lambda f_2 \\
c &\equiv \frac{1}{g_i+\rho+\theta-\lambda r-(1-\lambda)r/\chi} \\
f_1 &\equiv -\frac{\phi\bar{\delta}(k)^{\phi}}{(1-\bar{\delta}(k)^{\phi})} \\
&\quad + \frac{\lambda r}{g+\rho+\theta-\lambda r-(1-\lambda)r/\chi}(\phi-1)((1-\lambda)\kappa+1) - \frac{(1-\lambda)r/\chi}{g+\rho+\theta-\lambda r-(1-\lambda)r/\chi}\lambda\kappa(\phi-1) \\
f_2 &\equiv -\frac{\lambda r}{g+\rho+\theta-\lambda r-(1-\lambda)r/\chi}(1-\lambda)\kappa(\phi-1) + \frac{(1-\lambda)r/\chi}{g+\rho+\theta-\lambda r-(1-\lambda)r/\chi}(\phi-1)(\lambda\kappa+1)
\end{aligned}$$

where $\lambda \equiv \lambda_{11} = \lambda_{22}$. Solving for $d\ln k_1$ and $d\ln k_2$ give

$$d\ln k_1 = -\frac{1}{(1-a)^2 - b^2}\left[(1-a)\lambda + b(1-\lambda)\right] cdg_1 - \frac{1}{(1-a)^2 - b^2}\left[(1-a)(1-\lambda) + b\lambda\right] cdg_2$$
$$d\ln k_2 = -\frac{1}{(1-a)^2 - b^2}\left[(1-a)(1-\lambda) + b\lambda\right] cdg_1 - \frac{1}{(1-a)^2 - b^2}\left[(1-a)\lambda + b(1-\lambda)\right] cdg_2.$$

Assuming $dg_2 = 0$ without loss of generality, we have

$$d\ln k_1 - d\ln k_2 = \frac{1}{(1-4\lambda(1-\lambda)\kappa(\phi-1) + (2\lambda-1)(f_2-f_1))}(1-2\lambda)cdg_1. \qquad (.1.1)$$

The net foreign asset position of country 1 is

$$\begin{aligned} d\log(A_{21}/Y_1) - d\log(A_{12}/Y_1) &= -\lambda\kappa(\phi-1)(d\log k_1 - d\log k_2) - (f_2-f_1)(d\log k_1 - d\log k_2) - cdg_1 \\ &= -\left(\frac{\lambda\kappa(\phi-1)(2\lambda-3)+1}{(1-4\lambda(1-\lambda)\kappa(\phi-1)+(2\lambda-1)(f_2-f_1))}\right)cdg_1. \end{aligned}$$

The stability condition requires the denominator to be positive

$$(1-4\lambda(1-\lambda)\kappa(\phi-1) + (2\lambda-1)(f_2-f_1)) > 0,$$

which holds when the steady state interest rate r is sufficiently low and $\kappa(\phi-1) \leq 1$. Under this condition, an increase in the growth rate leads to positive net foreign asset position if and only if $\phi > \bar{\phi} \equiv 1 + \frac{1}{\lambda\kappa(2\lambda-3)}$.

.1.2 Proof of Proposition 15

The financial development in each country is $\delta_i = \bar{\delta}(k_i)^{\phi}$, where $k_i \equiv K_i/Y_i$. From expression (.1.1), it is sufficient to show

$$\alpha(\lambda) \equiv \frac{1}{(1-4\lambda(1-\lambda)\kappa(\phi-1) + (2\lambda-1)(f_2-f_1))}(1-2\lambda)$$

is decreasing in λ (since λ is decreasing in χ). With sufficiently small r, since we have

$$f_2 - f_1 \approx \frac{\phi\bar{\delta}(k)^{\phi}}{(1-\bar{\delta}(k)^{\phi})},$$

α is indeed decreasing in λ. This shows that the difference in the financial development increases with χ.

Now consider an environment with exogenous financial development. One can show that for sufficiently small r,

$$dA_{21} - dA_{12} \approx \beta_\delta d \ln(\bar{\delta}_1/\bar{\delta}_2) - \beta_g dg_1,$$

where

$$\beta_\delta \equiv \frac{\lambda(1-\lambda)\kappa}{1+4\lambda(1-\lambda)\kappa} + (1-\lambda)\frac{\lambda\kappa(3-2\lambda)+1}{1+4\lambda(1-\lambda)\kappa}\frac{\bar{\delta}}{1-\bar{\delta}} > 0$$
$$\beta_g \equiv \frac{\lambda\kappa(3-2\lambda)+1}{1+4\lambda(1-\lambda)\kappa}(1-\lambda) > 0.$$

Both β_δ and β_g are decreasing in λ. In the case of of endogenous financial development, $d\ln(\bar{\delta}_1/\delta_2)$ also decreases. Therefore the effect of financial liberalization is amplified.

Bibliography

ACEMOGLU, D. (2001): "Good Jobs versus Bad Jobs," *Journal of Labor Economics*, 19, 1–21.

ACEMOGLU, D. AND R. SHIMER (1999): "Holdups and Efficiency with Search Frictions," *International Economic Review*, 40, 827–849.

ADDARIO, S. D., P. KLINE, R. SAGGIO, AND M. SØLVSTEN (2020): "It Ain't Where You're from, It's Where You're at: Hiring Origins, Firm Heterogeneity, and Wages," *Working paper*.

AGHION, P. AND P. HOWITT (1992): "A Model of Growth through Creative Destruction," *Econometrica*, 60.

ANDERSON, R. W. AND K. JOEVEER (2012): "Bankers and bank investors: Reconsidering the economies of scale in banking," .

ANGELETOS, G.-M. AND V. PANOUSI (2011): "Financial integration, entrepreneurial risk and global dynamics," *Journal of Economic Theory*, 146, 863–896.

ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.

AUCLERT, A., B. BARDÓCZY, M. ROGNLIE, AND L. STRAUB (2019): "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," Working Paper 26123, National Bureau of Economic Research.

AUTOR, D., D. DORN, G. H. HANSON, G. PISANO, AND P. SHU (2017): "Foreign Competition and Domestic Innovation: Evidence from U.S. Patents," *mimeo MIT*.

AZARIADIS, C. (1975): "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, 83, 1183–1202.

BAGGER, J., F. FONTAINE, M. GALENIANOS, AND I. TRAPEZNIKOVA (2020): "Vacancies, Employment Outcomes and Firm Growth: Evidence from Denmark," SSRN Scholarly Paper ID 3599258, Social Science Research Network, Rochester, NY.

BAGGER, J., F. FONTAINE, F. POSTEL-VINAY, AND J.-M. ROBIN (2014): "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 104, 1551–1596.

BAILY, M. N. (1974): "Wages and Employment under Uncertain Demand," *The Review of Economic Studies*, 41, 37–50.

BALASSA, B. (1965): "Trade Liberalisation and "Revealed" Comparative Advantage," *The Manchester School*, May, 99–123.

BALEY, I., A. FIGUEIREDO, AND R. ULBRICHT (2019): "Mismatch Cycles," SSRN Scholarly Paper ID 3441280, Social Science Research Network, Rochester, NY.

BARLEVY, G. (2002): "The Sullying Effect of Recessions," *The Review of Economic Studies*, 69, 65–96.

BARNICHON, R. (2010): "Building a Composite Help-Wanted Index," *Economics Letters*, 109, 175–178.

BARRO, R. J. (1977): "Long-Term Contracting, Sticky Prices, and Monetary Policy," *Journal of Monetary Economics*, 3, 305–316.

BARTELME, D., A. COSTINOT, D. DONALDSON, AND A. RODRÍGUEZ-CLARE (2019a): "The Textbook Case for Industrial Policy: Theory Meets Data," *NBER Working Paper 26193*.

BARTELME, D., T. LAN, AND A. LEVCHENKO (2019b): "Specialization, Market Access and Real Income," *mimeo University of Michigan*.

BASU, S. AND C. L. HOUSE (2016): "Chapter 6 - Allocative and Remitted Wages: New Facts and Challenges for Keynesian Models," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 297–354.

BEAUDRY, P. AND J. DINARDO (1991): "The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data," *Journal of Political Economy*, 99, 665–688.

BERGER, D. W., K. F. HERKENHOFF, AND S. MONGEY (2019): "Labor Market Power," Working Paper 25719, National Bureau of Economic Research.

BEWLEY, T. F. (1999): *Why Wages Don't Fall during a Recession*, Harvard University Press.

BILAL, A., N. ENGBOM, S. MONGEY, AND G. VIOLANTE (2019): "Firm and Worker Dynamics in a Frictional Labor Market," Tech. Rep. w26547, National Bureau of Economic Research, Cambridge, MA.

BILBIIE, F. O. (2020): "Monetary Neutrality with Sticky Prices and Free Entry," *Review of Economics and Statistics*.

BILS, M., Y. CHANG, AND S.-B. KIM (2016): "How Sticky Wages In Existing Jobs Can Affect Hiring," Working Paper 2016rwp-91, Yonsei University, Yonsei Economics Research Institute.

BILS, M. J. (1985): "Real Wages over the Business Cycle: Evidence from Panel Data," *Journal of Political Economy*, 93, 666–689.

BLANCHARD, O. AND J. GALÍ (2010): "Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment," *American Economic Journal: Macroeconomics*, 2, 1–30.

BLANCHARD, O. J. (1985): "Debt, deficits, and finite horizons," *Journal of political economy*, 93, 223–247.

BLANCHARD, O. J. AND P. DIAMOND (1989): "The Beveridge Curve," *Brookings Papers on Economic Activity*, 1989, 1–76.

BLOESCH, J. AND B. TASKA (2019): "Which Workers' Wages Track Productivity? The Role of Position Specificity and On-the-Job Search," 44.

BLOOM, N., M. DRACE, AND J. V. REENEN (2016): "Trade induced technical change? The impact of Chinese imports on innovation, IT and productivity," *Review of Economic Studies*, 83, 87–117.

BOLDRIN, M. AND J. A. SCHEINKMAN (1988): "Learning-By-Doing, International Trade, and Growth: A Note," in *The Economy as an Evolving Complex System*, Perseus Book Publishing L.L.C, SFI Studies in the Sciences of Complexity.

BOLLARD, A., P. J. KLENOW, AND H. LI (2016): "Entry costs rise with development," *Working Paper*.

BONTEMPS, C., J.-M. ROBIN, AND G. J. V. DEN BERG (1999): "An Empirical Equilibrium Job Search Model With Search on the Job and Heterogeneous Workers and Firms," *International Economic Review*, 40, 1039–1074.

BOROVIČKA, J. AND K. BOROVIČKOVÁ (2018): "Risk Premia and Unemployment Fluctuations," .

BREZA, E., S. KAUR, AND Y. SHAMDASANI (2018): "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 133, 611–663.

BRODA, C. AND D. WEINSTEIN (2006): "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121, 541–585.

BRONER, F., T. DIDIER, A. ERCE, AND S. L. SCHMUKLER (2013): "Gross capital flows: Dynamics and crises," *Journal of Monetary Economics*, 60, 113–133.

BUERA, F. J. AND Y. SHIN (2017): "Productivity growth and capital flows: The dynamics of reforms," *American Economic Journal: Macroeconomics*, 9, 147–85.

BUERA, P. AND E. OBERFIELD (2017): "The Global Diffusion of Ideas," *mimeo Princeton University*.

BURDETT, K. AND M. COLES (2003): "Equilibrium Wage-Tenure Contracts," *Econometrica*, 71, 1377–1404.

——— (2010): "Wage/Tenure Contracts with Heterogeneous Firms," *Journal of Economic Theory*, 145, 1408–1435.

BURDETT, K. AND D. T. MORTENSEN (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257–273.

BURSTEIN, A. AND A. ATKESON (2019): "Aggregate Implications of Innovation Policy," *Journal of Political Economy*, 127.

BURSTEIN, A. AND G. GOPINATH (2014): "Chapter 7 - International Prices and Exchange Rates," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Elsevier, vol. 4 of *Handbook of International Economics*, 391–451.

CABALLERO, R. J., E. FARHI, AND P.-O. GOURINCHAS (2008): "An Equilibrium Model of," *The American Economic Review*, 98, 358–393.

CAI, X. (2020): "Efficiency of Wage Bargaining with On-the-Job Search," *International Economic Review*, 61, 1749–1775.

CALDWELL, S. AND N. HARMON (2019): "Evidence from Coworker Networks," 105.

CARD, D., A. MAS, E. MORETTI, AND E. SAEZ (2012): "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review*, 102, 2981–3003.

CARLSSON, M. AND A. WESTERMARK (2016): "Endogenous Separations, Wage Rigidities and Employment Volatility," SSRN Scholarly Paper ID 2871910, Social Science Research Network, Rochester, NY.

CHODOROW-REICH, G. AND L. KARABARBOUNIS (2016): "The Cyclicality of the Opportunity Cost of Employment," *Journal of Political Economy*, 124, 1563–1618.

COEURDACIER, N., S. GUIBAUD, AND K. JIN (2015): "Credit constraints and growth in a global economy," *American Economic Review*, 105, 2838–81.

COEURDACIER, N. AND H. REY (2013): "Home bias in open economy financial macroeconomics," *Journal of Economic Literature*, 51, 63–115.

COSTINOT, A. (2009): "On the Origins of Comparative Advantage," *Journal of International Economics*, 77, 255–264.

COSTINOT, A., D. DONALDSON, AND I. KOMUNJER (2012): "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies*, 79, 581–608.

COSTINOT, A., D. DONALDSON, N. M. KYLE, AND H. WILLIAMS (2016): "The More We Die, The More We Sell? A Simple Test of the Home-Market Effect," .

COSTINOT, A. AND A. RODRÍGUEZ-CLARE (2013): "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, New York: Elsevier, vol. 4.

DAVIS, S. J., R. J. FABERMAN, AND J. C. HALTIWANGER (2013): "The Establishment-Level Behavior of Vacancies and Hiring," *The Quarterly Journal of Economics*, 128, 581–622.

DIAMOND, D. W. AND P. H. DYBVIG (1983): "Bank runs, deposit insurance, and liquidity," *Journal of political economy*, 91, 401–419.

DIAMOND, P. A. (1982): "Wage Determination and Efficiency in Search Equilibrium," *The Review of Economic Studies*, 49, 217–227.

DIXIT, A. K. AND J. E. STIGLITZ (1977): "Monopolistic competition and optimum product diversity," *The American Economic Review*, 67, 297–308.

DONOVAN, K., J. LU, AND T. SCHOELLMAN (2018): "Labor Market Flows and Development," Tech. Rep. 976, Society for Economic Dynamics.

DUBE, A., L. GIULIANO, AND J. LEONARD (2019): "Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior," *American Economic Review*, 109, 620–663.

DUPRAZ, S., E. NAKAMURA, AND J. STEINSSON (2019): "A Plucking Model of Business Cycles," Working Paper 26351, National Bureau of Economic Research.

EECKHOUT, J. AND I. LINDENLAUB (2019): "Unemployment Cycles," *American Economic Journal: Macroeconomics*, 11, 175–234.

ELIAZ, K. AND R. SPIEGLER (2014): "Reference Dependence and Labor Market Fluctuations," *NBER Macroeconomics Annual*, 28, 159–200.

ENGBOM, N. (2019): "Firm and Worker Dynamics in an Aging Labor Market," Preprint, Working Paper.

——— (2020): "Recruiting Talent During a Recession," *Working paper*.

ENGBOM, N. AND C. MOSER (2017): "Earnings Inequality and the Minimum Wage: Evidence from Brazil," SSRN Scholarly Paper ID 2940013, Social Science Research Network, Rochester, NY.

FABERMAN, R. J., A. I. MUELLER, A. ŞAHIN, AND G. TOPA (2020): "Job Search Behavior among the Employed and Non-Employed," Working Paper 23731, National Bureau of Economic Research.

FAJGELBAUM, P., G. M. GROSSMAN, AND E. HELPMAN (2011): "Income distribution, product quality, and international trade," *Journal of political Economy*, 119, 721–765.

FEENSTRA, R. C., R. E. LIPSEY, H. DENG, A. C. MA, AND H. MO (2005): "World Trade Flows: 1962-2000," *NBER Working Paper 11040*.

FOSTER, L., C. GRIM, J. HALTIWANGER, AND Z. WOLF (2016): "Firm-Level Dispersion in Productivity: Is the Devil in the Details?" *The American Economic Review*, 106, 95–98.

GAUTIER, P. A., C. N. TEULINGS, AND A. V. VUUREN (2010): "On-the-Job Search, Mismatch and Efficiency," *Review of Economic Studies*, 77, 245–272.

GERTLER, M., C. HUCKFELDT, AND A. TRIGARI (2020): "Unemployment Fluctuations, Match Quality, and the Wage Cyclicality of New Hires," *The Review of Economic Studies*, 39.

GERTLER, M. AND A. TRIGARI (2009): "Unemployment Fluctuations with Staggered Nash Wage Bargaining," *Journal of Political Economy*, 117, 38–86.

GOLOSOV, M., P. MAZIERO, AND G. MENZIO (2013): "Taxation and Redistribution of Residual Income Inequality," *Journal of Political Economy*, 121, 1160–1204.

GOUIN-BONENFANT, E. (2020): "Productivity Dispersion, Between-Firm Competition and the Labor Share," Tech. rep.

GOURINCHAS, P.-O. AND O. JEANNE (2013): "Capital flows to developing countries: The allocation puzzle," *Review of Economic Studies*, 80, 1484–1515.

GOURINCHAS, P.-O. AND H. REY (2014): "External Adjustment, Global Imbalances, Valuation Effects," *Handbook of International Economics*, 4, 585–645.

GRIGSBY, J., E. HURST, AND A. YILDIRMAZ (2019): "Aggregate Nominal Wage Adjustments: New Evidence from Administrative Payroll Data," Working Paper 25628, National Bureau of Economic Research.

GROSSMAN, G. M. AND E. HELPMAN (1990): "Trade, Innovation, and Growth," *American Economic Review Papers and Proceedings*, 80, 86–91.

——— (1995): "Technology and Trade," in *Handbook of International Economics*, ed. by G. M. Grossman and K. Rogoff, Elsevier, vol. 3.

GUTIÉRREZ, G. AND T. PHILIPPON (2018): "How European Markets Became Free: A Study of Institutional Drift," Working Paper 24700, National Bureau of Economic Research.

HAEFKE, C., M. SONNTAG, AND T. VAN RENS (2013): "Wage Rigidity and Job Creation," *Journal of Monetary Economics*, 60, 887–899.

HAGEDORN, M. AND I. MANOVSKII (2008): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited," *American Economic Review*, 98, 1692–1706.

HALL, R. E. (2017): "High Discounts and High Unemployment," *American Economic Review*, 107, 305–330.

Hall, R. E. and A. B. Krueger (2012): "Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search," *American Economic Journal: Macroeconomics*, 4, 56–67.

Hall, R. E. and P. R. Milgrom (2008): "The Limited Influence of Unemployment on the Wage Bargain," *American Economic Review*, 98, 1653–1674.

Haltiwanger, J. C., H. R. Hyatt, L. B. Kahn, and E. McEntarfer (2018): "Cyclical Job Ladders by Firm Size and Firm Wage," *American Economic Journal: Macroeconomics*, 10, 52–85.

Hanson, G. H., N. Lind, and M.-A. Muendler (2016): "The Dynamics of Comparative Advantage," *mimeo UCSD*.

Harris, M. and B. Holmstrom (1982): "A Theory of Wage Dynamics," *The Review of Economic Studies*, 49, 315–333.

Hausman, R., C. A. Hidalgo, S. Bustos, M. Coscia, S. Chung, J. Jimenez, A. Simoes, and M. Yildirim (2013): *The Atlas of Economic Complexity: Mapping Paths to Prosperity*, MIT press, 2 ed.

Hausman, R., J. Hwang, and D. Rodrik (2007): "What you export matters," *Journal of Economic Growth*, 12, 1–25.

Hazell, J. and B. Taska (2019): "Downward Rigidity in the Wage for New Hires," *Mimeo, MIT Economics Department*.

Heise, S. and T. Porzio (2019): "Spatial Wage Gaps and Frictional Labor Markets," SSRN Scholarly Paper ID 3467682, Social Science Research Network, Rochester, NY.

Herrendorf, B., R. Rogerson, and A. Valentinyi (2013): "Two Perspectives on Preferences and Structural Transformation," *American Economic Review*, 103, 2752–2789.

Hidalgo, C. A. and R. Hausman (2009): "The Building Blocks of Economic Complexity," *Proceedings of the National Academy of Sciences*, 106, 10570–10575.

Hosios, A. J. (1990): "On the Efficiency of Matching and Related Models of Search and Unemployment," *The Review of Economic Studies*, 57, 279–298.

Hughes, J. P. and L. J. Mester (2013): "Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function," *Journal of Financial Intermediation*, 22, 559–585.

IACHAN, F. S., P. T. NENOV, AND A. SIMSEK (2015): "The Choice Channel of Financial Innovation," .

JAROSCH, G. (2015): "Searching for Job Security and the Consequences of Job Loss," 57.

JAROSCH, G., J. S. NIMCZIK, AND I. SORKIN (2019): "Granular Search, Market Structure, and Wages," Tech. Rep. w26239, National Bureau of Economic Research, Cambridge, MA.

JIN, K. (2012): "Industrial structure and capital flows," *The American Economic Review*, 102, 2111–2146.

JONES, C. I. (1999): "Growth: With or Without Scale Effects?" *American Economic Review, Papers and Proceedings*, 89, 139–144.

KAAS, L. AND P. KIRCHER (2015): "Efficient Firm Dynamics in a Frictional Labor Market," *American Economic Review*, 105, 3030–3060.

KARAHAN, F., S. OZKAN, AND J. SONG (2019): "Anatomy of Lifetime Earnings Inequality: Heterogeneity in Job Ladder Risk vs. Human Capital," *SSRN Electronic Journal*.

KEHOE, P. J., P. LOPEZ, V. MIDRIGAN, AND E. PASTORINO (2019): "Asset Prices and Unemployment Fluctuations," 54.

KENNAN, J. (2010): "Private Information, Wage Bargaining and Employment Fluctuations," *The Review of Economic Studies*, 77, 633–664.

KRUGMAN, P. (1979): "A Model of Innovation, Technology Transfer, and the World Distribution of Income," *Journal of Political Economy*, 87, 253–266.

——— (1980): "Scale economies, product differentiation, and the pattern of trade," *The American Economic Review*, 70, 950–959.

——— (1986): *A 'Technology Gap' Model of International Trade*, Structural Adjustment in Advanced Economies, London: Mcmillan.

——— (1987): "The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher," *Journal of Development Economics*, 27, 41–55.

KUCHERYAVYY, K., G. LYN, AND A. RODRÍGUEZ-CLARE (2016): "Grounded by gravity: A well-behaved trade model with industry-level economies of scale," .

KUDLYAK, M. (2014): "The Cyclicality of the User Cost of Labor," *Journal of Monetary Economics*, 68, 53–67.

LAMADON, T. (2016): "Productivity Shocks, Long-Term Contracts and Earnings Dynamics," *Working paper*.

LAMADON, T., M. MOGSTAD, AND B. SETZLER (2019): "Imperfect Competition, Compensating Differentials and Rent Sharing in the U.S. Labor Market," Tech. Rep. w25954, National Bureau of Economic Research.

LEDERMAN, D. AND W. F. MALONEY (2012): *Does What You Export Matter? In Search of Empirical Guidance for Industrial Policies*, Washington, D.C.: The World Bank.

LEVCHENKO, A. A. (2007): "Institutional Quality and International Trade," *Review of Economic Studies*, 74, 791–819.

LEVCHENKO, A. A. AND J. ZHANG (2016): "The Evolution of Comparative Advantage: Measurement and Welfare Implications," *Journal of Monetary Economics*, 78, 96–111.

LINDER, S. B. (1961): *An essay on trade and transformation*, Wiley.

LISE, J. AND J.-M. ROBIN (2017): "The Macrodynamics of Sorting between Workers and Firms," *American Economic Review*, 107, 1104–1135.

LJUNGQVIST, L. AND T. J. SARGENT (2017): "The Fundamental Surplus," *American Economic Review*, 107, 2630–2665.

LUENBERGER, D. G. (1969): "Optimization by Vector Space Methods | Wiley," https://www.wiley.com/en-us/Optimization+by+Vector+Space+Methods-p-9780471181170.

MANNING, A. (2003): *Monopsony in Motion*.

MARTELLINI, P., G. MENZIO, AND L. VISSCHERS (2020): "Revisiting the Hypothesis of High Discounts and High Unemployment," *Working paper*.

MARTIN, P. AND H. REY (2004): "Financial super-markets: size matters for asset trade," *Journal of International Economics*, 64, 335–361.

MAS-COLELL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic Theory*, Oxford University Press.

MATSUYAMA, K. (2004): "Financial Market Globalization, Symmetry-Breaking, and Endogenous Inequality of Nations," *Econometrica*, 72, 853–884.

——— (2005): "Credit Market Imperfections and Patterns of International Trade and Capital Flows," *Journal of the European Economic Association*, 3, 714–723.

——— (2013): "Endogenous Ranking and Equilibrium Lorentz Curve Across (Ex Ante) Identical Countries," *Econometrica*, 81, 2009–2031.

——— (2018): "Engel's Law in the Global Economy: Demand-induced Patterns of Structural Change, Innovation, and Trade," *Econometrica*.

MELITZ, M. AND A. CUNAT (2012): "Volatility, labor market flexibility, and the pattern of comparative advantage," *Journal of the European Economic Association*, 10, 225–254.

MELITZ, M. J. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 71, 1695–1725.

MENDOZA, E. G., V. QUADRINI, AND J.-V. RIOS-RULL (2009): "Financial integration, financial development, and global imbalances," *Journal of Political economy*, 117, 371–416.

MENZIO, G. (2004): "High-Frequency Wage Rigidity," *Working paper*.

MENZIO, G. AND E. R. MOEN (2010): "Worker Replacement," *Journal of Monetary Economics*, 57, 623–636.

MENZIO, G. AND S. SHI (2011): "Efficient Search on the Job and the Business Cycle," *Journal of Political Economy*, 119, 468–510.

MOEN, E. R. (1997): "Competitive Search Equilibrium," *Journal of Political Economy*, 105, 385–411.

MONGEY, S. (2017): "Market Structure and Monetary Non-Neutrality," Tech. Rep. 558, Federal Reserve Bank of Minneapolis.

MORTENSEN, D. (1982): "The Matching Process as a Noncooperative Bargaining Game," NBER Chapters, National Bureau of Economic Research, Inc.

MOSCARINI, G. AND F. POSTEL-VINAY (2016a): "Did the Job Ladder Fail after the Great Recession?" *Journal of Labor Economics*, 34, 55–93.

——— (2016b): "Wage Posting and Business Cycles: A Quantitative Exploration," *Review of Economic Dynamics*, 19, 135–160.

——— (2018): "On the Job Search and Business Cycles," SSRN Scholarly Paper ID 3261712, Social Science Research Network, Rochester, NY.

——— (2019): "The Job Ladder: Inflation vs. Reallocation," .

MUKOYAMA, T. (2009): "A Note on Cyclical Discount Factors and Labor Market Volatility," .

——— (2014): "The Cyclicality of Job-to-Job Transitions and Its Implications for Aggregate Productivity," *Journal of Economic Dynamics and Control*, 39, 1–17.

NAKAMURA, A., E. NAKAMURA, K. PHONG, AND J. STEINSSON (2019): "New Evidence on the Cyclicality of Employer-to-Employer Flows from Canada," *AEA Papers and Proceedings*, 109, 456–460.

NICKELL, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

OBERFIELD, E. AND D. RAVAL (2014): "Micro Data and Macro Technology," *NBER Working Paper 20452*, princeton working paper.

OBSTFELD, M. (2018): "Addressing Global Imbalances Requires Cooperation," https://blogs.imf.org/2018/07/24/addressing-global-imbalances-requires-cooperation/, [Online; accessed 26-January-2019].

OKAWA, Y. AND E. VAN WINCOOP (2012): "Gravity in international finance," *Journal of international Economics*, 87, 205–215.

PERLA, J., C. TONETTI, AND M. E. WAUGH (2015): "Equilibrium Technology Diffusion, Trade, and Growth," *NBER Work. Pap. 20881*.

PHELAN, G. (2017): "Correlated default and financial intermediation," *The Journal of Finance*, 72, 1253–1284.

PISSARIDES, C. A. (1985): "Short-Run Equilibrium Dynamics of Unemployment Vacancies, and Real Wages," *American Economic Review*, 75, 676–690.

——— (2009): "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?" *Econometrica*, 77, 1339–1369.

REDDING, S. J. AND D. E. WEINSTEIN (2018): "Accounting for Trade Patterns," *NBER Working Paper 24051*.

REITER, M. (2009): "Solving Heterogeneous-Agent Models by Projection and Perturbation," *Journal of Economic Dynamics and Control*, 33, 649–665.

ROMER, P. M. (1990): "Endogenous technological change," *Journal of political Economy*, 98, S71–S102.

ROTEMBERG, J. J. (1983): "Aggregate Consequences of Fixed Costs of Price Adjustment [Sticky Prices and Disequilibrium Adjustment in a Rational Model of the Inflationary Process]," *American Economic Review*, 73, 433–436.

RUDANKO, L. (2009): "Labor Market Dynamics under Long-Term Wage Contracting," *Journal of Monetary Economics*, 56, 170–183.

——— (2019): "Firm Wages in a Frictional Labor Market," SSRN Scholarly Paper ID 3321608, Social Science Research Network, Rochester, NY.

SAMPSON, T. (2016): "Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth," *Quarterly Journal of Economics*, 131, 315–380.

SAMUELSON, P. A. (1939): "The Gains from International Trade," *Canadian Journal of Economics*, 5, 195–205.

SCHAAL, E. (2017): "Uncertainty and Unemployment," *Econometrica*, 85, 1675–1721.

SCHETTER, U. (2019): "A Structural Ranking of Economic Complexity," *CID Working Paper No. 119*.

——— (2020): "Quality Differentiation, Comparative Advantage, and International Specialization Across Products," *CID Working Paper No. 126*.

SCHOEFER, B. (2016): "The Financial Channel of Wage Rigidity," Tech. Rep. 1605, Society for Economic Dynamics.

SHIMER, R. (2005): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95, 25–49.

——— (2010): "Labor Markets and Business Cycles," Tech. rep., Princeton University Press.

——— (2012): "Reassessing the Ins and Outs of Unemployment," *Review of Economic Dynamics*, 15, 127–148.

SNELL, A. AND J. P. THOMAS (2010): "Labor Contracts, Equal Treatment, and Wage-Unemployment Dynamics," *American Economic Journal: Macroeconomics*, 2, 98–127.

SONG, Z., K. STORESLETTEN, AND F. ZILIBOTTI (2011): "Growing like china," *The American Economic Review*, 101, 196–233.

STOKEY, N. (1991): "Human Capital, Product Quality, and Growth," *Quarterly Journal of Economics*, 106, 587–616.

SUTTON, J. AND D. TREFLER (2016): "Capabilities, Wealth, and Trade," *Journal of Political Economy*, 124, 826–878.

THOMAS, J. AND T. WORRALL (1988): "Self-Enforcing Wage Contracts," *The Review of Economic Studies*, 55, 541–554.

VAN DEN BERG, G. J. AND G. RIDDER (1998): "An Empirical Equilibrium Search Model of the Labor Market," *Econometrica*, 66, 1183–1221.

WANG, O. AND I. WERNING (2020): "Dynamic Oligopoly and Price Stickiness," *National Bureau of Economic Research Working Paper Series*.

WHEELOCK, D. C. AND P. W. WILSON (2012): "Do large banks have lower costs? New estimates of returns to scale for US banks," *Journal of Money, Credit and Banking*, 44, 171–199.

WRIGHT, R., P. KIRCHER, B. JULIEN, AND V. GUERRIERI (2019): "Directed Search and Competitive Search Equilibrium: A Guided Tour," *Journal of Economic Literature*.

YAARI, M. E. (1965): "Uncertain lifetime, life insurance, and the theory of the consumer," *The Review of Economic Studies*, 32, 137–150.

YAMAGUCHI, S. (2010): "Job Search, Bargaining, and Wage Dynamics," *Journal of Labor Economics*, 28, 595–631.

YASHIV, E. (2000): "Hiring as Investment Behavior," *Review of Economic Dynamics*, 3, 486–522.

YOUNG, A. (1991): "Learning by Doing and the Dynamic Effects of International Trade," *Quarterly Journal of Economics*, 106, 396–406.