Essays in Macroeconomics and International Economics

By

Dohyeon Lee

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN – MADISON

2020

Date of final oral examination: June 5, 2020

The dissertation is approved by the following members of the Final Oral Committee:

Charles Engel, Professor, Economics

Kenneth D. West, Professor, Economics

Menzie D. Chinn, Professor, Public Affairs and Economics

Dmitry Mukhin, Assistant Professor, Economics

Abstract

The first chapter [**"Optimal Monetary Policy in an Open Economy with Shocks to UIP"**] studies optimal cooperative monetary policy between two symmetric countries where shocks to UIP (uncovered interest parity) lead to deviation from the UIP condition. UIP shock results in welfare loss because it distorts the relative consumption between the two countries, which would propagate into inefficient levels of output. Optimal monetary policy, while unable to affect the path of relative consumption, can improve efficiency compared to the flexible price allocation by reducing the distortions in output at the expense of a modest increase in price dispersion. Optimal capital control, in the form of discriminating the interest rates faced by the households and the financial intermediaries, would nullify the impact of UIP shock.

The second chapter [**"Offshoring and Segregation by Skill: Theory and Evidence"**] **(with Gueyon Kim)** examines the labor market consequences of offshoring. We use the Danish employer-employee matched data together with the newly constructed skill measures to evaluate the effect of offshoring on wages and reallocation of workers within offshorable occupations. Offshoring reduces domestic worker wages; and increases the probability of reallocation away from the high-productivity firms to the low-productivity ones. The least skilled workers further face a greater risk of switching out to a less competitive sector. On the firm-side, offshoring improves the average skill of in-house workers at a lower cost. By estimating a worker-firm matching model, we examine the mechanisms of how offshoring affects labor market inequality and further

assess the quantitative importance of various competing hypotheses such as technological change and the expansion of higher education, in addition to offshoring. We find substantially different effects: technology mainly increases the inequality between firms in terms of worker skill quality and average wages, while offshoring mitigates this rising trend.

In the third chapter [**"Selective Accumulation of Ideas: Accounting for the Decline in Entry Rate"**], I explain the secular decline in entry rate of new firms using the mechanism of selective accumulation of ideas over time. In the model, an idea is a blueprint for a new product that arrives exogenously. An individual finds an idea of random quality drawn from an exogenous distribution, and makes an occupational choice of whether to become an entrepreneur using that idea, or to work for other entrepreneurs while discarding the idea. As ideas accumulate over time, the equilibrium threshold idea endogenously rises over time, and this would lower the rate of entry. With an expanding set of industries, the model also explains the industry life cycle, where the number of firms in each industry first increases and then decreases over time.

Acknowledgements

I am very grateful to Charles Engel, Ken West, and Menzie Chinn for their guidance and support, as well as valuable feedback. I also thank Dmitry Mukhin for helpful comments and discussions. Thanks to my parents, and especially to my wife Gueyon Kim who is also a great coauthor, for their continued encouragement, patience, and support in many aspects.

List of Figures

List of Tables

Contents

Chapter 1

Optimal Monetary Policy in an Open Economy with Shocks to UIP

1.1 Introduction

Uncovered interest parity (UIP) puzzle is one of the most long-standing and empirically robust puzzles in international macro. Although standard macro models predict that the expected returns to two bonds be the same in equilibrium (no arbitrage), this UIP condition is not a good description of the real world economy.[1] One immediate explanation for the deviation from UIP is to assume that there are shocks to premium for a bond denominated in one currency over another, which may arise from various sources including differences in the riskiness or in the liquidity value across different bonds.[2]

Several authors have noted the importance of UIP shocks. It turns out that the UIP shock is useful in explaining not only the UIP puzzle but also several other empirical puzzles. For example, Engel (2014b) shows that persistent but stationary UIP shocks

[1]A part of this discrepancy is due to the feature of the standard approach in DSGE models that considers only small perturbations around the steady state. Hence the violation of UIP does not immediately imply the rejection of these models.

[2]See Engel (2014b) for an extensive review of recent empirical as well as theoretical literature related to the UIP puzzle.

would account for the long-run forecastability and short-run unforecastability of exchange rates. Itskhoki and Mukhin (2017) also concludes that the UIP shock is the only shock among the numerous candidate shocks that can explain the exchange rate disconnect puzzle, while not being inconsistent with other puzzles such as the Backus-Smith puzzle and the UIP puzzle.

If the UIP shock is so prevalent and empirically important, what is its impact on welfare, and through what mechanism? Is there something that monetary policy can do to alleviate the impact of this shock? The purpose of this paper is to provide answers for these questions, using the conventional New Keynesian toolkit for optimal monetary policy analysis. In particular, this paper focuses on the cooperative monetary policy under commitment. The model features incomplete market, as well as segmented financial market à la Itskhoki and Mukhin (2017), on top of a standard, symmetric two country open economy. UIP shock is modeled as an exogenous shock to the demand of noise traders for one country's bond over the other country's, which may capture temporary liquidity needs.[3]

An immediate equilibrium impact of the UIP shock is that it distorts the expected paths of consumption in both countries. As a result of households' intertemporal optimization (Euler equation), a country with higher expected return would make an increasing consumption profile, and vice versa. A distinctive feature of the UIP shock is that it affects the growth rate of consumption. As a result, a positive shock to excess return that is expected to persist for multiple periods would make the consumption path

[3]One important feature of the UIP shock is that it is exogenous to the optimizing households. A related shock could be a shock to the preference of the households to prefer one bond over another. But such a shock would also alter the first best allocation, as would the productivity shock.

increasing over these periods, and the contemporaneous drop in the level of consumption would be large as it includes the cumulative effects of consumption growth over all expected future periods. In a sense, the UIP shock gives rise to a forced net saving of the households without affecting economic fundamentals such as productivity.

The first best allocation that attains the highest welfare is that of the complete markets where all risks are insured. The welfare loss compared to the first best allocation (loss function) can be approximated as the sum of squares of relative output gap, relative consumption, and relative inflation. In the current environment with incomplete markets, lack of state contingent claims is a fundamental source of inefficiency that gives rise to difference in allocation in response to shocks. Facing this constraint, the planner would try to improve welfare by reducing these differences upon the realization of shocks and thereby achieve an allocation closer to that under complete markets.

Under flexible price equilibrium, or under inflation targeting policy that replicates the flexible price equilibrium, any difference in consumption necessarily propagates into a difference in output gap in the same period, because higher consumption implies higher marginal rate of substitution between leisure and consumption, which implies higher real wages and thus higher price for the goods produced in the country that optimally chooses to consume more than the other. While the monetary policy can do little to affect the difference in consumption, it can reduce the difference in output gap by taking advantage of the price stickiness. As a result, optimal monetary policy would reduce the difference in output gap at the expense of allowing for a relatively small difference in inflation compared to the flexible price allocation.

This paper belongs to the literature that studies optimal monetary policy in an open economy. In an open economy with two countries,[4] Clarida et al. (2002) famously established the baseline result that optimal monetary policy in an open economy with complete asset markets would be identical to that in a closed economy. Subsequent works have shown that, as the economy deviates from this knife-edge case, optimal monetary policy in an open economy would be different from that in a closed economy. One important dimension of this deviation is the local currency pricing (LCP), where prices set in the importing country's currency are sticky. The literature has shown that introduction of LCP in place of producer currency pricing (PCP) would give rise to different price dynamics and thus different optimal policy prescriptions.[5] Another dimension is the role of different asset market structures (complete markets, incomplete markets, or financial autarky). Corsetti et al. (2018) as well as Engel (2014a) find that such different asset market structures would call for different monetary policies. The current paper is along this line of literature, where uninsurable financial shocks require incomplete markets to begin with. To my knowledge, this is the first paper that explores an optimal monetary policy response to the UIP shocks.

A separate but related line of literature explores the optimal Taylor type rule to be used for conducting monetary policy. This literature numerically finds the welfare-maximizing monetary policy rule, for example, optimal coefficient on inflation in the Taylor rule. A few papers, including Kollmann (2004), and Wang (2010), have considered optimal response to UIP shocks in this context. The current paper complements this literature by finding analytical solutions facing UIP shocks in a microfounded model, which is a more general targeting rule as opposed to an instrument rule.

[4]For optimal monetary policy analyses in a small open economy, see Gali and Monacelli (2005), Faia and Monacelli (2008), de Paoli (2009a,b), etc.

[5]See for example Devereux and Engel (2003), and Engel (2011).

This paper is also related to the literature that models financial frictions or imperfections in an international context. In many cases, financial intermediaries are allowed to exploit the expected return differential between bonds denominated in different currencies. The optimal solution of the intermediaries often ends up being proportional to the size of excess return. For example, Gabaix and Maggiori (2015), as well as Itskhoki and Mukhin (2017), introduce intermediaries that invests in carry trade along with segmented financial markets. These papers show that introduction of such financial imperfections turns out to be useful for explaining several well-known puzzles related to exchange rate. In the current paper, apart from the UIP shock, the segmented financial market itself features an element of financial imperfections that gives rise to imperfect capital mobility, where the degree of capital (im)mobility is governed by the degree of risk aversion of the financial intermediaries or the size of financial sector.

This class of model has also been used to study optimal policy, particularly regarding optimal foreign intervention. Fanelli and Straub (2019) studies optimal foreign intervention in a small open economy, with a model that is similar in spirit. While Fanelli and Straub focuses on how to optimally smooth the convex cost of excess return over time while abstracting from nominal rigidities, this paper ignores the former by assuming that the planner can tax most of the profits earned by the intermediaries and focuses on the role of nominal rigidities along with the particular dynamics of current account generated by the optimizing behavior of the intermediaries. In a sense, this paper complements this literature, by focusing on different aspects while faced with similar questions.

The rest of the paper is organized as follows. In section 2, the model environment is described, and the equilibrium is characterized. In section 3, the optimal monetary policy is characterized and solved, both numerically and analytically, starting from simpler special cases to capture intuitions. In addition to the optimal response to shocks, a targeting rule is also derived following Giannoni and Woodford (2017). Section 4 discusses the optimal capital control, which makes the assumption that the planner can set different interest rates to different types of agents. Section 5 compares the welfare across different policy in response to different shocks. Section 6 concludes.

1.2 Model

As explained above, most of the model elements follow the canonical two-country New Keynesian model. The only exception is the segmented financial market and the behavior of intermediaries.

1.2.1 Households

The representative household in the Home country maximizes the following standard intertemporal utility function:

$$V_0 = E_0 \sum_{t=0}^{\infty} \beta^t \Big(\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{N_t^{1+\eta}}{1+\eta} \Big)$$

subject to the budget constraint: $P_t C_t + B_t \leq \Omega_t N_t + R_{t-1} B_{t-1} + \Pi_t + T_t$. C_t is aggregate consumption, N_t is aggregate labor supply, B_t is the quantity of Home bond in nominal units held by Home households, which yields gross risk-free nominal interest rate of R_t, P_t is the aggregate consumer price level, Ω_t is the nominal wage, Π_t is the firm's profit in the aggregate (returned to households because they own these firms), and T_t is lumpsum net transfer to households. The foreign country's representative household solves a similar optimization problem in Foreign variables, which are typically denoted with an asterisk superscript (*). Note that the Home households do not have access to the Foreign bond, and vice versa.

In both Home and Foreign country, there is a continuum of firms whose measure is normalized to 1. Each firm produces a single variety indexed by j that is imperfectly substitutable with one another. The preference of both Home and Foreign households is a nested CES over all varieties, first over the varieties within each country with elasticity of substitution of $\epsilon > 1$, and then between the country-specific aggregate goods with elasticity of substitution $\phi < \epsilon$. In addition, there is home bias in preferences: $a \in [\frac{1}{2}, 1]$, where $a = \frac{1}{2}$ implies no home bias, and $a = 1$ means full home bias. The aggregate consumption of the Home household can be expressed as:

$$C_t = \left(a^{\frac{1}{\phi}} C_{Ht}^{\frac{\phi-1}{\phi}} + (1-a)^{\frac{1}{\phi}} C_{Ft}^{\frac{\phi-1}{\phi}} \right)^{\frac{\phi}{\phi-1}}$$

where

$$C_{Ht} = \left(\int C_{Ht}(j)^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}}, \qquad C_{Ft} = \left(\int C_{Ft}(j)^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}}$$

Likewise, the aggregate consumption of the Foreign household C_t^* can be expressed as follows, with analogous definitions for C_{Ft}^* and C_{Ht}^*:

$$C_t^* = \left(a^{\frac{1}{\phi}} (C_{Ft}^*)^{\frac{\phi-1}{\phi}} + (1-a)^{\frac{1}{\phi}} (C_{Ht}^*)^{\frac{\phi-1}{\phi}} \right)^{\frac{\phi}{\phi-1}}$$

1.2.2 Firms

The production technology requires one type of factor input, labor, and has constant returns to scale. In addition, all firms in the same country face the same aggregate productivity. Thus the production function is the following linear function: $Y_t(j) =$

$A_t N_t(j)$, where A_t is the aggregate productivity common to all firms in Home country, and $N_t(j)$ is the quantity of labor hired in firm j. Likewise, $Y_t^*(j) = A_t^* N_t^*(j)$ for a firm in Foreign country.

Each household supplies a differentiated type of labor that is imperfectly substitutible yet symmetric with one another. Each firm need to hire the entire bundle of different types of labor which is combined as CES over all types of labor:

$$N_t(j) = \Big(\int N_t(j,h)^{\frac{\widetilde{\mu}_t - 1}{\widetilde{\mu}_t}} dh \Big)^{\frac{\widetilde{\mu}_t}{\widetilde{\mu}_t - 1}}$$

The elasticity of substitution between different types of labor is equal to $\widetilde{\mu}_t > 1$,[6] which is exogenous and time-varying. This allows households to have bargaining power over their wages, and earn a time-varying markup of $\frac{\widetilde{\mu}_t}{\widetilde{\mu}_t - 1}$ over their marginal disutility.[7] Unlike the prices for goods which are sticky, wages are assumed to be set in a flexible manner.

The profit of each firm j in Home country in each period can be expressed as

$$\Pi_t(j) = P_{Ht}(j) C_{Ht}(j) + \mathcal{E}_t P_{Ht}^*(j) C_{Ht}^*(j) - (1 - \tau) \Omega_t N_t(j)$$

where $P_{Ht}(j)$ is the domestic price set in Home currency, $P_{Ht}^*(j)$ is the export price set in Foreign currency, $\mathcal{E}_t$ is the nominal exchange rate, Ω_t is the nominal wage, and $C_{Ht}(j)$ and $C_{Ht}^*(j)$ are the domestic and export demand respectively. Market clearing for each good implies $Y_t(j) = A_t N_t(j) = C_{Ht}(j) + C_{Ht}^*(j)$. In the presence of sticky prices à la Calvo, firms that are able to set price in each period set prices to maximize the expected

[6]The tilde above μ_t is to indicate that this is in level as opposed to the log deviation. Later, the log-deviation of $\frac{\widetilde{\mu}_t}{\widetilde{\mu}_t - 1}$ will be denoted as μ_t.

[7]This is one of the standard ways of introducing a cost shock in a baseline New Keynesian model. $\widetilde{\mu}_t$ has a natural interpretation of time-varying monopoly power of labor unions.

discounted profits:

$$\max_{P_{Ht}(j),P^*_{Ht}(j)} E_t \sum_{k=0}^{\infty} \beta^k \left(\frac{C_{t+k}}{C_t}\right)^{-\sigma} \alpha^k \Pi_{t+k}(j)$$

where $(1-\alpha)$ is the probability that each firm can reset prices in each period.

1.2.3 Financial Sector

The financial sector is borrowed from Itskhoki and Mukhin (2017). There are two additional representative agents: noise traders, and financial intermediaries, both of which are owned by the Foreign households.[8] To make the problem simple, it is assumed that they operate period by period without accumulating equity, and transfer any net profits realized in each period to the Foreign households.[9]

Noise traders simply take the positions of Home and Foreign bonds as determined by exogenous forces, regardless of the expected returns. This can be interpreted as, for example, temporary liquidity needs for a particular country's bond.[10] Based on the zero capital assumption, they take a long position in the bond that they need more, and short position in the other bond. Denote the noise trader's holding of Home nominal bond as N_t, and that of Foreign nominal bond as N^*_t. Then $N_t + \mathcal{E}_t N^*_t = 0$ holds. The value of N_t is determined as

$$N_t = -n\widetilde{f}_t$$

where $\widetilde{f}_t$ is the current shock to liquidity of Foreign bond relative to Home bond, and

[8] This is exactly the same assumption made by Itskhoki and Mukhin. Because the profits are of second order, it turns out that this assumption does not cause asymmetry up to first order approximation, and not affect the results using LQ-approximation. This point will be further elaborated in the next subsection.

[9] It can be thought of as new generations of intermediaries born in each period, who operate only for one period and then retire.

[10] As explained in the introduction, there is now a large literature that discusses the liquidity value of bonds, as well as its impact on exchange rates.

n is the mass of noise traders. A positive shock to $\widetilde{f}_t$ implies that the noise traders currently have a greater demand for Foreign bond.

Financial intermediaries have CARA (Constant Absolute Risk Aversion) utility function over the stochastic profit denominated in Foreign goods. They only consider the next period's profit by investing in one-period bonds. Like the noise traders, intermediaries' also use zero capital strategy, but based on the expected returns: long in the bond with higher expected return, and short in the other bond. Denote the intermediaries's holding of Home nominal bond as D_t, and that of Foreign nominal bond as D_t^*. Then zero capital investment implies that $D_t + \mathcal{E}_t D_t^* = 0$. The ex post profit of the financial intermediaries in Foreign currency is

$$\Pi_{t+1}^{\mathcal{I}*} = R_t^* D_t^* + R_t \frac{D_t}{\mathcal{E}_{t+1}} = \Big(R_t^* - R_t \frac{\mathcal{E}_t}{\mathcal{E}_{t+1}}\Big) D_t^*$$

As a result of the optimal investment choice of the intermediaries that maximize their expected utility, which derivation is shown in the appendix, size of their investment is proportional to the expected excess return. In particular, up to first order approximation,

$$D_t = \frac{m}{\gamma \sigma^2} E_t er_{t+1}$$

where er_{t+1} is the first order approximation of $(R_t^* - R_t \frac{\mathcal{E}_t}{\mathcal{E}_{t+1}})$ around the steady state of zero, m is the mass of financial intermediaries, γ is the absolute risk aversion parameter, and σ^2 is the variance of the expected return.[11]

Lastly, bond market clearing conditions in each country implies $B_t + D_t + N_t = 0$.[12]

[11]Given the i.i.d. nature of shocks, the conditional variance is the same as the unconditional variance.

[12]Likewise, $B_t^* + D_t^* + N_t^* = 0$ holds for the Foreign bond. Together with the zero capital condition for both noise traders and financial intermediaries, it follows that $B_t + \mathcal{E}_t B_t^* = 0$. Generally speaking, given financial segmentation, the sum of Home and Foreign households' net saving $(B_t + \mathcal{E}_t B_t^*)$ need not be equal to zero without the zero capital assumption. But this zero capital assumption results in

Combining with the first order approximation of $B_t = Y^{ss} b_t$, along with other equations shown above,

$$Y^{ss} b_t + \frac{m}{\gamma\sigma^2} E_t er_{t+1} - n\widetilde{f}_t = 0$$

Rearranging,

$$E_t er_{t+1} = -\frac{\gamma\sigma^2 Y^{ss}}{m} b_t + \frac{\gamma\sigma^2 n}{m}\widetilde{f}_t \equiv -\chi b_t + f_t$$

where $\chi \equiv \frac{\gamma\sigma^2 Y^{ss}}{m}$, and $f_t \equiv \frac{\gamma\sigma^2 n}{m}\widetilde{f}_t$ is simply a renormalization of the shock. This is the key equation that summarizes the difference between the economies with and without segmented financial markets as described above. Without segmentation, the usual UIP condition would hold: $E_t er_{t+1} = 0$. In line with this condition under the benchmark environment, the above equation will be called "modified UIP" condition henceforth.

One usefulness of this model of financial sector à la Itskhoki and Mukhin is that, in principle, χ can take any positive value. $\chi \to 0$ implies a perfect mobility of capital without financial segmentation, while $\chi \to \infty$ represents a financial autarky. The ratio $\frac{\gamma\sigma^2 n}{m}$ would govern the responsiveness of excess return with respect to the size of liquidity shock (per each noise trader) in terms of quantity. That is, the liquidity shock would have a more pronounced effect on the expected excess return if the mass of noise trader is larger compared to the financial intermediaries, or if the intermediaries are more risk averse. Holding this ratio $\frac{\gamma\sigma^2 n}{m}$ constant, $\chi \to 0$ as $\frac{n}{Y^{ss}} \to \infty$, and $\chi \to \infty$ as $\frac{n}{Y^{ss}} \to 0$.[13]

a natural correspondence to the case without financial segmentation where $B_t + \mathcal{E}_t B_t^* = 0$ necessarily holds.

[13]If we were to study the effect of changes in χ, a particular normalization of the shock is not innocuous. For example, if the object of interest is the effect of changes in the degree of risk aversion of the financial intermediaries, the shock f_t would also need to be multiplied by χ. In that case, as the intermediaries become risk neutral in the limit, the UIP shock f_t would simply vanish from all equations. Because the focus of the paper is to study the impact of UIP shocks, holding the variance of shocks to UIP constant while varying χ seems more relevant. A relevant interpretation of a decrease in χ would then be a simultaneous increase in the size of intermediaries and the noise traders, while holding other things constant.

Based on the current structure of the financial market, financial intermediaries are expected (i.e., on average) to make positive profit. Noise traders are expected to make losses because their demand is inelastic, and the intermediaries would operate towards positive profits in expectation. However, whether the households make profits or losses is ambiguous, and depends on which shock is in effect.

First, if the economy faces a liquidity shock, households' choice of net saving is purely induced by the excess return that arise from the intermediaries' attempt to meet the demand for the noise traders. For example, suppose that the noise traders want to hold Foreign bond, so that there is now excess supply of Home bond and excess demand for Foreign bond. Then the price of Home bond would fall, and the expected return of Home bond would be higher than if there were no liquidity shock, as well as higher than the current expected return of Foreign bond. Facing this expected excess return, the financial intermediaries would hold the Home bond while shorting on the Foreign bond. Likewise, facing a higher expected return of Home bond, Home households would lend to the financial intermediaries and delay consumption, whereas facing a lower expected return, Foreign households would borrow from the intermediaries and consume more in the current period. Because the households' positions are in the same direction as that of the financial intermediaries, households would also make profits in expectation as do intermediaries.

Suppose the economy faces a positive shock in UIP, induced by an increase in the relative demand for Home bond by the noise traders. Facing this exogenous demand, the equilibrium price of Home bond would rise, resulting in a lower expected return. As the expected excess return for Home bond falls, financial intermediaries would supply Home bond while holding Foreign bond. At the same time, facing the lower real return, Home households would reduce their net saving and consume more today, and thereby supplying additional quantity of Home bond. Since both of these supply curves for Home bond are monotonically decreasing in the expected excess return for Home bond, the equilibrium expected excess return will be determined at the point where the supply meets the exogenously given quantity of Home bond demanded by the noise traders.

Second, if it is a non-liquidity shock that is in effect, noise traders would not take part in the bond market, and it is only the households' optimal response to the shock that is to be intermediated by the financial intermediaries. In other words, households' choice of net saving is induced by the self needs in the presence of profit-seeking intermediaries. For example, suppose that the Home household faces a positive shock in relative productivity. Home households would like to save the currently high income for future consumptions, and the opposite would hold for the Foreign households. But to achieve these net saving activities, the expected return of Home bond must be lower than that of Foreign bond, in order to induce the financial intermediaries to take the opposite position and clear the bond market. And facing this lower expected return of Home bond, the amount of saving would be lower than what would have been desired by the households without financial segmentation. Lastly, if these two shocks are both in effect, the two mechanisms would work in the opposite direction, and whether the households make profits or losses would depend on the relative size of each shock.

Note that the ex post profit to each of the three types of agents is the product of realized return er_t and the size of respective positions (B_t, D_t, N_t). Because both the position variables and excess return are zero in the steady state, the product of the two is second order. Thus the profits do not appear in any of the equations up to first order approximation, and the only modification from Corsetti et al. (2018) would be the modified UIP equation in approximation.

The assumption that the second order profits in the financial sector is transferred to households is crucial for the subsequent analysis of optimal policy. It makes the world budget constraint exactly the same as without the financial sector. Note that even if this profit is second order, the derivation of the loss function involves second order approximation of the world budget constraint (which is essentially the sum of two market clearing conditions). If the profits are consumed by some agents other than households, it must be properly taken into account and the loss function would need to be altered.[14] Lastly, the assumption that the profits are transferred only to the Foreign household is made only for convenience, and does not alter any of the subsequent analysis, as in Itskhoki and Mukhin (2017).

1.2.4 Equilibrium

The Home households' demand for the aggregate goods from each country are:

$$C_{Ht} = a\Big(\frac{P_{Ht}}{P_t}\Big)^{-\phi} C_t, \qquad C_{Ft} = (1-a)\Big(\frac{P_{Ft}}{P_t}\Big)^{-\phi} C_t$$

where

$$P_t = \Big(aP_{Ht}^{1-\phi} + (1-a)P_{Ft}^{1-\phi}\Big)^{\frac{1}{1-\phi}}$$

[14] An alternative assumption where the profits in the financial sector are not transferred to the households but instead consumed by the agents in the financial sector would involve further complications. A crucial question is how to think of the welfare of the financial intermediaries. For example, the social planner might take into account the consumption of these agents, or simply ignore their consumption and regard the profits as deadweight loss to the economy. Both are plausible approaches, and in fact Fanelli and Straub (2019) is an example of the environment where the profits in the financial sector is regarded as deadweight loss. In this paper, in order to focus on the monetary policy aspect of the model, I take the simplest assumption of transferring the profits to the households, which allows a close connection with the existing literature.

The demand for each variety are:

$$C_{Ht}(j) = \Big(\frac{P_{Ht}(j)}{P_{Ht}}\Big)^{-\epsilon} C_{Ht} = \Big(\frac{P_{Ht}(j)}{P_{Ht}}\Big)^{-\epsilon} \Big(\frac{P_{Ht}}{P_t}\Big)^{-\phi} C_t$$

$$C_{Ft}(j) = \Big(\frac{P_{Ft}(j)}{P_{Ft}}\Big)^{-\epsilon} C_{Ft} = \Big(\frac{P_{Ft}(j)}{P_{Ft}}\Big)^{-\epsilon} \Big(\frac{P_{Ft}}{P_t}\Big)^{-\phi} C_t$$

where

$$P_{Ht} = \Big(\int P_{Ht}(j)^{1-\epsilon} df\Big)^{\frac{1}{1-\epsilon}}, \qquad P_{Ft} = \Big(\int P_{Ft}(j)^{1-\epsilon} df\Big)^{\frac{1}{1-\epsilon}}$$

Analogous equations hold for the Foreign household's demand.

When the law of one price (LOOP) holds, $P_{Ht} = \mathcal{E}_t P^*_{Ht}$ and $P_{Ft} = \mathcal{E}_t P^*_{Ft}$ hold, where $\mathcal{E}_t$ is the nominal exchange rate. If the LOOP does not hold, prices are said to be misaligned. The price misalignment for the goods from each country is defined as

$$M_{Ht} \equiv \frac{\mathcal{E}_t P^*_{Ht}}{P_{Ht}}, \;\; M_{Ft} \equiv \frac{\mathcal{E}_t P^*_{Ft}}{P_{Ft}}$$

In general, $M_{Ht} = M_{Ft}$ need not hold. However, as shown in Engel (2011), if the initial condition is given as symmetric ($M_{H0} = M_{F0}$), the equilibrium misalignment would be always symmetric, i.e., $M_{Ht} = M_{Ft}$, $\forall t$. Since there is little gain from imposing asymmetry, I maintain the assumption of symmetric misalignment as do Corsetti et al., and proceed with $M_{Ht} = M_{Ft} \equiv M_t$.

Terms of trade is defined as the relative price between Foreign and Home goods when expressed in the same currency:

$$S_t \equiv \frac{P_{Ft}}{P_{Ht}} = \frac{P^*_{Ft}}{P^*_{Ht}}$$

where $\frac{P_{Ft}}{P_{Ht}} = \frac{P^*_{Ft}}{P^*_{Ht}}$ follows from $M_{Ht} = M_{Ft}$. The real exchange rate (RER) is defined as

$$Q_t \equiv \frac{\mathcal{E}_t P^*_t}{P_t}$$

In log-linearized form, the above set of equations can be expressed as:

$$m_t = e_t + p^*_{Ht} - p_{Ht} = e_t + p^*_{Ft} - p_{Ft}$$

$$s_t = p_{Ft} - p_{Ht} = p^*_{Ft} - p^*_{Ht}$$

$$p_t = ap_{Ht} + (1-a)p_{Ft}$$

$$p^*_t = ap^*_{Ft} + (1-a)p^*_{Ht}$$

$$q_t = e_t + p^*_t - p_t = (2a-1)s_t + m_t$$

Note that if $a = 1/2$ (no home bias), the consumption bundle for Home and Foreign households become identical, and the resulting real exchange rate q_t would reflect only the degree of misalignment m_t, if any. If in addition the LOOP holds, $q_t \equiv 0$.

Goods market clearing in the aggregate implies

$$Y_t = C_{Ht} + C^*_{Ht} = a\Big(\frac{P_{Ht}}{P_t}\Big)^{-\phi} C_t + (1-a)\Big(\frac{P^*_{Ht}}{P^*_t}\Big)^{-\phi} C^*_t$$

$$Y^*_t = C^*_{Ft} + C_{Ft} = a\Big(\frac{P_{Ft^*}}{P^*_t}\Big)^{-\phi} C^*_t + (1-a)\Big(\frac{P_{Ft}}{P_t}\Big)^{-\phi} C_t$$

In log-linearized form,

$$y_t = ac_{Ht} + (1-a)c^*_{Ht} = ac_t + (1-a)c^*_t + 2a(1-a)\phi s_t$$

$$y^*_t = ac^*_{Ft} + (1-a)c_{Ft} = ac^*_t + (1-a)c_t - 2a(1-a)\phi s_t$$

Household budget constraint in the Home country is

$$P_t C_t + B_t = \Omega_t N_t + \Pi_t + R_{t-1}B_{t-1} + T_t$$

Combined with the firm's profit in the aggregate ($\Pi_t = P_{Ht}Y_t - (1-\tau)\Omega_t N_t$), and the government budget constraint ($T_t + \tau\Omega_t N_t = 0$), the consolidated budget constraint in the Home country can be expressed as

$$P_t C_t + B_t = P_{Ht} Y_t + R_{t-1} B_{t-1}$$

or equivalently, net export equals net saving:

$$NX_t = P_{Ht} Y_t - P_t C_t = B_t - R_{t-1} B_{t-1}$$

In log-linearized form around the steady state of zero,

$$nx_t = y_t - c_t - (1-a)s_t = b_t - \beta^{-1} b_{t-1}$$

As shown in the appendix, the Foreign country's budget constraint is redundant given the Home country's budget constraint, as usual.

Intertemporal optimization of the household results in the following standard Euler equations:

$$\beta R_t E_t \Big(\frac{C_{t+1}}{C_t} \Big)^{-\sigma} \frac{1}{\pi_{t+1}} = 1$$

$$\beta R_t^* E_t \Big(\frac{C_{t+1}^*}{C_t^*} \Big)^{-\sigma} \frac{1}{\pi_{t+1}^*} = 1$$

where $\pi_t \equiv \frac{P_t}{P_{t-1}}$ and $\pi_t^* \equiv \frac{P_t^*}{P_{t-1}^*}$ are the CPI inflation in Home and Foreign country, respectively. Log-linearizing, and taking the difference,

$$r_t - r_t^* + E_t \big[-\sigma \Delta (c_{t+1} - c_{t+1}^*) - (\pi_{t+1} - \pi_{t+1}^*) \big] = 0$$

From the definition of real exchange rate ($Q_t \equiv \frac{\mathcal{E}_t P_t^*}{P_t}$, where $\mathcal{E}_t$ is the nominal exchange rate; $q_t = e_t + p_t^* - p_t$ in log-linearized form),

$$\pi_{t+1} - \pi_{t+1}^* = \Delta e_{t+1} - \Delta q_{t+1}$$

Following Corsetti et al. (2018), define the "wealth gap" as:[15]

$$W_t \equiv \frac{U_C^*(t)}{U_C(t)} \frac{1}{Q_t} = \left(\frac{C_t^*}{C_t}\right)^{-\sigma} \frac{1}{Q_t}$$

or in log-linearization,

$$w_t = \sigma(c_t - c_t^*) - q_t$$

Then

$$\underbrace{r_t - r_t^* - E_t \Delta e_{t+1}}_{E_t er_{t+1}} - \underbrace{E_t\Big[\Big(\sigma\Delta(c_{t+1} - c_{t+1}^*) - \Delta q_{t+1}\Big)\Big]}_{E_t \Delta w_{t+1}} = 0$$

As before, er_{t+1} is defined as the first order approximation of the ex post excess return of Home bond relative to Foreign. Combining with the financial market equilibrium condition yields the following modified UIP condition:

$$E_t \Delta w_{t+1} = E_t er_{t+1} = -\chi b_t + f_t$$

The firm's optimal price setting decision leads to standard Phillips curves. As is well known, there are different possibilities regarding in which currency the firms set prices. The first case is known as the producer currency pricing (PCP), where the firms set prices in domestic currency, while the export prices is simply determined by the equilibrium nominal exchange rate. The law of one price holds naturally: $\frac{\mathcal{E}_t P_{Ht}^*}{P_{Ht}} = \frac{\mathcal{E}_t P_{Ft}^*}{P_{Ft}} = 1$. Taking the first order condition for the firm's problem, the standard set of Phillips curves under PCP are derived:

$$\pi_{Ht} = \beta \pi_{Ht+1} + \delta\big(mc_t + \mu_t\big)$$

$$\pi_{Ft}^* = \beta \pi_{Ft+1}^* + \delta\big(mc_t^* + \mu_t^*\big)$$

[15]As a result of intertemporal optimization, the following expression is equal to the cross-country difference in the sum of current and expected discounted future stream of the value of consumption, which can be naturally called wealth gap. Note that in the absence of home bias ($a = 1/2$), $Q_t \equiv 1$, and the wealth gap is simply the consumption gap. With home bias, wealth gap includes the relative price between Home and Foreign goods in addition to the consumption gap.

where $\delta \equiv \frac{(1-\alpha\beta)(1-\alpha)}{\alpha}$, and $1-\alpha$ is the probability that a firm can reset its price. $mc_t \equiv \sigma c_t + \eta(y_t - a_t) + (1-a)s_t - a_t$ is the real marginal cost in the Home country, and analogously for the Foreign country.

The second case is the local currency pricing (LCP), where the firms also set prices in the exporting country's currency, which is also sticky in that currency. The law of one price does not hold in general, and the prices are "misaligned." Without loss of generality, the degree of misalignment can be assumed to be symmetric, in which case the following expression would measure the common degree of misalignment: $M_t = \frac{\mathcal{E}_t P^*_{Ht}}{P_{Ht}} = \frac{\mathcal{E}_t P^*_{Ft}}{P_{Ft}} \neq 1$ in general. Taking the first order conditions for the firm's problem, the standard set of Phillips curves under LCP are obtained:

$$\pi_{Ht} = \beta\pi_{Ht+1} + \delta\big(mc_t + \mu_t\big)$$

$$\pi^*_{Ft} = \beta\pi^*_{Ft+1} + \delta\big(mc^*_t + \mu^*_t\big)$$

$$\pi^*_{Ht} = \beta\pi^*_{Ht+1} + \delta\big(mc_t - m_t + \mu_t\big)$$

$$\pi_{Ft} = \beta\pi_{Ft+1} + \delta\big(mc^*_t + m_t + \mu^*_t\big)$$

where $m_t \equiv \log(M_t/M^{ss}) = \log(M_t)$ is the log-linearized value of the misalignment M_t. The first two Phillips curves are for domestic prices, and the latter two are for export prices. In addition, there is an identity that must hold under LCP: $\Delta s_t = \pi_{Ft} - \pi_{Ht}$. This is an equation that trivially holds under PCP, but under LCP this acts as a constraint to the monetary policy.

As usual, under sticky price, the equilibrium is indeterminate without some specification of monetary policy. There are two monetary policy tools available, one in each country. Under PCP, there are 9 endogenous variables: $\{y_t, y^*_t, c_t, c^*_t, s_t, w_t, b_t, \pi_{Ht}, \pi^*_{Ft}\}$.

There are 7 equations: 2 market clearing, budget constraint, definition of w_t, modified UIP, and 2 Phillips curves. This leaves room for 2 policy tools, as it should. Under LCP, there are 12 endogenous variables: $\{y_t, y_t^*, c_t, c_t^*, s_t, w_t, m_t, b_t, \pi_{Ht}, \pi_{Ft}^*, \pi_{Ht}^*, \pi_{Ft}\}$. There are 10 equations: 2 market clearing, budget constraint, definition of w_t, modified UIP, 4 Phillips curves, and the law of motion for s_t. Again, this leaves room for 2 policy tools.

Decomposing the Home and Foreign variables as World and Relative terms makes the problem much simpler, which is also a common practice in the literature. That is, for each pair of log-linearized variables x_t and x_t^*, define

$$x_t^W \equiv \frac{x_t + x_t^*}{2}, \qquad x_t^R \equiv \frac{x_t - x_t^*}{2}$$

The pair of market clearing conditions can be expressed as

$$y_t^W = c_t^W$$

$$y_t^R = (2a-1)c_t^R + 2a(1-a)\phi s_t$$

Definition of w_t would be

$$w_t = 2\sigma c_t^R - \big((2a-1)s_t + m_t\big)$$

In addition, define π_t^W, π_t^R as the World and Relative expression for π_{Ht}, π_{Ft}^*, which are the inflations in domestically sold goods' prices. Then the Phillips curves for domestic prices would be

$$\pi_t^W = \beta E_t \pi_{t+1}^W + \delta\big(mc_t^W + \mu_t^W\big)$$

$$\pi_t^R = \beta E_t \pi_{t+1}^R + \delta\big(mc_t^R + \mu_t^R\big)$$

where

$$mc_t^W \equiv \sigma c_t^W + \eta y_t^W - (1+\eta)a_t^W$$

$$mc_t^R \equiv \sigma c_t^R + \eta y_t^R + (1-a)s_t - (1+\eta)a_t^R$$

Turning to the Phillips curves for export prices, note that the sum of these two equations is identical to the sum of the two Phillips curves for domestic prices.[16] The Relative Phillips curve for export prices that holds under LCP would be

$$\pi_t^X = \beta E_t \pi_{t+1}^X + \delta\big(mc_t^R - m_t + \mu_t^R\big)$$

where $\pi_t^X \equiv \frac{\pi_{Ht}^* - \pi_{Ft}}{2}$.

1.2.5 First Best Allocation

Before proceeding further, it is useful to consider the "first best allocation," as typically done in the literature. This is particularly useful because it is convenient to express the welfare as relative to the first best allocation, which attains the highest welfare possible.

Typically, regarding optimal monetary policy, first best allocation involves three conditions or policy tools: (1) complete markets, (2) flexible prices, and (3) time-varying employment subsidy. It is a hypothetical economy without any elements that distort the allocation. In the presence of cost shocks, the time-varying employment subsidy is needed because it would act to cancel out these cost shocks. In the current environment, all the three conditions are necessary. In addition, there is a fourth condition to ensure the first best allocation in the presence of UIP shocks: an ability to differentiate and discriminate interest rates facing different agents of the economy, i.e., noise traders or financial intermediaries from typical households. With this ability, the government would impose a time-varying capital tax facing non-households. With this policy tool in hand, the modified UIP condition does not bind anymore because the free policy variable can

[16] Likewise, one of the four inflation variables is trivially redundant.

sustain any difference that arises in that equation. Section 4 (Optimal Capital Control) contains a more detailed description of this policy.

The first best allocation can be described using the following set of log-linearized equations around the deterministic steady state:

$$\text{(Market Clearing - W): } y_t^{Wfb} = c_t^{Wfb}$$

$$\text{(Market Clearing - R): } y_t^{Rfb} = (2a-1)c_t^{Rfb} + 2a(1-a)\phi s_t^{fb}$$

$$\text{(Definition of } w_t\text{): } w_t^{fb} \equiv 2\sigma c_t^{Rfb} - (2a-1)s_t^{fb}$$

$$\text{(Complete Market): } w_t^{fb} = 0$$

$$\text{(Flexible Prices - W): } \sigma c_t^{Wfb} + \eta y_t^{Wfb} = (1+\eta)a_t^W$$

$$\text{(Flexible Prices - R): } \sigma c_t^{Rfb} + \eta y_t^{Rfb} + (1-a)s_t^{fb} = (1+\eta)a_t^R$$

where the flexible price condition and complete market condition implicitly assume the availability of corresponding policy tools, i.e., time-varying employment subsidy and time-varying capital tax, respectively. The superscripts "fb" indicates the values being under first best allocation.

It is straightforward to solve for the first best allocation. Combining the two World equations,

$$y_t^{Wfb} = c_t^{Wfb} = \frac{1+\eta}{\sigma+\eta}a_t^W$$

Combining the first three Relative equations,

$$c_t^{Rfb} = \frac{2a-1}{D}y_t^{Rfb}$$

$$s_t^{fb} = \frac{2\sigma}{D}y_t^{Rfb}$$

where $D \equiv 4a(1-a)\sigma\phi + (2a-1)^2 = 4a(1-a)(\sigma\phi - 1) + 1$. Substituting this into (Flexible Prices),

$$y_t^{Rfb} = \frac{1+\eta}{\frac{\sigma}{D}+\eta} a_t^R$$

and c_t^{Rfb} and s_t^{fb} can be solved accordingly.

One important feature of this first best allocation is that it depends only on the productivity shock a_t^R, but not on the other shocks such as the cost shock μ_t^R or the UIP shock f_t. The productivity shock affects the fundamental production process, so it affects the first best allocation. The other shocks are distortionary shocks, and it is best to nullify these shocks by appropriate policy tools, assuming the availability of sufficiently rich set of such policy tools as in the first best allocation derived above.

1.2.6 Flexible Price Equilibrium

In this section I derive an analytic solution under flexible price equilibrium. This part is important for at least two distinct reasons. First, the flexible price equilibrium clearly illustrates the key economic mechanisms of the current incomplete market environment facing various shocks. Second, the flexible price allocation can be attained by PPI (producer price index) targeting, and thus will serve as the benchmark economy when assessing the performance of optimal policy to be derived later. In this section, the analytic solution will focus on the case of $\chi = 0$, but it is not too difficult to obtain analytic solution even with $\chi > 0$, as described in the appendix. Although the main contribution of this paper is about the response to UIP shocks, other types of shocks (productivity shock, cost shock) will also be considered.

From this point, in many cases, the "gap" variables will be used for convenience. For each variable x_t, $\widetilde{x}_t \equiv x_t - x_t^{fb}$, which is the difference between the log-linearized value of a variable and the value under first best allocation also in log-linearized form. This is of course a very common practice in the literature.

The flexible price equilibrium in world variables can be described by the following set of equations in gap form:

$$\text{(Market Clearing): } \widetilde{y}_t^W = \widetilde{c}_t^W$$

$$\text{(Flexible Prices): } \sigma\widetilde{c}_t^W + \eta\widetilde{y}_t^W + \mu_t^W = 0$$

It follows immediately that

$$\widetilde{y}_t^W = \widetilde{c}_t^W = -\frac{1}{\sigma+\eta}\mu_t^W$$

Combined with the first best allocation,

$$y_t^W = c_t^W = \frac{1+\eta}{\sigma+\eta}a_t^W - \frac{1}{\sigma+\eta}\mu_t^W$$

Overall, it is evident that the world economy behaves in exactly the same way as a typical closed economy, as it should. Consumption and output are the same in all periods, and they efficiently increase facing a positive world productivity shock. But facing a positive world cost shock, both output and consumption decrease inefficiently.

The flexible price equilibrium in relative variables can be characterized by the following set of equations in gap form:

$$\text{(Market Clearing): } \widetilde{y}_t^R = (2a-1)\widetilde{c}_t^R + 2a(1-a)\phi\widetilde{s}_t$$

$$\text{(Definition of } w_t\text{): } \widetilde{w}_t \equiv 2\sigma\widetilde{c}_t^R - (2a-1)\widetilde{s}_t$$

$$\text{(modified UIP): } E_t\widetilde{w}_{t+1} - \widetilde{w}_t = -\chi b_t + f_t = f_t$$

$$\text{(Budget Constraint): } y_t - c_t - (1-a)s_t = nx_t = b_t - \beta^{-1}b_{t-1}$$

$$\text{(Flexible Prices): } \sigma\widetilde{c}_t^R + \eta\widetilde{y}_t^R + (1-a)\widetilde{s}_t + \mu_t^R = 0$$

For notational convenience, for each variable x_t, define $\overline{x}_t \equiv \lim_{k\to\infty} E_t\widetilde{x}_{t+k}$, that is, the long-run expected value conditional on the information available at t. Typically, all variables would converge to zero in log-deviation from steady state when there is no unit root or $\chi > 0$. However, when $\chi = 0$, there is no steady state, and these long-run expected values are typically non-zero. Strictly speaking, this perturbation around the steady state would not make sense when χ is exactly equal to zero, but we can always think of a limiting case where χ is sufficiently close to but not exactly equal to zero. In addition, by introducing the imperfect capital mobility, the case with $\chi > 0$ has a solid micro-foundation.

After some relatively straightforward algebra, as shown in the appendix, we can first obtain the permanent level of the wealth gap as

$$\overline{w}_t = \frac{1-\beta}{(1-\rho_f)(1-\beta\rho_f)} f_t + \frac{1-\beta}{A}\left[\beta^{-1}b_{t-1} + E_t\sum_{k=0}^{\infty}\beta^k nx_{t+k}^{fb} - \frac{2W_yD}{W_b(\sigma+\eta D)}E_t\sum_{k=0}^{\infty}\beta^k\mu_{t+k}^R\right]$$

where $A \equiv \frac{1}{W_b}(1 + \frac{W_y(D-2a+1)}{\sigma+\eta D})$ is the constant ratio between $\widetilde{nx}_t$ and $\widetilde{w}_t$: $\widetilde{nx}_t = -A\widetilde{w}_t + (shocks)$. Consequently, all $E_t\widetilde{w}_{t+k}$ as well as $E_t\widetilde{y}_{t+k}^R$ can be obtained as

$$E_t\widetilde{w}_{t+k} = \overline{w}_t - \frac{\rho_f^k}{1-\rho_f}f_t$$

$$E_t\widetilde{y}_{t+k}^R = -\frac{D-2a+1}{2(\sigma+\eta D)}E_t\widetilde{w}_{t+k} - \frac{D}{\sigma+\eta D}E_t\mu_{t+k}$$

Now a brief discussion about the effect of each shock is in order. First, consider the effect of a productivity shock:

$$\overline{w}_t = \frac{1-\beta}{A}\left[\beta^{-1}b_{t-1} + E_t\sum_{k=0}^{\infty}\beta^k nx_{t+k}^{fb}\right]$$

$$E_t\widetilde{w}_{t+k} = \overline{w}_t, \qquad E_t\widetilde{y}_{t+k}^R = -\frac{D-2a+1}{2(\sigma+\eta D)}\overline{w}_t$$

A positive shock in relative productivity gives rise to an increase in current income, which is then allocated as higher expected consumption across all future periods based on perfect smoothing of relative consumption (UIP). This evenly higher consumption in each period is transmitted period by period into an evenly higher output gap with the opposite sign, based on the flexible price condition which means zero labor wedge by simply equating the marginal rate of substitution between leisure and consumption to the real wage. Then the lifetime budget constraint determines the size of increase in relative consumption, while considering its effect on net export through output and relative price.

As it would become evident in the welfare analysis part, any nonzero value of the gap variables, or equivalently, any deviation from the first best allocation, would result in inefficiency in terms of welfare compared to the first best allocation. Under complete market, $\widetilde{w}_t = 0$ by construction, and the flexible price response to productivity shock is efficient. However, incomplete market results in $\widetilde{w}_t \neq 0$ facing productivity shocks, which acts as the source of inefficiency.

nx_t^{fb} is an important object when considering the relative productivity shock. In terms of log approximation, it can be expressed as $nx_t^{fb} = y_t^{fb} - c_t^{fb} - (1-a)s_t^{fb}$, and substituting the first best allocation under complete markets, $nx_t^{fb} = \frac{1+\eta}{\frac{\sigma}{D}+\eta} \frac{2W_y}{W_b} a_t^R$ is simply a linear transformation of the realized productivity shock a_t^R. This term plays a crucial role for determining the value of $\overline{w}_t$ based on the budget constraint. It represent the excess value of output over consumption under complete markets where the production is efficient and consumption is equalized by full risk sharing. Using the state contingent claims, whichever country that faces a high productivity shock in the future effectively promises to transfer a part of the increased production to the low productivity country so that the realized consumption of the two countries are the same in all realization of the state. And nx_t^{fb} captures precisely how much this transfer would be under complete markets.

Under incomplete markets, the lifetime budget constraint now involves ex post realization of the shocks. nx_t^{fb} can be thought of as an extra income at time t that can be used for consumption in all future periods in expectation. In other words, the discounted sum of nx_t is zero by the budget constraint, which itself is the sum of the discounted sums of $\widetilde{nx}_t$ and nx_t^{fb}. Facing a positive nx_t^{fb}, that country would enjoy a higher consumption net of production with $\widetilde{nx}_t < 0$, which turns out to be permanently constant in expectation under perfect capital mobility and flexible prices.

Second, consider the effect of a cost shock:

$$\overline{w}_t = -\frac{1-\beta}{A}\Big[\frac{2W_y D}{W_b(\sigma+\eta D)} E_t \sum_{k=0}^{\infty} \beta^k \mu_{t+k}^R\Big]$$

$$E_t\widetilde{w}_{t+k} = \overline{w}_t, \qquad E_t\widetilde{y}_{t+k}^R = -\frac{D-2a+1}{2(\sigma+\eta D)} E_t\overline{w}_t - \frac{D}{\sigma+\eta D} E_t\mu_{t+k}$$

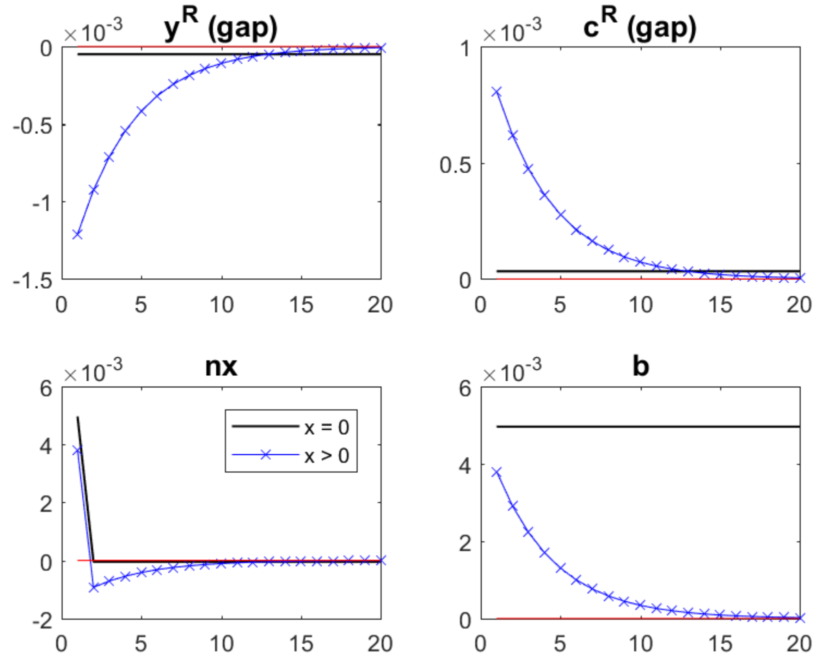
As in the previous case with productivity shock, UIP condition implies that relative consumption is perfectly smoothed in expectation. A positive cost shock in Home country is fully absorbed as higher relative price, which implies lower demand of Home goods. This leads to lower output and thus lower expected lifetime income of Home households relative to Foreign. Facing lower expected lifetime income, Home households lower the consumption in all current and future periods by the same amount.

Third, consider the effect of a UIP shock:

$$\overline{w}_t = \frac{1-\beta}{(1-\rho_f)(1-\beta\rho_f)} f_t$$

$$E_t\widetilde{w}_{t+k} = \overline{w}_t - \frac{\rho_f^k}{1-\rho_f} f_t, \qquad E_t\widetilde{y}_{t+k}^R = -\frac{D-2a+1}{2(\sigma+\eta D)} E_t\widetilde{w}_{t+k}$$

Figure 1: Impulse responses to a 1% iid ($\rho = 0$) shock on relative productivity under flexible price, across different degrees of capital mobility: $\chi \in \{0, 0.1\}$. The variables shown are: relative output gap $E_t\widetilde{y}^R_{t+k}$, relative consumption $E_t\widetilde{c}^R_{t+k}$, net export $E_t nx_{t+k}$, and net saving $E_t b_{t+k}$. [Solid black line]: perfect capital mobility ($\chi = 0$). [Blue 'x']: imperfect capital mobility ($\chi = 0.1$). For simplicity, no home bias ($a = 1/2$) and elastic labor supply ($\eta = 0$) are assumed.



The effect of a UIP shock is like a forced net saving.

In each period, relative output is immediately proportional to the relative consumption as a result of the flexible prices without cost shocks. However, unlike the case with productivity shock where perfectly smoothed consumption path implied equally smooth path of relative output in expectation, the UIP shock makes the expected path of relative consumption unequal across periods.

The figure shows the impulse response of key variables facing a shock to relative productivity. The permanent response under perfect capital mobility, as well as larger responses in the gap variables as well as convergence to the steady state under imperfect capital mobility, are evident from the figure. Note that the expected path of net export $E_t nx_{t+k} = E_t \left[\widetilde{nx}_{t+k} + nx^{fb}_{t+k}\right]$ need to satisfy the lifetime budget constraint: $E_t \sum_{k=0}^{\infty} \beta^k nx_{t+k} = -b_{t-1}$ (= 0 in the impulse response shown), regardless of the value of χ. This can be roughly verified from the impulse response of nx_t.

The distinct feature under $\chi = 0$ is that the impulse responses of all variables facing a productivity shock are flat and permanent in expectation, without converging to the steady state. This is of course a well known feature of the incomplete markets economy that exhibits a unit root, which arises as a result of perfect consumption smoothing over the infinite horizon. If $\chi > 0$, all variables would converge to the deterministic steady state.

1.3 Optimal Monetary Policy

1.3.1 Loss Function

The loss function measures the deviation of welfare from the first best allocation, typically approximated in the second-order:

$$(\text{Loss}) \equiv E_0 \sum_{t=0}^{\infty} \beta^t \Big(\underbrace{u(C_t^{fb}, N_t^{fb}) - u(C_t, N_t)}_{\approx \Psi_t} \Big) \geq 0$$

First best allocation is, as the name suggests, the optimal allocation that would prevail in the absence of any market frictions. In this model, as in most other New Keynesian models, it is equivalent to the allocation under the following three conditions: complete market, flexible prices, and optimal time-varying employment subsidy that is capable of nullifying any cost shocks. Importantly, it does not depend on the bond market structure (e.g., incomplete market), and the introduction of the modified UIP condition would not alter the loss function either. Hence it is the same as Corsetti et al. (2018) which contains the wealth gap as well as the misalignment term. I omit the derivation of the loss function here because it would be identical to what is derived in the previous literature.

Under PCP, $m_t \equiv 0$, and

$$\Psi_t = (\sigma+\eta)(\widetilde{y}_t^W)^2 + (\frac{\sigma}{D}+\eta)(\widetilde{y}_t^R)^2 + \frac{a(1-a)\phi}{D}(w_t)^2 + \frac{\epsilon}{2\delta}\Big((\pi_{Ht})^2 + (\pi_{Ft}^*)^2\Big)$$

Note that the quadratic terms on inflation can be expressed using W/R notation as

$$\frac{1}{2}\Big((\pi_{Ht})^2 + (\pi_{Ft}^*)^2\Big) = (\pi_t^W)^2 + (\pi_t^R)^2$$

Under LCP,

$$\begin{aligned}\Psi_t =& (\sigma+\eta)(\widetilde{y}_t^W)^2 + (\frac{\sigma}{D}+\eta)(\widetilde{y}_t^R)^2 + \frac{a(1-a)\phi}{D}(w_t+m_t)^2 \\ &+ \frac{\epsilon}{2\delta}\Big(a(\pi_{Ht})^2 + a(\pi_{Ft}^*)^2 + (1-a)(\pi_{Ht}^*)^2 + (1-a)(\pi_{Ft})^2\Big)\end{aligned}$$

The quadratic terms on inflation can be expressed as

$$\begin{aligned}\frac{1}{2}\Big(a(\pi_{Ht})^2 + a(\pi_{Ft}^*)^2 + (1-a)(\pi_{Ht}^*)^2 + (1-a)(\pi_{Ft})^2\Big) &= (\pi_t^W)^2 + a(\pi_t^R)^2 + (1-a)(\pi_t^X)^2 \\ &= (\pi_t^W)^2 + (\pi_t^{RC})^2 + a(1-a)(\Delta s_t)^2\end{aligned}$$

where $\pi_t^W \equiv \frac{1}{2}(\pi_{Ht}+\pi_{Ft}^*)$, $\pi_t^R \equiv \frac{1}{2}(\pi_{Ht}-\pi_{Ft}^*)$, $\pi_t^X \equiv \frac{1}{2}(\pi_{Ht}^*-\pi_{Ft})$, and π_t^{RC} is the relative CPI inflation.[17]

Comparing with the loss function under complete market as in Engel (2011), the only modification under incomplete market is the addition of term w_t^2, which turns out to

[17] All inflation variables are further explained in the appendix, along with the corresponding Phillips curves.

appear in the same manner as how m_t^2 appears for LCP. Under both LCP and incomplete market, these two terms appear together as $(w_t + m_t)^2$.[18] Setting $w_t = 0$, it is evident that the loss functions reduce to what is derived under complete market in Engel (2011).

The wealth gap appears in the loss function because the global planner cares both countries equally, and any inequality in consumption between the two countries is deemed inefficient. This channel was not present under complete market because there was full risk sharing, but it play a role here under incomplete market. As a result, there are now two distinct sources of inefficiency: the one arising from price dispersion due to sticky prices, and the one arising from the inequality in consumption due to imperfect risk sharing under incomplete market. In order to separate out the two distinct motives of the global planner, I also characterize the optimal monetary policy when only the inflation and output gap matters for the welfare and not the wealth gap. This would be a potentially useful policy as well, considering that the issue of redistribution between countries may be subject to more controversy compared to the other objectives.

1.3.2 Inefficiencies under Incomplete Markets

Based on the loss function derived above, we can now discuss the specific sources of inefficiencies present in this economy. The loss in welfare can arise from the world variables as well as from the relative variables:

$$\Psi_t = \underbrace{(\sigma+\eta)(\widetilde{y}_t^W)^2 + \frac{\epsilon}{\delta}(\pi_t^W)^2}_{\Psi_t^W} + \underbrace{(\frac{\sigma}{D}+\eta)(\widetilde{y}_t^R)^2 + \frac{a(1-a)\phi}{D}(\widetilde{w}_t)^2 + \frac{\epsilon}{\delta}(\pi_t^R)^2}_{\Psi_t^R}$$

[18]It is natural that these two terms appear in a parallel manner. A positive, symmetric misalignment implies that Home households are paying less than Foreign households for each good, so they can consume more just like when they have larger wealth. In fact, since $w_t = 2\sigma\widetilde{c}_t^R - (2a-1)\widetilde{s}_t^R - m_t$, w_t and m_t always appear in the form of $(w_t + m_t)$ except in the (relative) budget constraint.

As to be explained below, the world economy is like a closed economy, and as is well known, the monetary policy can cope with the productivity shock by stabilizing inflation, which also stabilizes the output by the divine coincidence. However, it turns out that the efficient level of welfare cannot be attained in terms of relative variables even with the optimal policy, which attains a higher welfare compared to the inflation (PPI) targeting policy.

Aside from the price dispersion, there are two sources of inefficiencies: the relative consumption $\widetilde{w}_t$, and the relative output $\widetilde{y}_t^R$. Under the efficient allocation that is attained under complete market, the wealth gap should be zero in all periods, and the level of output should be at the efficient level that reflects the productivity shock: $y_t^R = y_t^{Rfb} = \frac{(1+\eta)D}{\sigma+\eta D} a_t^R$.

Under incomplete markets, income risks due to the productivity shock are not shared ex ante. The country with higher productivity will be endowed with a higher lifetime income and thus higher lifetime consumption,[19] which is optimally allocated into higher expected consumption in all periods. Clearly, this gap in consumption itself is not efficient compared to the efficient allocation under complete market, where the two countries would have agreed to transfer the unexpected income upon a higher productivity shock to the low productivity country using state contingent claims, so that they reach the same level of consumption (net of real exchange rates). This is explicitly captured by the term $(\widetilde{w}_t)^2$ in the loss function.

This inefficient gap in consumption further propagates into inefficient levels of output. The country with higher productivity consumes more than the other, and this leads to

[19] This hold for $\phi > 1$ and the opposite holds for $\phi < 1$. Even in the latter case, the reason for the inefficiency would hold similarly.

higher marginal rate of substitution between leisure and consumption, which is equal to the real wages in equilibrium based on the household's optimal choice of labor supply. Then the higher wage in the high productivity country is passed on to the relative price, so that the goods produced in high productivity country are more expensive compared to the relative price under efficient allocation without the wage gap. The higher price leads to lower demand, and thus there is underproduction in the high productivity country and overproduction in the low productivity country. And this is the second source of inefficiency that is captured by the term $(\widetilde{y}_t^R)^2$ in the loss function.

Under flexible price, the passthrough from wages to prices occur immediately, and consequently the monetary policy cannot alter the allocation. In contrast, under sticky price, this passthrough from wages to prices can be delayed by using monetary policy that exploits the price stickiness. The gap in consumption is difficult or costly to reduce because it is an inherent feature of the incomplete markets. On the other hand, reducing the output gap while allowing for some price dispersion leads to an improvement in welfare compared to the PPI-targeting policy that replicates the flexible price allocation. In a sense, the optimal monetary policy does not reduce the first source of inefficiency (relative consumption), but achieves improvements in welfare by suppressing the transmission of inefficiencies from relative consumption to relative output, which is the second source of inefficiency.

1.3.3 Optimal Policy under PCP

One benefit of writing in terms of world and relative variables is that they can be entirely separated from each other in the dynamic optimization problem. That is, Lagrangian for the Ramsey planner's problem: $\mathcal{L} \equiv \mathcal{L}^W + \mathcal{L}^R$, where

$$\mathcal{L}^W = E_0 \sum_{t=0}^{\infty} \beta^t \left((\sigma + \eta)(\widetilde{y}_t^W)^2 + \frac{\epsilon}{\delta}(\pi_t^W)^2 + 2\gamma_t^W \left[\pi_t^W - \beta \pi_{t+1}^W - \delta \Big((\sigma + \eta) \widetilde{y}_t^W + \mu_t^W \Big) \right] \right)$$

$$\begin{aligned}
\mathcal{L}^R = E_0 \sum_{t=0}^{\infty} \beta^t & \Bigg((\frac{\sigma}{D} + \eta)(\widetilde{y}_t^R)^2 + \frac{a(1-a)\phi}{D} w_t^2 + \frac{\epsilon}{\delta}(\pi_t^R)^2 \\
& + 2\gamma_t^R \left[\pi_t^R - \beta \pi_{t+1}^R - \delta \Big((\frac{\sigma}{D} + \eta) \widetilde{y}_t^R + \frac{D - 2a + 1}{2D} w_t + \mu_t^R \Big) \right] \\
& + 2\lambda_t \Big[w_t - w_{t+1} - \chi b_t + f_t \Big] \\
& + 2\xi_t \Big[w_t - 2W_y \widetilde{y}_t^R + W_b \Big(b_t - \beta^{-1} b_{t-1} - nx_t^{fb} \Big) \Big] \Bigg) \\
\equiv E_0 & \sum_{t=0}^{\infty} \beta^t \mathcal{L}_t^R
\end{aligned}$$

where $W_y \equiv \frac{2a(\sigma\phi - 1) + 1 - \sigma}{2a(\phi - 1) + 1}$, $W_b \equiv \frac{D}{1-a} \frac{1}{2a(\phi - 1) + 1}$.

The first order necessary conditions for the world variables are:

$$(\pi_t^W) : \frac{\epsilon}{\delta} \pi_t^W + \gamma_t^W - \gamma_{t-1}^W = 0$$

$$(\widetilde{y}_t^W) : \widetilde{y}_t^W - \delta \gamma_t^W = 0$$

together with PC_t^W. The two FOC's $(\pi_t^W, \widetilde{y}_t^W)$ and the world Phillips Curve can be used to solve for $\{\pi_t^W, \widetilde{y}_t^W, \gamma_t^W\}$.

The first order necessary conditions for the relative variables are:

$$(\pi_t^R) : \frac{\epsilon}{\delta} \pi_t^R + \gamma_t^R - \gamma_{t-1}^R = 0$$

$$(\widetilde{y}_t^R) : (\frac{\sigma}{D} + \eta) \Big(\widetilde{y}_t^R - \delta \gamma_t^R \Big) - 2W_y \xi_t = 0$$

$$(w_t) : \frac{a(1-a)\phi}{D} w_t - \frac{D - 2a + 1}{2D} \delta \gamma_t^R + \Big(\lambda_t - \beta^{-1} \lambda_{t-1} \Big) + \xi_t = 0$$

$$(b_t) : W_b\Big(\xi_t - E_t\xi_{t+1}\Big) - \chi\lambda_t = 0$$

together with PC_t^R, modified UIP condition, and the budget constraint. The four FOC's $(\pi_t^R, \widetilde{y}_t^R, w_t, b_t)$ and the three constraints can be used to solve for $\{\pi_t^R, \widetilde{y}_t^R, w_t, b_t, \gamma_t^R, \lambda_t, \xi_t\}$.[20]

Optimal Policy in Relative Variables

The optimal monetary policy that involves the world variables would be exactly the same as that in a closed economy, and can be characterized by a simple targeting rule. In contrast, the optimal policy regarding the relative variables is far more complex. Facing this discrepancy, in this subsection, I will focus on the discussion of the optimal monetary policy for the relative variables. The discussion of the world policy is postponed until later where the targeting rules are discussed, both for the world and relative variables.

The degree of price stickiness is an important parameter for characterizing the optimal monetary policy. In particular, it directly affects the relative weights on the loss function as derived above. First, consider the flexible price case, where $\alpha = 0$, and $\delta = \frac{(1-\alpha)(1-\alpha\beta)}{\alpha} \to \infty$. At first glance, it looks as if the weight on the inflation goes to zero. However, considering the Phillips curve:

$$\pi_t^R = \beta E_t\pi_{t+1}^R + \delta mc_t^R = \delta E_t \sum_{k=0}^{\infty} \beta^k mc_{t+k}^R$$

the loss function for the relative terms can be rewritten as:

$$\Psi_t^R = (\frac{\sigma}{D} + \eta)(\widetilde{y}_t^R)^2 + \frac{a(1-a)\phi}{D}(\widetilde{w}_t)^2 + \epsilon\delta\Big(\sum_{k=0}^{\infty} \beta^k mc_{t+k}^R\Big)^2$$

where $mc_{t+k}^R = (\frac{\sigma}{D} + \eta)\widetilde{y}_{t+k}^R + \frac{D-2a+1}{2D} w_{t+k}$. When expressed in terms of inflation, the

[20] The number of variables can be reduced by plugging in the budget constraint for w_t, as well as FOC-(π_t^R): $\gamma_t^R = \frac{\epsilon}{\delta} p_t^R$, which is the CDL's approach. There will be 2 FOC's $(\widetilde{y}_t^R, b_t)$ and two constraints excluding the budget constraint, which can be used to solve for $\{\pi_t^R, \widetilde{y}_t^R, b_t, \lambda_t\}$.

weight on inflation was infinitesimally small. This reflects the fact that the price dispersion is very small given that almost all firms get to reset their prices. But in terms of the marginal cost, this weight is infinitely large under flexible price. This reflects the fact that it would be difficult to sustain the comparable deviations in the marginal costs under nearly flexible prices compared to the stick price. If such sizeable deviations in the marginal costs are actually to be implemented, it would involve a very large cost in terms of welfare. Facing this infinite weight on the marginal cost, the planner would focus only on minimizing the size of marginal costs down to zero, without considering the other two objectives at all which have infinitesimal weights compared to the marginal costs. In other words, the flexible price limit can be thought of as the planner minimizing the $(\sum_{k=0}^{\infty} \beta^k mc_{t+k}^R)^2$ part of the loss function.

Second, consider the case with fully sticky price, which corresponds to $\alpha = 1$, and $\delta = 0$. The situation is now exactly reversed. Inflation is infinitely costly, but the marginal cost is not costly at all. Getting towards the fully sticky price limit, the marginal costs can be easily manipulated by the monetary policy without affecting the inflation, because almost no firms can reset the price anyway. Facing this zero weight on the marginal costs, the planner would focus only on minimizing the non-inflation part of the loss function, i.e., $(\frac{\sigma}{D} + \eta)(\widetilde{y}_t^R)^2 + \frac{a(1-a)\phi}{D}(w_t)^2$.

As a result, under moderately sticky prices, the planner would minimize some weighted average of the two objectives — the inflation part versus the non-inflation part. A higher price stickiness would make the planner care more about the non-inflation part of the loss function and less about the inflation part, and vice versa. And this is why comparing the limiting case of fully sticky price is useful for understanding the optimal monetary policy under general degree of price stickiness, together with the PPI-targeting policy that minimizes the inflation part of the loss function and recovers the flexible price equilibrium.

In the following subsections, I will first show the optimal monetary policy under $\chi = 0$, which is the case of Corsetti et al., and then proceed with $\chi > 0$. In each of the two cases, I will characterize the special case of fully sticky price, which is indeed useful for understanding the intuition for the optimal policy. These will be the mechanical solutions to the optimization problem of the planner, and based on these quantitative solutions by themselves, it is hard to fully comprehend the underlying intuitions. Later, I will point out three "properties" of the optimal monetary policy, and explain in a qualitative manner why the output gap in the short run should be reduced at the expense of all other policy objectives,

Perfect capital mobility ($\chi = 0$)

The strategy for obtaining the analytic solution is: first, rewrite endogenous variables such as $\widetilde{y}_t^R$ in terms of γ_t^R as well as the constant expected paths of ξ_t and w_t; second, solve γ_t^R in terms of ξ_t and w_t using FOC-(π_t^R) and (PC); third, use the (BC) and FOC-(w_t) to subsitute out λ_t's and b_t's; and finally, solve for $\overline{\xi}_t$ and $\overline{w}_t$ by constructing $\sum\limits_{k=0}^{\infty} \beta^k \gamma_t^R$ in terms of $\overline{\xi}_t$ and $\overline{w}_t$ and combining with the lifetime versions of (BC) and FOC-(w_t). The full derivation is shown in the appendix.

Because the analytic solution in the fully general case (shown in the appendix) is too complicated to yield useful intuitions, here I focus on the special case of no home bias ($a = \frac{1}{2}$) and perfectly elastic labor supply ($\eta = 0$). The corresponding solution, which is still fairly involved, would be:

$$\bar{c}_t^R = \frac{\frac{1}{\sigma} + \frac{1-\nu_1}{1-\beta\nu_1}(\phi-1)}{D^{rp}}(1-\beta)\phi\Big(a_t^R + \frac{1}{\phi-1}\beta^{-1}b_{t-1}\Big) + \frac{\frac{\nu_1(1-\beta)}{1-\beta\nu_1}}{D^{rp}}\Big(\delta\gamma_{t-1}^R + 2(\phi-1)(1-\beta)\beta^{-1}\lambda_{t-1}\Big)$$

where $D^{rp} \equiv h + \frac{1}{\sigma(\phi-1)} + \frac{1-\nu_1}{1-\beta\nu_1}(1+\sigma\phi h)$ and $h \equiv (\phi-1) + \frac{1}{\sigma}$. Note that under fully sticky price, $\delta = 0$ implies $\nu_1 = 1$, and thus $\frac{1-\nu_1}{1-\beta\nu_1} = 0$. On the other hand, under flexible price, $\delta \to \infty$ implies $\nu_1 = 0$, and thus $\frac{1-\nu_1}{1-\beta\nu_1} = 1$.

It is useful to compare this solution with the flexible price case, which is the allocation with PPI-targeting policy instead of the optimal monetary policy: $\bar{c}_t^{R,na} = \dfrac{(1-\beta)(\phi-1)}{\sigma h}a_t^R$, where the superscript "na" stands for the "natural allocation" under flexible price, which is to be attained by the PPI-targeting policy. Denote the current solution of the Ramsey planner by the superscript "rp". Assuming that the economy starts at the efficient steady state with $b_{t-1} = \gamma_{t-1} = \lambda_{t-1} = 0$, which is without loss of generality,

$$\frac{\bar{c}_t^{R,rp}}{\bar{c}_t^{R,na}} = \frac{\phi h + \frac{1-\nu_1}{1-\beta\nu_1}\phi(\phi-1)\sigma h}{(\frac{1-\nu_1}{1-\beta\nu_1} - 1)(\phi-1) + \phi h + \frac{1-\nu_1}{1-\beta\nu_1}\phi(\phi-1)\sigma h}$$

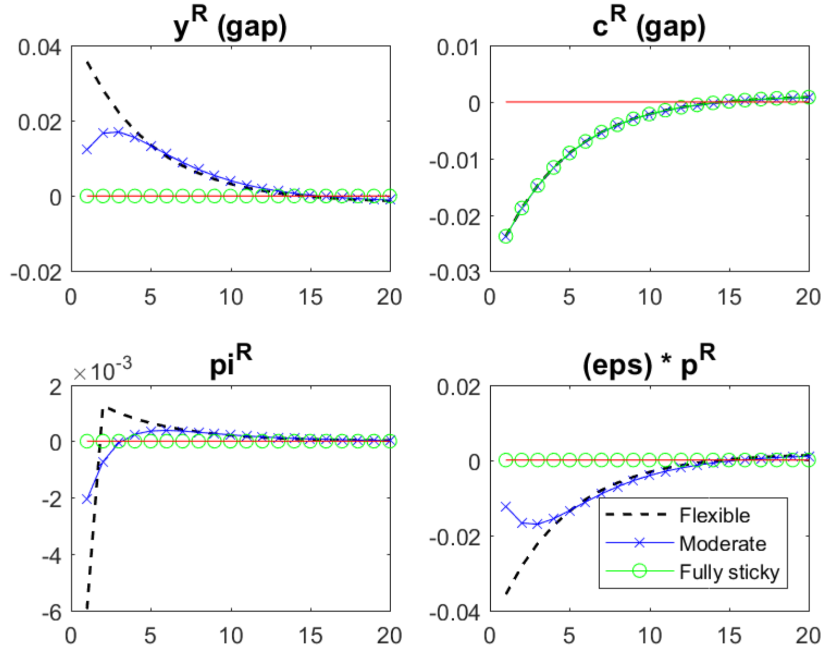
Clearly, $\frac{1-\nu_1}{1-\beta\nu_1} - 1 = -\frac{\nu_1(1-\beta)}{1-\beta\nu_1} < 0$. Given $\phi > 1$, which is a more realistic calibration,[21] it turns out that $|\bar{c}_t^{R,rp}| > |\bar{c}_t^{R,na}|$. That is, the optimal policy would allow for a larger gap in consumption, in addition to a trivially larger inflation, compared to the PPI-targeting policy. It can be inferred that the gains from reducing the relative output gap outweighs the costs in these two other objectives.

Given this solution for $\bar{c}_t^R$, all other variables can also be solved analytically, for example:

$$2\bar{\xi}_t = -\frac{\frac{1}{\sigma} + \frac{1-\nu_1}{1-\beta\nu_1}\phi}{D^{rp}}(1-\beta)\phi a_t^R$$

[21] As usual, $\phi < 1$ implies that high productivity country would end up with smaller wealth and consumption, which arises due to the complementarity between Home and Foreign goods. It turns out that when $\phi < 1$, $|\bar{c}_t^{R,rp}| < |\bar{c}_t^{R,na}|$.

Figure 2: Optimal policy responses to a 1% shock to UIP with persistence $\rho = 0.8$ under perfect capital mobility ($\chi = 0$), across different degrees of price stickiness: $\alpha \in \{0, 0.75, 1\}$. The variables shown are: relative output gap $E_t\widetilde{y}^R_{t+k}$, relative consumption $E_t\widetilde{c}^R_{t+k}$, relative inflation $E_t\pi^R_{t+k} \equiv \frac{1}{2}E_t[\pi_{Ht+k} - \pi^*_{Ft+k}]$, and relative price level $E_t p^R_{t+k} \equiv \frac{1}{2}E_t[p_{Ht+k} - p^*_{Ft+k}]$ multiplied by $\epsilon = 6$.[22] [Dashed black line]: flexible price ($\alpha = 0$). [Blue 'x']: moderately sticky price ($\alpha = 0.75$, baseline calibration). [Green 'o']: fully sticky price ($\alpha = 1$). For simplicity, no home bias ($a = 1/2$) and elastic labor supply ($\eta = 0$) are assumed.



$$E_t\gamma^R_{t+k} = -\frac{1-\nu_1^{k+1}}{\delta D^{rp}}(1-\beta)\phi a^R_t$$

$$E_t\widetilde{y}^R_{t+k} = -\frac{\phi\left(1 + \frac{1-\nu_1}{1-\beta\nu_1}\sigma(\phi-1)\right) - \nu_1^{k+1}}{D^{rp}}(1-\beta)\phi a^R_t$$

As noted above, flexible price implies $\nu_1 = 0$ and fully sticky price implies $\nu_1 = 1$. In both cases, $E_t\widetilde{y}^R_{t+k}$ is constant for all k. However, for an intermediate degree of price stickiness, $\nu_1 \in (0,1)$, and thus the expected path of relative output gap converges to a certain level different from zero in the long run. In particular, it turns out that $\lim_{k\to\infty} \frac{E_t\widetilde{y}^R_{t+k}}{\bar{c}_t} = -\sigma\phi$, which is the ratio under flexible price.

[22]By scaling the price level by ϵ, the "targeting rule" ($E_t\left[\widetilde{y}^R_{t+k} + \epsilon p^R_{t+k}\right] =$ constant) can be easily

Figure 3: Optimal policy responses to a 1% shock to relative productivity with persistence $\rho = 0.8$ under perfect capital mobility ($\chi = 0$), across different degrees of price stickiness: $\alpha \in \{0, 0.75, 1\}$. The variables shown are: relative output gap $E_t\widetilde{y}^R_{t+k}$, relative consumption $E_t\widetilde{c}^R_{t+k}$, relative inflation $E_t\pi^R_{t+k}$, and relative price level $E_t p^R_{t+k}$ multiplied by $\epsilon = 6$. [Dashed black line]: flexible price ($\alpha = 0$). [Blue 'x']: moderately sticky price ($\alpha = 0.75$, baseline calibration). [Green 'o']: fully sticky price ($\alpha = 1$). For simplicity, no home bias ($a = 1/2$) and elastic labor supply ($\eta = 0$) are assumed.

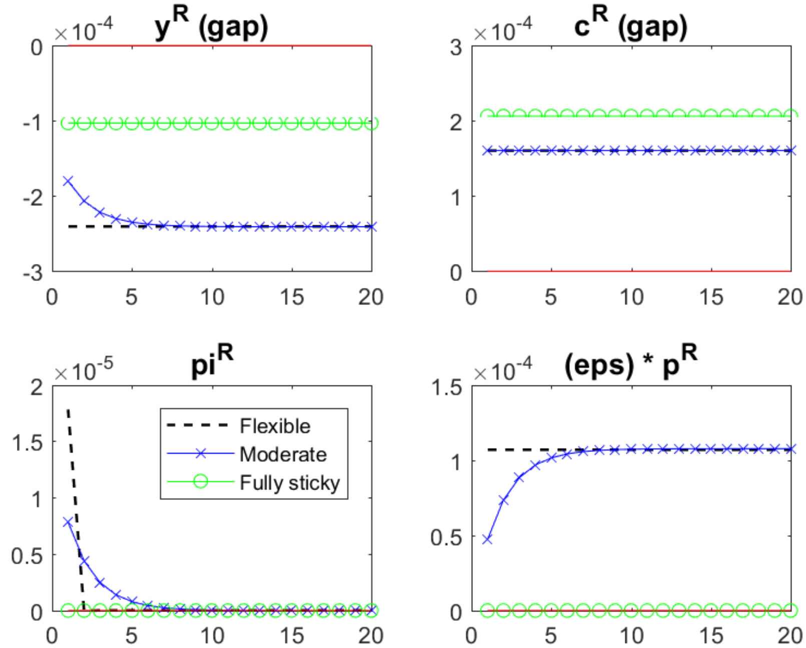
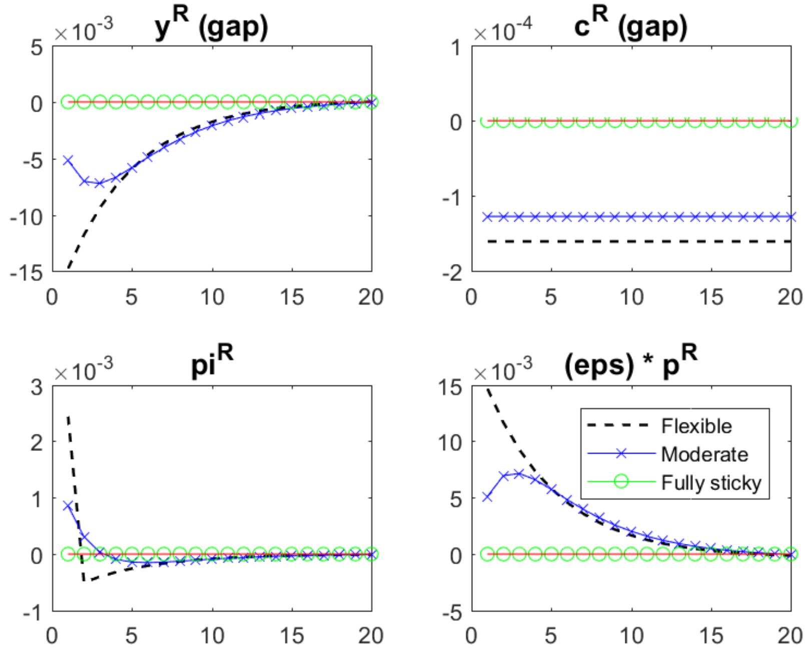


Figure 4 shows the impulse response of the optimal allocation in response to a 1% shock on relative productivity. Under fully flexible price, the price levels are irrelevant, but the impulse responses shown are the optimal policy in the limit as α approaches zero. As usual, the corresponding allocation $\{\widetilde{y}^R_t, \widetilde{c}^R_t\}$ under flexible price can be replicated under arbitrary degree of price stickiness by using the PPI-targeting policy.

Although $\widetilde{y}^R_t, \widetilde{c}^R_t$, and p^R_t all seem to converge to the same permanent level as the flexible price limit, there are small differences as can be seen in the analytic solution derived above. In order to sustain a larger difference, the value of $\frac{1-\nu_1}{1-\beta\nu_1}$ should be closer

verified.

Figure 4: Optimal policy responses to a 1% shock to relative cost with persistence $\rho = 0.8$ under perfect capital mobility ($\chi = 0$), across different degrees of price stickiness: $\alpha \in \{0, 0.75, 1\}$. The variables shown are: relative output gap $E_t\widetilde{y}^R_{t+k}$, relative consumption $E_t\widetilde{c}^R_{t+k}$, relative inflation $E_t\pi^R_{t+k}$, and relative price level $E_tp^R_{t+k}$ multiplied by $\epsilon = 6$. [Dashed black line]: flexible price ($\alpha = 0$). [Blue 'x']: moderately sticky price ($\alpha = 0.75$, baseline calibration). [Green 'o']: fully sticky price ($\alpha = 1$). For simplicity, no home bias ($a = 1/2$) and elastic labor supply ($\eta = 0$) are assumed.



to zero (fully sticky price) and farther away from one (flexible price). But based on this functional form, it is difficult to bring this value close to zero. For example, if the prices are 10 times more sticky ($\alpha = 1 - 0.25/10$), $\nu_1 \approx 0.95$, and $\frac{1-\nu_1}{1-\beta\nu_1} \approx 0.85$. In sum, within the realistic range of parameter values, these permanent levels of the impulse responses converge fairly close to those under flexible price, although never exactly the same.

Qualitative Properties of the Optimal Monetary Policy

In this subsection, I describe additional properties of the optimal monetary policy that would help to understand the quantitative results.

First, it is difficult for the planner to manipulate the expected path of relative consumption. When $\chi = 0$, this property follows trivially from the modified UIP condition, which implies $E_t \Delta w_{t+1} = f_t$ or $w_t = \overline{w}_t - \frac{1}{1-\rho_f} f_t$. If the planner wants to increase relative consumption in a certain period, it has to increase relative consumption in all periods in expectation. If $\chi > 0$, manipulation of the expected path of relative consumption is not infeasible, but still it is much more costly compared to manipulating the relative output gap.

Suppose the planner attempts to increase the relative consumption in period $t+1$ ($E_t w_{t+1}$) by X, while holding the value in all other periods $E_t w_{t+k}$ constant. The modified UIP condition:

$$b_t = \frac{1}{\chi}\big(w_t - E_t w_{t+1} + f_t\big)$$

implies that b_t should decrease by X/χ, and b_{t+1} should increase by X/χ. Then based on the budget constraint:

$$w_t - 2W_y \widetilde{y}_t^R + W_b\big(b_t - \beta^{-1} b_{t-1} - nx_t^{fb}\big) = 0$$

$\widetilde{y}_{t+1}^R = \frac{1}{2W_y} w_t + \frac{W_b}{2W_y}(b_{t+1} - \beta^{-1} b_t)$ would need to increase by $\frac{1}{2W_y}\big(1 + (1+\beta^{-1})W_b/\chi\big)X$. As $\chi \to 0$, the required increase in $\widetilde{y}_{t+1}^R$ is infinite, reflecting that it is infinitely costly to marginally change the relative consumption in a particular period. When χ is close to zero, such manipulation becomes feasible, but comes at a high cost in the sense that it requires a large change in $\widetilde{y}_t^R$.

Here is an interpretation of this result. An increase in relative consumption in a particular period means an increasing profile before this period and a decreasing one afterward. In order to sustain an increasing consumption profile, a lower excess return is required, and this would lead to a dissaving in that country. As a result, a country should decrease net saving before an increased consumption, and likewise increase net saving after an increased consumption. But this is actually an awkward direction, especially if the stream of income is smooth. In order to sustain this consumption profile, the relative output gap in that period should be increased by a more pronounced magnitude, that is, multiplied by $\frac{1}{2W_y}\left(1+(1+\beta^{-1})W_b/\chi\right)$.

The key consequence of this property is that the planner would mainly alter the expected path of $\widetilde{y}^R_{t+k}$ rather than w_{t+k}. The main policy tradeoff would then be between stabilizing the output gap and reducing inflation (or the relative price dispersion), the most typical tradeoff for optimal monetary policy.

The second property is that the optimal relative inflation is zero in the long run. If $\chi > 0$, all the real variables including mc^R_t are stationary, so the long run inflation is trivially zero *for any* monetary policy rule. In contrast, if $\chi = 0$, there exist monetary policy rules that sustains nonzero relative inflation in the long run. Still, the optimal monetary policy would aim to set the relative inflation equal to zero in the long run.

This property follows immediately from the first order condition:

$$E_t\left[\Delta\widetilde{y}^R_{t+1} + \epsilon\pi^R_{t+1}\right] = -\frac{2W_y}{\frac{\sigma}{D}+\eta}E_t\Delta\xi_{t+1} = \frac{2W_y}{\frac{\sigma}{D}+\eta}\frac{\chi}{W_b}\lambda_t$$

which is equal to zero if $\chi = 0$. Suppose that the long run inflation converges to a nonzero value, i.e., $\lim_{k\to\infty} E_t\pi_{t+k} = \overline{\pi}^R_t \neq 0$. Then $\lim_{k\to\infty} E_t\Delta\widetilde{y}^R_{t+k} = -\epsilon\overline{\pi}^R_t \neq 0$. This implies that the relative output gap would grow indefinitely in the long run, which is clearly suboptimal.

There is also a more intuitive explanation for this result. This conclusion can be drawn directly from the Phillips curve that the current inflation is the discounted sum of current and all future expected marginal costs:

$$\pi_t^R = E_t \beta \pi_{t+1}^R + \delta mc_t^R = \delta E_t \sum_{k=0}^{\infty} \beta^k mc_{t+k}^R$$

The current marginal cost contributes to the current inflation but not to any of the future inflations. But the expected marginal cost at $t+k$ contributes to all current and future inflations up to period $t+k$. This means that the cost of maintaining a nonzero marginal cost is disproportionately more costly in the long run. As a result, the optimal marginal cost is zero in the long run, and this in turn implies that the optimal inflation is zero in the long run.

Targeting Rule in World Variables

The optimal policy in world variables is straightforward to fully solve and characterize. A targeting rule is a necessary condition for the optimal policy, and summarizes the key relationship between output gap and price level under timeless perspective. Substituting $\gamma_t^W = \frac{1}{\delta}\widetilde{y}_t^W$ from FOC-($\widetilde{y}_t^W$) into FOC-(π_t^W),

$$\epsilon \pi_t^W + \widetilde{y}_t^W - \widetilde{y}_{t-1}^W = \left(\widetilde{y}_t^W + \epsilon p_t^W\right) - \left(\widetilde{y}_{t-1}^W + \epsilon p_{t-1}^W\right) = 0$$

$$\left(\widetilde{y}_t^W + \epsilon p_t^W\right) = E_t\left(\widetilde{y}_{t+1}^W + \epsilon p_{t+1}^W\right) = \cdots = \left(\widetilde{y}_{t-1}^W + \epsilon p_{t-1}^W\right)$$

That is, sum of the level of output gap and the level of world price level should match the previously committed value. This commitment rule also holds for all future periods. This sum rule is in fact the same as not only CDL, but also as Engel, CGG, or even the one in a closed economy. This is because the world economy as a whole is just like a closed economy. Regardless of whether the market is complete or not, the world budget constraint is just the world resource constraint, where there is no delaying or advancing of the use of real resource. As a result, even if there is a shock in world variables including the world cost shock, the world targeting rule implies that the value of $\widetilde{y}_t^W + \epsilon p_t^W$ should remain the same in all t and in all state of the economy.

The intuition behind this targeting rule under commitment can be understood by looking at the structure of Phillips curve. A marginal increase in π_t^W would increase the value of the discounted Phillips curve at t by β^t, and decrease the corresponding value at $t-1$ by β^t. As a result, the FOC with respect to π_t^W is simply $\frac{\epsilon}{\delta}\pi_t^W + \gamma_t^W - \gamma_{t-1}^W = 0$. From this FOC, γ_t^W is equal to the level of price p_t^W, up to a constant that depends on the commitment made in the distant past, which can be assumed to be zero without loss of generality. In turn, the FOC regarding output gap simply states that marginal cost of increasing $\widetilde{y}_t^W$ is the same as the marginal benefit from relaxing the Phillips curve at t, which is just $\delta\gamma_t^W = -\epsilon p_t^W$.

Targeting Rule in Relative Variables

Optimal monetary policies are typically characterized by some version of the "targeting rule," where the monetary authority can achieve optimal policy by managing output gap and inflation to satisfy the target criterion derived from the model. For example, in a closed economy under commitment, a target criterion is typically of the following form:

$$\pi_t + \phi(\widetilde{y}_t - \widetilde{y}_{t-1}) = 0$$

In the current model, while a similar targeting rule would be optimal for the World variables, such a simple targeting rule is not available for the relative variables.

From the first order conditions, it is straightforward to derive the following necessary condition for optimality:

$$\widetilde{y}_t^R + \epsilon p_t^R = \frac{2W_y}{\frac{\sigma}{D} + \eta} \xi_t$$

Note that in the corresponding equation for World variables, the right hand side was equal to zero, assuming initial values at the steady state. Combined with $\xi_t - E_t \xi_{t+1} = \frac{\chi}{W_b} \lambda_t$,

$$E_t \Big[\Delta \widetilde{y}_{t+1}^R + \epsilon \pi_{t+1}^R \Big] = -\frac{2W_y}{\frac{\sigma}{D} + \eta} \frac{\chi}{W_b} \lambda_t$$

When $\chi = 0$,[23] the above equation looks similar to the 'typical' targeting rule, collapsing to the one derived in Corsetti et al. As these authors have noted, this targeting rule involves expectation and does not hold in each state of the economy. This is different from the ones without expectation, typically derived under the assumption of complete markets.

This targeting rule follows from the FOC-(b_t): $\xi_t - E_t \xi_{t+1} = 0$. This describes precisely the nature of incomplete market where the risk cannot be fully insured. The current choice of b_t cannot respond to shocks realized at $t+1$, and thus the best that the planner (and of course the individual agent as well) can do is to choose the optimal b_t in expectation. As a result, the expected path of future optimal policy will actually be followed only when the realized values of future shocks are zero. In face of nonzero realization of future shocks, the actual path of output gap and inflation will need to adjust in each period.

This version of the targeting rule includes a Lagrange multiplier λ_t unless $\chi = 0$. This is not a desirable feature because policy advice that involves Lagrange multipliers would be difficult to interpret or implement.[24] Recently, Giannoni and Woodford (2017)

[23] This result coincides with the targeting rule derived in Corsetti et al. (2018), as it should.

[24] Also see Devereux et al. (2019), which offers an alternative approach when characterization of the optimal policy rule is too complex to be used as a guide to the policy in the real world.

showed that a characterization of the targeting rule is possible in a fairly general environment.[25] Following the procedure suggested in that paper, it is possible to characterize the targeting rule without involving the Lagrange multipliers or the shocks. First, the target variable z_t can be constructed as[26]

$$\begin{aligned}
z_t \equiv & \Big(\widetilde{y}_t^R - (2+\beta^{-1} - \frac{\chi}{W_b})\widetilde{y}_{t-1}^R + (2\beta^{-1} + 1 - \frac{\chi}{W_b})\widetilde{y}_{t-2}^R - \beta^{-1}\widetilde{y}_{t-3}^R\Big) \\
& + \epsilon\Big(\pi_t^R - \Big[1+\beta^{-1} + \frac{\chi}{W_b}(1 - \frac{D-2a+1}{\sigma+\eta D}W_y)\Big]\pi_{t-1}^R + \beta^{-1}\pi_{t-2}^R\Big) \\
& - \frac{2W_y\chi}{W_b}\frac{a(1-a)\phi}{\sigma+\eta D}\Big(w_{t-1} - w_{t-2}\Big) \\
= & \Big((1-L)(1-\beta^{-1}L)(\Delta\widetilde{y}_t^R + \epsilon\pi_t^R)\Big) \\
& + \frac{\chi}{W_b}\Big(\Delta\widetilde{y}_{t-1}^R - \Big[1 - \frac{D-2a+1}{\sigma+\eta D}W_y\Big]\epsilon\pi_{t-1}^R - 2W_y\frac{a(1-a)\phi}{\sigma+\eta D}\Delta w_{t-1}\Big)
\end{aligned}$$

where L is the lag operator. Then the optimal targeting rule can be derived as

$$E_t z_{t+1} = \phi_1\big(z_t - E_{t-1}z_t\big) + \phi_2\big(z_{t-1} + E_{t-2}z_{t-1}\big)$$

where $(\phi_1, \phi_2) = \big(-(1+\beta^{-1}), \beta^{-1}\big)$ for the current model. $\big(z_t - E_{t-1}z_t\big)$ can be interpreted as the forecast revision of the targeting variable at t. Because the predetermined part of z_t cancels out, $\big(z_t - E_{t-1}z_t\big)$ can be expressed more concisely as $\big(\widetilde{y}_t^R + \epsilon p_t^R\big) - E_{t-1}\big(\widetilde{y}_t^R + \epsilon p_t^R\big)$.

This result is consistent with the general form of the targeting rule shown in Giannoni and Woodford (2017). In this exercise, there are 3 constraints and 4 endogenous variables, requiring one monetary policy tool in relative terms. One of the 3 constraints is backward looking (budget constraint), and two are forward looking (modified UIP,

[25] However, it cannot be applied to the Corsetti et al.'s environment with $\chi = 0$ because it violates a regularity condition [Assumption 2-(b)] of Giannoni and Woodford (2017). This is not surprising because the steady state does not exist with $\chi = 0$, and that regularity condition was designed to preclude such a case.

[26] Derivation of this targeting rule is shown in the appendix.

Phillips curve). This combination leads to the above targeting rule that is forward looking by one period and backward looking (lags) by two periods.[27]

1.4 Optimal Capital Control

Suppose the planner can discriminate intermediaries and household regarding the interest rate for the bonds. In this case, the planner can separately choose the expected path of consumption which is governed by the expected path of interest rate that the households face, and the current account balance which is determined by the expected excess return that only the intermediaries face.

This capital control can be described as the households facing the nominal interest rate of R_t, R_t^*, respectively, while the intermediaries face $(1+\tau_t)R_t, (1+\tau_t^*)R_t^*$, where $\tau^{ss} = (\tau^*)^{ss} = 0$. The intermediaries' profit would then be:

$$\Pi_{t+1}^{\mathcal{I}} = (1+\tau_t)R_t D_t + \mathcal{E}_{t+1}(1+\tau_t^*)R_t^* D_t^* = \Big((1+\tau_t)R_t - (1+\tau_t^*)R_t^* \frac{\mathcal{E}_{t+1}}{\mathcal{E}_t}\Big) D_t$$

The log-linearized expected excess return is now

$$E_t er_{t+1} = \underbrace{(r_t - r_t^*) - E_t \Delta e_{t+1}}_{E_t \Delta w_{t+1}} + (\tau_t - \tau_t^*) = -\chi b_t + f_t$$

$$\therefore E_t \Delta w_{t+1} = -\chi b_t - (\tau_t - \tau_t^*) + f_t$$

Since it is always the difference in the tax rate $(\tau_t - \tau_t^*)$ that matters, from here on let τ_t stand for $\tau_t - \tau_t^*$ without loss of generality.

[27] The case of "typical" targeting rule, shown as an example in Giannoni and Woodford (2017), is indeed considerably simpler than what appears here. Although their example also consists of one backward looking constraint and two forward looking, "the reduced system of equations written in terms of the "essential" state variables contains no backward-looking structural relations." This results in the right hand side of the targeting rule being equal to zero, without involving any forecast revisions.

With the availbility of such time-varying capital tax,[28] the planner can now solve the optimization problem without the constraint on the evolution of w_t: $[w_t - w_{t+1} - \chi b_t + f_t = 0]$, because it will now be modified as $[w_t - w_{t+1} - \chi b_t + f_t - \tau_t = 0]$ where τ_t is a free policy variable that can be optimally chosen at the planner's discretion. In terms of FOC's, since this UIP constraint is no longer binding, the corresponding multiplier $\{\lambda_t\}$ would be zero.[29] Thus the relevant question would be, what would the optimal allocation look like, and how should the planner set the capital tax/subsidy facing intermediaries, in addition to the monetary policy?

Lagrangian for the world variables, $\mathcal{L}^W$, is exactly the same as before. Lagrangian for the relative variable:

$$\begin{aligned}
\mathcal{L}^R = E_0 \sum_{t=0}^{\infty} \beta^t \Bigg(& (\frac{\sigma}{D} + \eta)(\widetilde{y}_t^R)^2 + \frac{a(1-a)\phi}{D} w_t^2 + \frac{\epsilon}{\delta}(\pi_t^R)^2 \\
& + 2\gamma_t^R \Big[\pi_t^R - \beta \pi_{t+1}^R - \delta \underbrace{\Big((\frac{\sigma}{D} + \eta)\widetilde{y}_t^R + \frac{D - 2a + 1}{2D} w_t + \mu_t^R \Big)}_{mc_t^R} \Big] \\
& + 2\lambda_t \Big[w_t - w_{t+1} - \tau_t - \chi b_t + f_t \Big] \\
& + 2\xi_t \Big[w_t - 2W_y \widetilde{y}_t^R + W_b \Big(b_t - \beta^{-1} b_{t-1} - nx_t^{fb} \Big) \Big] \Bigg) \\
\equiv E_0 & \sum_{t=0}^{\infty} \beta^t \mathcal{L}_t^R
\end{aligned}$$

where $W_y \equiv \frac{2a(\sigma\phi - 1) + 1 - \sigma}{2a(\phi - 1) + 1}$, $W_b \equiv \frac{D}{1-a} \frac{1}{2a(\phi-1)+1}$. The addition of τ_t for the modified UIP condition (with the multiplier λ_t) is the only change compared to the problem without capital control.

[28] The availability of time-varying capital tax in the real world is somewhat controversial. Some authors advocate the use of sterilized intervention. See e.g. Liu and Spiegel (2015) and Prasad (2018).

[29] The Lagrangians as well as the first order conditions are shown in the appendix, which are quite similar to the problems without capital control.

The first order necessary conditions for the relative variables are:

$$(\pi_t^R) : \frac{\epsilon}{\delta}\pi_t^R + \gamma_t^R - \gamma_{t-1}^R = 0$$

$$(\widetilde{y}_t^R) : (\frac{\sigma}{D} + \eta)\Big(\widetilde{y}_t^R - \delta\gamma_t^R\Big) - 2W_y\xi_t = 0$$

$$(w_t) : \frac{a(1-a)\phi}{D}w_t - \frac{D-2a+1}{2D}\delta\gamma_t^R + \Big(\lambda_t - \beta^{-1}\lambda_{t-1}\Big) + \xi_t = 0$$

$$(b_t) : W_b\Big(\xi_t - E_t\xi_{t+1}\Big) - \chi\lambda_t = 0$$

$$(\tau_t) : \lambda_t = 0$$

together with PC_t^R, modified UIP condition, and the budget constraint. The five FOC's $(\pi_t^R, \widetilde{y}_t^R, w_t, b_t, \tau_t)$ and the three constraints can be used to solve for $\{\pi_t^R, \widetilde{y}_t^R, w_t, b_t, \gamma_t^R, \xi_t, \lambda_t, \tau_t\}$.

By inspection, it can be seen immediately that the parameter χ no longer appears in the characterization of the optimal capital control, because the constraint that contains χ is no longer in effect. This means that optimal capital control can completely nullify the inefficiency that arises from the financial segmentation. As a result, the optimal allocation would resemble the one with perfect capital mobility. In addition, because one of the constraints has been removed, the welfare can only increase at least weakly under this policy. Moreover, this policy can entirely insulate the economy from any financial shock f_t. As to be shown in the next section, the welfare gains from the capital control in this environment and in this particular form would be quite large.

The solution to this problem is relatively straightforward. From FOC-(b_t), it follows that $E_t\xi_{t+k} = \xi_t,\ \forall k \geq 1$. Substituting this into FOC-$(\widetilde{y}_t^R)$ and FOC-(w_t),

$$\widetilde{y}_{t+k}^R = \delta\gamma_{t+k}^R + \frac{2W_y}{\frac{\sigma}{D} + \eta}\xi_t$$

$$w_{t+k} = \frac{D-2a+1}{2a(1-a)\phi}\delta\gamma_{t+k}^R - \frac{D}{a(1-a)\phi}\xi_t$$

Substituting this result, as well as FOC-(π_t^R), into the Phillips curve,

$$
\begin{aligned}
\beta\gamma_{t+1}^R - (1+\beta)\gamma_t^R + \gamma_{t-1}^R = &\delta\Big((\frac{\sigma}{D}+\eta) \underbrace{(\delta\gamma_t^R + \frac{2W_y}{\frac{\sigma}{D}+\eta}\xi_t)}_{\widetilde{y}_t^R} \\
&+ \frac{D-2a+1}{2D} \underbrace{(\frac{D-2a+1}{2a(1-a)\phi}\delta\gamma_t^R - \frac{D}{a(1-a)\phi}\xi_t)}_{w_t} + \mu_t^R\Big)
\end{aligned}
$$

Rearranging, a second order expectational difference equation in γ_t^R can be obtained, where $E_t\xi_{t+k} = \xi_t$ can be used. Given the two real roots of the characteristic equation, $\nu_1 < 1 < \beta^{-1} < \nu_2$, $E_t\gamma_{t+k}$ can be solved in terms of ξ_t and μ_t. Finally, the single lifetime budget constraint can be used to pin down the value of ξ_t as a function of exogenous shocks.

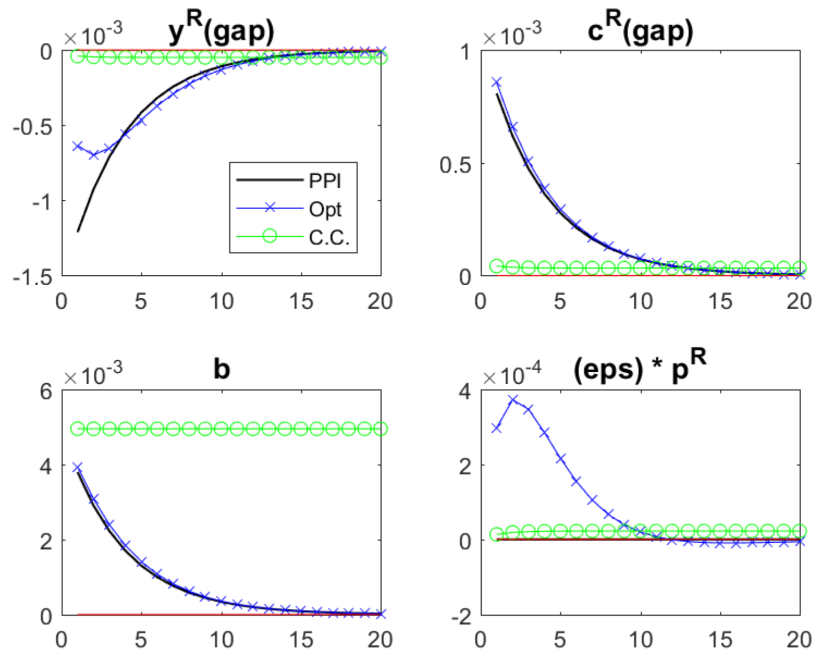
The figure shows the impulse response to a shock on relative productivity under optimal capital control, which is marked with green ‘o’. It can be seen that the expected path of the variables are nearly flat, which is what would happen qualitatively under perfect capital mobility.

1.5 Welfare Comparison

In this section, welfare across different policy regimes are compared while facing each different type of shocks. There are two typical measures of welfare, both of which are widely used in welfare analyses: conditional welfare and unconditional welfare.[30] Each of these measures has a distinct interpretation, and each one is better suited for answering particular questions than the other. The conditional welfare would measure the welfare difference while holding constant the state variables as exogenously given. From the

[30]See for example Schmitt-Grohe and Uribe (2007), and Lester et al. (2014).

Figure 5: Impulse responses to a 1% shock to relative productivity. Solid black line is the PPI-targeting, same as the flexible price allocation. Blue 'x' is the optimal monetary policy. Green 'o' is the optimal capital control.



optimal policy perspective, the planner's maximization problem solves an optimization conditional on the state realization at time 0, so the conditional welfare may be better suited for assessing the performance of a particular policy. The unconditional welfare further includes the welfare effect of state variables evolving stochastically over time, making it more suitable for measuring the welfare differences in two different economies in the long run.

One property of the conditional welfare is that it discounts the future utility by $\beta < 1$. The optimal policy maximizes the conditional welfare, so by construction higher weights are imposed on the near future compared to the far future. In contrast, a policy that maximizes unconditional welfare would minimize losses in all periods equally. Hence it is not uncommon that the welfare rankings are reversed based on which metric is used. For example, it has been noted that optimal policy under timeless perspective is not optimal in terms of welfare in the timeless sense (i.e., unconditional expectation) because of this discrepancy.[31] In fact, the difference between conditional and unconditional welfare is may be particularly pronounced under the current context — if χ is close to 0 so that the dynamic system has eigenvalues with magnitudes close to one. This would result in arbitrarily large unconditional variance of the state variables. This in turn implies that the unconditional welfare would be arbitrarily low when $\chi \to 0$.[32]

In the current model, it turns out that the consumption equivalent welfare loss is closely related to the loss function as previously derived:[33]

$$\lambda \approx \frac{1-\beta}{2} E_0 \sum_{t=0}^{\infty} \beta^t \Psi_t$$

$$\lambda^u \approx \frac{1}{2} E\Psi_t$$

where λ is the conditional welfare difference in consumption equivalent units with respect to the first best allocation, and λ^u is the unconditional welfare difference. Conditional welfare is evaluated as the welfare conditional on the state variables being equal to their steady state values. Unconditional welfare can be decomposed as the sum of conditional welfare and unconditional variance of the state variables, and hence always weakly greater than conditional welfare. Because the unconditional welfare losses would

[31] See for example Blake (2001), and Jensen and McCallum (2002).

[32] The intuition is simple. After sufficiently long time elapses, an economy with $\chi = 0$ will have arbitrarily large value of $|\widetilde{c}_t^R|$ with arbitrarily high probability, even if it started with $\widetilde{c}_0^R = 0$. Large $|\widetilde{c}_t^R|$ means one country would consume much less than the other country, because it would be indebted to the other country so much after accumulating all the negative shocks in the past in an optimal manner. But in this environment, first order approximation around the steady state would not make sense, where the steady state does not even exist. Although the variance would be finite with $\chi > 0$, $\chi \to 0$ would still inherit these properties qualitatively, including very low unconditional welfare.

[33] The derivation of the conditional and unconditional expectations of the loss function, as well as the calculation of the consumption equivalent welfare losses, are explained in detail in the appendix. The coefficient $\frac{1}{2}$ simply reflects the particular normalization used while deriving the loss function, i.e., $2u_t^W = u_t + u_t^*$ instead of u_t^W.

shock	productivity shock			cost shock		
χ	0	0.1	10	0	0.1	10
PPI-T	0.001	0.019	0.112	0.746	0.648	0.415
Ramsey	0.001	0.017	0.091	0.144	0.131	0.118
K-Ctrl	0.001			0.106		
shock	UIP shock ($\frac{\gamma\sigma_{er}^2 n}{m}$ = const)			UIP shock (n = const)		
$\chi(=\frac{\gamma\sigma_{er}^2}{m})$	0	0.1	10	0	0.1	10
PPI-T	0.309	0.205	0.006	0	0.002	0.642
Ramsey	0.160	0.109	0.005	0	0.001	0.468
K-Ctrl	0			0		

Table 1: Consumption equivalent (conditional) welfare losses compared to the first best allocation. Numbers are in units of percentage points. PPI-T: PPI targeting. Ramsey: Ramsey optimal policy. K-ctrl: optimal capital control.

blow up as $\chi \to 0$, and because the conditional welfare is consistent with the planner's objective, I will henceforth focus on the conditional welfare measure.[34]

Table 1 show the conditional welfare losses with respect to the first best allocation in terms of consumption equivalent units, under each policy and facing each different types of shocks. The size of the shocks are 1% with respect to the steady state values. Note that the optimal capital control policy results in the same allocation regardless of the value of χ, hence only one value is reported across χ's.

A few observations are in order. First, the welfare ranking between policy regimes are as expected. Ramsey policy dominates PPI targeting that attains the same allocation as the flexible price equilibrium. Capital control achieves higher welfare than the Ramsey policy because it can effectively eliminate one constraint that constrains the planner. Facing productivity shocks, Ramsey policy does slightly better than the PPI targeting,

[34]Numerically, the unconditional welfare loss is nearly identical to the conditional welfare loss when $\chi = 10$, still quite close (typically within 2% range) when $\chi = 0.1$, but infinity when $\chi = 0$ or under optimal capital control.

whereas it significantly reduces the welfare loss when facing cost shocks.

Looking across different values of χ's, it can be immediately seen that an increase in χ corresponds to an increasing welfare loss from productivity shocks, and a decreasing welfare loss from cost shocks. χ can be interpreted as the degree of frictions in the financial market, due to the risk aversion of the financial intermediaries. The value of $\chi = 0$ implies perfect capital mobility because the intermediaries are risk neutral, and $\chi \to \infty$ corresponds to the financial autarky where the intermediaries are so risk averse that they never take any positions, leaving the financial market completely segmented.

First, consider the effect of χ on the welfare facing productivity shocks. A larger value of χ makes it difficult for households to borrow or save, because households would need to pay premium to the financial intermediaries, by the amount proportional to the size of net saving. The consumption of the country with higher productivity would be higher on impact (and thus higher volatility) and decay at a faster rate when facing higher χ. As explained above, this higher consumption results in higher marginal rate of substitution between leisure and consumption and thus higher real wage in the country with higher productivity, and disproportionately more so with higher values of χ. Ultimately this leads to an underproduction in the country with higher productivity. Higher χ would act to exacerbate this distortion, both in terms of relative consumption and relative output gap. Indeed, the table shows that the welfare loss is higher when χ is higher. The optimal policy reduces the loss, but by a modest amount.

Second, consider the relative cost shocks. Suppose the monetary policy targets PPI so that natural flexible price allocation is in place. When $\chi \to 0$, households would seek perfect consumption smoothing in expectation $\left[w_t = E_t w_{t+1} = \cdots = \overline{w}_t\right]$, while the impact of cost shock is almost entirely borne by the current output gap $\left[(\frac{\sigma}{D} + \eta)\widetilde{y}_t^R =\right.$

$-\mu_t^R - \frac{D-2a+1}{2D}\overline{w}_t \approx -\mu_t^R]$. In contrast, when $\chi \to \infty$, households cannot attain any consumption smoothing (financial autarky; $b_t \equiv 0$). This results in $w_t \equiv 2W_y\widetilde{y}_t^R$, and an exogenous increase in cost will be borne by decreases in both relative output gap and relative consumption. In other words, when χ is large, an exogenous increase in relative cost is accompanied by lower relative consumption because it is costlier to borrow. This in turn lowers the real wage, which would act to mitigate the effect of the increase in relative cost. Although the loss from consumption gap would increase, it turns out that the overall loss would decrease with higher values of χ. This mechanism is present only for the cost shock, and is related to the fact that the impulse response of $\widetilde{y}_t^R$ and w_t move in the same direction, as opposed to the impact of other shocks. While this reduction in loss of welfare as χ increases is also present under the Ramsey optimal policy, the difference is smaller because the optimal policy would significantly reduce the losses compared to the flexible price allocation, especially when χ is smaller. Also note that the gains from the optimal policy compared to the flexible price appear to be much larger for the cost shock, compared to other shocks in consideration.

Third, consider the effect of χ on the welfare facing UIP shocks. There is an issue of how to scale the variance of the shock to UIP as χ changes, because the modified UIP condition derived from the financial sector imposed an arbitrary normalization for the variance of UIP shock. First, consider the case where the variance of the UIP shock f_t is unaffected, which is equivalent to assuming that $\frac{\gamma\sigma_{er}^2 n}{m}$ is constant while varying $\chi = \frac{\gamma\sigma_{er}^2}{m}$. This corresponds to the case where the size of the financial sector varies altogether relative to the size of the real economy (Y^{ss}). A larger χ immediately implies a smaller size of the UIP shock, and thus the welfare loss from the UIP shock becomes smaller.

Next, consider the case where the variance of the UIP shock f_t varies proportionally with χ, which is equivalent to assuming that n is constant. This corresponds to the case where the risk aversion of financial intermediaries (γ) changes. The modified UIP condition now takes the form of $E_t \Delta w_{t+1} = \chi(-b_t + f_t)$. When $\chi \to 0$, any demand for Home bond by the noise traders would be immediately met by the financial intermediaries without requiring any excess return. Thus the UIP shock would not affect any of the households' allocation, and the UIP condition would hold. In other words, the risk neutral financial intermediaries would perfectly insulate the household from the UIP shocks. On the other hand, suppose that $\chi \to \infty$, which is equivalent to the financial autarky case. Then it must be that $b_t = f_t$, i.e., households should meet the exogenous demand for bond on their own, because the financial intermediaries are too risk averse to take any positions. This would result in the largest distortions in allocation, and hence the largest losses in welfare.

1.6 Conclusion

In this paper, optimal cooperative monetary policy under commitment has been studied for a canonical two-country New Keynesian economy, with various shocks including shocks to UIP. UIP shock, as well as other shocks, leads to welfare loss because it distorts relative consumption and relative output. Optimal monetary policy, while not being able to alter the relative consumption, would reduce the distortion in relative output at the expense of allowing for some dispersion in prices.

Bibliography

Blake, A. P. (2001). A ‘timeless perspective’ on optimality in forward-looking rational expectations models. *Working paper*.

Clarida, R., Gali, J., and Gertler, M. (2002). A simple framework for international monetary policy analysis. *Journal of Monetary Economics*.

Corsetti, G., Dedola, L., and Leduc, S. (2018). Exchange rate misalignment, capital flows, and optimal monetary policy trade-offs. *Working paper*.

de Paoli, B. (2009a). Monetary policy and welfare in a small open economy. *Journal of International Economics*.

de Paoli, B. (2009b). Monetary policy under alternative asset market structures: The case of a small open economy. *Journal of Money, Credit and Banking*.

Devereux, M. B. and Engel, C. (2003). Monetary policy in the open economy revisited: Price setting and exchange-rate flexibility. *Review of Economic Studies*.

Devereux, M. B., Engel, C., and Lombardo, G. (2019). Implementable rules for international monetary policy coordination. *Working paper*.

Engel, C. (2011). Currency misalignments and optimal monetary policy: A reexaminztion. *American Economic Review*.

Engel, C. (2014a). Exchange rate stabilization and welfare. *Annual Review of Economics*.

Engel, C. (2014b). Exchange rates and interest parity. *Handbook of International Economics, Volume 4*.

Faia, E. and Monacelli, T. (2008). Optimal monetary policy in a small open economy with home bias. *Journal of Money, Credit and Banking*.

Fanelli, S. and Straub, L. (2019). A theory of foreign exchange interventions. *Working paper*.

Gabaix, X. and Maggiori, M. (2015). International liquidity and exchange rate dynamics. *Quarterly Journal of Economics*.

Gali, J. and Monacelli, T. (2005). Monetary policy and exchange rate volatility in a small open economy. *Review of Economic Studies*.

Giannoni, M. P. and Woodford, M. (2017). Optimal target criteria for stabilization policy. *Journal of Economic Theory*.

Itskhoki, O. and Mukhin, D. (2017). Exchange rate disconnect in general equilibrium. *NBER Working Paper 23401*.

Jensen, C. and McCallum, B. T. (2002). The non-optiality of proposed monetary policy rules under timeless perspective commitment. *Economic Letters*.

Kollmann, R. (2004). Welfare effects of a monetary union: The role of trade openness. *Journal of the European Economic Association*.

Lester, R., Pries, M., and Sims, E. (2014). Volatility and welfare. *Journal of Economic Dynamics and Control*.

Liu, Z. and Spiegel, M. M. (2015). Optimal monetary policy and capital account restrictions in a small open economy. *IMF Economic Review*.

Prasad, N. (2018). Sterilized interventions and capital controls. *Journal of International Money and Finance*.

Schmitt-Grohe, S. and Uribe, M. (2007). Optimal simple and implementable monetary and fiscal rules. *Journal of Monetary Economics*.

Wang, J. (2010). Home bias, exchange rate disconnect, and optimal exchange rate policy. *Journal of International Money and Finance*.

Chapter 2

Offshoring and Segregation by Skill: Theory and Evidence

2.1 Introduction

The last two decades ushered in a momentous shift in the paradigm of international trade from exchange in final goods to trade in tasks (Grossman and Rossi-Hansberg, 2006). Recent advances in Information and Communication Technology (ICT) together with changes in economic institutions have facilitated the fragmentation of production processes in disparate locations across borders, which is the essence of what is generally referred to as "offshoring." In response to changes in the nature of production, firms make adjustments, re-optimizing the mix of occupations and the skill-type of workers to keep in-house.

In this paper, we study the labor market consequences of offshoring, with a particular focus on worker-firm matching and wage inequality. We use the Danish employer-employee matched data and the Danish international trade registers to examine the effects of offshoring on wages and reallocation of workers in the occupations with tasks that are highly offshorable.[1] Since examining the effects of offshoring at the within-occupation-worker level requires having a detailed measure of worker's skill, we construct a novel measure of skill using rich information on individual education and job training records. Armed with the measure of worker's skill, we empirically examine predictions derived from the model, and further establish a causal effect of offshoring using an instrumental variable in the reduced-form analysis. The predictions are based on a simple matching model between workers and firms where offshoring affects labor market outcomes by changing the effective supply of workers. Finally, we quantify the equilibrium effects of offshoring by estimating the matching model extended with unobserved heterogeneity in preferences.

A key prerequisite for an empirical investigation of matching and sorting is a measure of the characteristics by which agents are sorted. In the context of worker-firm matching, measures of worker's skill and firm's productivity are required. An important contribution of this paper is in the construction of worker's skill, using the rich information in the occupation and education contained in the Danish administrative data. More specifically, we extract the skill components in various dimensions (e.g. cognitive, manual, interpersonal) from the education and occupation records by linking the raw textual descriptions to the O*NET scores using techniques in textual analysis.[2]

To illustrate the mechanism, we build a matching model in the spirit of Becker (1973) and Sattinger (1993) where for each occupation, workers and firms with heterogeneous

[1]Offshorable occupations are generally associated with routine tasks that are easily codifiable (e.g., Autor et al., 2003; Oldenski, 2012). The work performance in these jobs, in general, does not require direct physical contact; or geographic proximity (e.g., Blinder, 2009; Blinder and Krueger, 2013; Goos et al., 2014).

[2]There is an increasing number of recent studies that conduct empirical analysis using data based on textual information processed through machine learning techniques: Atalay et al. (2018), Gentzkow et al. (2018) Hoberg and Phillips (2016), Michaels et al. (2016), Gentzkow and Shapiro (2010), etc.

attributes competitively find matches to produce occupation-specific outputs. Due to the complementarity in the production function, there is positive assortative matching between workers and firms in equilibrium, and the jointly produced output is shared as wages and profits. In a global economy, firms have the option to match with foreign workers upon paying a fixed cost of offshoring in each occupation.[3] We focus on the North-South framework of offshoring (e.g., Feenstra and Hanson, 1997; Grossman and Rossi-Hansberg, 2008), where only the North finds offshoring a less expensive alternative for production (one-way offshoring).[4]

The model yields several intuitive predictions. First, with additional supply of workers from abroad, domestic workers in offshorable occupations would experience a reduction in wages and face a greater reallocation risk. The flip side of this prediction is that firms are better off as they are able to hire better quality workers at a lower cost. Second, the variance of the worker's wage within offshorable occupations would decrease. While the model mechanism operates within each occupation, since different occupations have different degrees of offshorability, the model also generates predictions across occupations — each of these predictions would be more pronounced in occupations with higher offshorability.

Consistent with the model predictions, we confirm in the Danish data that workers with low cognitive skills are hurt relatively more in terms of reallocation risks compared to high-cognitive workers. Firms improve the average cognitive skill of their in-house workers in response to offshoring, and the extent to which firms improve their quality of

[3]The notion of offshoring is similar to Antràs et al. (2006) and Kremer and Maskin (2006) in the sense that it effectively changes the aggregate supply of workers in offshorable occupations.

[4]Burstein and Vogel (2010) and Grossman and Rossi-Hansberg (2012) explore a North-North framework where offshoring occurs between similar countries.

workers is greater for low-productivity firms relative to high-productivity firms. We also confirm at the industry level that offshoring increases occupational segregation, which is measured as the variance in the share of offshorable occupations at the firm-level in the Danish data.[5] In order to address the simultaneity concerns, we also use instrumental variables based on the China's export supply to the world excluding Denmark.

A key departure from the offshoring models using a matching framework is the unobservable preference shocks introduced in the matching problem where we follow the marriage market literature (Choo and Siow, 2006; Dupuy and Galichon, 2014) in the assumptions. The main purpose of this extension is to estimate the model and perform counterfactual exercises, which allows to assess the quantitative impact of offshoring relative to other competing concurrent channels, such as technological change (e.g., Acemoglu and Autor, 2011; Lindenlaub, 2017) and the expansion of higher education (Kremer and Maskin, 1996), on worker-firm matching and between-firm wage inequality.[6] Using the joint distribution of workers and firms in the Danish data, we estimate the matching model by a moment-matching procedure. The main challenge in the identification lies in the number of offshored matches, which is essentially not observed in the data. To recover the number of offshored matches that are unobserved in the data, we assume that the value-added per domestic worker is the same as the value of offshoring per foreign worker composites. The counterfactual experiments show that technology mainly drives firms to become more different in terms of worker quality and average

[5]Changes in between-firm inequality requires jointly examining both the within-occupation and between-occupation channels. As the model cannot incorporate any interactions across different occupation types due to properties of matching models, we consider this to be beyond the scope of this paper.

[6]Several papers aim to disentangle the impact of technological change and globalization on labor market outcomes (Autor et al., 2015; Bahar Baziki et al., 2015; Hakanson et al., 2015).

wages, while offshoring offsets these differences between firms.

This study is related to a small yet growing trade literature that uses matching models[7] to study the distributional effects of globalization: heterogeneous effects of international trade within sector, firm, occupation etc. (e.g., Kremer and Maskin, 2006; Costinot and Vogel, 2010; Grossman et al., 2017). However, these matching models are seldom estimated, particularly in the context of offshoring. In this paper, by introducing unobserved preference shocks à la Dupuy and Galichon (2014), we are able to bring the matching framework to data. To the best of our knowledge, this is the first paper to estimate a worker-firm matching model with offshoring.

Next, this study contributes to the trade literature examining the labor market effects of offshoring. Previous studies (e.g., Feenstra and Hanson, 1997; Hsieh and Woo, 2005; Biscourp and Kramarz, 2007) have focused on changes in wage and employment outcomes in response to offshoring, comparing across broad categories: occupations, education groups etc. More recently, the focus has shifted to further examine the impact of offshoring at a more disaggregate level using administrative data on firms and workers (e.g., Baumgarten et al., 2013; Becker et al., 2013; Hummels et al., 2014). The novelty of our findings is based on the high-quality Danish data together with a full characterization of worker skill, which enable us to examine the distributional effects of offshoring on workers within offshorable occupations and further study changes in the skill composition at the firm-level.

Finally, this project also contributes to the burgeoning literature on worker sorting or segregation of workers by skill. Previous studies have documented evidence of

[7] Grossman and Maggi (2000) employ a matching model to study the type of production technology determining the pattern of specialization.

growing segregation by skill in recent decades notably in developed countries.[8] The potential mechanisms proposed in the literature include: technological change (e.g., Autor et al., 2003; Acemoglu and Autor, 2011), outsourcing (e.g., Abraham and Taylor, 1996; Goldschmidt and Schmieder, 2017), international trade (e.g., Helpman et al., 2010; Davidson et al., 2014)[9], and rising skill dispersion (e.g., Kremer and Maskin, 1996; Acemoglu, 1999). We propose offshoring as an important channel that affects between-firm inequality through the occupation composition as well as the worker mix within occupations.

The remainder of the paper proceeds as follows. Section 2 introduces the worker-firm matching model with a fixed cost of offshoring. Section 3 provides data descriptions with details on the skill construction and other measures. Section 4 presents the estimation strategy and results of the reduced-form analysis. Section 5 presents the structural estimation of the matching model and the results of the counterfactual experiments. The last section concludes.

[8]e.g., United States (Song et al., 2019), United Kingdoms (Faggio et al., 2007), Germany (Card et al., 2013), France (Abowd et al., 1999), Sweden (Bahar Baziki et al., 2015; Hakanson et al., 2015), Denmark (Bagger et al., 2013; Bagger and Lentz, 2018)

[9]Davidson et al. (2014) empirically examines the idea that globalization improves matching for high-productivity firms in the exporting sector. This is in line with Helpman et al. (2010) that show how worker-firm matching is affected by exporting firms' intensity in screening their workers to gain competitiveness.

2.2 Model

2.2.1 Baseline Economy

We build a Becker-type matching model (Becker, 1973) where for each occupation, workers and firms with heterogeneous attributes competitively find matches to produce occupation-specific outputs. In a global economy, firms have the option to form international teams upon paying a fixed cost of offshoring in each occupation. The notion of offshoring is similar to Antràs et al. (2006) and Kremer and Maskin (2006) where offshoring effectively changes the aggregate supply of workers in offshorable occupations. For estimation purposes, we introduce random preference shocks in the matching problem following Choo and Siow (2006) and Dupuy and Galichon (2014).

Economic Environment There are two sectors (manufacturing and traditional) and multiple occupations in the economy. The manufacturing sector is endowed with a continuum of heterogeneous firms with productivity y, which is a realization of $Y \subseteq \mathbb{R}_+$ with p.d.f. of $\overline{g}(y)$. In each occupational category ($o \in O$), there exists a continuum of inelastically supplied[10] heterogeneous workers characterized by their skills x that contribute to the production process: a realization of $X \subseteq \mathbb{R}_+$, denoted by x with p.d.f of $\overline{f}(x)$. Workers can either participate in the manufacturing sector where they match with a firm to produce a task output and earn wages; or sort into the traditional sector where they are offered a constant wage $\underline{w}$ regardless of their skills. Firms may also choose not to operate in the manufacturing sector, which allows them a constant outside option

[10]While interesting analysis arises with matching in relation to workers' choice of occupation, we consider this to be beyond the scope of this paper, as in many previous studies including Grossman et al. (2017).

of zero. The distributions of those who take the outside option are denoted as $f_0(x)$ and $g_0(y)$; and those of workers and firms in the manufacturing sector, $f(x)$ and $g(y)$ respectively. By construction, $f_0(x) + f(x) = \overline{f}(x)$, $g_0(y) + g(y) = \overline{g}(y)$.

Production Technology Production in the traditional sector requires workers only; however, in the manufacturing sector, it requires output from each occupation which is generated through matching between a firm and a worker.

$$q(x, y) = xy \tag{2.1}$$

Occupation-specific outputs are required to produce a final good and there is no complementarity between different occupations in production.[11] The functional form of the task output is a simplified version of a bilinear production technology $\mathbf{x}'\Gamma\mathbf{y}$ where $\mathbf{x} = [x_1, x_2, ...x_n]'$ and $\mathbf{y} = [y_1, y_2, ..., y_m]'$ provide characteristics of workers and firms respectively combined through a production technology Γ, an n-by-m matrix that captures the complementarity between workers and firms across different characteristics. In this section, we use the one-dimensional matching model for simplicity; however, when we structurally estimate the model in Section 2.5, we use the fully developed multidimensional matching model with unobservable preferences.[12]

Unobserved Preferences In order to allow deviations from pure positive assortative

[11]Such abstraction is necessary to ensure "existence and tractability" of matching models (Eeckhout and Kircher, 2018).

[12]Note that it is also possible to include worker or firm-specific effects, capturing the extent to which workers or firms contribute to the output independent of the matches. These components can even take nonlinear functions (Dupuy and Galichon, 2014).

matching[13] that is rarely observed in the real world, unobserved components are introduced closely following Dupuy and Galichon (2014).[14] A worker with skill x maximizes his or her utility, which includes wages and unobserved preferences.

$$\max\ [\underline{w}+\varepsilon_1^o,\ \{\max_y\ w(x,y)+\varepsilon_1(y)\}] \tag{2.2}$$

$\varepsilon_1(y)$ is the unobserved, idiosyncratic preference of the worker for each firm of productivity y; and ε_1^o is the utility the worker gets by sorting into the traditional sector. Analogously, a firm with productivity y maximizes its surplus, which includes profits and unobserved preferences.

$$\max\ [\varepsilon_2^o,\ \{\max_x\ r(x,y)+\varepsilon_2(x)\}] \tag{2.3}$$

$\varepsilon_2(x)$ is the unobserved, idiosyncratic preference of the firm for each worker of skill x, and ε_2^o is the utility the firm receives by exiting the manufacturing sector. Random preference shocks $\varepsilon_1(y)$, ε_1^o and $\varepsilon_2(x)$, ε_2^o are assumed to follow an extreme value stochastic process with scale parameters λ_x, λ_y capturing the extent to which unobserved heterogeneity plays a role.[15] Each worker with observed skill x has a set composed of random realization of "acquaintances," which follows a Poisson point process on $Y \times R$ of intensity $\exp(-\varepsilon_1)d\varepsilon_1 dy$. As a consequence of the Poisson point process assumption, each individual has an infinite but countable number of acquaintances. Note that a competitive equilibrium requires $w(x,y)+r(x,y)=q(x,y)$.

[13]Pure matching denotes the case where each x is matched with a unique y, and vice versa. In this case, for each given x or y, there is only one value of y or x for which $\pi(x,y)$ is nonzero, and there exists a one-to-one matching function $\mu : X \to Y$ and $\mu^{-1} : Y \to X$.

[14]Their study is a continuous generalization of Choo and Siow (2006) which introduced standard multinomial logit over discrete types into the matching problem.

[15]In order to maintain the analytical tractability of the standard logit problem with discrete choice, the same scale parameter σ_x is used for both the outside option and for each of the matching options y. See Appendix C for further details on the random shock process.

Equilibrium Matching and Wages Using properties of generalized extreme value distributions, the equilibrium matching between workers and firms as well as equilibrium wages are given as follows,[16]

$$\pi(x,y) = \hat{a}(x)\hat{b}(y)\exp(\frac{q(x,y)}{\lambda}) \tag{2.4}$$

$$w(x,y) = \frac{\lambda_x(q(x,y) - b(y)) + \lambda_y a(x)}{\lambda} \tag{2.5}$$

where $\lambda_x + \lambda_y = \lambda$; $\hat{a}(x) = \exp(-\frac{a(x)}{\lambda})$; and $\hat{b}(y) = \exp(-\frac{b(y)}{\lambda})$.[17] Note that the wage depends not only on x but also on y due to the unobserved heterogeneity components.[18] Greater values of λ generate a matching that is closer to a random match whereas small λ implies a matching that primarily relies on observed characteristics. Also, $a(x)$ and $b(y)$ correspond to Lagrange multipliers on the scarcity constraint $f(x) = \int \pi(x,y)dy$ and $g(y) = \int \pi(x,y)dx$. Therefore, higher values of $a(x)$ indicate scarcity in workers with observed characteristics x which results in greater extraction of the produced task output while a large $b(y)$ would benefit firms' profits. See Appendix C for the details on the characterization of the exogenously given marginal distributions and also on how we solve the model equilibrium.[19]

[16] See Appendix B for derivation of equilibrium matching and wages.

[17] $a(x) = \lambda_x \log \frac{\int \exp(\frac{w(x,y)}{\lambda_x})dy}{f(x)}$; $b(y) = \lambda_y \log \frac{\int \exp(\frac{r(x,y)}{\lambda_y})dx}{g(y)}$;

[18] As this unobserved heterogeneity converges to zero, $w(x,y) \to w(x)$ and $r(x,y) \to r(y)$.

[19] There is a straightforward iterative algorithm proposed in Bojilov and Galichon (2016); referred as the "Iterated Proportional Fitting Procedure (IPFP)" or "Sinkhorn's algorithm." to recover all other endogenous objects, including the matching function $\pi(x,y)$ as well as wages $w(x,y)$ and profits $r(x,y)$.

2.2.2 Global Economy Equilibrium

With globalization, domestic firms have the option to match with foreign workers upon paying an occupation-specific fixed cost of offshoring. Similar to Antràs et al. (2006), we refer to improvements in the information communication technology (ICT), or economic reforms in China or Eastern European countries followed by increased participation in global economic activities as forces of *globalization* that reduce the cost of offshoring.

Global Economic Environment Foreign is endowed with workers with observed skill x_F a realization of X_F with p.d.f of $\overline{h}(x_F)$ in each occupation o. For simplicity, only the traditional sector exists in Foreign, which is populated with self-employed workers that earn a constant income of $\underline{w}_F$. The marginal distributions of agents in a global economy are denoted as follows: $\overline{f}(x) = f_0(x) + f(x)$, $\overline{h}(x_F) = h_0(x_F) + h(x_F)$, and $\overline{g}(y) = g_0(y) + g(y) + g_F(y)$ where those with a subscript zero denote agents who are not in the manufacturing sector.[20]

Production with Offshoring The output when matched with a Foreign worker, $q_F(x_F, y)$, is as follows:

$$q_F(x_F, y) = q(x_F, y) - C = x_F y - C \tag{2.6}$$

Here, we assume that the firm's productivity level y does not change with respect to the location of operation while the model can incorporate a more general form of technology.[21] A simple way would be to allow for the coefficient on $x_F y$ to be different from

[20] In addition, $\int f(x)dx = \int g(y)dy = \int\int \pi(x,y)dxdy$, and $\int h(x_F)dx_F = \int g_F(y)dy = \int\int \pi_F(x_F,y)dx_F dy$, where $\pi(x,y)$ and $\pi_F(x_F,y)$ are the density functions that describe the realized pattern of matching with Home and Foreign workers, respectively.

[21] Ramondo and Rodríguez-Clare (2013) examine multinational firm activities and how the firm's

1. We maintain this simple form to understand the mechanism in this section; however, when we do a structural estimation in Section 2.5, we estimate the value of coefficients using data.

In the global economy, firms have the option to match with foreign workers upon paying a fixed cost of offshoring in each occupation where the cost is associated with managing production processes of each intermediate good overseas that often involves a significant level of organizational complexity as it limits the opportunities for monitoring and coordinating workers (Grossman and Rossi-Hansberg, 2008). The per-match aspect of the cost reflects the model mechanism where offshoring firms seek to match with the best possible foreign workers[22] in the global economy (Alchian and Allen, 1983).[23]

Each firm's decision to offshore depends on the benefit of matching with foreign workers and the cost associated with hiring them. Conditional on the cost of offshoring, which decreases with globalization, firms would only find offshoring profitable when foreign workers demonstrate competitive skill-levels to their domestic workers. However, it is difficult to define a skill measure that is comparable across countries nor is it available in the data. So instead, we focus on the notion of worker composites from Foreign that can be comparable to one Danish worker that the firm hires.

Firms solve the profit maximization problem by optimally choosing the best possible worker from each country and comparing profits.

$$\max\ [\varepsilon_2^o,\ \max_x\{r(x,y)+\varepsilon_2(x)\},\ \max_{x_F}\{r_F(x_F,y)+\varepsilon_{F2}(x_F)\}] \tag{2.7}$$

technology potentially differs depending on the location of operation .

[22]Note that the foreign worker endowment should be interpreted as worker composites whose skills are comparable to domestic ones in a one-to-one manner. Therefore, the notion of "quality" of foreign workers is in efficiency units of labor, which consistently applies to firms hiring foreign labor in greater quantities taking advantage of the low cost.

[23]The Alchian-Allen effect demonstrates how in the presence of a per unit cost consumption shifts towards high quality goods, and in the context of international trade,"shipping the good apples out."

$\varepsilon_2(x)$, $\varepsilon_{F2}(x_F)$, and ε_2^o are the unobserved, idiosyncratic random shocks of the firm for a Home worker x, a Foreign worker x_F, and exiting the sector, respectively. Analogously, each worker from Home and Foreign maximizes his or her utility, which includes wages and preferences as follows.

$$\begin{aligned} \max\ [\underline{w} + \varepsilon_1^o,\ \ \max_y \{w(x,y) + \varepsilon_1(y)\}] \\ \max\ [\underline{w}_F + \varepsilon_{F1}^o,\ \ \max_y \{w_F(x_F,y) + \varepsilon_{F1}(y)\}] \end{aligned} \tag{2.8}$$

$\varepsilon_1(y)$ and ε_1^o ($\varepsilon_{F1}(y)$ and ε_{F1}^o) denote the unobserved, idiosyncratic random shocks of the Home (Foreign) worker for firm y and sorting into the traditional sector respectively. Note that $w(x,y)$, $w_F(x_F,y)$, $r(x,y)$, $r_F(x_F,y)$ are endogenous objects to be determined in equilibrium. Again, a competitive equilibrium requires $q(x,y) = w(x,y) + r(x,y)$ and $q_F(x_F,y) = w_F(x_F,y) + r_F(x_F,y)$.

While features of globalization lower costs related to transportation and communication or even institutional factors such as tariffs, the extent to which the cost C decreases is occupation-specific, which depends on the nature of the task. That is, a decline in C would be trivial for occupations that perform nonroutine tasks that require direct physical contact and geographic proximity, i.e. *non-offshorable* occupations. Even for occupations that demonstrate high offshorability, there exists a cost component that remains high for firms to operationalize offshoring: the inherent cost associated with managing production processes of each intermediate goods overseas, which involves a significant level of organizational complexity. In fact, it is often observed to be mainly concentrated in firms that are more productive, larger, older, and capital-intensive (Hummels et al., 2014; Monarch et al., 2017). The implied cost may be even higher if countries where offshoring is performed do not have the institutions that effectively enforce intellectual property rights (IPR) on firm-specific innovations embodied in the production process. To reflect these empirical regularities, we further assume the following.

Assumption 1 *The cost of offshoring is greater than the traditional sector's wage gap:* $C > \underline{w} - \underline{w}_F$.

It is worth mentioning that, it may not be sensible to make one-to-one comparisons of workers' talent across countries, especially in a North-South framework that we intend to bring to data, in the context of offshoring. For example, the fact that a Danish firm chooses to hire workers from low wage countries through offshoring should not have the interpretation that workers from low wage countries are more skilled than Danish ones. Thus, in order to model offshoring in a way that indicates the possibility of substituting home workers, we characterize the foreign worker endowment as *worker composites* whose skills are comparable to the domestic ones at a fixed ratio that we exogenously impose.[24] The value of the ratio does not affect the analysis as the final quality of skill provided through a match is what counts, whether it is a single worker or a bundle of workers.[25]

Global Economy Equilibrium Matching and Wages Again, using properties of generalized extreme value distributions, the equilibrium matching between workers and firms as well as equilibrium wages and profits are given as follows,[26]

[24] In a fully developed model, the ratio would depend on the wage differences between two different locations.

[25] The same framework can be applied to examining changes in worker-firm matching when firms gain opportunities to adopt automation technology. See Appendix C for further analysis on the worker-firm matching problem when, instead of foreign workers, there is a machine that replaces a subset of workers with certain level of skills.

[26] See Appendix C for the derivation of equilibrium matching and wages.

$$\pi(x,y) = \hat{a}(x)\hat{b}(y)\exp(\frac{q(x,y)}{\lambda}) \quad \text{and} \quad \pi_F(x_F,y) = \hat{c}(x_F)\hat{b}_F(y)\exp(\frac{q_F(x_F,y)}{\lambda_F}) \tag{2.9}$$

$$w(x,y) = \frac{\lambda_x(q(x,y) - b(y)) + \lambda_y a(x)}{\lambda} \quad \text{and} \quad w_F(x_F,y) = \frac{\lambda_{x_F}(q(x_F,y) - b_F(y)) + \lambda_y c(x_F)}{\lambda_F} \tag{2.10}$$

where $\lambda \equiv \lambda_x + \lambda_y$, $\lambda_F \equiv \lambda_{x_F} + \lambda_y$, $\hat{a}(x) \equiv \exp(-\frac{a(x)}{\lambda})$, $\hat{b}(y) \equiv \exp(-\frac{b(y)}{\lambda})$, $\hat{c}(x_F) \equiv \exp(-\frac{c(x_F)}{\lambda_F})$, $\hat{b}_F(y) \equiv \exp(-\frac{b_F(y)}{\lambda_F})$. We show in Appendix C that $\hat{b}_F(y) = \hat{b}(y)^{\frac{\lambda}{\lambda_F}}$ must hold, which allows for a simple characterization of the equilibrium under offshoring.[27]

Numerical Exercise Here, we examine the model implications derived using an example setting the parameter values as $\lambda = 1$ and $\sigma = 1$. For simplicity, we additionally assume uniform distributions $X \sim U[0,1]$, $X_F \sim U[0,1]$, and $Y \sim U[0,1]$; and further impose $\underline{w} = \underline{w}_F$.

First, due to the fixed cost associated with offshoring, firms with higher values of y face a greater probability of matching with foreign workers (Figure 6). Note that in the special case of the model where $\lambda = 0$ and the upper bound of the Foreign endowment is greater compared to that of Home, the model predictions regarding "who offshores" are consistent with Helpman et al. (2004): high-productivity firms strictly prefers to offshore. In particular, if the wage differences between home and foreign are large, allowing domestic firms to hire foreign workers in greater quantities, it is possible that the skill output of these foreign worker composites is high enough that there are no home workers to compete with the corresponding level of skill output.

Next, firms improve upon the quality of their domestic worker match while workers

[27] See Appendix C for full derivation of the equilibrium.

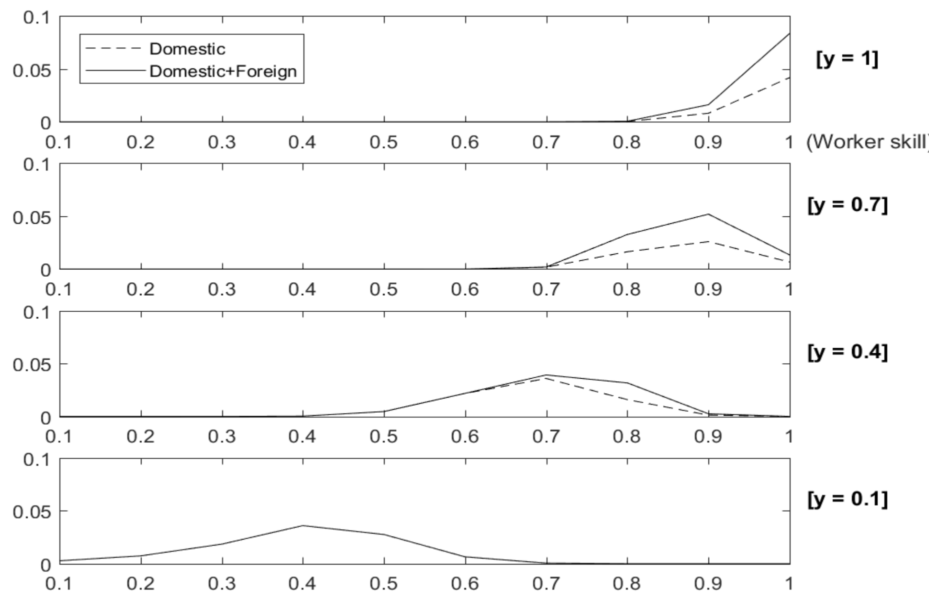


Figure 6: Equilibrium matching of workers for each firm-type

Each panel shows the probability mass of workers for each firm-type y under closed economy and global economy. The share of foreign workers is captured in the area between the dashed and the solid lines.

undergo a downward transition in the match quality (Figure 7). Due to the formation of international teams in offshorable occupations, the demand for domestic workers within these jobs decreases, which consequently drives out the least productive workers at the bottom end of the worker distribution to the traditional sector. Thus, within occupations that are highly exposed to offshoring, the less skilled workers face a greater risk of reallocation with globalization.

Finally, as workers in offshorable occupations become less expensive with a decline in the cost C, the overall wage-level falls for these workers domestically. Note that when the cost of offshoring becomes negligible ($C \to 0$), which indicates a convergence to a perfectly integrated world economy, the wage profile does not differ between workers

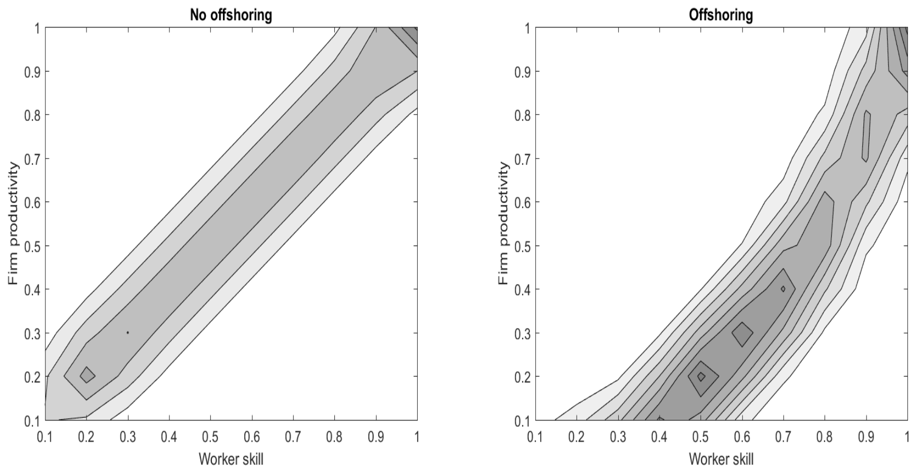


Figure 7: Equilibrium matching between firms and domestic workers

from home and foreign within these jobs.[28] This is the labor supply effect identified in Grossman and Rossi-Hansberg (2008) where "factor prices respond to factor supplies." Further, comparing across occupations that differ in their offshorability,[29] offshoring facilitated by features of globalization magnifies inequality in wages, which resonates with the wage inequality results in Feenstra and Hanson (1996), Zhu and Trefler (2005), and Costinot and Vogel (2010).[30]

[28]See Appendix C for analysis with the economy setting $\lambda = 0$.

[29]If firms increase their demand for non-offshorable occupations as a result of an expansion in size, this would increase the wage for non-offshorable occupations, which would amplify the wage inequality between offshorable and non-offshorable occupations.

[30]Firms, on the other hand, gain from greater exposure to offshoring as they are able to not only hire better quality workers at a lower cost but also increase profits. See Appendix A for changes in profits in the simulation exercise.

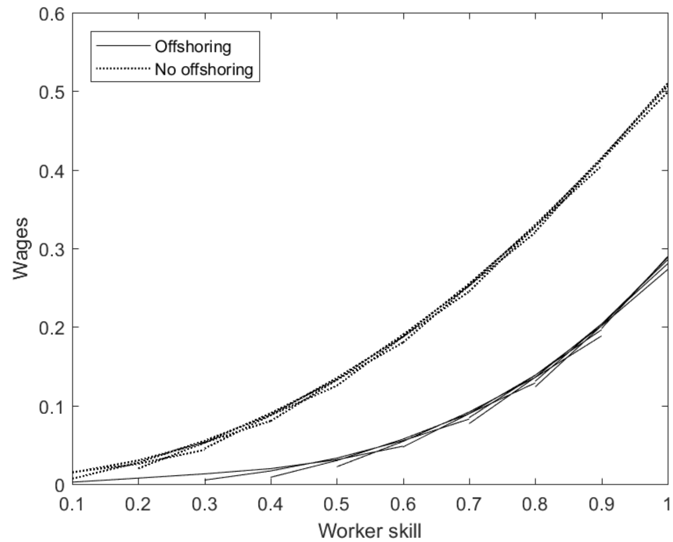


Figure 8: Equilibrium wage profile by worker's skill

2.2.3 Between-Firm Inequality in a Global Economy

So far, we have examined how offshoring by high-productivity firms, facilitated by globalization, increases competition from foreign workers in offshorable occupations domestically. As a result, within offshorable jobs: (i) domestic workers undergo a wage loss, (ii) switch down their firm matches, and (iii) the least skilled workers reallocate to the traditional sector. What implications does the model provide in terms of between-firm inequality? We discuss the extent to which firms diverge or converge in their occupational composition, their worker composition within occupations, and the average wages they pay.

Occupational Segregation With globalization, firms become increasingly different in their occupation composition. That is, occupational segregation across firms increases with offshoring. As noted earlier, channels of globalization only affect the cost of offshoring by lowering the transactional component (Fort, 2017; Benfratello et al., 2015). As a result, high productivity firms that are both technologically and managerially better equipped to produce outside the boundaries of the firm can engage in offshoring activities and replace their in-house workers while those that are not keep their offshorable occupations. In a sense, offshoring technology functions as one of the key mechanisms that drives firms to differentiate themselves in their demand for occupations, causing higher degrees of occupational segregation across firms.[31]

The model also generates predictions on the between-occupation channel (Figure 9). High-productivity firms replace their in-house workers in offshorable occupations whereas low-productivity firms that cannot afford offshoring have no choice but to keep all occupation types within their firm boundaries. Thus, with offshoring possibilities, high-productivity firms become more homogeneous in their occupation mix by replacing the offshorable occupations whereas low-productivity firms keep both offshorable and non-offshorable occupations in-house.

Within-Occupation Segregation by Skill With greater exposure to offshoring, firms are able to hire better domestic workers within offshorable jobs in terms of skill levels than before; and the type of workers they hire become more similar across the

[31] As examined in previous studies such as Goldschmidt and Schmieder (2017) and Handwerker (2015), domestic outsourcing is one of the other important channels through which occupational segregation increases in the economy.[32] However, there are important distinctions between the two in terms of *how* the distribution of occupations across firms is affected. In comparison to outsourcing, offshoring involves a higher cost due to monitoring and managing production overseas in addition coordinating differences in institutions that affect economic activities, etc. As a result, a firm's decision to participate in offshoring hinges on the firm's productivity, which subsequently affects its occupational demand differentially even among firms that have the same core competency. Domestic outsourcing, however, mainly shapes occupational segregation in a way that results in an economy with firms specialized in what they identify as the core of their production.

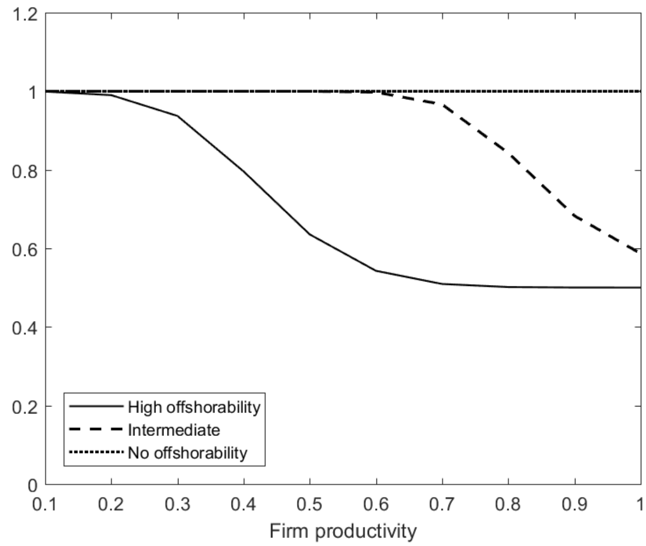


Figure 9: Occupation composition across firms under global economy

We show, for each firm-type y, the probability of keeping different occupation categories in-house, which vary in offshorability.

distribution of firms. In other words, within offshorable jobs, there is a decrease in segregation by skill, in addition to skill upgrading. As shown in the results earlier, domestic workers within offshorable jobs that are exposed to competition from foreign workers switch down their firm matches, and the least skilled ones reallocate to the traditional sector. As a result, firms face a pool of domestic workers that are better in their overall quality and that demonstrate increased homogeneity. The implications are reminiscent of Melitz (2003) where increased forces of competition driving out the least productive firms.

Between-Firm Wage Inequality For domestic workers with occupations that are vulnerable to foreign competition under the global economy, the average wage firms pay

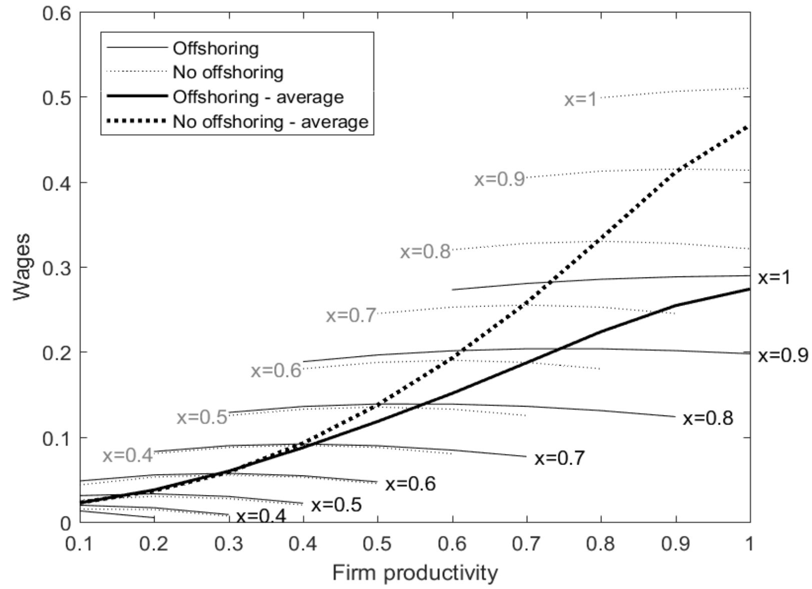


Figure 10: Equilibrium wages by firm's productivity

also becomes similar across. As mentioned earlier, workers within offshorable jobs undergo a wage loss, with high-skill workers that face direct foreign worker competition losing more. That is, the wage dispersion within offshorable jobs decreases. Note that the overall between-firm wage inequality combining across occupations should depend on the magnitude of within-occupation versus between-occupation channel. While the within-occupation channel operates in the direction of making firms become more similar in their average wages to workers in offshorable occupations, the between-occupation channel potentially amplifies the differences across firms. That is, as high-productivity firms trim down their offshorable occupations in-house, the average wage for their domestic workers increasingly depend on workers in non-offshorable jobs. Low-productivity firms, on the other hand, keep both the non-offshorable jobs and offshorable jobs in-house, and therefore, the average wage they pay reflects the overall wage loss workers with offshorable occupations have undergone with globalization taking place.[33] In the following section, we use the Danish matched employer-employee data to empirically test the predictions of the model.

Offshoring by high-productivity firms, facilitated by globalization, generates distributional effects in the labor market:

1. Occupational segregation across firms increases.

2. Within offshorable occupations, firms improve upon their worker matches, thereby decreasing between-firm inequality in average worker quality and average wages.

2.3 Data

Denmark is a small open economy, which is part of the European Union (EU), and has limited power in affecting trade policies according to its domestic economic environment. Thus, changes in the economic environment faced by Danish firms due to China's entry to WTO and subsequent changes in the quota policies on Chinese products or the enlargement of the EU to include Eastern European countries are exogenous variations that affect Danish firms' incentives to offshore their production. The labor market impact of these shocks facilitated by globalization is more prominent when the labor market demonstrates flexibility compared to a centralized market where collective bargaining prevails (Hummels et al., 2014).

[33]Although the model does not explicitly feature across-occupation interactions that would allow us to quantitatively evaluate the magnitudes of each channel, it is still useful to think about how the two channels qualitatively operate in different directions in terms of between-firm wage inequality.

Since the major labor market decentralization in 1989, Denmark has been shifting away from centralized collective bargaining to a decentralized system: between the years 1995 and 2004, firm-level wage bargaining grew from a coverage of 11% to 22%. The decentralization process was initiated by the firm side to negotiate wage contracts at the worker-firm level as they found the standard-rate system[34] not flexible enough to incorporate changes led by forces of globalization or technological change (Dahl et al., 2013). In fact, the Danish labor market currently exhibits great flexibility with average tenure comparable to Anglo-Saxon countries; and these high turnover rates are accompanied by a well-designed social security system which provides generous unemployment benefits yet incentivizes the unemployed to search for jobs actively.[35] Hence, Denmark is a good candidate country to examine labor market responses to changes in the global economic environment in the past two decades.

2.3.1 Data Source and Baseline Sample

We use the Danish register-based Matched Employer-Employee panel (1995-2011), which provides the universe of private firms and the population of individuals matched through their unique identifiers. The database includes variables on standard individual socioeconomic characteristics and detailed firm characteristics. Data on international trade comes from UHDI that records from Denmark's customs in addition to firm-level reports to Statistics Denmark regarding any trade activities (1993-2013) (Keller and Utar,

[34]The wage bargaining occurred at the industry level.

[35]It is characterized by a "flexicurity model," which comprises three components: (i) considerable flexibility for firms right to hire and fire employees; (ii) an extensive social safety net in case of unemployment; and (iii) active labour market policies where the entitlement to compensation in the event of unemployment is countered by the obligation to actively seek a job and to participate in job-related activities (Kristofferson, 2016).

2016). It contains firm-level international transactions of goods (weight and value) observed at a triplet of year-country-product (8-digit product classifications according to the Combined Nomenclature (CN) system). We further utilize the U.S. Department of Labor Occupational Characteristics Database (O*NET), the successor of the Dictionary of Occupation Titles (DOT), to obtain factor descriptions of occupation-specific skill and task requirements in the skill construction. More specifically, O*NET provides information on key features of occupation-specific requirement for knowledge, skills, and abilities in standardized measures on almost 1,000 occupations covering the entire U.S. economy. The data is collected by surveying job incumbents or occupation experts, and updated frequently to keep up with changes in the occupation structure over time.

The empirical analysis uses the Matched Employer-Employee panel as the baseline panel focusing on the period 1995-2004, with the unit of observation as an individual each year. We trim the data in the following way: we drop observations with missing identification codes for individual, firm or industry. We also disregard individuals with age below 20 and above 65 and those with missing occupation codes or military-related occupations. We focus on the manufacturing sector only, leaving out retail, service and public sectors. we follow (Bagger et al., 2013) and further trim the top and bottom 1% of each education and experience subgroup. Merging the trade register, with observations provided at the product-level transaction for each firm-year, to the baseline sample, we work with 5,305,975 observations with the unit of observation as an individual each year that include 28,276 firms through 1995-2004. Wages are CPI-adjusted to the level of 1995.

2.3.2 Construction of Measures

Skill Supply We construct a vector of skills (cognitive, manual and interpersonal skills) for each individual using (i) the highest obtained education (*hfaudd*), and (ii) the highest completed professional training (*erhaudd*)[36] together with one's occupation tracked throughout 1995-2004.[37] There are 2449 different types of education and job training records in the Danish data described in detailed textual information (e.g. B.A. in Engineering, Jewelry Designing, etc.). The strength of this measure lies in the great heterogeneity of worker skill, reflecting rich information on both education and occupation, which allows us to examine the quality of worker skills at the firm-level within different occupation categories. It is particularly useful investigating the skill quality of workers between firms since, unlike wages, the measure is independent of the firm component.

In order to construct quantifiable measures of skill, we proceed in the following three steps. First, we create a mapping between the education records and the most relevant occupation, assuming that an individual's educational attainment or job training reflects his or her ability to perform tasks required in a particular occupation. The mapping is generated employing the Continuous Bag of Words (CBOW) Model, which examines the similarity of the context in which education records and occupations appear in Wikipedia.[38] We further do a robustness check of the data construction using the *O*NET*

[36]Vocational education in Denmark tends to be between 2.5 to 5 years long which includes periods of formal schooling and apprenticeships (Keller and Utar, 2016). For example, vocational training for a blacksmith involves 2.5 to 5 years of education depending on the specialization choice. And the baseline training period is followed by formal schooling which includes blocks of internships providing opportunities to practice in actual workplaces (`https://www.ug.dk`).

[37]The idea is similar to Lindenlaub (2017) where education records are considered as a reflection of individual capabilities to perform tasks in particular jobs.

[38]The algorithm employed in the skill construction can be utilized in different data sets with rich textual information. Note that there is an increasing number of recent empirical studies that use

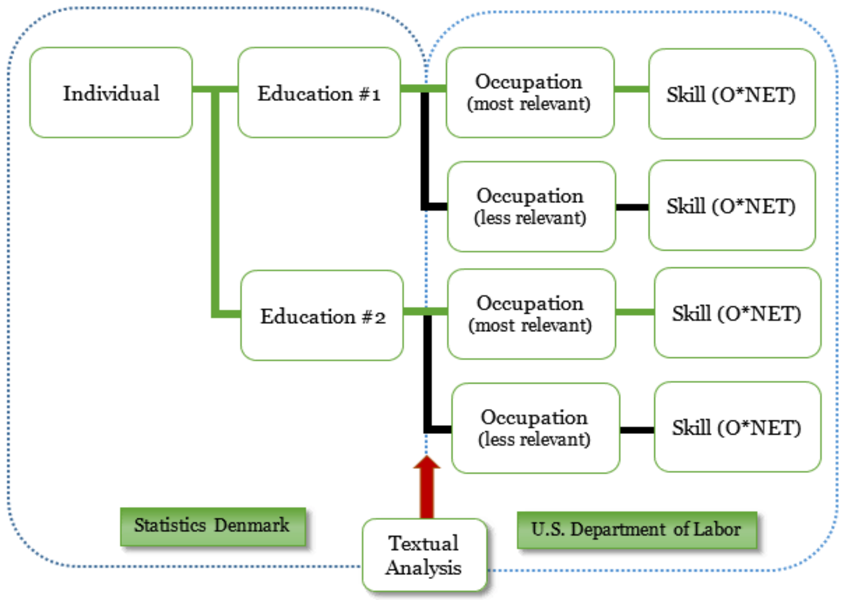


Figure 11: Mapping individual's education to skill

code connector where we feed in keywords collected from education and job training records and obtain the most relevant occupation. See Appendix A for further details.

Second, we use the O*NET data to obtain occupation-specific factor descriptions. There is no Danish version of O*NET data that we can utilize, hence, we assume that the occupation-specific task and skill requirements measured in the U.S. are similar to those in Denmark. Then, we reduce down the dimensionality of the data on occupation-specific standardized descriptors to cognitive, manual, and interpersonal skills, using principal component analysis. More specifically, we collect standardized descriptors in the following categories, in importance scales, reported at the O*NET-SOC-level: "cognitive abilities" in O*NET Abilities (1.A.1.a.1 - 1.A.1.g.2), "psychomotor and physical abilities" in O*NET Abilities (1.A.2.a.1 - 1.A.3.c.4), and "social skills" in O*NET Skills (2.B.1.a - 2.B.1.f). Then, we perform principal component analysis in each category and

textual data processed through machine learning techniques: Atalay et al. (2018), Gentzkow et al. (2018), Hoberg and Phillips (2016), Michaels et al. (2016), Gentzkow and Shapiro (2010), etc.

reduce the dimensions by taking the first principal component.[39]

So far, we have explained how we obtain a vector of cognitive, manual, and interpersonal skills using individuals' records in education attainment and professional training. In the final steps of skill assignment in the data, we incorporate the occupation-specific skills obtained using information on one's occupation choice at the 2 digit-level. For each point in time t, we add the two skill vectors to impute the individual skill scores:

$$s_{it} = \alpha s_{it}^{o} + (1-\alpha)s_{it}^{e} \quad \text{where} \quad \alpha > 0, \quad s_{it}^{o} = [c^{o}, m^{o}, p^{o}], \;\; s_{it}^{e} = [c^{e}, m^{e}, p^{e}] \tag{2.11}$$

Due to occupation switching or additional training, the skill values under the current construction may vary over time. Assuming that each experience through a job or institutional training provides value-added in one's skill accumulation, we compare the skill vector in time t (s_{it}) with that in time $t+1$ (s_{it+1}) and take the maximum of each skill component and replace the skill vector in time $t+1(s'_{it+1})$.

Firm Productivity We proxy firms using various measures: size, value-added, capital-intensity, and total factor productivity in the initial year of operation, which we obtain using methods in Olley and Pakes (1996). See Appendix A for details on the construction of measures.

Offshoring The measure of offshoring in this study is constructed using the value of

[39]The methodology we follow is adopted from existing studies such as Postel-Vinay and Lise (2015) and Goos et al. (2014), etc. that utilize the O*NET database in constructing multidimensional skill or task measures. To provide robustness on how the skill construction does not rely on the *a priori* skill categorization assigned in O*NET, we take all the descriptors and perform PCA obtaining the principal components, which demonstrate high correlation with the cognitive, manual, and interpersonal skills constructed above.

firm-level imported intermediate and final goods from abroad that are utilized in the production process and potentially substitute in-house workers, as in Hummels et al. (2014). In doing so, we exclude imports of raw materials[40] and only consider transactions that are in the same industry category as the firm's final good production: *narrow offshoring* (Feenstra and Hanson, 1999). Furthermore, focusing on the manufacturing sector in the case of Denmark ensures that the purpose of these purchases is not reselling for direct consumption (Hummels et al., 2014).[41] We also note that this measure of offshoring does not distinguish between carrying out production at Danish firms' own affiliates in a foreign country versus producing through arm's-length contracts with foreign firms.

The measure of offshoring in the reduced-form analysis is the industry-level offshoring exposure from low wage countries (Bernard et al., 2006). Following the literature (Feenstra and Hanson, 1999), we define and identify offshoring using firm-level data on intermediate and final good purchases from abroad in the Danish trade registers.[42] In particular, we utilize purchases that are used as inputs in the final good production and also serve as potential substitutes for in-house workers (Hummels et al., 2014). We aggregate these firm-level offshoring at the industry-level.

We focus on the North-South framework of offshoring (Feenstra and Hanson, 1997;

[40]To identify raw materials, we follow Eurostat and use the definitions of goods according to the fourth revision of Standard International Trade Classification (SITC rev. 4) section 2 (crude materials, inedible, except fuels) and section 4 (animal and vegetable oils, fats and waxes).

[41]There are challenges faced by empirical studies examining offshoring as they lack data sources that comprehensively cover offshoring activities. Looking into offshoring through imported intermediate and final goods, for example, fails to capture any offshoring activities through multinational activities of firms or the final assembly offshoring (Park, 2018).

[42]Offshoring in this paper means that the firm chooses to match with foreign workers instead of home workers. In bringing this notion to data, we use firm-level purchases of intermediate and final goods, which are interpreted as the embodiment of foreign workers' human capital and value-added. For example, if Danish firms purchase industrial robots from South Korea, this can be interpreted as Danish firms matching with South Korean workers whose value-added is captured in the form of robots.

Grossman and Rossi-Hansberg, 2008) where only the North finds offshoring a less expensive production alternative (one-way offshoring), and look into the intermediate and final goods purchases from low wage and eastern European countries (henceforth, simply 'South'). Low wage countries are defined as those with less than 5% GDP per capita relative to the U.S. during 1972-2001 (Bernard et al., 2006). The eastern European countries of interest are those that were included in the European Union through the Eastern Enlargement: Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, Slovenia, Malta, and Cyprus. Separately looking into offshoring from low wage countries provides substantive importance in thinking about changes in the feasible worker-firm matches. Due to recent institutional changes that integrate these countries into the global economy (e.g. China's accession to WTO (December 11th, 2001) and the Eastern Enlargement of EU (May 1st, 2004)), labor endowments that demonstrate clear differences in terms of skill composition, skill abundance, etc. have become easily accessible and further expanded economic activities worldwide.[43] In our sample, the share of offshoring from low wage and eastern European countries more than doubled, from 1.8% to 4.1%, and from 4% to 10% each. As for low wage countries, the share of offshoring from China increased from 47% to 82%.

Occupational Offshorability We mainly follow Blinder and Krueger (2013) to measure occupational offshorability at the ISCO two-digit level and to categorize occupations as offshoreable or non-offshoreable. Blinder and Krueger (2013) utilizes household survey

[43] In particular, the rise of China constitutes perhaps one of the most important trade shocks from low wage countries to hit the northern economies (Keller and Utar, 2016). China's share of imports to the United States and 12 EU countries more than doubled between 2000 and 2007 from 5.7% to 12.4% (Bloom et al., 2016)

	Variables	Mean	Standard Deviation
Worker	log(wage)	5.187	0.352
	Education (years)	16.249	4.548
	Experience (years)	13.047	6.158
Firm	Size	30.363	179.359
	Average wage	5.100	0.281
	Share of high-skilled	0.337	0.296
	Share of female	0.276	0.309
	log(value of imports)	14.062	2.829
	log(value of exports)	14.066	2.819
Aggregate	Share of exporters	0.346	0.0405
	Share of importers	0.350	0.052

Table 2: Summary Statistics for the Baseline Sample

measurements of job offshorability.[44] Offshorable occupations are generally associated with routine tasks that are easily codifiable (Autor et al., 2003; Oldenski, 2012) and the work performance in these jobs does not require direct physical contact; and geographic proximity is less important (Blinder, 2009; Blinder and Krueger, 2013; Goos et al., 2014). Also, offshorable jobs are not necessarily low in skill content: anecdotally, offshorable tasks that require high skills such as software programming, reading X-rays, or preparing tax forms have been offshored to low wage countries (Baumgarten, 2015). However, as the focus of this paper is on offshoring activities in the manufacturing sector, we classify the following occupations as offshorable: stationary plant and related operators; other craft and related trades workers; precision, handicraft, craft printing and related trades workers; machine operators and assemblers.

[44]The literature offers several different ways to capture the degree of offshorability at the occupation-level: Autor et al. (2003), Blinder (2009), Goos et al. (2014). To ensure that our results are not sensitive to the offshorability measure used to categorize occupations, we also try different measures following these existing studies. See Appendix A for details regarding comparisons between the measures.

2.4 Empirical Evidence

Here, we examine how greater exposure to offshoring from the South affects changes in the distribution of occupations across firms at the industry level, the employment composition at the firm level, and reallocations at the worker level.

Identification Strategy for Offshoring In each of the analysis below, the exogenous variation is the over time reduction in the cost of offshoring (e.g. tariffs) at the industry or the product level due to institutional changes: China's accession to the WTO and the eastern enlargement of EU. However, lowering of tariffs through such institutional changes has also increased trade flows at the final goods level and thereby affected labor market outcomes through the labor demand channel as well. We complement this strategy in the following two ways: First, we add time-varying and time-invariant controls to capture other factors at the industry level or firm level that potentially correlate with offshoring. Second, we employ an instrumental variable to address concerns regarding unobserved industry-level adjustments or industry-specific characteristics that change firms' incentives to offshore, re-ogranize the production process or the workforce, which also affect workers' reallocation risks. The instrument is similar to Hummels et al. (2014) and Baumgarten et al. (2013),

$$I_{kt} = \sum_h s_{hk0} \times \text{WES}_{ht} \tag{2.12}$$

where WES_{ht} is the export supply of product h from China to the world excluding Denmark in time t and $s_{hk0} = \frac{\text{Offshoring}_{hk0}}{\text{Offshoring}_{h0}} = \frac{\text{Offshoring}_{hk0}}{\sum_k \text{Offshoring}_{hk0}}$ is industry k's contribution to total offshored product h in the pre-sample period in Denmark where Offshoring_{hk0} is the value of offshoring in product h for industry k in the pre-sample period and aggregating this measure across all industries, $\sum_k \text{Offshoring}_{hk0}$ generates the total value of offshoring in product h in that year, which we denote as Offshoring_{h0}. In a nutshell, we combine the product-level export supply from China to the world excluding Denmark weighted by initial industry shares in offshoring of each product in Denmark. Note that we primarily focus on China as most of the change in offshoring from the South is driven by China. Therefore, the instrument, which has a product-time variation, is correlated with the value of Danish firms' purchases from low wage countries, but is external to the firm-level or worker-level labor market outcomes in Denmark. We further discuss how we address potential threats to the validity of the instrumental variable in each of the regressions below.

2.4.1 Occupational Segregation

Industry-level Analysis In order to examine the distributional effects of offshoring in terms of the degrees of occupational segregation by offshorability, we begin with the following industry-level regression.

$$\text{Segregation}_{kt} = \alpha_0 + \alpha_1 \text{Offshoring}_{kt} + \eta_k + \eta_t + \varepsilon_{kt} \tag{2.13}$$

Offshoring_{kt} is the share of narrow offshoring from low wage countries and eastern European countries in the aggregate value of offshoring in industry k. The dependent variable Segregation_{kt} captures the degrees of occupational segregation across firms in industry k using the segregation index (Kremer and Maskin, 1996), which is the ratio of the between-firm variance and the total variance.

$$\rho_{kt} = \frac{\text{Between-firm variance in industry } k \text{ in time } t}{\text{Total variance in industry } k \text{ in time } t} = \frac{\sum_{i,j}(\overline{x}_{jkt} - \overline{x}_{kt})^2}{\sum_{i,j}(x_{ijkt} - \overline{x}_{kt})^2} \tag{2.14}$$

We assign $x_{ijkt} = 1$ if worker i in firm j and industry k has an offshorable occupation according to the previously defined categories in time t. So $\overline{x}_{jkt}$ indicates the share of offshorable occupations in firm j that operates in industry k in time t, and $\overline{x}_{kt}$, the share of offshorable occupations in industry k in time t.[45] Thus, the segregation index captures the *variance in the share of offshorable occupations at the firm-level*, $\frac{\mathrm{Var}(\overline{x}_{jt})}{\overline{x}_t(1-\overline{x}_t)}$. In attempts to address concerns that certain industries that are inherently more segregated in their occupational structure potentially facing greater exposure to offshoring from the South, industry fixed effects (η_k) are included to control for time-invariant industry characteristics. We also add year fixed effects (η_t) to control for time-varying macroeconomic shocks such as the business cycle that potentially affect both offshoring intensity and the distribution of occupations. Thus, the coefficient α_1 captures the within-industry-over-time variation in the degree of matching due to changes in offshoring, net of aggregate time trends.

The degree of occupational segregation can be affected by time-varying industry components such as (i) the level of technology adoption (Acemoglu (1999), Albrecht and Vroman (2002)); (ii) other major global engagement activities such as exporting (Davidson et al., 2014); and (iii) domestic outsourcing (Goldschmidt and Schmieder, 2017). To disentangle the effect of offshoring from the effects of these channels, we further add technology intensity, export intensity, and domestic outsourcing intensity as controls. Technology intensity is constructed by taking the share of technical equipment and machinery in the capital stock. We use the sum of export value normalized by aggregate production for each industry to construct export intensities. Finally, we identify

[45] As for segregation by skill in general, we denote worker i's skill as x_{ijkt}, so $\overline{x}_{jkt}$ indicates the average skill-level at firm j that operates in industry k in time t, and $\overline{x}_{kt}$, the average skill-level in industry k in time t.

Table 3: Offshoring and Occupational Segregation

	OLS				IV			
	(1)[b]	(2)[c]	(3)	(4)	(1)	(2)	(3)	(4)
Offshoring[d]	0.0934***	0.0320*	0.0251	0.0228*	0.864***	0.0885	0.00716	0.120**
	(0.0246)	(0.0174)	(0.0188)	(0.0123)	(0.175)	(0.0644)	(0.0716)	(0.0532)
Observations[a]	2,097	2,097	2,097	2,097	2,097	2,097	2,097	2,097
R^2	0.757	0.790	0.850	0.833	0.399	0.785	0.850	0.819
Industry FE & Year FE	yes	yes	yes	yes	yes	yes	yes	yes
Industry Controls	yes	yes	yes	yes	yes	yes	yes	yes

Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

[a] The unit of observation is an industry (4-digit) in a given year (1995-2004).
[b] Segregation index computed for each industry where $x_{ijt} = 1$ is assigned if a worker has an offshorable occupation.
[c] Segregation index computed for each industry where x_{ijt} is cognitive (column (2)), manual (column (3)), and interpersonal (column (4)) skills of a worker, respectively.
[d] Offshoring is the share of relevant intermediate and final good purchases (narrow offshoring) from low wage countries Bernard et al. (2006) and eastern European countries (Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, Slovenia, Malta, and Cyprus) to aggregate offshoring constructed at the industry level (4-digit).

outsourcing activities[46] using the sum of variables on the cost of intermediate goods, the cost of subcontractors, and the cost of temporary employment agencies normalized by value-added.[47] To control for any time-varying industry-specific demand or technology shocks that demonstrate correlations with offshoring from the South, we employ the instrumental variables.

Baseline results (Table 2) show a positive and statistically significant correlation between the share of offshoring activities from low wage and eastern European countries in aggregate offshoring and the degree of occupational segregation across firms. We find qualitatively similar results for segregation by skill in response to the offshoring shock. That is, industries that are exposed to high offshoring from the South demonstrate a more segregated occupational structure, and a more sorted workforce. Quantitatively,

[46]As an alternative way to control for outsourcing, we identify firms that perform outsourcing following Goldschmidt and Schmieder (2017) and exclude them from the sample and repeat the exercise above.

[47]We exactly follow Bagger et al. (2014) to construct value-added from the Danish data. See Appendix C for details.

one standard deviation increase in the share of offshoring from the South is associated with an increase of 0.143 standard deviations in occupational segregation, which corresponds to a 7.5% increase in the degree of segregation relative to the mean value of 0.24. This demonstrates economic significance in the association between changes in the distribution of occupations across firms and changes in offshoring shares from the South. Furthermore, the quantitative magnitude is greater employing the instrumental variable.

2.4.2 Within-Occupation Quality of Skill

We present within-occupation evidence using both firm-level and worker-level regressions on how offshoring exposure from the South affects the firm-level average skill quality and the reallocation of workers within these occupations across firms. Note that the regression is conducted for offshorable occupations. In order to avoid issues related to product-level import penetration that directly affects industry size, we focus on *export-oriented* industries where an industry is defined as *export-oriented* if (i) the change in the value of net exports is greater than zero and it has a positive net export value in 2004; or (ii) it continues to have positive net export values between the years 1995 and 2004. Otherwise, it is identified as an *import-oriented* sector.

Firm-level Analysis The following firm-level regression examines how the average skill of workers hired in-house responds to changes in industry-specific exposure to offshoring from low-wage and eastern European countries. Note that the interaction term captures additional information on the between-firm inequality in average skills.

$$\begin{aligned}\text{Average Skill}^s_{jkt} =& \alpha^m_0 + \alpha^m_1 \text{Offshoring}_{kt} + \alpha^m_2 (\text{Offshoring}_{kt} \times \text{TFP}_{jk0}) + \alpha^m_3 \text{TFP}_{jk0} \\ &+ \text{Firm}_{jkt} + \text{Industry}_{kt} + \eta^m_{k't} + \varepsilon^m_{jkt}\end{aligned} \tag{2.15}$$

Average Skill^s_{jkt} is the average skill s of workers in firm j and industry k in time t where s = cognitive, manual, or interpersonal skills. TFP_{jk0} is the firm-specific total factor productivity in the initial year of operation obtained using Olley and Pakes (1996). Time-varying firm controls include size, share of high-skilled workers, share of female workers, capital intensity, outsourcing intensity, and exporting intensity. We also add time-varying industry controls, which include capital intensity, exporting intensity, and domestic outsourcing intensity. Note that adding the firm's size controls for any changes in average skills due to an expansion in employment size as a result of productivity gains from performing offshoring (Hummels et al., 2014). We further include sector-by-year fixed effects ($\eta_{k't}$) to control for any sector-specific time-varying exogenous demand or technology shocks. Therefore, the coefficient α^m_2 for the interaction term together with α^m_1 provides implications for changes in the average skill of firms and the magnitude of change based on the firm's initial productivity levels. For example, $\alpha^m_1 > 0$ and $\alpha^m_2 < 0$ indicates that with an increase in the offshoring shock from the South, the average quality of in-house workers improves and the extent to which firms improve the quality of worker skills is greater for low productivity firms compared to high productivity ones. the between-firm inequality across firms in their average workers' skills decreases. As the industry-level offshoring measure is constructed using the sum of firm-level offshoring, simultaneity concerns potentially arise in the presence of industries with a high

Table 4: Offshoring and Average Quality of Skill

South		OLS			IV	
	(1)[c]	(2)	(3)	(1)	(2)	(3)
Offshoring[b]	0.025**	0.0139	0.0202*	0.179***	-0.0533	0.0691
	(0.009)	(0.013)	(0.011)	(0.037)	(0.049)	(0.043)
Offshoring × TFP	-0.006	-0.003	-0.002	-0.0001	0.00429	0.00598
	(0.0081)	(0.002)	(0.002)	(0.005)	(0.007)	(0.005)
Observations[a]	13,614	13,614	13,614	13,614	13,614	13,614
R^2	0.115	0.415	0.164	0.069	0.448	0.159
Firm & Industry Controls	yes	yes	yes	yes	yes	yes
Industry × Year FE	yes	yes	yes	yes	yes	yes
Low Wage Countries		OLS			IV	
	(1)	(2)	(3)	(1)	(2)	(3)
Offshoring[c]	0.123***	-0.0659	0.130***	0.656***	-0.200	0.246
	(0.061)	(0.083)	(0.074)	(0.140)	(0.187)	(0.162)
Offshoring × TFP	-0.0207***	0.0098	-0.0155**	-0.0131	0.0179	0.0148
	(0.006)	(0.007)	(0.006)	(0.018)	(0.025)	(0.021)
Observations[a]	13,614	13,614	13,614	13,614	13,614	13,614
R^2	0.115	0.415	0.164	0.079	0.448	0.164
Firm & Industry Controls	yes	yes	yes	yes	yes	yes
Industry × Year FE	yes	yes	yes	yes	yes	yes

Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

[a] The unit of observation is a firm in a given year (1995-2004).
[b] Offshoring is the share of relevant intermediate and final good purchases (narrow offshoring) from low wage and eastern European countries constructed at the industry-level (4-digit).
[c] Columns (1) in each panel correspond to results based on the average of workers' cognitive skills as the dependent variable; columns (2), their manual skills; columns (3), their interpersonal skills.

concentration ratio. Again, we employ an instrument to address this concern.

Results (Table 3) show that, within offshorable occupations, exposure to offshoring from low wage and eastern European countries is positively associated with the average quality of workers' cognitive and interpersonal skills. In particular, the coefficient α_1^m maintains statistical significance for average cognitive skills of workers across different specifications. Note that the magnitude of improvement is greater in response to offshoring exposure from low wage countries only, compared to that of all countries in the South including eastern European countries. Also, we find evidence that the extent to which firms improve their average quality of worker skill is greater for the low-productivity firms compared to the high-productivity firms. That is, firms with an initial TFP below 5.85 (i.e. TFP $< \frac{0.123}{0.021} = 5.85$) would benefit from the offshoring shock in terms of average quality of workers' cognitive skills that they hire while those above would not.[48] We find a negative sign for α_2^m across all specifications for average cognitive skills while the statistical significance holds only for the fixed effect specification with low wage countries. This can be explained by how high-productivity firms offshore and find substitutes for their high-skilled workers allowing lower productivity firms to match with the next best workers that the offshoring firms release.

Worker-level Analysis If firms are improving in the average quality of workers' skills in export-oriented industries in response to offshoring from the South, does it mean that workers face a greater risk of undergoing downward transitions? Here, we examine whether workers are more likely to move down the firm ladder or switch out to a less competitive sector, in response to offshoring exposure from the South. We use the following worker-level regression to investigate these hypotheses, changing definitions of the dependent variable accordingly.

$$\begin{aligned} C_{ijkt} =& \alpha_0^c + \alpha_1^c \text{Offshoring}_{kt} + \alpha_2^c (\text{Offshoring}_{kt} \times \text{ Skill}_i^s) + \alpha_3^c \text{Skill}_i^s \\ &+ \text{Worker}_{ijkt} + \text{Firm}_{jkt} + \text{Industry}_{kt} + \eta_{k't}^c + \varepsilon_{ijkt}^c \end{aligned} \tag{2.16}$$

C_{ijkt} is a dummy variable set equal to 1 if a worker experiences *downward transitions* between time t and $t+1$: (i) a worker is reallocated to a firm with lower TFP than his or her previously matched firm; (ii) a worker switches out to the import-oriented sector. Skill_i^s is the level of worker i's skill s where s = cognitive, manual, or interpersonal skills. In addition to the time-varying controls for firms and industries described in

[48] The average firm with a TFP of 6.923 is expected to experience an increase in average cognitive skills by 0.002 standard deviations in response to an increase in offshoring from the South by 1 standard deviation.

the previous regression exercise, we further include time-varying controls for workers: years of experience and years of education. Again, we control for sector-by-year fixed effects ($\eta^c_{k't}$). Therefore, the coefficient α^c_2 for the interaction term together with α^c_1 provides implications for changes in the probabilities workers face in terms of descending transitions with further information on the magnitude of change by worker skill-levels. For example, $\alpha^c_1 > 0$ and $\alpha^c_2 < 0$ indicate that workers with lower skill levels are more likely to switch down their firm or sector in response to an increase in the offshoring shock from the South. Simultaneity or reverse causality is less of an issue in worker-level regressions as it is unlikely to have individual workers affect industry level offshoring (Ebenstein et al., 2014; Baumgarten et al., 2013). However, if high-skilled workers in Danish manufacturing sort into industries with high offshoring activities from the South, this may not be a negligible issue. Again, we employ the instrumental variable in this analysis to address this concern.

In terms of the qualitative implications of the results (Table 4), an increase in offshoring from the South increases the probability that workers in offshorable occupations undergo a descending transition in workplace as well as sectors where those with low cognitive skills face a relatively greater risk. Again, the magnitude of the coefficients is greater when employing the instrumental variable. Quantitatively, the average worker with cognitive skills of 0.38 faces a greater probability of experiencing a transition in sector by 0.006 standard deviations in response to an increase in offshoring from the South by one standard deviation while for workers in the bottom quartile on average, by 0.023 standard deviations. As for reallocation across firms, workers with cognitive skills below 0.44 face a positive probability of moving down to a firm with lower TFP in response to an offshoring shock. Comparing the magnitudes across workers by their

Table 5: Offshoring and Worker Reallocations across Sectors

Firms	OLS			IV		
	(1)[c]	(2)	(3)	(1)	(2)	(3)
Offshoring[b]	0.0827***	0.0417***	0.0596***	0.0666**	-0.0403	0.0416
	(0.011)	(0.009)	(0.009)	(0.029)	(0.025)	(0.030)
Offshoring × Skill	-0.0469**	0.0437***	0.00896	-0.149***	0.0991***	-0.0770**
	(0.023)	(0.016)	(0.017)	(0.046)	(0.033)	(0.039)
Observations[a]	312,353	312,353	312,353	312,353	312,353	312,353
R^2	0.072	0.072	0.072	0.072	0.072	0.072
Controls	yes	yes	yes	yes	yes	yes
Industry × Year FE	yes	yes	yes	yes	yes	yes
Sector	OLS			IV		
	(1)	(2)	(3)	(1)	(2)	(3)
Offshoring	0.0085**	-0.0024	0.0043	0.0850***	0.0452***	0.0823***
	(0.004)	(0.003)	(0.003)	(0.012)	(0.011)	(0.013)
Offshoring × Skill	-0.0124**	0.0117***	-0.0018	-0.0621***	0.0031**	-0.0482***
	(0.008)	(0.005)	(0.006)	(0.017)	(0.013)	(0.014)
Observations	312,353	312,353	312,353	312,353	312,353	312,353
R^2	0.055	0.055	0.055	0.052	0.052	0.052
Controls	yes	yes	yes	yes	yes	yes
Industry × Year FE	yes	yes	yes	yes	yes	yes

Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

[a] The unit of observation is a worker in a given year (1995-2004).
[b] Offshoring is the share of relevant intermediate and final good purchases (narrow offshoring) from low wage and eastern European countries constructed at the industry-level (4-digit).
[c] Columns (1) in each panel correspond to results based on workers' cognitive skills which are interacted with the share of offshoring; columns (2), their manual skills; columns (3), their interpersonal skills.

cognitive skills, those in the top quartile on average face a greater probability of switching down by 0.069 standard deviations in response to an increase in offshoring from the South by one standard deviation, whereas for those in the bottom quartile on average, by 0.104 standard deviations.

With high-productivity firms performing offshoring, high-skilled workers hired in these firms face direct competition from foreign workers, which effectively alters worker-firm matching at the top of the distributions for both workers and firms. This impact of foreign labor supply subsequently spills over down the distributions, eventually affecting even those who are not directly exposed to offshoring from the South. In other words, offshoring causes workers in offshorable occupations to move down the firm ladder, and ultimately the least productive workers are reallocated to the less competitive sector. As in Hummels et al. (2014), a potential threat to the instrumental variable in the analyses above is whether domestic demand shocks from Denmark affect the world export supply from China. In the industry-level analysis, it is unlikely that the industry-level demand shocks in Danish manufacturing affect the world export supply as Denmark is a small open economy. In the firm-level and worker-level analyses, this threat is less of a concern as it is unlikely that firm-level or worker-level demand meaningfully affects the aggregate world export supply from China.

2.5 Structural Estimation

In this section, we use the Danish matched employer-employee data to estimate the key model parameters and examine the quantitative impact of the globalization channel on: (i) how worker-firm matching evolves, and (ii) how the between-firm inequality in wages is affected. The value-added in analyzing through the lens of a structural model lies in assessing the quantitative importance of offshoring in comparison to competing hypotheses and identifying the main channel that drives labor market inequality.

Along with channels of globalization that lowers the cost of matching with foreign workers, there are important concurrent changes that potentially affect worker-firm matching and further distributional labor market outcomes: changes in the skill distribution of workers and that in the production technology. If the supply of workers' skills has evolved toward higher skill dispersion, then for a fixed production technology, worker-firm matching is affected in a way that increases segregation by skill (Kremer and Maskin, 1996).

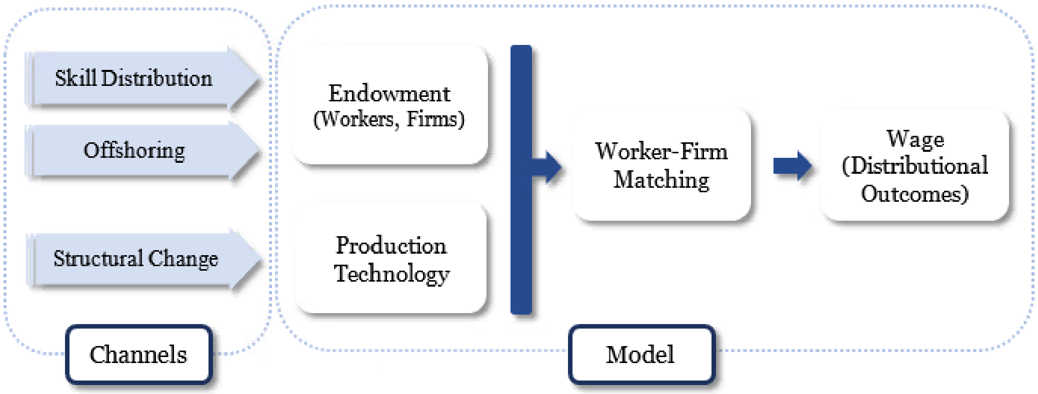


Figure 12: Structural estimation: decomposing the contributions of different sources to wage inequality

Additionally, structural changes in the production technology such as skill-biased technological change (Acemoglu and Autor, 2011; Lindenlaub, 2017) can alter worker-firm matching: cognitive skills relative to manual skills of workers become more important, and thus, the assortative matching between firms and workers is greater on worker's cognitive skills. In other words, firms' productivity becomes more *complementary* with the cognitive skill of workers, resulting in a greater degree of assortative matching on worker's cognitive skills.

Multidimensional Skills Introducing multidimensional skills, the bilinear production technology assumed in Section 3 extends as follows:

$$q(x_c, x_m, x_p, y) = (\gamma_c x_c + \gamma_m x_m + \gamma_p x_p)y + u_c(x_c) + u_m(x_m) + u_p(x_p) \tag{2.17}$$

The production technology parameters $\gamma_c, \gamma_m, \gamma_p$ represent the *strength of complementarities* between firms' productivities and different dimensions of workers' skills in each occupation, which indicates whether workers' cognitive, manual, or interpersonal skills

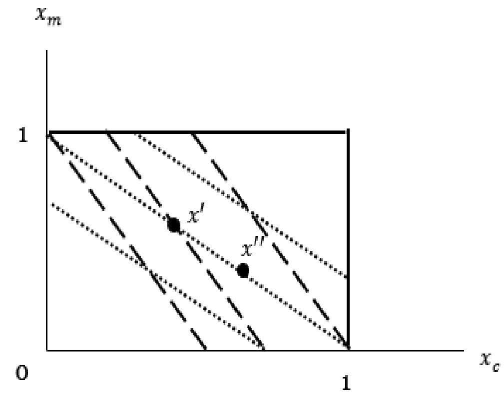


Figure 13: Matching with multidimensional skills and technological change

Suppose there exist workers $x' = x'' \equiv \hat{x}$ with skill bundles $x' = [x'_c, x'_m]$ and $x'' = [x''_c, x''_m]$ where $x'_c < x''_c$, $x'_m > x''_m$ with γ_c, γ_m. If there is an increase in cognitive skill complementarity γ'_c (dashed line) compared to the previous one $> \gamma_c$ (dotted line), then the following holds: $\hat{x}' = \hat{x} + x'_c(\gamma'_c - \gamma_c) < \hat{x} + x''_c(\gamma'_c - \gamma_c) = \hat{x}''$. Thus, workers x' and x'' that were previously matched with the same firm-type $\hat{y}$, are now separated into different firms $\hat{y}'$ and $\hat{y}''$, respectively, where $\hat{y}' < \hat{y}''$.

are complements or substitutes to firms' productivity.[49] we also include skill-specific main effects, capturing the extent to which each skill component of the workers contributes to the task output independent of the firm's productivity.[50]

2.5.1 Estimation Strategy

We first identify the skill complementarity parameter $\Gamma = [\gamma_c, \gamma_m, \gamma_p]$ that would generate the same model moments as those observed in the data $(i = c, m, p)$[51]:

$$E_{\pi^{\Gamma}}[X_i Y] = E_{\pi^{\hat{\Gamma}}}[X_i Y] \tag{2.18}$$

[49]While it is possible to have multiple characteristics on both firms and workers to examine complementarity across different attributes, we keep the firm-side as unidimensional in order to maintain the focus on the worker-side attributes in explaining chages in matching. See Lindenlaub (2017) for a two-sided multidimensional matching models with analytical solutions.

[50]Note that the skill-specific main effect components can also take nonlinear functions (Dupuy and Galichon, 2014).

[51]Dupuy et al. (2017) proves that the log-likelihood function in MLE is equivalent to the moment matching estimator.

More specifically, we derive the marginal densities $f(x)$ and $g(y)$ from the data and use the iterated proportional fitting procedure (IPFP) to recover the Lagrangian multipliers $a(x)$ and $b(y)$ for a particular assumed Γ, which simulates a corresponding worker-firm matching. Optimal $\hat{\Gamma}$ is obtained by a moment matching procedure where we iterate this process until the difference between the model moments and the data moments is minimized.[52] Using the estimated $\hat{\Gamma}$ together with data on workers' wages, we non-parametrically estimate the scale parameter λ_x capturing the extent to which workers' unobserved characteristics matter in the matching process, which subsequently determines λ_y, and the parameters for the worker's skill-specific main effects.

Results Recall that the parameters in the production technology represent the strength of complementarities between firms' productivity and different dimensions of workers' skill. Here, we highlight features of the estimated skill complementarities in the Danish data for the following ISCO 1-digit occupation category (Table 5). First, for most occupations, cognitive and interpersonal skills relative to manual skills demonstrate greater importance in worker-firm matching for both 1995 and 2004. As for workers in craft occupations, for example, increasing the workers' cognitive (interpersonal) skills and the firms' productivity by one standard deviation increases the task output by 0.134 (0.074) units in 1995 and 0.116 (0.115) units in 2004. However, increasing their manual skills and the firms' productivity by one standard deviation increases the task output by 0.084 units in 1995 and 0.072 units in 2004.

Second, the manual skills of workers become more substitutable for managers, clerical workers, and those in sales and services while complementary for the rest. A negative

[52] For estimation purposes, we consider each job the firm holds as a unit of firm.

		Crafts	Elementary	Managers	Clerical	Machines	Professionals	Service, Sales	Associates
1995	Cognitive	0.134	0.045	0.045	-0.014	0.068	0.036	0.063	0.046
		(0.010)	(0.017)	(0.040)	(0.015)	(0.008)	(0.040)	(0.030)	(0.023)
	Manual	0.084	-0.015	0.007	-0.065	0.033	0.048	0.058	0.001
		(0.010)	(0.016)	(0.036)	(0.013)	(0.008)	(0.036)	(0.029)	(0.022)
	Interpersonal	0.074	0.000	-0.051	-0.010	0.023	0.059	0.077	-0.024
		(0.009)	(0.014)	(0.034)	(0.018)	(0.007)	(0.039)	(0.027)	(0.018)
2005	Cognitive	0.116	0.012	-0.041	-0.019	0.072	0.095	0.061	0.031
		(0.010)	(0.016)	(0.034)	(0.019)	(0.007)	(0.025)	(0.035)	(0.015)
	Manual	0.072	0.033	-0.025	-0.109	0.062	0.044	-0.021	0.012
		(0.009)	(0.016)	(0.034)	(0.017)	(0.007)	(0.025)	(0.034)	(0.016)
	Interpersonal	0.115	0.070	0.142	-0.011	0.017	0.062	0.008	0.060
		(0.011)	(0.015)	(0.034)	(0.030)	(0.007)	(0.027)	(0.040)	(0.016)

Bootstrapped standard errors in parentheses

Table 6: Estimated complementarity by ISCO (1-digit)

value in the estimated coefficients indicates how an increase in the firm's productivity increases the task output of the matched pair whose workers are relatively less skilled in their manual ability. Consistent with what existing studies find (Autor et al., 2003; Autor and Dorn, 2013; Lindenlaub, 2017), how talented workers are in their physical or psychomotor ability became less important for the firms in hiring workers due to changes in the production technology (e.g.automation or mechanization).

Third, for most occupations, the importance of cognitive and interpersonal skills grew over time. Not only do cognitive and interpersonal skills remain important characteristics across time, but also their importance relative to manual skills increased. For example, for workers in professional occupations in 1995, increasing their cognitive skills and the firms' productivity by one standard deviation increased the task output by 0.036 units, and to achieve the equivalent increment in the task output, workers' manual skill and firms' productivity both had to increase by 0.87 ($=\sqrt{\frac{0.036}{0.048}}$) standard deviations. However, in 2004 the corresponding increment is 1.47 ($=\sqrt{\frac{0.095}{0.044}}$) standard deviations for both workers and firms, which indicates the rising importance of cognitive skills relative to manual skills in task output. Using a similar argument for workers in craft occupations, increasing their interpersonal skills and firms' productivity by one standard deviation raises output by 0.074, which takes an increase by 0.94 ($= \sqrt{\frac{0.074}{0.084}}$) standard deviations for workers' manual skills and the productivity of firms in 1995 to obtain the equivalent amount; however, it becomes 1.26 ($= \sqrt{\frac{0.115}{0.72}}$) standard deviations in 2004.

2.5.2 Estimation with Offshoring

In the estimation results so far, we assume that matches between workers and firms observed in the data capture the full population of firms and workers in Danish manufacturing; however, what we observe in the data are worker-firm pairs that chose to match domestically, which fails to capture the international matches. More specifically, the problem with the data associated with estimating the model with offshoring comprises two parts: for each occupational category, (i) firms that match with foreign workers are not observed in the data; and (ii) the skill characteristics of foreign workers are not provided. In the following, we elaborate on how we approach this problem and quantitatively capture the effects of offshoring.

Offshoring The measure of offshoring in this study is constructed using the value of firm-level imported intermediate and final goods from abroad that are utilized in the production process and potentially substitute in-house workers. Conceptually, offshoring is the formation of international teams in production: firms' choice to match with foreign workers instead of home workers. In operationalizing this notion of offshoring to data, we give the following interpretation of the worker-firm international matches: firms, through their purchases of intermediate or final goods, are essentially matching with foreign workers whose value-added is encapsulated in the form of intermediate or final goods. For example, if Danish firms purchase industrial robots from South Korea, this can be interpreted as Danish firms matching with South Korean workers whose value-added is captured in the form of robots.

Identification Strategy with Offshoring As previously discussed, firms that face workers with the exact same skill qualities at home and abroad are indifferent between domestic matching and offshoring. Thus, the model implies that the per worker value-added for each firm,[53] which captures the average task output of a domestic worker through a successful match, should be equal to the per worker value-added of an offshored match: $q(x, y) = q(x_F, y)$. To obtain a relevant measure in the data, we equate the per worker value-added to the constructed firm-level offshoring measure per worker composite overseas.

$$\frac{\text{Value-Added}}{\text{Number of Domestic Workers}} = \frac{\text{Intermediate and Final Good Purchases Abroad}}{\text{Number of Offshored Composite Workers}} \tag{2.19}$$

Thus, the task output of one Danish worker corresponds to the equivalent output provided by a composite of workers from the South.[54] Note that there are firms that

[53] Here, we assume that the value-added shares produced by each occupation are equal to the firm-level occupational shares.

[54] What matters in the estimation is the final quality of skill provided through a match, whether it is a single worker or a bundle of workers. Therefore, the identification strategy is not sensitive to the ratio of workers between two different origins which potentially depends on wage differences.

perform offshoring, yet do not hire any workers in offshorable occupations. This potentially stems from the definition of offshorability that is used for categorizing occupations or from a situation where these firms only keep the non-offshorable jobs in-house and rely on purchasing intermediates and final goods for the rest of the production process. These firms comprise a negligible portion of the sample; however, to ensure robustness, we try several different things: dropping these firms in the estimation or performing data imputation.[55] To summarize the estimation strategy with offshoring, we add on the number of offshored matches recovered using the strategy above to the supply of offshorable occupations for each firm. Then, the data moments including the offshored matches are derived, with which we match the model moments and obtain the estimated skill complementarity for each occupation with offshoring for the years 1995 and 2004.

Estimation Results with Offshoring There are several notable features from the estimation results using the dichotomous categorization of occupations (Table 6). While workers' interpersonal skills are complements with firms' productivity in both occupations, cognitive and manual skills demonstrate opposite patterns. That is, workers in non-offshorable occupations show complementarity (substitutability) in their cognitive (manual) skills with the qualities of the matched firms while those in offshorable occupations exhibit substitutability (complementarity). Taking into account that the occupations categorized as offshorable are stationary plant and related operators; precision, handicraft, craft printing and related trades workers; machine operators and assemblers, etc., it makes sense that workers' manual skills are important and complementary in the

[55]For example, we assume that these firms match with foreign workers that provide the quality of skill equivalent to that of the average of their in-house workers with non-offshorable jobs.

	Offshorable Occupations			Non-offshorable Occupations		
	Cognitive	Manual	Interpersonal	Cognitive	Manual	Interpersonal
1995	-0.270	0.231	0.380	0.048	-0.088	0.061
	(0.007)	(0.007)	(0.008)	(0.004)	(0.004)	(0.004)
2005	-0.160	0.145	0.283	0.091	-0.091	0.048
	(0.007)	(0.007)	(0.006)	(0.004)	(0.004)	(0.004)

Bootstrapped standard errors in parentheses

Table 7: Estimated complementarity with offshoring

worker-firm matching process. Nonetheless, it is worth mentioning that the evolution of skill complementarity with firms' productivity in both occupations demonstrates that cognitive skills are becoming more complementary or less substitutable (i.e. the value of the coefficients increases) while manual skills become less complementary or more substitutable (i.e. the value of the coefficients decreases) over time.

2.5.3 Counterfactual Exercises

Here, we use the estimated model to separately quantify the effect of offshoring on changes in the worker-firm matching and the between-firm inequality for offshorable occupations.[56] As discussed earlier, changes in the feasible worker-firm matches occur not only due to access to additional workers via offshoring (O), but also due to concurrent changes in the economy such as the structural changes ($\hat{\Gamma}$) as well as shifts in the supply of skill (S) in the economy. Therefore, we disentangle the three channels and quantify the effects of each through a decomposition exercise, in which we allow only one channel to change at a time while shutting down the rest. We employ this method to examine

[56]While there exist effects of offshoring that change the nature of firms' in-house production (e.g. greater intensity in R&D activities, better management and efficient organization, etc.), this lies beyond the scope of this study.

measures of labor market inequality such as the between-firm wage inequality as well as segregation by skill.

$$\begin{aligned}\hat{\pi}_{04}(\hat{\Gamma}_{04}, O_{04}, S_{04}) - \hat{\pi}_{95}(\hat{\Gamma}_{95}, O_{95}, S_{95}) &= \underbrace{\{\hat{\pi}_{04}(\hat{\Gamma}_{04}, O_{04}, S_{04}) - \hat{\pi}(\hat{\Gamma}_{04}, O_{95}, S_{04})\}}_{\text{offshoring}} \\ + \underbrace{\{\hat{\pi}(\hat{\Gamma}_{04}, O_{95}, S_{04}) - \hat{\pi}(\hat{\Gamma}_{95}, O_{95}, S_{04})\}}_{\text{structural change}} &+ \underbrace{\{\hat{\pi}(\hat{\Gamma}_{95}, O_{95}, S_{04}) - \hat{\pi}_{95}(\hat{\Gamma}_{95}, O_{95}, S_{95})\}}_{\text{skill distribution}}\end{aligned} \tag{2.20}$$

Worker-Firm Matching Looking into changes in matching by each skill dimension within offshorable jobs over time (Table 7), we find that the degree of assortative matching of workers with firms has increased in terms of workers' cognitive and interpersonal skills while it has decreased in their manual skills. Comparing the magnitudes across different channels, structural change plays a major role in affecting changes in worker-firm matching among other channels. While offshoring affects matching in a qualitatively similar way, the magnitude is quite small. Note that the skill supply of workers in the manufacturing sector for offshorable jobs decreases the degree of assortative matching on workers' cognitive and interpersonal skills.

Between-Firm Wage Inequality and Segregation by Skill How do changes in worker-firm matching due to globalization affect between-firm inequality in wages? In the graphs below, we show changes in log wages by decile of firms' productivity where a positive slope indicates an increase in between-firm wage inequality as high-productivity firms increase wages for their workers more than low-productivity firms do. Results from the baseline model demonstrate that the change in the average wage for each decile of firms by their productivity increases over time, which indicates an increase in

	Initial	ΔSkill Supply	ΔStructural	ΔOffshoring	**Final**
Cognitive	-0.135	-0.009	0.080	0.002	-0.064
	(0.004)	(0.002)	(0.006)	(0.000)	(0.005)
Manual	0.141	0.005	-0.074	-0.002	0.072
	(0.004)	(0.002)	(0.006)	(0.000)	(0.005)
Interpersonal	0.101	-0.007	0.010	0.000	0.113
	(0.004)	(0.002)	(0.006)	(0.000)	(0.004)

Bootstrapped standard errors in parentheses

Table 8: Decomposition in changes in worker-firm matching

between-firm wage inequality. However, in a counterfactual economy where changes in offshoring do not take place, the slope is even greater. That is, without the channel of globalization supplying additional workers for firms to match with, firms demonstrate greater differences in terms of the average wage they pay.[57]

As shown in Figure, examining each channel provides sharp comparisons in the effect on between-firm inequality. That is, the structural change channel mainly drives changes in the wage across firms to become more different whereas offshoring functions in a way that reduces the wage gap between firms. In terms of wage levels, formations of international teams by high-productivity firms impose direct competition in the high-skilled workers they hire, which generates an overall wage loss that is greater in magnitude for these workers with direct exposure to foreign worker competition. In support of the analysis by the decile of firms, we also compute changes in the average and variance of wages together

[57]See Appendix A for the decomposition analysis using both offshorable and non-offshorable occupations.

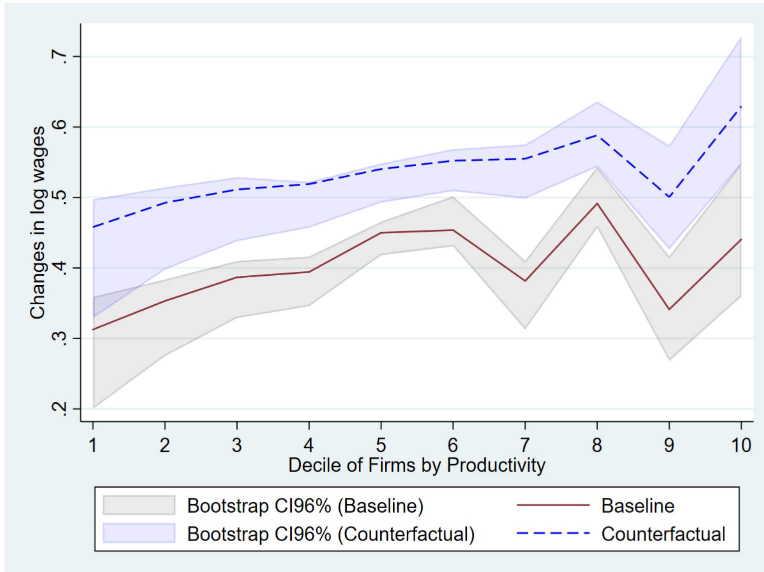


Figure 14: Changes in log(wage) by decile of firms (productivity)

with changes in the segregation index (Kremer and Maskin, 1996),[58] which captures how sorted the economy is in terms of the distribution of average wage payment of each firm (i.e. between-firm wage inequality). Consistent with the theoretical predictions, the offshoring channel lowers the average wage of workers in offshorable occupations; however, it brings about a decrease in wage dispersion across firms. Also, the segregation index indicates how the introduction of international teams brings about a decrease in between-firm wage inequality.

58

$$\rho_t = \frac{\text{Between-firm variance}}{\text{Total variance}} = \frac{\sum_{i,j}(\overline{x}_{jt} - \overline{x}_t)^2}{\sum_{i,j}(x_{ijt} - \overline{x}_t)^2} \tag{2.21}$$

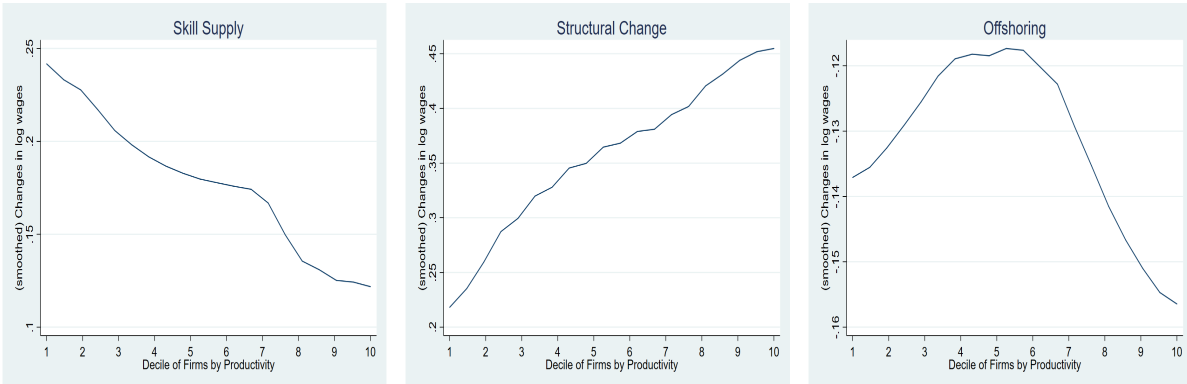


Figure 15: Changes in log(wage) by decile of firms (productivity) through each channel

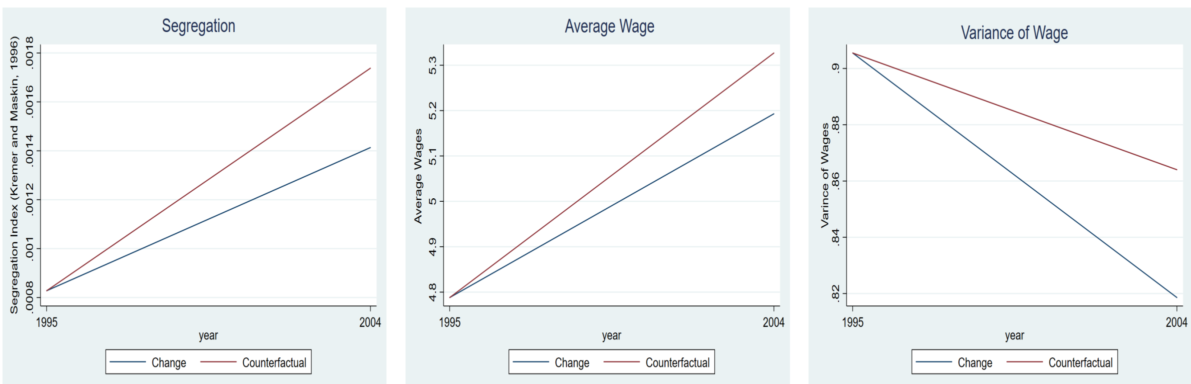


Figure 16: Changes in Segregation, Average and Dispersion in Wages

2.6 Conclusion

In this paper, we examine the mechanisms of how offering affects labor market inequality by altering the reallocation of workers across firms. We use the Danish employer-employee matched data together with the newly constructed skill measures to evaluate the effect of offshoring on workers across the skill distribution within offshorable occupations. Using both the model and data, we find that offshoring reduces domestic worker wages; and increases the probability of reallocation away from the high-productivity firms to the low-productivity ones. The least skilled workers further face a greater risk of switching out to a less competitive sector. On the firm-side, offshoring improves the average skill of in-house workers at a lower cost. Analyzing through the lens of a structural model, we examine the mechanisms of how offshoring affects labor market inequality and further assess the quantitative importance of various competing hypotheses such as technological change and the expansion of higher education, in addition to offshoring. We actually find substantially different effects: technology mainly increases the inequality between firms in terms of worker skill quality and average wages, while offshoring mitigates this rising trend.

The novelty in the analysis lies in examining the effects of offshoring at the within-occupation-worker level using the newly constructed skill measure. Together with the matched Danish data, we further examine changes in the skill mix of workers observed at the firm-level. This potentially provides significant implications for setting objectives and designing specific curriculums of job training or trade adjustment assistance promgrams. It may also serve as useful guidelines for individuals on making human capital investment decisions and help designing effective education policies that prepare individuals to demonstrate competitiveness as workers in a global economy setting. The structural framework, which extend the Becker-type matching model, demonstrates how offshoring contributes to recent trends in labor market inequality where we see a significant portion being explained by between-firm inequality. The significance of the model lies in not only disentangling the effects of concurrent and paramount forces affecting labor market inequality but also evaluating the quantitative importance of each channel.

Bibliography

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67:251–333.

Abraham, K. G. and Taylor, S. K. (1996). Firms' Use of Outside Contractors: Theory and Evidence. *Journal of Labor Economics*, 14(3):394–424.

Acemoglu, D. (1999). Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence. *American Economic Review*, 89(5):1259–1278.

Acemoglu, D. and Autor, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. *Handbook of Labor Economics*, 4:1043–1171.

Albrecht, J. and Vroman, S. (2002). A Matching Model with Endogenous Skill Requirements. *International Economic Review*, 43(1):282–305.

Alchian, A. A. and Allen, W. R. (1983). *Exchange & Production: Competition, Coordination & Control.* Wadsworth Pub, Belmont, CA.

Antràs, P., Garicano, L., and Rossi-Hansberg, E. (2006). Offshoring in a Knowledge Economy. *Quarterly Journal of Economics*, 121(1):31–77.

Atalay, E., Phongthiengtham, P., Sotelo, S., and Tannenbaum, D. (2018). New Technologies and the Labor Market. *Journal of Monetary Economics*, 97:48–67.

Autor, D., Levy, F., and Murnane, R. J. (2003). The Skill Content of Recent

Technological Change: An Empirical Exploration. *Quarterly Journal of Economics*, 118(4):1279–1333.

Autor, D. H. and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5):1553–1597.

Autor, D. H., Dorn, D., and Hanson, G. H. (2015). Untangling Trade and Technology: Evidence from Local Labour Markets. *The Economic Journal*, 125(584):621–646.

Bagger, J., Christensen, B. J., and Mortensen, D. T. (2014). Wage and Labor Productivity Dispersion: The Roles of Total Factor Productivity, Labor Quality, Capital Intensity, and Rent Sharing. In *2014 Meeting Papers*.

Bagger, J. and Lentz, R. (2018). An Empirical Model of Wage Dispersion with Sorting. *Review of Economic Studies, Forthcoming*.

Bagger, J., Sørensen, K. L., and Vejlin, R. (2013). Wage Sorting Trends. *Economics Letters*, 118(1):63–67.

Bahar Baziki, S., Ginja, R., and Borota Milicevic, T. (2015). Trade Competition, Technology and Labor Reallocation. *Working Paper*.

Baumgarten, D. (2015). Offshoring, the Nature of Tasks, and Occupational Stability: Empirical Evidence for Germany. *World Economy*, 38:479–508.

Baumgarten, D., Geishecker, I., and Görg, H. (2013). Offshoring, Tasks, and the Skill-Wage Pattern. *European Economic Review*, 61:132–152.

Becker, G. (1973). A Theory of Marriage. *Journal of Political Economy*, 81(4):813–846.

Becker, S. O., Ekholm, K., and Muendler, M. (2013). Offshoring and the Onshore Composition of Tasks and Skills. *Journal of International Economics*, 90:91–106.

Benfratello, L., Razzolini, T., and Sembenelli, A. (2015). Does ICT Investment Spur or Hamper Offshoring? Empirical Evidence from Microdata. *Working Paper*.

Bernard, A. B., Jensen, J. B., and Schott, P. K. (2006). Survival of the Best Fit: Exposure to Low-Wage Countries and the (uneven) Growth of U.S. Manufacturing Plants. *Journal of International Economics*, 68(1):219–237.

Biscourp, P. and Kramarz, F. (2007). Employment, Skill Structure and International Trade: Firm-level Evidence for France. *Journal of International Economics*, 72:22–51.

Blinder, A. S. (2009). How Many US Jobs Might be Offshorable? *World Economics*, 10(2):41.

Blinder, A. S. and Krueger, A. B. (2013). Alternative Measures of Offshorability: A Survey Approach. *Journal of Labor Economics*, 31(2):s97–s128.

Bloom, N., Draca, M., and Van Reenen, J. (2016). Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity. *Review of Economic Studies*, 83(1):87–117.

Bojilov, R. and Galichon, A. (2016). Matching in Closed-Form: Equilibrium, Identification, and Comparative Statics. *Economic Theory*, 61(4):587–609.

Burstein, A. and Vogel, J. (2010). Globalization, Technology, and the Skill Premium: A Quantitative Analysis. *NBER Working Paper*.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics*, 128(3):967–1015.

Choo, E. and Siow, A. (2006). Who Marries Whom and Why. *Journal of political Economy*, 114(1):175–201.

Costinot, A. and Vogel, J. (2010). Matching and Inequality in the World Economy. *Journal of Political Economy*, 118(4):747–786.

Dahl, C. M., le Maire, D., and Munch, J. R. (2013). Wage Dispersion and Decentralization of Wage Bargaining. *Journal of Labor Economics*, 31(3):501–533.

Davidson, C., Heyman, F., Matusz, S., Sjöholm, F., and Zhu, S. C. (2014). Globalization and Imperfect Labor Market Sorting. *Journal of International Economics*, 94(2):177–194.

Dupuy, A. and Galichon, A. (2014). Personality Traits and the Marriage Market. *Journal of Political Economy*, 122(6):1271–1319.

Dupuy, A., Galichon, A., and Sun, Y. (2017). Estimating Matching Affinity Matrix under Low-Rank Constraints. *IZA Discussion Paper No. 10449*.

Ebenstein, A., Harrison, A., McMillan, M., and Phillips, S. (2014). Estimating the Impract of Trade and Offshoring on American Workers using the Current Population Surveys. *Review of Economics and Statistics*, 96(4):581–595.

Eeckhout, J. and Kircher, P. (2018). Assortative Matching with Large Firms. *Econometrica*, 86(1):85–132.

Faggio, G., Salvanes, K. G., and Van Reenen, J. (2007). The Evolution of Inequality in Productivity and Wages: Panel Data Evidence. *Working Paper*.

Feenstra, R. and Hanson, G. (1996). Foreign Investment, Outsourcing and Relative Wages. *Political Economy of Trade Policy: Essays in Honor of Jagdish Bhagwati*, pages 89–127.

Feenstra, R. and Hanson, G. (1999). The Impact of Outsourcing and High-technology Capital on Wages: Estimates for the United States, 1979-1990. *Quarterly Journal of Economics*, 114(3):907–40.

Feenstra, R. C. and Hanson, G. H. (1997). Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras. *Journal of International Economics*, 42:371–393.

Fort, T. (2017). Technology and Production Fragmentation: Domestic versus Foreign Sourcing. *Review of Economic Studies*, 84:650–687.

Gentzkow, M. and Shapiro, J. M. (2010). What Drives Media Slant? Evidence from U.S. Daily Newspapers. *Econometrica*, 78(1):35–71.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2018). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *NBER Working Papers*.

Goldschmidt, D. and Schmieder, J. F. (2017). The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure. *Quarterly Journal of Economics*, 132(3):1165–1217.

Goos, M., Manning, A., and Salomons, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review*, 104(8):2509–2526.

Grossman, G. and Rossi-Hansberg, E. (2006). The Rise of Offshoring: It's not Wine for Cloth Anymore. *The New Economic Geography: Effects and Policy Implications*.

Grossman, G. M., Helpman, E., and Kircher, P. (2017). Matching, Sorting and the Distributional Effects of Internatinal Trade. *Journal of Political Economy*, 125(1):224–264.

Grossman, G. M. and Maggi, G. (2000). Diversity and Trade. *American Economic Review*, 90(5):1255–1275.

Grossman, G. M. and Rossi-Hansberg, E. (2008). Trading Tasks: A Simple Theory of Offshoring. *American Economic Review*, 98(5):1978–1997.

Grossman, G. M. and Rossi-Hansberg, E. (2012). Task Trade between Similar Countries. *Econometrica*, 80(2):593–629.

Hakanson, C., Lindqvist, E., and Vlachos, J. (2015). Firms and Skills: the Evolution of Worker Sorting. *Discussion paper*.

Handwerker, E. W. (2015). Increased Concentration of Occupations, Outsourcing, and Growing Wage Inequality in the United States. *Working Paper*.

Helpman, E., Itskhoki, O., and Redding, S. (2010). Inequality and Unemployment in a Global Economy. *Econometrica*, 118(4):1239–1283.

Helpman, E., Melitz, M., and Yeaple, S. (2004). Export versus FDI with Heterogeneous Firms. *American Economic Review*, 94(1):300–316.

Hoberg, G. and Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, 124(5):1423–1465.

Hsieh, C. T. and Woo, K. T. (2005). The Impact of Outsourcing to China on Hong Kong's Labor Market. *Quarterly Journal of Economics*, 95(5):1673–1687.

Hummels, D., Jørgensen, R., Munch, J., and Xiang, C. (2014). The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data. *American Economic Review*, 104(6):1597–1629.

Keller, W. and Utar, H. (2016). International Trade and Job Polarization: Evidence at the Worker-level. *Working Paper*.

Kremer, M. and Maskin, E. (1996). Wage Inequality and Segregation by Skill. *Working Paper*.

Kremer, M. and Maskin, E. (2006). *Globalization and Inequality*. Harvard.

Kristofferson, M. S. (2016). Geographical Job Mobility and Wage Flexibility. *Danmarks Nationalbank Monetary Review*, pages 1279–1333.

Lindenlaub, I. (2017). Sorting Multidimensional Types: Theory and Application. *Review of Economic Studies*, 84(2):718–789.

Melitz (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Inudstry Productivity. *Econometrica*, 71(6):1695–1725.

Michaels, G., Rauch, F., and Redding, S. J. (2016). Tasks and Technology in the United States 1880-2000. *NBER Working Papers*.

Monarch, R., Park, J., and Sivadasan, J. (2017). Domestic Gains from Offshoring? Evidence from TAA-lined U.S. Microdata. *Journal of International Economics*, 105:105–173.

Oldenski, L. (2012). The Task Composition of Offshoring by US Multinationals. *International Economics*, 131:5–21.

Olley, S. G. and Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263–1297.

Park, J. (2018). The Cleansing Effect of Offshoring in an Analysis of Employment. *Eastern Economic Journal*, 44:242–272.

Postel-Vinay, F. and Lise, J. (2015). Multidimensional Skills, Sorting, and Human Capital Accumulation. *2015 Meeting Papers 386*.

Ramondo, N. and Rodríguez-Clare, A. (2013). Trade, Multinational Production, and the Gains from Openness. *Journal of Political Economy*, 121(2):273–322.

Sattinger, M. (1993). Assignment Models of the Distribution of Earnings. *Journal of Economic Literature*, 31(2):831–880.

Song, J., Price, D. J., Guvenen, F., Bloom, N., and Wachter, T. v. (2019). Firming up Inequality. *Quarterly Journal of Economics*, 134:1–50.

Zhu, S. and Trefler, D. (2005). Trade and Inequality in Developing Countries: A General Equilibrium Analysis. *Journal of International Economics*, 65(1):21–48.

Chapter 3

Selective Accumulation of Ideas: Accounting for the Decline in Entry Rate

3.1 Introduction

Entrepreneurship and the start of new businesses are very important sources of aggregate growth of an economy. These new businesses are often the most innovative ones at the time of entry, and indeed some of these firms become very successful and become the market leader. As such, the new firms have been a significant source of reallocation in an economy, creating a mass of new and more productive jobs, hence resulting in the so-called "creative destruction."

Recently, there has been a rising concern about this role of entrepreneurship. Numerous studies, including Decker et al. (2014), have documented a decline in the entry rates of new firms in the U.S., along with other signs of decline in the business dynamism. This phenomenon may be universal and probably not confined to the U.S., as more empirical evidence become available in other countries (e.g. Cao et al. (2017) for Canada). At the same time, there has been a separate literature that document a slowdown of the productivity growth in the aggregate economy (e.g., Fernald (2015)).

In general, it is difficult to explain such a decline in the entry rate or a decline in the productivity growth rate using a stationary model, without assuming some ad hoc changes over time in exogenous parameters. For example, one way to sustain a declining entry rate within the framework of a typical stationary model of firm dynamics would be to assume that the fixed cost of entry is increasing over time. But there is little justification for why it would increase over time, and if it does, a subsequent challenge would be to determine the rate of increase of the entry cost.

In this paper, I illustrate a simple mechanism that can explain the declining entry rate. The key mechanism can be summarized as a selective accumulation of the best ideas over time, assumed to be drawn from an exogenous distribution. An idea is like a blueprint for a new product, and the quality of idea is scaled so that it corresponds to the productivity for producing that good. An idea of random quality arrives exogenously to each individual as a Poisson process with a constant arrival rate. Upon observing the quality of the idea, the individual makes an occupational choice of whether to become an entrepreneur and start up a new business by using that idea, or to discard that idea and just keep working for other firms run by existing entrepreneurs.

Suppose that the productivity of each incumbent firm does not change over time. During a short interval of time, a selective entry would occur by those who found sufficiently good ideas. The threshold quality of idea for entry should be at least as high as the exit threshold of the incumbent firms.[1] If no firms exit during this period, the average productivity of all incumbent firms would weakly rise, and the mass of firms (or

[1] This point is trivial, because otherwise the newly entered firm with an idea below the exit threshold would simply exit immediately. There is no point of such an entry.

equivalently the measure of varieties) would strictly increase during this interval. But even if the average productivity remains constant, an increase in the product variety results in a higher welfare and higher real wages, thereby decreasing the profits of all firms while increasing the outside option of working for wages. As a result of this fiercer competition, the exit threshold must strictly rise over time. This selection process implies that, even if the productivity of all firms remain constant, average productivity would rise over time because it is always the least productive firms that would exit. If in addition the incumbent firms' productivity improve over time, this selection effect would be even more pronounced. Likewise, facing this lower profitability while holding the quality of idea constant, the threshold idea that a potential entrant would find indifferent to working for wages would rise over time. This selection mechanism is similar to Sampson (2016).

According to this mechanism, a decline in the entry rate simply reflects the growing maturity of the state of an economy, which is a very natural process. This may offer a more relieving perspective on the recent concerns. For example, a mechanical accounting approach based on a stationary environment may show a very concerning picture of the economy (e.g., Alon et al. (2018)). But according to the model in this paper, we need to take into account that the currently active firms are the result of a selective survival and accumulation of the fittest businesses among the entire set of ideas that has ever arrived to an economy, and this selection process continues to occur indefinitely over time. In a sense, this decline in entry rate can be viewed as a natural consequence of "time" that one needs not worry too much about.

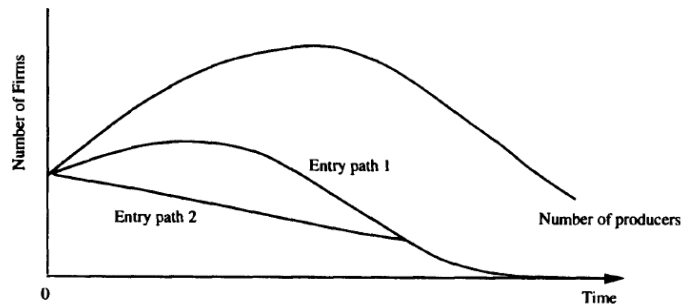
This decline in aggregate entry rate has an analogous counterpart in the industry life cycle. Although it is only recent that the declining entry rate in the aggregate economy started to receive academic attention, a parallel phenomenon at a narrowly defined product level is much more pronounced and well-established, which dates back to at least Gort and Klepper (1982). This old literature on industry life cycle has shown that, in each narrowly-defined industry, there is a salient pattern in the entry-exit dynamics. In particular, entry rates are typically high in the earlier stage of the industry, but as the industry matures over time, only the fittest or the most productive firms survive, and the entry rate declines over time, close to zero in the long run.[2]

Considering only this within-industry perspective, it is only natural that the aggregate entry rate is declining; rather it would be a puzzle that the entry rates are not low enough. One additional dimension that need to be considered in an aggregate economy is that the variety of goods and services may be expanding over time. In addition, the rate of introduction of new industry needs not be constant over time. From the perspective of industry life cycle, this may be one source of time-varying entry rate in the aggregate economy. For example, there was a surge in entry rate around 2000 that resulted in many productive high-tech firms. This is likely a consequence of the IT boom at that time. This kind of industry specific evidence underlines the importance of industry life cycle and the introduction of new industries when explaining the entry rates, either at the industry level or in the aggregate.

In this paper, the baseline model with a single industry is extended to incorporate the expanding set of industries over time. The extended model also accounts for the industry life cycle pattern in terms of the number of firms, growing at first until it hits a peak and declining thereafter, in addition to the declining entry rate within each

[2]This pattern may be difficult to capture in a typical data where the definition of industry is not narrow enough, for example, with 3 digits of industry classification. Any new product introduced at each point in time will be classified as one of the broadly defined existing industries.

Figure 17: Industry life cycle: temporal patterns of entry and number of producers within a typical industry. Excerpt from Klepper (1996)



industry. Figure 1, which is excerpt from Klepper (1996), shows the temporal patterns of entry and number of producers within a typical industry. In terms of the aggregate entry rate, if the new industries arrive at a sufficiently fast rate, a non-declining entry rate can be sustained in the aggregate.

This paper is closely related to several different literatures. First of all, this model directly speaks to the literature that documents declining entry rates, as well as contributes to the subsequent literature that tries to explain this feature. Decker et al. (2014) is among the first papers to document this decline in entry rate. Hopenhayn et al. (2018) and Karahan et al. (2019) attribute the decline in entry rate to the decrease in the population or labor supply growth, while Pugsley and Sahin (2019) also suggests the import penetration as a potential source of declining entry rate in addition to the demographic changes. Akcigit and Ates (2019a), Akcigit and Ates (2019b), on the other hand, argue that decline in the knowledge diffusion can account for the decline in entry rates, along with other signs of declining business dynamism. Kozeniauskas (2018) and Salgado (2019) explain the decline in aggregate entry rate, together with a disproportionately more decline in high skill individual's entry, by fundamental structural changes in the economy, including skill-biased technical change, rise in entry cost, or the cost of capital. The current paper contributes to this growing literature that aims to explain the decline in aggregate entry rate, showing that it may occur even without any exogenous structural changes in the economy.

In addition, there are at least two distinct literatures that are potentially related to the declining entry rates: the literature that examines the slowdown of productivity, and the one that studies the rising cost of R&D over time. A set of recent studies, including Fernald (2015), Byrne et al. (2016), Cette et al. (2016), and Syverson (2017), all document a slowdown of productivity growth. This literature mainly focuses on the U.S. economy, but as shown by Cette et al., this slowdown in productivity is also observed in a set of developed countries in Europe. This set of papers carefully takes into account the possibility of mismeasurement of productivity, and still concludes that there has been a significant slowdown in productivity growth. Alon et al. (2018) shows that this slowdown in productivity may be related to the decline in entry rate. Although this paper maintains an agnostic stance about the evolution of the productivity of the incumbent firms, the mechanism of selective entry is able to explain at least a part of the slowdown of productivity growth. In this paper, the slowdown of productivity is intrinsically related to the decline in entry rate as in Alon et al., and both of these trends are natural consequences of the growing maturity of an industry or the aggregate economy as a whole.

Even if the productivity growth were maintained at a constant rate, it is sustained by an increasingly higher resource dedicated to R&D activities over time, suggesting decreasing returns in R&D. For example, Jones (1995) and Kortum (1997) find that the number of scientists and engineers has increased rapidly while the TFP growth or other measures of research output have remained at nearly constant levels, and build models that can justify this finding. Jones (2009) makes a related point that it becomes increasingly more costly over time for the innovators to obtain the required depth of knowledge. Bloom et al. (2020) shows a variety of recent evidence that the ideas are getting harder to find over time, which is consistent with the argument made by Jones (1995). The key mechanism of the current model, selective accumulation of ideas, is also closely related to this line of literature, as it shows in almost a trivial manner why it would become harder to find good ideas over time.

There is also a small but growing literature that studies the changes in entry cost over time. For example, Bollard et al. (2016) argues that entry costs rise with development. This argument is based on the assumption of free entry, which interprets the average discounted profits as the cost of entry. This finding can be re-interpreted in the context of the current model as the outside option for the potential entrant, which is the wage income, rising over time.

This paper is also related to the vast literature on economic growth that models ideas as the source of growth. For example, in **?**, R&D leads to an increasing set of varieties, whereas in Grossman and Helpman (1991), successful R&D results in a superior blueprint that allowed the firm to capture the entire market under perfect substitutability across goods. Klette and Kortum (2004) extends the Grossman and Helpman's model to the environment with a continuum of products (product lines), which spurred a large literature that links firm dynamics, innovation, and creative destruction.[3] More recently, Perla and Tonetti (2014) and Sampson (2016) studied a new class of endogenous growth

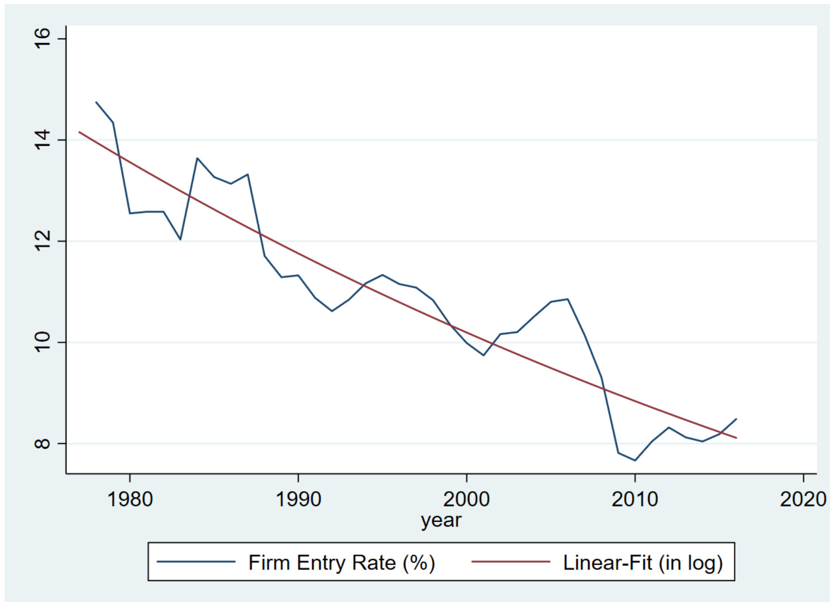
[3]See for example Aghion et al. (2014) and the reference therein.

models, under the assumption that the pool of current incumbent firms serves as the distribution from which a potential entrant draws the productivity. An externality arises because becoming a high productivity firm benefits the potential entrants by improving the expected productivity upon entry. The current paper does not assume that the ideas are drawn from the distribution of the current incumbent firms, which is an important distinction, but otherwise closely related to these models in terms of equilibrium mechanism.

Lastly, this paper also contributes the older literature on industry life cycle. Gort and Klepper (1982) and Agarwal and Gort (1996), among many others, empirically studied the patterns of industry life cycle. Hopenhayn (1993), Jovanovic and MacDonald (1994) and Klepper (1996) are examples of the papers that build theoretical models to explain the patterns of industry life cycle. The current paper provides a new perspective as well as a very simple alternative mechanism that can explain the declining entry rate as an industry matures, which is an important part of the industry life cycle.

The rest of the paper is organized as follows. In section 2, I show some evidence of decline in entry rates. In section 3, I present the model of a single industry in two steps. First, I describe and solve for an equilibrium under static environment. Second, I show what would be different under dynamic environment and how it may differ from the static equilibrium. Then I show that, by introducing a minor assumption that does not alter the aggregate properties, the equilibrium condition under dynamic environment exactly coincides with that under static environment. In section 4, I extend the model to an environment with a growing set of industries. In section 5, I use the data and calibrate the model. Section 6 concludes.

Figure 18: Time series evidence of declining entry rate of firms in the U.S. The point estimate of the linear fit coefficient (in log) is -0.0143, or 1.43% decline per year. (Data source: Business Dynamics Statistics)



3.2 Evidence of Decline in Entry Rates

Figure 2 shows the time series evidence of declining entry rate of firms in the U.S. As evidenced by several existing empirical works, there has been a clear downward trend in the aggregate entry rate, which appears together with fluctuations at the business cycle frequency. The linear fit in logs[4] shows that the average rate of decline in the entry rate is 1.43% per year, which would accumulate to 43.7% decline over the span of 40 years. The objective of this paper is to explain this steady yet pronounced decline in entry rate.

[4]Taking logs is appropriate because entry rate cannot fall below zero.

3.3 A Simple Model of Industry Equilibrium

3.3.1 A Static Equilibrium

The economy is populated with a continuum of individuals with mass equal to $\overline{L}$. Each individual is endowed with the same amount of skill, normalized to 1,[5] to be used when working for wages, and an idea of quality a that may differ across individuals. The unit of the quality of ideas is scaled such that when an idea of quality a is implemented to production by the entrepreneur, the productivity of that firm is equal to a.

Each individual would earn a wage of w upon choosing to work for wages.[6] The potential profit to the entrepreneur upon a business start-up with an idea of quality a is $\pi(a)$, deterministic and known ex ante. As it turns out later, this profit function also implicitly depends on the aggregate state of the economy. The density of a over the entire population is $h(a)$, where $\int h(a)da = \overline{L}$. There is no disutility of supplying labor, so there is always full employment, either as a worker or as an entrepreneur.

An individual with an idea a will choose to become an entrepreneur if $\pi(a) \geq w$, and vice versa.[7] As long as higher productivity firms earn higher profits, there will be a cutoff threshold for the decision rule: to become entrepreneur if $a > a^*$ and become a wage worker if $a < a^*$.[8] Each entrepreneur corresponds to each firm, which makes optimal decisions regarding price and labor demand. The mass of firms is thus $M =$

[5]It is straightforward to extend to the case with heterogeneous skills, but it is not implemented here because it is not the focus of this paper. This feature is implemented in the other paper in preparation.

[6]Because this is a real model, the nominal wage can be rescaled using any numeraire. For example, it is innocuous to set $w = 1$.

[7]It is implicitly assumed that the entrepreneur does not contribute to the labor input, so the production involves the combination of an idea of the entrepreneur and labor input of workers.

[8]Note that in this static environment, there is no transition between entrepreneur and worker. Thus the notion of entry cost is absent, and there is no "hysteresis" in terms of the difference between entry and exit threshold.

$\int h(a)1(a \geq a^*)da$, and the mass of workers is $L = \overline{L} - M$. The normalized density of firms (as a probability density function) can be expressed as $\frac{1}{M}h(a)1(a \geq a^*)$.

The economic environment is standard. The preference of all consumers are identical, and exhibits constant elasticity of substitution across all existing varieties:

$$\mathbb{C} = \Big(\int_{\Omega} c(j)^{\frac{\epsilon-1}{\epsilon}} dj \Big)^{\frac{\epsilon}{\epsilon-1}}$$

where j stands for each variety, $c(j)$ is the quantity of consumption of variety j, and $\epsilon > 1$ is the elasticity of substitution across different varieties. Each firm produces a single variety j. The demand for all varieties are symmetric.[9]

The technology for production requires a single type of input, labor, and exhibits constant returns to scale. The production function is then simply $y(j) = a(j)l(j)$, where $a(j)$ is the productivity of firm j, and $l(j)$ is the labor demand. Each firm sets its optimal monopoly price, taking the consumer's demand function as given. The demand function for firm j is

$$c(j) = \Big(\frac{p(j)}{\mathbb{P}}\Big)^{-\epsilon} \mathbb{C} = y(j)$$

where $p(j)$ is the price set by the firm j, $\mathbb{P} = \big(\int_{\Omega} p(j)^{1-\epsilon} dj \big)^{\frac{1}{1-\epsilon}}$ is the aggregate price index, and $\mathbb{C}$ is the aggregate consumption. Since all the firms with a same productivity behave exactly the same way, we can index each firm by the productivity a instead of the variety j. The profit of a firm with productivity a is

$$\pi(a) = p(a)y(a) - wl(a) = \Big(p(a) - \frac{w}{a}\Big)y(a)$$

[9]The preference for each variety may be scaled by constant scalars, indicating demand differences across varieties that can be interpreted as "quality" in addition to the "productivity" of each product. It is well known that, in this type of model with CES demand, productivity and quality are exactly isomorphic, and any existing difference in quality can be subsumed into productivity to yield exactly the same equilibrium outcome.

Because each firm is infinitesimally small and cannot affect any of the aggregate variables, profit maximizing choice of price is simply a constant markup over the marginal cost:

$$p(a) = \frac{\epsilon}{\epsilon - 1}\frac{w}{a}$$

Consequently, all the quantities are determined using this optimal choice of price:

$$\begin{aligned}
y(a) &= \left(\frac{p(a)}{\mathbb{P}}\right)^{-\epsilon}\mathbb{C} = \left(\frac{\epsilon}{\epsilon-1}w\right)^{-\epsilon} a^{\epsilon}\mathbb{P}^{\epsilon}\mathbb{C} \\
l(a) &= \frac{y(a)}{a} = \left(\frac{\epsilon}{\epsilon-1}w\right)^{-\epsilon} a^{\epsilon-1}\mathbb{P}^{\epsilon}\mathbb{C} \\
r(a) &= p(a)y(a) = \left(\frac{\epsilon}{\epsilon-1}w\right)^{1-\epsilon} a^{\epsilon-1}\mathbb{P}^{\epsilon}\mathbb{C} \\
\pi(a) &= \frac{1}{\epsilon}r(a) = \frac{w}{\epsilon-1}\left(\frac{\epsilon}{\epsilon-1}w\right)^{-\epsilon} a^{\epsilon-1}\mathbb{P}^{\epsilon}\mathbb{C}
\end{aligned}$$

Thus a firm's profit, revenue, and labor demand are all proportional to $a^{\epsilon-1}$. As noted by Melitz (2003),[10] regardless of the underlying distribution of productivity, all aggregate variables can be characterized in a simple way using the "average productivity" of all firms currently in operation:

$$\mathbb{A} \equiv \left(\int a^{\epsilon-1}\mu(a)da\right)^{\frac{1}{\epsilon-1}} = \left(\frac{1}{M}\int a^{\epsilon-1}h(a)da\right)^{\frac{1}{\epsilon-1}} = M^{\frac{1}{1-\epsilon}}\left(\int a^{\epsilon-1}\right)^{\frac{1}{\epsilon-1}}$$

$$\therefore \left(\int a^{\epsilon-1}\right)^{\frac{1}{\epsilon-1}} = M^{\frac{1}{\epsilon-1}}\mathbb{A}$$

Using this expression, the aggregate price can be rewritten as

$$\mathbb{P} = \left(\int p(a)^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}} = \left(\frac{\epsilon}{\epsilon-1}w\right)\left(\int a^{\epsilon-1}\right)^{\frac{1}{1-\epsilon}} = \left(\frac{\epsilon}{\epsilon-1}w\right)\left(M^{\frac{1}{\epsilon-1}}\mathbb{A}\right)^{-1}$$

The dependence of aggregate variables on the variety of goods M, or equivalently the mass of firms, is an important property in this class of models, which has been well known

[10]The current economic environment is very similar to the closed economy version of Melitz (2003). A minor difference would be the absence of fixed cost. Melitz used fixed cost of operation to induce low productivity firms to exit. Here, each individual's outside option as a wage worker serves as a natural fixed cost that leads to optimal exit decisions.

at least since Dixit and Stiglitz (1977) and re-emphasized in several studies including Melitz (2003). Holding the average productivity $\mathbb{A}$ constant, a larger mass of firms M leads to a lower aggregate price by a power of $\frac{1}{\epsilon-1}$, which in turn implies a higher real wage $w/\mathbb{P}$. Holding the aggregate supply of labor constant, which would result in the same aggregate revenue, the aggregate welfare would increase by a factor of $M^{1/(\epsilon-1)}$.

Now it remains to determine the equilibrium threshold quality of ideas, a^*. This can be done using the indifference condition at $a = a^*$:

$$\pi(a^*) = \frac{w}{\epsilon-1}\Big(\frac{\epsilon}{\epsilon-1}w\Big)^{-\epsilon}(a^*)^{\epsilon-1}\mathbb{P}^{\epsilon}\mathbb{C} = \frac{w}{\epsilon-1}\big(M^{\frac{1}{\epsilon-1}}\mathbb{A}\big)^{-\epsilon}(a^*)^{\epsilon-1}\mathbb{C} = w$$

The labor demanded by this threshold firm satisfies

$$l(a^*) = \frac{\epsilon-1}{w}\pi(a^*) = \epsilon - 1$$

Using this, the aggregate labor demand should satisfy

$$L = \int l(a) = \int l(a^*)\Big(\frac{a}{a^*}\Big)^{\epsilon-1} = \frac{l(a^*)}{(a^*)^{\epsilon-1}}\int a^{\epsilon-1} = \frac{l(a^*)}{(a^*)^{\epsilon-1}}M\mathbb{A}^{\epsilon-1}$$

Define $\mu(a^*) \equiv A/a^*$ as the ratio between average productivity of the currently surviving firms (i.e. conditional on $a \geq a^*$) and the cutoff threshold a^*. By construction, this ratio is independent of M and depends only on the shape of the distribution $h(a)$ and the threshold a^*. Then

$$L = (\epsilon-1)M\mu(a^*)^{\epsilon-1}$$

Finally, the labor market should be cleared:

$$\overline{L} = M + L = \Big(1 + (\epsilon-1)\mu(a^*)^{\epsilon-1}\Big)M(a^*)$$

Clearly, $M'(a^*) = -h(a^*) < 0$. In addition, for a broad class of distributions, $\mu'(a^*) \leq$

0.[11] As a result, the right hand side is a strictly monotonically decreasing function of a^* on the support of a^*. In addition, it takes the value of zero when $a^* \to \infty$, and certainly greater than $\overline{L}$ when $a^* \to 0$. So there is always a unique equilibrium in this model.

Pareto Distribution

The model described above has a well defined unique equilibrium for an arbitrary distribution of ideas or productivity. However, in almost all subsequent analyses, I will focus exclusively on the case where the distribution follows a Pareto distribution. There are at least three reason for making this specific choice. First, the equilibrium will always involve truncations of the distribution of ideas. Pareto distribution has a very nice property that any truncation of a Pareto distribution is still (or has a density proportional to) a Pareto distribution. Second, by choosing the Pareto distribution, $\mu(a^*)$ becomes a fixed constant. This means that the mass of firms becomes invariant regardless of the industry life cycle, or the accumulation of ideas over time. Since $\mu(a^*)$ depends only on the distributional assumptions, the equation (??) implies that the evolution of the mass of firms over the industry life cycle would entirely depend on the choice of distribution. Because it is ex ante not clear whether the mass of firms should increase or decrease over the industry life cycle, it seems desirable to take such an agnostic stance for this variable that does have significant impact on welfare. Third, the Pareto distribution is also capable of generating analytically tractable results, which is very useful.

The distribution of ideas follows a Pareto distribution with the shape parameter θ that is strictly greater than $(\epsilon - 1)$. Instead of specifying the cutoff threshold, which

[11]Typically, thick tailed distributions (slowly decaying density) have a better chance of maintaining an increasing $\mu(a^*)$. However, Pareto distribution, whose tail is quite thick, implies a constant $\mu(a^*)$, and distributions with tails thinner than Pareto would in general satisfy $\mu'(a^*) \leq 0$.

should be determined endogenously, I specify the value of the density function evaluated at $a = 1$ (extrapolated if not in the support), which is an arbitrary point chosen without loss of generality, and denote this parameter h. That is,

$$h(a) \equiv h a^{-\theta-1} \cdot \mathbb{1}(a \geq a_0)$$

where a_0 would specify the point where $\int_{a_0}^{\infty} h(a)da = \overline{L}$. Because there will always be individuals who work for wages whose ideas do not matter, this lower bound a_0 bears little importance when h is given. When discussing the evolution of an industry in later sections, h will be the key time-varying object that summarizes the maturity of an industry.

Under this distributional assumption,

$$M = \int_{a^*}^{\infty} h a^{-\theta-1} = \frac{h}{\theta}(a^*)^{-\theta}$$

$$\mathbb{A}^{\epsilon-1} = \int_{a^*}^{\infty} a^{\epsilon-1} \cdot \theta (a^*)^{\theta} a^{-\theta-1} da = \frac{\theta}{\theta - (\epsilon - 1)} (a^*)^{\epsilon-1}$$

It becomes obvious that, in order to have a finite value of $\mathbb{A}$, a key variable throughout the analysis, the distribution should be sufficiently thin-tailled: $\theta > \epsilon - 1$. In addition, since $\mathbb{A}$ uses the normalized probability density rather than the density function $h(\cdot)$ itself, it does not depend on the value of h. Other related aggregate objects can be expressed as:

$$\mu(a^*) = \frac{\mathbb{A}}{a^*} = \Big(\frac{\theta}{\theta - (\epsilon - 1)}\Big)^{\frac{1}{\epsilon-1}}$$

$$\int_{a^*}^{\infty} a^{\epsilon-1} = M\mathbb{A}^{\epsilon-1} = \frac{h}{\theta}(a^*)^{-\theta} \frac{\theta}{\theta - (\epsilon - 1)} (a^*)^{\epsilon-1} = \frac{h}{\theta - (\epsilon - 1)} (a^*)^{-(\theta-(\epsilon-1))}$$

$$\mathbb{P} = \Big(\frac{\epsilon}{\epsilon - 1} w\Big)\Big(M^{\frac{1}{\epsilon-1}} \mathbb{A}\Big)^{-1} = \Big(\frac{\epsilon}{\epsilon - 1} w\Big)\Big(\frac{h}{\theta - (\epsilon - 1)}\Big)^{\frac{1}{1-\epsilon}} (a^*)^{\frac{\theta-(\epsilon-1)}{\epsilon-1}}$$

Determination of a^* is straightforward, following the procedure shown in the previous section:

$$L(a^*) = \frac{l(a)}{a^{\epsilon-1}} M\mathbb{A}^{\epsilon-1} = (\epsilon-1)M\mu(a^*)^{\epsilon-1} = (\epsilon-1)\Big(\frac{h}{\theta}(a^*)^{-\theta}\Big)\frac{\theta}{\theta-(\epsilon-1)}$$

$$\overline{L} = M + L = \frac{h}{\theta-(\epsilon-1)}\Big(\frac{\theta-(\epsilon-1)}{\theta} + (\epsilon-1)\Big)(a^*)^{-\theta} = \frac{h}{\theta}\ \frac{\epsilon\theta-(\epsilon-1)}{\theta-(\epsilon-1)}(a^*)^{-\theta}$$

$$\therefore a^* = \Big(\frac{h}{\theta}\ \frac{\epsilon\theta-(\epsilon-1)}{\theta-(\epsilon-1)}\Big)^{\frac{1}{\theta}}\overline{L}^{-\frac{1}{\theta}}$$

Consequently,

$$\frac{\epsilon}{\epsilon-1}\frac{w}{\mathbb{P}} = \frac{\mathbb{C}}{L} = M^{\frac{1}{\epsilon-1}}\mathbb{A} = \Big(\frac{h}{\theta-(\epsilon-1)}\Big)^{\frac{1}{\theta}}\Big(\frac{\theta}{\epsilon\theta-(\epsilon-1)}\overline{L}\Big)^{\frac{\theta-(\epsilon-1)}{\theta(\epsilon-1)}}$$

In this environment, the only exogenous objects are h which captures the maturity of an industry, and the total population $\overline{L}$. All other aggregate quantities including $\mathbb{C}$ and $\mathbb{P}$ are endogenous. The real wage increases in the maturity of the industry, h, as well as in the population $\overline{L}$. The former is natural. The latter is due to the fact that a larger population leads to a greater variety of goods, leading to a higher welfare and higher real wages. This is reminiscent of Jones (1995), where population growth sustains a long run growth because more ideas are available. Also note that a^* decreases in $\overline{L}$ when holding h constant, meaning that average productivity would be lower if $\overline{L}$ is larger. This shows that, with a larger population, the benefit from greater varieties M would exceed the cost from lower average productivity $\mathbb{A}$.

Finally, note that

$$M = \frac{h}{\theta}(a^*)^{-\theta} = \frac{\theta-(\epsilon-1)}{\epsilon\theta-(\epsilon-1)}\overline{L}$$

$$L = \overline{L} - M = \frac{\theta(\epsilon-1)}{\epsilon\theta-(\epsilon-1)}\overline{L}$$

So the share of entrepreneurs and the share of workers are constant, which is a feature of the Pareto distribution.

3.3.2 Dynamic Model

In the dynamic version of the model, an individual considers not only the current income but also the expected future income. The productivity $a(j)$ of an incumbent firm j is assumed to stay at a constant level,[12] identical to the quality of idea that it started with. Over time, as new entrants enter, the least productive incumbent firms will be driven out as their profits fall below their reservation wages, and the cutoff threshold as well as average productivity would rise over time.[13]

Time is continuous. An individual k's expected lifetime utility at time t is given as

$$V^k(t) = E_t \int_0^\infty e^{-\rho\tau} c^k(t+\tau) d\tau$$

where ρ is the discount factor, and $c^k(t+\tau)$ is the consumption for the individual k at time $t+\tau$. Intertemporal consumption smoothing is not allowed, and the only choice one can make is the occupation choice between wage worker and entrepreneur, that is, whether to take or discard the given idea randomly drawn from the distribution of ideas, upon arrival of an idea that occurs at a constant rate. In addition, there is no uncertainty except for receiving the idea shock, and all the aggregate variables follow the predetermined paths which all individuals correctly foresee.

[12] This is clearly an unrealistic assumption, but again this is to illustrate the evolution of an industry, in particular, an endogenous rise in the cutoff threshold, that can be sustained even in an absence of any randomness. From the individual's choice of entry, adding uncertainty would act to lower the value of entry, and still there will be a threshold level of idea that the individual finds indifferent between entering and not.

[13] One natural consequence is that, the size and profitability of each firm are at its peak at the time of entry, and then monotonically decline over time. This is seemingly not consistent with the evolution of firm's characteristics in the real world, where the large firms typically start small at the time of entry. However, if one considers the "average" growth of firms, taking into account the exits due to selection, the incumbent firms are necessarily shrinking on average as long as there is a positive mass of new firms at each point in time. Absent any idiosyncratic shocks, it is not only natural but also necessary that all firms are shrinking over time. In fact, this feature is very similar to Caballero and Hammour (1994), which presents a vintage model of creative destruction. In that paper, older firms are monotonically less productive because the production units cannot improve over time once they are built, while the technology advances over time.

All wage workers consume identical quantity of aggregate composite good $c^W(t) = \frac{w}{\mathbb{P}(t)}$. Recall that the share of aggregate consumption that goes to the workers is $\frac{\epsilon-1}{\epsilon}\mathbb{C}(t) = \frac{w}{\mathbb{P}(t)}L(t)$, and this is equally distributed among the measure $L(t)$ of workers. The rest of aggregate consumption $\frac{1}{\epsilon}\mathbb{C}(t)$ is consumed by the entrepreneurs. An entrepreneur with productivity a consumes flow profit period by period: $c^E(a,t) = \frac{\pi(a,t)}{\mathbb{P}(t)}$. Note that unlike the static case, in this dynamic environment, it is possible that $\pi(a^*(t),t) \neq w$, depending on the assumptions on the arrival of ideas for the entrepreneurs and workers, where this threshold a^* is time-varying. Moreover, the entry and exit threshold may be different from each other.

First, consider the individuals who are currently working for wages. At each time t, a small fraction of workers would find an idea of random quality, and decide whether to start up a business using that idea. Because an idea of higher quality leads to a higher profit, there will be a cutoff threshold $a_E^*(t)$, such that an idea with $a \geq a_E^*(t)$ would be immediately realized as new businesses,[14] and the ones with $a < a_E^*(t)$ would be discarded. This threshold $a_E^*(t)$ depends on the current state of the aggregate economy. The value function for the wage worker thus satisfies

$$\rho V^W(t) = \frac{w}{\mathbb{P}(t)} + \widetilde{\eta}_t E\big[\max\big(0, V^E(\widetilde{a}_t, t) - X_E - V^W(t)\big)\big] + \dot{V}^W(t)$$

where $V^W(t)$ is the value of being a worker at time t, and $V^E(a,t)$ is the value of being an entrepreneur with productivity a at time t, and X_E is the entry cost. $\widetilde{\eta}_t$ is the rate of arrival of an idea at t, and the probability distribution of ideas is given as $\widetilde{h}_t(a)$. The tilde over a_t indicates that it is a random variable, and the expectation is taken over this variable. If the support of this distribution includes $a_E^*(t)$, $\widetilde{\eta}_t$ would include the arrival of

[14] Obviously, postponing entry will not be optimal.

inferior ideas below $a_E^*(t)$. The max operator indicates that the worker would disregard such an inferior idea and just keep working for wages. Considering only the ideas that are worth entering, the observed entry rate will be $\eta_t = \widetilde{\eta}_t P(\widetilde{a}_t \geq a_E^*(t))$. Likewise, the entry distribution can be redifined as $h_t(a) = \frac{\widetilde{h}_t(a)1(\widetilde{a}_t \geq a_E^*(t))}{P(\widetilde{a}_t \geq a_E^*(t))}$, which is simply the conditional distribution. Then the worker's value function becomes

$$\rho V^W(t) = \frac{w}{\mathbb{P}(t)} + \eta_t E\big[V^E(\widetilde{a}_t, t) - X_E - V^W(t))\big] + \dot{V}^W(t)$$

The entrepreneur's problem is described as follows. Suppose that the firms' productivity levels are permanent and there are no shocks to the productivity. Because entry occurs only selectively above the current threshold, the average productivity rises over time. As a result, aggregate price would fall over time, and the profit of each firm decline over time. An entrepreneur chooses its optimal stopping time based on the following problem:

$$V^E(a, t) = \max_{T \geq t} \left[\int_0^{T-t} e^{-\rho\tau} \frac{\pi(a, t+\tau)}{\mathbb{P}(t+\tau)} d\tau + e^{-\rho(T-t)} V^W(T) \right]$$

where T denotes the optimally chosen exit time. The entrepreneur consumes $\frac{\pi(a,t+\tau)}{\mathbb{P}(t+\tau)}$ between time t and T, and then exits and becomes a worker. Let $T(a)$ denote the optimal exit time for an entrepreneur with productivity a.[15] The inverse of the mapping $T(\cdot)$ can be defined as the exit threshold at each point in time: $a_X^*(t) \equiv T^{-1}(t)$, so that $T(a_X^*(t)) = t$. If the current profit is already too low, the firm will immediately exit at $T = t$, which would happen whenever it is already past the optimal exit time ($t \geq T(a)$). The value function can be written as $\rho V^E(a, t) = \frac{\pi(a,t)}{\mathbb{P}(t)} + \dot{V}^E(a, t)$ for $t \leq T(a)$ and $\rho V^E(a, t) = \rho V^W(t)$ for $t \geq T(a)$.

[15] Regardless of the date of entry, all entrepreneurs with the same productivity a will exit at the same date $T(a)$.

Suppose in addition that the entrepreneurs also face new ideas $\widehat{a}_t$ drawn from the distribution $\widehat{h}_t(a)$ with the same support as $h_t(a)$ (i.e. $\widehat{a}_t \geq a_t^*$) at rate δ_t, and that their productivity immediately change to the new productivity without the choice of keeping the old productivity even if the new productivity is lower than the previous one ($\widehat{a}_t < a$). Then the value function satisfies

$$\rho V^E(a,t) = \frac{\pi(a,t)}{\mathbb{P}(t)} + \delta_t\big(E[V^E(\widehat{a}_t,t)] - V^E(a,t)\big) + \dot{V}^E(a,t)$$

for $t \leq T(a)$. Note that, by construction, $V^E(a,t) \geq V^W(a), \forall t$, because all entrepreneurs have an option to exit at any point in time, and this inequality holds with equality for $t \geq T(a)$, when it is optimal to just exit right away and become a wage worker.

Existence of positive entry cost, $X_E > 0$, would give rise to an interesting dynamics in this environment. In particular, it would result in a "hysteresis," where the entry threshold is strictly higher than the exit threshold. That is, operating firms just above the exit threshold remain in business because they enjoy a higher value compared to being a worker; however, a potential entrant may choose not to enter if it falls below the value of being a worker after subtracting the entry cost. But apart from this hysteresis, the entry cost does not play any role in this environment. Given that this hysteresis is not the focus of this paper, I assume $X_E = 0$ in all subsequent analysis. As a result, the entry threshold would always coincide with the exit threshold, i.e., $a_E^*(t) = a_X^*(t) \equiv a^*(t)$.

In addition, the threshold rule becomes particularly simple if the arrival rate and the distribution of productivity draw are the same for both workers and entrepreneurs, i.e., $\delta_t = \eta_t$ and $\widehat{h}_t(a) = h_t(a)$. Under these assumptions, combining the two value function equations yields

$$(r+\eta_t)\big[V^E(a,t)-V^W(t)\big]=\frac{\pi(a,t)-w}{\mathbb{P}(t)}+\big[\dot{V}^E(a,t)-\dot{V}^W(t)\big]$$

In particular, at $t=T(a)$, $\pi(a,T(a))=w$ holds for all a, since $V^E(a,T(a))=V^W(T(a))$ and $\dot{V}^E(a,T(a))=\dot{V}^W(T(a))$.[16] Note that this result is equivalent to $\pi(a^*(t),t)=w,\forall t$, because the optimal exit time $T(a)$ is simply an inverse of the exit threshold function $a^*(t)$.

This is a very useful result because the solution to the occupation choice problem in this dynamic environment becomes identical to the static problem at each point in time. This result is intuitive in the sense that the option value of new ideas for wage workers is exactly the same as the option value for incumbent firms. For example, if the entrepreneurs are not allowed to get new ideas while in operation, which may be captured by setting $\delta_t\equiv 0$, the entrepreneurs that currently earn profits strictly above the wage would optimally choose to exit before reaching the exact static threshold $\pi(a^*,t)=w$ in order to start getting better ideas for the expected future gains at the expense of current flow losses. While this is what must happen in that version of the dynamic model, whether the exact threshold should be above or below the static outside wage option is not the question of interest here. In addition, it is hard to take a strong stance on a particular process of idea arrival, for example based on an empirical observation, especially when the precise comparison between that of the workers and entrepreneurs is the issue. On the other hand, because the firms face only the distribution above the threshold $a^*(t)$, that is, without being driven to exit, this is simply a stochastic reshuffling of productivity among the incumbent firms. And with a continuum of, i.e.,

[16] These conditions are a version of what is sometimes called "value matching" and "smooth pasting" respectively. See e.g. Stokey (2008).

uncountably many firms, the aggregate distribution of productivity remains unaltered with this reshuffling process.

Modifying or dispensing with these potentially unaesthetic assumptions on the arrival of ideas for the entrepreneurs does not qualitatively change any of the intuitions or mechanisms of the model, but only under this set of knife-edge assumptions is the solution exactly the same as the static version of the model. The sufficient conditions proposed here is to ensure that the new value function after making the transition upon an arrival of new idea is exactly the same between the worker and any entrepreneur so that it can be nicely cancelled out. One of the potentially undesirable consequence of this "trick" is that an entrepreneur cannot maintain its previous idea even if the new idea is inferior to the previous one. This is a small cost paid for getting a fully analytic solution.

Discussion on the Entry Assumption

The assumptions made on entry process in this paper are different from those made in the majority of previous works. A more common approach is to assume that an entry cost must be paid in full before drawing the productivity, and the free entry condition ensures that the entry cost is the same as expected discounted value upon entry, taking into account the probability of having to exit immediately upon a bad draw. It seems that this assumption is perceived as an economically appealing one, and has been used in the majority of models of firm dynamics as well as endogenous growth. The current paper assumes that ideas arrive exogenously, so it stands in contrast to the free entry assumption. However, this approach with an exogenous pool of potential entrants, which is probably simpler than with free entry condition, is of course not new, and has been adopted in several papers including Chaney (2008) and Monte (2011). The static version of the current model differs from the Chaney's entry assumption only in the sense that more entry due to lower threshold would lead to a reduction in labor supply by the same amount.

Recently, modified versions of entry process have also been used. For example, Lee and Mukoyama (2018) employs a two-step entry procedure, where a potential entrant pays a cost and get an idea, then after observing an idea, decides whether to pay an additional cost to implement the idea. The current paper can be viewed as a limiting case of that modified entry process, where getting an idea does not require any resource in the first stage, so that ideas may arrive while working as a wage worker. In addition, there is no additional fixed cost of entry that would correspond to the second stage, which is a relatively minor modification introduced to avoid hysteresis.

If we consider the real world, mediocre ideas may come with little or no cost, but good ideas often do require high costs. Consider a high-tech firm that requires a profound Ph.D. level knowledge of the founder. The founder takes a Ph.D. education, finds an idea and develops a blueprint, and launches a business based on that blueprint. But another plausible way of interpreting this process is that the founder had a hint of the idea before taking this education, and chose to take the education in order to develop and implement this idea. This story would be observationally equivalent to the other possibility where the idea was obtained only after taking the Ph.D. education. And this would be the suitable interpretation of exogenous arrival of ideas without incurring any cost. Again, after observing an idea, an entry cost — taking the Ph.D. education in this example — can be introduced to the model, at the cost of added complexity that arises from the entry threshold being higher than the exit threshold.

The current model has both similarities and differences with respect to the recent growth models based on technology diffusion or immitation (e.g. Perla and Tonetti (2014) and Sampson (2016)). While this literature assumes that new ideas are drawn from the current incumbent's distribution, in the current paper, new ideas are drawn from an exogenous distribution that may or may not evolve proportionally with the incumbent's distribution. Although this literature makes important theoretical contributions while sustaining a nice balanced growth path, the entry assumptions made in this literature are typically not supported by the data. For example, incumbent firms typically do have certain advantages over entrants, based on technological advances from R&D activities that are not fully disclosed as public knowledge, as well as building up of their own customer base. The declining entry rate would be one feature of the data that these models cannot explain. The current model can be viewed as a generalization of this literature where the entry distribution may evolve over time but not necessarily tied to the existing incumbent firms' productivity distribution. Thus it is capable of sustaining non-stationarity results as observed in the data, while at the same time it can still sustain the results of the literature by making specific assumptions on how this entry distribution evolves over time.

Another feature of the current model is that it ignores the stochastically time-varying aspect of the firms' productivity. This feature is also similar to Perla and Tonetti (2014) and Sampson (2016), but there is even less uncertainty here because the entry decision is made after observing the productivity with certainty. Like these papers, ignoring some of these uncertainties makes the analytic solution simpler, but introducing additional level of uncertainty would not alter any of the qualitative properties of the equilibrium.

Example: Time-Invariant Distribution of Ideas

Suppose the economy starts at $t = 0$, when the arrival of ideas starts taking place. The ideas are drawn from a time-invariant Pareto distribution with shape parameter θ and lower bound a_0:

$$h(a) = \theta a_0^{\theta} a^{-\theta-1}$$

Each individual has a Poisson arrival rate of η of getting an idea randomly drawn from this distribution. During a short interval of dt, $\eta L\ dt$ is the total measure of ideas found in this economy. The entrepreneurs also find new ideas, but because of the "reshuffling," arrival of new ideas for the entrepreneurs have no aggregate effect, and hence the entry depends on L and not $\overline{L} = M + L$.

First, assume that at $t = 0$, the economy is already populated by the equilibrium mass M(0) of firms with the productivity distribution identical to the distribution of ideas, as in the static economy discribed above. Then new ideas from the same distribution would arrive continuously in all $t \geq 0$. The unnormalized density of firms is

$$g_0(a) = Mh(a) = \frac{\theta - (\epsilon - 1)}{\epsilon\theta - (\epsilon - 1)} \overline{L} \cdot \theta a_0^{\theta} a^{-\theta-1}$$

where $\int g_0(a) = M$ and $a^* = a_0$.

At $t > 0$, the "hypothetical" density of firms, including entry of new firms but before considering exits, can be expressed as

$$\begin{aligned} \widetilde{g}_t(a) &= g_0(a) + \eta L h(a) \cdot t \\ &= (M + \eta L \cdot t)\theta a_0^{\theta} a^{-\theta-1} \end{aligned}$$

Using this, the actual density can be described as $g_t(a) = \widetilde{g}_t(a) \cdot 1(a \geq a^*(t))$. Among these firms, only the most productive firms will operate, and the exit threshold $a^*(t)$

would arise endogenously. In particular, given the Pareto distribution, the equilibrium mass of firms M is constant, independent of t. Thus $a^*(t)$ should satisfy

$$\int_{a^*(t)}^{\infty} \widetilde{g}_t(a) = (M + \eta L \cdot t)\theta a_0^{\theta} \cdot \frac{a^*(t)^{-\theta}}{\theta} = M$$

$$\therefore a^*(t) = \big(1 + \eta \frac{L}{M} \cdot t\big)^{\frac{1}{\theta}} a_0 = \Big(1 + \frac{\theta(\epsilon - 1)}{\theta - (\epsilon - 1)} \eta t\Big)^{\frac{1}{\theta}} a_0$$

That is, the threshold $a^*(t)$ rises over time naturally as a result of accumulation of the best ideas over time. In addition, $a^*(t)$ increases over time roughly at a constant power less than 1,[17] so the growth rate of $a^*(t)$ decreases over time ($\ddot{a}^*(t) < 0$). Despite this slow rate of growth, $a^*(t) \to \infty$ as $t \to \infty$, so the cutoff threshold $a^*(t)$ would grow without bound in the long run.

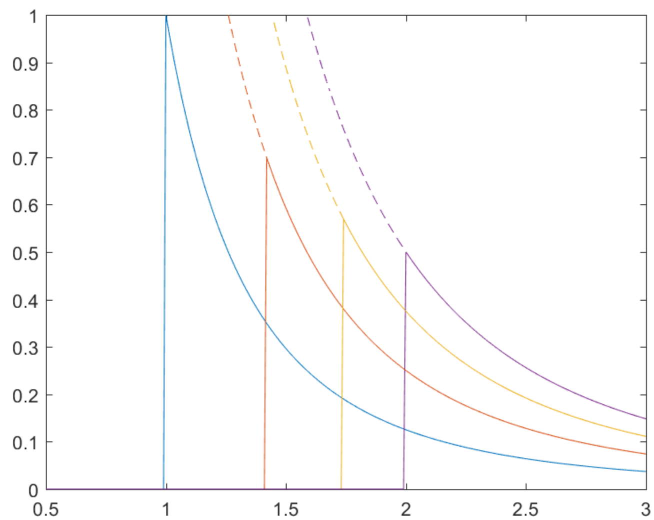
Note that average productivity of this economy $\mathbb{A}(t) = \big(\frac{\theta}{\theta - (\epsilon - 1)}\big)^{\frac{1}{\epsilon - 1}} a^*(t)$ is exactly proportional to $a^*(t)$. Thus the over-time changes in the threshold productivity $a^*(t)$ immediately reveals the rate of growth in the aggregate productivity, which grows over time due to the selective accumulation of the best ideas. In particular, the current example shows that aggregate productivity would grow indefinitely over time even in the absence of any productivity growth of the incumbent firms, as long as ideas keep arriving from the same distribution while driving out the least productive ones.

Now consider the entry rate. The entry rate (normalized by the current mass of operating firms) is the rate of finding any idea, multiplied by the probability that it is better than the current threshold $a^*(t)$, divided by the mass of firms. Therefore

$$\begin{aligned} \eta(t) &= \frac{1}{M} \cdot \eta L \int_{a^*(t)}^{\infty} \theta a_0^{\theta} a^{-\theta - 1} da \\ &= \frac{\eta L}{M} \Big(\frac{a_0^{\theta}}{a^*(t)}\Big)^{\theta} = \frac{(\eta L / M)}{1 + (\eta L / M) \cdot t} \end{aligned}$$

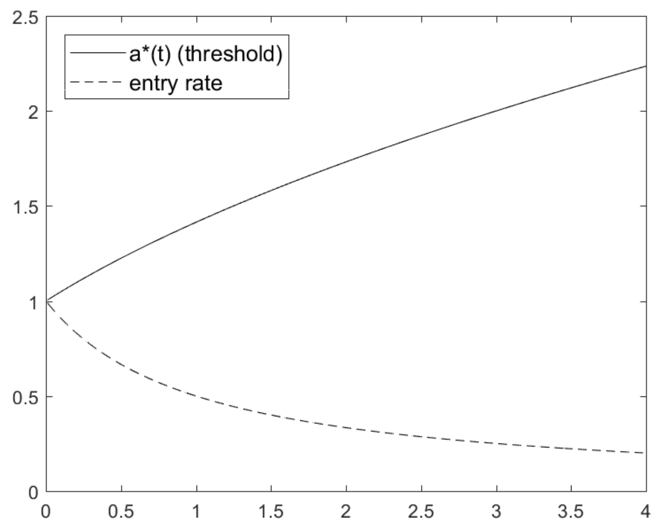
[17] Note that $\theta > (\epsilon - 1)$ must hold for finite A, and $\theta > 1$ must hold for finite M.

Figure 19: An example of the distribution of ideas at $t \in \{0, 1, 2, 3\}$, using a time-invariant Pareto distribution. The entry rate at each t is proportional to $\frac{1}{1+t}$ in this example.



That is, the entry rate would be the highest at $t = 0$ where all ideas are worth entering, and declines over time as the probability of getting an idea better than the current threshold decreases over time. In particular, $\eta(0) = \eta L/M$, and $\eta(t) \to 0$ as $t \to \infty$. This pattern of declining entry rate is consistent with the old literature on the declining entry rate within each industry based on industry life cycle, as well as the recent literature on the declining entry rate in the aggregate economy. Even if the incumbent firms are not improving, the entry that had occured in the past gives rise to an endogenous selection effect, which makes entry more difficult over time. If in addition the incumbent firms can improve productivity compared to the potential entrants for whatever reason — learning by doing, active R&D, or by preempting the customer base, entry would be even more difficult.

Figure 20: Evolution of the threshold idea $a^*(t) = (1+t)^{\frac{1}{\theta}} a_0$, and the entry rate $= \frac{1}{1+t}$, in the same example with time-invariant Pareto distribution. The mass of firms is kept constant as a result of the Pareto distribution.



Example: Time-Varying Distribution of Ideas and Productivity

Now consider more general cases. First, suppose that the distribution of ideas changes exogenously over time. One possibility is that the potential entrants can learn from the current incumbent firms.[18] Another possibility is that there may be an improvement in the general technology, which the incumbent firms cannot take advantage of.

The simplest way to implement this would be to let the lower bound of new ideas a_0 grow at a constant rate, say ν, so that $a_0(t) = e^{\nu t} a_0$. A potential complication arises because it is possible that the rate of improvement of potential entrant's ideas $a_0(t)$ is higher than the rate of increase of the cutoff threshold $a^*(t)$.[19] This is not desirable in the current context because the productivity distribution of incumbent firms will cease to be

[18] Recent literature has explored this assumption theoretically, for example, Perla and Tonetti (2014) and **?**.

[19] For example, $\frac{d}{dt} a^*(t)|_{t=0} = \frac{\eta}{\theta}$, so $\nu > \frac{\eta}{\theta}$ would give rise to this problem.

Pareto. A simple resolution of this minor issue is to extrapolate the Pareto distribution of ideas so that there are always abundant ideas in the lower end, which is probably a realistic description if we think of the nature of ideas in the real world. In this case, it becomes clear that the extrapolated density at some constant, say $a = 1$, is sufficient for describing any "status" of Pareto distribution, including the idea distribution of potential entrants and the productivity distribution of the incumbent firms.

With this additional assumption, the previous approach is readily available again. As the distribution of ideas improves at a rate ν,

$$h_t(a) = \theta(e^{\nu t}a_0)^\theta a^{-\theta-1} = e^{\theta\nu t}\theta a_0^\theta a^{-\theta-1}$$

The density of firms before truncation can be expressed as

$$\begin{aligned}\widetilde{g}_t(a) &= g_0(a) + \eta L h_0(a)\int_0^t e^{\theta\nu t}dt \\ &= \Big(M + \eta L\frac{e^{\theta\nu t}-1}{\theta\nu}\Big)\theta a_0^\theta a^{-\theta-1}\end{aligned}$$

And the mass of firms in operation should be equal to M:

$$\int_{a^*(t)}^\infty \widetilde{g}_t(a) = \Big(M + \eta L\cdot\frac{e^{\theta\nu t}-1}{\theta\nu}\Big)\theta a_0^\theta\cdot\frac{a^*(t)^{-\theta}}{\theta} = M$$

$$\therefore a^*(t) = \Big(1 + \eta\frac{L}{M}\cdot\frac{e^{\theta\nu t}-1}{\theta\nu}\Big)^{\frac{1}{\theta}}a_0$$

If $\nu > 0$, and when t is large,

$$a^*(t) \approx \Big(\frac{\eta}{\theta\nu}\frac{L}{M}\Big)^{\frac{1}{\theta}}e^{\nu t}a_0$$

That is, if the underlying distribution of ideas improves at a constant rate $\nu > 0$, the cutoff threshold also increases at the same rate ν when t is large. This in turn implies that the rate of entry would converge to some constant. Indeed, the entry rate can be expressed as

$$\begin{aligned}\eta(t) &= \int_{a^*(t)}^{\infty} \left(e^{\theta\nu t}\right)\theta a_0^{\theta} a^{-\theta-1} da \cdot \eta \frac{L}{M} \\ &= \frac{\eta \frac{L}{M} e^{\theta\nu t}}{1 + \eta \frac{L}{M} \frac{e^{\theta\nu t}-1}{\theta\nu}}\end{aligned}$$

At $t = 0$, $\eta(0) = \eta L/M$, and as t gets large, $\eta(t)$ would converge to $\theta\nu > 0$. That is, in so far as the distribution of new ideas grows at some constant rate, a positive entry rate in the long run can be sustained. Recall that this result was obtained when the productivity of the incumbent firms were assumed to be fixed while the new ideas are improving.

Second, suppose that the productivity of incumbent firms improve exogenously over time. As mentioned before, even in the absence of active R&D, incumbent firms may enjoy benefits that are increasing over time, possibly because of learning by doing or by preempting the customer base. Again, a simple example is to assume that the productivity of all incumbent firms grow at a constant rate of γ. However, because all firms are improving, the aggregate price would fall taking into account this aggregate growth, and profit of each firm would not increase. But in so far as the potential entrants do not benefit from this exogenous growth, the effect would be the same as when the distribution of new ideas gets worse over time. As a result, the profit of each incumbent firm would decrease at a slower rate compared to the case without this exogenous growth, and the entry rate would decrease more rapidly.

In fact, the dynamics of the economy would be very similar to the case where the distribution of new ideas deteriorates over time at the ratio of γ, or in other words, $\nu < 0$

in the above example. The key object is the evolution of cutoff threshold:

$$a^*(t) = \Big(1 + \eta \frac{L}{M} \cdot \frac{e^{\theta\nu t} - 1}{\theta\nu}\Big)^{\frac{1}{\theta}} a_0$$

If $\nu < 0$,

$$\lim_{t \to \infty} a^*(t) = \Big(1 + \frac{\eta}{\theta\nu} \frac{L}{M}\Big)^{\frac{1}{\theta}} a_0$$

That is, there is an upper limit to the cutoff threshold.

It is also straightforward to consider both effects. This can be thought of as a race between the incumbent firms and potential entrants. It is the difference in the growth rates of the ideas and the productivity, $(\nu - \gamma)$, that matters. If the entrants' ideas are improving at a faster rate than the rate of improvement of the incumbents' productivity $(\nu - \gamma > 0)$, the entry rate would converge to a strictly positive value in the long run, and the cutoff threshold would also keep increasing at a constant rate. If $\nu - \gamma < 0$, the entry rate would converge to zero in the long run, and the cutoff threshold would converge to a constant (net of the exogenous growth rate of incumbents γ). If $\nu - \gamma = 0$, the entry rate would converge to zero in the long run but at a slower rate ($\propto t^{-1}$ as opposed to an exponential decay), and cutoff threshold would grow at the speed of a polynomial ($\propto t^{\frac{1}{\theta}}$). When defined in terms of maintaining a constant positive entry rate, balanced growth path can be sustained only if $\nu > \gamma$, that is, the entrant's idea should improve at a faster rate than the rate of improvement in productivity.

Example: Population Growth

In this economy, population growth results in two different effects. First, it allows for a greater mass of better ideas, as if the distribution of new ideas are improving. Second, it is able to sustain an increasing mass of firms to operate in equilibrium.

Suppose that the population grows at a constant rate of λ:

$$\overline{L}_t = e^{\lambda t}\overline{L}_0$$

Based on the Pareto distribution of firm's productivity, $M_t = e^{\lambda t}M_0$ and $L_t = e^{\lambda t}L_0$ also holds. Suppose the distribution of ideas or the productivity of incumbent firms do not change over time, so that $h_t(a) = h(a) = \theta a_0^\theta a^{-\theta-1}$.

Between $(t, t+dt)$, the density of new entrants with idea of quality a would be $\eta L_0 e^{\lambda t} h(a)dt$. The density of all operating firms at t (before truncation) would then be

$$\begin{aligned}\widetilde{g}_t(a) &= M_0 h(a) + \int_0^t \eta L_0 e^{\lambda t} h(a) dt \\ &= \Big(M_0 + \eta L_0 \frac{e^{\lambda t}-1}{\lambda}\Big) h(a)\end{aligned}$$

The mass of surviving firms at t is equal to

$$M_0 e^{\lambda t} = \int_{a^*(t)}^{\infty} \widetilde{g}_t(a) = \Big(M_0 + \eta L_0 \frac{e^{\lambda t}-1}{\lambda}\Big)\theta a_0^\theta \frac{a^*(t)^{-\theta}}{\theta}$$

$$\therefore a^*(t) = \Big(e^{-\lambda t} + \frac{\eta L_0}{\lambda M_0}(1-e^{-\lambda t})\Big)^{\frac{1}{\theta}} a_0$$

It turns out that $a^*(t)$ converges to some constant $(\frac{\eta L_0}{\lambda M_0})^{1/\theta} a_0$ as $t \to \infty$. And lower the population growth rate λ, the threshold would increase up to a higher constant. The reason for an upper limit on this cutoff threshold is that the growing population can sustain a growing number of firms in equilibrium. When normalized by the existing mass of firms, exit rate is simply proportional to $\frac{d}{dt}a^*(t)$, so the exit rate near this limit would converge to zero. Note that the exit rate in units of mass of firms (not normalized) converges to a nonzero constant.[20]

[20] $\lim_{t\to\infty} M_t \frac{d}{dt} a^*(t) = \frac{1}{\theta}(\frac{\eta L_0}{\lambda M_0})^{1/\theta - 1}(\frac{\eta L_0}{\lambda M_0} - 1)\lambda M_0 > 0$

Now consider the entry rate:

$$\eta(t) = \frac{1}{M_0 e^{\lambda t}} \int_{a^*(t)}^{\infty} \eta L_0 e^{\lambda t} h(a) da = \frac{\eta L_0}{M_0} \Big(\frac{a^*(t)}{a_0} \Big)^{-\theta}$$
$$= \frac{\frac{\eta L_0}{M_0}}{e^{-\lambda t} + \frac{\eta L_0}{\lambda M_0}(1 - e^{-\lambda t})}$$

At $t = 0$, $\eta(0) = \frac{\eta L_0}{M_0}$, which is natural. As $t \to \infty$, $\eta(t) \to \lambda$. In the long run, the entry rate would be a constant, equal to the population growth rate, whereas the exit rate converges to zero. This simply means that an increasingly larger population gives rise to proportionally larger mass of ideas, which can all be sustained by the growing population itself.

3.4 Multiple Industries

3.4.1 Static Equilibrium

Now consider the case where there are multiple industries indexed by i, each of which has a continuum of varieties indexed by j. The aggregate consumption is now defined as

$$\mathbb{C} = \Big(\int \mathbb{C}_i^{\frac{\phi-1}{\phi}} di \Big)^{\frac{\phi}{\phi-1}} = \Big(\int \Big[\int c_i(j)^{\frac{\epsilon-1}{\epsilon}} dj \Big]^{\frac{\epsilon}{\epsilon-1} \frac{\phi-1}{\phi}} di \Big)^{\frac{\phi}{\phi-1}}$$

where the previously defined $\mathbb{C}$ is now indexed by i, and the aggregate consumption $\mathbb{C}$ is a CES aggregate of each $\mathbb{C}_i$ with elasticity of substitution equal to $\phi \in [1, \epsilon]$. If $\phi = \epsilon$, the goods of different industries are equally substitutable as goods within a same industry, in which case there is effectively no distinction of industry.

Most of the results obtained under a single industry extends to the case with multiple industries in a straightforward way. The prices of individual goods are $p_i(a) = \frac{\epsilon}{\epsilon-1} \frac{w}{a}$.

The aggregate price index of an industry i is

$$\mathbb{P}_i = \Big(\int p_i(j)^{1-\epsilon}dj\Big)^{\frac{1}{1-\epsilon}} = \frac{\epsilon}{\epsilon-1}w\Big(M_i^{\frac{1}{\epsilon-1}}\mathbb{A}_i\Big)^{-1}$$

and the aggregate price index across all industries is

$$\mathbb{P} = \Big(\int \mathbb{P}_i^{1-\phi}di\Big)^{\frac{1}{1-\phi}} = \frac{\epsilon}{\epsilon-1}w\Big(\int (M_i^{\frac{1}{\epsilon-1}}\mathbb{A}_i)^{\phi-1}\Big)^{\frac{1}{1-\phi}}$$

The aggregate productivity in industry i is

$$\mathbb{A}_i = \Big(\frac{1}{M_i}\int a(j)^{\epsilon-1}dj\Big)^{\frac{1}{\epsilon-1}}$$

Output of a firm in industry i with productivity a is

$$y_i(a) = \Big(\frac{p_i(a)}{\mathbb{P}_i}\Big)^{-\epsilon}\Big(\frac{\mathbb{P}_i}{\mathbb{P}}\Big)^{-\phi}\mathbb{C} = \Big(\frac{\epsilon}{\epsilon-1}\frac{w}{a}\Big)^{-\epsilon}\mathbb{P}_i^{\epsilon-\phi}\mathbb{P}^{\phi}\mathbb{C}$$

and the subsequent profit is

$$\pi_i(a) = \frac{w}{\epsilon-1}\frac{y_i(a)}{a} = \frac{w}{\epsilon-1}\Big(\frac{\epsilon}{\epsilon-1}w\Big)^{-\epsilon}a^{\epsilon-1}\mathbb{P}_i^{\epsilon-\phi}\mathbb{P}^{\phi}\mathbb{C}$$

Note that if $\phi = \epsilon$, firms in different industries with same productivity a would charge the same price, produce an equal amount, and earn the same profit. However, if $\phi < \epsilon$, firms in different industries with same productivity would produce a different amount while charging the same price. The firm in an industry with lower aggregate price $\mathbb{P}_i$ would produce less and earn less profit because that industry is more competitive than the other, which reflects a higher average productivity or a greater mass of firms.

Indifference condition in each industry would pin down the equilibrium cutoff threshold productivity by industry. By setting $\pi(a_i^*) = w$,

$$(a_i^*)^{\epsilon-1} = (\epsilon-1)\Big(\frac{\epsilon}{\epsilon-1}\frac{w}{\mathbb{P}_i}\Big)^{\epsilon}\Big(\frac{\mathbb{P}_i}{\mathbb{P}}\Big)^{\phi}\mathbb{C}^{-1} = (\epsilon-1)\Big(M_i^{\frac{1}{\epsilon-1}}\mathbb{A}_i\Big)^{\epsilon}\mathbb{C}_i^{-1}$$

which is the same as the previously derived a^* except that it now depends on industry specific characteristics.

The above equations do not depend on distributional assumptions on the productivity. Now assume that the distribution of productivity or ideas are Pareto. As before, I assume that the density of productivity is given as

$$h_i(a) \equiv h_i a^{-\theta-1} \cdot \mathbb{1}(a \geq a_i^*)$$

where h_i indicates the degree of maturity of an industry i that would typically increase as ideas accumulate over time. It is assumed that the shape parameter θ is same for all industries. Since a_i^* is endogenously determined, the only exogenous parameter is h_i, which summarizes all the relevant industry-specific characteristics. Then the average productivity of an industry, and the mass of firms in an industry, can be expressed as

$$\mathbb{A}_i = \left(\frac{\theta}{\theta - (\epsilon - 1)}\right)^{\frac{1}{\epsilon-1}} a_i^*$$

$$M_i = \frac{h_i}{\theta}(a_i^*)^{-\theta}$$

Now consider two different industries $i \neq k$. Using the indifference condition derived above, the ratio between equilibrium cutoff thresholds in these two industries can be expressed as

$$\left(\frac{a_i^*}{a_k^*}\right)^{\epsilon-1} = \left(\frac{\mathbb{P}_i}{\mathbb{P}_k}\right)^{\phi-\epsilon} = \left(\frac{M_i}{M_k}\right)^{\frac{\epsilon-\phi}{\epsilon-1}} \left(\frac{A_i}{A_k}\right)^{\epsilon-\phi} = \left(\frac{h_i}{h_k}\right)^{\frac{\epsilon-\phi}{\epsilon-1}} \left(\frac{a_i^*}{a_k^*}\right)^{-\frac{\theta-(\epsilon-1)}{\epsilon-1}(\epsilon-\phi)}$$

Rearranging,

$$\frac{a_i^*}{a_k^*} = \left(\frac{h_i}{h_k}\right)^{\frac{\epsilon-\phi}{K}}$$

where $K \equiv \theta(\epsilon - \phi) + (\epsilon - 1)(\phi - 1)$. Using this, the equilibrium cutoff threshold can be expressed as

$$a_i^* = X \cdot h_i^{\frac{\epsilon-\phi}{K}}$$

where X is some constant common to all industries.

It remains to find the value of X. Since the number of workers in industry i, L_i, is proportional to $\mathbb{P}_i \mathbb{C}_i \propto \mathbb{P}_i^{1-\phi}$,

$$L_i = \Big(\frac{\mathbb{P}_i}{\mathbb{P}}\Big)^{1-\phi} L = \Big(\frac{h_i}{\mathbb{H}}\Big)^{\frac{(\epsilon-1)(\phi-1)}{K}} L$$

where $\mathbb{H} \equiv \big(\int h_i^{\frac{(\epsilon-1)(\phi-1)}{K}} di\big)^{\frac{K}{(\epsilon-1)(\phi-1)}}$ measures the "average" maturity of existing industries, which also increases as the mass or number of industries grows. Because it would appear frequently, define $\rho \equiv \frac{(\epsilon-1)(\phi-1)}{K} \in [0,1]$. Note that $\phi = \epsilon$ implies $\rho = 1$, and $\phi = 1$ results in $\rho = 0$.

In addition, the mass of firms in industry i is

$$M_i = \frac{1}{\theta} X^{-\theta} \cdot h_i^{\rho}$$

The mass of all workers is

$$L = \frac{\theta(\epsilon-1)}{\epsilon\theta - (\epsilon-1)} \overline{L}$$

and the ratio between the mass of entrepreneurs (firms) and workers is

$$\frac{L_i}{M_i} = \frac{L}{M} = \frac{\theta(\epsilon-1)}{\theta - (\epsilon-1)}$$

Combining all these equations,

$$X = \Big(\frac{\epsilon\theta - (\epsilon-1)}{\theta(\theta - (\epsilon-1))} \mathbb{H}^{\rho} \overline{L}^{-1}\Big)^{\frac{1}{\theta}}$$

Finally, the equilibrium cutoff threshold in industry i, and the mass of firms in industry i can be expressed as

$$a_i^* = X \cdot h_i^{\frac{\epsilon-\phi}{K}} = \Big(\frac{\epsilon\theta - (\epsilon-1)}{\theta(\theta - (\epsilon-1))}\Big)^{\frac{1}{\theta}} \overline{L}^{-\frac{1}{\theta}} \mathbb{H}^{\frac{1}{\theta}} \Big(\frac{h_i}{\mathbb{H}}\Big)^{\frac{1-\rho}{\theta}}$$

$$M_i = \frac{h_i}{\theta}(a_i^*)^{-\theta} = \frac{\theta - (\epsilon - 1)}{\epsilon\theta - (\epsilon - 1)}\overline{L}\Big(\frac{h_i}{\mathbb{H}}\Big)^{\rho} = M\Big(\frac{h_i}{\mathbb{H}}\Big)^{\rho}$$

As ϕ varies from 1 to ϵ, $\frac{\epsilon-\phi}{K}$ varies from $\frac{1}{\theta}$ to 0. If $\phi = \epsilon$, the industry specific maturity h_i have no impact on a_i^*, and it is only the aggregate maturity $\mathbb{H}$ that determines the common cutoff threshold for all industries. This is a natural result, considering that the between-industry substitutability is exactly the same as within-industry substitutability. If $\phi = 1$, the aggregate maturity $\mathbb{H}$ has no impact on individual threshold a_i^*, and only the maturity of that specific industry fully determines this threshold. This is also a natural consequence of Cobb-Douglas specification where total expenditure in each industry is fixed. For intermediate values of ϕ, both the economy-wide average maturity and the industry-specific maturity relative to the economy-wide average maturity affect the industry-specific cutoff threshold.

3.4.2 Dynamic Environment

Based on the above characterization in a static environment, now I move on to the dynamic properties where the set of industries increase over time. Consistent with how ideas arrive within each industry, I assume that the industry itself arrive exogenously over time. There is a continuum of industries, each characterized by the time of arrival and the pattern of arrival of new ideas within industry.

For simplicity, I assume that all industries evolve over time in the same manner once they are introduced. As shown above, the variable h_i captures the degree of maturity of an industry i, which is now time-varying as new ideas arrive within that industry. In particular, I assume that $h_i(t) = \eta(t - T_i) \cdot \mathbb{1}(t \geq T_i)$, where T_i is the time when industry i was first discovered. Starting from $t = T_i$, ideas start accumulating in industry i, drawn from a Pareto distribution with shape parameter θ, and with a constant rate of accumulation η that is common to all industries. T_i is then a sufficient statistic that captures all the industry specific characteristics. $(t - T_i) \equiv s_i$ measures the age of the industry i at time t, so older industries are more mature in terms of mass of ideas (for some fixed quality of idea) accumulated up to t.

The distribution of industry can be characterized by the time of arrival T_i, of which density is denoted by $g_T(t)$. I will consider two specific examples: First, a constant arrival of new industry, where the economy starts from time 0: $\big[g_T(t) = \gamma \cdot \mathbb{1}(t \geq 0)\big]$. Second, a gradually increasing rate of arrival of new industries, where the economy starts from time $-\infty$: $\big[g_T(t) = \gamma e^{\lambda t}\big]$. Among all potential industries, only the ones with $T_i \leq t$ exist at t.

Given $g_T(t)$, the density in terms of maturity at each point in time t can be expressed as $g_H(h;t) = \frac{1}{\eta} g_T(t - \frac{h}{\eta})$. The key variable that summarizes the aggregate state of maturity $\mathbb{H}(t)$ can be found from

$$\mathbb{H}(t) \equiv \Big(\int_0^\infty h^\rho g_H(h;t) dh \Big)^{\frac{1}{\rho}}$$

where $\rho \equiv \frac{(\epsilon-1)(\phi-1)}{(\epsilon-\phi)\theta+(\epsilon-1)(\phi-1)} \in [0,1]$, as defined previously.

Constant (Linear) Expansion of Industries

In this section, $g_T(t) = \gamma \cdot \mathbb{1}(t \geq 0)$ will be assumed, which means that the same mass $\gamma\, dt$ of industries are newly introduced at each point in time. Changing the variable from the time of entry T_i into the measure of maturity h, $g_H(h;t) = \frac{\gamma}{\eta} \cdot \mathbb{1}(h \in (0, \eta t))$ is a uniform distribution over $(0, \eta t)$, where the oldest industry has $h = \eta t$ and the youngest one $h = 0$. Integrating over all industries,

$$\mathbb{H}(t) = \Big(\int_0^{\eta t} h^\rho \, \frac{\gamma}{\eta} \, dh \Big)^{\frac{1}{\rho}} = \eta \Big(\frac{\gamma}{1+\rho} \Big)^{\frac{1}{\rho}} t^{\frac{1+\rho}{\rho}}$$

Naturally, the aggregate maturity would grow proportional to t if there is no introduction of new industries. Additional industry makes the aggregate distribution of ideas to improve at a faster rate than linear in t.

Consistent with the previous notation, $h_i(t)$ is normalized as the density of ideas that have arrived up to time t at $a = 1$ in each industry i. The density of cumulative ideas in industry i at time t can be expressed as $f_i(a,t) = h_i(t) a^{-\theta-1} = \eta(t - T_i) a^{-\theta-1} \cdot \mathbb{1}(t \geq T_i)$, and the density of operating firms at t would be $f_i(a,t) \cdot \mathbb{1}(a \geq a_i^*(t))$.[21]

As derived above, the mass of firms in industry i is

$$M_i(t) = M \Big(\frac{h_i(t)}{\mathbb{H}(t)} \Big)^\rho = M \frac{1+\rho}{\gamma} t^{-(1+\rho)} (t - T_i)^\rho$$

and the cutoff threshold satisfies

$$a_i^*(t)^\theta = \frac{1}{\theta M} \mathbb{H}(t)^\rho h_i(t)^{1-\rho} = \frac{1}{\theta M} \frac{\eta \gamma}{1+\rho} t^{1+\rho} (t - T_i)^{1-\rho}$$

Entry rate in industry i is the mass of additional ideas above the threshold $a_i^*(t)$ that arrives during the marginal lapse of period dt:

$$\eta_i(t) = \int_{a_i^*(t)}^\infty \frac{\partial f_i(a,t)}{\partial t} da = \frac{\eta}{\theta} a_i^*(t)^{-\theta} = \frac{M(1+\rho)}{\gamma} t^{-(1+\rho)} (t - T_i)^{\rho-1}$$

Exit rate in industry i is the mass of firms at $a_i^*(t)$ that are driven out of business because of a marginal increase in $a_i^*(t)$:

$$\begin{aligned} \delta_i(t) &= f_i(a_i^*(t), t) \frac{d a_i^*(t)}{dt} = \eta(t - T_i) a_i^*(t)^{-\theta-1} \frac{d a_i^*(t)}{dt} = -\frac{\eta}{\theta} (t - T_i) \frac{d}{dt} \big(a_i^*(t)^{-\theta} \big) \\ &= \frac{M(1+\rho)}{\gamma} t^{-2-\rho} (t - T_i)^{\rho-1} \Big(2t - (1+\rho) T_i \Big) \end{aligned}$$

[21] Again, because of this truncation by endogenous selection, not specifying the lower bound for the Pareto distribution of ideas is innocuous and sustains a finite aggregate entry rate, as long as $\phi > 1$.

The rate of change in the mass of firms in an industry is equal to the entry rate minus the exit rate, as it should be:

$$\frac{d}{dt}M_i(t) = \eta_i(t) - \delta_i(t) = \frac{M(1+\rho)}{\gamma}t^{-2-\rho}(t-T_i)^{\rho-1}\Big((1+\rho)T_i - t\Big)$$

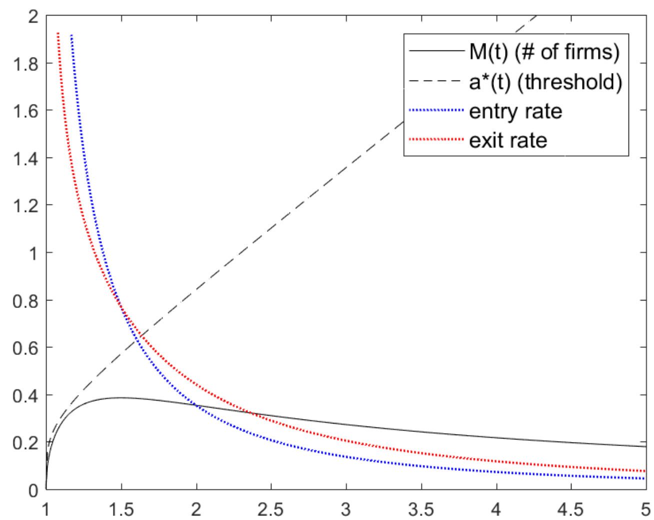
Note that $\frac{dM_i(t)}{dt} = 0$ implies $t = (1+\rho)T_i$.

If $\rho > 0$, $M_i(T_i) = 0$. That is, an industry begins with mass zero at $t = T_i$. And then the number of firms increases until $t = (1+\rho)T_i$ where it hits the maximum, and gradually decreases thereafter. This is in fact constistent with the well-established empirical result in the older literature on industry life cycle, for example, Gort and Klepper (1982), that the number of firms in an industry first increases over time and then gradually decreases (i.e. the "shakeout" phase). The current model, which is extremely simple where ideas arrive exogenously from a time-invariant distribution, is able to explain this robust and well-known pattern of industry life cycle.

If $\rho = 0$, $M_i(t) = \frac{M(1+\rho)}{\gamma}t^{-1}$, and in particular, $M_i(T_i) = \frac{M(1+\rho)}{\gamma}T_i^{-1} > 0$. This implies that firms not only enter at a rate of infinity when an industry is first discovered, but it immediately attains a strictly positive mass which turns out to be its maximum. This stark result is an outcome of two assumptions. First, the Cobb-Douglas preference implies that even the lowest productivity industry gets an equal share of expenditure. Second, it is assumed that there is no lower bound to the distribution of ideas, and the Pareto distribution with $\theta > 0$ implies that $\int_0^x a^{-\theta-1}da = \infty$ for any strictly positive value of x. So there are sufficiently many individuals who have some ideas that are arbitrarily close to zero, and they immediately enter into business to satisfy the demand for firms in this industry based on the Cobb-Douglas preferences.

Having characterized the entry rates and the mass of firms in each industry, now it

Figure 21: The time-series pattern of an industry. The industry was first discovered at $t = 1$, and since then ideas arrive constantly from a time-invariant distribution.



remains to characterize the aggregate entry and exit rates. The aggregate rate of entry into existing industries can be expressed as

$$\eta(t) = \int_0^t \eta_i(t) g_T(\widetilde{T}) d\widetilde{T} = \frac{M(1+\rho)\eta}{\gamma} t^{-\rho-1} \int_0^t \gamma(t-\widetilde{T})^{\rho-1} d\widetilde{T} = \frac{M(1+\rho)}{\rho} t^{-1}$$

That is, the aggregate entry rate is simply proportional to t^{-1}, which is consistent with the result obtained from the single industry case. It turns out that the elasticity of substitution between industries, ϕ, affects the overall level but not its time-series pattern.

Increasing (Exponential) Expansion of Industries

Now suppose that $g_T(t) = \gamma e^{\lambda t}$, which means that the number of industries introduced at each t is exponentially increasing. The assumption on the evolution of maturity in each industry is maintained, so that $h_i = \eta(t - T_i) \cdot \mathbb{1}(t \geq T_i)$. The corresponding density of industries in terms of maturity would be $g_H(h;t) = \frac{\gamma}{\eta} e^{\lambda(t - \frac{h}{\eta})}$, which is now an exponential distribution with support $(0, \infty)$. Integrating over all industries,

$$\mathbb{H}(t)^\rho = \int_{-\infty}^{t} \big[\eta(t - T_i)\big]^\rho \gamma e^{\lambda \widetilde{T}} d\widetilde{T} = \gamma \eta^\rho e^{\lambda t} \underbrace{\int_0^\infty s^\rho e^{-\lambda s} ds}_{\equiv H_0}$$

where the time of introduction $\widetilde{T}$ has been replaced by age of the industry $s = t - \widetilde{T}$ for convenience. Using this measure of time-varying aggregate maturity, the cutoff threshold as well as the mass of firms can be expressed as

$$a_i^*(t)^\theta = \frac{1}{\theta M} \mathbb{H}(t)^\rho h_i(t)^{1-\rho} = \frac{\eta\gamma}{\theta M} e^{\lambda t} s_i^{1-\rho} H_0$$

$$M_i(t) = M\Big(\frac{h_i(t)}{\mathbb{H}(t)}\Big)^\rho = \frac{M}{\gamma} e^{-\lambda t} \frac{s_i^\rho}{H_0}$$

Consequently, the entry and exit rates can be found as

$$\eta_i(t) = \int_{a_i^*(t)}^{\infty} \frac{\partial f_i(a,t)}{\partial t} da = \frac{\eta}{\theta} a_i^*(t)^{-\theta} = \frac{M}{\gamma} e^{-\lambda t} \frac{s_i^{\rho-1}}{H_0}$$

$$\begin{aligned} \delta_i(t) &= f_i(a_i^*(t), t) \frac{da_i^*(t)}{dt} = -\frac{\eta}{\theta}(t - T_i) \frac{d}{dt}\big(a_i^*(t)^{-\theta}\big) \\ &= \frac{M}{\gamma H_0} e^{-\lambda t} s_i^{\rho-1} \Big((1-\rho) + \lambda s_i\Big) \end{aligned}$$

Again, it can be confirmed that

$$\frac{d}{dt} M_i(t) = \eta_i(t) - \delta_i(t) = \frac{M}{\gamma H_0} e^{-\lambda t} s_i^{\rho-1} \big(\rho - \lambda s_i\big)$$

Setting $\frac{d}{dt} M_i(t) = 0$, $s_i = t - T_i = \frac{\rho}{\lambda}$ is the age of an industry at which it attains the maximum number of firms. If the rate of introduction of new industries λ is higher, this peak occurs at an earlier time, which implies that an industry is crowded out more rapidly because of the emergence of competing industries. As before, ρ closer to 0, or equivalently, ϕ closer to 1, implies an immediate expansion of demand for firms in a new industry even at a very low maturity (and thus low productivity), which is an artifact of the near Cobb-Douglas utility.

Looking at the mass of firms $M_i(t) \propto e^{-\lambda t} s_i^{\rho}$, it has a similar form with the previous example with constant expansion of industries $M_i(t) \propto t^{-(1+\rho)} s_i^{\rho}$. Unless $\rho = 0$, the mass of firms depends positively on the age of an industry $s_i \equiv (t - T_i)$. Because age increases over time, this leads to an increase in the mass of firms over time, and is simply an outcome of elapse of time which allow ideas to arrive and selectively accumulate. In addition, the mass of firms negatively depends on the time t in a direct manner. This term captures the dependence of an industry on the growth of other competing industries. Combination of these two forces leads to a natural time path of number of firms in each industry, initially increasing and then decreasing. Although the assumption on the distribution of arrival of new industries matter for the exact time path, such a general pattern would be visible regardless of the timing of arrival of each industries.

Finally, the aggregate entry and exit rates:

$$\begin{aligned}\delta(t) &= \int_{-\infty}^{t} \delta_i(t) g_T(\widetilde{T}) d\widetilde{T} = \frac{M}{\gamma H_0} \int_0^{\infty} e^{-\lambda s}\Big((1-\rho)s^{\rho-1} + \lambda s^{\rho}\Big) ds \\ &= \frac{M}{\gamma H_0}\Big((1-\rho)\frac{\lambda}{\rho} H_0 + \lambda H_0\Big) = \frac{\lambda M}{\rho \gamma}\end{aligned}$$

where $H_0 \equiv \int_0^{\infty} s^{\rho} e^{-\lambda s} ds = \frac{\rho}{\lambda} \int_0^{\infty} s^{\rho-1} e^{-\lambda s} ds$ has been used. It turns out that the aggregate entry or exit rate is simply a time-invariant function of the key parameters. Higher λ, or faster expansion of industries, implies higher aggregate entry rate. Greater γ, or larger mass of industries leads to lower aggregate entry rate, because the cutoff threshold would be higher in each industry, and only the ideas above this cutoff threshold would enter. Lower value of ρ, or lower elasticity of substitution between industries, results in a higher entry rate for the reasons elucidated above.

3.5 Conclusion

This paper proposes a model of accumulation of ideas. Market equilibrium leads to a natural selection where inferior ideas are discarded, which results in a rising threshold idea over time. The model successfully explains the decline in the entry rate if the distribution of ideas is fixed and the arrival rate is constant. In addition, the extended model with an expanding set of industries successfully replicates the industry life cycle pattern, in which the number of firms in an industry first increases and then decreases over time.

Bibliography

Agarwal, R. and Gort, M. (1996). The evolution of markets and entry, exit and survival of firms. *Review of Economics and Statistics*.

Aghion, P., Akcigit, U., and Howitt, P. (2014). What do we learn from Schumpeterian growth theory? *Handbook of Economic Growth*.

Akcigit, U. and Ates, S. T. (2019a). Ten facts on declining business dynamism and lessons from endogenous growth theory. *NBER Working Paper*.

Akcigit, U. and Ates, S. T. (2019b). What happened to U.S. business dynamism? *NBER Working Paper*.

Alon, T., Berger, D., Dent, R., and Pugsley, B. (2018). Older and slower: The startup deficit's lasting effects on aggregate productivity growth. *Journal of Monetary Economics*.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*.

Bollard, A., Klenow, P. J., and Li, H. (2016). Entry costs rise with development. *Working paper*.

Byrne, D. M., Fernald, J. G., and Reinsdorf, M. B. (2016). Does the United States have a productivity slowdown or a measurement problem? *Brookings Papers on Economic Activity*.

Caballero, R. J. and Hammour, M. L. (1994). The cleansing effect of recessions. *American Economic Review*.

Cao, S., Salameh, M., Seki, M., and St-Amant, P. (2017). Trends in firm entry and new entrepreneurship in Canada. *Canadian Public Policy*.

Cette, G., Fernald, J., and Mojon, B. (2016). The pre-Great Recession slowdown in productivity. *European Economic Review*.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*.

Decker, R., Haltiwanger, J., Jarmin, R., and Miranda, J. (2014). The secular decline in business dynamism in the U.S. *working paper*.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*.

Fernald, J. G. (2015). Productivity and potential output before, during, and after the Great Recession. *NBER Macroeconomics Annual*.

Gort, M. and Klepper, S. (1982). Time paths in the diffusion of product innovations. *Economic Journal*.

Grossman, G. M. and Helpman, E. (1991). Quality ladders in the theory of growth. *Review of Economic Studies*.

Hopenhayn, H., Neira, J., and Singhania, R. (2018). From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share. *NBER Working Paper No. 25382*.

Hopenhayn, H. A. (1993). The shakeout. *Working paper.*

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Review of Economic Studies.*

Jones, C. I. (1995). R&d-based models of economic growth. *Journal of Political Economy.*

Jovanovic, B. and MacDonald, G. M. (1994). The life cycle of a competitive industry. *Journal of Political Economy.*

Karahan, F., Pugsley, B., and Sahin, A. (2019). Demographic origins of the startup deficit. *NBER Working Paper No. 25874.*

Klepper, S. (1996). Entry, exit, growth, and innovation over the product life cycle. *American Economic Review.*

Klette, T. J. and Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of Political Economy.*

Kortum, S. S. (1997). Research, patenting, and technological change. *Econometrica.*

Kozeniauskas, N. (2018). What's driving the decline in entrepreneurship? *Working paper.*

Lee, Y. and Mukoyama, T. (2018). A model of entry, exit, and plant-level dynamics over the business cycle. *Journal of Economic Dynamics and Control.*

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica.*

Monte, F. (2011). Skill bias, trade, and wage dispersion. *Journal of International Economics*.

Perla, J. and Tonetti, C. (2014). Equilibrium imitation and growth. *Journal of Political Economy*.

Pugsley, B. W. and Sahin, A. (2019). Grown-up business cycles. *Review of Financial Studies*.

Salgado, S. (2019). Technical change and entrepreneurship. *Working paper*.

Sampson, T. (2016). Dynamic selection: An idea flows theory of entry, trade, and growth. *Quarterly Journal of Economics*.

Syverson, C. (2017). Challenges to mismeasurement explanations for the US productivity slowdown. *Journal of Economic Perspectives*.

Appendix A

Appendix to Chapter 1

A.1 Financial Intermediaries' Problem

The position taken by the financial intermediaries is denoted as $D_t^* = \frac{D_t}{\mathcal{E}_t}$. The ex post profit in units of real Foreign goods at $t+1$ can be written as

$$\frac{\Pi_{t+1}^{\mathcal{I}*}}{P_{t+1}^*} = \Big[\frac{P_t^*}{P_{t+1}^*}\big(R_t^* - R_t\frac{\mathcal{E}_t}{\mathcal{E}_{t+1}}\big)\Big]\frac{D_t^*}{P_t^*} \equiv \widehat{R}_{t+1}\frac{D_t^*}{P_t^*}$$

The term $\widehat{R}_{t+1} = \frac{P_t^*}{P_{t+1}^*}\big(R_t^* - R_t\frac{\mathcal{E}_t}{\mathcal{E}_{t+1}}\big)$ contains two sources of uncertainty: $\mathcal{E}_{t+1}$ and P_{t+1}^*. Assume that this stochastic excess return $\widehat{R}_{t+1}$ conditional on information at t is normally distributed with mean μ_t and variance σ^2.[1] The intermediaries are assumed to have CARA utility: $u(x) = -\exp(-\gamma x)$. Their expected utility can be expressed as:

$$E_t u\big(\frac{\Pi_{t+1}^{\mathcal{I}*}}{P_{t+1}^*}\big) = -E_t \exp\big(-\widehat{R}_{t+1}\frac{D_t^*}{P_t^*}\big) = -\exp\big(-\gamma\mu_t\frac{D_t^*}{P_t^*} + \frac{1}{2}(\gamma\sigma\frac{D_t^*}{P_t^*})^2\big)$$

where the second equality makes use of the property of log-normal distribution. Maximization of this expected utility is thus equivalent to minimizing the term that goes into the exponential function, which is just a deterministic, quadratic function of $\frac{D_t^*}{P_t^*}$. The utility-maximizing choice implies

$$\frac{D_t^*}{P_t^*} = \frac{\mu_t}{\gamma\sigma^2} = \frac{1}{\gamma\sigma^2}E_t\frac{P_t^*}{P_{t+1}^*}\big(R_t^* - R_t\frac{\mathcal{E}_t}{\mathcal{E}_{t+1}}\big)$$

[1] Strictly speaking, this excess return is not necessarily exactly normally distributed. See Itskhoki and Mukhin (2019) for the derivation under continuous-time environment that gets over this problem. It turns out that the result ends up being almost the same.

Log-linearing around the steady state with $(P^*)^{ss} = \mathcal{E}^{ss} = 1$, $R^{ss} = (R^*)^{ss} = \beta^{-1}$, and $(D^*)^{ss} = -(B^*)^{ss} = 0$, while normalizing $D_t^* \approx (Y^*)^{ss} d_t^*$,

$$d_t^* = -d_t = -\frac{1}{\gamma\sigma^2 \cdot (Y^*)^{ss}} E_t er_{t+1} \equiv \frac{1}{\chi} E_t er_{t+1}$$

$$b_t = -b_t^* = -\frac{1}{\chi} E_t er_{t+1}$$

It follows that $E_t er_{t+1} = E_t \Delta w_{t+1} = -\chi b_t$. This equation is called "modified UIP" condition throughout the paper.

A.2 World Budget Constraint

The consolidated budget constraint in the Home country can be expressed as:

$$P_t C_t + B_t = P_{Ht} Y_t + R_{t-1} B_{t-1}$$

Likewise, the consolidated budget constraint in the Foreign country would be:

$$P_t^* C_t^* + B_t^* = P_{Ft}^* Y_t^* + R_{t-1}^* B_{t-1}^* + \Pi_t^{\mathcal{I}*}$$

where $\Pi_t^{\mathcal{I}*}$ denotes profits to the financial intermediaries that is transferred to the Foreign households.

Adding the two country budget constraints, with the Foreign budget constraint multiplied by $\mathcal{E}_t$,

$$\big(P_t C_t + \mathcal{E}_t P_t^* C_t^*\big) + \big(B_t + \mathcal{E}_t B_t^*\big) = \big(P_{Ht} Y_t + \mathcal{E}_t P_{Ft}^* Y_t^*\big) + \big(R_{t-1} B_{t-1} + \mathcal{E}_t R_{t-1}^* B_{t-1}^*\big) + \mathcal{E}_t \Pi_t^{\mathcal{I}*}$$

As shown above, $B_t + \mathcal{E}_t B_t^* = 0$ holds. In addition, using $\Pi_t^{\mathcal{I}*} = \big(R_{t-1}^* - R_{t-1}\frac{\mathcal{E}_{t-1}}{\mathcal{E}_t}\big)(-B_{t-1}^*)$,

$$\big(R_{t-1} B_{t-1} + \mathcal{E}_t R_{t-1}^* B_{t-1}^*\big) + \mathcal{E}_t \Pi_t^{\mathcal{I}*} = \mathcal{E}_t R_{t-1}^* \big(B_{t-1}^* + \frac{B_{t-1}}{\mathcal{E}_{t-1}}\big) = 0$$

Thus the world budget constraint is equivalent to the following natural form, which is trivial since the profits to financial intermediaries are transferred to the households:

$$P_t C_t + \mathcal{E}_t P_t^* C_t^* = P_{Ht} Y_t + \mathcal{E}_t P_{Ft}^* Y_t^*$$

Rearranging,

$$\underbrace{P_{Ht} Y_t - P_t C_t}_{NX_t} + \mathcal{E}_t \underbrace{\left(P_{Ft}^* Y_t^* - P_t^* C_t^* \right)}_{NX_t^*} = 0$$

This simply shows that the net export of this two country economy as a whole, or equivalently its net saving, is equal to zero, as it should. This is a confirmation of a version of the Walras' law applicable in this economy.

A.3 Steady State of the Ramsey Planner's Allocation

From FOC-$(\pi_t^W), (\pi_t^R)$,

$$\pi^W = \pi^R = 0$$

From FOC-(b_t),

$$\lambda = 0$$

From modified UIP constraint,

$$b = 0$$

From constraint-$\mathrm{PC}^W, \mathrm{PC}^R$,

$$mc^W = mc^R = 0$$

$$\widetilde{y}^W = 0$$

$$\Big(\frac{\sigma}{D}+\eta\Big)\widetilde{y}^R+\Big(\frac{D-2a+1}{D}W_y\Big)w=0$$

From budget constraint,

$$w-2W_y\widetilde{y}^R=0$$

$$\widetilde{y}^R=w=0$$

From FOC-$(\widetilde{y}^W_t),(\widetilde{y}^R_t)$,

$$\gamma^W=\gamma^R=0$$

The above result shows that the steady state of the planner's problem coincides with the efficient steady state. Together with the simplifying assumptions, the existence of intermediaries which replaces household's access to international financial market does not alter the steady state of this economy.

A.4 Unconditional and Conditional Expectations of the Loss Function

To find the unconditional and conditional expectations of the loss function derived in section 2.8, it requires to find $E[x_t^2+\beta x_{t+1}^2+\cdots]$ and $E_{t-1}[x_t^2+\beta x_{t+1}^2+\cdots]$, respectively, for each $x\in\{\widetilde{y}^R,w,\pi^R\}$.[2] Let s_t denote the vector of state variables in each model, e.g., $s_t\equiv(b_t,\lambda_t,\gamma^R_t)'$ under the Ramsey optimal policy. In addition, consider the impact of each shock one at a time: $\varepsilon_t\in\{a^R_t,\zeta^R_t,\mu^R_t,f_t\}$.

[2] To be precise, the planner maximizes the expectation at t after observing the shocks at t. Expectation at $t-1$ would be taking the same state variable b_{t-1} as given, while taking expectation over the realization of shocks at t. If the conditional expectation is taken at t, the only change is the slight decrease in the variance because it is evaluated after observing the shocks at t.

Evolution of the vector of state variables s_t as well as each of the variables in the loss function x_t can be expressed as:

$$s_t = B_b s_{t-1} + B_e \varepsilon_t$$

$$x_t = X_b s_{t-1} + X_e \varepsilon_t$$

where B_b is a square transition matrix, B_e is a column vector, X_b is a row vector, and X_e is a scalar. Let $E_{t-1}\varepsilon_t^2 = E\varepsilon_t^2 = \sigma^2$. Since all variables are log-linearized with respect to the deterministic steady state, means of all x_t and s_t (and of course ε_t) are zero. Also note that the shocks are i.i.d. (no persistence) and thus previous shocks are not counted as state variables, although it would be straightforward include these as a part of the state variables.

Unconditional expected loss through $\{x_t\}$ can be found as

$$E[x_t^2 + \beta x_{t+1}^2 + \cdots] = E[x_t^2 + \beta x_t^2 + \cdots] = \frac{1}{1-\beta} E[x_t^2]$$

$$\begin{aligned} E[x_t^2] &= E\big[(X_b s_{t-1} + X_e \varepsilon_t)(X_b s_{t-1} + X_e \varepsilon_t)'\big] \\ &= X_b E[s_t s_t'] X_b' + \sigma^2 X_e X_e' \end{aligned}$$

The unconditional variances of the state variables can be found using the above representation:

$$\begin{aligned} E[s_t s_t'] &= E\big[(B_b s_{t-1} + B_e \varepsilon_t)(B_b s_{t-1} + B_e \varepsilon_t)'\big] \\ &= B_b E[s_t s_t'] B_b' + \sigma^2 B_e B_e' \end{aligned}$$

This equation is precisely in the form of Lyapunov's equation ($X = FXF' + S$), which can be solved numerically after obtaining the coefficients B_b, B_e in the transition rule.[3]

[3]Of course, this is done automatically through dynare. It is well known that there exists a unique solution X if and only if the linear system $s_{t+1} = B_b s_t$ is globally asymptotically stable.

Denote the numerical solution to this equation as $E[s_t s_t'] \equiv V_s$, which is simply the unconditional variance-covariance matrix of the state variables. Then the relevant portion ($x \in \{\widetilde{y}^R, w, \pi^R\}$) of the unconditional expectation of the loss function

$$\frac{1}{1-\beta}E[x_t^2] = \frac{1}{1-\beta}\Big(X_b V_s X_b' + \sigma^2 X_e X_e'\Big)$$

can be obtained numerically.

Expected loss through x_t conditional on the information at $t-1$ is slightly more complicated. First, solving for each term $E_{t-1}x_{t+k}^2$ of the expected loss function separately:

$$\begin{aligned} E_{t-1}x_t^2 &= E_{t-1}\big[(X_b s_{t-1} + X_e \varepsilon_t)(X_b s_{t-1} + X_e \varepsilon_t)'\big] \\ &= X_b s_{t-1} s_{t-1}' X_b' + \sigma^2 X_e X_e' \end{aligned}$$

$$\begin{aligned} E_{t-1}x_{t+1}^2 &= E_{t-1}\big[(X_b^2 s_{t-1} + X_b X_e \varepsilon_t + X_e \varepsilon_{t+1})(X_b^2 s_{t-1} + X_b X_e \varepsilon_t + X_e \varepsilon_{t+1})'\big] \\ &= X_b^2 s_{t-1} s_{t-1}' X_b'^2 + \sigma^2\Big(X_b X_e X_e' X_b' + X_e X_e'\Big) \end{aligned}$$

$$\begin{aligned} E_{t-1}x_{t+2}^2 &= E_{t-1}\big[(X_b^3 s_{t-1} + X_b^2 X_e \varepsilon_t + X_b X_e \varepsilon_{t+1} + X_e \varepsilon_{t+2})\cdot \\ &\qquad (X_b^3 s_{t-1} + X_b^2 X_e \varepsilon_t + X_b X_e \varepsilon_{t+1} + X_e \varepsilon_{t+2})'\big] \\ &= X_b^3 s_{t-1} s_{t-1}' X_b'^3 + \sigma^2\Big(X_b^2 X_e X_e' X_b'^2 + X_b X_e X_e' X_b' + X_e X_e'\Big) \\ &\;\;\vdots \end{aligned}$$

Adding all these terms while discounting by β,

$$\begin{aligned} E_{t-1}\big[x_t^2 + \beta x_{t+1}^2 + \cdots\big] &= \Big(X_b s_{t-1} s_{t-1}' X_b' + \beta X_b^2 s_{t-1} s_{t-1}' X_b'^2 + \cdots\Big) \\ &\quad + \sigma^2\Big(X_e X_e'\big(1 + \beta + \cdots\big) + X_b X_e X_e' X_b'\big(\beta + \beta^2 + \cdots\big) + \cdots\Big) \\ &= \Big(X_b s_{t-1} s_{t-1}' X_b' + \beta X_b^2 s_{t-1} s_{t-1}' X_b'^2 + \cdots\Big) \\ &\quad + \frac{\sigma^2}{1-\beta}\Big(X_e X_e' + \beta X_b X_e X_e' X_b' + \beta^2 X_b^2 X_e X_e' X_b'^2 + \cdots\Big) \end{aligned}$$

In so far as we are interested in the perturbation from the deterministic steady state, it is natural to consider the case with $s_{t-1} = 0$, which is without loss of generality. The second bracket is obtained numerically by summing over the terms up to $T = 10^5$.[4] This is how I calculate the conditional epectations of the loss function in section 5.

The conditional and unconditional expectations are of course linked by the law of iterated expectation. Taking the unconditional expectation of the conditional expectation,

$$E\Big[E_{t-1}\big[x_t^2 + \beta x_{t+1}^2 + \cdots\big]\Big] = E\Big(X_b s_{t-1} s_{t-1}' X_b' + \beta X_b^2 s_{t-1} s_{t-1}' X_b'^2 + \cdots\Big) \\ + \frac{\sigma^2}{1-\beta}\Big(X_e X_e' + \beta X_b X_e X_e' X_b' + \beta^2 X_b^2 X_e X_e' X_b'^2 + \cdots\Big)$$

where the terms in the second bracket is now constant. Using $E\big[s_t s_t'\big] \equiv V_s$, the first bracket is simply

$$X_b V_s X_b' + \beta X_b^2 V_s X_b'^2 + \beta^2 X_b^3 V_s X_b'^3 + \cdots$$

Again, this expression is evaluated by numerically summing over the terms up to $T = 10^5$. I confirmed that the sum of these two expressions in each bracket are numerically identical to the unconditional expectation found above.

A.5 Consumption Equivalent Welfare Cost

The consumption equivalent welfare difference λ between some benchmark allocation and an arbitrary policy can be obtained as follows. The World welfare under the first best allocation and a policy x can be expressed as

$$V^{W,fb} = E_0 \sum_{t=0}^{\infty} \beta^t \Big(\frac{(C_t^{fb})^{1-\sigma} + (C_t^{*fb})^{1-\sigma}}{2(1-\sigma)} - \frac{(N_t^{fb})^{1+\eta} + (N_t^{*fb})^{1+\eta}}{2(1+\eta)} \Big)$$

[4]Increasing T to 10^6 did not result in any changes up to a meaningful precision.

$$
\begin{aligned}
V^{W,x} &= E_0 \sum_{t=0}^{\infty} \beta^t \Big(\frac{(C_t^x)^{1-\sigma} + (C_t^{*x})^{1-\sigma}}{2(1-\sigma)} - \frac{(N_t^x)^{1+\eta} + (N_t^{*x})^{1+\eta}}{2(1+\eta)} \Big) \\
&= E_0 \sum_{t=0}^{\infty} \beta^t \Big(\frac{(1-\lambda_x)^{1-\sigma}[(C_t^{fb})^{1-\sigma} + (C_t^{*fb})^{1-\sigma}]}{2(1-\sigma)} - \frac{(N_t^{fb})^{1+\eta} + (N_t^{*fb})^{1+\eta}}{2(1+\eta)} \Big) \\
&= V^{W,fb} - \Big(1 - (1-\lambda_x)^{1-\sigma}\Big) E_0 \sum_{t=0}^{\infty} \beta^t \frac{(C_t^{fb})^{1-\sigma} + (C_t^{*fb})^{1-\sigma}}{2(1-\sigma)} \\
&\approx V^{W,fb} - \Big(1 - (1-\lambda_x)^{1-\sigma}\Big) \frac{1}{(1-\beta)} \frac{\overline{C}^{1-\sigma}}{1-\sigma}
\end{aligned}
$$

Note that the level of steady state consumption $\overline{C}^{fb} = \overline{C}^{*fb} = \overline{C}$ would dominate all first order effect of shocks, and hence the above approximation holds for small shocks. Recall that the loss function was previously defined as

$$
\mathcal{L}^x = V^{W,fb} - V^{W,x} = \frac{\overline{C}^{1-\sigma}}{2} E_0 \sum_{t=0}^{\infty} \beta^t \Psi_t
$$

where the welfare loss in each period was derived as

$$
\Psi_t = \underbrace{(\sigma+\eta)(\widetilde{y}_t^W)^2 + \frac{\epsilon}{\delta}(\pi_t^W)^2}_{\Psi_t^W} + \underbrace{(\frac{\sigma}{D} + \eta)(\widetilde{y}_t^R)^2 + \frac{a(1-a)\phi}{D}(w_t)^2 + \frac{\epsilon}{\delta}(\pi_t^R)^2}_{\Psi_t^R}
$$

Combining with the above expression for V^x,

$$
\lambda_x = 1 - \Big[1 - \frac{1-\sigma}{2}(1-\beta)\Big(E_0 \sum_{t=0}^{\infty} \beta^t \Psi_t\Big)\Big]^{\frac{1}{1-\sigma}}
$$

That is, the consumption-equivalent welfare difference with respect to the first best allocation can be readily obtained from the loss function. This value of λ_x as defined above corresponds to the *conditional* welfare difference. The *unconditional* welfare difference can be constructed similarly, where the conditional expectation on the right hand side is simply replaced by the unconditional expectation. If $\sigma = 1$,

$$
\lambda_x = 1 - \exp\Big(- \frac{1-\beta}{2} E_0 \sum_{t=0}^{\infty} \beta^t \Psi_t \Big)
$$

For both cases, if $(1-\beta)E_0 \sum_{t=0}^{\infty} \beta^t \Psi_t$ is small,

$$\lambda_x \approx \frac{1-\beta}{2} E_0 \sum_{t=0}^{\infty} \beta^t \Psi_t$$

In addition, the unconditional welfare loss can be expressed simply as

$$\lambda_x^u \approx \frac{1}{2} E \Psi_t$$

A.6 Derivation of analytical solutions

In this section, analytical solutions for the equilibrium and the optimal policy are derived, for the cases of: (1) Flexible price allocation — (a) $\chi = 0$, (b) $\chi > 0$; (2) Optimal monetary policy — (a) $\chi = 0$, (b) $\chi > 0, \alpha = 1$ (fully sticky prices), (c) $\chi > 0$, $\alpha < 1$; (3) Optimal Capital Control (Independent of χ).

I. (a) Flexible Price Allocation under $\chi = 0$

The flexible price equilibrium in relative variables can be characterized by the following set of equations in gap form:

$$\text{(Market Clearing): } \widetilde{y}_t^R = (2a-1)\widetilde{c}_t^R + 2a(1-a)\phi\widetilde{s}_t$$

$$\text{(Definition of } w_t\text{): } \widetilde{w}_t \equiv 2\sigma\widetilde{c}_t^R - (2a-1)\widetilde{s}_t$$

$$\text{(modified UIP): } E_t\widetilde{w}_{t+1} - \widetilde{w}_t = -\chi b_t + f_t = f_t$$

$$\text{(Budget Constraint): } y_t - c_t - (1-a)s_t = nx_t = b_t - \beta^{-1}b_{t-1}$$

$$\text{(Flexible Prices): } \sigma\widetilde{c}_t^R + \eta\widetilde{y}_t^R + (1-a)\widetilde{s}_t + \mu_t^R = 0$$

The first step is to express $\widetilde{c}_t^R$ and $\widetilde{s}_t$ in terms of $\widetilde{y}_t^R$ and $\widetilde{w}_t$, which turns out to be useful. This can be easily done by inverting the 2 by 2 matrix implicit in the (Market Clearing) and (Definition of w_t):

$$\widetilde{c}_t^R = \frac{2a-1}{D}\widetilde{y}_t^R + \frac{2a(1-a)\phi}{D}\widetilde{w}_t$$

$$\widetilde{s}_t = \frac{2\sigma}{D}\widetilde{y}_t^R - \frac{2a-1}{D}\widetilde{w}_t$$

where again $D \equiv 4a(1-a)\sigma\phi + (2a-1)^2 = 4a(1-a)(\sigma\phi - 1) + 1$ is equal to the determinant of the aforementioned 2 by 2 matrix.

Substituting these expressions in the Flexible Price condition,

$$\widetilde{y}_t^R = -\frac{D-2a+1}{2(\sigma+\eta D)}\widetilde{w}_t - \frac{D}{\sigma+\eta D}\mu_t^R$$

Now express the net export in gap form:

$$\begin{aligned}\widetilde{nx}_t &= \widetilde{y}_t^R - \widetilde{c}_t^R - (1-a)\widetilde{s}_t \\ &= \underbrace{\frac{(D-2a+1)-(1-a)2\sigma}{D}}_{\frac{2W_y}{W_b}}\widetilde{y}_t^R - \underbrace{\frac{(1-a)(2a\phi-2a+1)}{D}}_{\frac{1}{W_b}}\widetilde{w}_t\end{aligned}$$

where $W_b \equiv \frac{D}{1-a}\frac{1}{2a(\phi-1)+1}$, and $W_y \equiv \frac{2a(\sigma\phi-1)+1-\sigma}{2a(\phi-1)+1}$.[5] In addition, combining with the above expression from the flexible price condition,

$$\begin{aligned}\widetilde{nx}_t &= \frac{2W_y}{W_b}\widetilde{y}_t^R - \frac{1}{W_b}\widetilde{w}_t = -\underbrace{\frac{1}{W_b}\left(1+\frac{W_y(D-2a+1)}{\sigma+\eta D}\right)}_{\equiv A}\widetilde{w}_t - \frac{2W_yD}{W_b(\sigma+\eta D)}\mu_t^R \\ &= nx_t - nx_t^{fb} = b_t - \beta^{-1}b_{t-1} - nx_t^{fb}\end{aligned}$$

[5] As a result, $\widetilde{w}_t = 2W_y\widetilde{y}_t^R - W_b\widetilde{nx}_t = 2W_y\widetilde{y}_t^R - W_b\big(b_t - \beta^{-1}b_{t-1} - nx_t^{fb}\big)$.

Now consider the modified UIP condition. Assuming for simplicity that f_t follows an AR(1) process ($E_t f_{t+1} = \rho_f f_t$),

$$\begin{aligned}
E_t\widetilde{w}_{t+1} &= \widetilde{w}_t + f_t \\
E_t\widetilde{w}_{t+2} &= \widetilde{w}_t + f_t + E_t f_{t+1} = \widetilde{w}_t + (1+\rho_f) f_t \\
&\vdots \\
\lim_{k\to\infty} E_t\widetilde{w}_{t+k} \equiv \overline{w}_t &= \widetilde{w}_t + \frac{1}{1-\rho_f} f_t
\end{aligned}$$

Given $\overline{w}_t$, it is straightforward to show that

$$E_t\widetilde{w}_{t+k} = \overline{w}_t - \frac{\rho_f^k}{1-\rho_f} f_t, \ \forall k \geq 0$$

Using this, all $E_t\widetilde{w}_{t+k}$ can be written in terms of $\overline{w}_t$ and the current shock f_t. Adding up the budget constraint side by side, discounting by β to cancel out b_t's, and making use of the transversality condition,

$$\begin{aligned}
E_t \sum_{k=0}^{\infty} \beta^k \widetilde{nx}_{t+k} &= -\frac{A}{(1-\beta)}\overline{w}_t + \frac{A}{(1-\rho_f)(1-\beta\rho_f)} f_t - \frac{2W_y D}{W_b(\sigma+\eta D)} E_t \sum_{k=0}^{\infty} \beta^k \mu_{t+k}^R \\
&= -\beta^{-1} b_{t-1} - E_t \sum_{k=0}^{\infty} \beta^k nx_{t+k}^{fb} \\
&= -\beta^{-1} b_{t-1} - \frac{1+\eta}{\frac{\sigma}{D}+\eta} \frac{2W_y}{W_b} E_t \sum_{k=0}^{\infty} \beta^k a_{t+k}^R
\end{aligned}$$

$$\therefore \overline{w}_t = \frac{1-\beta}{(1-\rho_f)(1-\beta\rho_f)} f_t + \frac{1-\beta}{A}\left[\beta^{-1} b_{t-1} + E_t \sum_{k=0}^{\infty} \beta^k nx_{t+k}^{fb} - \frac{2W_y D}{W_b(\sigma+\eta D)} E_t \sum_{k=0}^{\infty} \beta^k \mu_{t+k}^R\right]$$

where A is the constant ratio between $\widetilde{nx}_t$ and $\widetilde{w}_t$. Consequently, all $E_t\widetilde{w}_{t+k}$ as well as $E_t\widetilde{y}_{t+k}^R$ can be obtained as

$$\begin{aligned}
E_t\widetilde{w}_{t+k} &= \overline{w}_t - \frac{\rho_f^k}{1-\rho_f} f_t \\
E_t\widetilde{y}_{t+k}^R &= -\frac{D-2a+1}{2(\sigma+\eta D)} E_t\widetilde{w}_{t+k} - \frac{D}{\sigma+\eta D} E_t \mu_{t+k}
\end{aligned}$$

I. (b) Flexible Price Allocation under $\chi > 0$

The flexible price equilibrium under imperfect capital mobility ($\chi > 0$) is characterized by the following equations:

$$\text{(Market Clearing): } \widetilde{y}_t^R = (2a-1)\widetilde{c}_t^R + 2a(1-a)\phi\widetilde{s}_t$$

$$\text{(Definition of } w_t) : \widetilde{w}_t \equiv 2\sigma\widetilde{c}_t^R - (2a-1)\widetilde{s}_t$$

$$\text{(Modified UIP): } E_t\Delta\widetilde{w}_{t+1} = -\chi b_t + f_t$$

$$\text{(BC): } y_t - c_t - (1-a)s_t = nx_t = b_t - \beta^{-1}b_{t-1}$$

$$\text{(Flexible Prices): } \sigma\widetilde{c}_t^R + \eta\widetilde{y}_t^R + (1-a)\widetilde{s}_t + \mu_t^R = 0$$

In order to obtain an analytic solution, it is assumed that all shocks follow AR(1) process: $E_t nx_{t+1} = \rho_a nx_t$, $E_t\mu_{t+1} = \rho_\mu\mu_t$, $E_t f_{t+1} = \rho_f f_t$.

The following equations are the same as under $\chi = 0$:

$$\widetilde{y}_t^R = -\frac{D-2a+1}{2(\sigma+\eta D)}\widetilde{w}_t - \frac{D}{\sigma+\eta D}\mu_t^R$$

$$\widetilde{nx}_t = \frac{2W_y}{W_b}\widetilde{y}_t^R - \frac{1}{W_b}\widetilde{w}_t = -\underbrace{\frac{1}{W_b}\left(1 + \frac{W_y(D-2a+1)}{\sigma+\eta D}\right)}_{\equiv A}\widetilde{w}_t - \frac{2W_y}{W_b}\frac{D}{\sigma+\eta D}\mu_t^R$$

However, with $\chi > 0$,

$$\begin{aligned}\widetilde{nx}_t = -A\widetilde{w}_t &= b_t - \beta^{-1}b_{t-1} - nx_t^{fb} \\ &= \frac{1}{\chi}(w_t - w_{t+1}) - \beta^{-1}b_{t-1} - nx_t^{fb}\end{aligned}$$

And for $k \geq 1$,

$$\begin{aligned}\widetilde{nx}_{t+k} = -A\widetilde{w}_{t+k} &= b_{t+k} - \beta^{-1}b_{t+k-1} \\ &= \frac{1}{\chi}\Big((w_{t+k} - w_{t+k+1}) - \beta^{-1}(w_{t+k-1} - w_{t+k})\Big)\end{aligned}$$

Rearranging,

$$\widetilde{w}_{t+1} - (1 + A\chi)\widetilde{w}_t = -\chi n x_t^{fb}$$

$$\widetilde{w}_{t+2} - (1 + \beta^{-1} + A\chi)\widetilde{w}_{t+1} + \beta^{-1}\widetilde{w}_t = 0$$

$$\vdots$$

Define the characteristic equation

$$\nu^2 - (1 + \beta^{-1} + A\chi)\nu + \beta^{-1} = 0$$

and its solutions $0 < \nu_1 < 1 < \beta^{-1} < \nu_2$. Then,

$$(\widetilde{w}_{t+2} - \nu_1\widetilde{w}_{t+1}) - \nu_2(\widetilde{w}_{t+1} - \nu_1\widetilde{w}_t) = 0$$

$$\nu_2^{-1}(\widetilde{w}_{t+3} - \nu_1\widetilde{w}_{t+2}) - (\widetilde{w}_{t+2} - \nu_1\widetilde{w}_{t+1}) = 0$$

$$\vdots$$

$$\therefore (\widetilde{w}_{t+1} - \nu_1\widetilde{w}_t) = \nu_2^{-1}(\widetilde{w}_{t+2} - \nu_1\widetilde{w}_{t+1}) = \cdots = 0$$

That is,

$$E_t\widetilde{w}_{t+k} = \nu_1^k\widetilde{w}_t, \ \forall k \geq 0$$

Finally, using the budget constraint at t, and noting that $1 + A\chi = \nu_1 + \nu_2 - \beta^{-1}$,

$$\nu_1\widetilde{w}_t - (\nu_1 + \nu_2 - \beta^{-1})\widetilde{w}_t = -(\nu_2 - \beta^{-1})\widetilde{w}_t = -\chi n x_t^{fb}$$

$$\widetilde{w}_t = \frac{\chi}{\nu_2 - \beta^{-1}} n x_t^{fb}$$

As $\chi \to 0$, $(\nu_2 - \beta^{-1}) \to 0$. Through a simple algebra, it can be shown that

$$\lim_{\chi \to 0} \frac{\nu_2 - \beta^{-1}}{\chi} = \frac{A}{(1 - \beta)}$$

Thus, as $\chi \to 0$,

$$\widetilde{w}_t = \frac{\chi}{\nu_2 - \beta^{-1}} n x_t^{fb} \to \frac{(1 - \beta)}{A} n x_t^{fb}$$

which indeed coincides with the solution found for the case of $\chi = 0$.

II. (a) Optimal Monetary Policy under $\chi = 0$

There are seven equations that pin down seven variables. The first two equations involve permanent expected levels for two of the variables, while these levels would later need to be determined endogenously:

$$\text{(M-UIP)} : E_t\widetilde{w}_{t+1} - \widetilde{w}_t = -\chi b_t + f_t = f_t$$

$$\text{FOC-}(b_t) : \xi_t = E_t\xi_{t+1} = \overline{\xi}_t$$

For simplicity, assume that $f_t \sim \text{AR}(1)$: $E_t f_{t+1} = \rho f_t$. Then

$$\text{(M-UIP)} : E_t\widetilde{w}_{t+k} = \overline{w}_t - \frac{\rho^k}{1-\rho} f_t, \ \forall k \geq 0$$

Next equation expresses $\widetilde{y}_t^R$ in terms of γ_t^R as well as $\overline{\xi}_t$:

$$\text{FOC-}(\widetilde{y}_t^R) : \widetilde{y}_t^R = \delta\gamma_t^R + \frac{2W_y}{\frac{\sigma}{D} + \eta}\overline{\xi}_t$$

Next two equations can be used to simply cancel out b_t and λ_t, leaving just one lifetime constraint for each. These are then used to pin down the values of $\overline{c}_t^R$ and $\overline{\xi}_t$:

$$\text{FOC-}(w_t) : \frac{a(1-a)\phi}{D}\widetilde{w}_t + \overline{\xi}_t = \frac{D-2a+1}{2D}\delta\gamma_t^R - (\lambda_t - \beta^{-1}\lambda_{t-1})$$

$$\text{(BC)} : \widetilde{w}_t - 2W_y\widetilde{y}_t^R + W_b(b_t - \beta^{-1}b_{t-1} - nx_t^{fb}) = 0$$

Lastly, the two remaining equations related to inflation can be used to solve for γ_t^R:

$$\text{FOC-}(\pi_t^R) : \pi_t^R = -\frac{\delta}{\epsilon}\Delta\gamma_t^R$$

$$\text{(PC)} : \pi_t^R - \beta\pi_{t+1}^R = \delta\Big((\frac{\sigma}{D} + \eta)\widetilde{y}_t^R + \frac{D-2a+1}{2D}\widetilde{w}_t + \mu_t^R\Big)$$

Again, for simplicity, assume that $\mu_t^R \sim \text{AR}(1)$: $E_t\mu_{t+1}^R = \rho_\mu\mu_t^R$.

Now proceed to simplifying this system of equations. From FOC-(w_t),

$$\frac{D-2a+1}{2D}\delta E_t\sum_{k=0}^{\infty}\beta^k\gamma_{t+k} = \frac{1}{1-\beta}\Big(\frac{a(1-a)\phi}{D}\overline{w}_t + \overline{\xi}_t\Big) - \frac{a(1-a)\phi}{D}\frac{1}{(1-\rho)(1-\beta\rho)}f_t - \beta^{-1}\lambda_{t-1}$$

From (BC),

$$2W_y\sum_{k=0}^{\infty}\beta^k\widetilde{y}^R_{t+k} = \frac{1}{1-\beta}\overline{w}_t - \frac{1}{(1-\rho)(1-\beta\rho)}f_t - W_b\big(E_t\sum_{k=0}^{\infty}\beta^k nx^{fb}_{t+k} + \beta^{-1}b_{t-1}\big)$$

Using FOC-($\widetilde{y}^R_t$) to express $\widetilde{y}^R_t$ in terms of γ^R_t,

$$2W_y\delta E_t\sum_{k=0}^{\infty}\beta^k\gamma_{t+k} + \frac{1}{1-\beta}\frac{(2W_y)^2}{\frac{\sigma}{D}+\eta}\overline{\xi}_t = \frac{1}{1-\beta}\overline{w}_t - \frac{1}{(1-\rho)(1-\beta\rho)}f_t - W_b\big(E_t\sum_{k=0}^{\infty}\beta^k nx^{fb}_{t+k} + \beta^{-1}b_{t-1}\big)$$

Combining the above two equations,

$$\begin{aligned}
&(1-\beta)\frac{D-2a+1}{2D}\delta E_t\sum_{k=0}^{\infty}\beta^k\gamma_{t+k}\\
&= \Big(\frac{a(1-a)\phi}{D}\overline{w}_t + \overline{\xi}_t\Big) - \frac{a(1-a)\phi}{D}\frac{1-\beta}{(1-\rho)(1-\beta\rho)}f_t - (1-\beta)\beta^{-1}\lambda_{t-1}\\
&= \frac{D-2a+1}{(2D)(2W_y)}\Big(\overline{w}_t - \frac{(2W_y)^2}{\frac{\sigma}{D}+\eta}\overline{\xi}_t - \frac{1-\beta}{(1-\rho)(1-\beta\rho)}f_t\\
&\qquad - (1-\beta)W_b\big(E_t\sum_{k=0}^{\infty}\beta^k nx^{fb}_{t+k} + \beta^{-1}b_{t-1}\big)\Big)
\end{aligned}$$

Rearranging,

$$\underbrace{\big(1+\frac{W_y(D-2a+1)}{\sigma+\eta D}\big)}_{W_bA}\overline{\xi}_t = \big(\frac{D-2a+1}{4W_yD} - \frac{a(1-a)\phi}{D}\big)\big(\overline{w}_t - \frac{1-\beta}{(1-\rho)(1-\beta\rho)}f_t\big) + (1-\beta)\beta^{-1}\lambda_{t-1} - \frac{W_b(D-2a+1)}{4W_yD}(1-\beta)\big(E_t\sum_{k=0}^{\infty}\beta^k nx^{fb}_{t+k} + \beta^{-1}b_{t-1}\big)$$

Now, combining the two equations regarding inflation,

$$\beta\gamma^R_{t+1} - (1+\beta)\gamma^R_t + \gamma^R_{t-1} = \epsilon\Big[(\frac{\sigma}{D}+\eta)\delta\gamma^R_t + 2W_y\bar{\xi}_t + \frac{D-2a+1}{2D}\big(\overline{w}_t - \frac{f_t}{1-\rho}\big) + \mu^R_t\Big]$$

$$\gamma^R_{t+1} - \big\{1+\beta^{-1}\big[1+\epsilon\delta(\frac{\sigma}{D}+\eta)\big]\big\}\gamma^R_t + \beta^{-1}\gamma^R_{t-1} = \beta^{-1}\epsilon\Big[2W_y\bar{\xi}_t + \frac{D-2a+1}{2D}\big(\overline{w}_t - \frac{f_t}{1-\rho}\big) + \mu^R_t\Big]$$

Define the characteristic equation

$$\nu^2 - \big(1+\beta^{-1}+\frac{\epsilon\delta}{\beta}(\frac{\sigma}{D}+\eta)\big)\nu + \beta^{-1} = 0$$

and its solutions $0 < \nu_1 < 1 < \beta^{-1} < \nu_2$. Then,

$$(\gamma^R_{t+1} - \nu_1\gamma^R_t) - \nu_2(\gamma^R_t - \nu_1\gamma^R_{t-1}) = \beta^{-1}\epsilon\Big[2W_y\bar{\xi}_t + \frac{D-2a+1}{2D}\big(\overline{w}_t - \frac{1}{1-\rho}f_t\big) + \mu^R_t\Big]$$

$$\nu_2^{-1}(\gamma^R_{t+2} - \nu_1\gamma^R_{t+1}) - (\gamma^R_{t+1} - \nu_1\gamma^R_t) = \nu_2^{-1}\beta^{-1}\epsilon\Big[2W_y\bar{\xi}_t + \frac{D-2a+1}{2D}\big(\overline{w}_t - \frac{\rho}{1-\rho}f_t\big) + \rho_\mu\mu^R_t\Big]$$

$$\vdots$$

Define

$$\overline{z}_t \equiv \frac{\nu_2^{-1}}{1-\nu_2^{-1}}\beta^{-1}\epsilon\Big[2W_y\bar{\xi}_t + \frac{D-2a+1}{2D}\overline{w}_t\Big]$$

Then

$$E_t(\gamma^R_{t+k} - \nu_1\gamma^R_{t+k-1}) = -\overline{z}_t + \beta^{-1}\epsilon\frac{D-2a+1}{2D(1-\rho)}\frac{\nu_2^{-1}\rho^k}{1-\nu_2^{-1}\rho}f_t - \beta^{-1}\epsilon\frac{\nu_2^{-1}\rho_\mu^k}{1-\nu_2^{-1}\rho_\mu}\mu^R_t$$

Using $\gamma^R_{t+1} - \nu_1\gamma^R_t = \overline{z}_t = \frac{1}{1-\nu_1}\overline{z}_t - \frac{\nu_1}{1-\nu_1}\overline{z}_t$,

$$\begin{aligned}
E_t\gamma^R_{t+k} - \frac{1}{1-\nu_1}\overline{z}_t &= \nu_1^{k+1}\big(\gamma^R_{t-1} - \frac{1}{1-\nu_1}\overline{z}_t\big) + [...]\big(\nu_1^k + \nu_1^{k-1}\rho + \cdots + \rho^k\big)f_t \\
&\quad + [...]\big(\nu_1^k + \cdots + \rho_\mu^k\big)\mu^R_t \\
&= \nu_1^{k+1}\big(\gamma^R_{t-1} - \frac{1}{1-\nu_1}\overline{z}_t\big) + \beta^{-1}\epsilon\frac{D-2a+1}{2D(1-\rho)}\frac{1}{\nu_2-\rho}\frac{\nu_1^{k+1}-\rho^{k+1}}{\nu_1-\rho}f_t \\
&\quad - \beta^{-1}\epsilon\frac{1}{\nu_2-\rho_\mu}\frac{\nu_1^{k+1}-\rho_\mu^{k+1}}{\nu_1-\rho_\mu}\mu^R_t
\end{aligned}$$

$$\begin{aligned}\therefore E_t\gamma^R_{t+k} &= -\frac{1-\nu_1^{k+1}}{1-\nu_1}\overline{z}_t + \nu_1^{k+1}\gamma^R_{t-1} + [...]f_t + [...]\mu^R_t \\ &= -\frac{1-\nu_1^{k+1}}{(1-\nu_1)(\nu_2-1)}\beta^{-1}\epsilon\Big[2W_y\overline{\xi}_t + \frac{D-2a+1}{2D}\overline{w}_t\Big] + \nu_1^{k+1}\gamma^R_{t-1} + [...]f_t + [...]\mu^R_t\end{aligned}$$

Using $\nu_1+\nu_2 = 1+\beta^{-1}+\frac{\epsilon\delta}{\beta}(\frac{\sigma}{D}+\eta)$ and $\nu_1\nu_2 = \beta^{-1}$, $(1-\nu_1)(\nu_2-1) = \frac{\epsilon\delta}{\beta}(\frac{\sigma}{D}+\eta)$. Therefore,

$$E_t\gamma^R_{t+k} = -\frac{(1-\nu_1^{k+1})}{\delta(\frac{\sigma}{D}+\eta)}\Big[2W_y\overline{\xi}_t + \frac{D-2a+1}{2D}\overline{w}_t\Big] + \nu_1^{k+1}\gamma^R_{t-1} + [...]f_t + [...]\mu^R_t$$

Now, to find the value of the present discounted sum of γ^R_{t+k},

$$\sum_{k=0}^{\infty}\beta^k(1-\nu_1^{k+1}) = \underbrace{\big(1+\beta+\cdots\big)}_{\frac{1}{1-\beta}} - \underbrace{\nu_1\big(1+\beta\nu_1+\cdots\big)}_{\frac{\nu_1}{1-\beta\nu_1}} = \frac{1-\nu_1}{(1-\beta)(1-\beta\nu_1)}$$

$$\sum_{k=0}^{\infty}\beta^k(\nu_1^{k+1}-\rho^{k+1}) = \underbrace{\nu_1\big(1+\beta\nu_1+\cdots\big)}_{\frac{\nu_1}{1-\beta\nu_1}} - \underbrace{\rho\big(1+\beta\rho+\cdots\big)}_{\frac{\rho}{1-\beta\rho}} = \frac{\nu_1-\rho}{(1-\beta\nu_1)(1-\beta\rho)}$$

$$\begin{aligned}\therefore E_t\sum_{k=0}^{\infty}\beta^k\gamma^R_{t+k} &= -\frac{1-\nu_1}{(1-\beta)(1-\beta\nu_1)}\frac{1}{\delta(\frac{\sigma}{D}+\eta)}\Big[2W_y\overline{\xi}_t + \frac{D-2a+1}{2D}\overline{w}_t\Big] + \frac{\nu_1}{1-\beta\nu_1}\gamma^R_{t-1} \\ &+\beta^{-1}\epsilon\frac{D-2a+1}{2D(1-\rho)(\nu_2-\rho)}\frac{1}{(1-\beta\nu_1)(1-\beta\rho)}f_t - \beta^{-1}\epsilon\frac{1}{\nu_2-\rho_\mu}\frac{1}{(1-\beta\nu_1)(1-\beta\rho_\mu)}\mu^R_t\end{aligned}$$

Now we have constructed two equations that contains two unknown variables $\overline{c}^R_t, \overline{\xi}_t$ to be solved in terms of exogenous shock a^R_t, and the state variables (Lagrange multipliers in the previous period) that represent precommitments. For better readability, assume $f_t = \mu^R_t = 0$, although introducing both shocks is trivial. Combining with the equations derived previously,

$$\begin{aligned}-\frac{1-\nu_1}{1-\beta\nu_1}\frac{1}{\frac{\sigma}{D}+\eta}\Big[2W_y\overline{\xi}_t + \frac{D-2a+1}{2D}\overline{w}_t + \frac{\nu_1(1-\beta)}{1-\beta\nu_1}\delta\gamma^R_{t-1} \\ = \frac{2D}{D-2a+1}\Big[\frac{a(1-a)\phi}{D}\overline{w}_t + \overline{\xi}_t - (1-\beta)\beta^{-1}\lambda_{t-1}\Big]\end{aligned}$$

Plugging in the above expression for $\overline{\xi}_t$:

$$2W_bA\cdot\overline{\xi}_t=\Big(\frac{D-2a+1}{2W_yD}-\frac{2a(1-a)\phi}{D}\Big)\overline{w}_t+[...]$$

and rearranging, we can finally obtain the full analytic solution for $\overline{w}_t$:

$$\begin{aligned}D^{rp}\cdot\overline{w}_t=&W_b\Big(\frac{1}{W_y}+\frac{1-\nu_1}{1-\beta\nu_1}\frac{D-2a+1}{\sigma+\eta D}\Big)(1-\beta)\big(nx_t^{fb}+\beta^{-1}b_{t-1}\big)\\&+\frac{\nu_1(1-\beta)}{1-\beta\nu_1}\Big[\frac{2W_y}{\frac{\sigma}{D}+\eta}2(1-\beta)\beta^{-1}\lambda_{t-1}-2\big(1+\frac{W_y(D-2a+1)}{\sigma+\eta D}\big)\delta\gamma_{t-1}^R\Big]\end{aligned}$$

where $D^{rp}\equiv\Big(\frac{1}{W_y}+\frac{4a(1-a)\phi W_y}{\sigma+\eta D}\Big)+\frac{1-\nu_1}{1-\beta\nu_1}\Big(W_y\big(\frac{D-2a+1}{\sigma+\eta D}\big)^2+2\frac{D-2a+1}{\sigma+\eta D}-\frac{4a(1-a)\phi W_y}{D-2a+1}\Big)$. Given this solution for $\overline{w}_t$, all other variables can also be solved analytically.

II. (b) Imperfect capital mobility ($\chi>0$), Fully Sticky Prices ($\alpha=1$)

Having characterized the fully general solution under $\chi=0$, now consider the imperfect capital mobility case with $\chi>0$, first with fully sticky price. The first order necessary conditions are:

$$(\widetilde{y}_t^R):\widetilde{y}_t^R=\frac{2W_y}{\frac{\sigma}{D}+\eta}\xi_t$$

$$(w_t):\frac{a(1-a)\phi}{D}w_t+\big(\lambda_t-\beta^{-1}\lambda_{t-1}\big)+\xi_t=0$$

$$(b_t):W_b\big(\xi_t-E_t\xi_{t+1}\big)-\chi\lambda_t=0$$

$$(\text{M-UIP}):w_t-E_tw_{t+1}-\chi b_t+f_t=0$$

$$(\text{BC}):w_t-2W_y\widetilde{y}_t^R+W_b\big(b_t-\beta^{-1}b_{t-1}-nx_t^{fb}\big)=0$$

Using FOC-(b_t) and modified UIP to substitute out λ_t and b_t in the equations FOC-(w_t) and (BC) respectively, using FOC-($\widetilde{y}_t^R$) to convert all $\{\xi_t\}$ into $\{\widetilde{y}_t^R\}$, and assuming that the all shocks are iid,

$$\widetilde{y}_{t+2}^R-\big(1+\beta^{-1}+\frac{\chi}{W_b}\big)\widetilde{y}_{t+1}^R+\beta^{-1}\widetilde{y}_t^R=\frac{a_H(1-a_H)\phi}{\sigma+\eta D}\frac{2W_y\chi}{W_b}w_{t+1}$$

$$w_{t+2} - (1 + \beta^{-1} + \frac{\chi}{W_b})w_{t+1} + \beta^{-1}w_t = -\frac{2W_y\chi}{W_b}\widetilde{y}^R_{t+1} - \beta^{-1}f_t$$

where expectation at t is implicit for all forward looking variables (and for all equations below, unless otherwise specified). This system of equations can be simplified analytically because the two variables can be described using a common characteristic equation. Multiplying the second equation by $j \cdot \sqrt{\frac{a_H(1-a_H)\phi}{\sigma+\eta D}} \equiv j \cdot \kappa$ ($j^2 = -1$) and adding side by side, and defining $z_t \equiv \widetilde{y}^R_t + j\kappa w_t$,

$$z_{t+2} - \big(1 + \beta^{-1} + \frac{\chi}{W_b} - j\kappa\frac{2W_y\chi}{W_b}\big)z_{t+1} + \beta^{-1}z_t = -j\kappa\beta^{-1}f_t$$

The characteristic equation for this second-order expectational difference equation is the quadratic equation $\nu^2 - \big(1 + \beta^{-1} + \frac{\chi}{W_b} - j\kappa\frac{2W_y\chi}{W_b}\big)\nu + \beta^{-1} = 0$, which contains a complex coefficient. Denote the two roots as ν_1, ν_2, where $|\nu_1| < 1 < \beta^{-1} < |\nu_2|$. The above equation at t can be written as

$$\big(z_{t+2} - \nu_1 z_{t+1}\big) - \nu_2\big(z_{t+1} - \nu_1 z_t\big) = -j\kappa\beta^{-1}f_t$$

The corresponding equations for future periods in expectation at t, multiplied by multiples of ν_2^{-1}, is

$$\nu_2^{-1}\big(z_{t+3} - \nu_1 z_{t+2}\big) - \big(z_{t+2} - \nu_1 z_{t+1}\big) = -\nu_2^{-1} \cdot j\kappa\beta^{-1}E_t f_{t+1} = 0$$

$$\vdots$$

Since the right hand sides of all these equations are zero, it follows that

$$\big(z_{t+2} - \nu_1 z_{t+1}\big) = \nu_2^{-1}\big(z_{t+3} - \nu_1 z_{t+2}\big) = \cdots = 0$$

$$\therefore z_{t+k} = \nu_1^{k-1}z_{t+1}$$

Back to the equation at t,

$$-\nu_2\big(z_{t+1}-\nu_1 z_t\big) = -j\kappa\beta^{-1}f_t$$

Since $\nu_1\nu_2=\beta^{-1}$, it follows that $z_{t+1}=\nu_1\big(z_t+j\kappa f_t\big)$.

The above derivation used equations FOC-(w_t) and (BC) at $t+1$ and onward, but not at t. The problem is slightly different at t because the state variables b_{t-1},λ_{t-1} are taken as given. The corresponding equation using z_t at t is

$$\begin{aligned}
z_{t+1}-\big(1+\frac{\chi}{W_b}+j\kappa\frac{2W_y\chi}{W_b}\big)z_t &= -\big(1+\frac{\chi}{W_b}+j\kappa\frac{2W_y\chi}{W_b}-\nu_1\big)z_t \\
&= -\frac{2W_y\chi}{W_b}\frac{1}{\frac{\sigma}{D}+\eta}\beta^{-1}\lambda_{t-1}-j\kappa\big(\chi\beta^{-1}b_{t-1}+\chi n x_t^{fb}-f_t\big)
\end{aligned}$$

The right hand side is simply the sum of functions of state variables at $t-1$ and the shocks at t. When multiplied by the reciprocal of the complex coefficient on z_t, the real and imagenary components pin down the values of $\widetilde{y}_t^R$ and w_t respectively.

By construction, z_t contains all the relevant information in the variables $(\widetilde{y}_t^R, w_t)$.[6] κ, the ratio between the coefficients for $\widetilde{y}_t^R$ and w_t in the definition of z_t, is in accordance with the relative weight of these terms in the loss function, which is natural. $E_t z_{t+k+1}=\nu_1 E_t z_{t+k}$ implies that the optimal trade-off involves an exponential decay in $|z_t|$ by a factor of $|\nu_1|<1$ in each period. The rate of decay $|\nu_1|$ ranges from 1 when $\chi=0$ to 0 when $\chi\to\infty$, which is consistent with the intuition that the impact of a shock is more persistent or closer to permanent (random walk) when χ is smaller, and vice versa. In addition, there is an oscillatory component that pertains to the angle of ν_1. This angle ranges from zero when $\chi=0$ to $\arctan(2W_y\kappa)\in(0,\pi/2)$ when $\chi\to\infty$. Note that

[6] Here the matrix $\begin{pmatrix}0 & -1\\ \kappa^2 & 0\end{pmatrix}$ has two eigenvalues $\pm j\kappa$ that are complex conjugate with each other. Clearly, the information contained in the complex conjugate of z_t is the same as what is in z_t itself.

$2W_y\kappa$ is independent of χ, and takes a relatively small value (≈ 0.3) under the baseline calibration, but it can take a larger value when η is small, ϕ is large, or σ is large.

II. (c) Imperfect capital mobility ($\chi > 0$), General Degree of Price Stickiness ($\alpha < 1$)

Finally, consider the general degree of price stickiness under imperfect capital mobility. The solution approach is similar to the case of fully sticky price, but the analytic solution will be significantly more complicated because the coefficients that were same in the previous case, which led to a huge simplification, do not coincide in this general case. The first order necessary conditions for the Relative variables are:

$$\text{FOC-}(\widetilde{y}_t^R) : \widetilde{y}_t^R - \delta\gamma_t^R = \frac{2W_y}{\frac{\sigma}{D} + \eta}\xi_t$$

$$\text{FOC-}(b_t) : \lambda_t = \frac{W_b}{\chi}\big(\xi_t - E_t\xi_{t+1}\big)$$

$$\text{FOC-}(w_t) : \frac{a(1-a)\phi}{D}w_t - \frac{D - 2a_H + 1}{2D}\delta\gamma_t^R + \big(\lambda_t - \beta^{-1}\lambda_{t-1}\big) + \xi_t = 0$$

$$(\text{M-UIP}) : b_t = \frac{1}{\chi}\big(w_t - E_t w_{t+1} + f_t\big)$$

$$(\text{BC}) : w_t - 2W_y\widetilde{y}_t^R + W_b\big(b_t - \beta^{-1}b_{t-1} - nx_t^{fb}\big) = 0$$

$$\text{FOC-}(\pi_t^R) : \pi_t^R = \frac{\delta}{\epsilon}\big(\gamma_{t-1}^R - \gamma_t^R\big)$$

$$(\text{PC}) : \pi_t^R - \beta\pi_{t+1}^R - \delta\big((\frac{\sigma}{D} + \eta)\widetilde{y}_t^R + \frac{D - 2a_H + 1}{2D}w_t + \mu_t^R\big) = 0$$

For convenience, define $F(x_t) \equiv x_{t+1} - \big(1 + \beta^{-1} + \frac{\chi}{W_b}\big)x_t + \beta^{-1}x_{t-1}$. Combining the first three equations,

$$F\big(\widetilde{y}_{t+1}^R - \delta\gamma_{t+1}^R\big) = \frac{\chi}{W_b}\big(\frac{a_H(1-a_H)\phi}{D}w_{t+1} - \frac{D - 2a_H + 1}{2D}\delta\gamma_{t+1}^R\big)$$

Combining the next two equations,

$$F(w_{t+1}) = -\frac{2W_y\chi}{W_b}\widetilde{y}^R_{t+1} - \chi n x^{fb}_{t+1} + f_{t+1} - \beta^{-1}f_t$$

Combining the last two equations, and subtracting $\frac{\chi}{W_b}\gamma^R_{t+1}$ from both sides,

$$F(\gamma^R_{t+1}) = \beta^{-1}\epsilon\Big((\frac{\sigma}{D}+\eta)\widetilde{y}^R_{t+1} + \frac{D-2a_H+1}{2D}w_{t+1} + \mu^R_{t+1}\Big) - \frac{\chi}{W_b}\gamma^R_{t+1}$$

Define $x_t \equiv (\widetilde{y}^R_t, w_t, \gamma^R_t)'$. By appropriately defining the matrices A and B, the above three equations can be summarized as

$$F(Bx_{t+1}) = Ax_{t+1} \quad \Leftrightarrow \quad F(x_{t+1}) = B^{-1}Ax_{t+1}$$

Using the eigenvalue decomposition of $B^{-1}A = Q\Lambda Q^{-1}$, it can be expressed as

$$Q^{-1}F(x_{t+1}) = F(Q^{-1}x_{t+1}) = \Lambda Q^{-1}x_{t+1}$$

Defining $z_t \equiv Q^{-1}x_t$,

$$F(z_{t+1}) = \Lambda z_{t+1}$$

$$F(z_{i,t+1}) = \lambda_i z_{i,t+1},\ \forall i \in \{1,2,3\}$$

Then each $z_{i,t}$ can be solved using

$$F(z_{i,t+1}) - \lambda_i z_{i,t+1} = z_{t+2} - \big(1 + \beta^{-1} + \frac{\chi}{W_b} - \lambda_i\big)z_{i,t+1} + \beta^{-1}z_t = 0$$

Note that each eigenvalue λ_i of $B^{-1}A$ can either be real or complex. Since $B^{-1}A$ is a real matrix, complex eigenvalues must appear in pairs of conjugates, if any. In principle, there can be either two or no complex roots, but considering the case with $\alpha = 1$, it seems more likely that there will be one pair of complex conjugate roots and one real root. Then each quadratic characteristic equations would yield two complex roots (ν_{i1}, ν_{i2}).

Although it is not proved explicitly, it is probably the case that $|\nu_{i1}| < 1 < \beta^{-1} < |\nu_{i2}|$, implying $z_{i,t+k} = \nu_{i1}^{k-1} z_{i,t+1}$. Using the equations at t, and combining with the state variables at $t-1$ and shocks at t, $z_{i,t}$ will be found, and $x_{i,t}$ can be backed out using $x_t = Qz_t$.

Suppose there is a pair of complex conjugate eigenvalues, say (λ_1, λ_2). Then clearly ν_{11} and ν_{21} are complex conjugates as well because $(1 + \beta^{-1} + \frac{\chi}{W_b})$ is real-valued. This means that $\{z_{1,t}\}$ and $\{z_{2,t}\}$ will decay at the same rate with an exactly opposite angle. Any linear combination of these two series would then decay at the common rate $|\nu_{11}| = |\nu_{21}| < 1$ but have an arbitrary (constant) change in angle between each period based on the relative weights on $z_{1,t}$ and $z_{2,t}$. On the other hand, λ_3 is a real root, and it can result in either two real roots or two complex conjugate roots (ν_{31}, ν_{32}). But if these roots are complex, the system would be unstable because $|\nu_{31}| \cdot |\nu_{32}| \geq \nu_{31}\nu_{32} = \beta^{-1} > 1$ results in $|\nu_{31}| = |\nu_{32}| > 1$. Therefore it can be inferred that ν_{31} is real and is smaller than 1 assuming that there is a stable solution.

Overall, each of the variables $x_{i,t} \in \{\widetilde{y}_t^R, w_t, \gamma_t^R\}$ would be a linear combination of two exponential decaying series, where one at rate ν_{31} is without oscillatory component and the other at rate $|\nu_{11}| = |\nu_{21}|$ can have an oscillatory component.

III. Optimal Capital Control (Independent of χ)

Compared to the optimal monetary policy case, there is one less variable (λ_t), and one less equation (modified UIP). But otherwise the solution can be found in a similar way. There are now six equations that pin down six variables. The first equation, FOC-(b_t), sets permanent expected levels for $\xi_t = E_t\xi_{t+1} = \overline{\xi}_t$, which would later be determined endogenously.

The next two equations allows $\widetilde{y}_t^R$ and $\widetilde{c}_t^R$ to be expressed in terms of γ_t^R and $\overline{\xi}_t$:

$$\text{FOC-}(\widetilde{y}_t^R) : \widetilde{y}_t^R = \delta\gamma_t^R + 2\sigma(\phi - 1)\overline{\xi}_t$$

$$\text{FOC-}(w_t) : \frac{1}{2}\widetilde{c}_t^R + \overline{\xi}_t = \frac{1}{2}\delta\gamma_t^R$$

As before, the budget constraint can be combined into a lifetime version to cancel out b_t, which would eventually pin down the value of $\overline{\xi}_t$:

$$(\text{BC}) : \widetilde{c}_t^R - (1 - \frac{1}{\phi})\widetilde{y}_t^R + (b_t - \beta^{-1}b_{t-1}) = (\phi - 1)a_t^R$$

Lastly, the two remaining equations related to inflation can be used to solve for γ_t^R:

$$\text{FOC-}(\pi_t^R) : \pi_t^R = -\frac{\delta}{\epsilon}\Delta\gamma_t^R$$

$$(\text{PC}) : \pi_t^R - \beta\pi_{t+1}^R = \delta\big(\frac{1}{\phi}\widetilde{y}_t^R + \sigma\widetilde{c}_t^R\big)$$

Combining the two equations,

$$\beta\gamma_{t+1}^R - (1+\beta)\gamma_t^R + \gamma_{t-1}^R = \epsilon\big[\frac{1}{\phi}\big(\delta\gamma_t^R + 2\sigma(\phi-1)\overline{\xi}_t\big) + \sigma\big(\delta\gamma_t^R - 2\overline{\xi}_t\big)\big] = \epsilon\big[(\frac{1}{\phi} + \sigma)\delta\gamma_t^R - \frac{2\sigma}{\phi}\overline{\xi}_t\big]$$

$$\gamma_{t+1}^R - \big[1 + \beta^{-1} + \beta^{-1}\epsilon\delta(\frac{1}{\phi} + \sigma)\big]\gamma_t^R + \beta^{-1}\gamma_{t-1}^R = -\frac{\epsilon\sigma}{\beta\phi}2\overline{\xi}_t$$

Define the characteristic equation

$$\nu^2 - \big[1 + \beta^{-1} + \frac{\epsilon\delta}{\beta}(\frac{1}{\phi} + \sigma)\big]\nu + \beta^{-1} = 0$$

and its solutions $0 < \nu_{k1} < 1 < \beta^{-1} < \nu_{k2}$. Then,

$$(\gamma_{t+1}^R - \nu_{k1}\gamma_t^R) - \nu_{k2}(\gamma_t^R - \nu_{k1}\gamma_{t-1}^R) = -\frac{\epsilon\sigma}{\beta\phi}2\overline{\xi}_t$$

$$\nu_{k2}^{-1}(\gamma_{t+2}^R - \nu_{k1}\gamma_{t+1}^R) - (\gamma_{t+1}^R - \nu_{k1}\gamma_t^R) = -\nu_{2k}^{-1}\frac{\epsilon\sigma}{\beta\phi}2\overline{\xi}_t$$

$$\vdots$$

$$\therefore (\gamma_t^R - \nu_{k1}\gamma_{t-1}^R) = (\gamma_{t+1}^R - \nu_{k1}\gamma_t^R) = \cdots = -\frac{\nu_{k2}^{-1}}{1-\nu_{k2}^{-1}}\frac{\epsilon\sigma}{\beta\phi}2\bar{\xi}_t \equiv \bar{z}_{kt}$$

From $\gamma_{t+1}^R - \nu_{k1}\gamma_t^R = \bar{z}_{kt} = \dfrac{1}{1-\nu_{k1}}\bar{z}_{kt} - \dfrac{\nu_{k1}}{1-\nu_{k1}}\bar{z}_{kt}$, and assuming $\gamma_{t-1}^R = 0$ without loss of generality,

$$\gamma_{t+1}^R - \frac{1}{1-\nu_{k1}}\bar{z}_{kt} = \nu_{k1}\big(\gamma_t^R - \frac{1}{1-\nu_{k1}}\bar{z}_{kt}\big) = -\frac{\nu_{k1}^2}{1-\nu_{k1}}\bar{z}_{kt}$$

Iterating forward,

$$E_t\gamma_{t+k}^R - \frac{1}{1-\nu_{k1}}\bar{z}_{kt} = \nu_{k1}^k\big(\gamma_t^R - \frac{1}{1-\nu_{k1}}\bar{z}_{kt}\big) = -\frac{\nu_{k1}^{k+1}}{1-\nu_{k1}}\bar{z}_{kt},\ \forall k \geq 0$$

$$\therefore E_t\gamma_{t+k}^R = -\frac{1-\nu_{k1}^{k+1}}{1-\nu_{k1}}\bar{z}_{kt} = \frac{1-\nu_{k1}^{k+1}}{(1-\nu_{k1})(\nu_{k2}-1)}\frac{\epsilon\sigma}{\beta\phi}2\bar{\xi}_t,\ \forall k \geq 0$$

Using $\nu_{k1}+\nu_{k2} = 1+\beta^{-1}+\dfrac{\epsilon\delta}{\beta}(\dfrac{1}{\phi}+\sigma)$ and $\nu_{k1}\nu_{k2} = \beta^{-1}$, $(1-\nu_{k1})(\nu_{k2}-1) = \dfrac{\epsilon\delta}{\beta}(\dfrac{1}{\phi}+\sigma)$. Therefore,

$$E_t\gamma_{t+k}^R = (1-\nu_{k1}^{k+1})\frac{2\sigma}{\delta(1+\sigma\phi)}\bar{\xi}_t$$

As before,

$$\sum_{k=0}^{\infty}\beta^k(1-\nu_{k1}^{k+1}) = \frac{1-\nu_{k1}}{(1-\beta)(1-\beta\nu_{k1})}$$

$$\therefore E_t\sum_{k=0}^{\infty}\beta^k\gamma_{t+k}^R = \frac{1-\nu_{k1}}{(1-\beta)(1-\beta\nu_{k1})}\frac{2\sigma}{\delta(1+\sigma\phi)}\bar{\xi}_t$$

Now back to (BC),

$$\begin{aligned}
\widetilde{nx}_t &= (1-\frac{1}{\phi})\tilde{y}_t^R - \tilde{c}_t^R \\
&= (1-\frac{1}{\phi})\big(\delta\gamma_t^R + 2\sigma(\phi-1)\bar{\xi}_t\big) - \delta\gamma_t^R + 2\bar{\xi}_t \\
&= -\frac{\delta}{\phi}\gamma_t^R + \big(\frac{\sigma}{\phi}(\phi-1)^2+1\big)2\bar{\xi}_t
\end{aligned}$$

$$\sum_{k=0}^{\infty}\beta^k\widetilde{nx}_t^R = -\frac{\delta}{\phi}\sum_{k=0}^{\infty}\beta^k\gamma_{t+k}^R + \frac{1}{1-\beta}\big(1+\frac{\sigma}{\phi}(\phi-1)^2\big)2\bar{\xi}_t = -(\phi-1)a_t^R$$

Combining with the equation derived previously,

$$\Big[-\frac{\delta}{\phi}\frac{\sigma}{\delta(1+\sigma\phi)}\frac{1-\nu_{k1}}{1-\beta\nu_{k1}}+1+\frac{\sigma}{\phi}(\phi-1)^2\Big]2\bar{\xi}_t=-(\phi-1)a_t^R$$

$$\therefore 2\bar{\xi}_t=-\frac{(1-\beta)\phi(\phi-1)}{1+\sigma(\phi-1)h-\frac{1-\nu_{k1}}{1-\beta\nu_{k1}}\frac{\sigma}{1+\sigma\phi}}a_t^R\equiv-\frac{(1-\beta)\phi(\phi-1)}{D^{kp}}a_t^R$$

where $\dfrac{\phi}{\sigma}+(\phi-1)^2=(\phi-1)h+\dfrac{1}{\sigma}$ has been used. Given this solution for $\bar{\xi}_t^R$, all other variables can also be solved analytically.

Appendix B

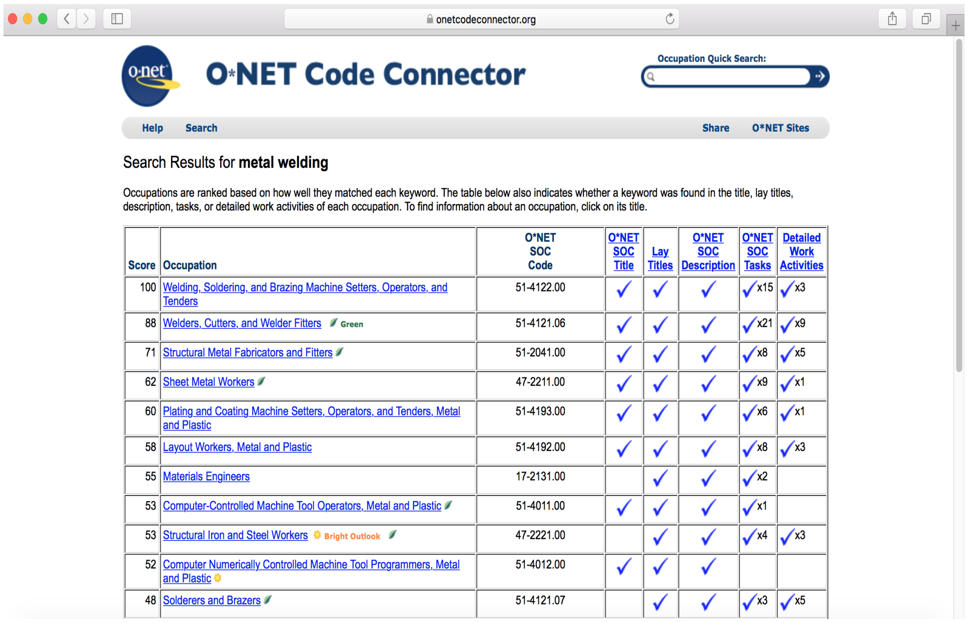
Appendix to Chapter 2

B.1 Data

B.1.1 Skill Construction

We scrape education and vocational training records from variables *hfaudd* and *erhaudd* respectively where we obtain 2449 different types of records, and clean the textual information: make all words to lower-case, remove unnecessary abbreviations, replace punctuation characters with blank spaces, etc. Then, we translate the Danish words into english and manually examine words that are not directly translatable. We end up with 523 corresponding O*NET-SOC occupations that correspond to the educational information provided in the two variables. As the goal lies in identifying textual information in each education record that can be useful in relating the skill sets of an individual worker, we further use the education guide provided by the Ministry of Education in Denmark (`https://www.ug.dk`) to capture key words that characterize the task/skill content of academic education as well as vocational training records.

Next, we feed in the cleaned textual information of each education entry to the O*NET code connector (`https://www.onetcodeconnector.org`, which provides a list



onetcodeconnector.org

O*NET Code Connector

Occupation Quick Search:

Help Search Share O*NET Sites

Search Results for **metal welding**

Occupations are ranked based on how well they matched each keyword. The table below also indicates whether a keyword was found in the title, lay titles, description, tasks, or detailed work activities of each occupation. To find information about an occupation, click on its title.

Score	Occupation	O*NET SOC Code	O*NET SOC Title	Lay Titles	O*NET SOC Description	O*NET SOC Tasks	Detailed Work Activities
100	Welding, Soldering, and Brazing Machine Setters, Operators, and Tenders	51-4122.00	✓	✓	✓	✓x15	✓x3
88	Welders, Cutters, and Welder Fitters Green	51-4121.06	✓	✓	✓	✓x21	✓x9
71	Structural Metal Fabricators and Fitters	51-2041.00	✓	✓	✓	✓x8	✓x5
62	Sheet Metal Workers	47-2211.00	✓	✓	✓	✓x9	✓x1
60	Plating and Coating Machine Setters, Operators, and Tenders, Metal and Plastic	51-4193.00	✓	✓	✓	✓x6	✓x1
58	Layout Workers, Metal and Plastic	51-4192.00	✓	✓	✓	✓x8	✓x3
55	Materials Engineers	17-2131.00		✓	✓	✓x2	
53	Computer-Controlled Machine Tool Operators, Metal and Plastic	51-4011.00	✓	✓	✓	✓x1	
53	Structural Iron and Steel Workers Bright Outlook	47-2221.00		✓	✓	✓x4	✓x3
52	Computer Numerically Controlled Machine Tool Programmers, Metal and Plastic	51-4012.00	✓	✓	✓		
48	Solderers and Brazers	51-4121.07		✓	✓	✓x3	✓x5

of relevant corresponding occupations. The criteria we use for finding a match is that, (i) relevance scores are higher than 90; (ii) the education key words checks off with the occupation title, the lay title, the job description, the task content and work activities. For education entries that fail these criteria go through a second set of algorithm, which requires (i) relevance scores are higher than 90; (ii) the education key words checks off with the occupation title or the task content or work activities. Python codes and files are available upon request. For educational records that are too general and abstract are not included in the algorithm.

B.1.2 Construction of Other Measures

Value Added We exactly follow Bagger et al. (2014) to construct the value added Y:

1. 1995-1998

$$Y = (\text{OMS}+\text{AUER}+\text{ADR}+\text{DLG})$$
$$-(\text{KRH}+\text{KENE}+\text{KLEO}+\text{UDHL}+\text{UASI}+\text{OEEU}+\text{SEUD})$$

2. 1999-2001

$$Y = (\text{OMS}+\text{AUER}+\text{ADR}+\text{DLG}+\text{TGT}\times 0.0079)$$
$$-(\text{KRH}+\text{KENE}+\text{KLEO}+\text{UDHL}+\text{UASI}+\text{UDVB}+\text{ULOL}+\text{ANEU}+\text{SEUD})$$

3. 2002-2003

$$Y = (\text{OMS}+\text{AUER}+\text{ADR}+\text{DLG})$$
$$-(\text{KRH}+\text{KENE}+\text{KLEO}+\text{UDHL}+\text{UASI}+\text{UDVB}+\text{ULOL}+\text{ANEU}+\text{SEUD})$$

4. 2004-2013

$$Y = (\text{OMS}+\text{AUER}+\text{ADR}+\text{DLG})$$
$$-(\text{KVV}+\text{KRHE}+\text{KENE}+\text{KLEO}+\text{UDHL}+\text{UASI}+\text{UDVB}+\text{ULOL}+\text{ANEU}+\text{SEUD})$$

OMS is revenue, AUER is work conducted at own expense, ADR is other operating revenue, DLG is ultimo inventory minus primo inventory; KRH is cost of intermediates, KENE is cost of energy, KLEO is costs of subcontractors, UDHL is housing rents, UASI is purchases of minor equipment, OEEU is other external costs, SEUD is secondary costs, TGT is total credits, UDVB is purchases of temporary employment agency, ULOL is costs of long-term leasing, ANEU is other external costs, KVV is purchases of goods for resale, and KRHE is costs of intermediates.

Occupational Offshorability There are several different methods the literature has established ways of capturing the degrees of offshorability at the occupation-level. Autor et al. (2003) and Acemoglu and Autor (2011) rely on occupation characteristics provided

	Blinder and Krueger (2013)	Autor, Levy, Murnane (2003)
Offshorable	Machine operators and assemblers	Office clerks
	Precision, handicraft, craft printing and related trade workers	Precision, handicraft, craft printing and related trade workers
	Stationary plant and related operators	Customer service clerks
	Other craft and related trade workers	Other craft and related trade workers
	Physical, mathematical and engineering professionals	Machine operators and assemblers
Non-Offshorable	Drivers and mobile plant operators	Managers of small enterprises
	Personal and protective service workers	Drivers and mobile plant operators
	Extraction and building trades workers	Life science and health professionals
	Models, salespersons and demonstrators	Physical, mathematical and engineering professionals
	Sales and service elementary occupations	Corporate managers
	Blinder (2009)	**Goos, Manning, Salomons (2014)**
Offshorable	Physical, mathematical and engineering professionals	Machine operators and assemblers
	Precision, handicraft, craft printing and related trade workers	Stationary plant and related operators
	Machine operators and assemblers	Office clerks
	Other craft and related trade workers	Laborers in mining, construction, manufacturing and transport
	Stationary plant and related operators	Metal, machinery and related trade work
Non-Offshorable	Models, salespersons and demonstrators	Life science and health associate professionals
	Teaching associate professionals	Models, salespersons and demonstrators
	Teaching professionals	Life science and health professionals
	Drivers and mobile plant operators	Personal and protective service workers
	Personal and protective service workers	Drivers and mobile plant operators

Table 9: Offshorability of occupations as constructed in the previous literature

in O*NET to capture occupational offshorability while Blinder (2009) adds his subjective judgement to further categorize offshorable occupations. Blinder and Krueger (2013) utilizes household survey measurements of the "offshorability" of jobs while Goos et al. (2014) uses all of the pre-existing measures of offshorability listed above and also compare their own construction, which is based on the European Restructuring Monitor (ERM) contains summaries of news reports about cases of offshoring by companies located in Europe.

B.1.3 Industry and Occupation Classifications

Industry Classification We employ variables *gf_branche_93*, *gf_branche_03*, and *gf_branche_07* in MEE, which provide the Danish Industrial Classification (Dansk Branchekode; abbreviated DB) at the six-digit level to identify firms' industry categories.

This classification follows the NACE system where DB93, DB03, DB07 demonstrate correspondence with the NACE Rev. 1, the NACE Rev. 1.1, and the NACE Rev. 2, respectively.

Occupation Classification We use variables *discoalle_indk* (1991-2009) and *disco08alle_indk* (2010-) in MEE to obtain information regarding individuals' occupations. Denmark assigns the Danish version of ISCO (International Standard Classification of Occupations), DISCO, to these variables, which at the first two-digit level demonstrates high correspondence with the ISCO.[1] While *discoalle_indk* provides six-digit level occupational categories, we use this variable at the two-digit level, as there have changes over time in the way the codes are formulated in the higher digits.[2]

B.2 Counterfactuals (Additional Results)

Between and Within Occupation Channel What happens if we consider average wages combined across offshorable and non-offshorable jobs? As discussed earlier, the overall between-firm wage inequality combining across occupations depends on the relative magnitudes of the within-occupation versus between-occupation channels. As high-productivity firms trim down their offshorable occupations in-house, the average wage for their domestic workers largely accounts for those in non-offshorable jobs. Low-productivity firms, on the other hand, keep both the non-offshorable jobs and offshorable jobs in-house so that the average wage they pay reflects the overall wage loss workers

[1] http://www.ilo.org/public/english/bureau/stat/isco/isco88/publ3.htm
https://www.dst.dk/en/Statistik/dokumentation/nomenklaturer/disco-88

[2] https://www.dst.dk/da/Statistik/dokumentation/Times/personindkomst/discoalle-indk

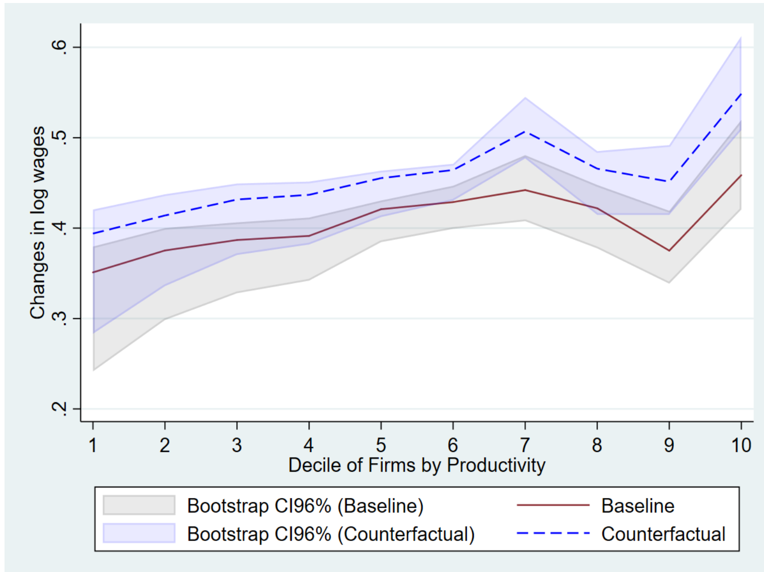


Figure 22: Changes in log(wage) by decile of firms (productivity)

with offshorable occupations have undergone due to globalization taking place. Thus, in the following we show changes in average wages weighted by occupation shares of offshorable and non-offshorable occupations by decile of firms' productivity, which we compare with the counterfactual economy where changes in occupations shares are not reflected.

B.3 Theories and Proofs

B.3.1 Unobserved Preferences

We closely follow Dupuy and Galichon (2014) in the assumptions for unobserved heterogeneities. Each worker with observable skill x has a set composed of random realization of "acquaintances," which follows a Poisson point process on $Y \times R$ of intensity

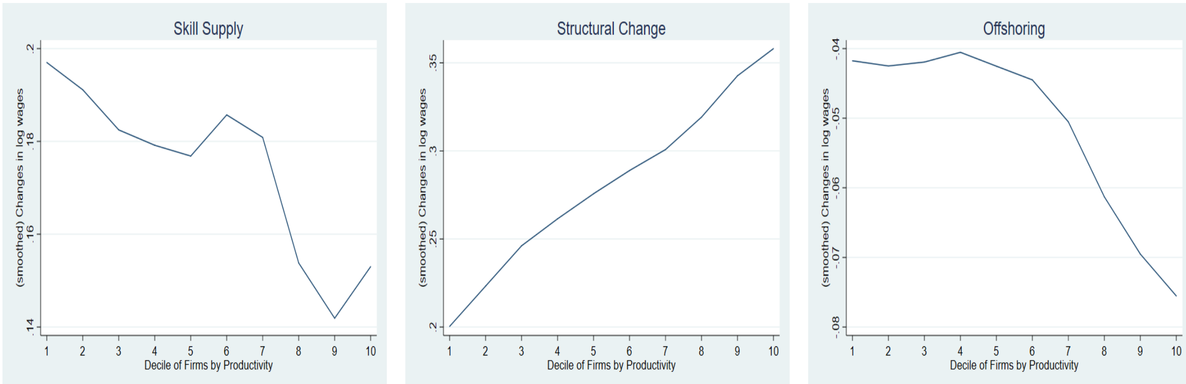


Figure 23: Changes in log(wage) by decile of firms (productivity) in each channel

$\exp(-e)dedy$. If a point (y, e) is in the acquaintance set, this implies that the individual's unobserved preference for firm with productivity y is equal to e.[3] As a consequence of the Poisson point process assumption, each individual has infinite but countable number of acquaintances. For all values of y that are not in the acquaintance set, negative infinity is assigned to the preference shock, which is a natural assumption in the current context, implying that they can never be optimally chosen. In sum,

$$e_x(y) = \begin{cases} \max_k e_k & \text{if } y_k \in \text{acquaintance set} \\ -\infty & \text{otherwise} \end{cases}$$

In addition, each individual has a random preference on the choice of outside option of being unmatched, also given by an analogous Poisson point process $\exp(-e)de$. Taking maximum of the given points, this assumption is exactly the same as assuming a Gumbel distribution for this preference shock.[4] Note that in order to maintain the analytical tractability of standard logit problem with discrete choice, the same scale parameter λ_x is used for both outside option and each of the matching option y. All of the above

[3] When there are multiple e's that coincide with the same y, the maximum of these e's are taken. However, since such events occur with probability zero, this consideration is in fact immaterial.

[4] Dupuy and Galichon (2014) also consider this case with outside option in their appendix D, and show that there is no meaningful changes in their results.

description of preference shocks apply equally to both Home and Foreign workers.

Firms also face exactly the same preference shocks, defined by the same Poisson point process if matched and a Gumbel distribution if unmatched, as described above. The only potential difference is that the degree of unobserved heterogeneity may be different: $\lambda_x \neq \lambda_y$ in general. It turns out that only $\lambda = \lambda_x + \lambda_y$ matters for the matching function $\pi(x, y)$, but each λ_x, λ_y does matter for determining how the output is shared between workers and firms.[5]

B.3.2 Derivation of Closed Economy Equilibrium

Each worker makes an idiosyncratic decision, but because of the convenient assumptions of Poisson point process (which leads to an continuous analog of the standard logit models), the distributions of these choices conditional on observable type x can be characterized as follows. Recall that the skill distribution of workers that choose to be unmatched is $f_0(x)$, and the distribution of matched workers is $f(x)$. Then

$$\frac{f_0(x)}{\overline{f}(x)} = \frac{\exp(\frac{w}{\lambda_x})}{\exp(\frac{w}{\lambda_x}) + \int \exp(\frac{w(x,y)}{\lambda_x})dy}, \quad \frac{f(x)}{\overline{f}(x)} = \frac{\int \exp(\frac{w(x,y)}{\lambda_x})dy}{\exp(\frac{w}{\lambda_x}) + \int \exp(\frac{w(x,y)}{\lambda_x})dy}$$

holds in equilibrium, which is directly comparable to the conditional choice probabilities in standard logit case. Similarly, the distribution of unmatched and matched firms $g_0(y)$ and $g(y)$ satisfy

$$\frac{g_0(y)}{\overline{g}(y)} = \frac{1}{1 + \int \exp(\frac{r(x,y)}{\lambda_y})dx}, \quad \frac{g(y)}{\overline{g}(y)} = \frac{\int \exp(\frac{r(x,y)}{\lambda_y})dx}{1 + \int \exp(\frac{r(x,y)}{\lambda_y})dx}$$

[5]When $\lambda \to 0$, the equilibrium converges to a perfect sorting or pure matching. When $\lambda \to \infty$, the equilibrium converges to a purely random matching, i.e., $\pi(x, y) = f(x)g(y)$. Any finite, positive value of λ describes an environment in between these extreme cases: a smaller value of λ indicates more sorted economy, and vice versa.

where 1's correspond to the firm's outside option of zero. The conditional probability of a worker with observable type x choosing firm y satisfies:

$$\pi(y|x) = \frac{\pi(x,y)}{f(x)} = \frac{\exp(\frac{w(x,y)}{\lambda_x})}{\int \exp(\frac{w(x,y)}{\lambda_x})dy}$$

Taking logs on both sides and rearranging, collecting terms independent of y,

$$\log \pi(x,y) = \frac{w(x,y)}{\lambda_x} - \log \frac{\int \exp(\frac{w(x,y)}{\lambda_x})dy}{f(x)} = \frac{w(x,y) - a(x)}{\lambda_x}$$

where $a(x) \equiv \lambda_x \log \frac{\int \exp(\frac{w(x,y)}{\lambda_x})dy}{f(x)}$. Similarly on the firm's side,

$$\pi(x|y) = \frac{\pi(x,y)}{g(y)} = \frac{\exp(\frac{r(x,y)}{\lambda_y})}{\int \exp(\frac{r(x,y)}{\lambda_y})dx}$$

$$\log \pi(x,y) = \frac{r(x,y)}{\lambda_y} - \log \frac{\int \exp(\frac{r(x,y)}{\lambda_y})dx}{g(y)} = \frac{r(x,y) - b(y)}{\lambda_y}$$

and $b(y) \equiv \lambda_y \log \frac{\int \exp(\frac{r(x,y)}{\lambda_y})dx}{g(y)}$.

Combining the above two equation and using $q(x,y) = w(x,y) + r(x,y)$, the endogenous objects $w(x,y)$ and $r(x,y)$ can be cancelled out to yield $(\lambda_x + \lambda_y)\log \pi(x,y) = q(x,y) - a(x) - b(y)$. Denoting $\lambda \equiv \lambda_x + \lambda_y$,

$$\pi(x,y) = \exp(-\frac{a(x)}{\lambda})\exp(-\frac{b(y)}{\lambda})\exp(\frac{q(x,y)}{\lambda}) = \hat{a}(x)\hat{b}(y)\exp(\frac{q(x,y)}{\lambda})$$

where $\hat{a}(x) \equiv \exp(-\frac{a(x)}{\lambda})$ and $\hat{b}(y) \equiv \exp(-\frac{b(y)}{\lambda})$ are defined for notational convenience. In addition,

$$\exp(-\frac{a(x)}{\lambda_x}) = \frac{f(x)}{\int \exp(\frac{w(x,y)}{\lambda_x})dy} = \frac{\overline{f}(x)}{\exp(\frac{w}{\lambda_x}) + \int \exp(\frac{w(x,y)}{\lambda_x})dy} = \hat{a}(x)^{\frac{\lambda}{\lambda_x}}$$

$$\exp(-\frac{b(y)}{\lambda_y}) = \frac{g(y)}{\int \exp(\frac{r(x,y)}{\lambda_y})dx} = \frac{\overline{g}(y)}{1 + \int \exp(\frac{r(x,y)}{\lambda_y})dx} = \hat{b}(y)^{\frac{\lambda}{\lambda_y}}$$

From these equations, it immediately follows that

$$f_0(x) = \frac{\exp(\frac{w}{\lambda_x})\overline{f}(x)}{\exp(\frac{w}{\lambda_x}) + \int \exp(\frac{w(x,y)}{\lambda_x})dy} = \hat{a}(x)^{\frac{\lambda}{\lambda_x}} \exp(\frac{w}{\lambda_x})$$

$$g_0(y) = \frac{\overline{g}(y)}{1 + \int \exp(\frac{r(x,y)}{\lambda_y})dx} = \hat{b}(y)^{\frac{\lambda}{\lambda_y}}$$

Now the exogenously given marginal distributions can be expressed as

$$\overline{f}(x) = \hat{a}(x)^{\frac{\lambda}{\lambda_x}} \exp(\frac{w}{\lambda_x}) + \hat{a}(x) \int \hat{b}(y) \exp(\frac{q(x,y)}{\lambda})dy$$

$$\overline{g}(y) = \hat{b}(y)^{\frac{\lambda}{\lambda_y}} + \hat{b}(y) \int \hat{a}(x) \exp(\frac{q(x,y)}{\lambda})dx$$

where $f(x) = \int \pi(x,y)dy$ and $g(y) = \int \pi(x,y)dx$ have been used.

This last set of equations is crucial for solving the equilibrium. The endogenous objects $\hat{a}(x), \hat{b}(y)$ need to be solved, taking as given: marginal distributions $\overline{f}(x), \overline{g}(y)$, production function $q(x,y)$, and the degrees of unobserved heterogeneity λ_x, λ_y. There is a straightforward iterative algorithm to solve for $\hat{a}(x), \hat{b}(y)$: start with an initial $\hat{b}(y)$, plug in to (1) to obtain $\hat{a}(x)$, then plug in to (2) to obtain $\hat{b}(y)$, and repeat until convergence.[6] Once $\hat{a}(x), \hat{b}(y)$ are found, it is straightforward to recover all other endogenous objects, including the matching function $\pi(x,y)$ as well as wages and profits $w(x,y), r(x,y)$.

[6] This algorithm is proposed in Bojilov and Galichon (2016), which they refer to as "Iterated Proportional Fitting Procedure (IPFP)" or "Sinkhorn's algorithm."

B.3.3 Derivation of Global Economy Equilibrium

Home worker with skill x maximizes his utility:

$$\max(w + \lambda_x e, \max_y\{w(x,y) + \lambda_x e(y)\})$$

Likewise, Foreign worker with skill z maximizes his utility:

$$\max(w_F + \lambda_{x_F} e, \max_y\{w_F(x_F,y) + \lambda_{x_F} e(y)\})$$

The utility shocks are defined in the same way as before. Foreign workers may have different outside wage option of w_F, different wage $w_F(x_F, y)$, and different degree of heterogeneity λ_{x_F}.

A (Home) firm with productivity y maximizes its utility as before, but now it has a third option: to offshore by matching with a Foreign worker. The problem thus becomes: $max(\lambda_y e, max_x\{r(x,y) + \lambda_y e(x)\}, max_z\{V_F(x_F,y) + \lambda_y e(x_F)\})$. As before, competitive equilibrium implies $w_F(x_F,y) + V_F(x_F,y) = q_F(x_F,y)$.

The conditional distributions of choices for Home and Foreign workers are the same as before:

$$\frac{f_0(x)}{\overline{f}(x)} = \frac{\exp(\frac{w}{\lambda_x})}{\exp(\frac{w}{\lambda_x}) + \int \exp(\frac{w(x,y)}{\lambda_x})dy}, \quad \frac{f(x)}{\overline{f}(x)} = \frac{\int \exp(\frac{w(x,y)}{\lambda_x})dy}{\exp(\frac{w}{\lambda_x}) + \int \exp(\frac{w(x,y)}{\lambda_x})dy}$$

$$\frac{h_0(x_F)}{\overline{h}(x_F)} = \frac{\exp(\frac{w_F}{\lambda_{x_F}})}{\exp(\frac{w_F}{\lambda_{x_F}}) + \int \exp(\frac{w_F(x_F,y)}{\lambda_{x_F}})dy}, \quad \frac{h(x_F)}{\overline{h}(x_F)} = \frac{\int \exp(\frac{w_F(x_F,y)}{\lambda_{x_F}})dy}{\exp(\frac{w_F}{\lambda_{x_F}}) + \int \exp(\frac{w_F(x_F,y)}{\lambda_{x_F}})dy}$$

Distributions for firms need to be modified to incorporate the newly available choice of offshoring:

$$\frac{g_0(y)}{\overline{g}(y)} = \frac{1}{D}\ , \quad \frac{g(y)}{\overline{g}(y)} = \frac{\int \exp(\frac{r(x,y)}{\lambda_y})dx}{D}\ , \quad \frac{g_F(y)}{\overline{g}(y)} = \frac{\int \exp(\frac{V_F(x_F,y)}{\lambda_y})dz}{D}$$

where $D \equiv 1 + \int \exp(\frac{r(x,y)}{\lambda_y})dx + \int \exp(\frac{V_F(x_F,y)}{\lambda_y})dz$.

The conditional probabilities for those who form matches are:

$$\pi(y|x) = \frac{\pi(x,y)}{f(x)} = \frac{\exp(\frac{w(x,y)}{\lambda_x})}{\int \exp(\frac{w(x,y)}{\lambda_x})dy}, \quad \pi(x|y) = \frac{\pi(x,y)}{g(y)} = \frac{\exp(\frac{r(x,y)}{\lambda_y})}{\int \exp(\frac{r(x,y)}{\lambda_y})dx}$$

$$\pi_F(y|z) = \frac{\pi_F(x_F,y)}{h(x_F)} = \frac{\exp(\frac{w_F(x_F,y)}{\lambda_{x_F}})}{\int \exp(\frac{w_F(x_F,y)}{\lambda_{x_F}})dy}, \quad \pi_F(z|y) = \frac{\pi_F(x_F,y)}{g(y)} = \frac{\exp(\frac{V_F(x_F,y)}{\lambda_y})}{\int \exp(\frac{V_F(x_F,y)}{\lambda_y})dz}$$

Taking logs on both sides and rearranging,

$$\log \pi(x,y) = \frac{w(x,y)}{\lambda_x} - \log \frac{\int \exp(\frac{w(x,y)}{\lambda_x})dy}{f(x)} = \frac{w(x,y) - a(x)}{\lambda_x}$$

$$\log \pi(x,y) = \frac{r(x,y)}{\lambda_y} - \log \frac{\int \exp(\frac{r(x,y)}{\lambda_y})dx}{g(y)} = \frac{r(x,y) - b(y)}{\lambda_y}$$

$$\log \pi_F(x_F,y) = \frac{w_F(x_F,y)}{\lambda_{x_F}} - \log \frac{\int \exp(\frac{w_F(x_F,y)}{\lambda_{x_F}})dy}{h(x_F)} = \frac{w_F(x_F,y) - c(x_F)}{\lambda_{x_F}}$$

$$\log \pi_F(x_F,y) = \frac{V_F(x_F,y)}{\lambda_y} - \log \frac{\int \exp(\frac{V_F(x_F,y)}{\lambda_y})dz}{g(y)} = \frac{V_F(x_F,y) - b_F(y)}{\lambda_y}$$

where $a(x) \equiv \lambda_x \log \frac{\int \exp(\frac{w(x,y)}{\lambda_x})dy}{f(x)}$, $b(y) \equiv \lambda_y \log \frac{\int \exp(\frac{r(x,y)}{\lambda_y})dx}{g(y)}$,
$c(x_F) \equiv \lambda_{x_F} \log \frac{\int \exp(\frac{w_F(x_F,y)}{\lambda_{x_F}})dy}{h(x_F)}$, and $b_F(y) \equiv \lambda_y \log \frac{\int \exp(\frac{V_F(x,y)}{\lambda_y})dz}{g(y)}$.

Again, it is straightforward to obtain $\lambda \log \pi(x,y) = q(x,y) - a(x) - b(y)$ and $\lambda_F \log \pi_F(x_F,y) = q_F(x_F,y) - c(x_F) - b_F(y)$, and thus

$$\pi(x,y) = \exp(-\frac{a(x)}{\lambda}) \exp(-\frac{b(y)}{\lambda}) \exp(\frac{q(x,y)}{\lambda}) = \hat{a}(x)\hat{b}(y) \exp(\frac{q(x,y)}{\lambda})$$

$$\pi_F(x_F,y) = \exp(-\frac{c(x_F)}{\lambda_F}) \exp(-\frac{b_F(y)}{\lambda_F}) \exp(\frac{q_F(x_F,y)}{\lambda_F}) = \hat{c}(x_F)\hat{b}_F(y) \exp(\frac{q_F(x_F,y)}{\lambda_F})$$

where $\lambda \equiv \lambda_x + \lambda_y$, $\lambda_F \equiv \lambda_{x_F} + \lambda_y$, $\hat{a}(x) \equiv \exp(-\frac{a(x)}{\lambda})$, $\hat{b}(y) \equiv \exp(-\frac{b(y)}{\lambda})$, $\hat{c}(x_F) \equiv \exp(-\frac{c(x_F)}{\lambda_F})$, $\hat{b}_F(y) \equiv \exp(-\frac{b_F(y)}{\lambda_F})$.

Note that following equations hold, as before:

$$\exp(-\frac{a(x)}{\lambda_x}) = \frac{f(x)}{\int \exp(\frac{w(x,y)}{\lambda_x})dy} = \frac{\overline{f}(x)}{\exp(\frac{w}{\lambda_x}) + \int \exp(\frac{w(x,y)}{\lambda_x})dy} = \hat{a}(x)^{\frac{\lambda}{\lambda_x}}$$

$$\exp(-\frac{c(x_F)}{\lambda_{x_F}}) = \frac{h(x_F)}{\int \exp(\frac{w_F(x_F,y)}{\lambda_{x_F}})dy} = \frac{\overline{h}(x_F)}{\exp(\frac{w_F}{\lambda_{x_F}}) + \int \exp(\frac{w_F(x_F,y)}{\lambda_{x_F}})dy} = \hat{c}(x_F)^{\frac{\lambda_F}{\lambda_{x_F}}}$$

$$\exp(-\frac{b(y)}{\lambda_y}) = \frac{g(y)}{\int \exp(\frac{r(x,y)}{\lambda_y})dx} = \frac{\overline{g}(y)}{1 + \int \exp(\frac{r(x,y)}{\lambda_y})dx + \int \exp(\frac{V_F(x_F,y)}{\lambda_y})dz} = \hat{b}(y)^{\frac{\lambda}{\lambda_y}}$$

$$\exp(-\frac{b_F(y)}{\lambda_y}) = \frac{g_F(y)}{\int \exp(\frac{V_F(x_F,y)}{\lambda_y})dz} = \frac{\overline{g}(y)}{1 + \int \exp(\frac{r(x,y)}{\lambda_y})dx + \int \exp(\frac{V_F(x_F,y)}{\lambda_y})dz} = \hat{b}_F(y)^{\frac{\lambda_F}{\lambda_y}}$$

As is evident from above, it turns out that $\hat{b}_F(y) = \hat{b}(y)^{\frac{\lambda}{\lambda_F}} = \hat{b}(y)^{\lambda_R}$ must hold, which allows for a simple characterization of the equilibrium under offshoring.

Combining these results, the exogenously given marginal distributions can be expressed as

$$\overline{f}(x) = \hat{a}(x)^{\frac{\lambda}{\lambda_x}} \exp(\frac{w}{\lambda_x}) + \hat{a}(x) \int \hat{b}(y) \exp(\frac{q(x,y)}{\lambda})dy$$

$$\overline{h}(x_F) = \hat{c}(x_F)^{\frac{\lambda_F}{\lambda_{x_F}}} \exp(\frac{w_F}{\lambda_{x_F}}) + \hat{c}(x_F) \int \hat{b}(y)^{\lambda_R} \exp(\frac{q_F(x_F,y)}{\lambda_F})dy$$

$$\overline{g}(y) = \hat{b}(y)^{\frac{\lambda}{\lambda_y}} + \hat{b}(y) \int \hat{a}(x) \exp(\frac{q(x,y)}{\lambda})dx + \hat{b}(y)^{\lambda_R} \int \hat{c}(x_F) \exp(\frac{q_F(x_F,y)}{\lambda_F})dz$$

Similar to the previous section, the equilibrium boils down to solving for the endogenous objects $\hat{a}(x), \hat{b}(y), \hat{c}(x_F)$, taking as given: marginal distributions $\overline{f}(x), \overline{g}(y), \overline{h}(x_F)$, production functions $q(x,y), q_F(x_F,y)$, and the degree of unobserved heterogeneity λ_x, λ_y, λ_{x_F}. The previously described algorithm can still be used – this time, start with an initial $\hat{b}(y)$, plug in to (1) and (2) to obtain $\hat{a}(x), \hat{c}(x_F)$, then plug in to (3) to obtain $\hat{b}(y)$, and repeat until convergence. Again, once $\hat{a}(x), \hat{b}(y), \hat{c}(x_F)$ are found, it is straightforward to recover all other endogenous objects, including the matching function $\pi(x,y), \pi_F(x_F,y)$ as well as wages and profits $w(x,y), r(x,y), w_F(x_F,y), V_F(x_F,y)$.

B.3.4 Global Economy Equilibrium with $\lambda = 0$

Here, we demonstrate the model with $\lambda = 0$ where worker-firm matching is realized on the observable characteristics of workers and firms and draw clear predictions of the model mechanisms. In order to characterize the equilibrium with no unobserved heterogeneities, we begin by showing the following: for any firm that obtains the same profits through offshoring and domestic hires, the best possible domestic or foreign worker match in equilibrium must have exactly the same skill.

Lemma B.1 *Let* $b^* = \inf\{y \mid r(y) = r_F(y)\}$ *and* $b^{**} = \sup\{y \mid r(y) = r_F(y)\}$. *Then,* $x(y) = x_F(y)$, $\forall\, y \in (b^*, b^{**})$.

Proof We begin by showing that if there exists any two distinct firms y_1 and y_2 that are indifferent between offshoring and domestic hires, then any firm y between y_1 and y_2 is also indifferent. And if so, workers that each firm optimally finds from Home and Foreign are equally talented.

Result 1 Suppose $r(y) = r_F(y)$ holds for $y \in \{y_1, y_2\}$. Then, $r(y) = r_F(y)$, $\forall y \in [y_1, y_2]$.

Suppose $\exists y \in (y_1, y_2)$ s.t. $r(y) \neq r_F(y)$. Without loss of generality, we can assume that $\forall y \in (y_1, y_2), r(y) \neq r_F(y)$ by redefining the interval $[y_1, y_2]$. Without loss of generality, assume $r(y) > r_F(y), \forall y \in (y_1, y_2)$. This implies that all firms between y_1 and y_2 matches with domestic workers only. The matching function with foreign workers $x_F(y)$ stays constant: $x'_F(y) = 0$.

Since $x_F(y) = r'_F(y)$, $x'_F(y) = r''_F(y) = 0$. So $r_F(y)$ must be linear in the range of (y_1, y_2). In addition, $r(y)$ must be strictly convex in (y_1, y_2). Combining these two statements, $r(y) < r_F(y), \forall y \in (y_1, y_2)$. But this contradicts with the assumption that $r(y) > r_F(y), \forall y \in (y_1, y_2)$. Q.E.D.

Result 2 Suppose $r(y) = r_F(y)\ \forall y \in [y_1, y_2]$. Then, $x(y) = x_F(y)$, $\forall y \in [y_1, y_2]$.

Suppose that firm y is indifferent between matching with the best possible worker x from Home and x_F from Foreign given the wage schedules $w(x)$ and $w_F(x_F)$. Recall the matching function $\mu : X \to Y$ where $\mu(X) = \tilde{G}^{-1}(\tilde{F}(X))$ and $\mu^{-1}(Y) = \tilde{F}_P^{-1}(\tilde{G}(Y))$ with $\tilde{G}(y) \equiv 1 - G(y)$, $\tilde{F}(x) \equiv 1 - F(x)$. And we similarly define a matching function assigning foreign workers to doemstic firms, $\mu_F(X_F) \to Y$ where $\mu(X_F) = \tilde{G}^{-1}(\tilde{F}_R(X_F))$ and $\mu^{-1}(Y) = \tilde{F}_R^{-1}(\tilde{G}(Y))$ with $\tilde{F}_R(x_F) \equiv 1 - F_R(x_F)$. Using the properties of the wage schedule together with the market clearing conditions, the following two conditions must hold.

$$\mu(x) = w'(x) \text{ and } \mu_F(x_F) = w'_F(x_F) \tag{B.1}$$

$$f_Y(y)dy = f_X(x)dx + f_Z(x_F)dx_F \quad \Leftrightarrow \quad \tilde{F}_Y(y) = \tilde{F}_X(x) + \tilde{F}_Z(x_F) \tag{B.2}$$

As the matching functions can also be written as $\mu'(x) = \frac{dy}{dx}$ and $\mu'_F(x_F) = \frac{dy}{dx_F}$, the market clearing condition can be re-formulated as follows.

$$f_Y(y) = \frac{f_X(x)}{\mu'(x)} + \frac{f_Z(x_F)}{\mu'_F(x_F)} \tag{B.3}$$

Plugging in the assumed matching functions into the indifference condition, the following must hold at all x and x_F that are indifferent to some firm y: $x\mu(x) - w(x) = x_F\mu_F(x_F) - w_F(x_F) - C$. Taking derivatives of both sides where the following is obtained.

$$(x\mu'(X) + \mu(x) - w'(x))dx = (x_F\mu_F'(x_F) + \mu_F(x_F) - w_F'(x_F))dx_F \quad \text{(B.4)}$$

As $\mu(x) = w'(x)$ and $\mu_F(x_F) = w'(x_F)$, as well as $\mu'(x) = \frac{dy}{dx}$ and $\mu_F'(x_F) = \frac{dy}{dx_F}$, it follows that $x = x_F$ must hold. That is, when a firm y optimally chooses the best domestic worker x and the best foreign worker x_F, in equilibrium these workers must have exactly the same skill. Q.E.D.

Next, we show that even when $\overline{x} = \overline{x}_F = \infty$, the firm with the highest productivity that finds indifference between Home and Foreign optimally matches with the best possible workers.

Result 3 Suppose $\overline{x} = \overline{x}_F = \infty$. Then, $y(\overline{x}) = y(\overline{x}_F) = \overline{y}$.

Suppose $\overline{x} = \infty$ and $y(\overline{x}) = y^* < \overline{y}$ such that all $y \geq y^*$ is matched with x_F. Using the wage functions, $y(\overline{x}) = w'(x) = y^*$, $w(x) \approx y^*x + k$. For $\forall y > y^*$, $r(y) = \max_x xy - w(x) = \max_x (y - y^*)x + k = \infty$ holds. All $y > y^*$ chooses x_F with $r_F(y) > r(y)$. As there is positive assortative matching between $(y^*, \overline{y})$ and $(x_F^*, \overline{x}_F = \infty)$. That is, any finite y is matched with finite z. Thus, $r_F(y)$ is finite, which means $r(y) > r_F(y)$. Then, $\overline{x} = \infty$ is a better match to $y > y^*$ than some finite x_F. This is a contradiction. Therefore, $\overline{x}$

must be matche with $\overline{y}$. Q.E.D.

Use Result 3 and Result 1 to show that $r(y) = r_F(y)$, $\forall y \geq b^*$. Then, using Result 2 we can show that, $x(y) = z(y)$, $\forall y \geq b^*$. Q.E.D.

It also follows from Lemma 1 that the wage difference between the two countries should equal the fixed cost C. Thus, for each occupation, when the cost of offshoring becomes negligible ($C \to 0$), which indicates a convergence to a perfectly integrated world economy, the wage profile does not differ between workers from home and foreign within these jobs.

Corollary B.2 *The wage difference between Home and Foreign is equivalent to the cost of offshoring.*

$$w(x) - w_F(x) = C \quad \textit{for} \quad a_2^* \leq x$$

Proof From Lemma 1, it follows immediately that $q(x, y) = q(x_F, y)$ holds. Using the indifference condition, it is straightforward to see that, $w(x) = w_F(x_F) + C$ holds. Thus, the two wage schedules must be parallel with each other with a constant gap equal to the fixed cost of offshoring C. Q.E.D.

Hence, the equilibrium matching functions $\mu(x)$ and $\mu_F(x_F)$ that map domestic workers and foreign workers to domestic firms, respectively, are as follows.

$$\mu(x) = \begin{cases} \tilde{G}^{-1}(\tilde{F}(x) - \tilde{F}(a_1^*) + \tilde{G}(\underline{y})) & \text{for} \quad a_1^* \leq x \leq a_2^* \\ \tilde{G}^{-1}(\tilde{H}(x)) & \text{for} \quad a_2^* \leq x \end{cases} \tag{B.5}$$

$$\mu_F(x_F) = \tilde{G}^{-1}(\tilde{H}(x_F)) \quad \text{for} \quad a_2^* \leq x_F \tag{B.6}$$

We denote the c.d.f of workers and firms from Home as $F(x)$ and $G(y)$ respectively. Also, $H(x)$ indicates the aggregate worker endowment; and $\tilde{G}(y) \equiv 1 - G(y)$, $\tilde{F}(x) \equiv 1 - F(x)$. That is, the densities in the aggregate worker endowment in the global economy $h_w(x)$ is the sum of densities from home and foreign.[7] Next, equilibrium wages in the global economy are given as follows, where $q(x, y) = w(x) + r(y)$ and $q(x_F, y) = w_F(x_F) + r(y)$ hold.

$$w(x) = \begin{cases} c_1 + \int_{a_2^*}^{x} \frac{\partial q}{\partial x}(t, \tilde{G}^{-1}(\tilde{H}(t)))dt & \text{when} \quad a_2^* \leq x \\ c_2 + \int_{a_1^*}^{x} \frac{\partial q}{\partial x}(t, \tilde{G}^{-1}(\tilde{F}(t) - \tilde{F}(a_1^*) + \tilde{G}(\underline{y})))dt & \text{when} \quad a_1^* \leq x \leq a_2^* \end{cases} \tag{B.7}$$

$$w_F(x_F) = c_3 + \int_{a_2^*}^{x_F} \frac{\partial q}{\partial x}(t, \tilde{G}^{-1}(\tilde{H}(t)))dt \quad \text{when} \quad a_2^* \leq x_F \tag{B.8}$$

$$r(y) = \begin{cases} c_4 + \int_{b^*}^{y} \frac{\partial q}{\partial y}(\tilde{H}^{-1}(\tilde{G}(s)), s)ds & \text{when} \quad b^* \leq y \quad \text{(offshoring)} \\ c_5 + \int_{b^*}^{y} \frac{\partial q}{\partial y}(\tilde{H}^{-1}(\tilde{G}(s)), s)ds & \text{when} \quad b^* \leq y \quad \text{(domestic hires)} \\ c_6 + \int_{\underline{y}}^{y} \frac{\partial q}{\partial y}(\tilde{F}_P^{-1}(\tilde{G}(s) - \tilde{G}(\underline{y}) + \tilde{F}(a_1^*), s)ds & \text{when} \quad \underline{y} \leq y \leq b^* \quad \text{(domestic hires)} \end{cases} \tag{B.9}$$

Here, $w(a_1^*) = \underline{w}$, $w_F(a_2^*) = \underline{w}_F$, $c_1 + c_5 = q(a_2^*, b^*)$, $c_3 + c_4 = q(a_2^*, b^*) - C$, and $c_2 + c_6 = q(a_1^*, \underline{y})$ hold. Since firms above b^* are indifferent between offshoring and matching with domestic workers, which imposes $c_4 = c_5$, it follows that the difference in wages between foreign and home must equal to the fixed cost of offshoring.[8]

[7] Note that we do not normalize the mass of workers in the aggregate endowment to 1 since an increase in the mass of workers in the global economy is an important channel that alters worker-firm matches. As before, the following notation is used: $\tilde{H}(x) \equiv 2 - H(x)$.

[8] This equation determines the threshold b^*, which also subsequently pins down a_1^* and a_2^*. See Appendix C for derivations.

As a result of offshoring, workers above a_2^* match with firms that are indifferent between matching with workers from home and foreign; those between a_1^* and a_2^* match with firms that do not offshore; and those below a_1^* switch out to the traditional sector.[9]

Theorem B.3 *For a given offshoring cost C, there exists a unique equilibrium of the global economy: there exists a threshold $b^* = \tilde{G}^{-1}(\tilde{H}(a_2^*))$ above which firms perform offshoring and below which firms domestically hire workers. The equilibrium demonstrates positive assortative matching with strictly convex profiles of wages and profits.*

Proof From Lemma 1, some firm y that is indifferent between offshoring and domestic hires and optimally chooses either x or x_F, finds the same quality of workers, $x = x_F$. Also, from Corollary 1, the wage difference between Homa and Foriegn is equal to the cost of offshoring C. Using the indifference condition, it immediately follows that, $\mu(x) = \mu_F(x_F)$ for these firms that are indifferent. Therefore, the market clearing condition can be re-arranged and simplified as follows:

$$g(y) = \frac{f_P(x)}{\mu'(x)} + \frac{f_R(x_F)}{\mu'_F(x_F)} = \frac{f_P(x) + f_R(x)}{\mu'(x)} \quad \Leftrightarrow \quad \mu'(x) = \frac{f_P(x) + f_R(x)}{g(y)} \tag{B.10}$$

Thus, for a particular offshoring cost C, there exists a threshold b^* above which firms find indifference in workers from home and foreign, $a_2^* \leq x$ and $a_2^* \leq x_F$ where $\tilde{G}(b^*) = \tilde{H}(a_2^*)$ with $\tilde{H}(x) = \tilde{F}(x) + \tilde{F}_R(x_F)$. As for firms

[9] With firms offshoring, the traditional sector expands and improves in the workers' skill quality as those who become unmatched in the competitive sector and sort into the traditional sector are more skilled than those who were in the traditional sector under the closed economy assumptions.

below b^* who match with domestic workers $a_1^* \le x \le a_2^*$, as they cannot afford the cost of offshoring, the market clears, $\tilde{G}(b^*) - \tilde{G}(\underline{y}) = \tilde{F}(a_2^*) - \tilde{F}(a_1^*)$ and the following holds.

$$g(y) = \frac{f_P(x)}{\mu'(x)} \quad \text{and} \quad \mu(x) = w'(x) \text{ for } a_1^* \le x \le a_2^* \tag{B.11}$$

Taking anti-derivatives of $\mu'(x)$ and subsequently for $w'(x)$ provides the equilibrium matching and wage schedules for each interval of firms and the corresponding workers. Note that the profile of wages and profits are strictly convex as $w'(x)$ increases in x and $r'(y)$ inreases in y due to the nondecreasing matching function $\mu(x)$ derived from the supermodular production function $q(x, y)$. This indicates the existence of the equilibrium. Furthermore, Lemma 1 shows that $x = x_F$ is a necessary condition for an equilibrium in order to sustain the indifference condition in this range of firms. Thus, the equilibrium is unique. Q.E.D.

Derivation of Wage and Profits In matching problems, wage and profits are endogenously shared as follows:

$$w(x) = c_1 + \int_{x_0}^{x} \frac{\partial q}{\partial x}(t, \mu(t))dt \tag{B.12}$$

$$r(y) = c_2 + \int_{\mu(x_0)}^{y} \frac{\partial q}{\partial y}(\mu^{-1}(s), s)ds \tag{B.13}$$

with $c_1 + c_2 = q(x_0, \mu(x_0))$. As shown in Galichon (2016), we can use properties of *comonotonicity* of random variables to derive the wage and profit equations with different distributional assumptions for the endowment. Suppose that X has a c.d.f of $F(X)$ and Y, a c.d.f of $G(Y)$.

Definition B.4 *Random variables X and Y are comonotone if there is U following uniform distribution such that $X = F_P^{-1}(U)$ and $Y = G^{-1}(U)$. Equivalently, X and Y are said to exhibit positive assortative matching.*

Thus, $X = F_P^{-1}(U)$ and $Y = G^{-1}(U)$ match in a positively assortative way with a nondecreasing assignment function $Y = G^{-1}(F(X))$: equilibrium matching between X and Y and that between $F(X)$ and $G(Y)$ are equivalent. Wages and profits can be solved numerically using c.d.f's of whichever distribution we assume to have as follows with a matching function $\mu(X) = G^{-1}(F(X))$.

B.3.5 Global Economy Equilibrium: Uniform Distribution

Assuming uniform distributions for domestic and foreign endowments, domestic worker's skill x, firm's productivity y, and skill of foreign worker composites x_F are realizations of X, Y, and X_F, respectively where $X \sim U[0,1]$, $Y \sim U[0,1]$, and $X_F \sim U[0,\sigma]$.[10] with mass $\rho > 1$ where $\sigma > 1$. Using the same production technology as before, the optimal assignment of workers to firms in a global economy equilibrium is given as,

$$\mu(x) = \begin{cases} \frac{\rho+\sigma}{\sigma}x - \rho & \text{when} \quad a_2 \le x \le 1 \\ x - \frac{(\sigma - b_1)\rho}{\sigma+\rho} & \text{when} \quad a_1 \le x \le a_2 \end{cases}$$

$$\mu_F(x_F) = \begin{cases} \frac{\rho}{\sigma}x_F + (1-\rho) & \text{when} \quad 1 \le x_F \le \sigma \\ \frac{\rho+\sigma}{\sigma}x_F - \rho & \text{when} \quad a_2 \le x_F \le 1 \end{cases}$$

[10]The situation where $\sigma > 1$ arises when the worker composites from the foreign country delivers a higher quality of skill compared to home workers. In particular, if the wage differences between home and foreign is significantly large, allowing domestic firms to hire foreign workers in greater quantities, it is possible that the skill output of these foreign worker composites is high enough that there are no home workers to compete with the corresponding level of skill output.

where $a_1 = \underline{y} + \frac{(\sigma - b_1)\rho}{\sigma+\rho}$; $a_2 = \frac{(\rho+b_1)\sigma}{\sigma+\rho}$ and $b_2 = \frac{\sigma - \rho\sigma + \rho}{\sigma}$. b_1 will be determined using wages and profits below. Again, equilibrium wages and profits under global economy are given as follows where $q(x,y) = w(x) + r(y)$ and $q(x_F, y) = w_F(x_F) + r(y)$ hold.

$$w(x) = \begin{cases} \frac{1}{2}(\frac{\sigma+\rho}{\sigma})x^2 - \rho x + d_1 & \text{when} \quad a_2 \leq x \leq 1 \\ \frac{1}{2}x^2 - \frac{(\sigma-b_1)\rho}{\sigma+\rho}x + d_2 & \text{when} \quad a_1 \leq x \leq a_2 \end{cases}$$

$$w_F(x_F) = \begin{cases} \frac{1}{2}\frac{\rho}{\sigma}x_F^2 + (1-\rho)x_F + d_3 & \text{when} \quad 1 \leq x_F \leq \sigma \\ \frac{1}{2}(\frac{\sigma+\rho}{\sigma})x_F^2 - \rho x_F + d_4 & \text{when} \quad a_2 \leq x_F \leq 1 \end{cases}$$

$$r(y) = \begin{cases} \frac{1}{2}\frac{\sigma}{\rho}y^2 + \frac{\sigma(\rho-1)}{\rho}y + d_5 & \text{when} \quad b_2 \leq y \leq 1 \quad \text{(offshore)} \\ \frac{1}{2}(\frac{\sigma}{\sigma+\rho})y^2 + (\frac{\sigma\rho}{\sigma+\rho})y + d_6 & \text{when} \quad b_1 \leq y \leq b_2 \quad \text{(offshore)} \\ \frac{1}{2}(\frac{\sigma}{\sigma+\rho})y^2 + (\frac{\sigma\rho}{\sigma+\rho})y + d_7 & \text{when} \quad b_1 \leq y \leq b_2 \quad \text{(domestically hire)} \\ \frac{1}{2}y^2 + \frac{(\sigma-b_1)\rho}{\sigma+\rho}y + d_8 & \text{when} \quad \underline{y} \leq y \leq b_1 \quad \text{(domestically hire)} \end{cases}$$

where $d_1 = \frac{\sigma}{\sigma+\rho}\frac{\rho^2 - b_1^2}{2} + \frac{b_1^2}{2} - \frac{\underline{y}^2}{2} + \underline{w}$; $d_2 = \frac{1}{2}(\frac{(\sigma-b_1)\rho}{\sigma+\rho})^2 - \frac{\underline{y}^2}{2} + \underline{w}$; $d_3 = \frac{\sigma}{\sigma+\rho}\frac{\rho^2-b_1^2}{2} + \underline{w}_F - \frac{1}{2}$; $d_4 = \frac{\sigma}{\sigma+\rho}\frac{\rho^2-b_1^2}{2} + \underline{w}_F$; $d_5 = \frac{1}{2}\frac{(\sigma+\rho-\sigma\rho)^2}{(\sigma+\rho)\rho} + \frac{b_1^2}{2}\frac{\sigma}{\sigma+\rho} - \underline{w}_F - C$; $d_6 = \frac{b_1^2}{2}\frac{\sigma}{\sigma+\rho} - \underline{w}_F - C$; $d_7 = \frac{b_1^2}{2}\frac{\sigma}{\sigma+\rho} - \frac{b_1^2}{2} + \frac{\underline{y}^2}{2} - \underline{w}$; and $d_8 = \frac{\underline{y}^2}{2} - \underline{w}$. Using the indifference condition that firms between b_1 and b_2 are indifferent between offshoring and hiring domestically (i.e. $d_6 = d_7$), b_1 is pinned down: $b_1 = \sqrt{2(C + \frac{1}{2}\underline{y}^2 + \underline{w}_F - \underline{w})}$.

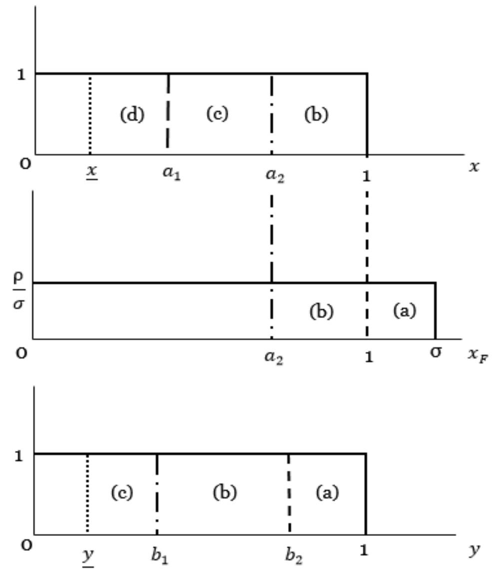


Figure 24: Equilibrium Matching with Offshoring

Each panel in the figure above illustrates endowments of domestic workers, foreign workers, and firms, respectively. In the bottom panel, area (a) corresponds to firms that only offshore; area (b), firms that are indifferent between offshoring and domestic hires; and area (c), firms that do not offshore. Firms below $\underline{y}$ exit the market. Corresponding labels in the workers' endowment indicate the respective worker matches. Area (d) indicate the mass of workers driven out to the traditional sector due to worker competition from abroad.

B.3.6 Foreign Endowment with Lower Quality

The result of the model based on the assumed mechanism holds even when we assume an introduction of a foreign endowment that provides an inferior quality of skill output where the bottom end of the domestic distribution of workers face competition from abroad. While the high productive firms find no incentives to offshore, the low productive ones would be indifferent between offshoring and domestic matching, as they face the same quality of workers from home and foreign, assuming they can afford the cost of offshoring.[11] Suppose for a particular fixed cost of offshoring, there exists an

[11]When the cost of offshoring is high enough or the skill quality of foreign endowment is far too inferior that there is no overlap with the domestic workers, the result of the model is equivalent to that of a closed economy.

interval of firms with relatively low productivity that choose to offshore. While those who face foreign competition in the distribution of workers have changed, the fact that the economy is provided with additional supply of workers and therefore driving out the least productive ones in the domestic labor market does not change. Again, the result of the model delivers skill-upgrading and improved homogeneity; however, the magnitude of the change from globalization would be smaller compared to a situation where the economy faces a more competitive foreign labor force, as assumed in the main part of the model.

B.3.7 Comparison with Technological Change: Automation

Many studies that investigate the phenomena of "hollowing out" of the labor market or the disappearance of middle-skill occupations discuss such job replacements with respect to automation technology or offshoring. Although the focus of this paper mainly lies in examining the effect of offshoring on labor market outcomes, here, we briefly discuss how the model can be applied to examining a situation where occupations are replaced by machines instead of a distribution of workers.

Assuming constant productivity of machines brings about identical structure as to thinking about domestic firms facing a distribution of domestic workers and a *Dirac delta* measure of constant productivity generated by machines,

$$\delta(x_T) = \begin{cases} +\infty, & x_T = \kappa \\ 0, & x_T \neq \kappa \end{cases}$$

where $\int_{-\infty}^{+\infty} \delta(x_T)dx_T = 1$. Thus, firms can buy a machine instead of hiring a worker with occupation o, which generates a productivity of κ with cost C_T. Thus, when a firm decides to employ a machine instead of a worker, the profit is $y\kappa - C_T$ where $\kappa < \overline{\kappa}$, and $C_T \geq w(\underline{x})$, and the firm's choice of technology adoption depends on the following:

$$\max\left[y\kappa - C_T, xy - w(x)\right] \quad \text{where} \quad \kappa < \overline{\kappa}, \quad C_T \geq w(\underline{x}) \tag{B.14}$$

Then, we solve for an equilibrium where technology adoption decision differs across the distribution of firms. Firms at the top-end who face workers with better productivity than the machine's output and the firms in the bottom-end who cannot afford the cost of adopting machines would continue to match with workers; and firms in the middle will end up adopting machines. There exist thresholds a_3, b_3 such that positive assortative matching is optimal with matching:

$$\mu(x) = \begin{cases} \tilde{G}^{-1}(\tilde{F}(x) - \tilde{F}(a_3^*)) & \text{for} \quad a_3^* \leq x \leq a_4^* \\ \tilde{G}^{-1}(\tilde{F}(x)) & \text{for} \quad x \geq a_4^* \end{cases} \tag{B.15}$$

where the least productive workers $x \leq a_3^*$ remains unmatched due to technology adoption by firms in the middle of the distribution. Next, we derive the profiles of wage and profit using optimal conditions above. We set the value of outside option to remain unmatched for workers to be 0. Since $h(x, y) = w(x) + r(y)$ must hold, equilibrium wages and profits are as follows:

$$w(x) = \begin{cases} c_1' + \int_{a_3^*}^{x} \frac{\partial h}{\partial x}(t, \tilde{G}^{-1}(\tilde{F}(t) - \tilde{F}(a_3^*)))dt & \text{when} \quad a_3^* \leq x \leq a_4^* \\ c_2' + \int_{b_3}^{x} \frac{\partial h}{\partial x}(t, \tilde{G}^{-1}(\tilde{F}(t)))dt & \text{when} \quad x \geq a_4^* \end{cases} \tag{B.16}$$

$$r(y) = \begin{cases} c_3' + \int_{\underline{y}}^{y} \frac{\partial h}{\partial y}(a_3^* + \tilde{F}^{-1}(\tilde{G}(s)), s)ds & \text{when} \quad y \leq b_1^* \\ c_4' + \int_{b_4}^{y} \frac{\partial h}{\partial y}(\tilde{F}^{-1}(\tilde{G}(s)), s)ds & \text{when} \quad y \geq b_2^* \end{cases} \tag{B.17}$$

where $c_1' + c_3' = h(a_3^*, \underline{y})$ and $c_2' + c_4' = h(a_4^*, b_2^*)$. Examining how firms $y = b_1^* = \tilde{G}^{-1}(\tilde{F}(a_4^*) - \tilde{F}(a_3^*))$ and $y = b_2^* = \tilde{G}^{-1}(\tilde{F}(a_4^*))$ should be indifferent between employing a worker versus adopting technology, allows the model to determine the output and cost of the machine that defines such equilirbium:

$$\begin{aligned} \tilde{G}^{-1}(\tilde{F}(a_4^*) - \tilde{F}(a_3^*))\kappa - C_T = a_4^*(\tilde{G}^{-1}(\tilde{F}(a_4^*) - \tilde{F}(a_3^*))) - w(a_4^*) \\ \tilde{G}^{-1}(\tilde{F}(a_4^*))\kappa - C_T = a_4^*(\tilde{G}^{-1}(\tilde{F}(a_4^*))) - w(a_4^*) \end{aligned} \tag{B.18}$$

Solving this system of two equations gives the thresholds a_3, b_3 as a function of technological output κ and its cost C_T where the interval of firms that adopt technology $\tilde{G}^{-1}(\tilde{F}(a_4^*) - \tilde{F}(a_3^*)) \leq y \leq \tilde{G}^{-1}(\tilde{F}(a_4^*))$ increases with κ and decreases with C_T.

Example: Uniform Distribution Again, we assume uniform distributions for both skill endowments: $X \sim U[0,1]$, $Y \sim U[0,1]$. Analytically solving for the matching function, we obtain the following:

$$\mu(x) = \begin{cases} x \quad \text{for} \quad B \leq x \leq 1 \\ x - (B - A) \quad \text{for} \quad B - A \leq x \leq B \end{cases} \tag{B.19}$$

Again, when technology is costless with productivity greater than any worker in this economy, such occupation is completely substituted by automation while in the opposite case, there will be no technology adoption. Firms in the top ($B \leq y \leq 1$) and the bottom ($0 \leq y \leq A$) continue to employ occupation o while those in the middle ($A \leq y \leq B$) do not.

Equilibrium profiles of wages and profits are pinned down as follows:

$$w(x) = \begin{cases} \frac{1}{2}x^2 - \frac{1}{2}(B^2 - A^2) \quad \text{when} \quad B \leq x \leq 1 \\ \frac{1}{2}x^2 - (B - A)x + \frac{1}{2}(B - A)^2 \quad \text{when} \quad B - A \leq x \leq B \end{cases}$$

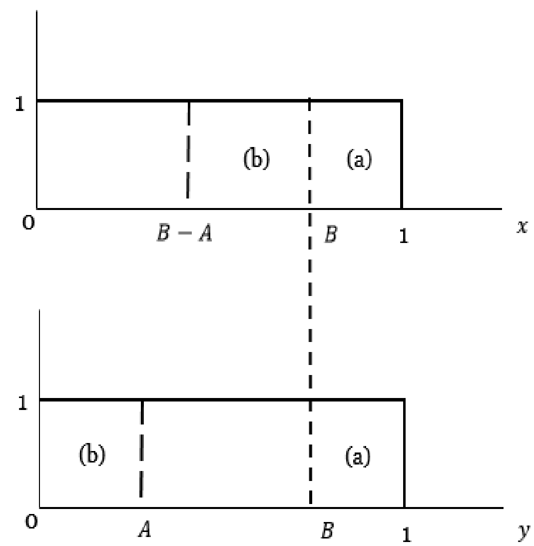


Figure 25: Equilibrium Matching with Technology Adoption

$$r(y) = \begin{cases} \frac{1}{2}y^2 + (B-A)B \quad \text{when} \quad B \le y \le 1 \\ y\kappa - C_T \quad \text{when} \quad A \le y \le B \\ \frac{1}{2}y^2 + (B-A)y \quad \text{when} \quad 0 \le y \le A \end{cases}$$

Furthermore, using indifferent conditions for firms between matching with workers and employing machines at thresholds A and B allows the model to analytically pin down values of A and B:

$$A = \sqrt{2C_T}, \quad B = \kappa$$

Due to changes in the demand for workers with technological change, workers $0 \le x \le A$ are unmatched, and even for those who are matched $A \le x \le 1$ undergo a wage loss. Firms in the top and bottom continue to hire workers with occupation o, and the mass of firms that decide to employ machines increases as the productivity of machine goes up ($\kappa \uparrow$) and the cost goes down ($C_T \downarrow$). Also, firms in the bottom match with better quality workers than before.