FIRM-TO-FIRM RELATIONSHIPS
IN INTERNATIONAL TRADE

KEVIN LIM

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ECONOMICS
ADVISER: STEPHEN REDDING

MAY 2016
Abstract

This dissertation studies the role of firm-to-firm relationships in international trade, with a focus on two margins: (1) the dynamics of arms-length buyer-seller relationships, and (2) technology transfers within multinational firms.

Chapter 1 studies the quantitative implications of frictions in the creation and destruction of firm-to-firm trading relationships for aggregate patterns of trade. I develop a structural model of trade in which the network of firm-level input-output linkages is endogenously determined, and use data on relationships between US firms to estimate the model's parameters. Counterfactual simulations of the model show that the endogenous adjustment of relationships dynamically amplifies the effects of changes in trade costs on trade volumes and welfare by more than three times.

Chapter 2 documents that the hazard rate of relationship termination between firms in the US decreases with the age of the relationship as well as with buyer and seller size. I develop a general equilibrium model of trade which replicates this finding through a simple mechanism: firms engaged in trade receive relationship-specific productivity shocks that are both persistent and age-dependent. Using numerical simulations of the model, I show that these two features of the productivity process matter for the gains from trade, with welfare responding more strongly to trade costs when productivity growth is slower and when productivity shocks are more persistent.

Chapter 3 studies the interaction between intellectual property rights (IPR) protection by governments and technology transfer (TT) by multinational firms to their subsidiaries. I develop a two-country model of trade and foreign direct investment, and show analytically the existence of two kinds of inefficiencies: one arising from governments' choices of IPR policies, and another from multinational firms' choices of TT. I find that the IPR and TT inefficiencies are characterized by under-provision of IPR protection and TT respectively, and that they are comparable in magnitude, amounting to as much as 4% of the gains from openness to trade and FDI.
Acknowledgements

I am extremely grateful to Stephen Redding and Gene Grossman, whose remarkable passion and perspicacity for the field of economics not only influenced but inspired my own love for the discipline. I learned from them not only about economics but about the world of economists, and how to navigate the currents of both. More than providing sage advice, they constantly encouraged me to better myself, to push the boundaries of possibility, and to never stop learning. They were and will continue to be role models whose standards of excellence I will long aspire towards.

I am also thankful for the extraordinary trade community at Princeton, which provided an intellectual environment that was second-to-none. Esteban Rossi-Hansberg, Oleg Itskhoki, and Ezra Oberfield in particular challenged my understanding of economics on a weekly basis with razor-sharp critical thinking and thoughtful comments. I greatly benefited also from interactions with the many brilliant minds that visited the international economics section over the last three years - Davin Chor in particular, who has become both colleague and friend.

I count myself fortunate to be graduating from Princeton alongside a stellar cohort of economists. David Nagy, Mark Razhev, Ricardo Reyes-Heroles, and Sharon Traiberman were not only amazing intellectual companions with similar research interests, but also fast friends with whom I endured and survived the vagaries of life as a graduate student. Pedro Olea, Moshe Katzwer, Olivier Darmouni, and Sungki Hong were amongst many wonderful friends with whom I shared unforgettable times, in the basement of the economics department, on the football field, and in the handful of decent places to dine around town.

Finally, I am indebted to my family for their love and support over the years, and to Maple for being a constant reminder of what it is that is truly beautiful and important in life. Without them, this dissertation would not only have been impossible to complete - it would have been a meaningless endeavor as well.
## Contents

1. **Firm-to-firm Trade in Sticky Production Networks**  
   1.1 Introduction  
   1.2 Data and Empirical Regularities  
   1.2.1 Data  
   1.2.2 Empirical regularities  
   1.3 Static Model  
   1.3.1 Basic environment  
   1.3.2 Static market equilibrium  
   1.3.3 Embedding geography  
   1.4 Dynamics and Endogenous Network Formation  
   1.4.1 Dynamics of firm matching  
   1.4.2 Dynamic market equilibrium  
   1.4.3 Properties of the steady-state  
   1.5 Numerical Analysis  
   1.5.1 Parametric assumptions  
   1.5.2 Parameter estimation  
   1.5.3 Results  
   1.6 Counterfactuals  
   1.6.1 Trade cost shocks  
   1.6.2 Relationship cost shocks  
   1.6.3 Idiosyncratic fluctuations and aggregate dynamics  
   1.6.4 The importance of rational expectations
C Appendix to Chapter 3 187

C.1 Derivation of the Aggregate Price Index for the Case with General CES Preferences 187

Bibliography 190
Chapter 1

Firm-to-firm Trade in Sticky Production Networks

1.1 Introduction

Many of the goods and services that are traded between firms lack centralized markets or intermediaries facilitating their exchange.\(^1\) Such firm-to-firm trade is therefore contingent on firms’ active management of direct relationships with their customers and suppliers. This can often be an integral yet costly aspect of operations. Market analysts estimate, for example, that firms in the United States spent more than $10bn in 2014 on customer relationship management (CRM) and supply chain management (SCM) software systems alone.\(^2\) Motivated by this observation, this paper studies the quantitative implications of frictions in the creation and destruction of firm-to-firm trading relationships (henceforth referred to as relationship stickiness) for aggregate output and trade across locations. When it is costly to form and adjust trading relationships, how do firms vary their selection of trade partners in response to changes in the economic environment? Consequently, how do these decisions translate into the responses of aggregate output and trade to macroeconomic shocks?

To answer these questions, I develop a structural model of trade between heterogeneous

---

\(^1\)This is a point dating back at least to Rauch (1999), who was one of the first to argue using empirical evidence for a view of trade as characterized by networks of buyers and sellers rather than by frictionless markets.

\(^2\)See for instance the reports by Gartner, Inc. (2014a, 2014b). Software platforms marketed by industry leaders such as Salesforce and SAP offer solutions for a wide range of relationship management tasks, such as the organization of contact databases, monitoring of customer and supplier financial information, tender and contract management, supplier performance assessment, and so on. This highlights the potentially complex nature of the costs that firms face in managing business relationships, of which expenses on software are only one particular facet.
firms in which the network of firm-level input-output linkages is determined both dynamically and endogenously. In the model, monopolistically-competitive firms in different locations produce output using a technology exhibiting constant returns to scale and a constant elasticity of substitution across inputs. Access to additional customers therefore increases the variable profit of a firm, while access to additional suppliers lowers its marginal cost. These incentives to form trading relationships are counterbalanced by assuming that firms face a fixed cost per active relationship, and that the opportunity to activate or terminate each relationship arrives randomly over time. The static fixed cost creates a meaningful tradeoff for firms in their selection of relationships, while the dynamic opportunity cost makes these selection problems forward-looking. These assumptions therefore allow the model to generate rich predictions regarding the distributions of customers and suppliers across firms, the assortativity of matching between firms, the persistence of firm-to-firm relationships across time, as well as the differential responses of these patterns to aggregate shocks in the short-versus the long-run.

At the same time, the model remains computationally tractable. Cross-sectional firm-level variables are pinned down by sufficient statistics that are easily computed for any input-output architecture, and solving for the model's transition dynamics under rational firm expectations typically requires about one hour on a standard personal computer. Computational tractability in turn permits structural estimation of the model and the quantitative analysis of counterfactual exercises. Using both cross-sectional and panel data on firm-level trading relationships in the United States (obtained from Standard and Poor's Capital IQ and Compustat platforms), I estimate the model's parameters via a simulated method of moments technique. I show that the model is able to replicate the majority of empirical regularities that I document in the paper, with larger firms tending to: (1) have more suppliers and customers; (2) trade with larger and more connected firms; and (3) have

---

3The fixed relationship cost is analogous to the fixed cost of exporting in Melitz (2003), except that here it is paid at the firm-to-firm level. The random arrival of opportunities to reset the status of a relationship is analogous to the price reset shock in Calvo (1983), except that here firms are constrained in their ability to adjust relationships along the extensive rather than the intensive margin.
trading relationships that are more persistent. I then study the quantitative responses of trade patterns and welfare to counterfactual changes in trade costs, changes in relationship costs, and idiosyncratic firm-level fluctuations.

The key findings of this paper are as follows. First, the endogenous adjustment of firm-to-firm trading relationships dynamically amplifies the effects of changes in variable trade costs on aggregate interfirm trade and welfare. Intuitively, when relationships are sticky, a fall in trade costs induces firms to not only buy more from existing trade partners but also to accumulate more trade partners over time. Quantitatively, the magnitude of this amplification effect is large: the elasticities of aggregate trade and welfare with respect to trade costs are estimated to be between three to four times higher in the long-run than in the short-run. This suggests that taking into account the timing of policies aimed at reducing trade costs can be important, and in particular provides a rationale for quick rather than gradual reduction of trade barriers.

Second, reductions in relationship fixed costs have stronger effects on aggregate trade and welfare than cost-equivalent reductions in variable trade costs. Consider a planner with an exogenous subsidy budget who can choose to either subsidize the intensive margin of trade (through export or import subsidies for example) or to subsidize the fixed cost of each active relationship (by mitigating communication or meeting costs for instance). The model’s counterfactuals predict that the latter option would generate increases in aggregate trade and welfare that are more than 50% larger in the long-run than the gains that would be realized under the former option, with similar rates of dynamic adjustment. This implies that policy measures which reduce the frictions that firms face in establishing trading relationships can be equally as if not more cost-effective than traditional trade policy instruments in terms of their ability to increase trade and welfare. This may be of particular interest for policymakers who find the direct promotion of firm-to-firm relationships to be less politically objectionable.

---

4The magnitude of this dynamic amplification effect is similar to the size of the amplification effect that Alessandria, Choi, and Ruhl (2015) estimate, which in their model is generated by firm-level investments in lowering export costs that respond endogenously to changes in trade barriers.
than adjustments in tariff barriers.

Third, when firm relationships are sticky, both macroeconomic shocks as well as idiosyncratic fluctuations in firm-level characteristics can have effects on aggregate trade and output that are not only large but persistent as well. Following a decline in trade or relationship costs, the dynamic adjustments of trade volumes and welfare typically exhibit half-lives of around two years. Similarly, idiosyncratic shocks to firm-level characteristics generate declines in trade and welfare that dissipate gradually with a half-life of around two years, even when such shocks leave the aggregate distribution of firm characteristics unchanged (and therefore would have no aggregate effect in a frictionless model). The endogenous adjustment of firm-to-firm relationships due to relationship stickiness therefore imparts a high degree of inertia to the dynamics of aggregate outcomes, whether these dynamics are driven by macroeconomic shocks or by idiosyncratic firm-level fluctuations.

Finally, a simple policy exercise shows that subsidies to the cost of maintaining relationships with customers financed by a tax on imports can improve welfare. This suggests that firms in the market equilibrium are trading too much at the intensive margin and too little at the extensive margin relative to the social optimum. I show analytically that inefficiency of the market equilibrium stems from two sources. The first is the standard markup distortion arising from firm monopoly power. The second is a novel source of inefficiency generated by the network structure of production (often referred to as a network externality): firms select relationships based only on profit-maximizing criteria and do not internalize the value of each relationship to all other firms in the network.

The modeling of frictions in firm-level trading relationships in this paper is most closely related to the models of Oberfield (2015) and Chaney (2014, 2015). In both of these models, potential buyer-supplier pairs also receive trading opportunities at a finite rate, and the network of firm-level input-output linkages is an endogenous and dynamic outcome of this exogenous stochastic process. However, there are two key differences between these frameworks and the model that I develop. First, I introduce a fixed cost to relationship
formation, whereas activating a trading relationship is costless for firms in both Oberfield (2015) and Chaney (2014, 2015).\textsuperscript{5} It is this costly nature of relationship formation that generates the dynamic amplification of shocks discussed above, which I find to be quantitatively large. Explicitly modeling the costs of relationship formation also allows me to study the effect of reductions in such costs on aggregate patterns of output and trade.\textsuperscript{6} Second, both Oberfield (2015) and Chaney (2014, 2015) only partially model variations in the extensive margin of firm-to-firm trading relationships.\textsuperscript{7} These models therefore lose identifying power that might otherwise be gained by exploiting richer heterogeneity in empirically observed networks of firm-to-firm trade. For this reason, I construct a model that simultaneously generates non-trivial predictions about the distributions of both customers and suppliers across firms.

The theory developed in this paper is also related to the broader theoretical literature on social and economic network formation, within which there are two qualitatively different approaches to modeling the formation of ties between atomistic agents.\textsuperscript{8} The first approach posits an exogenous stochastic algorithm for the formation of links, and then proceeds to study the resulting network properties.\textsuperscript{9} As these models of network formation are non-structural, however, they cannot be used to study how networks of trade between firms respond to changes in economic incentives. The second approach to modeling network formation assumes that the creation and destruction of links are the result of strategic in-

\textsuperscript{5}In Oberfield (2015), firms always have the option of buying from suppliers that they could have traded with in the past, while in Chaney (2014, 2015), trade occurs automatically once a potential seller acquires contact with a buyer. In both models, there are no fixed costs of trade between firms.

\textsuperscript{6}Bernard, Moxnes, and Saito (2015) and Bernard, Moxnes, and Ulltveit-Moe (2015) explicitly model fixed relationship costs between firms in the same way that I do here. However, these papers model only the static formation of relationships between one group of buyers and one group of sellers - in essence capturing only one tier of the static network of trade between firms.

\textsuperscript{7}In Oberfield (2015), the number of suppliers per firm is exogenously fixed, while in Chaney (2014, 2015), every firm has the same number of suppliers even though the number of suppliers per firm grows over time.

\textsuperscript{8}See Jackson (2005, 2011) for more in-depth surveys of the network formation literature.

teractions between agents. These game-theoretic approaches therefore explicitly take into account optimizing behavior by the agents constituting the network, but the complexity of solving these models beyond simple illustrative examples precludes quantitative analysis.

The modeling of network formation in this paper can thus be viewed as a combination of the two approaches discussed above, or in the terminology of Currarini, Jackson, and Pin (2010), a combination of “chance and choice”: firms receive the opportunity to adjust relationships according to an exogenous stochastic process, but the activation or termination of a trading relationship conditional on having the opportunity to do so is an endogenous outcome. This hybrid approach is similar in spirit to the dynamic network formation models in Bala and Goyal (2000), Watts (2001), and Jackson and Watts (2002), but within the context of a structural model of trade between heterogeneous producers that can be used for quantitative analysis.

Finally, this paper contributes to several other areas of research. In studying quantitatively how firm-level relationship stickiness affects the responses of aggregate trade to shocks across different time horizons, this paper adds to the already-vast literature on the dynamics of firm-level trade and the estimation of trade elasticities. Although the concept of trade studied in this paper focuses on trade between firms and is not explicitly international in nature, the notion of relationship stickiness applies to firm-to-firm trade in general, whether goods cross national borders or not. Understanding the effects of these frictions on trade within a country therefore also adds to our understanding of their effects on trade between countries. This paper also contributes to the study of how microeconomic shocks translate

10Aumann and Myerson (1988) and Myerson (1991) model network formation as extensive-form and simultaneous move games respectively. Jackson and Wolinsky (1996) adopt a cooperative game theoretic approach, while Kranton and Minehart (2001) study buyer-seller networks in which ascending-bid auctions are used to determine the formation of links.

11Bala and Goyal (2000), Watts (2001), and Jackson and Watts (2002) also assume for tractability that agents are myopic in their decisions about which links to form, whereas firms are my model optimally select relationships given rational expectations about the future costs and benefits of each relationship.

into aggregate fluctuations. Gabaix (2011) and Acemoglu et al (2012) argue that the firm size distribution and the network structure of linkages between sectors matter for how idiosyncratic firm- and sector-level shocks translate into aggregate movements, but do not seek to explain what determines these characteristics of the economy in the first place. The model that I develop endogenizes both the firm size distribution as well as the firm-level input-output architecture, and therefore can be used to study the two-way interaction between these characteristics and aggregate fluctuations.

The outline of this paper is as follows. In section 1.2, I describe the data and document empirical regularities in the US production network. In section 1.3, I develop a static version of the theoretical model, in which the set of buyer-supplier relationships is taken as given. I characterize how firm size, trade volumes, and household welfare depend on the existing production network, and show how to solve the market equilibrium of the model for any given network of relationships. In section 1.4, I then endogeneize the formation of linkages between firms in the economy by introducing a dynamic matching process between potential buyers and sellers. I examine in detail the steady-state of the model, and show how to construct theoretical counterparts to the empirical moments described in section 1.2. In section 1.5, I take the model to data and estimate its parameters via simulated method of moments. Section 1.6 uses these parameter estimates to quantitatively study the model’s predicted responses of trade volumes and welfare to counterfactual shocks. Finally, section 1.7 concludes.

1.2 Data and Empirical Regularities

1.2.1 Data

Before describing the theoretical model, I first present several stylized facts about production networks in the US economy. These empirical regularities are documented using two overlapping datasets. The first is obtained from Standard and Poor’s Capital IQ plat-
form, which collects fundamental data on a large set of companies worldwide, covering over 99% of global market capitalization. For a subset of these firms, both public and private but located mostly in the US, the database also records supplier and customer relationships based on a variety of sources, such as publicly available financial forms, company reports, and press announcements. From this database, I select all firms in the continental US for which relationship data is available and average revenue from 2003-2007 is positive. This gives me a dataset comprising 8,592 firms with $16.3 trillion in total revenue, comparable to the value of $30.0 trillion in total non-farm US business revenue as reported in the Census Bureau’s 2007 survey of business owners. The Capital IQ platform also provides the headquarters address of the majority of firms in this sample, which I geocode to obtain estimates of a firm’s location. Using these estimated locations, I then compute estimated distances between every supplier-customer pair in the dataset. Figure 1 shows the Capital IQ network for illustration, where each circle (node) represents a firm and each line (edge) represents a trading relationship.

Figure 1: The network of firm-to-firm trade in the continental United States, Capital IQ dataset

The second dataset is based on information from the Compustat platform, which is also operated by Standard and Poor. The Compustat database contains fundamental information for publicly-listed firms in the US, compiled solely from financial disclosure forms, and includes firms’ own reports of who their major customers are. In accordance with Financial
Accounting Standards No. 131, a major customer is defined as a firm that accounts for at least 10% of the reporting seller’s revenue. The Compustat relationship data was processed and studied by Atalay et al (2011), from whom the dataset was obtained. It contains 103,379 firm-year observations from 1979 to 2007.

Both the Capital IQ and Compustat datasets have their advantages and disadvantages. The Capital IQ platform offers greater coverage of firms with relationship data, as the database includes both public and private firms and records relationships based on sources other than financial disclosure forms. However, the main drawback of the dataset is that it is not possible to tell whether a particular relationship reported in a given year is still active at a later date. The Compustat data, on the other hand, is in panel form and therefore allows one to track the creation and destruction of trading relationships across time. The main weakness of the Compustat data is the 10% truncation level, which implies that a firm cannot have more than 10 customers reported in a given year, although there is still substantial variation in the number of recorded suppliers a firm has. For these reasons, I treat the capital IQ data as cross-sectional and primarily use it to estimate the steady-state of the model. I use the Compustat data to measure dynamic moments that are also used in the estimation.

1.2.2 Empirical regularities

In what follows, I document several empirical regularities characterizing the production network between firms in the data sample. In section 1.5, a subset of these moments will be used to estimate the theoretical model by simulated method of moments, and it is therefore useful at this point to formalize notation. Denoting the set of firms by \( S \), I first define \( N_{\text{bin}} \) evenly-spaced quantile bins \( \{B_b\}_{b \in \{1, \cdots, N_{\text{bin}}\}} \), where:

\[
B_b = \begin{cases} 
[q_{b-1}, q_b) , & b \in \{1, \cdots, N_{\text{bin}} - 1\} \\
[q_{b-1}, q_b] , & b = N_{\text{bin}}
\end{cases} 
\] (1.2.1)
with $q_b \equiv \frac{b}{N_{\text{bin}}}$, and define $\bar{q}_b \equiv \frac{q_{b-1} + q_b}{2}$ as the midpoint of bin $b$. I then compute for each variable of interest $X$ the quantile of this variable for firm $s$, $q^X(s)$, and define $b^X(s) \equiv \{b|q^X(s) \in B_b\}$ as the quantile bin of variable $X$ for firm $s$. Finally, I define $S^X_b \equiv \{s \in S|b^X(s) = b\}$ as the set of firms for which variable $X$ falls in quantile bin $b$.

1.2.2.1 Firm-level distributions

I begin by documenting the high degree of firm heterogeneity along several dimensions. To do so, I first compute for each variable $X$ the Kaplan-Meier estimate of the cumulative distribution function of the normalized variable:

$$\tilde{X}(s) \equiv \tilde{X}(s) - \min_{s' \in S} X(s') \max_{s' \in S} X(s') - \min_{s' \in S} X(s')$$

I then evaluate the inverse empirical CDF at the points $\{\bar{q}_b\}_{b \in \{1, \ldots, N_{\text{bin}}\}}$ via linear interpolation, obtaining estimates of the quantile levels $\{\bar{X}_b\}_{b \in \{1, \ldots, N_{\text{bin}}\}}$ for each quantile bin. Figure 2 shows these moments for the distributions of log revenue, log employment, in-degree (number of suppliers), and out-degree (number of customers) across all firms in the Capital IQ dataset.

To gain some sense about the parametric form of the distributions, I first compare the revenue and employment distributions to log-normal distributions with the same mean and variance by Monte Carlo simulation. As can be seen from the graphs, the distributions are relatively well-modeled by log-normal distributions, as is a common finding in the literature on firm size distributions.\(^{14}\) The lognormal approximation slightly overstates the fraction of firms with revenue below a given amount, however, and does the opposite for the firm employment distribution.

Next, to characterize the firm-level degree distributions, I compare these to two distributions that play central roles in network theory. It is well-known that in random graph

\(^{13}\)This normalization is employed so as to make computed moments of the univariate firm-level distributions scale-invariant, and therefore directly comparable to corresponding moments in the theoretical model.

\(^{14}\)See for example Cabral and Mata (2003) and Rossi-Hansberg and Wright (2007).
models, where links form between nodes with a constant probability, the degree distribution is approximately Poisson. On the other hand, in preferential attachment graph models, where nodes with a greater number of links form new links with a greater probability, the degree distribution exhibits a power law. I therefore compare the degree distributions to Poisson and Pareto distributions. From this, we see that the Poisson distribution is a poor approximation to the empirical degree distributions, strongly suggesting that relationships between firms are far from random, as might be expected. The Pareto distribution is a somewhat better approximation, although the approximation is also far from perfect.

1.2.2.2 Bivariate distributions

Next, I study how firm-level variables vary with firm size. Toward this end, I compute:

\[ R\hat{Q}_b^X \equiv \frac{1}{|S_b^R|} \sum_{s \in S_b^R} q^X(s) \]

(1.2.3)

as the average quantile of variable \( X \) for all firms with revenue falling in quantile bin \( b \). These moments are displayed in Figure 3 for employment, in-degree, and out-degree for all firms in the Capital IQ dataset. As expected, firm revenue and employment are highly correlated, but it is also clear from the graphs that larger firms tend to have larger numbers of customers and suppliers on average, with the rate of increase in degree also increasing in firm size. Firm-level variation in the numbers of suppliers and customers as well as the covariance of these measures with firm size will speak to the magnitude of the static aspect of relationship stickiness in the theoretical model.

\[^{15}\text{The Poisson parameter is chosen to match the mean of the empirical distribution, while the tail index of the Pareto distribution is computed using the Hill estimation procedure and the lower bound is set to match the mean of the empirical distribution.}\]
Figure 2: Firm-level distributions
Figure 3: Bivariate distributions
1.2.2.3 Matching distributions

Having characterized both the distributions and correlations of revenue, employment, in-degree, and out-degree across firms, I now examine what kinds of firms match up with what kinds of firms in the network. In particular, I study how matching between firms varies with firm size by first computing $\bar{q}^{S,X}(s)$ and $\bar{q}^{C,X}(s)$ as the quantile of the mean level of variable $X$ amongst suppliers and customers respectively of firm $s$ (conditional on firm $s$ having positive in- or out-degree). Next, as in section (1.2.2.2), I compute the averages of these firm-level measures within each revenue quantile bin:

$$R\bar{Q}^{S,X}_b \equiv \frac{1}{|S^R_b|} \sum_{s \in S^R_b} \bar{q}^{S,X}(s)$$

(1.2.4)

$$R\bar{Q}^{C,X}_b \equiv \frac{1}{|S^R_b|} \sum_{s \in S^R_b} \bar{q}^{C,X}(s)$$

(1.2.5)

Figure 5 shows these moments for supplier and customer revenue, employment, in-degree, and out-degree, for all firms in the Capital IQ dataset. From these graphs, we see that the assortativity of matching between firms is unambiguously positive, whether measured in terms of firm size or connectivity. On average, larger firms tend to buy and sell from firms that are also larger and better connected. This finding stands in contrast with the report of negative assortative matching in Bernard et al (2015) between exporting Norwegian firms and their trade partners, but agrees with the finding of Sugita et al (2014) that matching assortativity is positive between textile firms in Mexico selling to firms in the US. These patterns of firm matching will be important in identifying the shape of the distribution of relationship fixed cost shocks in the theoretical model.

1.2.2.4 Relationship geography

In addition to characterizing the assortativity of firm matching, the geocoded locations of firms in the Capital IQ dataset allow me to examine the geographic distribution of a firm’s
Figure 4: Matching distributions
suppliers and customers. To do so, I first compute $D_S(s)$ and $D_C(s)$ as the average distance between firm $s$ and its suppliers and customers respectively, normalized by the maximum trading distance in the Capital IQ dataset.\footnote{The maximum distance is 4,415 kilometers, which is approximately equal to the horizontal width of the continental United States. Again, this normalization is employed do as to make empirical moments directly comparable to the simulated moments in the theoretical model.} I then compute:

$$\bar{D}_b^S \equiv \frac{1}{|S^R_b|} \sum_{s \in S^R_b} D_S(s)$$

(1.2.6)

$$\bar{D}_b^C \equiv \frac{1}{|S^R_b|} \sum_{s \in S^R_b} D_C(s)$$

(1.2.7)

as the averages of the supplier and customer distance measures respectively for all firms with revenue falling in quantile bin $b$. These moments are shown in Figure 5. Perhaps somewhat surprisingly, larger firms tend to sell to customers that are located nearer by, while average supplier distance does not appear to vary much with firm size.\footnote{This finding is surprising in the context of trade models featuring fixed costs of exporting, for example, since these models predict that larger firms are more likely to sell to customers in more distant locations. On the other hand, it is perhaps less surprising in the context of models featuring agglomeration effects.}

1.2.2.5 Relationship dynamics

Finally, I make use of the panel nature of the Compustat data to study the dynamics of firm-to-firm relationships, which will be used to infer the magnitude of the dynamic aspect of relationship stickiness in the theoretical model. In particular, I examine how the rates at which firms retain existing suppliers and customers vary with firm size. To address this, I first
compute for every firm $s$ that exists in the dataset in both periods $t - 1$ and $t$ the variables $\rho^S_{t,ret}(s)$ and $\rho^C_{t,ret}(s)$, which denote the fraction of that firm’s suppliers and customers at date $t - 1$ respectively that are retained in period $t$. I then compute the following cross-sectional averages:

$$\bar{\rho}^S_{b,t} \equiv \frac{1}{|S^R_{b,t}|} \sum_{s \in S^R_{b,t}} \rho^S_{t,ret}(s)$$

$$\bar{\rho}^C_{b,t} \equiv \frac{1}{|S^R_{b,t}|} \sum_{s \in S^R_{b,t}} \rho^C_{t,ret}(s)$$

where $S^R_{b,t}$ denotes the set of firms in revenue quantile bin $b$ at date $t$ (relative to the cross-sectional revenue distribution at that date). Finally, I compute the time-series averages of these moments across time:

$$\bar{\rho}^S_b \equiv \frac{1}{T} \sum_{t=1}^{T} \bar{\rho}^S_{b,t}$$

$$\bar{\rho}^C_b \equiv \frac{1}{T} \sum_{t=1}^{T} \bar{\rho}^C_{b,t}$$

where $T = 29$ is the number of years in the Compustat dataset.

These moments are shown in Figure 6. From these graphs, we see that larger firms tend to retain a larger fraction of both existing suppliers and customers, and by implication, the average duration of relationships is longer for relationships involving larger firms. The mean duration of trading relationships across all firms in the Compustat dataset is 1.74 years, and the average rate at which suppliers and customers are terminated year-to-year are 38.4% and 30.1% respectively.

1.2.2.6 Summary of stylized facts

In sum, the production network between firms in the data sample can be characterized by the following stylized facts:
1. The firm size distribution is approximately log-normal, and the degree distributions deviate from both the Poisson and Pareto distributions predicted by statistical network formation models.

2. Larger firms tend to have more suppliers and customers.

3. The assortativity of matching between firms in terms of revenue, employment, and degree is unambiguously positive.

4. Larger firms tend to buy from and sell to suppliers and customers that are located nearer by.

5. Larger firms retain a larger fraction of suppliers and customers year-to-year.

Having documented these empirical regularities, I now turn to development of a simple model of trade between heterogeneous firms featuring sticky trading relationships, in which the firm-level degree distributions and matching between firms are endogenous outcomes. I return to the data in section 1.5 when I make use of the moments described above to estimate the model.

1.3 Static Model

I begin by describing a static version of the model in which the network of trading relationships between firms is fixed, and show how to characterize and solve for the static
equilibrium conditional on the network. Having done so, I then focus attention on endo-
geneizing dynamic formation of the production network in section 1.4.

1.3.1 Basic environment

The economy consists of a representative household and an exogenously-given unit con-
tinuum of heterogeneous firms that each produce a unique variety of a differentiated product. Firms are heterogeneous over states $\chi = (\phi, \delta)$, where $\phi$ and $\delta$ are what I refer to as the fundamental productivity of a firm’s production process and the fundamental quality of a firm’s product respectively, to be defined below. The exogenous cumulative distribution function over firm states is denoted by $F_\chi$, with density $f_\chi$ and support $S_\chi$ a bounded subset of $\mathbb{R}^2_+$. For brevity, I also refer to firms with state $\chi$ as $\chi$-firms. I begin by studying a simplified version of the model in which all firms belong to a single location. In section 1.3.3, I show how it is straightforward to incorporate multiple locations into the model, and in particular I embed geography which will allow the model to speak to the geographic distribution of firm-to-firm trade discussed in section 1.2.2.4.

1.3.1.1 Households

The representative household supplies $L$ units of labor inelastically and has CES prefer-
ces over all varieties of the differentiated product, given by:

$$U = \left[ \int_{S_\chi} [\delta x_H(\chi)]^{\frac{\sigma-1}{\sigma}} dF_\chi(\chi) \right]^{\frac{\sigma}{\sigma-1}}$$

(1.3.1)

where $\sigma$ is the elasticity of substitution across varieties, and $x_H(\chi)$ is the household’s con-
sumption of $\chi$-firm varieties.\footnote{Note that given the assumed unit mass of firms, integrals of all firm-level variables over the distribution $F_\chi$ are equal to both the average as well as the total value of that variable across firms.} Given the price $p_H(\chi)$ charged by $\chi$-firms to the household,
household demand is given by:

\[ x_H(\chi) = \Delta_H^{\delta - 1} [p_H(\chi)]^{-\sigma} \] (1.3.2)

Note that conditional on prices, households demand a greater amount of varieties for which fundamental quality \( \delta \) is higher. As opposed to buyer-seller specific components of quality, I assume here that \( \delta \) is a characteristic of the firm that is common across all customers. The household’s demand shifter can then be written as:

\[ \Delta_H \equiv U P_H^\sigma \] (1.3.3)

and the consumer price index is equal to:

\[ P_H = \left[ \int_{S_\chi} \left( \frac{p_H(\chi)}{\delta} \right)^{1-\sigma} dF_\chi(\chi) \right]^{\frac{1}{1-\sigma}} \] (1.3.4)

### 1.3.1.2 Firm production technology

Each firm produces its variety of the differentiated product using labor and the output of other firms. I assume, however, that firm-to-firm trade is characterized by relationship frictions, such that every \( \chi \)-firm is only able to purchase inputs from a given \( \chi' \)-firm with probability \( m(\chi, \chi') \). Given that there exists a continuum of firms of every state, this implies that \( m(\chi, \chi') \) is also equal to the fraction of \( \chi' \)-firms that supply a given \( \chi \)-firm, as well as the fraction of \( \chi \)-firms that purchase from a given \( \chi' \)-firm. I refer to \( m \) as the matching function of the economy, which completely specifies the extensive margin of firm-to-firm trading relationships in the economy. I take \( m \) as given in this section, and endogeneize formation of firm-to-firm trading relationships once dynamics are introduced into the model in section 1.4.

Given the matching function, the output of a \( \chi \)-firm is given by the following constant
returns to scale CES production function:

\[
X(\chi) = \left[ \phi l(\chi) \frac{\sigma - 1}{\sigma} + \int_{S_{\chi}} m(\chi, \chi') \left[ \alpha x(\chi, \chi') \right]^{\frac{\sigma - 1}{\sigma}} dF_{\chi}(\chi') \right]^{\frac{1}{\sigma - 1}} \tag{1.3.5}
\]

where \(l(\chi)\) is the quantity of labor demanded and \(x(\chi, \chi')\) is the quantity of each \(\chi'\)-variety used as inputs.\(^{19}\) The parameter \(\alpha\) is a measure of input-suitability, which I take as constant across firm pairs for now. Once I introduce geography into the model in section 1.3.3, \(\alpha\) will be a natural means of incorporating trade costs across firms in different locations.\(^{20}\) As is standard in the literature, I assume that the elasticity of substitution across inputs for intermediate demand is the same as that for final demand.

Taking the wage as the numeraire and given prices \(\{p(\chi, \chi')\}_{\chi' \in S_{\chi}}\) charged by other firms, the marginal cost of each \(\chi\)-firm is therefore given by:

\[
\eta(\chi) = \left[ \phi^{\sigma - 1} + \alpha^{\sigma - 1} \int_{S_{\chi}} m(\chi, \chi') \left[ p(\chi, \chi') \right]^{1-\sigma} dF_{\chi}(\chi) \right]^{\frac{1}{1-\sigma}} \tag{1.3.6}
\]

while the quantities of labor and intermediate inputs demanded are given respectively by:

\[
l(\chi) = X(\chi) \eta(\chi)^{\sigma} \phi^{\sigma - 1} \tag{1.3.7}
\]

\[
x(\chi, \chi') = X(\chi) \eta(\chi)^{\sigma} \alpha^{\sigma - 1} p(\chi, \chi')^{-\sigma} \tag{1.3.8}
\]

Note that conditional on prices, firms with greater fundamental productivity \(\phi\) have lower marginal costs.

\(^{19}\)In the appendix, I show how the model is isomorphic to one in which firms face convex input costs rather than a production function exhibiting “love of variety”.

\(^{20}\) In section A.3.1 of the appendix, I also discuss how \(\alpha\) can be used to capture differences in input suitability across industries and to match industry-level input-output shares, although I do not pursue this extension in the numerical analysis.
1.3.1.3 Relationship costs

It is evident from equation (1.3.6) that as long as prices are finite, access to additional suppliers always lowers the marginal cost of a firm, which follows from the CES property of the production function. Furthermore, since the production function exhibits constant returns to scale, access to additional customers always increases a firm’s variable profit. These forces generate incentives for firms to form as many upstream and downstream trading relationships as possible. To allow for the endogenous selection of relationships in the dynamic model studied in section 1.4, I therefore impose a cost of forming relationships by assuming that a link between any two firms requires $f$ units of labor. This can be interpreted as the cost of resources needed to manage ongoing relationships, such as expenditures on customer and supplier management systems as alluded to in the introduction to this paper or as more general man-hour costs.

In what follows, I further assume that this fixed relationship cost is paid fully by the selling firm. As we will see, this assumption implies that firm pricing decisions which are optimal in the static market equilibrium remain optimal in the dynamic market equilibrium, and that decisions about which relationships to keep active need to be analyzed only from the perspective of selling firms. In section A.3.2 of the appendix, I discuss how this assumption might be relaxed to allow for the buying firm to pay a positive share of the fixed relationship cost.

1.3.1.4 Market clearing

The labor market clearing condition can be written as:

$$\int_{S_{\chi}} l(\chi) dF_{\chi}(\chi) = L - L_f$$  \hspace{1cm} (1.3.9)
where $L_f$ is the total amount of labor used to pay the fixed relationship costs in the economy:

$$L_f = f \int_{S_x} \int_{S_x} m(\chi, \chi') dF_x(\chi) dF_x(\chi') \quad (1.3.10)$$

If we define the total mass of a $\chi$-firm’s suppliers and customers respectively as:

$$M_S(\chi) \equiv \int_{S_x} m(\chi, \chi') dF_x(\chi') \quad (1.3.11)$$
$$M_C(\chi) \equiv \int_{S_x} m(\chi', \chi) dF_x(\chi') \quad (1.3.12)$$

then total fixed labor costs can be written equivalently as $L_f = \int_{S_x} M_S(\chi) dF_x(\chi) = \int_{S_x} M_C(\chi) dF_x(\chi)$.

Since variable labor $l(\chi)$ must be non-negative, we see that labor market clearing can be satisfied for any arbitrary matching function $m : S_x \times S_x \rightarrow [0, 1]$, including the matching function $m(\chi, \chi') = 1$ for all $\chi, \chi' \in S_x$ specifying the complete network, if and only if the following assumption holds.

**Assumption 1.** The fixed relationship cost $f$ is less than the total labor supply $L$.

Finally, market clearing for the output of a $\chi$-firm requires:

$$X(\chi) = x_H(\chi) + \int_{S_x} m(\chi', \chi) x(\chi', \chi) dF_x(\chi') \quad (1.3.13)$$

### 1.3.1.5 Firm pricing and market structure

The market structure for all firm sales is assumed to be monopolistic competition. Given that the household and all purchasing firms face a continuum of sellers of every state and have demand functions (1.3.2) and (1.3.8) exhibiting a constant price elasticity, the profit-maximizing price charged by each firm is equal to the standard CES markup over marginal
As I discuss in section 1.4.1.3, the assumption that selling firms pay the entire share of the fixed relationship cost implies that the costly nature of relationships has no effect on the optimal price charged by firms. In section A.3.2 of the appendix, I also discuss how the model might be enriched by allowing for a form of bargaining between buyers and sellers, so that the markups charged by firms remain constant but are not completely determined by the elasticity of substitution \( \sigma \).

### 1.3.2 Static market equilibrium

#### 1.3.2.1 Firm network characteristics

As described above, the parameters \( \phi \) and \( \delta \) capture exogenous productivity and quality characteristics that are fundamental to the firm, in the sense that they are independent of the firm’s connection to other firms. Conditional on prices, firms with greater values of \( \phi \) and \( \delta \) enjoy lower marginal costs and greater final demand respectively. Firm-level outcomes in equilibrium, however, such as the overall size and profit of a firm, depend not only on a firm’s fundamental characteristics but also on the characteristics of other firms that it is connected to in the production network. For an arbitrary matching function, a given firm-level outcome may therefore in principle be a function of very complicated moments of the production network, which would render solution of the model intractable.

Fortunately, however, we can rely on the structure of the CES production function specified in (1.3.5) to derive sufficient statistics at the firm level that will allow us to compute all variables of interest with minimal computational difficulty. In contrast with firm fundamen-
tal characteristics $\phi$ and $\delta$, it is therefore useful to characterize the static market equilibrium of the model in terms of what I call a $\chi$-firm’s network productivity and quality, defined respectively by:

$$
\Phi (\chi) \equiv \eta (\chi)^{1-\sigma} \quad (1.3.17)
$$

$$
\Delta (\chi) \equiv \frac{1}{\Delta_H} X (\chi) \eta (\chi)^{\sigma} \quad (1.3.18)
$$

Note that $\Phi (\chi)$ is an inverse measure of a $\chi$-firm’s marginal cost, while $\Delta (\chi)$ is the demand shifter of a $\chi$-firm in the intermediate demand function (1.3.8) relative to the household’s demand shifter $\Delta_H$.

In what sense do $\Phi (\chi)$ and $\Delta (\chi)$ capture the characteristics of a $\chi$-firm in the production network as a whole, and how are these quantities determined? Combining the demand equations (1.3.2) and (1.3.8), the firm marginal cost equation (1.3.6), the goods market clearing condition (1.3.13), and the pricing conditions (1.3.14) and (1.3.15), we obtain the following system of equations:

$$
\Phi (\chi) = \phi^{\sigma-1} + \mu^{1-\sigma} \alpha^{\sigma-1} \int_{s_\chi} m (\chi, \chi') \Phi (\chi') dF_\chi (\chi') \quad (1.3.19)
$$

$$
\Delta (\chi) = \mu^{-\sigma} \delta^{\sigma-1} + \mu^{-\sigma} \alpha^{\sigma-1} \int_{s_\chi} m (\chi', \chi) \Delta (\chi') dF_\chi (\chi') \quad (1.3.20)
$$

Given the matching function, (1.3.19) and (1.3.20) specify a pair of decoupled linear functional equations in $\Phi (\cdot)$ and $\Delta (\cdot)$ respectively, and show how a firm’s network characteristics depend on both its fundamental characteristics as well as on the network characteristics of its suppliers and customers. Conditional on $\phi$ and $\delta$, firms that are connected to firms with larger network productivities and qualities also have higher network productivities and qualities themselves.\footnote{Note that $\Phi$ and $\Delta$ are conceptually similar to the measure of weighted average productivity in Melitz (2003), but in my model, these are measures at the firm-level on both the buyer and seller sides, and depend on the network structure specified by the matching function.}
The following proposition shows that as long as input-suitability \( \alpha \) is not too large relative to the markup \( \mu \), there exist unique solutions to the equations (1.3.19) and (1.3.20) for any matching function, and that starting from any arbitrary (but bounded) guesses for \( \Phi (\cdot) \) and \( \Delta (\cdot) \), iterating on (1.3.19) and (1.3.20) converges to these unique solutions with a known rate.\(^{22}\) The proof of Proposition 1, relegated to the appendix, entails showing that the functional equations (1.3.19) and (1.3.20) constitute contraction mappings with Lipschitz constants \( \left( \frac{\alpha}{\mu} \right)^{\sigma-1} \) and \( \frac{\alpha^{\sigma-1}}{\mu^\sigma} \) respectively.

**Proposition 1.** Under assumption 2, there exist unique network productivity and quality functions \( \Phi : S_\chi \to \mathbb{R}_+ \) and \( \Delta : S_\chi \to \mathbb{R}_+ \) for any matching function \( m : S_\chi \times S_\chi \to [0,1] \). Furthermore, starting from any arbitrary functions \( \tilde{\Phi} : S_\chi \to \mathbb{R}_+ \) and \( \tilde{\Delta} : S_\chi \to \mathbb{R}_+ \), iteration on equations (1.3.19) and (1.3.20) converges to \( \Phi \) and \( \Delta \) at rates \( \left( \frac{\alpha}{\mu} \right)^{\sigma-1} \) and \( \frac{\alpha^{\sigma-1}}{\mu^\sigma} \) respectively.

**Assumption 2.** Input suitability \( \alpha \) is less than the markup \( \mu \).

Under assumption 2, we can also rewrite equations (1.3.19) and (1.3.20) to express the network productivity and quality of a \( \chi \)-firm respectively as:

\[
\Phi (\chi) = \int_{S_\chi} \left[ \sum_{d=0}^\infty \left( \frac{\alpha}{\mu} \right)^d \frac{\sigma}{\mu^\sigma} m^{(d)} (\chi, \chi') \right] (\phi')^{\sigma-1} \mu^\sigma \mu dF_\chi (\chi') \tag{1.3.21}
\]

\[
\Delta (\chi) = \mu^{-\sigma} \int_{S_\chi} \left[ \sum_{d=0}^\infty \left( \frac{\alpha^{\sigma-1} \mu}{\mu^\sigma} \right)^d m^{(d)} (\chi', \chi) \right] (\phi')^{\sigma-1} \mu^\sigma \mu dF_\chi (\chi') \tag{1.3.22}
\]
where \( m^{(d)} \) is the \( d^{th} \)-degree matching function, defined recursively by:

\[
m^{(0)}(\chi, \chi') = \begin{cases} 
\frac{1}{J_\chi(\chi)}, & \text{if } \chi = \chi' \\
0, & \text{if } \chi \neq \chi'
\end{cases} \quad (1.3.23)
\]

\[
m^{(1)}(\chi, \chi') = m(\chi, \chi') \quad (1.3.24)
\]

\[
m^{(d)}(\chi, \chi') = \int_{S_\chi} m^{(d-1)}(\chi, \chi'') m(\chi'', \chi') dF_\chi(\chi'') \quad (1.3.25)
\]

Intuitively, one can think of \( m^{(d)}(\chi, \chi') \) for \( d \geq 1 \) as the probability that a \( \chi \)-firm buys indirectly from a \( \chi' \)-firm through a supply chain that is of length \( d \). With this interpretation, equations (1.3.21) and (1.3.22) show how the network productivity and quality of a firm depend on its connections to all other firms via supply chains of all lengths. Note that the rate at which the value of an indirect relationship decays with the length of the supply chain is decreasing in input suitability \( \alpha \) and increasing in the markup \( \mu \).

### 1.3.2.2 Firm size and inter-firm trade

Once the fundamental and network characteristics of a firm are known, the total revenue, variable profit, and variable employment of a \( \chi \)-firm are completely determined up to the scale factor \( \Delta_H \), and are given respectively by:

\[
R(\chi) = \mu \Delta_H \Delta(\chi) \Phi(\chi) \quad (1.3.26)
\]

\[
\pi(\chi) = (\mu - 1) \Delta_H \Delta(\chi) \Phi(\chi) \quad (1.3.27)
\]

\[
l(\chi) = \Delta_H \Delta(\chi) \phi^{\sigma-1} \quad (1.3.28)
\]

Intuitively, if a firm is twice as productive and produces a product that is twice as good from the perspective of the entire networked economy, its revenue and profit gross of fixed
relationship costs quadruples. Total firm profit and employment are given by:

\[
\Pi(\chi) = \pi(\chi) - fMC(\chi)
\]  

(1.3.29)

\[
L(\chi) = l(\chi) + fMC(\chi)
\]  

(1.3.30)

Total output of a \(\chi\)-firm is also completely determined by firm fundamental and network characteristics up to a scale factor:

\[
X(\chi) = \Delta_H \Delta (\chi) \Phi (\chi)^{\frac{\sigma}{\sigma-1}}
\]  

(1.3.31)

as are the value and quantity of output traded from \(\chi'\)- to \(\chi\)-firms:

\[
r(\chi, \chi') = \left( \frac{\alpha}{\mu} \right)^{\sigma-1} \Delta_H \Delta (\chi) \Phi (\chi')
\]  

(1.3.32)

\[
x(\chi, \chi') = \alpha^{\sigma-1} \mu \Delta_H \Delta (\chi) \Phi (\chi')^{\frac{\sigma}{\sigma-1}}
\]  

(1.3.33)

### 1.3.2.3 Household welfare and demand

To complete characterization of the static market equilibrium, it remains to determine the scale factor \(\Delta_H\). From the labor market clearing condition (1.3.9) and the firm variable employment equation (1.3.28), this is given by:

\[
\Delta_H = \frac{L - L_f}{\int_{S_\chi} \Delta (\chi) \phi^{\sigma-1}dF_\chi(\chi)}
\]  

(1.3.34)

Equations (1.3.3) and (1.3.4) then give the CPI and household welfare respectively as:

\[
P_H = \mu \left[ \int_{S_\chi} \Phi (\chi) \delta^{\sigma-1}dF_\chi (\chi) \right]^{\frac{1}{\frac{1}{\sigma}}}
\]  

(1.3.35)

\[
U = \mu^{-\sigma} (L - L_f) \left[ \int_{S_\chi} \Phi (\chi) \delta^{\sigma-1}dF_\chi (\chi) \right]^{\frac{1}{\sigma-1}}
\]  

(1.3.36)
while household demand is given by:

\[ x_H(\chi) = \mu^{-\sigma} \Delta_H \delta^{\sigma-1} \Phi(\chi) \frac{\sigma}{\sigma-1} \]  

(1.3.37)

Using equations (1.3.21) and (1.3.22) to substitute for \( \Phi(\chi) \) and \( \Delta(\chi) \) respectively, we see that the numerator and denominator of (1.3.36) are identical except for the terms \( \left( \frac{\alpha}{\mu} \right)^{d(\sigma-1)} \) and \( \left( \frac{\alpha^{\sigma-1}}{\mu^{\sigma-1}} \right)^d \), with the difference going to zero exponentially as \( d \) increases. An intuitive approximation to the value of household welfare is therefore:

\[
U \approx (L - L_f) \left[ \int_{S_S} \int_{S_S} \sum_{d=0}^{\infty} \left( \frac{\alpha}{\mu} \right)^{d(\sigma-1)} m^{(d)}(\chi, \chi') \right] \left( \delta \phi' \right)^{\sigma-1} \frac{dF_\chi(\chi) dF_\chi(\chi')}{\sigma-1} \]  

(1.3.38)

which is exact in the limit as \( \mu \to 1 \) (perfect competition). Equation (1.3.38) suggests that household welfare is greater when buyers of greater fundamental quality \( \delta \) are connected with sellers of greater fundamental productivity \( \phi' \), with the cost to welfare of additional relationships appearing in the term \( L - L_f \). When \( \mu > 1 \), the same general intuition applies, although household utility is only given exactly by the slightly more complicated expression (1.3.36).

### 1.3.2.4 Static market equilibrium definition

Given the matching function \( m \), the exogenous distribution over fundamental firm characteristics \( F_\chi \), and the model parameters \( \{L, \sigma, \alpha, f\} \), we can now define a static market equilibrium of the economy as follows. In section A.1.1 of the appendix, I describe the computational algorithm used to solve for the static market equilibrium.

**Definition 1.** A static market equilibrium of the economy is a pair of firm network characteristic functions \( \Phi : S_\chi \to \mathbb{R}_+ \) and \( \Delta : S_\chi \to \mathbb{R}_+ \) satisfying equations (1.3.19) and (1.3.20), a scalar household demand shifter \( \Delta_H \) satisfying (1.3.34), and allocation functions \( \{l(\cdot), X(\cdot), x(\cdot, \cdot), x_H(\cdot)\} \) given respectively as side equations by (1.3.28), (1.3.31), (1.3.33), and (1.3.37).
1.3.2.5 Static market equilibrium efficiency

To characterize the efficiency of a static market equilibrium, we can compare the resulting allocation with the allocation that would be chosen by a social planner seeking to maximize household welfare subject to the same exogenous matching function, production technology, and resource constraints. The following proposition (proved in section A.2.1 of the appendix) summarizes the solution to the planner’s problem.

**Proposition 2.** Given a matching function \( m : S_\chi \times S_\chi \rightarrow [0, 1] \), the network characteristic functions under the social planner’s allocation satisfy:

\[
\Phi^{SP}(\chi) = \phi^{\sigma-1} + \alpha^{\sigma-1} \int_{S_\chi} m(\chi, \chi') \Phi^{SP}(\chi') \, dF_\chi(\chi') \quad (1.3.39)
\]

\[
\Delta^{SP}(\chi) = \delta^{\sigma-1} + \alpha^{\sigma-1} \int_{S_\chi} m(\chi', \chi) \Delta^{SP}(\chi') \, dF_\chi(\chi') \quad (1.3.40)
\]

and the allocations of output and labor are given by equations (1.3.28), (1.3.31), (1.3.33), and (1.3.37) with \( \mu \) set equal to 1.

This result shows that any static market equilibrium allocation coincides with the corresponding planner’s allocation if and only if all firms in the decentralized equilibrium are perfectly competitive. With monopolistically-competitive firms, the static market equilibrium allocation is therefore inefficient relative to the planner’s allocation. This result can be interpreted as implying that the introduction of relationship frictions into the model through the exogenous matching function \( m \) imposes no additional inefficiency beyond the standard monopoly markup distortion. Once the matching function is endogeneized in section 1.4, this will no longer be true, as firm’s decisions about which relationships to keep active generate an additional dynamic source of inefficiency.
1.3.3 Embedding geography

Before introducing dynamics and endogenizing the formation of firm-to-firm trading relationships, it is useful to first describe how geography can be embedded into the model to study how relationship stickiness affects trade patterns across different locations, as this will be one area of focus of the numerical analysis and counterfactuals in sections 1.5 and 1.6. Toward this end, I assume that the unit mass of firms is evenly distributed along a unit circle, with each point on the circle indicating a different location. The distribution over firm states $F_\chi$ is assumed to be identical in all locations, and we can therefore focus on characterizing the market equilibrium in a single location.

To model trade costs, I assume that trade between two locations separated by a distance $D$ along the unit circle is subject to iceberg trade costs equal to $\tau(D) \geq 1$, with $\tau(0) = 1$, $\tau'(D) > 0$, and $\tau$ log-subadditive.\(^{23}\) Since all locations are identical, we can assume for notational simplicity and without loss of generality that firms in any one location can only sell to locations located clockwise of their own location. Given these assumptions, the static market equilibrium with geography embedded is simply characterized by analogous equations for the network productivity and quality functions:

\[
\Phi (\chi) = \phi^{\sigma-1} + \left( \frac{\alpha}{\mu} \right)^{\sigma-1} \int_0^1 \int_{S_\chi} \tau(D)^{1-\sigma} m [\chi, \chi'|\tau(D)] \Phi (\chi') dF_{\chi}(\chi') dD \tag{1.3.41}
\]

\[
\Delta (\chi) = \mu^{-\sigma} \delta^{\sigma-1} + \mu^{-\sigma} \alpha^{\sigma-1} \int_0^1 \int_{S_\chi} \tau(D)^{-\sigma} m [\chi', \chi|\tau(D)] \Delta (\chi') dF_{\chi}(\chi') dD \tag{1.3.42}
\]

where the matching function is now allowed to depend on distance through the trade cost $\tau(D)$.\(^{24}\)

As in the model without geography, there exist unique solutions to equations (1.3.41)

\(^{23}\)That is, $\log \tau(D_1) + \log \tau(D_2) \geq \log \tau(D_1 + D_2)$ for any $D_1, D_2 \in [0, 1]$, which is equivalent to the assumption that trade costs satisfy the triangle inequality.

\(^{24}\)Note that by writing equation (1.3.42) in this way, we are implicitly assuming that the representative household in each location purchases goods only from firms in its own location. Making the alternative assumption that households also purchase directly from firms in other locations subject to the same trade costs would simply require multiplying the first term on the right-hand side of (1.3.42) by the term $\tilde{\tau} \equiv \int_0^1 \tau(D)^{-\sigma} dD$, and would add nothing of qualitative substance to the model.
and \((1.3.42)\) for the functions \(\Phi\) and \(\Delta\). Given these, the value of trade between a \(\chi\)-buyer and a \(\chi'\)-seller separated by a distance \(D\) is then given by:

\[
R(\chi, \chi' | D) = \left(\frac{\alpha}{\mu}\right)^{\sigma^{-1}} \tau(D)^{1-\sigma} \Delta_H \Delta(\chi) \Phi(\chi')
\]

Notice that equation \((1.3.43)\) resembles a gravity equation for trade volumes at the firm level, where \(\Delta_H \Delta(\cdot)\) and \(\Phi(\cdot)\) capture the economic size of the importer and exporter respectively. The total value of trade between locations a distance \(D\) apart, however, also depends on the mass of firms that match between the two locations, and is given by:

\[
\bar{R}(D) = \left(\frac{\alpha}{\mu}\right)^{\sigma^{-1}} \tau(D)^{1-\sigma} \Delta_H \int_{S_{\chi}} \int_{S_{\chi'}} m[\chi, \chi' | \tau(D)] \Delta(\chi) \Phi(\chi') dF_{\chi}(\chi) dF_{\chi'}(\chi')
\]

Observe that if the matching function is held fixed, then as in models of trade with CES roundabout production such as Melitz (2003), the elasticity of trade volumes with respect to trade costs depends only on the elasticity of substitution \(\sigma\). However, once the matching function is endogenously determined as the result of firms’ decisions to trade or not to trade with other firms in various locations, the response of trade volumes to trade costs also depends on the extent of relationship frictions between firms.

### 1.4 Dynamics and Endogenous Network Formation

Analysis of the static version of the model shows that given any arbitrary matching function \(m\), numerical solution of all firm-level variables of interest is straightforward and tractable. It is the matching function \(m\), however, that captures all the relevant information determining the empirical moments in which we are interested, as described in section 1.2. Endogeneizing formation of the network is therefore crucial to my analysis, and I accomplish this by introducing a dynamic process of firm matching, as described below.
1.4.1 Dynamics of firm matching

Time is discrete and the representative household has preferences at date $t$ defined by:

$$V_t = \sum_{s=t}^{\infty} \beta^{s-t} U_s$$

(1.4.1)

where $U_t$ is given by the date $t$ equivalent of (1.3.1). Since the household’s value function is linear in per-period utility, household decisions every period are characterized exactly as in the static model, and the discount factor $\beta$ exists only to characterize how firms (which are owned by the household) discount the future. To economize on notation, I first describe the dynamic model without geography embedded, and reintroduce geography once I conduct the numerical analysis and study counterfactuals. The dynamics of firm matching are modeled based on three main assumptions.

1.4.1.1 Random fixed relationship costs

First, I assume that the fixed relationship cost $f_t$ is a random variable given by $f_t = f \xi_t$, where $\xi_t$ is independent and identically distributed across firm pairs and time with cumulative distribution function $F_\xi$ and unit mean. As in the static model, I assume that regardless of the realization of $\xi_t$, the selling firm always pays the full share of the fixed cost. The stochastic nature of the fixed relationship cost is the mechanism that generates the creation of new linkages between firms and the destruction of existing relationships, even in the steady-state of the model.

The assumption that $\xi_t$ exhibits no serial correlation is made primarily for tractability, and might jar with one’s intuition that relationship costs should be persistent. Nonetheless, the model generates non-trivial predictions about the persistence of relationships via assumptions about how often firms can reset relationships, described next.
1.4.1.2 Sticky relationships

I assume that firm-to-firm trading relationships are temporally sticky in the following sense. At each date, a firm receives the opportunity to sell to each firm that it did not sell to in the previous period with probability $1 - \nu$, and also receives the opportunity to terminate trading relationships with each of its existing customers with probability $1 - \nu$. I refer to this as the reset shock, and assume that it is independent across all firm pairs. Although the model can easily accommodate differences in the probabilities with which a firm can create and destroy relationships, I assume for parsimony that these probabilities are the same. Furthermore, I assume that regardless of whether a reset shock is received, selling firms can costlessly adjust prices every period, so that firm-to-firm relationships are sticky only along the extensive margin.

The assumption that firms can only sell to new customers with a finite probability may be interpreted as modeling the fact that potential trading partners take time to meet and learn about the suitability of their output for each other’s production processes or to negotiate new trading arrangements. Similarly, the assumption that firms cannot costlessly terminate existing relationships may be interpreted as either legal barriers to reneging on pre-negotiated contractual obligations, or more simply as the notion that winding down trading relationships also takes time. Allowing firms to costlessly adjust the intensive but not the extensive margin of trade may be interpreted as assuming that contracts between firms mandate only the provision of a good by the seller and not the price at which that good is sold.

Note that since the selling firm always pays the full share of the fixed relationship cost, the buying firm is always agreeable to any trading relationship, and therefore the decision to terminate or activate relationships only needs to be analyzed from the perspective of the selling firm. Under these assumptions, the matching function evolves according to the following law of motion:

$$m_t(\chi, \chi') = \nu m_{t-1}(\chi, \chi') + (1 - \nu) a_t(\chi, \chi')$$

(1.4.2)
where $a_t(\chi, \chi')$ is the probability that a $\chi'$-firm sells to a $\chi$-firm in period $t$ conditional on being given the opportunity to reset that relationship. I refer to $a_t$ as the acceptance function and characterize this in the following section. In any steady-state of the model, the matching function is simply equal to the acceptance function:

$$m(\chi, \chi') = a(\chi, \chi')$$ (1.4.3)

Note that $f$ and $\nu$ capture respectively the static and dynamic aspects of relationship stickiness alluded to in the introduction of this paper. There are several qualitatively different cases that one can consider. First, in the absence of the dynamic friction ($\nu = 0$), the matching function converges immediately to its steady-state value of $a(\cdot, \cdot)$, and the short- and long-run elasticities of trade volumes with respect to aggregate shocks are equal. Second, when the dynamic friction is extreme ($\nu = 1$), the production network exhibits no dynamics along the extensive margin. Third, in the presence of extreme static relationship costs ($f \to \infty$ and $\nu \in [0, 1)$), any steady-state of the model features an empty network in which no inter-firm trade occurs. Fourth, in the absence of the static friction ($f = 0$ and $\nu \in [0, 1)$), any steady-state of the model features a complete network in which all firms trade with one another. Trade therefore does not respond to external shocks along the extensive margin. Finally, with moderate static and dynamic frictions ($f \in (0, \infty)$ and $\nu \in (0, 1)$), the model exhibits both non-trivial steady-state production networks as well as non-trivial transition dynamics between steady-states.

1.4.1.3 Dynamic relationship activation decisions

The third and final assumption regards how and when firms decide to reset trading relationships conditional on having the opportunity to do so. First, note that the assumption that buying firms pay none of the fixed cost implies that it is never optimal for the selling firm to deviate from the standard CES markup pricing. Therefore, the variable profit earned
by a $\chi'$-firm from selling to a $\chi$-firm at date $t$ is the same as in the static market equilibrium, given by equations (1.3.20) and (1.3.27) as:

$$
\pi_t(\chi, \chi') = \mu^{-\sigma} (\mu - 1) \alpha^{-1} \Delta_{H,t} \Delta_t(\chi) \Phi_t(\chi')
$$

(1.4.4)

where $\Phi_t(\cdot)$, $\Delta_t(\cdot)$, and $\Delta_{H,t}$ are defined by the date $t$ equivalents of equations (1.3.19), (1.3.20), and (1.3.34).

Now, let $V^+_t(\chi, \chi'|\xi_t)$ denote the value to a $\chi'$-firm of selling to a $\chi$-firm in period $t$ conditional on the realization of the relationship cost shock $\xi_t$, and let $V^-_t(\chi, \chi')$ denote the value to the firm of not selling.\textsuperscript{25} These value functions are given by the following Bellman equations:

$$
V^+_t(\chi, \chi'|\xi_t) = \pi_t(\chi, \chi') - f\xi_t + \beta (1 - \nu) \mathbb{E}_t \left[ V^+_t(\chi, \chi'|\xi_{t+1}) \right] + \beta\nu \mathbb{E}_t \left[ V^+_t(\chi, \chi'|\xi_{t+1}) \right]
$$

(1.4.5)

$$
V^-_t(\chi, \chi') = \beta (1 - \nu) \mathbb{E}_t \left[ V^+_t(\chi, \chi'|\xi_{t+1}) \right] + \beta\nu V^-_{t+1}(\chi, \chi')
$$

(1.4.6)

where $V^O_t(\chi, \chi'|\xi_t)$ denotes the value to a $\chi'$-firm of having the option to reset its relationship with a $\chi$-firm customer given the relationship cost shock $\xi_t$:

$$
V^O_t(\chi, \chi'|\xi_t) = \max \left\{ V^+_t(\chi, \chi'|\xi_t), V^-_t(\chi, \chi') \right\}
$$

(1.4.7)

Note that the assumption of sticky relationships makes the activation and termination decisions facing a given firm forward-looking. If a firm chooses not to terminate a relationship given the chance to do so, it may find itself wishing to terminate the relationship in the future but lacking the opportunity to do so. Similarly, if a firm chooses not to sell to a potential customer despite having the chance to do so, it may be forced to wait several

\textsuperscript{25}Note that since the relationship cost shocks are i.i.d. over time, the value of not selling at date $t$ does not depend on $\xi_t$. Furthermore, since there is no aggregate uncertainty in the model, this implies that there is no uncertainty over the value of $V^-_t$ at any date for any pair of firms.
periods before being able to activate the relationship. Observe that if relationships are not sticky ($\nu = 0$) or firms are completely myopic ($\beta = 0$), then $V_t^+ (\chi, \chi' | \xi_t) \geq V_t^- (\chi, \chi')$ if and only if $\pi_t (\chi, \chi') \geq f \xi_t$. In these two special cases, relationships are activated as long as the static profits accruing to selling firms are enough to cover the fixed relationship costs. The probability that a $\chi'$-firm sells to a $\chi$-firm at date $t$ once it has the chance to do so is then given by:

$$\tilde{a}_t (\chi, \chi') = F \xi \left[ \frac{\pi_t (\chi, \chi')}{f} \right]$$

(1.4.8)

From (1.4.4), this implies that firms with larger network productivities and qualities are more likely to form downstream and upstream trading relationships respectively. The assumption of myopic agents in models of network formation is in fact somewhat standard in the network literature, and might seem to be a reasonable first approximation to firms’ decision making processes.\(^{26}\) We can, however, go further in characterizing the dynamic activation decisions of firms in this model.

It is instructive to first consider a steady-state of the model in which the functions $\pi_t$, $V_t^+$, $V_t^-$, and $V_t^O$ are all constant. From equations (1.4.5) and (1.4.6), it is straightforward to verify that:

$$\mathbb{E} \left[ V^O (\chi, \chi' | \xi) \right] = \begin{cases} \frac{\pi (\chi, \chi') - f}{1 - \beta}, & \forall (\chi, \chi') \in S_+ \\ 0, & \forall (\chi, \chi') \notin S_+ \end{cases}$$

(1.4.9)

where $S_+ \equiv \{(\chi, \chi') \subset S^2_\chi | \pi (\chi, \chi') - f \geq 0 \}$. This tells us that the option value of a relationship is positive if and only if the profit from that relationship exceeds the relationship cost on average. Substituting (1.4.9) into (1.4.5) and (1.4.6), we then find:

$$V^+ (\chi, \chi' | \xi) - V^- (\chi, \chi') = \frac{\pi (\chi, \chi') - \beta \nu f}{1 - \beta \nu} - f \xi$$

(1.4.10)

\(^{26}\)See for example Bala and Goyal (2000) and Jackson (2005).
and therefore the probability that a \( \chi' \)-firm sells to a \( \chi \)-firm conditional on having the chance to do so is given by:

\[
a(\chi, \chi') = F_\xi \left[ \frac{\pi(\chi, \chi') - \beta \nu f}{(1 - \beta \nu) f} \right]
\]  

(1.4.11)

Comparing this expression with equation (1.4.8), we again see that firms with greater network productivities and qualities are more likely to form downstream and upstream trading relationships respectively, but once the option values of relationships are taken into account, this effect becomes more pronounced. In particular, for firm pairs such that \( \pi(\chi, \chi') > f \), there is a positive probability, equal to \( a(\chi, \chi') - \tilde{a}(\chi, \chi') \), that temporarily-unprofitable relationships will still be activated because the relationship is profitable enough on average. Similarly, for firm pairs such that \( \pi(\chi, \chi') < f \), there is a positive probability, given by \( \tilde{a}(\chi, \chi') - a(\chi, \chi') \), that temporarily-profitable relationships will not be activated because the relationship is not profitable enough on average. Furthermore, note that (1.4.11) implies that firm pairs with \( \pi(\chi, \chi') < \beta \nu f \) will never form trading relationships in steady-state.

How do we characterize the activation and termination decisions of firms outside the steady-state? Iterating forward on equations (1.4.5), (1.4.6), and (1.4.7), we can write the difference in the values of selling and not selling as:

\[
V^+_t (\chi, \chi' | \xi_t) - V^-_t (\chi, \chi') = \pi_t (\chi, \chi') - f \xi_t
\]

\[
+ \sum_{s=1}^{\infty} (\beta \nu)^s \left[ \pi_{t+s} (\chi, \chi') - f \right]
\]

(1.4.12)

which can be interpreted as the expected future stream of profits net of fixed costs until the relationship can be reset. The acceptance function at date \( t \) is therefore given by:

\[
a_t (\chi, \chi') = F_\xi \left[ \frac{\pi_t (\chi, \chi')}{f} + \sum_{s=1}^{\infty} (\beta \nu)^s \left[ \frac{\pi_{t+s} (\chi, \chi')}{f} - 1 \right] \right]
\]

(1.4.13)

From this, we see that solving for the acceptance function at date \( t \) outside of the steady-state requires solving for the profit functions \( \pi_{t+s} \) for all \( s \geq 1 \). In section A.1.1.1 of the
appendix, I describe the computational algorithm that I employ to accomplish this, which essentially involves iterating on the path of profit functions \( \{ \pi_{t+s} \}_{s=1}^{T} \) for some value of \( T \) large enough such that \( m_{t+T} \) is close to the eventual steady-state matching function. This allows me to solve exactly for the model’s transition dynamics between steady-states under rational firm expectations. In section 1.6.4, I show why this is important, as the assumption of myopic firms leads to model predictions that are both qualitatively and quantitatively different from the rational expectations case.

Note that even though \( \xi_t \) is assumed to have unit mean, firms in the dynamic market equilibrium select relationships based on the realized values of the relationship cost shocks. Therefore, the average cost of active relationships is no longer equal to \( f \) as it was in the static model, and the total mass of labor used to pay relationship fixed costs is now given by:

\[
L_{f,t} = f \int_{S_{\chi}} \int_{S_{\chi}} \left[ \nu m_{t-1} (\chi, \chi') + (1 - \nu) \bar{\xi}_t (\chi, \chi') \right] dF_{\chi} (\chi) dF_{\chi} (\chi') \tag{1.4.14}
\]

The first term in the integral reflects the cost of relationships that cannot be reset (and hence for which there is no selection on \( \xi_t \)), while the second term reflects the cost of relationships that are voluntarily selected by firms. The term \( \bar{\xi}_t (\chi, \chi') \) denotes the average value of the idiosyncratic component of the cost shock amongst \( \chi - \chi' \) firm pairs that receive the reset shock:

\[
\bar{\xi}_t (\chi, \chi') = \int_{0}^{\xi_{\max,t} (\chi, \chi')} \xi dF_{\xi} (\xi) \tag{1.4.15}
\]

and \( \xi_{\max,t} (\chi, \chi') \) is the maximum value of the cost shock for which \( \chi - \chi' \) relationships are voluntarily selected:

\[
\xi_{\max,t} (\chi, \chi') = \max \left\{ \frac{\pi_t (\chi, \chi')}{f} + \sum_{s=1}^{\infty} (\beta \nu)^s \left[ \frac{\pi_{t+s} (\chi, \chi')}{f} - 1 \right], 0 \right\} \tag{1.4.16}
\]
1.4.2 Dynamic market equilibrium

1.4.2.1 Dynamic market equilibrium definition

Having characterized the dynamics of firm matching, we can now define a dynamic market equilibrium as follows.

**Definition 2.** Given an initial matching function \( m_0 : S_x \times S_x \to [0, 1] \), a dynamic market equilibrium of the model is a list of sequences of matching functions \( \{m_t\}_{t=1}^\infty \), acceptance functions \( \{a_t\}_{t=0}^\infty \), profit functions \( \{\pi_t\}_{t=0}^\infty \), and network characteristic functions \( \{\Phi_t, \Delta_t\}_{t=0}^\infty \), as well as a list of scalars \( \{\Delta_{Ht}\}_{t=0}^\infty \), all of which satisfy equations (1.3.19), (1.3.20), (1.3.34), (1.4.2), (1.4.4), and (1.4.13). Given the matching function \( m_t \), the allocation at date \( t \) in a dynamic equilibrium is as defined in the static model.

Similarly, we can define a steady-state of the dynamic model as a dynamic market equilibrium in which all variables in Definition 2 are constant.

**Definition 3.** A steady-state equilibrium of the dynamic model is a matching function \( m \), an acceptance function \( a \), a profit function \( \pi \), network characteristic functions \( \{\Phi, \Delta\} \), as well as a scalar \( \Delta_H \), all of which satisfy equations (1.3.19), (1.3.20), (1.3.34), (1.4.3), (1.4.4), and (1.4.11). Given the steady-state matching function \( m \), the allocation in a steady-state equilibrium is as defined in the static model.

In section A.1.1.1 of the appendix, I describe the computational algorithms used to solve for both the model’s transition dynamics as well as its steady-state.

1.4.2.2 Dynamic market equilibrium efficiency

To what extent are the dynamic relationship selection decisions made by firms socially optimal? Recall that the results of Proposition 2 showed how the static market equilibrium is inefficient relative to the social planner’s allocation because of the monopoly markups charged by firms. Similarly, we can characterize the dynamic efficiency of the model by
comparing the market equilibrium allocation with the dynamic allocation that would be
chosen by a social planner subject to the same static and dynamic frictions faced by firms.
In particular, we can compare the cutoff value for the relationship cost shock chosen by
firms, given by equation (1.4.16), to the cutoff value that would be chosen by the planner.
In section A.2.2 of the appendix, I show that the planner’s solution is characterized by the
following proposition.

**Proposition 3.** The cutoff value for the cost shock at date \( t \) chosen by the social planner is
given by:

\[
\xi_{\text{max},t}^{\text{SP}}(\chi, \chi') = \max \left\{ \frac{\pi_{t}^{\text{SP}}(\chi, \chi')}{f} + \sum_{s=1}^{\infty} (\beta \nu)^{s} \left( \frac{C_{t+s}}{C_{t}} \right) \left[ \frac{\pi_{t+s}^{\text{SP}}(\chi, \chi')}{f} - 1 \right], 0 \right\} \tag{1.4.17}
\]

where \( \pi_{t}^{\text{SP}} \) is the planner’s analog of the profit function:

\[
\pi_{t}^{\text{SP}}(\chi, \chi') \equiv \frac{\alpha \sigma^{-1}}{\sigma - 1} \Delta_{H,t}^{\text{SP}} \Delta_{t}^{\text{SP}}(\chi^*) \Phi_{t}^{\text{SP}}(\chi'^*) \tag{1.4.18}
\]

and \( C_{t} \) is a measure of the total connectivity between firms in the economy:

\[
C_{t} \equiv \left[ \int_{S_{x}} \int_{S_{x}} \left[ \sum_{d=0}^{\infty} \frac{\alpha d^{(\sigma-1)} m_{t}^{\text{SP},(d)}(\chi, \chi')}{\delta \phi^{(\sigma)}} \right] \frac{1}{\sigma - 1} dF_{\chi}(\chi) dF_{\chi}(\chi') \right]^{\frac{1}{\sigma - 1}} \tag{1.4.19}
\]

Comparing equations (1.4.16) and (1.4.17), we now see that the criterion by which firms
select relationships in the market equilibrium differs from the socially-optimal criterion in
two ways. First, because of the monopoly markup distortion discussed in section 1.3.2.5,
the static social value of a given relationship relative to its cost (measured by \( \frac{\pi_{t}^{\text{SP}}}{f} \)) differs
from the ratio of profits to fixed costs (\( \frac{\pi_{t}}{f} \)) that are faced by selling firms in the market
equilibrium. Note that holding fixed the network productivity of the selling firm and the
network quality of the buying firm, the function \( \pi_{t}^{\text{SP}} \) differs from the profit function \( \pi_{t} \) only
by a constant fraction \( \mu^{-\sigma} \).

Second, the planner internalizes the effect of each relationship on all other firms in the
production network (often referred to as network externalities) whereas firms in the market equilibrium do not. To better understand this effect, it is useful to consider the social value of a given relationship at date $t$, which can be characterized by the static marginal change in household utility resulting from a marginal increase in the mass of active relationships between firms of given states. In the proof of Proposition 3, I show that this is given by:

$$
\frac{dU_t}{d\bar{m}_t(\chi, \chi')} = C_t \left[ \pi_t^{SP} (\chi, \chi') - f \right]
$$

(1.4.20)

where $\bar{m}_t(\chi, \chi') \equiv m_t(\chi, \chi') f_\chi(\chi) f_\chi(\chi')$ denotes the total mass of connections between $\chi$-firm buyers and $\chi'$-firm sellers. From equation (1.4.20), we see that the social value of each relationship is equal to the difference $\pi_t^{SP} - f$ amplified by the aggregate connectivity measure $C_t$. Intuitively, when firms are more connected to each other ($C_t$ is larger), the activation or termination of a single relationship has larger aggregate effects. Since the amplification term $C_t$ potentially varies across time, the planner values changes in the extensive margin of firm relationships accordingly. This effect appears through the term $\frac{C_{t+s}}{C_t}$ in equation (1.4.17) but is absent in firms’ decision making processes about which relationships to activate and terminate at each date.

### 1.4.3 Properties of the steady-state

#### 1.4.3.1 Firm-level distributions

In our analysis of the static market equilibrium, we saw how the revenue and employment of a firm are completely determined (up to a scale factor) by the fundamental and network characteristics of that firm. I now show that variation in firm in-degrees (measured by $M_S$) and out-degrees (measured by $M_C$) is also completely determined by variation in network characteristics. To see this, first observe from equations (1.4.3), (1.4.4), and (1.4.11) that variations across firm-pairs in the profit, activation, and matching functions depend only on
variations in the product $\Delta (\chi) \Phi (\chi)$.\textsuperscript{27} In particular, the matching function in steady-state can be written as:

$$m (\chi, \chi') = \tilde{m} \left[ \Delta_H \Delta (\chi) \Phi (\chi') \right]$$  \hspace{1cm} (1.4.21)

where $\tilde{m} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is an increasing scalar function defined by:

$$\tilde{m} (x) = F_\xi \left[ \frac{x - \beta \nu \bar{f}}{(1 - \beta \nu) \bar{f}} \right]$$  \hspace{1cm} (1.4.22)

with $\bar{f} \equiv \frac{\mu^\sigma}{(\mu - 1)^{1 - \sigma}}$. As a result, the network quality and productivity of a $\chi$-firm are sufficient statistics for its in- and out-degrees respectively:

$$M_S (\chi) = \tilde{M}_S [\Delta (\chi)] \equiv \int_{\chi} \tilde{m} \left[ \Delta (\chi) \Phi (\chi') \Delta_H \right] dF_\chi (\chi')$$  \hspace{1cm} (1.4.23)

$$M_C (\chi) = \tilde{M}_C [\Phi (\chi)] \equiv \int_{\chi} \tilde{m} \left[ \Delta (\chi') \Phi (\chi) \Delta_H \right] dF_\chi (\chi')$$  \hspace{1cm} (1.4.24)

Since firm revenue is proportional to the product of firm network productivity and quality, this implies that firms with larger masses of suppliers and customers also tend to have larger revenue.

Figure 7 shows an example of the network productivity and quality functions in a steady-state of the model obtained through numerical solution, as well as the supplier and customer functions $M_S (\cdot)$ and $M_C (\cdot)$ defined by equations (1.3.11) and (1.3.12). Note that even though fundamental firm productivities and qualities $\phi$ and $\delta$ may be uncorrelated, a firm’s network productivity $\Phi (\chi)$ is still increasing in $\delta$ because a firm with higher fundamental quality offers greater profit opportunities to potential suppliers, and therefore is more likely to form upstream trading relationships. Similarly, a firm’s network quality $\Delta (\chi)$ is increasing

\textsuperscript{27}Given that each firm has a continuum of both suppliers and customers of each state, these functions do not depend on idiosyncratic realizations of the fixed cost shock $\xi_t$.  

43
in both its fundamental productivity and and quality.

1.4.3.2 Matching assortativity

What determines the assortativity of matching between firms in the model? The average supplier and customer revenue of a $\chi$-firm are given respectively by:

$$
\bar{R}_S(\chi) = \frac{\int s_x m(\chi, \chi') R(\chi') dF_X(\chi')}{M_S(\chi)} \quad (1.4.25)
$$

$$
\bar{R}_C(\chi) = \frac{\int s_x m(\chi', \chi) R(\chi') dF_X(\chi')}{M_C(\chi)} \quad (1.4.26)
$$

Given the analysis in the previous section, the matching between a $\chi$-firm and its suppliers and customers depends only on $\Delta(\chi)$ and $\Phi(\chi)$ respectively, and therefore we can alternatively consider the average supplier and customer revenue of firms with network quality $\Delta$.
and productivity $\Phi$ respectively (which I henceforth refer to as $\Delta$- and $\Phi$-firms), given by:

$$
\tilde{R}_S (\Delta) = \int_{S} \bar{m} \left[ \Delta \Phi (\chi') \Delta H \right] \frac{R (\chi') dF_{\chi} (\chi')}{M_s (\Delta)}
$$

(1.4.27)

$$
\tilde{R}_C (\Phi) = \int_{S} \bar{m} \left[ \Delta (\chi') \Phi \Delta H \right] \frac{R (\chi') dF_{\chi} (\chi')}{M_C (\Phi)}
$$

(1.4.28)

Since firms with higher network productivity and quality also tend to have higher revenue, the assortativity of firm matching (in terms of revenue) can be characterized by the gradients of the functions $\tilde{R}_S$ and $\tilde{R}_C$. Differentiating equation (1.4.27), for example, we obtain:

$$
\tilde{R}_S' (\Delta) = \frac{\Delta}{M_s (\Delta)} \int_{S} \left[ R (\chi') \right] \varepsilon_{\bar{m}} \left[ \Delta \Phi (\chi') \Delta H \right] \bar{m} \left[ \Delta \Phi (\chi') \Delta H \right] dF_{\chi} (\chi')
$$

(1.4.29)

where $\varepsilon_{\bar{m}}$ is the elasticity of the scalar matching function $\bar{m}$. From equation (1.4.29) and the equivalent derivative of equation (1.4.28), we make the following observation: if the elasticity $\varepsilon_{\bar{m}}$ is constant, then $\tilde{R}_S (\cdot)$ and $\tilde{R}_C (\cdot)$ are constant functions, and in this sense the assortativity of matching between firms is neutral, with average customer and supplier revenue independent of firm size. This suggests that the elasticity $\varepsilon_{\bar{m}}$ plays a crucial role in shaping the assortativity of matching between firms in general.

We can characterize this even further by considering the average revenue of $\Delta'$-firms that supply a $\Delta$-firm, the derivative of which with respect to $\Delta$ is:

$$
\tilde{R}_S' (\Delta | \Delta') = \frac{\Delta}{M_s (\Delta)} \int_{S} \left[ R (\chi') \right] \varepsilon_{\bar{m}} \left[ \Delta \Phi (\chi') \Delta H \right] \bar{m} \left[ \Delta \Phi (\chi') \Delta H \right] dF_{\chi} (\chi')
$$

(1.4.30)

Since $\bar{m}$ is an increasing function, then from this equation we can make an even stronger observation about the role of $\varepsilon_{\bar{m}}$: the assortativity of matching between $\Delta$-buyers and $\Delta'$-sellers is positive if $\varepsilon_{\bar{m}}$ is increasing, and is negative if $\varepsilon_{\bar{m}}$ is decreasing. The same is also true regarding the assortativity of matching between $\Phi$-buyers and $\Phi'$-sellers.

This analysis then begs the question: what determines the elasticity of the matching
function? From equation (1.4.22), the matching function elasticity is equal to:

\[ \varepsilon_{\tilde{m}}(x) = \frac{x}{(1 - \beta \nu) \tilde{f}} \left[ \frac{F'_{\xi} \left[ \frac{x - \beta \nu \tilde{f}}{(1 - \beta \nu) \tilde{f}} \right]}{F_{\xi} \left[ \frac{x - \beta \nu \tilde{f}}{(1 - \beta \nu) \tilde{f}} \right]} \right] \] (1.4.31)

In the special case when \( \nu = 0 \), so that the model is completely static, the elasticity of the matching function is completely determined by the elasticity of the distribution function \( F_{\xi} \) of the relationship cost shock. Consequently, this implies that the assumed parametric form for \( F_{\xi} \) will be crucial for determining the model’s predictions regarding the assortativity of matching between firms, an issue that we will return to when we discuss numerical estimation of the model in section 1.5.

### 1.4.3.3 Geographic distribution of trade partners

Reintroducing geography into the model simply requires rewriting the matching function as:

\[ m \left[ \chi, \chi' | \tau (D) \right] = \tilde{m} \left[ \frac{\Delta (\chi) \Phi (\chi') \Delta_H}{\tau (D)^{\sigma - 1}} \right] \] (1.4.32)

and using equations (1.3.41) and (1.3.42) to specify the network characteristic functions. We can then easily compute the average supplier and customer distance of a \( \chi \)-firm as follows:

\[ D_S (\chi) = \frac{\int_0^1 \int_{S_{\chi}} Dm \left[ \chi, \chi' | \tau (D) \right] dF_{\chi} (\chi') dD}{M_S (\chi)} \] (1.4.33)

\[ D_C (\chi) = \frac{\int_0^1 \int_{S_{\chi}} Dm \left[ \chi, \chi' | \tau (D) \right] dF_{\chi} (\chi') dD}{M_C (\chi)} \] (1.4.34)

Exactly the same analysis as in section 1.4.3.2 can be used to show that the matching function elasticity plays a key role in determining whether larger firms tend to have suppliers and customers that are located further or nearer by. When the elasticity is increasing, larger firms tend to have closer trade partners than smaller firms.
1.4.3.4 Relationship dynamics

Even in the steady-state of the model, there is churning of firm relationships due to the stochastic nature of the fixed relationship cost. First, note that the unconditional probabilities that a $\chi$-firm will retain any one of its suppliers or customers are given by:

$$\rho_{ret}^S(\chi) = \nu + (1 - \nu) \int_{S_x} a(\chi, \chi') dF_x(\chi')$$

(1.4.35)

$$\rho_{ret}^C(\chi) = \nu + (1 - \nu) \int_{S_x} a(\chi', \chi) dF_x(\chi')$$

(1.4.36)

Since these probabilities are constant in steady-state, the unconditional duration of relationships between a $\chi$-firm and its suppliers and customers follows a geometric distribution, with means $\frac{1}{1 - \rho_{ret}^S(\chi)}$ and $\frac{1}{1 - \rho_{ret}^C(\chi)}$ respectively. Furthermore, since the matching function is equal to the acceptance function in the steady-state of the model, then equations (1.4.35) and (1.4.36) deliver sharp predictions about the relation between the retention probabilities and the masses of a firm’s suppliers and customers:

$$\rho_{ret}^S(\chi) = \nu + (1 - \nu) M_S(\chi)$$

(1.4.37)

$$\rho_{ret}^C(\chi) = \nu + (1 - \nu) M_C(\chi)$$

(1.4.38)

Firms with more suppliers and customers are therefore more likely to retain existing trading relationships.

Note that firms in the model are also more likely to trade with existing partners than new ones because of the sticky nature of relationships. If a $\chi - \chi'$ relationship was active in the previous period, the probability that it will be maintained in the current period is equal to $\nu + (1 - \nu) a(\chi, \chi')$, whereas the probability that it will be newly-formed is equal to $(1 - \nu) a(\chi, \chi')$. The fractions of suppliers and customers that are new for a $\chi$-firm every
period are therefore given respectively by:

\[
\rho_{S}^{\text{new}}(\chi) = \frac{\int_{S_{\chi}} (1 - \nu) a(\chi, \chi') [1 - m(\chi, \chi')] dF_{\chi}(\chi')} {M_{S}(\chi)}
\]  
(1.4.39)

\[
\rho_{C}^{\text{new}}(\chi) = \frac{\int_{S_{\chi}} (1 - \nu) a(\chi', \chi) [1 - m(\chi', \chi)] dF_{\chi}(\chi')} {M_{C}(\chi)}
\]  
(1.4.40)

Finally, it is useful to point out that the parameter \(\nu\) controls the rate of convergence between steady-states. As an illustrative example, consider an economy that is in steady-state at \(t = 0\) with both the relationship fixed cost \(f\) and the reset friction \(\nu\) being finite, and denote the matching function in this economy by \(m_{ss}\). Suppose then that the fixed relationship cost \(f\) becomes either infinite or zero, and denote the new steady-state matching function by \(m'_{ss}\) (identically zero or one respectively). From equations (1.4.2) and (1.4.11), the matching function evolves according to:

\[
\hat{m}_{t}(\chi, \chi') = \nu \hat{m}_{0}(\chi, \chi')
\]  
(1.4.41)

where \(\hat{m}_{t}(\chi, \chi') \equiv m_{t}(\chi, \chi') - m'_{ss}(\chi, \chi')\) is the deviation of the matching function from the new steady-state. When relationships are stickier (larger \(\nu\)), convergence between steady-states is slower.

1.5 Numerical Analysis

Having characterized the theoretical counterparts of the empirical moments described in section 1.2.2, I now take the model to data by estimating the steady-state of the model via simulated method of moments. I begin by specifying the remaining parametric assumptions in the model.
1.5.1 Parametric assumptions

First, given that the firm size distribution appears to be approximately log-normal (Figure 1.2.2.1), I assume that the log of fundamental firm productivities and qualities, $\phi$ and $\delta$, are jointly Gaussian with zero mean and covariance matrix given by:

$$
\Sigma = \begin{bmatrix}
    \phi^2 & \rho \phi \delta \\
    \rho \phi \delta & \delta^2
\end{bmatrix}
$$  \hspace{1cm} (1.5.1)

Note that in the empty network with $m(\chi, \chi') = 0$ for all $\chi, \chi' \in S_{\chi}$, this assumption would imply that firm revenue and employment are exactly log-normally distributed.

Parameterization of the distribution function $F_\xi$ of the relationship cost shock requires slightly more careful consideration. As discussed in section 1.4.3.2, the elasticity of $F_\xi$ plays a key role in determining qualitative properties of the model, and in particular the gradient of the elasticity of $F_\xi$ is directly related to the assortativity of matching between firms. As it turns out, almost all of the standard continuous distributions with support on $[0, \infty)$ feature a monotonically decreasing elasticity. One notable exception is the Gompertz or log-Weibull distribution, which is used extensively in survival analysis and has the following distribution function:

$$
F_\xi (x) = 1 - e^{-b_\xi (s_\xi x^s - 1)}
$$  \hspace{1cm} (1.5.2)

where $b_\xi$ is a scale parameter and $s_\xi$ characterizes the shape of the distribution. From a mathematical point of view, assuming that the relationship cost shock follows a Gompertz distribution is desirable because the sign of the elasticity gradient of the distribution is variable when $s_\xi \in (0, 1)$, which therefore allows for flexibility in the model’s predictions regarding the assortativity of firm matching.

From an economic standpoint, a Gompertz-distributed relationship cost shock can be interpreted as follows. Suppose that upon meeting, a pair of firms takes a random amount of

\footnotetext{28}{These include (at least) the Fréchet, Weibull, log-normal, Gamma, generalized Pareto, and log-logistic distribution.}
time (within the period) to negotiate the potential arrangements of the trading relationship, and that the fixed cost of the relationship is proportional to the amount of time that it takes for negotiations to be completed. Suppose also that the probability with which negotiations continue to drag on conditional on no agreement having been reached at a given point in time declines with time. If this process is characterized by the negotiation time having an exponential hazard rate, then the fixed cost of the relationship has a Gompertz distribution. Based on these considerations, I parameterize the relationship cost shock according to (1.5.2). With the mean of $\xi_t$ fixed at 1, this pins down the scale parameter $b_\xi$ given a choice of the shape parameter $s_\xi$.

Finally, trade costs are parameterized according to:

$$\tau(D) = (1 + \kappa D)^\epsilon$$

(1.5.3)

where $\kappa$ measures the overall level of trade costs and $\epsilon$ measures the elasticity of trade costs with respect to distance.\(^29\) Since the maximum possible trading distance in the model is normalized to 1, $\kappa$ can also be interpreted as the cost of trading with the most distant firms relative to trading with firms that are right next door. Note that trade costs are non-existent when either $\kappa = 0$ or $\epsilon = 0$.

1.5.2 Parameter estimation

The above parameterization of the model gives us a total of 12 parameters: the elasticity of substitution $\sigma$; input suitability $\alpha$; mean $f$ and shape $s_\xi$ of the relationship fixed cost; reset friction $\nu$; parameters of the $\chi$ distribution, $v_\phi$, $v_\delta$, and $\rho$; parameters of the trade cost function $k$ and $\epsilon$; labor supply $L$; and the household discount factor $\beta$.

Since the Compustat data is of annual frequency, I set $\beta = .96$. Also, note that the total labor supply $L$ only enters the set of equilibrium conditions through equation (1.3.34). If we

\(^{29}\)Note that with this parameterization, $\tau$ is log-subadditive for any $\kappa, \epsilon \geq 0$, and therefore trade costs satisfy the triangle inequality.
write the magnitude of the fixed relationship cost $f$ as a fraction $\hat{f}$ of the total labor supply, then from equations (1.4.4) and (1.4.13), we see that the activation function $a$ is independent of $L$. Equation (1.4.3) then implies that the matching function is also independent of $L$, and therefore so are the network characteristic functions defined by (1.3.19) and (1.3.20). In other words, the parameter $L$ affects equilibrium variables only by scaling firm size one-to-one. I therefore fix $L = 1$ and compare normalized moments of the model to the corresponding normalized moments of the data, as described in section 1.2.2.1.

The remaining 10 parameters of the model are estimated using simulated method of moments. Recall that the five sets of empirical moments discussed in sections 1.2.2.1-1.2.2.5 were respectively:

1. $\bar{X}_b$, the normalized quantile level of variable $X$ evaluated at the midpoint of quantile bin $b$;

2. $\bar{Q}_b^X$, the average quantile of variable $X$ for all firms with revenue falling in quantile bin $b$, given by equation (1.2.3);

3. $\bar{Q}_b^{S,X}$ and $\bar{Q}_b^{C,X}$, the average quantile of variable $X$ amongst all suppliers and customers respectively of all firms with revenue falling in quantile bin $b$, given by equations (1.2.4) and (1.2.5);

4. $\bar{D}_b^S$ and $\bar{D}_b^C$, the average normalized supplier and customer distances respectively amongst all firms with revenue falling in quantile bin $b$, given by equations (1.2.6) and (1.2.7);

5. $\bar{\rho}_b^{S,ret}$ and $\bar{\rho}_b^{C,ret}$, the dynamic moments capturing the rates at which firms retain old trading partners, given by equations (1.2.10) and (1.2.11).

One option for the estimation procedure is to target all of the moments described above. Since employment is highly correlated with revenue in the data, however, I choose to omit targeting the firm employment distribution ($\bar{L}_b$), as well as the correlation between revenue
and employment \((RQ_b^L)^{\prime}\). Furthermore, instead of targeting all of the moments that characterize firm-to-firm matching, I target only the revenue quantiles of suppliers and customers across firms \((RQ_b^{S,R} \text{ and } RQ_b^{C,R})\), and use the remaining matching moments as overidentification tests of model fit. This leaves \(13 \times N_{bin}\) sets of moments for estimating 10 parameters.

The estimation procedure is as follows. First, to reduce simulation error, I generate \(N_{sim}\) random seeds \((\tilde{\varepsilon}_\phi, \tilde{\varepsilon}_\delta)\) from a two-dimensional standard multivariate normal distribution.\(^{30}\) Then, for every candidate set of parameter values, I compute the theoretical moments corresponding to the targeted moments described above for a set of \(N_{sim}\) simulated firms. To do so, I first solve for the values of the steady-state network characteristic and matching functions at a set of \(N_{grid} \times N_{grid}\) points using the algorithm described in the appendix. I then solve for the functions \(R(\cdot), M_S(\cdot), M_C(\cdot), D_S(\cdot), D_C(\cdot), \rho_{ret}^S(\cdot), \text{ and } \rho_{ret}^C(\cdot)\) at these same grid points using equations (1.3.26), (1.3.11), (1.3.12), (1.4.33), (1.4.34), (1.4.35), and (1.4.36). Given the current values of \(v_\phi, v_\delta, \text{ and } \rho\), I then compute:

\[
\begin{bmatrix}
\log \phi \\
\log \delta
\end{bmatrix} =
\begin{bmatrix}
v_\phi \sqrt{1 - \rho^2} & \rho \nu_\phi \\
0 & \nu_\delta
\end{bmatrix}
\begin{bmatrix}
\tilde{\varepsilon}_\phi \\
\tilde{\varepsilon}_\delta
\end{bmatrix}
\tag{1.5.4}
\]

for each simulated firm (thereby maintaining consistency with the desired covariance matrix (1.5.1)), and then use bilinear interpolation to obtain the theoretical values of \(R, M_S, M_C, D_S, D_C, \rho_{ret}^S, \text{ and } \rho_{ret}^C\) for each firm.

Having computed the theoretical counterparts of the target moments, I then compute the distance between these and the empirical moments according to:

\[
D = (|\mathcal{M}_{data} - \mathcal{M}_{model}|)^T \mathcal{W} \ (|\mathcal{M}_{data} - \mathcal{M}_{model}|) \tag{1.5.5}
\]

where \(\mathcal{M}_{data}\) and \(\mathcal{M}_{model}\) are vectors containing the stacked empirical and model moments respectively, and \(\mathcal{W}\) is the pseudo-inverse of the covariance matrix of the empirical moment

\(^{30}\text{In order to obtain bounded support for the joint distribution of } \phi \text{ and } \delta, \text{ which is necessary for numerical solution of the model, I truncate the distributions of both } \tilde{\varepsilon}_\phi \text{ and } \tilde{\varepsilon}_\delta \text{ at the } 95^{th} \text{ percentiles.}\)
vector, estimated by bootstrapping techniques. Starting from an arbitrary initial choice of parameter values, I then execute a simulated annealing algorithm to minimize $D$. Standard errors are computed using a bootstrap procedure, in which I repeat the estimation procedure described above after replacing $M_{data}$ by the corresponding moments from a bootstrap re-sampling of the original data. To account for simulation error, I also regenerate the random seeds ($\tilde{\varepsilon}_\phi, \tilde{\varepsilon}_\delta$) each time the estimation is performed.

1.5.3 Results

1.5.3.1 Parameter estimates

The parameter values obtained using the estimation procedure described above are shown in Table 1. From this, we make several observations.

First, the estimated value of the mean static relationship cost $f$ appears to be small, but recall that total labor supply is normalized to 1 in the estimation, and therefore the estimate implies that around 7% of total production labor is used for managing relationships. At the firm-level, the model predicts that labor costs associated with managing existing trade relationships within a firm account for around 1.3% of total labor costs on average.

Second, the reset friction parameter $\nu$ affects the rate at which firms form new trading relationships and destroy existing ones. At these parameter estimates, the model predicts that the mean duration of a firm’s relationships with its suppliers and customers is around 1.9 years, which is very close to the empirically-measured mean relationship duration of 1.74 years. The model also predicts that the average relationship termination rate across firms is around 34%, which again is very close to the empirical supplier and customer termination rates of 38.4% and 30.1% respectively.

Third, although the substitution elasticity $\sigma$ is not very precisely estimated, the point estimate plus or minus one standard error falls well within the range of values typically

---

31I resample with replacement 2000 times from the set of firms for both the Capital IQ and Compustat datasets, and compute the covariance matrix resampled data. I do not perform resampling along the time dimension for the Compustat data, although in principle this is possible using block bootstrapping techniques.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean of relationship cost</td>
<td>0.070</td>
<td>0.02</td>
</tr>
<tr>
<td>meeting friction</td>
<td>0.647</td>
<td>0.03</td>
</tr>
<tr>
<td>shape of relationship cost shock</td>
<td>0.585</td>
<td>0.11</td>
</tr>
<tr>
<td>elasticity of substitution</td>
<td>3.02</td>
<td>0.27</td>
</tr>
<tr>
<td>input suitability</td>
<td>0.347</td>
<td>0.09</td>
</tr>
<tr>
<td>variance of fundamental productivity</td>
<td>0.364</td>
<td>0.06</td>
</tr>
<tr>
<td>variance of fundamental quality</td>
<td>0.544</td>
<td>0.06</td>
</tr>
<tr>
<td>correlation between φ and δ</td>
<td>0.241</td>
<td>0.07</td>
</tr>
<tr>
<td>trade cost level</td>
<td>0.688</td>
<td>0.18</td>
</tr>
<tr>
<td>elasticity of trade cost with distance</td>
<td>0.348</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Table 1: Estimated parameter values

estimated in the literature.\textsuperscript{32} This is reassuring given that the estimation is based on data in which the intensive margin of trade (transaction values) is unobserved.

Finally, the parameters governing the distribution of fundamental firm characteristics appear to be well identified, with relatively small standard errors, but the trade cost parameters are less precisely estimated. As discussed below, this is perhaps related to the inability of the model to match the qualitative relationship between firm size and trading partner distance.

1.5.3.2 Model fit

To examine the model’s fit with data, Figures 8-12 reproduce the graphs characterizing the empirical moments described in section 1.2.2, but with the model’s simulated moments overlaid. With regard to the firm-level distributions shown in Figure 8, we see that the theoretical firm revenue distribution closely approximates the empirical distribution, and takes on the same log-normal shape. The firm in-degree and out-degree distributions, on the other hand, are harder for the model to match exactly, although the theoretical and empirical distributions share the same convex shape. Comparing the theoretical degree distributions to the Poisson (random matching) and Pareto (preferential attachment) approximations described in section 1.2.2.1, we see that the model’s predicted distributions lie somewhere

\textsuperscript{32}See for example Broda and Weinstein (2006).
between the distributions of the two parametric forms. This is perhaps not surprising, given
that the structural model features both elements of random reset shocks as well as preferential
activation (and non-termination) with larger suppliers and customers. The firm employment
distribution predicted by the model (which is untargeted in the estimation) resembles the
empirically-observed employment distribution in terms of the log-normal shape, but the fit
is poorer compared to the revenue distribution.

Figure 9 shows the model’s fit with regard to the correlation of firm revenue with em-
ployment, in-degree, and out-degree. As in the data, the model predicts that firms with
larger revenue also tend to have larger employment, more suppliers, and more customers.
Furthermore, the model closely matches the specific quantiles of these variables for firms in
each revenue quantile bin, even for the untargeted employment distribution.

Next, we examine the model's fit with regard to the assortativity of matching between
firms, shown in Figure 10. From these graphs, we see that the model is able to reproduce the
positive assortative matching between firms documented in the data, whether with regard to
revenue (targeted), or employment, in-degree, and out-degree (untargeted). However, in each
case, the model fit is better for firms at the upper-end of the revenue distribution. The fit
with regard to matching between firms and their suppliers in terms of revenue, for example,
is almost perfect for firms with revenue above the median, but is poorer for firms with
revenue below the median. This suggests that the economic tradeoffs involved in forming
and terminating trading relationships may be significantly different for small versus large
firms. In particular, the empirical moments of the matching distributions imply that small
firms are likely to match with suppliers and customers that are larger than the theoretical
mechanism in the model suggests.

With regard to the geographic distribution of a firm's suppliers and customers, Figure
11 shows that the model is unable to replicate the qualitative feature of the data that larger
firms tend to match with trade partners that are located closer to themselves, although in
terms of levels the average normalized distances to suppliers and customers predicted by the
Figure 8: Model fit: firm-level distributions
Figure 9: Model fit: Bivariate distributions
Figure 10: Model fit: matching distributions
model for larger firms are not too far off from the corresponding empirical moments. This discrepancy between model and data suggests that additional theoretical mechanisms beyond the relationship frictions studied in this paper are needed to generate both positive assortative matching between firms as well as average trade partner distances that decline with firm size. The pattern observed in Figure 11 might be generated by a trade model featuring an endogenous geographic distribution of firms with positive externalities in each location, for example, so that larger firms tend to be located closer to larger firms. Embedding endogenous geography, however, is beyond the scope of this paper.

Finally, Figure 12 shows the model’s fit with respect to the moments characterizing firm relationship dynamics. Here, we see that the model replicates the empirically-observed positive relation between firm size and the rate at which firms retain existing suppliers and customers, although the exact moments do not line up perfectly. Nonetheless, as discussed above, the predicted relationship durations and relationship termination rates are very close to their empirical counterparts on average.
1.6 Counterfactuals

Having estimated the parameters of the model, I now return to addressing the key question initially posed in the introduction to this paper: what are the quantitative implications of stickiness in firm-to-firm relationships for the responses of aggregate trade patterns and welfare to shocks? To answer these questions, I study the model’s transition dynamics in response to three kinds of counterfactual changes: declines in trade costs (section 1.6.1), declines in relationship costs (section 1.6.2), and idiosyncratic fluctuations in firm-level characteristics (section 1.6.3). I also examine the importance of accounting for rational firm expectations in computing these counterfactual dynamics (section 1.6.4), and revisit the efficiency of the dynamic market equilibrium by studying a simple policy exercise in which the fixed relationship cost is subsidized by a planner who obtains revenue from an ad valorem import tax (section 1.6.5).

1.6.1 Trade cost shocks

To examine how sticky relationships affect the dynamic responses of aggregate trade volumes and welfare to trade cost shocks, I study the model’s transition dynamics following a change in the overall trade cost level $\kappa$ to some counterfactual level, starting from the steady-state of the model with parameters set at the SMM estimates. I assume that the shock hits the economy at $t = 0$ after all relationship cost shocks have been realized and all activation and termination decisions have been made, so that firms can readjust the intensive margin of trade in the initial period post-shock but not the extensive margin. In other words, the initial response of the economy to the trade cost shock takes the network of firm trade as fixed. From $t = 1$ onwards, firms adjust both the intensive and extensive margins of trade in response to the shock.

Recall that the aggregate value of imports at date $t$ from a location a distance $D$ away
is given by:

\[
\bar{R}_t(D) = \left( \frac{\alpha}{\mu} \right)^{\sigma - 1} \tau(D)^{1-\sigma} \Delta_{H,t} \int_{S_\chi} \int_{S_\chi} m_t \left[ \chi, \chi' | \tau(D) \right] \Delta_t(\chi) \Phi_t(\chi') dF_\chi(\chi) dF_\chi(\chi')
\] (1.6.1)

A decline in the cost of trade \( \tau(D) \) therefore affects trade volumes statically through a direct reduction in the cost of inputs purchased (via the term \( \tau(D)^{1-\sigma} \)), as well as dynamically through changes in the incentives that firms face in forming and terminating relationships (via the matching function \( m_t \)). In the initial period of the shock, the matching function is assumed to be fixed, and the short-run change in trade therefore occurs only through the static channel. In the long-run, the total effect of the trade cost shock on trade volumes incorporates adjustments of firm-to-firm trade along both the intensive and extensive margins.

Figure 13 shows the dynamic responses of trade and welfare following a uniform 5% decline in gross trade costs across all locations.\(^{33}\) The first graph shows the transition paths of exports from a given location (measured as the percentage change relative to the pre-shock steady-state) to locations integrated over each quadrant of the unit circle.\(^{34}\) The second and third graphs decompose these changes in trade volumes into changes along the extensive and intensive margins respectively, while the fourth graph shows changes in welfare. From these graphs, we observe the following. First, in the initial period of the shock, exports to all locations increase, with the total value of exports rising by around 8%. Since the set of active trading relationships is assumed to be fixed, all of these gains are generated by firms selling more to existing customers. Notice also that the initial increase in exports is larger for locations that are further away, so that the geographic distribution of trade immediately becomes more dispersed following the shock.

After the initial period, the decline in trade costs induces firms to accumulate more

\(^{33}\)Specifically, a change in \( \kappa \) corresponding to a 5% decline in the average trade cost measure \( \int_0^1 (1 + \kappa D)^\epsilon dD \).

\(^{34}\)Since all locations are symmetric, the values of exports and imports between any pair of locations are identical.
trading partners. Over time, the value of exports to all locations therefore continues to grow. Observe that along the transition path, the growth in the mass of active relationships is accompanied by a decline in the amount of trade per active relationship. The dynamic gains in aggregate trade are therefore driven solely by increases in the extensive margin of firm-to-firm trade. Once firms have fully adjusted their trading relationships in response to the shock, total exports to all locations are almost 30% higher relative to the pre-shock steady-state. The endogenous adjustment of firm-level relationships therefore amplifies the elasticity of aggregate trade with respect to trade costs by more than three times. Similarly, the welfare gains from the reduction in trade costs are close to four times higher in the post-shock steady-state than in the initial period of the shock (although the absolute welfare gains are small). Note that the dynamic amplification effect is larger for exports to more distant locations, so that the geographic dispersion of trade also increases over time.

In addition to studying a uniform decline in the cost of trade across all locations, we can also use the model to study the effects of a bilateral reduction in the costs of trade between a given pair of locations. Since the set of locations is continuous, a change in trade costs between a single pair of locations leaves aggregate variables in each location unchanged.\footnote{One can think of this as a small open economy assumption but applied to a pair of locations.} The response of trade is therefore given by equation (1.6.1) with $\Delta_{H,t}$, $\Delta_t (\cdot)$ and $\Phi_t (\cdot)$ held fixed at their respective pre-shock steady-states. Nonetheless, the economic mechanisms remain the same: the bilateral decline in trade costs affects trade volumes both statically and dynamically.

Figure 14 shows the responses of trade following a 5% decline in gross bilateral trade costs for different distances between importing and exporting locations.\footnote{Specifically, a change in $\kappa$ corresponding to a 5% decline in $(1 + \kappa D)^{\epsilon}$ for each value of $D$.} Again, we see that the initial increase in trade is dynamically amplified by the accumulation of additional trading partners by firms in response to the trade cost shock, and that the magnitude of the amplification is around a factor of three for all locations but is larger for more distant locations. Note that the response of trade in the initial period of the shock (the x-intercept
in the first graph) is determined solely by the elasticity of substitution $\sigma$, as it would be in the frictionless model.

### 1.6.2 Relationship cost shocks

Lower variable trade costs reduce the cost of firm-to-firm trade along the intensive margin. How do trade patterns and welfare respond to changes in the cost of firm-to-firm trade along the extensive margin when firm relationships are sticky? To study this, I examine the model’s transition dynamics following a change in the average value $f$ of the relationship cost shock. Again, I assume that the shock hits the economy at $t = 0$ after all relationships have been set, and only allow firms to create and terminate relationships from $t = 1$ onwards. Furthermore, to enable consistent quantitative comparison with the results of the previous section, I compute the magnitude of the change in $f$ in the following way.

Consider a decline in variable trade costs across all locations corresponding to a change in $\kappa$ to some counterfactual level $\kappa'$. The cost of this change across steady-states if it were
Figure 14: Responses of trade and welfare to 5% decline in bilateral trade costs

to be implemented by an ad valorem subsidy to exports would be given by:

\[
T_\kappa \left( \kappa, \kappa' \right) = \int_0^1 \left[ (1 + \kappa D)^\epsilon - (1 + \kappa' D)^\epsilon \right] \tilde{R} \left( D | \kappa' \right) dD
\]  

(1.6.2)

where \( \tilde{R} \left( \cdot | \kappa' \right) \) is the aggregate value of trade in the steady-state corresponding to \( \kappa' \). Similarly, the cost of a decline in \( f \) to some counterfactual value \( f' \) if it were to be implemented by a subsidy to the cost of maintaining relationships would be equal to:

\[
T_f \left( f, f' \right) = \left( f - f' \right) L_f \left( f' \right)
\]  

(1.6.3)

where here \( L_f \left( f' \right) \) is the total mass of labor used to pay relationship fixed costs in the steady-state corresponding to \( f' \). With \( \kappa \) and \( f \) set at the SMM parameter values, I therefore compute the value of \( f' \) such that \( T_f \left( f, f' \right) = T_\kappa \left( \kappa, \kappa' \right) \) for a given value of \( \kappa' \).

Figure 15 shows the responses of aggregate trade and welfare in response to a decline in \( f \) corresponding to the 5% decline in global variable trade costs studied in section 1.6.1.\textsuperscript{37}

From these graphs, we see that the effects of lower relationship costs are qualitatively similar.

\textsuperscript{37}In terms of parameter values, the comparison is between a 50% decline in \( \kappa \) versus an 18% decline in \( f \).
to the effects of lower variable trade costs: exports to all locations increase over time, driven by growth in the mass of active relationships and accompanied by a decline in the intensive margin of trade. Quantitatively, however, the effects of a decrease in $f$ on aggregate trade and welfare are much larger than the corresponding effects following a decrease in $\kappa$. The increase in total exports in the post-shock steady-state relative to the pre-shock steady-state is around 50% higher than the corresponding increase resulting from the decline in variable trade costs. Similarly, the long-run welfare gains are around 75% higher. Since the rates of adjustment in response to the shocks are similar in the two cases, these results suggest that policy measures targeting the frictions that firms face in establishing trading relationships can be equally as if not more cost-effective than ad valorem trade subsidies.

As in section 1.6.1, we can also study the effects of a decline in the bilateral cost of relationships between firms in a given pair of locations. The results (not shown) are similar, with a decline in $f$ generating larger gains in trade and welfare than a cost-equivalent decline in $\kappa$. 

Figure 15: Responses of trade and welfare to global decline in relationship costs equivalent to 5% decline in trade costs
1.6.3 Idiosyncratic fluctuations and aggregate dynamics

To study how shocks to firm-level fundamental characteristics translate into aggregate dynamics, I next consider the following counterfactual exercise. Suppose that at $t = 0$, the economy is initially in steady-state. Next, suppose that all firms receive an unexpected but permanent shock to their fundamental characteristics that leaves the distribution of states across firms unchanged. In particular, suppose that the post-shock fundamental productivities and qualities of a firm are given respectively by:

$$
\log \hat{\phi} = \sqrt{1 - s} \log \phi + \sqrt{s} \hat{\omega}_\phi
$$

(1.6.4)

$$
\log \hat{\delta} = \sqrt{1 - s} \log \delta + \sqrt{s} \hat{\omega}_\delta
$$

(1.6.5)

where the idiosyncratic shocks $\hat{\omega}_\phi$ and $\hat{\omega}_\delta$ are jointly normal with the same covariance matrix as $\log \phi$ and $\log \delta$, and where the parameter $s$ captures the ratio of the shock variance to the variance of pre-shock firm states. Under this specification, it is straightforward to verify that the distribution of $\hat{\phi}$ and $\hat{\delta}$ across firms is identical to the pre-shock distribution of $\phi$ and $\delta$. It is immediately obvious from this that in a model without costly relationships ($f = 0$), this shock would have no effect on the aggregate economy at all. In a world with sticky relationships, however, even such idiosyncratic fluctuations have aggregate effects.

As before, I assume that the shock hits the economy at $t = 0$ after all relationships have been set. Even though individual firm pairs cannot activate new relationships or terminate existing ones, however, the matching function still responds instantaneously to the fluctuation shock, not because firms adjust the identity of their trading partners, but because the states of individual firms change. In particular, the matching function at date 0 adjusts instantaneously to:

$$
\tilde{m}_0 (\hat{x}, \hat{x}') = \frac{\int_{S_x} \int_{S_x} m_{ss} (x, x') q (\hat{x} | x) q (\hat{x}' | x') dF_x (x) dF_x (x')}{\int_0^\infty \int_0^\infty q (\hat{x} | x) q (\hat{x}' | x') dF_x (x) dF_x (x')}
$$

(1.6.6)
where $q$ is the transition function between pre- and post-shock states implied by (1.6.4) and (1.6.5). Since the structural parameters of the model remain unchanged, the steady-state of the economy is the same as before the shock. However, firm relationships are “scrambled” by the idiosyncratic fluctuation in firm fundamental characteristics, and it takes time for the economy to return to its steady-state as firms readjust their relationships.

Figure 16 shows the responses of trade and welfare to the fluctuation shock for different values of the relative shock variance $s$. We observe that when $s$ is very small, the fluctuation in firm states has little effect on aggregate quantities. However, as $s$ starts to increase, the responses of trade and welfare grow quickly. With relative shock variances of 10% and 20%, aggregate trade falls immediately by about 10% and 30% respectively. Welfare also falls as firm states are scrambled, although again the magnitude of the effect is small. Furthermore, the economy only gradually returns to the steady-state, with the half-life of the trade and welfare responses being approximately two years.

This effect of idiosyncratic fluctuations on aggregate dynamics in the model can be considered complementary to the effects studied in Acemoglu et al (2012), where the authors examine the role of sector-level input-output structures in translating idiosyncratic shocks into aggregate fluctuations. In the model studied here, idiosyncratic shocks generate aggregate dynamics because the input-output structure of the economy at the firm level is endogenous, and responds to shocks that would have no aggregate effects in a model without relationship frictions.

### 1.6.4 The importance of rational expectations

Being able to solve for the model’s exact transition dynamics under rational expectations allows us to compare the model’s predictions to what would be obtained under the assumption that firms are myopic. As previously discussed, a common approach to modeling strategic network formation between atomistic agents is to assume that agents receive the chance to create or destroy links with finite probability, but that given the chance to
change a relationship, the decision is made myopically based only on the static changes to the agent’s payoff.

To study the implications of myopia and therefore the importance of taking rational firm expectations into account, I study the model’s predictions under the alternative assumption that the relationship acceptance function is given by (1.4.8) instead of (1.4.13), and compute the transition dynamics in response to the same global decline in variable trade costs discussed in section 1.6.1. Figure 17 shows the transition paths of trade and welfare (analogous to Figure 13), from which we observe the following. First, the short-run change in trade and welfare under both myopia and rational expectations is the same, because the matching function is held fixed. However, once firms are allowed to adjust the extensive margin of trade, the transition dynamics and the eventual steady-state of the model differ substantially under myopia relative to the rational expectations equilibrium. In particular, myopic firms form too many relationships relative to the rational expectations equilibrium, and welfare initially declines following the trade cost shock before increasing to a steady-state level that is about 25% lower than the rational expectations equilibrium steady-state. This divergence in both the qualitative as well as quantitative properties of the model under myopia clearly
shows that taking agents’ rational expectations into account can have a crucial impact on theoretical predictions.

Figure 17: Responses of trade and welfare to 5% decline in global trade costs with myopic firms

1.6.5 Trade policy and sticky relationships

Given the central role of relationship stickiness in this paper, a natural policy question to ask is: can household welfare be improved by subsidies to the cost of forming relationships? To provide a first look into the effects of trade policy under sticky firm relationships, I consider the following stylized counterfactual. Suppose that for every relationship formed by a seller in each location, the policymaker in that location pays a fraction $S_f$ of the fixed relationship cost, financed fully by an ad valorem import tax $T_M$. In other words, policymakers tax the intensive margin of trade to subsidize the extensive margin. Without transport costs ($\kappa = 0$), for example, the steady-state matching function under such a combination of policies would be:

$$m \left( \chi, \chi' \right) = \tilde{m} \left[ \frac{\Delta \left( \chi \right) \Phi \left( \chi' \right) \Delta_H}{(1 - S_f) (1 + T_M)^{\sigma - 1}} \right]$$  \hspace{1cm} (1.6.7)
where $\tilde{m}$ is as defined by (1.4.22), and where the firm network characteristic functions are given by:

$$\Phi(\chi) = \phi^{\sigma-1} + \left[ \frac{\alpha}{\mu (1 + T_M)} \right]^{\sigma-1} \int_{S_\chi} m(\chi, \chi') \Phi(\chi') dF_{\chi}(\chi')$$  \hspace{1cm} (1.6.8)

$$\Delta(\chi) = \mu^{-\sigma} \delta^{\sigma-1} + [\mu (1 + T_M)]^{-\sigma} a^{\sigma-1} \int_{S_\chi} m(\chi', \chi) \Delta(\chi') dF_{\chi}(\chi')$$  \hspace{1cm} (1.6.9)

Balanced budgets in each location then require:

$$S_f L_f = T_M \bar{R}$$  \hspace{1cm} (1.6.10)

where $L_f$ is given by equation (1.4.14) and $\bar{R}$ is total import expenditure:

$$\bar{R} = \left[ \frac{\alpha}{\mu (1 + T_M)} \right]^{\sigma-1} \Delta_H \int_{S_\chi} m(\chi, \chi') \Delta(\chi) \Phi(\chi') dF_{\chi}(\chi) dF_{\chi}(\chi')$$  \hspace{1cm} (1.6.11)

Figure 18 shows the percentage change in household welfare across steady-states relative to the no-policy equilibrium for different values of $S_f$. Evidently, the model implies that firm relationship cost subsidies can be welfare improving even when financed by import taxes that distort the intensive margin of trade. This is a result of the fact that the market equilibrium is inefficient relative to the social planner’s allocation, as characterized by Propositions 2 and 3.

### 1.7 Conclusion

This paper set out to study and quantify the effects of stickiness in firm-to-firm trading relationships on aggregate patterns of trade. The theoretical model developed to address these questions is able to adeptly match the majority of empirical moments relating to the distributions of relationships across firms, the correlation between firm connectivity and firm size, the assortativity of matching between firms, and the persistence of firm-to-firm
relationships. Numerical estimation and counterfactual simulation of the model then suggest that firm-level relationship frictions matter for understanding patterns of aggregate trade in several key ways. First, endogenous adjustment of sticky firm relationships dynamically amplifies the response of trade and welfare to macroeconomic shocks. Second, subsidies to the cost of firm-level trade along the extensive margin can be a more cost-effective means of increasing aggregate trade and welfare than subsidies along the intensive margin. Third, idiosyncratic fluctuations at the firm-level can generate large and persistent aggregate trade dynamics when firm relationships are sticky. Finally, selection of trading relationships by profit-maximizing firms in the presence of relationship stickiness can be socially sub-optimal, with scope for welfare-improving subsidies to the formation of firm-level linkages.

The issues confronted in this paper also provide scope for future research. In particular, the model’s inability to fit the matching distributions of firms at the lower-end of the revenue distribution suggest that more nuanced theory regarding the matching process may be needed to resolve this discrepancy. Extensions of the model, for instance, may consider the role of information in firm network formation, how such information propagates across firms, and how informational frictions may affect smaller versus large firms differentially. Furthermore, the empirical finding that larger firms tend to trade with partners that are closer by on average goes against not only predictions of the model developed in this paper, but also

Figure 18: Effect of relationship cost subsidies on household welfare
the standard intuition arising from heterogeneous-firm models of international trade that larger firms are more likely to export to more costly locations. This hints at a role for economic geography models in exploring the potentially-rich interaction between sticky firm relationships and the endogenous geographic locations of firms.
Chapter 2

Firm-to-firm Productivity Dynamics and International Trade Flows

2.1 Introduction

Firm-to-firm buyer-seller relationships are the threads that constitute the fabric of international trade flows. Understanding what shapes the resilience of these threads can therefore be useful for understanding aggregate trade patterns. In particular, how persistent are firm-to-firm relationships across different locations, and what enables some relationships to survive over time while others terminate? How do the dynamics of these relationships matter for the responses of welfare and trade flows to shocks?

In this paper, I study how relationship persistence is determined by the dynamic properties of relationship-specific productivity fluctuations, as well as how these properties matter for the gains from trade. Using data on trading relationships amongst US firms, I first show that the hazard rate of relationship termination decreases with both the age of the relationship as well as buyer and seller size, even after controlling for firm exit rates. An increase in the age of a trading relationship by one year or a 1% increase in either buyer or seller size is associated with a 2% decline in the hazard rate of relationship termination.

I then propose a simple theory that rationalizes this empirical observation: firms engaged in a trading relationship receive productivity shocks that are specific to the relationship, and these shocks are both persistent and age-dependent. Persistence of productivity shocks may be interpreted as modeling the fact that there are unobserved characteristics of trading relationships that enable some to survive instead of others, and that these characteristics endure over time. Age-dependence of the shock process, on the other hand, may be interpreted as a
reduced-form means of capturing learning within the relationship that leads to improvements in productivity, or as the accumulation of intangible capital specific to the relationship. I show that both these features of the productivity process - persistence and age-dependence - are needed to generate the prediction that both age and firm size matter separately for the persistence of the relationship.

To study the implications of this mechanism for patterns of aggregate trade flows and the gains from trade, I then embed it in a general equilibrium model of international trade. The model features search frictions between buyers and sellers in the style of Mortensen and Pissarides (1994): firms pay search costs at each date to engage in trading relationships, and once a relationship forms, it receives idiosyncratic productivity shocks that lead to endogenous relationship survival and termination. Using numerical simulations of the model, I then show that both the persistence of the relationship productivity process as well as the rate at which relationship capital accumulates matter for the responses of trade and welfare to changes in trade costs. In particular, more persistent shocks and slower rates of relationship capital growth lead to stronger responses of welfare.

To examine the sources of these effects, I derive an analytic decomposition of changes in welfare in the model that consists of three components. The first accounts for changes in the mass of firms that search for relationships (the entry effect), the second accounts for changes in average productivity of active relationships (the selection effect), and the third accounts for changes in the home trade share (the trade effect). I then examine how the properties of the relationship productivity process matter for the strengths of these various effects.

This paper builds on a growing literature that examines trade flows at the margin of the buyer-seller relationship. In modeling how the hazard rate of relationship termination varies with age, the theory developed in this paper is most closely related to work by Monarch and Schmidt-Eisenlohr (2016), who offer a model of importer learning to explain

\footnote{For recent work, see for example Oberfield (2015), Lim (2016), Chaney (2014, 2015), Bernard, Moxnes, and Ulltveit-Moe (2015), Bernard, Moxnes, and Saito (2015), Monarch (2014), and Macchiavello and Morjaria (2015).}
the dynamics of trading relationships involving US firms and foreign trade partners. This paper also contributes to the already-vast literature on firm-level dynamics in export markets, although the emphasis in what follows is on dynamics at the firm-to-firm level instead.39

The outline of this paper is as follows. I begin in section 2.2 by providing an empirical motivation for the theory developed in the model, using data on trading relationships amongst US firms to show that the hazard rate of relationship termination declines with age, even after controlling for buyer and seller size and exit rates. In section 2.3, I then describe a closed-economy version of the model to introduce the key theoretical mechanisms, which are (1) search frictions between suppliers and producers, and (2) relationship-specific productivity fluctuations that are persistent with a trend that grows with the age of the relationship. In section 2.4, I then extend the model to include multiple locations. Section 2.5 discusses calibration of the model’s parameters and results of numerical simulations, while section 2.6 concludes.

2.2 Empirical Motivation

To study the dynamics of firm-to-firm trading relationships, I employ data from Standard and Poor’s Compustat platform, which collects fundamental information for publicly-listed firms in the US. For a subset of these firms, the database also records supplier and customer relationships based on financial disclosure forms that firms are required to file. These forms include firms’ own reports of who their major customers are, where in accordance with Financial Accounting Standards No. 131, a major customer is defined as a firm that accounts for at least 10% of the reporting seller’s revenue. These customer reports are then matched back to the set of reporting firms, to generate a dataset with 103,379 firm-year observations and 33,873 relationship-year observations from 1979 to 2007.40

Using this dataset, I then compute the age of each relationship as follows: a relationship

---

40 The original Compustat relationship data was cleaned and studied by Atalay et al (2011).
is defined to be of age one if it is absent from the dataset in the preceding year, and to be of age $a + 1$ if it is of age $a$ in the preceding year. Furthermore, a relationship is defined to be terminated in year $t$ if it is absent from the dataset in year $t + 1$. Using these measures, I then examine how the hazard rate of relationship termination varies with age.\footnote{In the baseline analysis, I set the age of all relationships in the initial year of the dataset to one. The empirical results discussed in this section, however, are robust to excluding the first $n$ years of data at least up to $n = 15$ (approximately half of the dataset).}

First consider Figure 19, which shows how the average relationship termination rate (i.e. the fraction of relationships that are terminated) varies with age. Evidently, relationships that are older exhibit a smaller rate of termination, with most of the decline in termination rates occurring over the first five to ten years of the relationship.\footnote{This finding is consistent with similar patterns documented using international customs data for relationships between US importers and Chinese exporters - see Monarch and Schmidt-Eisenlohr (2016).} Furthermore, this relation is distinct from the well-known negative correlation between age and exit at the firm level. Relationship age is associated with a lower termination rate amongst relationships that see both the buyer and seller surviving, but not amongst relationships that terminate because either the buyer or seller (or both) exit from the dataset.

A potential explanation for the negative association between relationship age and the hazard rate of termination is that both the buying and selling firms tend to grow as the relationship ages, and that relationships involving larger firms are more likely to survive. To examine this hypothesis, I estimate different specifications of a binary response model with relationship termination as the dependent variable, and with relationship age and buyer and seller sizes as explanatory variables. Table 2 shows the coefficient estimates from this estimation exercise, while Table 3 shows the corresponding sample average marginal effects of each covariate on relationship termination.

From these results, we observe the following. First, without firm size or year controls, relationship age is negatively associated with relationship termination (column 1), even when the sample is restricted to the set of relationships without exit by either buyer or seller (column 2). This is consistent with the pattern documented in Figure 19. Second, controlling
for buyer and seller size (column 3) shows that relationships involving larger firms are less likely to terminate. The coefficient on age is smaller in magnitude but still significantly negative. Third, the negative association between relationship termination on the one hand and age and firm size on the other is robust to the inclusion of year fixed-effects (columns 4), and both probit and logit specifications (columns 5 and 6) yield similar results. Fourth, in terms of magnitudes (Table 3), an increase in the age of a relationship by one year is associated with around a 2% decline in termination probability, similar to the marginal effect of a 1% increase in either buyer or seller revenue. Finally, to allow for non-linear effects of age, I also estimate a variant of the probit model in column 4 with separate dummy indicators for each age value. The marginal mean termination rate for each age (i.e. the predicted termination rate holding age fixed at a given value) is shown in Figure 20, which closely resembles the unconditional termination rates shown in Figure 19.

In sum, both the age of a trading relationship as well as the size of the firms involved in the relationship appear to matter for predicting the hazard rate of relationship termination. I now turn attention towards developing a general equilibrium model of international trade, in which these features are generated by the properties of stochastic process governing the evolution of relationship-specific productivities.
Figure 20: Marginal response of relationship termination rate to relationship age (estimates from probit model with buyer/seller size controls and year fixed effects)
### Table 2: Regression estimates for binary response models of relationship termination

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationship age</td>
<td>−.056***</td>
<td>−.086***</td>
<td>−.067***</td>
<td>−.073***</td>
<td>−.127***</td>
<td>−.135***</td>
</tr>
<tr>
<td></td>
<td>(.014)</td>
<td>(.012)</td>
<td>(.013)</td>
<td>(.004)</td>
<td>(.024)</td>
<td>(.022)</td>
</tr>
<tr>
<td>Log seller revenue</td>
<td>−.062***</td>
<td>−.059***</td>
<td>−.103***</td>
<td>−.098***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.006)</td>
<td>(.004)</td>
<td>(.010)</td>
<td>(.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log buyer revenue</td>
<td>−.054***</td>
<td>−.055***</td>
<td>−.086***</td>
<td>−.090***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.005)</td>
<td>(.004)</td>
<td>(.009)</td>
<td>(.010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>probit</td>
<td>probit</td>
<td>probit</td>
<td>probit</td>
<td>logit</td>
<td>conditional logit</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>(yes)</td>
</tr>
<tr>
<td>Sample</td>
<td>all relationships</td>
<td>no buyer or seller exit</td>
<td>no buyer or seller exit</td>
<td>no buyer or seller exit</td>
<td>no buyer or seller exit</td>
<td>no buyer or seller exit</td>
</tr>
<tr>
<td>N</td>
<td>38,373</td>
<td>33,854</td>
<td>33,504</td>
<td>33,504</td>
<td>33,504</td>
<td>33,504</td>
</tr>
</tbody>
</table>

**Notes:** Table reports results of binary response model regression coefficients, where the dependent variable is relationship termination and the main explanatory variable is relationship age. Each observation is a buyer-seller relationship in a given year, and each column reports results for a different model specification. Standard errors clustered by year are shown in parentheses, where *** denotes significance at the 5% level.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationship age</td>
<td>−0.021***</td>
<td>−0.029***</td>
<td>−0.015***</td>
<td>−0.024***</td>
<td>−0.025***</td>
<td>−0.026***</td>
</tr>
<tr>
<td></td>
<td>(.005)</td>
<td>(.004)</td>
<td>(.004)</td>
<td>(.003)</td>
<td>(.004)</td>
<td>(.004)</td>
</tr>
<tr>
<td>Log seller revenue</td>
<td>−0.014***</td>
<td>−0.019***</td>
<td>−0.020***</td>
<td>−0.019***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.006)</td>
<td>(.002)</td>
<td>(.002)</td>
<td>(.002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log buyer revenue</td>
<td>−0.019***</td>
<td>−0.018***</td>
<td>−0.017***</td>
<td>−0.017***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.003)</td>
<td>(.002)</td>
<td>(.002)</td>
<td>(.002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>probit</td>
<td>probit</td>
<td>probit</td>
<td>probit</td>
<td>logit</td>
<td>conditional logit</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>(yes)</td>
</tr>
<tr>
<td>Sample</td>
<td>all relationships</td>
<td>no buyer or seller exit</td>
<td>no buyer or seller exit</td>
<td>no buyer or seller exit</td>
<td>no buyer or seller exit</td>
<td>no buyer or seller exit</td>
</tr>
<tr>
<td>N</td>
<td>38,373</td>
<td>33,854</td>
<td>33,504</td>
<td>33,504</td>
<td>33,504</td>
<td>33,504</td>
</tr>
</tbody>
</table>

**Notes:** Table reports sample average marginal effects from regression estimates of binary response models, where the dependent variable is relationship termination and the main explanatory variable is relationship age. Each observation is a buyer-seller relationship in a given year, and each column reports results for a different model specification. Standard errors computed using the delta method are shown in parentheses, where *** denotes significance at the 5% level.

Table 3: Sample average marginal effects on relationship termination
2.3 Closed Economy Model

To elucidate the key features of the theory, I first describe a version of the model in which there is trade between buyers and sellers, but all firms operate in one location with a common wage so that there is no trade in an international sense. In section 2.4, I then extend the model to incorporate multiple heterogeneous locations.

2.3.1 Model environment

2.3.1.1 Production Technology

At each date, there is a continuum of upstream suppliers and a continuum of downstream producers. Each final good in the economy is produced by an active relationship between one supplier and one producer, and each firm can be involved in at most one relationship at a time. At date $t$, a downstream producer charging a price $p^D_t$ receives final demand for its product given by:

$$x^D_t = A_t (p^D_t)^{-\sigma}$$

(2.3.1)

where $A_t$ is a demand shifter that is determined in general equilibrium, described in section 2.3.1.4.

Now consider a pair of firms engaged in an active relationship at date $t$. The upstream supplier can produce intermediate inputs for the relationship using labor alone with a constant returns to scale technology:

$$x^U_t = z_t l^U_t$$

(2.3.2)

where $l^U_t$ denotes the quantity of labor used by the supplier. The productivity $z_t$ of the technology is assumed to consist of two components:

$$\log z_t = \log \epsilon_t + \log \kappa (a_t)$$

(2.3.3)

The first component $\epsilon_t$ captures idiosyncratic productivity variations across relationships.
and follows a first-order Markov process with properties listed in assumption 3. To differentiate $\epsilon_t$ from the overall productivity of the relationship, I henceforth refer to it as the efficiency of the relationship, which may be interpreted as capturing unobserved characteristics of the relationship that vary over time. The initial value of efficiency for new relationships is assumed to be drawn from the stationary distribution of the Markov process, which is unique by part (i) of Assumption 3. Boundedness of the state space for $\epsilon_t$ (part (ii)) and the assumed properties of the Markov transition function (parts (iii) and (iv)) will allow us to use standard dynamic programming techniques to establish existence and uniqueness of equilibria, while the assumption that the conditional efficiency distribution has finite $b^{th}$ moment (part (v)) ensures that firm averages remain bounded.

**Assumption 3.** The stochastic process for $\epsilon_t$ is first-order Markov with (i) a unique stationary distribution, and (ii) support on the interval $[\epsilon, \bar{\epsilon}]$ with $0 \leq \epsilon < \bar{\epsilon} < \infty$. The transition function of the Markov process satisfies (iii) the Feller property and (iv) monotonicity, and (v) the conditional distribution of $\epsilon_t$ given any $\epsilon_{t-1} \in [\epsilon, \bar{\epsilon}]$ is in $L^b$ with $b \equiv (1 - \alpha)(\sigma - 1)$.

The second component in (2.3.3), $\kappa(a_t)$, captures how productivity changes with the age of the relationship, denoted by $a_t$. This can be interpreted as a reduced-form measure of intangible assets specific to the relationship, and thus I henceforth refer to $\kappa(a_t)$ as relationship capital. The function $\kappa$ is assumed to satisfy the properties specified in Assumption 4. Part (i) of this assumption implies that relationships become more productive on average as they age, while part (ii) ensures that newly-formed relationships can always produce. Part (iii) assumes that relationship capital stops growing after some maximum age $\bar{a}$, which is not essential for the qualitative properties of the model but makes numerical solution of the model simpler.

**Assumption 4.** The relationship capital function $\kappa$ is such that (i) $\kappa(a) > \kappa(a - 1)$ for all $a \in \{1, \cdots, \bar{a}\}$, (ii) $\kappa(0) = \kappa > 0$, and (iii) $\kappa(a) = \bar{\kappa}$ for all $a \geq \bar{a}$.

The production function for the downstream firm, on the other hand, is Cobb-Douglas
in labor and the supplier’s input:

\[ x_t^D = \left( \frac{l_t^D}{\alpha} \right)^{\alpha} \left( \frac{x_t^M}{1 - \alpha} \right)^{1-\alpha} \]  \hspace{1cm} (2.3.4)

where \( l_t^D \) and \( x_t^M \) denote the quantities of labor and intermediates used by the producer respectively, and the parameter \( \alpha \) captures the share of value-added. In addition, trade between the supplier and producer is subject to an iceberg trade cost \( \tau \geq 1 \), so that market clearing of intermediates requires:

\[ x_t^M = \frac{1}{\tau} x_t^U \]  \hspace{1cm} (2.3.5)

Finally, production in each period is assumed to require a fixed number of workers denoted by \( f \). This fixed operating cost will be what generates the margin of endogenous relationship destruction in the model, as relationships with low productivity will exit while others survive.

Note that all the dynamics of relationship specific productivity are embedded in the upstream supplier’s production process. This assumption is made without loss of generality if in addition it is also assumed that the downstream producer has perfect information about the productivity of the relationship, so that contracting frictions of the form discussed in Antras (2003) do not arise. Since the focus of this paper is on the dynamics of relationship-specific productivity rather than features of the contractual environment, I maintain the assumption of perfect information as a benchmark.

Given the production technology, each active supplier-producer pair then engages in generalized Nash bargaining every period over the surplus of the relationship, which is described below in section 2.3.1.3. Specifically, the supplier-producer pair choose the following: a price \( p_t^D \) charged by the downstream firm per unit of output sold to the household; a price \( p_t^U \) charged by the upstream firm per unit of intermediates produced; quantities \( x_t^D \), \( x_t^M \), and \( x_t^U \); labor inputs \( l_t^D \) and \( l_t^U \); and the share \( s_t \) of the fixed operating cost paid for by the
downstream firm.\(^{43}\) The flow payoff to the downstream producer can then be written as:

\[
\pi^D_t = p^D_t x^D_t - w^D_t l^D_t - p^U_t x^U_t - s_t f
\]  \hspace{1cm} (2.3.6)

where dependence of relationship-specific terms on the state \((\epsilon, a)\) of the relationship has been omitted for brevity.\(^{44}\) Similarly, the flow payoff to the upstream producer is given by:

\[
\pi^U_t = p^U_t x^U_t - w^U_t l^U_t - (1 - s_t) f
\]  \hspace{1cm} (2.3.7)

2.3.1.2 Search and matching

At the start of every period, each active supplier-producer pair observes its draw of \(\epsilon\). Both parties then decide whether to remain in the relationship or not. Regardless of the draw of \(\epsilon\), the relationship terminates with exogenous probability \(\delta \geq 0\).\(^{45}\) If both parties are willing to remain in the relationship and the exogenous termination shock is not received, production occurs.\(^{46}\)

Value of a relationship \hspace{1cm} Given the assumptions about the relationship productivity process, the state of a relationship can be completely summarized by its efficiency \(\epsilon\) and age \(a\). The value to a downstream producer at date \(t\) of being in an active relationship with state

\(^{43}\)Note that since prices and quantities are jointly chosen, the choice of the upstream firm’s price \(p^U_t\) is equivalent to a lump-sum transfer between the two firms. Therefore, allowing for two-part pricing with an explicit lump-sum transfer in addition to the ad valorem price is redundant.

\(^{44}\)In the closed economy, one can choose the wage as numeraire. In a model with multiple countries, however, it will be important to keep track of relative wages, and therefore I include notation for \(w_t\) throughout.

\(^{45}\)This ensures that a steady-state of the model exists even without endogenous relationship destruction.

\(^{46}\)In equilibrium, the solution to the Nash bargaining problem within a relationship ensures that the buyer and seller always agree about whether to maintain the relationship or not.
$(\epsilon, a)$ can then be written as:

\[
R_t^D (\epsilon, a) = \pi_t^D (\epsilon, a) + (1 - \delta) \int_{\xi}^{\epsilon} \max \left\{ R_{t+1}^D (\epsilon', a + 1), V_{t+1}^D \right\} Q \left( d\epsilon' \mid \epsilon \right) + \delta V_{t+1}^D
\]

where $V_{t+1}^D$ denotes the value to a downstream firm at date $t$ of not having an upstream supplier, and $Q (\cdot \mid \epsilon)$ denotes the conditional distribution of efficiency next period given current efficiency equal to $\epsilon$.\footnote{Note that as in Melitz (2003), there is no time discounting, since the exogenous exit rate $\delta$ effectively causes firms to discount the future.} Similarly, the value to the upstream supplier in the relationship satisfies:

\[
R_t^U (\epsilon, a) = \pi_t^U (\epsilon, a) + (1 - \delta) \int_{\xi}^{\epsilon} \max \left\{ R_{t+1}^U (\epsilon', a + 1), V_{t+1}^U \right\} Q \left( d\epsilon' \mid \epsilon \right) + \delta V_{t+1}^U
\]

where $V_{t+1}^U$ denotes the value to an upstream firm at date $t$ of not having a downstream customer.

**Matching technology** At date $t$, there are endogenously-determined masses $d_t$ and $u_t$ of downstream and upstream firms respectively that search for trading partners. The mass of matches that form is then given by $m (d_t, u_t)$, where the matching function $m$ satisfies the following properties.

**Assumption 5.** The matching function $m$ is (i) strictly increasing in both arguments, (ii) homogeneous of degree 1, and (iii) satisfies $\frac{m(d,u)}{d}, \frac{m(d,u)}{u} \in [0, 1]$ for all $d, u \in [0, \infty)$.
The probability that a downstream firm receives a match at date $t$ is therefore given by
\[
\frac{m(d_t,u_t)}{d_t} = m(1, \theta_t), \text{ where } \theta_t \equiv \frac{u_t}{d_t} \text{ is the market tightness. Similarly, the probability that an upstream firm receives a match is given by } \frac{m(d_t,u_t)}{u_t} = \frac{m(1, \theta_t)}{\theta_t}.
\]

**Value of searching** Each firm not in an active relationship has the option of searching for a trading partner, which entails the payment of a search cost in units of labor. The cost of search depends on the mass of firms searching, with costs for downstream and upstream firms at date $t$ denoted by $\gamma^D(d_t)$ and $\gamma^U(u_t)$ respectively. The search cost functions are assumed to satisfy the properties specified in Assumption 6. Part (i) of this assumption implies that there is congestion in the search process, which one can think of as the result of vacancies being purchased from perfectly competitive advertising firms that face convex costs of production. While this assumption is not essential for the closed economy model, together with part (ii) it ensures that every pair of locations in the open economy model will have positive masses of firms searching for trading relationships. Part (iii) will generate the result that market tightness is constant across time, which simplifies analysis of the model.

**Assumption 6**. The search cost functions $\gamma^D$ and $\gamma^U$ are (i) strictly increasing, (ii) satisfy $\lim_{d \to 0} \gamma^D(d) = \lim_{u \to 0} \gamma^U(u) = 0$, and (iii) are such that the ratio of upstream to downstream search costs $\gamma^U(u)/\gamma^D(d)$ is homogeneous of degree zero.

There is assumed to be free entry for both downstream and upstream firms, so that in equilibrium all firms not in active relationships will be indifferent between searching and exiting the economy completely. The value to a downstream producer at date $t$ of being out of a relationship is therefore given by:

\[
V_t^D = -\gamma^D (d_t) + m(1, \theta_t) \int_{\xi}^t \max \left\{ R_{t+1}^D \left( \epsilon', 0 \right) , V_{t+1}^D \right\} \bar{F} \left( d\epsilon' \right) \text{ max} \left\{ R_{t+1}^D \left( \epsilon', 0 \right) , V_{t+1}^D \right\} \bar{F} \left( d\epsilon' \right)
\]

\[
+ [1 - m(1, \theta_t)] V_{t+1}^D
\]

\[2.3.10\]
where $\bar{F}$ denotes the distribution of initial efficiencies. The corresponding value for an upstream supplier satisfies:

$$
V_t^U = -\gamma^U (u_t) + \left[ \frac{m(1, \theta_t)}{\theta_t} \right] \int_\epsilon^\epsilon \max \left\{ \bar{R}_{t+1}^U (\epsilon', 0), V_{t+1}^U \right\} F \left( d\epsilon' \right) 
+ \left[ 1 - \frac{m(1, \theta_t)}{\theta_t} \right] V_{t+1}^U
$$

(2.3.11)

Free entry then implies that the value of not having a trading partner must be equal to zero:

$$
V_t^D = V_t^U = 0
$$

(2.3.12)

2.3.1.3 Generalized Nash bargaining and relationship surplus

The surplus of a relationship with state $(\epsilon, a)$ can now be expressed as:

$$
S_t (\epsilon, a) = R_t^D (\epsilon, a) - V_t^D + R_t^U (\epsilon, a) - V_t^U
$$

(2.3.13)

It is then assumed that the buyer-seller pair in each active relationship chooses prices, output quantities, labor inputs, and the split of the fixed operating cost so as to maximize the Nash product:

$$
\left[ R_t^D (\epsilon, a) - V_t^D \right]^{\nu} \left[ R_t^U (\epsilon, a) - V_t^U \right]^{1-\nu},
$$

(2.3.14)

where $\nu$ measures the relative bargaining power of the downstream producer relative to its upstream supplier.

2.3.1.4 Households

The representative household inelastically supplies $L$ units of labor and has preferences over the set of goods produced by downstream firms at date $t$ given by:

$$
U_t = \left[ \int_\epsilon^\epsilon \sum_{a=0}^\infty \bar{n}_t (\epsilon, a) x_t^D (\epsilon, a)^{\frac{\sigma + 1}{\sigma}} d\epsilon \right]^{\frac{\sigma}{\sigma-1}}
$$

(2.3.15)
where $\tilde{n}_t(\epsilon, a)$ is the mass of relationships with state $(\epsilon, a)$, and $\sigma$ denotes the elasticity of substitution across goods. This form of the utility function gives rise to the demand function for final goods specified by (2.3.1), where the demand shifter is:

$$A_t = \frac{w_t L + \Pi_t}{P_t^{1-\sigma}}$$  \hspace{1cm} (2.3.16)

Here, $\Pi_t$ denotes aggregate profits in the economy, which are assumed to be rebated to the household and will turn out to be zero in equilibrium given free entry, while $P_t$ denotes the consumer price index:

$$P_t = \left[ \int_{\tilde{\epsilon}}^{\bar{\epsilon}} \sum_{a = 0}^{\infty} \tilde{n}_t(\epsilon, a) p_t^D(\epsilon, a)^{1-\sigma} d\epsilon \right]^{\frac{1}{1-\sigma}}$$  \hspace{1cm} (2.3.17)

### 2.3.2 Equilibrium

#### 2.3.2.1 Equilibrium conditions

**Generalized Nash bargaining solution** The solution to the Nash bargaining problem within each relationship is summarized by Proposition 4.\(^{48}\) Part (i) characterizes prices and the split of the relationship surplus and profits between the downstream and upstream firms. Note that the choices of the upstream price and shares of the fixed operating cost are equivalent means of transferring profit between the downstream and upstream firms, and therefore these variables are undetermined without further assumptions. Nonetheless, the shares of the relationship surplus and profits are pinned down by firms’ relative bargaining power. Furthermore, the price (2.3.18) charged by the downstream firm is identical to the price that a monopolistically-competitive downstream firm would charge if its upstream supplier charged an input price equal to its own marginal cost. Efficient bargaining between buyer and seller therefore avoids double marginalization within the relationship.

In the proof of Proposition 4, I also show that upstream firm size is proportional to

\(^{48}\)All proofs are relegated to the appendix.
downstream firm size if and only if the share of the fixed operating cost paid for by the downstream firm is equal to its relative bargaining power $\nu$. Since proportionality of revenues seems to be a natural benchmark for the model, I henceforth impose this in Assumption 7. This pins down the price charged by the upstream firm as given by (2.3.23), which is a constant markup over marginal cost. Note that the markups charged by both firms depend not only the elasticity of substitution $\sigma$, but also on bargaining power $\nu$ and value-added share $\alpha$. As $\nu \to 1$, the buyer has complete bargaining power and charges the standard CES markup of $\mu$ while the seller prices at marginal cost. As $\nu \to 0$ and $\alpha \to 0$, the downstream firm becomes irrelevant, so that the supplier charges the standard CES markup and the producer prices at marginal cost.

Part (ii) of Proposition 4 then characterizes the optimal choice of labor inputs, which are allocated in proportion to the share of value added $\alpha$ in the relationship, while part (iii) characterizes firm size. As mentioned above, Assumption 7 implies that the ratio of upstream firm size to downstream firm size is constant. The constant of proportionality, $\frac{b+1-\nu}{\sigma}$, again depends on firm bargaining power and value-added share. As $\nu \to 1$ and $\alpha \to 1$, the upstream firm becomes irrelevant so that upstream firm size goes to zero. As $\nu \to 0$ and $\alpha \to 0$, the downstream firm is irrelevant and simply passes on the output of the upstream firm to the household, and therefore upstream and downstream firm sizes are equal.

Assumption 7. Downstream firms pay a constant share $\nu$ of the operating cost in every active relationship.

Proposition 4. The solution to the Nash bargaining problem for a relationship of state $(\epsilon, a)$ at date $t$ is characterized as follows:

i. The downstream firm charges a final goods price given by:

$$p_D(t, \epsilon, a) = \mu w_t \left[ \frac{\tau}{eK(a)} \right]^{1-\alpha} \quad (2.3.18)$$
where \( \mu \equiv \frac{\sigma}{\sigma - 1} \), while the upstream firm price and split of the operating cost are chosen such that the relationship surplus is split in proportion to firms' bargaining power:

\[
R^D_t(\epsilon, a) - V^D_t = \nu S_t(\epsilon, a) \tag{2.3.19}
\]

\[
R^U_t(\epsilon, a) - V^U_t = (1 - \nu) S_t(\epsilon, a) \tag{2.3.20}
\]

Profits are also split in the same proportion, so that:

\[
\pi^D_t(\epsilon, a) = \nu \left[ B_t \left( \frac{\epsilon \kappa(a)}{\tau} \right)^b - f \right] \tag{2.3.21}
\]

\[
\pi^U_t(\epsilon, a) = (1 - \nu) \left[ B_t \left( \frac{\epsilon \kappa(a)}{\tau} \right)^b - f \right] \tag{2.3.22}
\]

where \( B_t \equiv \frac{1}{\nu} \mu^{1-\sigma} A_t w_t^{1-\sigma} \) and \( b \equiv (\sigma - 1)(1 - \alpha) \). If in addition Assumption 7 holds, then the upstream firm price is given by:

\[
p^U_t(\epsilon, a) = \left( \frac{b + 1 - \nu}{b} \right) \left[ \frac{w_t}{\epsilon \kappa(a)} \right] \tag{2.3.23}
\]

and the markups over marginal cost for the downstream and upstream firms are:

\[
\mu^D_t(\epsilon, a) = \mu \left( \frac{b + 1 - \nu}{b} \right)^{\alpha - 1} \tag{2.3.24}
\]

\[
\mu^U_t(\epsilon, a) = \frac{b + 1 - \nu}{b} \tag{2.3.25}
\]

ii. Labor inputs are chosen in proportion to the share of value-added \( \alpha \), and are given by:

\[
l^D(\epsilon, a) = \alpha \frac{\sigma}{\mu w_t} B_t \left[ \frac{\epsilon \kappa(a)}{\tau} \right]^{(\sigma - 1)(1 - \alpha)} \tag{2.3.26}
\]

\[
l^U(\epsilon, a) = (1 - \alpha) \frac{\sigma}{\mu w_t} B_t \left[ \frac{\epsilon \kappa(a)}{\tau} \right]^{(\sigma - 1)(1 - \alpha)} \tag{2.3.27}
\]
iii. Total revenue earned by the downstream firm from final goods sales is:

\[ r_t^D (\epsilon, a) = \sigma B_t \left[ \frac{\epsilon \kappa (a)}{\tau} \right]^b \]  

(2.3.28)

and under Assumption 7, total revenue earned by the upstream firm from intermediate sales is proportional to downstream firm revenue:

\[ r_t^U (\epsilon, a) = (b + 1 - \nu) B_t \left[ \frac{\epsilon \kappa (a)}{\tau} \right]^b \]  

(2.3.29)

**Free-entry condition**  
Substituting the free entry condition (2.3.12) and the constant surplus shares (2.3.19)-(2.3.20) into the Bellman equations characterizing the value of searching (2.3.10)-(2.3.11), we then obtain the following relation characterizing the mass of searching firms:

\[ \theta_t = \left(1 - \frac{\nu}{\nu} \right) \gamma^D (d_t) \]  

(2.3.30)

Under part (iii) of assumption 6, we can rewrite this as an equation in market tightness alone:

\[ \theta_t \gamma (1, \theta_t) = \frac{1 - \nu}{\nu} \]  

(2.3.31)

Since the left-hand side of (2.3.31) is strictly increasing in \( \theta_t \) from zero to \(+\infty\), this implies that there is a unique solution \( \theta_t \) that is constant across time and that is decreasing in the relative bargaining power of downstream firms. I henceforth denote this solution by \( \bar{\theta} \).

Consequently, the expected surplus of a newly-formed relationship is given by:

\[ \int_{\xi}^{\bar{\xi}} \max \{ S_{t+1} (\epsilon', 0), 0 \} \bar{F} (d\epsilon') = \frac{\gamma^D (d_t)}{\nu \bar{m}} \]  

(2.3.32)

where \( \bar{m} \equiv m (1, \bar{\theta}) \). Note that if the expected surplus of a newly-formed relationship is higher, the mass of searching firms must also be higher.
**Relationship creation, and destruction**  With free entry, constant surplus shares, and flow payoffs given by (2.3.21) and (2.3.22), equations (2.3.8)-(2.3.9) characterizing the value of an active relationship can be combined to give a single Bellman equation characterizing the evolution of the relationship surplus:

$$S_t(\epsilon, a) = B_t \left[ \frac{\epsilon \kappa(a)}{\tau} \right]^b - f + (1 - \delta) \int_\xi^\bar{\epsilon} \max \{ S_{t+1}(\epsilon', a + 1), 0 \} Q(\epsilon') \, d\epsilon' \quad (2.3.33)$$

Given the properties of the Markov efficiency process in Assumption 3, the surplus of a relationship in equilibrium will be strictly increasing in its efficiency $\epsilon$. The endogenous destruction of trading relationships is therefore characterized by age-dependent efficiency cutoffs $\epsilon^*_t(a)$, such that a relationship of age $a$ is willingly terminated at date $t$ if and only if the relationship receives a draw of efficiency less than $\epsilon^*_t(a)$.

If $S_t(\xi, a) < 0 < S_t(\bar{\epsilon}, a)$, then there exists a set of efficiency values for which the surplus of a relationship of age $a$ is negative, and a set of values for which the surplus is positive. In this case, the efficiency cutoff at age $a$ is defined implicitly by:

$$S_t [\epsilon^*_t(a), a] = 0 \quad (2.3.34)$$

Endogenous destruction of the relationship then occurs with some probability strictly between zero and one, and the efficiency cutoff $\epsilon^*_t(a)$ lies in the interior of the efficiency distribution support. However, if $S_t(\xi, a) \geq 0$, then any efficiency draw is good enough for the relationship to survive, so that $\epsilon^*_t(a) = \xi$. In this case, destruction of the relationship occurs only if the exogenous termination shock is received. Conversely, if $S_t(\bar{\epsilon}, a) \leq 0$, then no possible efficiency draw is good enough to ensure relationship survival, so that $\epsilon^*_t(a) = \bar{\epsilon}$ and all relationships terminate. Also, since $\kappa(a) = \bar{\kappa}$ for all $a \geq \bar{a}$, then evidently $\epsilon^*_t(a) = \epsilon^*_t(\bar{a})$ for all $a \geq \bar{a}$.

Now, note that the joint distribution of efficiency and age amongst active relationships
at date \( t \) can be expressed as:

\[
F_t(\epsilon, a) = \begin{cases} 
\frac{F_{\text{draw}}(\epsilon, a) - F_{\text{draw}}(\epsilon^*_t(a), a)}{1 - F_{\text{draw}}(\epsilon^*_t(a), a)} & \text{, if } \epsilon \in [\epsilon^*_t(a), \epsilon] \\
0 & \text{, if } \epsilon \notin [\epsilon^*_t(a), \epsilon]
\end{cases} \tag{2.3.35}
\]

where \( F_{\text{draw}}(\cdot, a) \) is the distribution of efficiency draws received by relationships of age \( a \) (before any exit occurs), given recursively by:

\[
F_{\text{draw}}^t(\epsilon, a) = \int_{\epsilon^*_t(a-1)}^{\epsilon^*_t(a)} Q(\epsilon|\mathbf{x}) F_{t-1}(\epsilon^*_t(a-1), a-1) \, d\epsilon
\tag{2.3.36}
\]

Therefore the law of motion for the distribution of efficiency draws received is:

\[
F_{\text{draw}}^t(\epsilon, a) = \frac{\int_{\epsilon^*_t(a-1)}^{\epsilon^*_t(a)} Q(\epsilon|\mathbf{x}) F_{\text{draw}}^{t-1}(\epsilon, a-1) \, d\epsilon}{1 - F_{\text{draw}}^{t-1}[\epsilon^*_t(a-1), a-1]} \tag{2.3.37}
\]

with initial condition \( F_{\text{draw}}^t(\cdot, 0) = \tilde{F}(\cdot) \) for all \( t \). The fraction of relationships that terminate at date \( t \) that were of age \( a - 1 \) in the previous period can thus be expressed as:

\[
\Delta_t^{\text{out}}(a) = F_{\text{draw}}^t[\epsilon^*_t(a), a] \tag{2.3.38}
\]

Equation (2.3.38) highlights how the unconditional hazard of relationship termination depends on the ages of relationships in two ways. First, because of growth in relationship capital, the efficiency cutoffs \( \epsilon^*_t \) are age-dependent. Second, the distribution of efficiency draws given by (2.3.37) also potentially varies with the age of a cohort of relationships because of the Markov efficiency process and selective survival of relationships at each date. Both of these features of the model could therefore in principle independently replicate the negative association between relationship termination rates and age shown in Figure 19.

However, note that the probability that a relationship survives up to age \( a - 1 \) and terminates at age \( a \) and date \( t \) conditional on lagged efficiency equal to \( \epsilon_{a-1} \) is simply given
by:

\[ \tilde{\Delta}_{t}^\text{out} (a|\epsilon_{a-1}) = Q[\epsilon_{t}^* (a) |\epsilon_{a-1}] \] (2.3.39)

Without growth in relationship capital, the efficiency cutoffs are independent of age, and equation (2.3.39) would therefore predict that once lagged efficiency is controlled for, the hazard rate of termination should not vary with age. Since there is a one-to-one mapping between efficiency and relationship size in the model, this implies that the hazard rate of relationship termination should not vary with age once the size of the relationship is controlled for. Conversely, with serially independent efficiency draws, the transition function Q does not depend on lagged efficiency, and therefore (2.3.39) would predict that once age is controlled for, firm size should no longer matter for the termination hazard rate. Both these predictions are contrary to the regression estimates documented in Table 2, which motivates the inclusion of both features in the model.

Of course, controlling for current relationship size in a model without relationship capital growth is no longer sufficient to eliminate the effect of age if the efficiency process follows a higher-order Markov process. In this case, however, including controls for relationship size with a number of lags equal to the order of the Markov process would again generate the prediction that age no longer matters. The results of Table 2, however, are robust to adding controls for buyer and seller size with at least 3 lags, suggesting that a higher-order Markov efficiency process is insufficient to explain the joint effects of age and firm size.

**Labor market clearing** In order to close the model, we need to keep track of relationships of different ages and cohorts. Let \( n_t (a) \) denote the mass of active relationships of age \( a \) at date \( t \). Accounting for both endogenous and exogenous relationship termination, the relationship age distribution evolves according to the following law of motion:

\[ n_t (a+1) = n_{t-1} (a) [1 - \Delta_{t}^\text{out} (a+1)] (1 - \delta), \forall a \geq 0 \] (2.3.40)
while the mass of newly-formed relationships is given by:

\[ n_t(0) = m(d_{t-1}, u_{t-1}) \left[ 1 - \Delta^{out}_t(0) \right] \]  

(2.3.41)

The mass of relationships with state \((\epsilon, a)\) is then given by:

\[ \tilde{n}_t(\epsilon, a) = n_t(a) f_t(\epsilon, a) \]  

(2.3.42)

where \(f_t\) is the density function implied by \(F_t\). Note that the initial conditions of the model are the initial relationship age distribution \(\{n_{-1}(a)\}_{a=0}^{\infty}\) and masses of downstream and upstream searching firms \(\{d_{-1}, u_{-1}\}\).

Finally, labor market clearing at date \(t\) requires:

\[ L = d_t \gamma^D(d_t) + u_t \gamma^U(u_t) + \sum_{a=0}^{\infty} n_t(a) \left[ \bar{\ell}_t(a) + f \right] \]  

(2.3.43)

where the average quantity of labor used for production in relationships of age \(a\) at date \(t\) is given by the sum of (2.3.26) and (2.3.27):

\[ \bar{\ell}_t(a) = \frac{\sigma}{\mu 
\bar{\ell}_t(a)} B_t \left[ \frac{\bar{\ell}_t(a) \kappa(a)}{\tau} \right]^b \]  

(2.3.44)

and \(\bar{\ell}_t(a)\) denotes the \(L^b\)-norm of efficiency of relationships of age \(a\):

\[ \bar{\ell}_t(a) = \left[ \int_{\epsilon_{t}^*(a)}^{\infty} e^b F_t(d\epsilon, a) \right]^{\frac{1}{b}} \]  

(2.3.45)

Alternatively, the labor market clearing condition can also be expressed as:

\[ L = d_t \gamma^D(d_t) + u_t \gamma^U(u_t) + \frac{\sigma}{w_t \mu} B_t \left( \frac{Z_t}{\tau} \right)^b + f N_t \]  

(2.3.46)
where $Z_t$ is the aggregate productivity of the economy:

\[
Z_t \equiv \left( \sum_{a=0}^{\infty} n_t(a) \left[ \bar{\epsilon}_t(a) \kappa(a) \right]^b \right)^{\frac{1}{b}}
\]  
(2.3.47)

and $N_t$ denotes the total mass of active relationships:

\[
N_t = \sum_{a=0}^{\infty} n_t(a)
\]
(2.3.48)

### 2.3.2.2 Equilibrium definition

We can now define an equilibrium of the closed economy as follows.

**Definition 4.** Given an initial mass function $n_{-1}$ and initial masses of downstream and upstream searching firms $\{d_{-1}, u_{-1}\}$, an equilibrium of the closed economy is a list of sequences of surplus functions $\{S_t\}_{t=0}^{\infty}$, efficiency cutoff functions $\{\epsilon_t^*\}_{t=0}^{\infty}$, efficiency-age distributions $\{F_t, F_t^{\text{draw}}\}_{t=0}^{\infty}$, mass functions $\{n_t\}_{t=0}^{\infty}$, masses of downstream and upstream searching firms $\{d_t, u_t\}_{t=0}^{\infty}$, demand shifters $\{A_t\}_{t=0}^{\infty}$, and wages $\{w_t\}_{t=0}^{\infty}$, all of which satisfy equations (2.3.31), (2.3.32), (2.3.33), (2.3.34), (2.3.37), (2.3.40), (2.3.41), and (2.3.43).

We can also define a steady-state equilibrium as one in which aggregate variables remain constant, even though there is constant creation and destruction of individual supplier-producer relationships.

**Definition 5.** A steady-state equilibrium of the closed economy is a surplus function $S$, an efficiency cutoff function $\epsilon^*$, a set of efficiency-age distributions $\{F, F^{\text{draw}}\}$, a mass function $n$, masses of downstream and upstream searching firms $\{d, u\}$, a demand shifter $A$, and a wage $w$, all of which satisfy the time-invariant versions of equations (2.3.31), (2.3.32), (2.3.33), (2.3.34), (2.3.37), (2.3.40), (2.3.41), and (2.3.43).
2.3.2.3 Existence and uniqueness

Existence and uniqueness of the steady-state closed economy equilibrium are established by Proposition 5. The proof of this proposition in section B.1.2 of the appendix also shows how to solve for the steady-state equilibrium. Intuitively, the solution procedure is as follows. Given the mass of downstream searching firms $d$, the Bellman equation for the relationship surplus (2.3.33) and the free entry condition (2.3.32) uniquely determine the demand shifter $A$, while the mass of upstream searching firms $u$ is pinned down by (2.3.31). The mass of downstream searching firms must then be such that labor market clearing (2.3.43) is satisfied.

**Proposition 5.** There exists a unique steady-state equilibrium of the closed economy. The efficiency cutoffs in the steady-state equilibrium satisfy $\epsilon^*(a + 1) \leq \epsilon^*(a)$ for all $a \geq 0$, so that the unconditional probability of relationship destruction is weakly decreasing with the age of the relationship.

2.3.2.4 Welfare

Given the free entry of firms, aggregate profits in the economy are zero. Household per capita welfare is therefore equal to the real wage, which is given by:

$$w_t\frac{P_t}{\mu(T)} = 1 - \epsilon_t\left(\frac{Z_t}{\tau}\right)^{1-a}$$

(2.3.49)

The aggregate productivity of the economy $Z_t$ is hence a sufficient statistic for welfare. Using equations (2.3.40), (2.3.41), and (2.3.47), we can express aggregate productivity as:

$$Z_t^b = \bar{m}_d_t \bar{Z}_t^b$$

(2.3.50)

where $\bar{Z}_t^b$ is a measure of average productivity:

$$\bar{Z}_t^b = \sum_{a=0}^{\infty} \phi_t(a) [\epsilon_t(a) \kappa(a)]^b$$

(2.3.51)
and $\phi(a)$ is the probability that a firm survives at least up to age $a$:

$$
\phi_t(a) = (1 - \delta)^a \prod_{s=0}^{a} [1 - \Delta_t^{out}(s)]
$$

Household welfare can then be written as:

$$
\log U_t = c - (1 - \alpha) \log \tau + \frac{1}{\sigma - 1} \log d_t + (1 - \alpha) \log \bar{Z}_t \tag{2.3.52}
$$

where $c \equiv -\log \mu + \frac{1}{\sigma - 1} \log \bar{m}$ is a constant. Equation (2.3.52) summarizes the three potential channels through which a rise in trade costs between suppliers and producers can affect welfare in the closed economy. First, it directly increases the costs of production, captured by the term $\log \tau$. The prices of goods become more expensive for the household which lowers welfare, with the strength of the effect decreasing in the share of value-added $\alpha$. I refer to this channel as the intensive margin effect. Second, changes in trade costs might lead to changes in entry, captured by the term $\log d$. If higher trade costs induce less entry, for example, welfare also falls. I refer to this as the entry effect. Finally, an increase in trade costs might also induce changes in average productivity, captured by the term $\bar{Z}$, through its effect on the endogenous decisions of firms in active relationships to exit or stay. I refer to this as the selection effect.

In the closed economy, however, only the intensive margin effect operates, with both the entry and selection effects absent. To see why this is so, first note that the free entry of firms implies zero aggregate profits. Therefore, with the wage as numeraire, the household has a fixed amount of expenditure to allocate across products in the economy. When $\tau$ increases, production costs for all relationships increase proportionally, and therefore no single relationship gains an advantage over others. Consequently, revenue earned by each relationship remains the same and the efficiency cutoffs are unchanged. This implies that average productivity $\bar{Z}$ does not change, so that the selection effect does not operate. Since entry is determined by expected future profits, this also implies that $d$ does not change, and
the entry effect is absent as well.

**Proposition 6.** In the steady-state of the closed economy, household welfare varies with \( \tau^{-(1-\alpha)} \). The masses of searching firms and active relationships as well as the distributions of efficiency amongst active relationships are independent of the trade cost.

An immediate corollary of Proposition 6 is that the elasticity of welfare with respect to trade costs is not affected by the rate of relationship capital growth, the properties of the Markov efficiency process, or the extent of search frictions in the economy. In the open economy, however, this will no longer be true, as changes in trade costs will affect relative production costs of relationships across different locations, and will induce changes in welfare through all three channels discussed above.

### 2.4 Open Economy Model

I now consider multiple locations indexed by \( i, j \in J \). The representative household in location \( i \) supplies \( L_i \) units of labor and has the same preferences as in the closed economy over goods produced by domestic downstream firms.\(^{49}\) Trade between upstream suppliers and downstream producers in different locations is subject to the same search and matching frictions, efficiency dynamics, and relationship capital growth as before. Search is assumed to be directed in the sense that firms can choose the location in which to look for trading partners, but undirected with respect to partners within a location.\(^{50}\) At date \( t \), there are then \( d_{ij,t} \) downstream firms in \( i \) searching for upstream suppliers in \( j \) and \( u_{ij,t} \) upstream suppliers searching for downstream customers in \( j \). I henceforth refer to an \( ij \)-relationship as one involving a downstream producer in \( i \) and an upstream supplier in \( j \).

Now let \( \tau_{ij} \) and \( f_{ij} \) denote the iceberg trade cost and operating cost respectively for \( ij \)-relationships. Also, let \( \gamma_{ij}^D \) and \( \gamma_{ij}^U \) denote the search cost functions for downstream firms in

\(^{49}\)Downstream firms therefore produce non-tradable goods, while upstream firms produce tradable goods.

\(^{50}\)Since all firms are ex-ante identical, this is equivalent to allowing directed search over partners, but assuming that a firm cannot observe the mass of other firms that are searching for its targeted partner (and therefore cannot condition its choice of targeted partner on this information).
and upstream rms in $j$ respectively that are searching for $ij$-relationships. It is assumed that the search costs are paid in units of labor in the location of the searching firm, and that the search cost functions for $ij$-relationships satisfy the same properties as in assumption 6.

The following assumption summarizes which model parameters are allowed to vary across locations and location pairs, and which are not.

**Assumption 8.** Trade costs $\{\tau_{ij}\}_{i,j \in J}$, operating costs $\{f_{ij}\}_{i,j \in J}$, and search cost functions $\{\gamma^D_{ij}, \gamma^U_{ij}\}_{i,j \in J}$ potentially vary across location pairs, and labor endowments $\{L_i\}_{i \in J}$ potentially vary across locations. The initial relationship efficiency distribution $\bar{F}$, Markov process for efficiency, and relationship capital function $\kappa$ are constant across location pairs, and household preferences are constant across locations.

### 2.4.1 Equilibrium

#### 2.4.1.1 Equilibrium conditions

**Generalized Nash bargaining solution** The solution to the Nash bargaining problem within each relationship in the open economy remains characterized by Proposition 4 (with the appropriate location-pair specific parameters).

**Free-entry condition** The surplus of an $ij$-relationship now satisfies the following Bellman equation:

$$S_{ij,t}(\epsilon, a) = B_{i,t} \left[ \frac{\epsilon \kappa(a)}{\omega_{ij,t} \tau_{ij}} \right]^b - f_{ij} + (1 - \delta) \int_{\tilde{\epsilon}}^{\epsilon} \max \{S_{ij,t+1}(\epsilon', a + 1), 0\} Q(\epsilon'|\epsilon)$$

(2.4.1)

where $\omega_{ij,t} \equiv \frac{w_{ij,t}}{w_{i,t}}$ is the relative wage between locations $i$ and $j$. The free entry of downstream firms in $i$ and upstream firms in $j$ then requires:

$$\int_{\tilde{\epsilon}}^{\infty} \max \{S_{ij,t+1}(\epsilon', 0), 0\} \tilde{F}(d\epsilon') = \frac{\gamma^D_{ij}(d_{ij,t})}{\nu \bar{m}_{ij}}$$

(2.4.2)
where $\bar{m}_{ij} \equiv m \left(1, \bar{\theta}_{ij}\right)$. Market tightness for $ij$-relationships is given by the unique solution to:

$$\bar{\theta}_{ij} \bar{\gamma}_{ij} \left(1, \bar{\theta}_{ij}\right) = \frac{1 - \nu}{\nu}$$ (2.4.3)

where as before, $\bar{\gamma}_{ij} \left(d, u\right) \equiv \frac{\gamma_{ij}^{U}(u)}{\gamma_{ij}^{D}(d)}$.

**Relationship creation and destruction** As in the closed economy, $ij$-relationships of age $a$ endogenously terminate at date $t$ if the efficiency draw received is less than a cutoff value $\epsilon_{ij,t}^{*}(a)$, which is defined implicitly by:

$$S_{ij,t} \left[\epsilon_{ij,t}^{*}(a), a\right] = 0$$ (2.4.4)

if $S_{ij,t}(\epsilon, a) < 0 < S_{ij,t}(\bar{\epsilon}, a)$, and is given by $\epsilon_{ij,t}^{*}(a) = \epsilon$ if $S_{ij,t}(\epsilon, a) \geq 0$ and $\epsilon_{ij,t}^{*}(a) = \bar{\epsilon}$ if $S_{ij,t}(\bar{\epsilon}, a) \leq 0$. The distribution of efficiency draws received by $ij$-relationships of age $a$ at date $t$ then obeys the analog of equation (2.3.37):

$$F_{\text{draw}ij,t}(\epsilon, a) = F_{\text{draw}ij,t}\left[\epsilon_{ij,t}^{*}(a), a\right]$$ (2.4.5)

with initial condition $F_{\text{draw}ij,t}(\cdot, 0) = F(\cdot)$ for all $i, j \in J$ and all $t$. The unconditional hazard rate of termination is given by:

$$\Delta_{ij,t}^{out}(a) = F_{\text{draw}ij,t}\left[\epsilon_{ij,t}^{*}(a), a\right]$$ (2.4.6)

and the joint distribution of efficiency and age amongst active relationships at date $t$ is:

$$F_{ij,t}(\epsilon, a) = \begin{cases} \frac{F_{\text{draw}ij,t}(\epsilon, a) - F_{\text{draw}ij,t}^{\epsilon_{ij,t}^{*}(a), a}}{1 - F_{\text{draw}ij,t}^{\epsilon_{ij,t}^{*}(a), a}}, & \text{ if } \epsilon \in \left[\epsilon_{ij,t}^{*}(a), \bar{\epsilon}\right] \\ 0, & \text{ if } \epsilon \notin \left[\epsilon_{ij,t}^{*}(a), \bar{\epsilon}\right] \end{cases}$$ (2.4.7)

The mass of active $ij$-relationships of age $a$, $n_{ij,t}(a)$, now evolves according to:
\[ n_{ij,t+1} (a + 1) = n_{ij,t} (a) \left[ 1 - \Delta_{ij,t+1}^{\text{out}} (a + 1) \right] (1 - \delta), \forall a \geq 0 \] (2.4.8)

and the mass of newly-formed \( ij \)-relationships is given by:

\[ n_{ij,t} (0) = m (d_{ij,t-1}, u_{ij,t-1}) \left[ 1 - \Delta_{ij,t}^{\text{out}} (0) \right] \] (2.4.9)

The initial conditions of the open economy model are now \( \{n_{ij,-1} (a)\}_{a=0}^\infty \) and \( \{d_{ij,-1}, u_{ij,-1}\} \) for all \( i, j \in J \).

**Labor market clearing and trade balance** To close the model, we need to consider both labor market clearing and the balance of trade. Labor market clearing in location \( i \) requires:

\[
L_i = \sum_{j \in J} d_{ij,t} \gamma_1^D (d_{ij,t}) + \sum_{j \in J} u_{ji,t} \gamma_1^U (u_{ji,t}) \\
+ \sum_{j \in J} \sum_{a=0}^\infty n_{ij,t} (a) \left[ \bar{l}_j^{D,t} (a) + \nu f \right] + \sum_{j \in J} \sum_{a=0}^\infty n_{ji,t} (a) \left[ \bar{l}_j^{S,t} (a) + (1 - \nu) f \right]
\] (2.4.10)

The first and second terms on the right-hand side of (2.4.10) are the quantities of labor used by searching downstream and upstream firms in \( i \) respectively, while the third and fourth terms are labor used for production (including shares of the operating cost) by downstream and upstream firms in \( i \) respectively. The average quantity of labor used for production in \( ij \)-relationships of age \( a \) can be expressed as the sum of the open economy equivalents of (2.3.26) and (2.3.27):

\[
\bar{l}_{ij,t} (a) = \mu^{-\sigma} A_{i,t} w_{i,t}^{-\sigma} \left[ \bar{e}_{ij,t} (a) \kappa (a) \right]^{b} \] (2.4.11)
where $\bar{\epsilon}_{ij,t}(a)$ denotes the $L^b$-norm of efficiency of $ij$-relationships of age $a$:

$$
\bar{\epsilon}_{ij,t}(a) = \left[ \int_{\epsilon_{ij,t}^{\min}(a)}^{\epsilon_{ij,t}^{\max}(a)} e^{b \epsilon_{ij,t} (de, a)} \right]^{\frac{1}{b}}
$$

(2.4.12)

The average quantities of labor used by downstream and upstream producers in $ij$-relationships are then given respectively by:

$$
\bar{\bar{l}}_{D,ij,t}(a) = \alpha \bar{l}_{ij,t}(a)
$$

(2.4.13)

$$
\bar{\bar{l}}_{U,ij,t}(a) = (1 - \alpha) \bar{l}_{ij,t}(a)
$$

(2.4.14)

As in the closed economy, we can rewrite the labor market clearing condition as:

$$
L_t = \sum_{j \in J} d_{ij,t} \gamma_{ij}^D (d_{ij,t}) + \sum_{j \in J} u_{ji,t} \gamma_{ji}^U (u_{ji,t})
$$

$$
+ \sum_{j \in J} \left[ \alpha \frac{\sigma}{w_{ij,t} \mu} B_{i,t} \left( \frac{Z_{ij,t}}{\omega_{ij,t} \tau_{ij}} \right)^b + \nu f N_{ij,t} \right]
$$

$$
+ \sum_{j \in J} \left[ (1 - \alpha) \frac{\sigma}{w_{ji,t} \mu} B_{j,t} \left( \frac{Z_{ji,t}}{\omega_{ji,t} \tau_{ji}} \right)^b + (1 - \nu) f N_{ji,t} \right]
$$

(2.4.15)

(2.4.16)

where the aggregate productivity of $ij$-relationships is defined as:

$$
Z_{ij,t} \equiv \left[ \sum_{a=0}^{\infty} n_{ij,t}(a) \left[ \bar{\epsilon}_{ij,t}(a) \kappa(a) \right]^b \right]^{\frac{1}{b}}
$$

(2.4.17)

and the total mass of $ij$-relationships is given by:

$$
N_{ij,t} = \sum_{a=0}^{\infty} n_{ij,t}(a)
$$

(2.4.18)

Finally, the nominal value of imports in $ij$-relationships of state $(\epsilon, a)$ is:

$$
M_{ij,t}(\epsilon, a) = (b + 1 - \nu) B_{i,t} \left[ \frac{e^{\epsilon}(a)}{\omega_{ij,t} \tau_{ij}} \right]^b
$$

(2.4.19)
which is the sum of variable revenue earned by the upstream firm and the lump-sum transfer.

The aggregate value of imports by location \(i\) from location \(j\) can therefore be written as:

\[
\bar{M}_{ij,t} = (b + 1 - \nu) B_{i,t} \left[ \frac{Z_{ij,t}}{\omega_{ij,t} T_{ij}} \right]^b
\] (2.4.20)

The trade deficit in location \(i\) is then given by:

\[
D_{i,t} = \sum_{j \in J} \bar{M}_{ij,t} - \sum_{j \in J} \bar{M}_{ji,t}
\] (2.4.21)

and trade is assumed to be balanced so that \(D_{i,t} = 0\) for all \(i \in J\) and \(t \geq 0\).

### 2.4.1.2 Equilibrium definition

We can now define an equilibrium of the open economy as follows.

**Definition 6.** Given an initial mass function \(n_{ij,-1}\) and initial masses of downstream and upstream searching firms \(\{d_{ij,-1}, u_{ij,-1}\}\) for all \(i,j \in J\), an equilibrium of the open economy is a list of sequences for each \(i,j \in J\) of surplus functions \(\{S_{ij,t}\}_{t=0}^{\infty}\), mass functions \(\{n_{ij,t}\}_{t=0}^{\infty}\), efficiency cutoff functions \(\{\epsilon_{ij,t}^*\}_{t=0}^{\infty}\), efficiency-age distributions \(\{F_{ij,t}, F_{ij,t}^{draw}\}_{t=0}^{\infty}\), masses of downstream and upstream searching firms \(\{d_{ij,t}, u_{ij,t}\}_{t=0}^{\infty}\), demand shifters \(\{A_{i,t}\}_{t=0}^{\infty}\), and wages \(\{w_{i,t}\}_{t=0}^{\infty}\) all of which satisfy equations (2.4.1), (2.4.2), (2.4.3), (2.4.4), (2.4.5), (2.4.8), (2.4.9), (2.4.10), and (2.4.21).

We can also define a steady-state equilibrium as one in which aggregate variables remain constant, even though there is constant creation and destruction of individual supplier-producer relationships across all locations.

**Definition 7.** A steady-state equilibrium of the open economy is a list of surplus functions \(\{S_{ij}\}_{i,j \in J}\), efficiency cutoff functions \(\{\epsilon_{ij}^*\}_{i,j \in J}\), efficiency-age distributions \(\{F_{ij}, F_{ij}^{draw}\}_{i,j \in J}\), mass functions \(\{n_{ij}\}_{i,j \in J}\), masses of downstream and upstream searching firms \(\{d_{ij}, u_{ij}\}_{i,j \in J}\),
demand shifters \{A_i\}_{i \in J} and wages \{w_i\}_{i \in J}, all of which satisfy the time-invariant versions of equations (2.4.1), (2.4.2), (2.4.3), (2.4.4), (2.4.5), (2.4.8), (2.4.9), (2.4.10), and (2.4.21).

2.4.1.3 Existence and uniqueness

In what follows, I focus on a version of the open economy model with symmetric locations, for which existence and uniqueness of the steady-state equilibrium are established by proposition 7.\textsuperscript{51} Symmetry across locations guarantees trade balance and therefore avoids having to solve for relative wages, although numerical solution of the model with heterogeneous locations is straightforward and points toward uniqueness of the equilibrium in general.

**Proposition 7.** There exists a unique steady-state equilibrium of the open economy with symmetric countries. The efficiency cutoffs in the steady-state equilibrium satisfy \(\epsilon^*_{ij}(a + 1) \leq \epsilon^*_{ij}(a)\) for all \(a \geq 0\) and \(i, j \in J\), so that the probability of relationship destruction is weakly decreasing with the age of the relationship.

2.4.1.4 Welfare

As in the closed economy, household welfare in each location is equal to the real wage, which is now given by:

\[
\frac{w_{i,t}}{P_{i,t}} = \frac{1}{\mu} \left[ \sum_{j \in J} \left( \frac{Z_{ij,t}}{\omega_{ij,t} \tau_{ij}} \right)^b \right]^{1 \over \sigma - 1} \tag{2.4.22}
\]

Furthermore, note from (2.4.20) that the home trade share of location \(i\) is given by:

\[
H_{i,t} \equiv \frac{\bar{M}_{i,t}}{\sum_{j \in J} M_{ij,t}} = \frac{\left( \frac{Z_{ii,t}}{\tau_{ii}} \right)^b}{\sum_{j \in J} \left( \frac{Z_{ij,t}}{\omega_{ij,t} \tau_{ij}} \right)^b} \tag{2.4.23}
\]

\[
= \frac{\left( \frac{Z_{ii,t}}{\tau_{ii}} \right)^b}{\sum_{j \in J} \left( \frac{Z_{ij,t}}{\omega_{ij,t} \tau_{ij}} \right)^b} \tag{2.4.24}
\]

\textsuperscript{51}The proof of this proposition is almost identical to the proof of proposition 5, and is therefore omitted from the appendix.
Analogous to equation (2.3.52), we can therefore express welfare in location $i$ as:

$$
\log U_{i,t} = c_i - (1 - \alpha) \log \tau_{ii} + \frac{1}{\sigma - 1} \log d_{ii,t}
$$

$$
+ (1 - \alpha) \log \bar{Z}_{ii,t} - \frac{1}{\sigma - 1} \log H_{i,t}
$$

(2.4.25)

where $c_i \equiv -\log \mu + \frac{1}{\sigma} \log \bar{m}_{ii}$ is a constant. Average productivity of $ij$-relationships is defined as:

$$
\bar{Z}_{ij,t}^b \equiv \sum_{a=0}^{\infty} \phi_{ij,t} (a) [\bar{\epsilon}_{ij,t} (a) \kappa (a)]^b
$$

(2.4.26)

where the probability that an $ij$-relationship survives at least up to age $a$ is:

$$
\phi_{ij,t} (a) = (1 - \delta)^a \prod_{s=0}^{a} [1 - \Delta_{ij,t}^\text{out} (s)]
$$

(2.4.27)

Again, equation (2.4.25) summarizes the potential channels through which changes in trade costs may affect welfare. As in the closed economy, the terms $\log \tau_{ii}$, $\log d_{ii,t}$, and $\log \bar{Z}_{ii,t}$ capture the intensive margin, entry, and selection effects. In the open economy, however, there is now a fourth channel captured by the term $\log H_{i,t}$, which I refer to as the trade effect. This channel reflects the fact that changes in trade costs may affect not only domestic prices, entry, and selection, but may also induce changes in the distribution of trade across locations. As highlighted by Arkolakis et al (2012), the home trade share plays an important role in many international trade models in determining the gains from trade.

In contrast with the class of models studied by Arkolakis et al (2012), however, the home trade share is no longer a sufficient statistic for welfare in equation (2.4.25). Instead, in response to a change in foreign trade costs, the entry and selection effects also matter. In section 2.5.2, I simulate the model to study how the dynamic properties of the relationship productivity process matter for these effects.
2.5 Numerical Analysis

2.5.1 Model calibration

2.5.1.1 Functional form assumptions

In order to calibrate the model, I first choose parametric forms for the various functions in the model. First, the relationship capital function is assumed to be given by:

\[
\kappa(a) = \kappa \left[ 1 + \left( \frac{a}{\bar{a}} \right)^\psi \left( \frac{\bar{\kappa}}{\kappa} - 1 \right) \right]
\] (2.5.1)

This function is chosen to ensure that the stock of relationship capital grows from \( \kappa \) to \( \bar{\kappa} \) as relationships grow from age 0 to age \( \bar{a} \), with a single parameter \( \psi \) governing the rate at which relationship-specific productivity grows with the age of the relationship. Note that as \( \psi \to 0 \), relationship capital is equal to \( \kappa \) for ages \( a < \bar{a} \) and is equal to \( \bar{\kappa} \) for ages \( a \geq \bar{a} \). Conversely, as \( \psi \to \infty \), relationship capital jumps to \( \bar{\kappa} \) instantaneously at age 0.

Second, the Markov process for relationship efficiency is assumed to be a discrete approximation to the following first-order auto-regressive process:

\[
\log e_{t+1} = \rho \log e_t + \sigma_e \log e_t
\] (2.5.2)

where \( \log e_t \) is a standard normal random variable that is independent across time and relationships. The parameter \( \rho \) governs the persistence of the efficiency process, while \( \sigma_e \) measures the dispersion of efficiency fluctuations. I then employ the method developed in Tauchen (1986) to discretize the state space for log efficiency over an equally-spaced grid on the interval \([-\sigma_e, \sigma_e]\), which also ensures that it is bounded.
Third, the search cost functions are assumed to have a constant and common elasticity:

\[ \gamma^D_{ij}(d) = \Gamma^D_{ij} d^\lambda \]  
(2.5.3)

\[ \gamma^U_{ij}(u) = \Gamma^U_{ij} u^\lambda \]  
(2.5.4)

where \( \Gamma^D_{ij} \) and \( \Gamma^U_{ij} \) are constants determining the overall level of search costs, and where \( \lambda \) governs the extent of congestion in the search process. Finally, the matching function is chosen such that it satisfies the properties in assumption 5:

\[ m(d, u) = d^{\nu_m} u^{1-\nu_m} (1 - e^{-u/d}) (1 - e^{-d/u}) \]

where \( \nu_m \in (0,1) \) measures the relative importance of downstream versus upstream firms in generating matches.\(^{52}\)

2.5.1.2 Parameter values

With these parametric assumptions, the model has 13 parameters \( \sigma, \beta, \alpha, \nu, \nu_m, \kappa, \bar{\kappa}, \bar{a}, \delta, \lambda, \rho, \sigma_e, \) and \( \psi \) that do not vary by location, 1 set of parameters \( L_i \) that varies by location, and 5 sets of parameters \( f_{ij}, \Gamma^D_{ij}, \Gamma^U_{ij}, \tau_{ij}, D_{ij} \) that vary by location pair. Parameter values used for a baseline calibration of an open economy with symmetric locations are shown in Table 4. Given the unavailability of detailed data on trading relationships across multiple countries, I calibrate the key parameters of the model to match the cross-sectional and dynamic properties of relationships amongst US firms in the Compustat data, where the underlying assumption is that the properties of relationship-specific productivity fluctuations are common for relationships regardless of the location of the buyer and seller. I now discuss the calibration of parameters in detail.

---

\(^{52}\)This matching function is similar to the Cobb-Douglas matching functions typically used in continuous time models of job search. The addition of the exponential terms are necessary to ensure that the mass of matches that form is always less than the mass of firms searching, so that the probability of receiving a match remains bounded away from 1, thus allowing us to avoid dealing with corner solutions.
First, the parameters \( \{\sigma, \beta, \alpha\} \) have been extensively studied elsewhere, and I choose values that are consistent with the relevant literature. Second, assuming that downstream and upstream firms are symmetric in the bargaining and matching processes implies \( \nu = \nu_m = \frac{1}{2} \). Third, I normalize \( \kappa = 1 \) and choose \( \bar{\kappa} \) to match the ratio of average relationship size (buyer plus seller size) for old relationships versus young relationships. The maximum age \( \bar{a} \) is chosen to be a large number. Fourth, the parameters of the efficiency distribution \( \{\rho, \sigma_e\} \) are chosen to match the serial correlation and variance of relationship sizes respectively.

Fifth, the fixed operating costs \( f_{ij} \) determine the average hazard rate of endogenous relationship termination, while \( \delta \) controls the exogenous termination rate. Since the relationship termination rate appears to plateau for large age values (see Figure 19), I choose \( \delta \) to match the exit rate for old domestic relationships and set \( f_{ii} \) to match the overall exit rate for domestic relationships. The parameter \( \psi \) is then chosen to match the curvature of the relationship age-hazard rate profile.

Sixth, the parameter \( \lambda \) controls the extent of congestion in the search process, and determines the relative mass of firms that search in foreign locations versus those that search at home.\(^53\) With data on the extensive margin of trade at the firm-to-firm level (e.g. from customs data), \( \lambda \) could then be chosen to match this moment. Here, I simply set \( \lambda = 1 \) as a benchmark.

Seventh, I set domestic trade costs to be one, so that \( \tau_{ii} = 1 \). Since the parameters \( \{f_{ij}, \Gamma_{ij}^D, \Gamma_{ji}^U, \tau_{ij}\} \) all jointly determine trade flows from location \( j \) to \( i \), separately identifying these parameters (fixed costs versus search costs versus variable trade costs) is not a trivial task. Without additional data, I set \( f_{ij} = f_{ii}, \Gamma_{ij}^D = \Gamma_{ii}^D, \) and \( \Gamma_{ij}^U = \Gamma_{jj}^U \), so that foreign operating and search costs are the same as those at home.\(^54\) Furthermore, assuming that downstream and upstream firms face symmetric search costs implies \( \Gamma_{ij}^D = \Gamma_{ij}^U \). The value of \( \Gamma_{ii}^D \) can then be chosen to match the mass of active firms in the economy, but here I simply

\(^53\)As \( \lambda \to 0 \), there is no congestion in search, and in a symmetric open economy with higher trade costs with foreign locations than at home, no firms would search abroad.

\(^54\)In a more realistic calibration, the level of search costs may be approximated by measures of language differences, ethnic composition, past colonial history, and other similar variables.
The foreign trade cost $\tau_{ij}$ is then allowed to vary and is studied extensively in the comparative static exercises discussed in section 2.5.2.

Finally, labor supply $L_i$ is normalized to 1 for all $i \in J$.

### 2.5.1.3 Age, size, and termination rates

Figure 21 shows the model’s predicted hazard rates of relationship termination conditional on both relationship size and age (equation (2.3.39)). Since relationship efficiency is persistent, firms within a given age cohort, relationships are larger are also more likely to have higher efficiency in the future because of persistence in the efficiency process, and therefore the hazard rate declines with size. Similarly, amongst relationships of the same size, relationships that are older have higher stocks of accumulated relationship capital, and hence are less likely to terminate. Both of these features are consistent with the empirical patterns documented in section 2.2.

![Figure 21: Predicted termination hazard rates conditional on relationship size and age](image)

### 2.5.2 Comparative static exercises

I now simulate the model to study how the dynamics of relationship productivity matter for the gains from trade. In particular, I examine how the various welfare effects
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Rationale/Moment Matched</th>
</tr>
</thead>
<tbody>
<tr>
<td>elasticity of substitution</td>
<td>$\sigma$ 5</td>
<td>literature</td>
</tr>
<tr>
<td>discount factor</td>
<td>$\beta$ .95</td>
<td>literature</td>
</tr>
<tr>
<td>value-added share</td>
<td>$\alpha$ .7</td>
<td>literature</td>
</tr>
<tr>
<td>downstream firm bargaining power</td>
<td>$\nu$ .5</td>
<td>symmetry between downstream/upstream firms</td>
</tr>
<tr>
<td>downstream firm matching weight</td>
<td>$\nu_m$ .5</td>
<td>symmetry between downstream/upstream firms</td>
</tr>
<tr>
<td>minimum relationship capital</td>
<td>$\kappa$ 1</td>
<td>normalization</td>
</tr>
<tr>
<td>maximum relationship capital</td>
<td>$\overline{\kappa}$ 1.5</td>
<td>ratio of average firm size for firms in old versus young relationships</td>
</tr>
<tr>
<td>maximum age</td>
<td>$\overline{a}$ 100</td>
<td>large number</td>
</tr>
<tr>
<td>efficiency distribution persistence</td>
<td>$\rho$ .6</td>
<td>serial correlation of firm size</td>
</tr>
<tr>
<td>efficiency shock std. dev.</td>
<td>$\sigma_e$ .1</td>
<td>variance of firm size distribution</td>
</tr>
<tr>
<td>rate of relationship capital growth</td>
<td>$\psi$ .25</td>
<td>curvature of domestic relationship age-hazard rate profile</td>
</tr>
<tr>
<td>exogenous exit rate</td>
<td>$\delta$ .15</td>
<td>exit rate for old domestic relationships</td>
</tr>
<tr>
<td>operating costs</td>
<td>$f_{ij}$ 6.5</td>
<td>average exit rate for domestic relationships</td>
</tr>
<tr>
<td>search costs</td>
<td>$\Gamma_{ij}$ 1</td>
<td>normalization</td>
</tr>
<tr>
<td>trade costs</td>
<td>$\tau_{ij}$ varied</td>
<td>aggregate trade flows</td>
</tr>
<tr>
<td>labor supply</td>
<td>$L_i$ 1</td>
<td>normalization</td>
</tr>
<tr>
<td>trade deficits</td>
<td>$D_{ij}$ 0</td>
<td>balanced trade</td>
</tr>
</tbody>
</table>

Table 4: Parameter values for baseline calibration of symmetric open economy
discussed in sections 2.3.2.4 and 2.4.1.4 depend on two key parameters: $\psi$ (which controls the rate of relationship capital growth) and $\rho$ (which controls the persistence of relationship efficiency). Recall that the decomposition of welfare under consideration here is:

$$\log U_i = \text{const.} - (1 - \alpha) \log \tau_{ii} + \frac{1}{\sigma - 1} \log d_{ii}$$

$$+ (1 - \alpha) \log \bar{Z}_{ii} - \frac{1}{\sigma - 1} \log H_i$$

(2.5.5)

(2.5.6)

where the terms $\log \tau_{ii}$, $\log d_{ii}$, $\log \bar{Z}_{ii}$, and $\log H_i$ capture the intensive margin, entry, selection, and trade effects respectively.

Given the complexity of solving for the transition dynamics of the model, I focus here on comparisons across steady-states in an economy with two symmetric countries, and examine the responses of welfare to symmetric changes in trade costs with the foreign location. Since transition paths are neglected, this analysis offers a partial characterization of the interaction between relationship-specific productivity dynamics and the gains from trade. Nonetheless, it sheds light on the channels through which productivity dynamics matter for trade and welfare responses to trade cost shocks.

First consider Figure 22, which summarizes the role played by the growth rate of relationship capital in determining the gains from trade. The first graph in the figure shows the overall change in welfare as trade costs increase from a position of free trade, while the subsequent graphs show corresponding changes in entry, selection, and trade shares. From this analysis, we see that slower growth in relationship capital leads to stronger responses of welfare to trade costs (graph 1). This effect partially operates by dampening the response of entry (graph 2): a rise in foreign trade costs induces higher entry at home, but with slower growth in relationship capital, the marginal increase in the expected surplus of a relationship resulting from the change in trade costs is lower, and therefore the gain in entry is also less.

The stronger response of welfare to trade costs is also induced by a dampening of the selection effect (graph 3). A rise in foreign trade costs increases the survival probability of
domestic relationships and therefore increases domestic average productivity. When relationship capital grows more slowly, however, survival becomes less important in determining average productivity. Finally, slower relationship capital growth dampens the response of trade shares to trade costs (graph 4). In a model where the home trade share is a sufficient statistic for welfare, this would imply that slower relationship capital growth should lead to weaker responses of welfare to trade costs. Here, however, the opposite is true because the trade effect is offset by the entry and selection effects, due to the accumulation of relationship capital.

Next, consider Figure 23, which summarizes the role played by the persistence of relationship efficiency in determining the gains from trade. Again, the first graph in the figure shows the overall change in welfare as trade costs increase from a position of free trade, while the subsequent graphs show corresponding changes in entry, selection, and trade shares. Here, we see that a higher degree of persistence in relationship efficiency has qualitatively similar effects as slower productivity growth: it generates stronger responses of welfare to changes in trade costs (graph 1), by dampening both the entry effect (graph 2) and the selection effect (graph 3). Again, the trade effect is dampened by more persistent productivity shocks, but this is offset by changes in the entry and selection effects so that the overall response of welfare is stronger.

In sum, both persistence as well as age-dependence of the relationship productivity process matter for the gains from trade, as welfare is more sensitive to trade costs when productivity shocks are persistent and when productivity growth is slower.

2.6 Conclusion

This paper develops a general equilibrium model of international trade featuring search frictions between suppliers and producers in different locations, and uses the model to study how the dynamic properties of relationship-specific productivity fluctuations affect relation-
Figure 22: Relationship capital growth and welfare effects
Figure 23: Persistence in relationship efficiency and welfare effects
ship termination rates at the micro level and the gains from trade at the macro level. Persistence in relationship productivity and growth in productivity with age generate a hazard rate of relationship termination that is declining with both age and firm size, which is consistent with empirical evidence on the dynamics of firm-to-firm relationships in the US. Numerical simulations of the model then show how persistence and age-dependence of productivity fluctuations affect the responses of welfare and trade volumes to changes in trade costs. In particular, welfare responds more strongly to trade costs when relationship capital grows more slowly and productivity shocks are more persistent.

These findings highlight the importance of understanding what shapes productivity dynamics within firm-to-firm relationships. The extent to which such productivity fluctuations are the result of endogenous decisions by firms within a relationship, however, is an open question. If firms respond to shocks by changing their investments in relationship-specific capital, for example, the interaction between firm-to-firm level productivity dynamics and aggregate trade flows could be more complicated than the model assumes.

Furthermore, the implications of the theory for transition dynamics following aggregate shocks are left unanswered. As pointed out by Samuelson (1978), the gains from trade integrated along transition paths may differ from the gains in steady-state not only quantitatively but qualitatively as well. Solving for the transition dynamics of the model is therefore essential for a complete characterization of welfare. A plausible conjecture is that the process of relationship capital accumulation may imply asymmetric effects of shocks, as negative shocks induce immediate termination of trading relationships and can be very costly, while positive shocks generate only gradual gains in welfare as new relationships take time to form and grow. This potential mechanism is qualitatively similar to the asymmetric dynamics of job creation and destruction present in search models in the style of Mortensen and Pissarides (1994), and echoes an older literature on hysteresis in international trade discussed by Baldwin and Krugman (1989).

Numerical solution of the model outside of steady-state equilibria remains challenging,
however. The main complication is that convergence between steady-states is potentially characterized by a series of corner solutions in which there is no endogenous exit until the point of convergence, which occurs in finite time. With age-dependent productivity growth, this implies that the solution for endogenous relationship exit may be at corners for some ages while remaining interior for others, and having to guess which is which can be computationally demanding. A potential remedy for this might be to introduce aggregate shocks and characterize stochastic steady-states of the model in which the endogenous exit solution is always interior, although this remains scope for future research.
Chapter 3

Incentives and Inefficiencies: Intellectual Property Rights Protection and Technology Transfer

3.1 Introduction

In this paper, I study the interaction between intellectual property rights (IPR) protection by governments and cross-border technology transfer (TT) by multinational firms. The importance of the relationship between IPRs and TT has been recognized by policymakers worldwide - the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), for example, which is administered by the World Trade Organization (WTO), contains the following two articles in its General Provisions and Basic Principles:

Article 7:
The protection and enforcement of intellectual property rights should contribute to the promotion of technological innovation and to the transfer and dissemination of technology, to the mutual advantage of producers and users of technological knowledge...

Article 8:
Appropriate measures... may be needed to prevent the abuse of intellectual property rights by right holders or the resort to practices which unreasonably restrain trade or adversely affect the international transfer of technology.

While Article 7 implicitly acknowledges that IPR protection can be conducive to the flow of technology across borders, Article 8 also concedes the possibility that the mere existence
of a strong IPR regime may be insufficient to ensure the optimal flow of technology across borders. Thus far, however, the academic literature on IPRs and TT has largely studied each of these two issues in isolation, and therefore existing economic models are ill-equipped to afford a rigorous analysis of the interdependencies alluded to above.

Here, I offer a step towards filling this gap by developing a structural model of trade and foreign direct investment (FDI) that allows for endogenous choices of both IPR protection by governments and TT by multinational firms. In particular, I use the model to study two key ideas. First, in deciding on the optimal level of IPR protection to offer, the government of one country may fail to internalize the effects that its choice might have on welfare in other countries, via firms that engage in international trade and FDI. Second, the incentives for IPR protection by governments and for cross-border TT by multinational firms can be mutually dependent, even though the objectives of governments and multinational firms may not be perfectly aligned. Each of these ideas suggests the potential for a distinct kind of inefficiency to arise when parties behave non-cooperatively, and therefore also for welfare improvements to be achieved by cooperative policies regarding IPR protection and TT.

Formal modeling of these inefficiencies is also useful because it allows us first of all to characterize their qualitative properties - to determine, for example, whether there is a sense in which there is under- or over-provision of IPR protection or TT in the non-cooperative equilibrium, and how the incentives for IPR protection and TT are affected by other parameters describing the global economy. Furthermore, a structural model enables us to estimate the potential welfare losses for consumers that result from these inefficiencies. A brief summary of the key findings of the paper is as follows.

First, the model predicts that in the Nash equilibrium, both the strength of IPR protection and the degree of TT by multinational firms are increasing in the degree of firm heterogeneity and the elasticity of substitution across varieties consumed by the household. Second, the IPR- and TT-related inefficiencies are both always characterized by under-provision of IPR protection and TT by multinational firms respectively. Finally, in terms of
magnitudes (measured by changes in household welfare), the model suggests that the IPR and TT inefficiencies are comparable in size, amounting in combination to as much as 4% of the overall gains from openness to trade and FDI. These results therefore suggest that Pareto-improving outcomes can be achieved not only from international cooperation with regard to protection of intellectual property, but also with regard to the cross-border transfer of technology between firms.

The results of this paper build on two distinct but related strands of literature. First, with regards to intellectual property, the study of optimal IPR policy has long been an area of keen interest in the economics profession. Since at least the work of Nordhaus (1969), economists have been aware of and have sought to formally model the fundamental tradeoff of IPR protection: on the one hand, stronger IPR protection creates dynamic gains by encouraging innovation, but also creates static deadweight losses by granting greater monopoly power to innovating firms. When there are multiple countries interacting with one another through trade and FDI, the tradeoffs associated with IPR protection are similar, but now IPR policy in one country may affect outcomes in other countries as well. This implies that, akin to other trade-related policies such as tariffs, IPR policy alone also has the potential to generate cross-border externalities and inefficient non-cooperative behavior. The existence of the TRIPS agreement is again evidence that IPR policy is an important concern in the international arena, and in fact is often seen as an effort by developed nations to ensure a minimum standard of IPR protection amongst their developing country trade partners.

More recent theoretical work on IPR policy in the international trade literature has built on these central ideas, and has studied the incentives for IPR protection in various contexts. Perhaps the first formal general equilibrium analysis of this sort was undertaken by Helpman (1993), who considers a North-South model in which all innovation takes place in the North, imitation occurs in the South, and the South chooses its optimal level of IPR protection. Using this model, Helpman identifies four channels via which IPR protection can affect consumer welfare: (1) the terms of trade, (2) the allocation of production across
countries; (3) the availability of products within countries; and (4) the timing and level of R&D investments. The model that I develop in this paper captures all four channels in a framework with heterogeneous firms, except for the timing of R&D investment, which the model cannot be used to study given its static nature.

Following the work of Helpman, another seminal contribution to the theory of optimal IPR policy in the international trade literature was made by Grossman and Lai (2004), who also consider a North-South model, but focus instead on the simultaneous choice of IPR protection in the two countries. By characterizing the non-cooperative Nash equilibrium outcome of the model and comparing this with the globally-efficient regime of IPR protection, Grossman and Lai show how inefficiencies might arise from the setting of IPR policy alone. In what follows, I replicate this finding in a different modeling framework that allows for firm heterogeneity and FDI, while at the same time characterizing a new type of inefficiency, one that results from the TT decisions of multinational firms.

While the papers by Helpman and Grossman and Lai highlight most of the key mechanisms via which IPR policy affects consumer welfare across countries, they largely abstract from the possibility that countries may interact with one another via FDI. In principle, allowing for FDI introduces a new channel via which IPR policy can affect the international allocation of production. On the one hand, foreign firms with better technology may be more willing to establish subsidiaries in developing economies where the strength of IPR protection is higher. On the other hand, the rate of local imitation may be reduced directly by more stringent IPR regulations, and indirectly by the influx of FDI via a crowd-out effect. Therefore, the optimal level of IPR protection with respect to its effects on the FDI margin is ambiguous.

These ideas have been studied formally by Branstetter and Saggi (2011), who develop a model that allows for FDI from North to South and use it to examine how the existence of this channel affects incentives for IPR protection in the South. Using data on IPR reforms in 16 countries, Branstetter et al (2011) also estimate the responses of multinational firms' FDI
decisions to categorical shifts in IPR regimes, and find evidence that FDI responds positively to a strengthening of IPR protection to an extent that outweighs the crowding-out of local imitation. That is, stronger IPR protection promotes industrial development in developing economies. In a related paper, Bilir (2014) examines how the effects of IPR policy on FDI may depend on product lifecycle lengths, finding that the response of multinational activity to stronger IPR protection is more pronounced in sectors where lifecycle lengths are longer. This work therefore highlights the importance of incorporating a channel for FDI in a model of IPR policy, although none of the above-mentioned papers address optimal IPR policy explicitly, nor do they allow for the study of technology transfer by multinational firms.

In general, the theoretical mechanisms by which IPR policy affects firms and consumers when countries interact with one another through both trade and FDI are relatively well-understood. However, with regard to the second strand of the literature that this paper builds upon - that dealing with technology transfer - there has thus far been little theoretical work examining the interaction between IPR policy and TT by multinational firms. But what precisely is meant by “technology transfer”?

Here, a useful schema for thinking about the role that multinational firms play in the flow of technology between countries is offered by the United Nations Conference on Trade and Development (UNCTAD). In UNCTAD’s World Investment Report, multinationals are seen as being involved in three separate stages of the flow of technology from one country to another: (1) technology generation, (2) technology transfer, and (3) technology dissemination. The first stage refers essentially to research and development (R&D) undertaken by the multinational firm. The second and third stages concern how innovations generated by multinational firms spread to other countries, but whereas technology transfer refers to the deliberate transmission of technology from firm to firm (embodied in both physical goods and tacit knowledge), technology dissemination refers to spillovers in the host economy resulting from multinational activity, via externalities that are seldom priced in well-functioning markets.
With regard to technology transfer more specifically, UNCTAD also distinguishes between two forms of TT: internal versus external. The former refers to transactions between multinational firms and their affiliates (either part- or fully-owned), while the latter refers to transactions between firms that do not share common ownership. Here, it is well-documented that the majority of payments for technology imports is of the internal variety. Data from UNCTAD show, for example, that from 1985 to 1997, intra-firm receipts of royalties and license fees accounted for between 65% to 85% of the value of all such transactions for firms in Germany, Japan, and the United States.\textsuperscript{55} Although these distinctions are rough, they are nonetheless useful to keep in mind: in this paper, I focus on the internal transfer of technology from multinational firms to their subsidiaries, and abstract from the effects of technological spillovers from subsidiaries of multinationals to local firms.

With regards to this sort of internal transfer of technology by multinational firms, the existing academic literature has been mostly empirical. Perhaps most relevant for the issues examined in this paper is another study by Branstetter et al (2006), which looks at the effects of IPR reforms on subsidiaries’ payments for technology imports from their parent firms. Using the same data on IPR reforms as in Branstetter et al (2011) as well as data on US multinational firms from surveys by the Bureau of Economic Analysis (BEA), the authors find that payments increased by up to 30% because of reforms, which seems to suggest that the response of multinational firms’ TT decisions to changes in IPR policy can be substantial. Other empirical work on internal technology transfer by multinationals has looked at how the extent of TT is affected by variables such as host country competition (Blomström et al (1995)), ownership sharing (Blomström and Sjoholm (1999)), and the existence of international treaties (Bilir et al (2011)).

In sum, while there is empirical evidence supporting the hypothesis that the incentives for TT by multinational firms depend on IPR policies in the respective host countries, there has thus far been no attempt to formally model this interaction by embedding an endogenous

\textsuperscript{55}UNCTAD (1999), p. 404. See also Grosse (1989) and UNCTC (1988) for earlier data.
TT margin for firms in a structural model of IPR policy. This is precisely the goal of this paper, which is organized as follows. In Section 3.2, I first describe a closed-economy version of the model, which embeds a mechanism for imitation and a margin for IPR protection in the closed-economy version of the heterogeneous-firm framework in Melitz (2003). I then show how the model allows for a simple closed-form characterization of the optimal IPR policy of the government, which enables us to highlight the key tradeoff in offering stronger protection of IPRs: creating greater incentives for innovation by monopolists, and increasing the mass of imitators that charge lower prices.

In Section 3.3, I then consider an open-economy version of the model with two symmetric countries. Section 3.3.1 first considers the scenario in which the two countries interact with one another only via exporting. In this environment, the level of IPR protection in one country affects consumers in the other country as well because of firms that export. I use this version of the model to highlight the following result: when governments choose their desired level of IPR protection without taking the above externality into account, the resulting Nash equilibrium is inefficient relative to the symmetric global optimum.

Section 3.3.2 then allows for the two countries to interact via FDI instead of exporting, introducing a TT margin for multinational firms where the profit-maximizing choice of TT depends on the strength of IPR protection in the host country. In this environment, I show that when governments take firms’ profit-maximizing TT choices as given, the Nash equilibrium level of IPR protection is again inefficient. In addition, however, the environment with FDI allows us to study the efficiency properties of firms’ TT decisions, and here I highlight a new kind of inefficiency that stems from the failure of multinational firms to internalize the effects of technology transfer on the marginal costs of successful imitators in the host country. Numerical simulations of the model show that the magnitude of this inefficiency is comparable to the inefficiency arising from governments’ choices of IPR policy.

As we will see, the versions of the model considered in Sections 3.3.1 and 3.3.2 allow for analytically tractable characterizations of the optimal IPR and TT decisions, despite the
potential complexities that are introduced by firm heterogeneity. However, this tractability depends in large part on a particular assumption regarding household preferences - that households in one country spend a fixed share of income on varieties invented in the foreign country. Therefore, to relax this assumption, I study a more general version of the model in Section 3.3.3, in which I also allow for firms in one country to serve the foreign market via either exporting or FDI. I solve the model numerically and show that the comparative static results are qualitatively similar to those obtained in previous sections. I also use this version of the model to quantify the welfare losses resulting from the IPR and TT inefficiencies.

Finally, Section 3.4 concludes.

3.2 Closed Economy Model

Here, we first study a closed-economy version of the model, in order to highlight the key tradeoff for the government in its choice of IPR policy. The model in this section embeds a mechanism for imitation and a margin for IPR protection in the heterogeneous-firm framework of Melitz (2003). The model is static (or equivalently, the steady-state of a dynamic model with exogenous firm death), and we take the wage as the numeraire.

3.2.1 Households

There is a representative household in the economy that is endowed with labor \( L \) and that has preferences over varieties of a differentiated product given by:

\[
U = \left[ \frac{Q^M}{\mu} \right]^{\mu} \left[ \frac{Q^I}{1 - \mu} \right]^{1 - \mu}
\]  
(3.2.1)

\[
Q^M = \left[ \int_{\omega \in \Omega^M} q^M(\omega) \frac{\sigma - 1}{\sigma} d\omega \right]^{\sigma / (\sigma - 1)}
\]  
(3.2.2)

\[
Q^I = \left[ \int_{\omega \in \Omega^I} q^I(\omega) \frac{\sigma - 1}{\sigma} d\omega \right]^{\sigma / (\sigma - 1)}
\]  
(3.2.3)
The interpretation of this preference specification is straightforward. $Q^M$ and $Q^I$ are CES bundles of non-imitated (i.e. produced under monopoly) and imitated varieties of the differentiated product respectively, where $\Omega^M$ and $\Omega^I$ denote the set of all such varieties that are available in the economy. The elasticity of substitution within each of these two bundles is $\sigma > 1$, and we define $\rho \equiv \frac{\sigma - 1}{\sigma}$. $Q$ is then a Cobb-Douglas aggregation of $Q^M$ and $Q^I$, so that a fraction $\mu$ of the household’s income is spent on non-imitated varieties. Here, it is assumed that households have different tastes for varieties that have been imitated and those that have not, which may be interpreted as a reduced-form way of capturing the fact that non-imitated products may have a higher “status” in the eyes of consumers than products for which there is a flood of imitations.

This preference specification implies the following demand functions for non-imitated and imitated varieties:

\[
q^M(\omega) = \mu L \frac{p^M(\omega)^{-\sigma}}{(P^M)^{1-\sigma}} \tag{3.2.4}
\]
\[
q^I(\omega) = (1 - \mu) L \frac{p^I(\omega)^{-\sigma}}{(P^I)^{1-\sigma}} \tag{3.2.5}
\]

where $p^M(\omega)$ and $p^I(\omega)$ are the prices charged by individual firms, and $P^M$ and $P^I$ are price indices for the non-imitated and imitated sectors that are defined as follows:

\[
P^M = \left[\int_{\omega \in \Omega^M} p^M(\omega)^{1-\sigma} \, d\omega\right]^{\frac{1}{1-\sigma}} \tag{3.2.6}
\]
\[
P^I = \left[\int_{\omega \in \Omega^I} p^I(\omega)^{1-\sigma} \, d\omega\right]^{\frac{1}{1-\sigma}} \tag{3.2.7}
\]

Note also that we have already imposed labor market clearing by treating the total expenditure of the household as being equal to the aggregate labor endowment $L$ (recall that the wage is the numeraire).
Per-capita utility can then be written as a function of prices in the economy:

\[ V \equiv \frac{U}{L} = P^{-1} \]  \hspace{1cm} (3.2.8)

where \( P \) is an aggregate price index defined by:

\[ P = (P^M)^\mu (P^I)^{1-\mu} \]  \hspace{1cm} (3.2.9)

In what follows, we will treat \( V \) as the objective function for the government. Equation (3.2.8) then implies that if it wishes to maximize per-capita utility, the government should choose IPR policy to minimize the price index \( P \).

### 3.2.2 Production technology and imitation

On the production side of the economy, there is an unbounded mass of potential entering firms. Firms wishing to enter the market choose the variety of the differentiated product that they wish to produce, and must then hire a fixed amount of labor \( f^E \) in order to receive an iid productivity draw \( \varphi \) from a continuous distribution \( G \). Henceforth, we will assume that the productivity distribution is Pareto with scale parameter \( \varphi \) and shape parameter \( \alpha \), so that:

\[ G(\varphi) = 1 - \left( \frac{\varphi}{\varphi} \right)^\alpha, \ \varphi \geq \varphi \]  \hspace{1cm} (3.2.10)

As is standard in such models, we make the following assumption to ensure boundedness of firms’ profits in expectation:

**Assumption 9.** \( \alpha > \sigma - 1 \).

Now, receipt of a productivity draw \( \varphi \) allows the firm to produce its chosen variety at a constant marginal cost of \( \frac{1}{\varphi} \) units of labor, so that conditional on production, a monopolist with productivity \( \varphi \) that maximizes its profits subject to the demand function (3.2.4) will choose its price to be a constant markup \( \frac{1}{\rho} \) of its marginal cost. We also assume that all
monopolists must pay a fixed cost \( f^D \) to operate, so that profits conditional on production are given by:

\[
\pi^D(\varphi) = \frac{\varphi^{\sigma-1}}{k(PM)^{1-\sigma}} - f^D
\]  

(3.2.11)

where \( k \equiv \frac{\sigma}{\mu L^\sigma r} \) is a constant. One can interpret the entry of firms in this model as the innovation of new varieties of the differentiated good, and \( f^E \) as including the costs of R&D needed to generate these innovations. It is also important to emphasize here that entry into the market gives the firm not only access to the production technology implied by \( \varphi \), but also ownership of monopoly rights to produce a specific variety using this technology. We will refer to such a technology-variety pair as a blueprint.

Next, to introduce a mechanism for imitation into the model, we will assume that a firm’s monopoly rights are only enforced by the government with some probability \( \gamma \). In what follows, we will treat \( \gamma \) as the sole choice variable for the government, and interpret it as a measure of the strength of IPR protection in the country. We will also abstract from any costs of enforcement of IPR protection, so that all costs and benefits to the government from choosing a particular value of \( \gamma \) accrue through the effects of this choice on consumers via its effects on the decisions of firms.

If a firm is hit by the imitation shock (which we assume is iid across blueprints), then its blueprint becomes available for use by all firms operating in the economy. Specifically, all firms can produce the imitated variety at the same marginal cost \( \frac{1}{\varphi} \) as the original monopolist. To keep things simple, we will also assume that production of imitated varieties requires no additional fixed costs, and that imitation is costless for firms.\(^{56} \) These assumptions imply that (i) a variety that is imitated is indeed produced in the economy, and (ii) all firms producing imitated varieties charge prices equal to marginal cost and therefore earn zero profits from these varieties.

Finally, after receiving its productivity draw and realizing its imitation shock, an entering

\(^{56}\)The fixed operating cost for monopolists may therefore be interpreted as partially capturing the cost of protecting its intellectual property.
firm can then decide whether to exit the market or whether to stay and produce its chosen variety. Firms whose blueprints are imitated are indifferent between producing and exiting because they receive zero profits, while those whose monopoly rights are enforced will produce if and only if their productivity satisfies \( \varphi \geq \varphi^{*D} \), where \( \varphi^{*D} \) is a cutoff value defined by:

\[
\pi^D(\varphi^{*D}) = 0 \Leftrightarrow \varphi^{*D} = (k f^D)^{\frac{1}{\sigma-1}} (P^M)^{-1}
\]  

(3.2.12)

To solve for the cutoff productivity level, we consider the free-entry condition for firms, which given our assumptions about the imitation process implies the following:

\[
\gamma \int_{\varphi^{*D}}^{\infty} \pi^D(\varphi) dG(\varphi) = f^E
\]  

(3.2.13)

Using equations (3.2.10)-(3.2.12), we can then solve (3.2.13) to obtain:

\[
\varphi^{*D} = \phi \gamma^{\frac{1}{\alpha}}
\]  

(3.2.14)

where \( \phi \equiv \left[ \frac{\sigma-1}{\alpha-(\sigma-1)} \cdot \frac{f^D}{f^E} \right]^{\frac{1}{\alpha}} \varphi \) is a constant.

### 3.2.3 Optimal IPR policy in the closed economy

In order to study the optimal choice of \( \gamma \) for the government, we must first characterize the aggregate price index \( P \) in the equilibrium of the model. Note that equations (3.2.12) and (3.2.14) immediately give us a solution for the monopolist price index \( P^M \) as a function of \( \gamma \):

\[
P^M = \left[ (k f^D)^{\frac{1}{\sigma-1}} \phi^{-1} \right] \gamma^{-\frac{1}{\alpha}}
\]  

(3.2.15)

Equation (3.2.15) shows that the monopolist price index is decreasing in \( \gamma \), the intuition for which is as follows. Stronger IPR protection tends to increase ex-ante profits for innovators, but in equilibrium, ex-ante profits must equal entry costs because of free-entry. Therefore,
the cutoff productivity value $\varphi^{*D}$ must adjust upwards to keep ex-ante profits constant - when all firms are more productive conditional on survival, the expected profit for an individual firm is lower. This increase in $\varphi^{*D}$ means that the average productivity of all monopolists operating in the economy is higher; since more productive firms charge lower prices, the average price charged by monopolists is also lower when IPR protection is stronger. Note that this effect operates solely through the selection of more productive firms: one can show that both the mass of entrants as well as the mass of monopolists operating in the economy are independent of $\gamma$, so stronger IPR protection raises the overall “quality” of monopolists but not their quantity.\(^{57}\)

To solve for the imitator price index $P^I$, we observe that our assumptions about the imitation process imply the following:

$$
\frac{P^I}{P^M} = \left[ \frac{1 - \gamma}{\gamma M^E \int_{\varphi^{*D}}^{\infty} \varphi^{\sigma-1} dG(\varphi)} \right]^{1/\sigma} \\
= \left( \frac{\gamma}{1 - \gamma} \right)^{1/\sigma - 1} \rho
$$

(3.2.16)

where here $M^E$ is the mass of entrants in the economy. Note that non-imitated varieties are priced at a markup $\frac{1}{\rho}$ of marginal cost, whereas for imitated varieties prices are equal to marginal cost. Equation (3.2.16) shows that given a value for the monopolist price index, stronger IPR protection tends to increase the imitator price index. The intuition for this result is straightforward: the productivity distribution of imitators is identical to the productivity distribution of monopolists, but an increase in $\gamma$ reduces the mass of imitators relative to the mass of monopolists.

Equations (3.2.12) and (3.2.14) therefore summarize the tradeoff for the government in

\(^{57}\)Using the fact that $(P^M)^{1-\sigma} = \gamma M^E \int_{\varphi^{*D}}^{\infty} (\rho \varphi)^{\sigma-1} dG(\varphi)$, we can substitute for $P^M$ and $\varphi^{*D}$ using equations (3.2.14) and (3.2.15), and solve for the mass of entrants in the economy, $M^E$. Doing so, we find that $M^E = \frac{\phi^{*D}}{\gamma \alpha \rho}$, which does not depend on $\gamma$. The mass of monopolists that actually produce is given by $M^M = \gamma M^E \left[ 1 - G(\varphi^{*D}) \right] = \gamma M^E \varphi^{\sigma} (\varphi^{*D})^{-\alpha}$. From the solution for $\varphi^{*D}$ given by equation (3.2.14), we see that $M^M$ also does not depend on $\gamma$.\]
its choice of IPR policy: stronger IPR protection reduces prices charged by monopolists, but also reduces the mass of imitators. More formally, we can now solve explicitly for the aggregate price index by substituting the expressions for \( P^M \) and \( P^I \) into equation (3.2.16), obtaining:

\[
P = \Phi_{\text{closed}} \frac{1}{\gamma^\frac{1}{\alpha} - \frac{1}{\alpha}} \frac{1}{(1 - \gamma)^\frac{1}{\sigma - 1}} \tag{3.2.17}
\]

where \( \Phi_{\text{closed}} \equiv \rho^{1-\mu} \left( k f^D \right)^{\frac{1}{\sigma - 1}} \phi^{-1} \) is a constant. In order to maximize per-capita utility, the government therefore chooses \( \gamma \) to maximize the denominator of the right-hand side of equation (3.2.17). This optimization problem has a unique interior solution under the following regularity assumption:

**Assumption 10.** \( \frac{\sigma - 1}{\alpha} > 1 - \mu. \)

Evidently, given Assumption 9, Assumption 10 requires the share of income spent on non-imitated varieties to be large enough. The optimal IPR policy for the government in the closed economy is then given by:

\[
\gamma_{\text{closed}}^* = 1 - \frac{\alpha(1 - \mu)}{\sigma - 1} \tag{3.2.18}
\]

and the following proposition summarizes the key results of the closed-economy model.

**Proposition 8.** The optimal strength of IPR protection in the closed economy \( \gamma_{\text{closed}}^* \) is:

1. strictly increasing in the share of income spent on non-imitated varieties (\( \mu \)),
2. strictly increasing in the degree of firm heterogeneity (as indexed by \( \frac{1}{\alpha} \)), and
3. strictly increasing in the elasticity of substitution across varieties (\( \sigma \)).

As discussed above, the tradeoff for the government in the closed economy is between keeping the monopolist price index low on the one hand, and keeping the imitator price index low on the other. When households spend a larger share of income on non-imitated
varieties, they value the first effect relatively more, which explains part (1) of Proposition 8. When the degree of firm heterogeneity is large (smaller $\alpha$), the free-entry condition dictates a larger increase in the cutoff productivity $\varphi^{*D}$ for a given increase in $\gamma$. This implies that the elasticity of the monopolist price index with respect to the level of IPR protection is larger (more negative) when $\alpha$ is smaller, which explains part (2) of Proposition 8. Finally, when varieties are more substitutable with respect to one another, households do not value an increase in the mass of imitators (which provides a larger variety of imitations) as much, which explains part (3).

3.3 Open Economy Model

Having discussed how the mechanism for imitation leads to a meaningful decision problem for the government in its choice of IPR policy in the closed economy, we now examine the incentives for IPR protection in an open economy. In what follows, we consider a world with two countries that are symmetric in all respects, except potentially with regard to the choice of IPR protection by governments and the choice of TT by multinational firms; having imposed symmetry with respect to parameters, however, we will also focus only on equilibria that are symmetric in choice variables.

Subsection 3.3.1 first considers a world in which the two countries interact with one another only via trade. Analysis of this version of the model will highlight the cross-country externality generated by a government’s choice of IPR protection, which in turn leads to an inefficiency when governments behave non-cooperatively. Subsection 3.3.2 then examines a case in which the two countries interact with one another only via FDI, introducing a margin for TT by firms and studying how the incentives for TT interact with the incentives for IPR protection. We will use this version of the model to highlight how TT by multinational firms generates a positive externality with respect to the marginal costs of successful imitating.

---

58 More specifically, the variance of the productivity Pareto distribution is equal to $\frac{\varphi^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$, which is strictly decreasing in $\alpha$. 

132
firms, which therefore leads to a second kind of inefficiency when multinational firms fail to internalize the benefits of this effect. Finally, Subsection 3.3.3 considers a more general version of the model that allows for both exports and FDI.

3.3.1 Exports

3.3.1.1 Households

As before, we assume that the representative household in each country is endowed with labor $\bar{L}$, but that preferences over varieties of the differentiated product for the household in country $i$ are now given by:

$$U_i = \left[ \frac{Q_i}{\eta} \right]^{\eta} \left[ \frac{\tilde{Q}_i}{1-\eta} \right]^{1-\eta}$$  \hfill (3.3.1)

$$Q_i = \prod_{j=1}^{2} \left[ \frac{Q_{ij}}{\beta_{ij}} \right]^{\beta_{ij}}$$  \hfill (3.3.2)

$$\beta_{ij} = \begin{cases} \beta , & i = j \\ 1-\beta , & i \neq j \end{cases}$$

$$Q_{ij} = \left[ \frac{Q_{ij}^M}{\mu} \right]^{\mu} \left[ \frac{Q_{ij}^I}{1-\mu} \right]^{1-\mu}$$  \hfill (3.3.3)

$$Q_{ij}^M = \left[ \int_{\omega \in \Omega_{ij}^M} q_{ij}^M(\omega) \frac{\alpha+1}{\sigma} d\omega \right]^{\frac{\sigma}{\sigma - 1}}$$  \hfill (3.3.4)

$$Q_{ij}^I = \left[ \int_{\omega \in \Omega_{ij}^I} \tilde{q}_{ij}(\omega) \frac{\alpha+1}{\sigma} d\omega \right]^{\frac{\sigma}{\sigma - 1}}$$  \hfill (3.3.5)

Here, $\tilde{Q}_i$ is a numeraire good that is produced in both countries under perfect competition one-for-one using labor, and is used to pin down wages (which therefore equal unity). Households therefore spend a constant amount $L \equiv \eta \bar{L}$ on the differentiated product. Note also that this preference specification allows consumers to differentiate between varieties based
on the country in which they were invented, and implies that a constant fraction $\beta$ of the household’s income is spent on varieties invented domestically, while the remaining $1 - \beta$ is spent on varieties invented abroad. Specifically, $Q_{ij}$ is a Cobb-Douglas aggregator of the non-imitated and imitated bundles of varieties invented in $j$ and sold in $i$, and $Q^M_{ij}$ are $Q^I_{ij}$ are again CES aggregators with elasticity of substitution $\sigma$ of the set of all non-imitated $(\Omega^M_{ij})$ and imitated $(\Omega^I_{ij})$ invented in $j$ and sold in $i$.

The demand in $i$ for non-imitated and imitated varieties invented in $j$ are now given respectively by:

$$q^M_{ij}(\omega) = \mu \beta_{ij} L P^M_{ij}(\omega)^{-\sigma} \left( P^M_{ij} \right)^{1-\sigma} \tag{3.3.6}$$

$$q^I_{ij}(\omega) = (1 - \mu) \beta_{ij} L P^I_{ij}(\omega)^{-\sigma} \left( P^I_{ij} \right)^{1-\sigma} \tag{3.3.7}$$

Per-capita utility is now given by:

$$V \equiv \frac{U}{L} = P^{-\eta} \tag{3.3.8}$$

where the aggregate price index is:

$$P_i = \prod_{j=1}^{2} \left[ (P^M_{ij})^\mu (P^I_{ij})^{1-\mu} \right]^{\beta_{ij}} \tag{3.3.9}$$

Again, $P^M_{ij}$ and $P^I_{ij}$ are price indices for the monopolistic and imitated sectors respectively, defined analogously to (3.2.6) and (3.2.7).
3.3.1.2 Production technology and imitation with exports

Given the preference specification described above, a firm from $i$ with productivity $\varphi$ that produces domestically earns profits given by:

$$\pi^D_i(\varphi) = \frac{\beta \varphi^{\sigma-1}}{k (P^M_{ii})^{1-\sigma}} - f^D$$  \hspace{1cm} (3.3.10)

In addition to producing for the domestic market, however, firms can now also export their varieties to the foreign country, paying an additional fixed cost $f^X$ and an iceberg trade cost $\tau > 1$. Conditional on exporting, a firm from $i$ with productivity $\varphi$ that exports to $j$ therefore earns profits given by:

$$\pi^X_{ji}(\varphi) = \frac{(1 - \beta) \tau^{1-\sigma} \varphi^{\sigma-1}}{k (P^M_{ji})^{1-\sigma}} - f^X$$  \hspace{1cm} (3.3.11)

The productivity cutoffs for domestic production and exporting are then defined respectively by:

$$\varphi^*_i = \left( \frac{k f^D}{\beta} \right)^{\frac{1}{\sigma-1}} (P^M_{ii})^{-1}$$  \hspace{1cm} (3.3.12)

$$\varphi^*_ji = \left( \frac{k f^X}{1 - \beta} \right)^{\frac{1}{\tau-1}} \tau (P^M_{ji})^{-1}$$  \hspace{1cm} (3.3.13)

With regards to the imitation of exported varieties, we again assume that the monopoly rights for a firm from $i$ to sell in $j$ are only enforced by the government in $j$ with probability $\gamma_j$. For simplicity, we assume that imitation of an imported variety also allows for production of that variety with the same productivity as the original monopolist, but absent the iceberg transport cost since the imitation is then produced domestically. This implies that the free-entry condition for entrants in $i$ can be written as:

$$\gamma_i \int_{\varphi^*_i}^{\infty} \pi^D_i(\varphi) dG(\varphi) + \gamma_j \int_{\varphi^*_j}^{\infty} \pi^X_{ji}(\varphi) dG(\varphi) = f^E$$  \hspace{1cm} (3.3.14)

135
Note that in order to successfully sell its variety in country \( j \), a monopolist from \( i \) must survive both the imitation shocks in \( i \) and \( j \). Therefore, it receives monopoly profits from exporting to \( j \) with probability \( \gamma_i \gamma_j \).

### 3.3.1.3 Solving for the price indices

As in the closed economy, analysis of the optimal choice of IPR protection by governments first requires a characterization of the aggregate price indices. To solve for these, we first observe that the monopolist price indices for domestic production in \( i \) and for exporting from \( i \) to \( j \), \( P_{ii}^M \) and \( P_{ji}^M \), are related according to:

\[
\left( \frac{P_{ji}^M}{P_{ii}^M} \right)^{1-\sigma} = \frac{\gamma_i \gamma_j M_i^E \int_{\varphi_i^D}^{\infty} \left( \frac{p_e}{\tau} \right)^{\sigma-1} dG (\varphi)}{\gamma_i M_i^E \int_{\varphi_i^D}^{\infty} (p \varphi)^{\sigma-1} dG (\varphi)}
\]

\[
= \frac{\gamma_i \tau^{1-\sigma} \left( \frac{\varphi_j^X}{\varphi_i^D} \right)^{\sigma-1}}{(\gamma_i \gamma_j)^{1-\alpha}} \tag{3.3.15}
\]

Equations (3.3.12)-(3.3.13), the analogs of (3.3.12)-(3.3.13) for \( \varphi_j^D \) and \( \varphi_i^X \), equation (3.3.14), the analog of (3.3.14) for country \( j \), equation (3.3.15), and the analog of (3.3.15) for \( P_{ji}^M \) then give us 8 equations in the 4 productivity cutoffs and the 4 monopolist price indices. We can solve these analytically to obtain the following expressions for the monopolist price indices relevant for consumption in \( i \):

\[
P_{ii}^M = \left[ \beta \gamma_i^{-1} \left( k f^D \right)^{1-\alpha} \phi^{-1} \right] \gamma_i^{-\frac{1}{\alpha}} \tag{3.3.16}
\]

\[
P_{ij}^M = \left[ \left( 1 - \beta \right) \left( k f^D \right)^{1-\alpha} \phi^{-1} \right] \left( \gamma_i \gamma_j \right)^{-\frac{1}{\alpha}} \tag{3.3.17}
\]

Note that equations (3.3.16) and (3.3.17) are almost identical to the expression for the monopolist price index in the closed economy (equation (3.2.15)), except that the price index for exports now depends on IPR policy in both countries. As in the closed economy, stronger IPR protection in country \( i \) lowers prices charged by monopolists (both domestic and foreign).
in that country, via the selection effect on firms. In a world with exports, however, weaker protection in \( j \) now also exerts a negative externality on households in country \( i \) - the average price paid for imports from \( j \) is higher, again because of the selection effect on firms in \( j \). It is this externality that generates inefficient non-cooperative behavior between governments in their choice of IPR policies, and as we will see, the Nash equilibrium results in too little protection of IPRs.

To complete the characterization of the aggregate price index, we observe that analogously to equation (3.2.16), the imitator price indices relevant for consumption in \( i \) are related to their monopolist counterparts by the following:

\[
\frac{P^I_{ii}}{P^M_{ii}} = \left( \frac{\gamma_i}{1 - \gamma_i} \right)^{\frac{1}{\sigma - 1}} \rho \tag{3.3.18}
\]

\[
\frac{P^I_{ij}}{P^M_{ij}} = \left( \frac{\gamma_i}{1 - \gamma_i} \right)^{\frac{1}{\sigma - 1}} \rho \tau^{-1} \tag{3.3.19}
\]

(Note that \( \tau \) appears in equation (3.3.19) because the iceberg trade cost is paid by monopolists that export from \( i \) to \( j \), but not by firms in \( j \) that successfully imitate varieties imported from \( i \)). As before, stronger IPR protection reduces the mass of imitators relative to the mass of monopolists, which increases the imitator price index relative to the monopolist price index.

Substituting equations (3.3.16)-(3.3.19) into the definition of the aggregate price index given by (3.3.9), we thus obtain:

\[
P_i = \Phi_{\text{exp}} \left( \frac{1 - \gamma_i}{1 - \gamma_i} \right)^{\frac{1}{\sigma - 1}} \gamma_i^{\frac{1}{\sigma - 1}} \gamma_j^{\frac{1}{\sigma - 1}} \phi^{-1}
\]

where \( \Phi_{\text{exp}} \equiv \rho^{1 - \mu} \tau^{\mu(1 - \beta)} (k f^{D})^{\frac{1}{1 - \beta}} \left[ \left( \frac{1}{\beta} \right) \beta \left( \frac{f^X}{f^D} \right) \right]^{\frac{1}{1 - \beta}} \phi^{-1} \) is a constant. Hence, we see that welfare in country \( i \) is affected not only by \( \gamma_i \), but by \( \gamma_j \) as well through its effect on firms that export from \( j \) to \( i \). In particular, weaker IPR protection in \( j \) unambiguously leads to an increase in \( P_i \) and therefore to a decrease in household welfare in \( i \).
3.3.1.4 Optimal IPR Policy in an open economy with exports

Suppose first that the government in $i$ takes IPR policy in $j$ as given and chooses $\gamma_i$ to maximize household welfare in $i$. Comparing equations (3.2.17) and (3.3.20), it is clear that the optimal choice of $\gamma_i$ for the government is the same as in the closed economy. The assumed symmetry between the two countries then implies that if the two governments choose their IPR policies simultaneously and non-cooperatively, the unique Nash equilibrium also results in the same level of protection as in the closed economy:

$$\gamma_{\text{exp,NE}} = \gamma_{\text{closed}} = 1 - \frac{\alpha (1 - \mu)}{\sigma - 1}$$ (3.3.21)

The equality between $\gamma_{\text{exp,NE}}$ and $\gamma_{\text{closed}}$ is a special result that depends largely on the assumption of constant expenditure shares for domestic and foreign varieties, and therefore should not be interpreted too literally. Equation (3.3.21) also implies, however, that the comparative statics described in Proposition 8 for the optimal level of IPR protection in the closed economy apply as well to Nash equilibrium level of IPR protection in the open economy with exports. This result is indeed robust to alternative assumptions about household preferences (such as those studied in Section 3.3.3), and we summarize it in the following Proposition:

**Proposition 9.** In the open economy with exports, the level of IPR protection that results in the non-cooperative Nash equilibrium, $\gamma_{\text{exp,NE}}$, is:

1. strictly increasing in the share of income spent on non-imitated varieties ($\mu$),
2. strictly increasing in the degree of firm heterogeneity (as indexed by $\frac{1}{\alpha}$), and
3. strictly increasing in the elasticity of substitution across varieties ($\sigma$).

Now, suppose instead that the two governments could cooperate and commit to a (common) level of IPR protection. In this case, symmetry between the two countries implies that
the optimal choice of the cooperative $\gamma$ is the value that maximizes the denominator of the right-hand side of equation (3.3.20), but with $\gamma_i = \gamma_j = \gamma$. Assumption 10 is sufficient to guarantee a unique interior solution for this optimization problem, and the resulting level of IPR protection, which we will refer to as the global optimum, is given by:

$$
\gamma_{exp,GO}^* = 1 - \frac{\alpha (1 - \mu)}{(2 - \beta)(\sigma - 1)} \tag{3.3.22}
$$

Since $\beta \in (0, 1)$, then clearly $\gamma_{exp,GO}^* > \gamma_{exp,NE}^*$, so that the globally-optimal level of IPR protection is higher than the level of protection that results in the non-cooperative Nash equilibrium. We will henceforth refer to this discrepancy between the Nash equilibrium and the globally-optimal level of IPR protection as the IPR inefficiency. Note that the difference between $\gamma_{exp,NE}^*$ and $\gamma_{exp,GO}^*$ is increasing in the share of income spent on varieties invented abroad $(1 - \beta)$. Given the discussion above regarding the source of the cross-country externality, the intuition for this result is straightforward: there is too little protection of IPRs in the Nash equilibrium because governments do not internalize the fact that weaker IPR protection in their country reduces welfare in the other country, and the larger the share of income that households spend on foreign goods, the larger the impact this externality has on household welfare.

What implications does the discrepancy between $\gamma_{exp,NE}^*$ and $\gamma_{exp,GO}^*$ have for household welfare? From the expression for the household’s indirect utility function (3.2.8), we see that the percentage welfare loss resulting from non-cooperative setting of IPR policies can be written as:

$$
\% \Delta V_{exp} \equiv 1 - \frac{V_{exp,NE}}{V_{exp,GO}} = 1 - \frac{P_{exp,GO}}{P_{exp,NE}} \tag{3.3.23}
$$

where $P_{exp,GO}$ and $P_{exp,NE}$ are the values of the aggregate price index under the global optimum and the Nash equilibrium respectively, and $V_{exp,GO}$ and $V_{exp,NE}$ are the resulting values of household utility. Substituting (3.3.21) and (3.3.22) into (3.3.20), we can then write
where straightforward differentiation shows that the welfare loss is indeed increasing in $1 - \beta$. We summarize our characterization of the IPR inefficiency in the following Proposition:

**Proposition 10.** In the open economy with exports, the globally-optimal level of IPR protection $\gamma^*_{\text{exp,GO}}$ is higher than the Nash equilibrium level, so that $\gamma^*_{\text{exp,GO}} > \gamma^*_{\text{exp,NE}}$. The magnitude of the IPR inefficiency (as measured by $\%\Delta V_{\text{exp}}$) is strictly increasing in the share of income spent on varieties invented abroad ($1 - \beta$).

To get a sense of the actual size of this inefficiency, Figure 24 plots the quantity $\%\Delta V_{\text{exp}}$ for various values of $\beta$ and $\mu$, setting $\alpha = 3.8$ and $\sigma = 3.8$.\textsuperscript{50} We see that the magnitude of the inefficiency is indeed increasing in the share of income spent on varieties invented abroad, as measured by $1 - \beta$, and is very sensitive to the choice of $\mu$. For instance, note that in this version of the model, the fraction of income spent on varieties *produced* domestically is given by $\beta + (1 - \mu) (1 - \beta)$. Choosing $\beta$ so that this fraction is equal to 0.93 (the share of income spent on domestic goods for the US in 2007 as reported by Arkolakis et al (2012)), the value of $\%\Delta V_{\text{exp}}$ ranges from 15.5% when $\mu = 0.2$ to 0.03% when $\mu = 0.8$.

In summary, the open economy model with exports highlights how non-cooperative setting of IPR policy can lead to an inefficient outcome, in which there is too little protection of IPRs in the Nash equilibrium. This result therefore provides a rationale for international agreements on IPR protection such as TRIPS, and is qualitatively identical to the inefficiency of the Nash equilibrium studied by Grossman and Lai (2004). Here, however, I have shown that the same kind of inefficiency can exist in a heterogeneous firm model, which therefore

\textsuperscript{50}The value used for the elasticity of substitution $\sigma$ is the same as that estimated in Bernard et al (2003), based on plant-level manufacturing data in the US. The choice of the productivity Pareto distribution shape parameter $\alpha$ is set so that $\alpha - (\sigma - 1) = 1$. In the model, this quantity would be the slope coefficient of a regression of the log of a firm’s size rank on the log of its actual size. A slope of 1 matches the data for most industries relatively well (see for example Helpman et al (2004)).
allows study of different kinds of comparative static results.

3.3.2 Foreign direct investment and technology transfer

Having demonstrated how the incentives for IPR protection can lead to inefficient non-cooperative behavior in an open economy, we now focus on a version of the model in which household preferences are the same as in the case with exports, but in which a firm can only sell in the foreign country by engaging in FDI. In doing so, the firm establishes a subsidiary to both produce and sell its variety in the foreign market. Focusing on this special case with no exports will allow us to highlight the mechanism via which IPR policy interacts with the TT decisions of multinational firms, and how the profit-maximizing choice of TT by firms can generate a new inefficiency distinct from the IPR inefficiency. In Section 3.3.3, we reintroduce exports in a more general version of the model.

3.3.2.1 Technology for FDI and technology transfer

To model the process of technology transfer between parent firms and their foreign subsidiaries, we assume that firms engaging in FDI can choose the fraction $T \in [0, 1]$ of technology transfer that they want. If a firm from $i$ with productivity $\varphi$ chooses a level of technology transfer $T$ for FDI in $j \neq i$, then workers employed in its subsidiary in $j$ can produce with productivity $T\varphi$. That is, the choice of $T$ acts as a productivity shifter. As with domestic production and exporting, we also assume that FDI requires payment of a fixed cost $f_F$.

To make the TT decision a meaningful one for the firm, however, we must also consider the cost to the TT process for the firm. Precisely what form should this cost of transferring technology take? In principle, technology is embodied not only in physical equipment (the machines that comprise a production line, for example), but also in the tacit knowledge of how that equipment should be used to produce a specific product (organization of the production line to produce as efficiently as possible). Therefore, while the transfer of

---

60See, for example, UNCTAD (1999) and Teece (1977) for discussions of this distinction.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>3.8</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.6</td>
</tr>
<tr>
<td>$f^D$</td>
<td>0.1</td>
</tr>
<tr>
<td>$f^A$</td>
<td>0.1</td>
</tr>
<tr>
<td>$f^E$</td>
<td>1</td>
</tr>
<tr>
<td>$f^E$</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>3.8</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>0.2</td>
</tr>
<tr>
<td>$\tau$</td>
<td>1.25</td>
</tr>
<tr>
<td>$L$</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>10</td>
</tr>
<tr>
<td>$\bar{c}$</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 5: Baseline Parameter Values

Figure 24: Welfare Loss Due to IPR Inefficiency in the Open Economy with Exports
technology embodied in physical capital might face only the same kinds of transport costs as applies to the shipping of finished products, the transfer of technology embodied in tacit knowledge might be subject to additional costs associated with, for example, the learning and absorption of this knowledge. In this paper, we focus on the latter kind of cost, and adopt the following approach to modeling the TT process for firms engaging in FDI.

Consider a firm that can produce with productivity $\varphi$ in the domestic market. We assume that in order for workers in the foreign subsidiary of this firm to produce with productivity $T\varphi$, the firm must pay a training cost of $C(T)$ per worker, where $C'(T) > 0$ and $C''(T) > 0$ for all $T \in (0, 1]$.

Furthermore, we assume that the training of workers in the foreign subsidiary occurs after realization of the imitation shock in the domestic market, but before realization of the imitation shock in the foreign market. The import of this key assumption regarding the timing of events is that the cost of TT for the firm must be paid regardless of whether its IPRs are protected in the foreign market or not, but the benefit to TT accrues to the firm only if its variety is not imitated. Therefore, this assumption immediately ties the incentives for TT to the strength of IPR protection in the host country.

Finally, we will assume that if the multinational firm’s IPRs are not enforced, then all firms in that market can produce the variety with productivity $T\varphi$.

Under these assumptions and given the demand function (3.3.6), we can write the expected profit from FDI in $j$ for a firm from $i$ with productivity $\varphi$, conditional on a choice of $T = T_j$, as follows:

---

61 One might interpret this cost as the need to hire $C(T)$ trainers per worker, for example, where trainers and workers are paid the same wage (equal to unity). Since wages are the same in both countries by assumption, then under this interpretation it is irrelevant whether trainers are hired from the domestic or the foreign labor market.

62 This need not be the only reason why such a dependence exists. For example, a weaker IPR regime might impose more direct costs of TT for a firm engaging in FDI, perhaps because it is more difficult to apply for patents when the level of IPR protection is low. Nonetheless, we will abstract from these other channels of interaction between IPR policy and TT decisions.
\[ E_{\pi_T}^F (\varphi | T_j) = \max_{l} \gamma_i \gamma_j \left[ \frac{(1 - \beta) \mu L}{P_j^{1-\sigma}} \right]^{\frac{1}{\sigma}} \left( T_j \varphi l \right)^{\frac{\sigma - 1}{\sigma}} - l - \frac{C(T_j)}{\gamma_j} \right]^{1-\sigma} \]  

That is, the firm chooses the quantity of labor \( l \) that it will hire in the event that its IPRs are protected, taking into consideration the fact that (i) each worker can produce with productivity \( T_j \varphi \), and (ii) for each worker hired, the firm must not only pay the wage of unity, but must also pay the training cost of \( C(T_j) \) prior to finding out whether its IPRs are enforced in \( j \) or not. Solving the optimization problem in (3.3.25), it is then straightforward to show that the firm’s expected profits from FDI can be written as:

\[ E_{\pi_T}^F (\varphi | T_j) = \gamma_i \gamma_j \left[ \frac{(1 - \beta) \mu L}{P_j^{1-\sigma}} \right]^{\frac{1}{\sigma}} \left( T_j \varphi l \right)^{\frac{\sigma - 1}{\sigma}} - l - \frac{C(T_j)}{\gamma_j} \right]^{1-\sigma} \]  

where we have defined:

\[ h(\gamma, T) \equiv \frac{T}{1 + \frac{C(T)}{\gamma}} \]  

3.3.2.2 The optimal choice of technology transfer for multinational firms

Equation (3.3.26) implies that if a firm from \( i \) engaging in FDI in \( j \) chooses its level of TT to maximize profits, then it will choose \( T_j \) to maximize \( h(\gamma_j, T_j) \) given IPR policy \( \gamma_j \) in \( j \). Under the assumption that the cost function \( C \) is strictly convex, the first order necessary condition for this maximization problem is also sufficient for a local optimum, and can be written as:

\[ C(T_j) \left[ \varepsilon_C (T_j) - 1 \right] = \gamma_j \]  

where \( \varepsilon_C \) is the elasticity of the function \( C \). Assuming that (3.3.28) is solved by an interior value of \( T_j \), the level of TT chosen by the firm is strictly increasing in the strength of IPR protection in the host country if and only if the left-hand side of (3.3.28) is also strictly
increasing in $T_j$ for $T_j \in (0,1)$. A sufficient condition, for example, is that the elasticity of the cost function is strictly greater than 1 and non-decreasing in $T$.

In what follows, we assume a simple parametric form for the cost function:

$$C(T) = \bar{c}T^\lambda$$  \hspace{1cm} (3.3.29)

so that $\varepsilon_C = \lambda$ is a constant. The profit-maximizing choice of TT for the firm is then given by:

$$T^* (\gamma_j) = \left[ \frac{\gamma_j}{\bar{c}(\lambda - 1)} \right]^{\frac{1}{\lambda}}$$  \hspace{1cm} (3.3.30)

and the following assumption on the parameters of the function $C$ ensures that $T^* (\gamma)$ is both interior and strictly increasing for all $\gamma \in (0,1)$.

**Assumption 11.** $\lambda \mu > 1$ and $\bar{c}(\lambda \mu - 1) > 1$.

In other words, technology transfer must be “costly enough” as measured by both the scale parameter $\bar{c}$ and the elasticity $\lambda$ of the cost function $C$, so that firms do not find it optimal to transfer the full extent of their technology to a foreign subsidiary. For convenience, we also define:

$$h^* (\gamma) \equiv h(\gamma, T^* (\gamma))$$

$$= \bar{h}\gamma^{\frac{1}{\lambda}}$$  \hspace{1cm} (3.3.31)

where $\bar{h} \equiv (\frac{\lambda-1}{\lambda}) \left[ \frac{1}{\bar{c}(\lambda-1)} \right]^{\frac{1}{\lambda}}$ is a constant.

Note that stronger IPR protection increases the value of profits for foreign monopolists producing in that country via three distinct channels. First, it reduces the probability of imitation, so that ex-ante profits are higher. Second, for a given value of $T$, it provides

---

63 Note that this also guarantees that (3.3.28) is solved by a unique value of $T$, and therefore that any local optimum is also the global optimum.

64 In fact, here we only need $\lambda > 1$ and $\bar{c}(\lambda - 1) > 1$. However, anticipating later results, the stronger conditions in Assumption 11 are needed to ensure that the *globally-optimal* choice of $T$ (equation (3.3.37)) is also interior and strictly increasing in $\gamma$. 

145
multinational firms with stronger incentives to hire more workers, which implies higher output, lower prices and higher profits in the event that the firm’s IPRs are enforced \((\frac{C(T)}{\gamma}\ is\ strictly\ decreasing\ as\ \gamma\ increases,\ so\ that\ h(\cdot, T)\ is\ strictly\ increasing)\). Third, it provides multinational firms with stronger incentives to transfer more technology \((h^{*}(\cdot)\ is\ strictly\ increasing)\).

Note also that in principle, firms may choose values of \(T\) that vary with their productivity: our assumptions about the TT process, however, imply that if foreign firms in \(j\) behave optimally, then they will all choose \(T = T^{*}(\gamma_{j})\). In what follows, we will therefore consider cases in which all foreign firms operating in a country \(j\) choose the same value of transfer \(T = T_{j}\), whether or not \(T_{j} = T^{*}(\gamma_{j})\).

### 3.3.2.3 Solving for the price indices

To solve for the monopolist price indices, we simply solve the system of equations defined by (3.3.12)-(3.3.15) for the case with exports, but replacing \(f^{X}\ with \(f^{F}\ and \(\tau\ with \(h(\gamma_{j}, T_{j})\) (or \(h(\gamma_{i}, T_{i})\)). Doing so, we find that \(P_{ij}^{M}\ is again given by (3.3.16), but that the price index for non-imitated varieties sold in \(i\ and invented in \(j\ is now given by:

\[
P_{ij}^{M} = \left[\left(1 - \beta\right) \left(\frac{f^{D}}{f^{X}}\right) \frac{1}{1 - \frac{1}{\sigma}} \left(k f^{D}\right) \frac{1}{\sigma - 1} \phi^{-1}\right] h(\gamma_{i}, T_{i})^{-1} (\gamma_{i} \gamma_{j})^{-\frac{1}{\sigma}} \right. (3.3.32)
\]

Here we see that equation (3.3.32) is almost identical to (3.3.17), except that it has \(h(\gamma_{i}, T_{i})^{-1}\ in place of \(\tau\). This highlights how allowing for FDI and TT introduces a new channel via which IPR policy can affect household welfare: since both \(h(\cdot, T_{i})\ and \(h^{*}(\cdot)\ are strictly increasing, stronger IPR protection in \(i\ increases profits for foreign firms operating in \(i\, which increases the average productivity of these firms via the same selection effect (from the free entry condition) as in the case with exports. This leads to lower prices charged by foreign monopolists in \(i\, therefore lowering the price index \(P_{ij}^{M}\).

As for the imitator price indices, first note that the ratio \(\frac{P_{ij}^{I}}{P_{ii}^{I}}\ is again given by (3.3.18).
However, the variety of a foreign firm in $i$ with productivity $\varphi$ and TT level $T_i$ is priced at $[\rho h(\gamma_i, T_i) \varphi]^{-1}$ if the firm’s IPRs are enforced, and at $(T \varphi)^{-1}$ if the firm’s IPRs are not enforced. Therefore, the ratio $\frac{P_{i}^{I}}{P_{i}^{M}}$ for the case with FDI and TT is given by:

$$\frac{P_{i}^{I}}{P_{i}^{M}} = \left( \frac{\gamma_i}{1 - \gamma_i} \right)^{\frac{1}{\alpha}} \left[ \frac{h(\gamma_i, T_i)}{T_i} \right] \rho$$

As in the case with exports, weaker IPR protection tends to decrease the imitator price index relative to the monopolist price index by increasing the relative mass of imitated varieties. Here, however, there are two additional effects. First, for a given value of $T_i$, foreign monopolists charge relatively higher prices than imitators when IPR protection is weaker. Second, for a given value of $\gamma_i$, imitators charge relatively lower prices than monopolists when $T_i$ is larger. Both these effects tend to reduce the ratio of $P_{i}^{I}$ to $P_{i}^{M}$. The second effect, in particular, highlights the fact that governments and foreign firms have different objectives with respect to the optimal choice of $T_i$, which, as we will see below, generates a new kind of inefficiency distinct from the IPR inefficiency studied in Section 3.3.1.

### 3.3.2.4 Optimal IPR and TT Policy in an open economy with FDI

Combining the expressions for the monopolist and imitator price indices, we can now write the aggregate price index in country $i$ as follows:

$$P_i = \Phi_{FDI} \frac{1}{(1 - \gamma_i)^{\frac{1}{\alpha}}} \left[ \left( \frac{1}{\beta} \right)^{\frac{1}{\alpha}} \left( \frac{k f^D}{1 - \beta} \right)^{\frac{1}{\beta}} \left[ h(\gamma_i, T_i)^{\mu} T_i^{1-\mu} \right]^{1-\beta} \right]$$

where $\Phi_{FDI} \equiv \rho^{1-\mu} \left( k f^D \right)^{\frac{1}{\beta}} \left[ \left( \frac{1}{\beta} \right)^{\frac{1}{\alpha}} \left( \frac{k f^D}{1 - \beta} \right)^{\frac{1}{\beta}} \right]^{\frac{1}{\beta-1}} \phi^{-1}$ is a constant. The objective for the government in $i$ is therefore again to maximize the denominator of the right-hand side of (3.3.34).

Now, first suppose that governments in both countries take the profit-maximizing behavior of multinational firms as given, so that $T = T^*(\gamma)$ and $h(\gamma, T) = h^*(\gamma)$ in both
i and j. In this case, non-cooperative setting of IPR policy again leads to same kind of inefficiency studied in the model with exports, where the unique Nash equilibrium choice of IPR protection is now:

\[
\gamma_{FDI,NE}^* = 1 - \left[ \frac{\lambda}{\lambda + \alpha (1 - \beta)} \right] \left[ \frac{\alpha (1 - \mu)}{\sigma - 1} \right] \tag{3.3.35}
\]

and the symmetric global optimum is:

\[
\gamma_{FDI,GO}^* = 1 - \left[ \frac{\lambda}{\lambda (2 - \beta) + \alpha (1 - \beta)} \right] \left[ \frac{\alpha (1 - \mu)}{\sigma - 1} \right] \tag{3.3.36}
\]

Note that in this version of the model, we have \( \gamma_{FDI,NE}^* > \gamma_{\text{closed}}^* \), so that the degree of IPR protection under the Nash equilibrium is stronger than in the closed economy. The intuition for this result is straightforward: with FDI and TT, multinational firms respond to stronger IPR protection by transferring more technology to their subsidiaries in the host country, which creates an additional incentive for the government to provide stronger IPR protection relative to the case in which the economy is closed to FDI.

When \( \lambda \) is large, however, the response of multinational firms’ TT decisions to changes in IPR policy are weaker (the elasticity of \( T^* \) with respect to \( \gamma \) is \( \frac{1}{\lambda} \)). Similarly, when \( 1 - \beta \) is small, the response of multinational firms’ TT decisions matter less for household welfare. Both these effects imply that the additional incentive for stronger IPR protection is weaker, leading to a lower value of \( \gamma_{FDI,NE}^* \). Comparative statics for \( \gamma_{FDI,NE}^* \) with respect to the other parameters of the model are the same as in the case with exports, as summarized by the following Proposition:

**Proposition 11.** In the open economy with FDI and TT by multinational firms, the level of IPR protection \( \gamma_{FDI,NE}^* \) that results in the non-cooperative Nash equilibrium when multinational firms choose the level of TT to maximize profits is:

1. strictly decreasing in the elasticity of the TT cost function (\( \lambda \)),

148
2. strictly increasing in the share of income spent on varieties invented abroad \((1 - \beta)\),

3. strictly increasing in the share of income spent on non-imitated varieties \((\mu)\),

4. strictly increasing in the degree of firm heterogeneity (as indexed by \(\frac{1}{\alpha}\)), and

5. strictly increasing in the elasticity of substitution across varieties \((\sigma)\).

Also, as in the case with exports, we see that \(\gamma_{FDI, GO}^* > \gamma_{FDI, NE}^*\), so that the IPR inefficiency again manifests itself here in the form of too little IPR protection in the Nash equilibrium relative to the global optimum. In addition to the IPR inefficiency, however, the possibility of FDI and TT introduces a second kind of inefficiency into the model, stemming from the choice of technology transfer \(T\). Note from equation (3.3.34) that the optimal choice of \(T_i\) from the perspective of the government in \(i\) is the value that maximizes \(h(\gamma_i, T_i)^\mu T_i^{1-\mu}\).

Evidently, this value is given by:

\[
T_{GO}^*(\gamma_i) = \left[\frac{\gamma_i}{c(\lambda \mu - 1)}\right]^{\frac{1}{\lambda}}
\]

which is interior and strictly increasing in \(\gamma_i\) under Assumption 11. However, multinational firms that carry out this technology transfer choose \(T_i = T^*(\gamma_i)\) only so as to maximize \(h(\gamma_i, T_i)\). Comparing the expressions for \(T^*\) and \(T_{GO}^*\) in equations (3.3.30) and (3.3.37) respectively, we then immediately see that \(T_{GO}^*(\gamma) > T^*(\gamma)\) for any value of \(\gamma \in (0, 1)\), with the discrepancy increasing in the share of income spent on imitated varieties \((1 - \mu)\).

In other words, what we will henceforth refer to as the TT inefficiency is characterized by too little technology transfer in the Nash equilibrium when multinational firms make their TT decisions to maximize profits, and governments take this profit-maximizing behavior as given. The intuition for this result is as follows. A higher level of TT by multinational firms not only makes the subsidiaries of these firms more productive, but also generates a positive externality by allowing successful imitators of these firms’ varieties to produce at a lower marginal cost. Both these effects reduce prices paid by consumers, but only the former
benefit is internalized by the multinational firm that makes the TT decision. Furthermore, the larger the share of income $1 - \mu$ that households spend on imitated goods, the greater the effect that this positive externality has on household welfare, and therefore the greater the welfare loss in the Nash equilibrium.

Now, it is straightforward to verify that the pair of values $(\gamma, T)$ which jointly minimizes the aggregate price index (3.3.34) is $(\gamma_{FDI,GO}^*, T_{GO}^* (\gamma_{FDI,GO}^*))$, so that the globally-optimal level of IPR protection is unaffected by removal of the restriction that multinational firms choose $T$ to maximize profits.\(^6\) Therefore, in studying the welfare losses to households that result from the IPR and TT inefficiencies, we can consider the following four cases. First, under the **global optimum**, governments mandate the globally-optimal level of TT and choose IPR policy cooperatively, resulting in $\gamma_{FDI,GO}^*$ and $T_{GO}^* (\gamma_{FDI,GO}^*)$. Second, under the **IPR + TT inefficiency** multinational firms choose the level of TT to maximize profits, and governments choose IPR policy non-cooperatively, resulting in $\gamma_{FDI,NE}^*$ and $T^* (\gamma_{FDI,NE}^*)$. Third, under the **IPR inefficiency** alone, governments mandate the globally-optimal level of TT, but choose IPR policy non-cooperatively, resulting in $\gamma_{FDI,NE}^*$ and $T_{GO}^* (\gamma_{FDI,NE}^*)$. Finally, under the **TT inefficiency** alone, multinational firms choose the level of TT to maximize profits, but governments choose IPR policy cooperatively, resulting in $\gamma_{FDI,GO}^*$ and $T^* (\gamma_{FDI,GO}^*)$.

Clearly, the first scenario yields the highest welfare to households in each of the two countries. Under the second scenario, both the IPR and TT inefficiencies generate welfare losses for households, while the third and fourth scenarios allow us to study the effects of the IPR or TT inefficiency in isolation. To quantify the magnitudes of these inefficiencies, we can then compute measures of the welfare losses relative to the global optimum (analogously to equations (3.3.23) and (3.3.24) for the case with exports) that arise from the IPR+TT, IPR

---

\(^6\)This is a special result that follows from the assumption of constant expenditure shares on varieties invented domestically and abroad. As we will see in Section 3.3.3, where we allow these expenditure shares to vary, allowing governments to jointly choose the welfare-maximizing pair $(\gamma, T)$ typically results in a different level of IPR protection than in the case when multinational firms choose $T$ to maximize profits and governments only choose $\gamma$ cooperatively.
and TT inefficiencies, which we denote by $\%\Delta V_{FDI}^{IPR+TT}$, $\%\Delta V_{FDI}^{TT}$, and $\%\Delta V_{FDI}^{IPR}$ respectively.

Figure 25, for example, shows how each of these measures vary with the share of income $\beta$ spent on varieties invented domestically, again setting $\alpha = 3.8$ and $\sigma = 3.8$. First, we see that all three inefficiency measures are decreasing in $\beta$. The intuition for this result should be clear from the preceding discussion: both the IPR and TT inefficiencies result from externalities across countries, so that the more households spend on foreign varieties, the more important these externalities are for household welfare. Again, straightforward differentiation of the inefficiency measures shows that this comparative static result holds more generally, as summarized by the following Proposition:

Proposition 12. In the open economy with FDI and TT by multinational firms, the globally-optimal level of IPR protection $\gamma_{FDI,GO}^*$ is strictly greater than the Nash equilibrium level $\gamma_{FDI,NE}^*$. Also, for any given level of IPR protection $\gamma \in (0, 1)$, the globally-optimal choice of technology transfer $T_{GO}^*(\gamma)$ is strictly greater than the value $T^*(\gamma)$ that maximizes profits for multinational firms. The magnitude of the IPR+TT, IPR, and TT inefficiencies (as measured by $\%\Delta V_{FDI}^{IPR+TT}$, $%\Delta V_{FDI}^{IPR}$, and $%\Delta V_{FDI}^{TT}$ respectively) are strictly increasing in the share of income spent on varieties invented abroad $(1 - \beta)$.

In sum, the open economy model with FDI and TT highlights how the TT decisions of
multinational firms can lead to a new kind of inefficiency in which there is too little transfer of technology in the Nash equilibrium. Furthermore, as shown in Figure 25, the TT inefficiency can be comparable in magnitude to the IPR inefficiency. The results of this section therefore suggest that in addition to international agreements regarding IPR policies, there is scope for cross-country cooperation on technology transfer by multinational firms as well.66

### 3.3.3 General CES preferences with both exports and FDI

In this section, we study a more general version of the open-economy model in which we relax the assumption of constant expenditure shares on varieties invented domestically and abroad, and allow for the coexistence of exports and FDI - in essence, embedding the mechanisms for imitation, IPR protection, and TT described previously in a framework of trade and FDI similar to that of Helpman et al (2004). This will allow us to study how our previous results are changed, if at all, by these modifications to the model, as well as to study a richer set of comparative statics (with respect to the iceberg trade cost, for example). The added complexity of the model is purchased at the cost of analytical tractability, however, and we therefore solve the model numerically in this section.

We now suppose that household preferences take the following form:

\[
U_i = \left[ \frac{Q_i^M}{\mu} \right]^\mu \left[ \frac{Q_i^I}{1 - \mu} \right]^{1 - \mu} \tag{3.3.38}
\]

\[
Q_i^M = \left[ \int_{\omega \in \Omega_i^M} q_i^M(\omega)^{\frac{\sigma - 1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma - 1}}
\]

\[
Q_i^I = \left[ \int_{\omega \in \Omega_i^I} q_i^I(\omega)^{\frac{\sigma - 1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma - 1}}
\]

---

66Note from equation (3.3.34) that the level of TT in one country does not affect welfare in the other country. This implies that even if governments were allowed to mandate T unilaterally instead of cooperatively, they would still choose the globally-optimal level of TT for a given value of \( \gamma \), suggesting that international cooperation is irrelevant. However, this is a special result that follows from the assumption of constant expenditure shares on varieties invented domestically and abroad. As we will see in Section 3.3.3, under more general household preferences, a change in the level of TT in one country affects welfare in the other country as well.

152
Relative to the preference specification considered in Sections 3.3.1 and 3.3.2, the consumer therefore no longer differentiates between varieties based on the country in which they are invented. Consequently, expenditure shares on varieties invented domestically and abroad are variable and endogenously determined in the model.

Also, to allow for greater flexibility in the choices of parameter values (again at the cost of analytical tractability), we assume a slightly different parametric form for the TT cost function:

$$C(T) = \frac{cT^\lambda}{1 - T} \quad (3.3.39)$$

Note that the inclusion of $1 - T$ in the denominator of (3.3.39) implies that it is prohibitively costly for multinational firms to transfer the full extent of their technology to their subsidiaries. Hence, it will never be optimal for firms (or governments) to choose $T = 1$, which automatically keeps the model away from corner solutions even without the second part of Assumption 11. It is straightforward to verify that as long as $\lambda > 1$, the cost function is still strictly convex, and the elasticity $\varepsilon_C$ is still greater than 1 and strictly increasing in $T$. Therefore, from the first-order condition (3.3.28) for the profit-maximizing value of $T$ for multinational firms, we see that this specification of the cost function again implies that firms respond to stronger IPR protection by choosing a higher level of TT. Furthermore, although the elasticity $\varepsilon_C$ is no longer identical to $\lambda$, it is still strictly increasing in this parameter, and therefore $\lambda$ plays the same role in the model as before.

Now, as in Melitz (2003) and Helpman et al (2004), we will focus on equilibria of the model in which the following are true: (i) exporting firms are more productive than non-exporting firms, and (ii) the set of exporting firms and the set of firms engaging in FDI are both non-empty, with firms engaging in FDI being more productive than exporting firms. The following assumption ensures that there exists a non-empty set of values for $(\gamma_i, \gamma_j, T_i, T_j)$ such that the unique equilibrium of the model exhibits both these characteristics:\footnote{Specifically, $\frac{f^X}{f^Y \tau^{\sigma - 1}} > 1$ is necessary and sufficient to guarantee that exporting firms are more productive than non-exporting firms. For the second characteristic to be true, we need $1 < h(\gamma, T) \tau < \left( \frac{f^X}{f^Y} \right)^{\frac{1}{\sqrt{\tau}}}$.}

153
Assumption 12. $\frac{f^X}{f^Y} > 1$ and $\frac{f^F}{f^E} > 1$.

In the appendix, I then show that in such an equilibrium of the model, the aggregate price index in country $i$ is given by the following system of equations:

$$P_i = (P_i^M)^\mu (P_i^I)^{1-\mu} \quad (3.3.40)$$

$$P_i^M = (kf^D)^{\frac{1}{\sigma-1}} \frac{1}{\phi} \left[ \frac{1 - \gamma_i H_0^i}{\gamma_i [1 - \gamma_i \gamma_j H_0^i H_0^j]} \right]^{\frac{1}{\sigma}} \quad (3.3.41)$$

$$P_i^I = \rho \left( \frac{\gamma_i}{1 - \gamma_i} \right)^{\frac{1}{\sigma-1}} \left[ \frac{M_R + \gamma_j H_1^j}{M_R + \gamma_j H_2^j} \right]^{\frac{1}{\sigma-1}} P_i^M \quad (3.3.42)$$

Here, $H_0^i$, $H_1^i$, and $H_2^i$ are functions of $(\gamma_i, T_i)$ that are defined in the appendix for the sake of brevity, $P_i^M$ and $P_i^I$ are the aggregate price indices for non-imitated and imitated varieties consumed in $i$ respectively, and $M_R \equiv \frac{M_E^i}{M_E^j}$ is the ratio of the mass of entrants in $i$ to the mass of entrants in $j$:

$$M_R = \frac{1 - \gamma_j (H_0^0 + H_1^j) + \gamma_i \gamma_j H_0^0 H_1^1}{1 - \gamma_i (H_0^0 + H_1^i) + \gamma_i \gamma_j H_0^0 H_1^j} \quad (3.3.43)$$

Note that the price indices $P_i^M$ and $P_i^I$ (and hence household welfare in $i$) depend on IPR policies and TT decisions in both countries.

In what follows, we solve the model numerically to study in more detail the Nash equilibrium levels of IPR protection and TT, as well as the various measures of inefficiency discussed in Section 3.3.2.4. In doing so, we set the baseline parameter values for the model as specified in Table 5. As before, the choice of $\sigma$ is based on the estimates in Bernard et al (2003), and $\alpha$ is set so that $\alpha - (\sigma - 1) = 1$. The entry cost $f^E$ amounts to a rescaling of the mass of entrants, and therefore the model’s results are invariant to the specific value chosen.

Here, we follow Bernard et al (2007) and choose the fixed costs for domestic production and exporting to be equal to 5% of the entry cost, and the minimum productivity $\phi$ to be 0.2. The trade cost is chosen so that each country’s own trade share is approximately equal to 0.93 in the Nash equilibrium when firms choose $T$ to maximize profits, and the FDI fixed in both countries in equilibrium, which is possible only if $\frac{f^E}{f^X} > 1$. 

154
cost $f^X$ is then adjusted so that there are both exports and FDI in equilibrium. The scale parameter of the TT cost function $\bar{c}$ is set so that the ratio of total TT costs to revenues for the subsidiaries of multinational firms is approximately 2%. The population size $L$ is a normalization that does not affect the computational results, and so we set this to 1 without loss of generality. Finally, we set $\mu = 0.6$ and $\lambda = 10$ as baseline values based on the numerical results depicted in Figures 24 and 25, and examine the comparative statics with respect to these parameters.

3.3.3.1 Numerical results

How are the comparative static results for the open economy discussed in Sections 3.3.1 and 3.3.2 (summarized by Propositions 9-12) changed when we allow for variable expenditure shares on domestic and foreign varieties, as well as both exports and FDI simultaneously? To address this question, we first solve the model at the baseline parametrization for each of the four cases considered in Section 3.3.2.4. Specifically, to compute the global optimum, we set $\gamma_i = \gamma_j = \gamma$ and $T_i = T_j = T$, and minimize the aggregate price index in equation (3.3.40) with respect to $(\gamma, T)$. To compute the equilibrium under the IPR+TT inefficiency, we set $T_i = T^*(\gamma_i)$ and $T_j = T^*(\gamma_j)$, and solve for the (symmetric) Nash equilibrium in $(\gamma_i, \gamma_j)$. To compute the equilibrium under the IPR inefficiency alone, we first set $T_i = T_j = T$ and compute the Nash equilibrium in $(\gamma_i, \gamma_j)$ for each value of $T$. We then find the value of $T$ that minimizes the aggregate price index. Finally, to compute the equilibrium under the TT inefficiency alone, we set $\gamma_i = \gamma_j = \gamma$ and $T_i = T_j = T^*(\gamma)$, and minimize the aggregate price index with respect to $\gamma$.

Figure 26 shows the objective functions for the global optimum, IPR inefficiency, and TT inefficiency equilibria, as well as the best response functions for the IPR+TT inefficiency case.

---

68Branstetter et al (2006) report statistics from the Bureau of Economic Analysis (BEA) and the World Intellectual Property Rights Organization (WIPO), showing that for the period 1982-1999, subsidiaries of US multinationals spent about 1% of sales revenues on royalty payments for technology licenses from their parent firms. In the model studied here, there is no direct counterpart to royalty payments, but the estimate of 1% suggests that total TT costs as a fraction of revenues should be around this number, and slightly higher.
Figure 26: Objective Functions and Best Response Functions for the Open Economy Model with Exports and FDI

These plots verify that in each case, the objective function exhibits a unique minimum within the allowable parameter space. Note also that the best response functions for the IPR+TT inefficiency equilibrium are downward-sloping at the point of intersection, indicating that the Nash equilibrium is stable and that IPR protection in one country is a strategic substitute for IPR protection in the other. The latter result is a new feature of the model with variable expenditure shares on domestic and foreign varieties, and to understand why it obtains, recall that the ex-ante profits from serving the foreign market depends on the strength of IPR protection both at home and abroad, whereas ex-ante profits from serving the domestic market depends only the strength of IPR protection domestically. Therefore, when the strength of IPR protection in the foreign market is low, a marginal increase in IPR protection at home increases the market share of home firms in the foreign market by a larger amount, which creates a stronger incentive for IPR protection by the home government.

Now, for each of the four types of equilibria described above, we investigate how the level of IPR protection $\gamma$ and the level of technology transfer $T$ vary with changes in several key parameters of the model. These results are summarized in Figures 27 and 28 respectively.
We also compute the size of the welfare losses under the IPR+TT, IPR and TT inefficiencies relative to the global optimum, in this case normalized by the overall gains from openness to trade and FDI, $1 - \frac{V_{closed}}{V_{GO}}$, and study the comparative statics for these measures. These results are summarized in Figure 29.

First, consider how the level of IPR protection $\gamma$ varies with the parameters of the model. Our previous results showed that the level of $\gamma$ in the Nash equilibrium when governments choose IPR policy non-cooperatively and multinational firms make TT decisions to maximize profits is: (i) increasing in the share of income spent on foreign varieties, and (ii) increasing in $\mu$, $\frac{1}{\alpha}$, and $\sigma$. With respect to the first result, we find similar behavior when we examine the comparative statics in response to changes in the iceberg trade cost $\tau$. As this parameter increases, the foreign trade share decreases, and as shown in Figure 27, the equilibrium level of $\gamma$ also decreases (whether under the global optimum or with the IPR and TT inefficiencies present). Similarly, Figure 27 displays comparative statics with respect to $\mu$, $\alpha$ and $\sigma$ that are qualitatively identical to previous results.\footnote{The only contrasting comparative static result here is with regard to variation in the parameter $\lambda$: we now find that as this parameter increases, the equilibrium level of IPR protection increases instead of decreases. This result obtains even if we use the same TT cost function as in Section 3.3.2.}

Next, we examine how the level of TT varies with the parameters of the model. Whenever the TT inefficiency is present (i.e. when multinational firms choose $T$ to maximize profits alone), variations in $T$ are tied to variations in $\gamma$ through equation (3.3.28). Figure 28 shows, however, that when this restriction is removed so that the TT inefficiency is absent, the globally-optimal level of $T$ can move in an opposite direction to the profit-maximizing value. As the share of income spent on non-imitated varieties ($\mu$) increases, for example, the globally-optimal value of $T$ decreases even though the profit-maximizing value increases. The intuition for this result is clear from the fact that the TT inefficiency stems from the positive externality that the transfer of technology has on successful imitating firms.

Now, is there still a sense in which we can say that the IPR and TT inefficiencies exhibit under- or over-provision of IPR protection and TT respectively? Recall that in Sections
3.3.1 and 3.3.2, the IPR inefficiency was always characterized by too little IPR protection by governments, and the TT inefficiency was always characterized by too little transfer of technology by multinational firms. Here, Figures 27 and 28 show that both results still hold in this version of the model: in equilibria where the IPR inefficiency is absent, the globally-optimal value of $\gamma$ is always strictly higher than the value that is chosen non-cooperatively by governments, and in equilibria where the TT inefficiency is absent, the globally-optimal value of $T$ is always strictly higher than the profit-maximizing value chosen by multinational firms.

Finally, we study in more detail the magnitudes of the IPR+TT, IPR and TT inefficiencies relative to the global optimum. Although the absolute size of the inefficiencies turns out to be rather small (on the order of 0.1% of welfare or less), it is also typically true that in models of this sort (see Arkolakis et al (2012)), the gains from trade are themselves small to begin with. For example, in our model, the overall gains from openness to trade and FDI amount to only about 2% of welfare, and as Figure 29 shows, the welfare losses from the IPR and TT inefficiencies appear to be as much as 4% of these gains. Furthermore, we see that the size of the TT inefficiency is comparable to (and in fact typically larger than) the size of the IPR inefficiency.

How do the magnitudes of the inefficiencies vary with the parameters of the model? Here, there are two main takeaways from Figure 29. First, note that for certain parameter values, it is possible for the magnitude of the IPR inefficiency to be zero even when governments choose IPR policy non-cooperatively. This result can best be understood with reference to the above discussion about how it is possible for the IPR inefficiency to be characterized by either too much or too little IPR protection, because weaker IPR protection in one country exerts a negative externality on the other country, but so too does stronger IPR protection. When these two effects exactly balance each other, the resulting IPR inefficiency is zero. As can be seen by comparing Figures 27 and 29, the parameter values at which the size of the IPR inefficiency is zero coincide with the values at which the non-cooperative value of $\gamma$ is
The second important takeaway is that the size of the overall IPR+TT inefficiency can vary non-monotonically with changes in the model’s parameters, which is largely due to the fact that the size of IPR efficiency varies non-monotonically for the reasons discussed above. We see, for example, that the magnitude of the IPR+TT inefficiency is largest for moderate values of $\tau$ and extreme (very small or very large) values of $\alpha$, $\mu$ and $\sigma$. In general, this complex behavior emphasizes that there are multiple general equilibrium effects at work, and that focusing on one mechanism in isolation can lead to misinformed conclusions about the optimal level of IPR protection or TT by multinational firms.
Figure 28: Comparative Static Results for the Level of Technology Transfer $T$
Figure 29: Comparative Static Results for the Magnitudes of the IPR and TT Inefficiencies
3.4 Conclusion

In this paper, I have developed a structural model of trade and foreign direct investment that allows for endogenously determined levels of imitation, protection of intellectual property rights, and technology transfer by multinational firms. I have then shown that equilibria of this model can exhibit two distinct kinds of inefficient behavior. The first kind - the IPR inefficiency - results when governments choose IPR policy without internalizing its effects on welfare in the other country. The second kind - the TT inefficiency - results when multinational firms choose the level of technology transfer only to maximize profits, without internalizing the effects of TT on the marginal costs of successful imitating firms in host country.

The model predicts that in the Nash equilibrium, both the strength of IPR protection and the degree of TT by multinational firms are increasing in the iceberg cost of trade, the degree of firm heterogeneity, the elasticity of substitution across varieties consumed by the household, and the fraction of income spent on imitated varieties. Furthermore, the both the IPR and TT inefficiencies are always characterized by under-provision of IPR protection by governments and of technology transfer by multinational firms respectively.

A simple calibration of the model suggests that the welfare losses from these two inefficiencies are small in absolute terms, but can amount to 5 – 10% of the overall gains from trade. Furthermore, numerical solutions of the model show that the magnitude of the TT inefficiency can be comparable to the magnitude of the IPR inefficiency, which implies that it is important to understand not only the incentives for governments to protect intellectual property, but also how these incentives interact with the incentives for multinational firms engaging in FDI to transfer technology to their subsidiaries.

The model developed in this paper does not consider the transition dynamics and the attendant implications for welfare that would result from starting in one of the hypothesized inefficient equilibria and switching to the global optimum. Study of this issue would most likely require a different kind of model that allows for greater flexibility in the integration of
time paths following a change in IPR or TT policy. Also, we have ignored wage effects in this model, and allowing for endogenously-determined wages might introduce additional, possibly countervailing, insights into the interaction between IPR policy and TT by multinational firms. Nonetheless, the model developed here suggests that in addition to international agreements on intellectual property rights such as TRIPs, international cooperation on cross-border technology transfer can achieve significant Pareto improvements as well.
Appendix A

Appendix to Chapter 1

A.1 Computational Algorithms

A.1.1 Static algorithm

Given the matching function $m$, the static market equilibrium specified in Definition 1 can be solved for easily using the following algorithm.

1. Make initial guesses $\hat{\Phi}$ and $\hat{\Delta}$ for the network productivity and quality functions, and iterate on equations (1.3.19) and (1.3.20) until convergence.

2. Solve for $\Delta_H$ using equations (1.3.10) and (1.3.34).

3. Compute the allocation $\{l(\chi), X(\chi), x(\chi, \chi'), x_H(\chi)\}_{\chi \in S_\chi}$ using (1.3.28), (1.3.31), (1.3.33), and (1.3.37) respectively.

Since the functional equations (1.3.19) and (1.3.20) constitute contraction mappings with Lipschitz constants $\left(\frac{\sigma}{\mu}\right)^{\sigma-1}$ and $\frac{\sigma-1}{\mu^\sigma}$ respectively, the iteration procedure in step 1 of the algorithm is guaranteed to converge at those rates. In practice, numerical solution of the model requires discretization of the state space $S_\chi$ into a mesh grid, of say $N_{grid} \times N_{grid}$ points. One can then solve for the functions $\Phi(\cdot)$ and $\Delta(\cdot)$ in step 1 at each point in the mesh grid, and then use bilinear interpolation to obtain numerical approximations of these functions as well as of the allocations $\{L(\chi), X(\chi), x(\chi, \chi'), x_H(\chi)\}$ for any desired value of $\chi \in S_\chi$.

A.1.1.1 Dynamic algorithm

I first describe the computational algorithm used to solve for the steady-state equilibrium specified in Definition 3, which is as follows.
1. Make initial guesses \( \hat{\Phi} \) and \( \Delta H \) for the network productivity function and the network quality function scaled by the household demand shifter.

2. Compute the implied profit function \( \tilde{\pi} \) from equation (1.4.4).

3. Compute the implied matching and acceptance functions, \( \tilde{m} \) and \( \tilde{a} \), from equations (1.4.3) and (1.4.11).

4. Compute the implied network productivity and quality functions, \( \tilde{\Phi} \) and \( \tilde{\Delta} \), from equations (1.3.19) and (1.3.20).

5. Compute the implied household demand shifter \( \tilde{\Delta} H \) from equations (1.3.34), (1.4.14), (1.4.15), and (1.4.16), and obtain the implied guess for the scaled network quality function, \( \Delta H \Delta = \tilde{\Delta} H \tilde{\Delta} \).

6. Compute the residual \( \mathcal{R} \equiv \max \{ \mathcal{R}_\Phi, \mathcal{R}_\Delta \} \) where

\[
\mathcal{R}_\Phi \equiv \max_{\chi \in S_\chi} \left| \hat{\Phi}(\chi) - \tilde{\Phi}(\chi) \right|
\]

\[
\mathcal{R}_\Delta \equiv \max_{\chi \in S_\chi} \left| \Delta H \Delta(\chi) - \Delta H \Delta(\chi) \right|
\]

and if \( \mathcal{R} > \epsilon \) for some tolerance level \( \epsilon \), update the guesses for the network productivity and scaled quality functions according to \( \hat{\Phi} = \frac{\hat{\Phi} + \tilde{\Phi}}{2} \) and \( \Delta H \Delta(\chi) = \frac{\Delta H \Delta(\chi) + \Delta H \tilde{\Delta}}{2} \), and repeat from step 1 until \( \mathcal{R} \leq \epsilon \).

I now discuss the computational algorithm used to solve for the model’s transition dynamics as specified in Definition 2. Suppose that the matching and profit functions at date 0 are given by \( m_0 \) and \( \pi_0 \) respectively, and that the economy is not in steady-state. The goal is to solve for the model’s transition path to the eventual steady-state characterized by the matching function denoted by \( m_{ss} \). Note that given the matching function \( m_t \), it is straightforward to solve for the static market equilibrium at date \( t \) using the algorithm discussed in section A.1.1. The challenge in solving the model’s transition dynamics therefore lies in computing
the matching function at date \( t \) given the matching function at date \( t - 1 \). As we see from equation (1.4.13), doing so while fully taking into account firm rational expectations requires solving for the profit functions \( \{ \pi_{t+s} \}_{s \geq 0} \). To accomplish this, I employ an algorithm that iterates on the path of profit functions \( \{ \pi_t \}_{t=1}^{T} \) for some value of \( T \) large enough such that the matching function at date \( T \) is close enough to the eventual steady-state matching function \( m_{ss} \). Formally, the algorithm is as follows.

1. Make a guess \( \hat{T} \) for the number of periods that it takes for convergence to the steady-state.

2. Make an initial guess for the profit functions \( \{ \hat{\pi}_t \}_{t=2}^{\hat{T}} \) (e.g. \( \hat{\pi}_t = \frac{1}{2}(\pi_0 + \pi_{ss}) \) for all \( t \in \{2, \cdots, \hat{T}\} \)).

3. At each date \( t \in \{1, \cdots, \hat{T}\} \), given \( \hat{m}_{t-1} \) (with \( m_0 = \hat{m}_0 \)):
   
   (a) Make initial guesses \( \hat{\Phi}_t \) and \( \Delta_t \Delta_t \) for the network productivity function and the network quality function scaled by the household demand shifter.

   (b) Compute the implied profit function \( \hat{\pi}_t \) from equation (1.4.4).

   (c) Compute the implied acceptance function \( \hat{a}_t \) (1.4.11), setting \( \pi_{t+s} = \hat{\pi}_{t+s} \) for \( s \in \{1, \cdots, \hat{T} - t\} \) and \( \pi_{t+s} = \pi_{ss} \) for \( s > \hat{T} - t \).

   (d) Compute the implied matching function \( \hat{m}_t \) from equation (1.4.2).

   (e) Compute the implied network productivity and quality functions, \( \hat{\Phi}_t \) and \( \hat{\Delta}_t \), from equations (1.3.19) and (1.3.20).

   (f) Compute the implied household demand shifter \( \hat{\Delta}_{H,t} \) from equations (1.3.34), (1.4.14), (1.4.15), and (1.4.16), and obtain the implied guess for the scaled network quality function, \( \Delta_h \Delta_t = \hat{\Delta}_{H,t} \Delta_t \).
(g) Compute the residual $R \equiv \max \{R_\Phi, R_\Delta\}$ where

$$R_\Phi \equiv \max_{\chi \in S_\chi} \left| \hat{\Phi}_t (\chi) - \tilde{\Phi}_t (\chi) \right|$$

$$R_\Delta \equiv \max_{\chi \in S_\chi} \left| \hat{\Delta}_H \Delta_t (\chi) - \tilde{\Delta}_H \Delta_t (\chi) \right|$$

and if $R > \epsilon$ for some tolerance level $\epsilon$, update the guesses for the network productivity and scaled quality functions according to

$$\hat{\Phi}_t'(\chi) = \frac{1}{2} \left[ \hat{\Phi}_t (\chi) + \tilde{\Phi}_t (\chi) \right]$$

$$\Delta_H \Delta_t'(\chi) = \frac{1}{2} \left[ \Delta_H \Delta_t (\chi) + \tilde{\Delta}_H \Delta_t (\chi) \right]$$

and repeat from step (a) until $R \leq \epsilon$, then set $\hat{m}_t = \tilde{m}_t$.

4. Compute the residual

$$R_\pi \equiv \max_{t \in \{2, \cdots, \hat{T}\}} \max_{(\chi, \chi') \in S_\chi^2} \left| \hat{\pi}_t (\chi, \chi') - \tilde{\pi}_t (\chi, \chi') \right|$$

and if $R_\pi > \epsilon_\pi$ for some tolerance level $\epsilon_\pi$, update the guesses for the profit functions according to $\hat{\pi}_t' = \frac{\hat{\pi}_t + \tilde{\pi}_t}{2}$ for all $t \in \{2, \cdots, \hat{T}\}$, and repeat from step 2 until $R_\pi \leq \epsilon$.

5. Compute the residual

$$R_m \equiv \max_{(\chi, \chi') \in S_\chi^2} \left| \hat{m}_T (\chi, \chi') - m_{ss} (\chi, \chi') \right|$$

and if $R_m > \epsilon_m$ for some tolerance level $\epsilon_m$, increment $\hat{T}$ and repeat from step 1.

As in solving for the static market equilibrium, numerical solution of the dynamic market equilibrium requires discretization of the state space $S_\chi$ into a mesh grid of $N_{grid} \times N_{grid}$ points, and bilinear interpolation can then be used to obtain numerical approximations of firm-level equilibrium variables off the grid points. Note that given the guess of future profit
functions, step 3 of the algorithm has the same computational complexity as solving for the model’s steady-state, and this part of the computation can be sped up by using the terminal guesses at the previous date when initializing the guesses for the network characteristic functions in step 3(a). Furthermore, upon increasing the guess for $\hat{T}$ to $\hat{T} + 1$ in step 5, the new guess for the profit functions up to date $\hat{T}$ used in step 2 can be set at the previous terminal guesses for the profit functions up to that date, which also speeds up the computation.

With a grid size of $N_{grid} = 20$ and tolerance levels $\epsilon = \epsilon_x = \epsilon_m = 10^{-4}$, executing the steady-state algorithm typically takes around 30 seconds, while solving for a transition path such as those discussed in the main text typically takes about one hour on a standard computer. Since estimation of the model’s parameters only requires solving for steady-state equilibria, the complexity of executing the dynamic algorithm does not factor into the tractability of estimating the model.

### A.2 Static and Dynamic Efficiency

#### A.2.1 Static efficiency

To characterize the efficiency of the static market equilibrium, I compare the resulting allocation with the allocation that would be chosen by a social planner whose goal is to maximize household welfare subject to the production technology and market clearing constraints. Given the matching function $m$, the social planner chooses the allocation $\mathcal{A} \equiv \left\{ l(\chi), X(\chi), \left\{ x(\chi, \chi') \right\}_{\chi' \in S\chi}, x_H(\chi) \right\}$ according to:

$$
U = \max_{\mathcal{A}} \left[ \int_{S\chi} \left[ \delta x_H(\chi) \right]^{\sigma - 1} dF_{\chi}(\chi) \right]^\frac{\sigma}{\sigma - 1}
$$

subject to the following constraints:
\[ X(\chi) = \left[ \phi l(\chi) \right]^{\sigma^{-1}} + \int_{S_\chi} m(\chi, \chi') \left[ \alpha x(\chi, \chi') \right]^{\sigma^{-1}} dF_\chi(\chi') \]  
(A.2.1)

\[ X(\chi) = x_H(\chi) + \int_{S_\chi} l(\chi', \chi) x(\chi', \chi) dF_\chi(\chi') \]  
(A.2.2)

\[
\int_{S_\chi} l(\chi) dF_\chi(\chi) = L - L_f \]  
(A.2.3)

where \( L_f = f \int_{S_\chi} \int_{S_\chi} m(\chi, \chi') dF_\chi(\chi) dF_\chi(\chi') \) is taken as given.

Denoting the Lagrange multipliers on constraints (A.2.2) and (A.2.3) by 
\[
\left( \frac{\nu}{\Delta H} \right)^{\frac{1}{\sigma}} \eta(\chi) f_\chi(\chi) \text{ and } \left( \frac{\nu}{\Delta H} \right)^{\frac{1}{\sigma}} \right] \text{ respectively, the first-order conditions for the planner’s problem can be expressed as:}
\]

\[ x_H(\chi) = \Delta_H \delta^{\sigma^{-1}} \eta(\chi)^{-\sigma} \]  
(A.2.4)

\[ l(\chi) = X(\chi) \eta(\chi)^{\sigma} \phi^{\sigma^{-1}} \]  
(A.2.5)

\[ x(\chi, \chi') = X(\chi) \eta(\chi)^{\sigma} \alpha^{\sigma^{-1}} \eta(\chi')^{-\sigma} \]  
(A.2.6)

Substituting these equations into (A.2.1) and (A.2.2), we get:

\[ \Phi(\chi) = \phi^{\sigma^{-1}} + \alpha^{\sigma^{-1}} \int_{S_\chi} m(\chi, \chi') \Phi(\chi') dF_\chi(\chi') \]  
(A.2.7)

\[ \Delta(\chi) = \delta^{\sigma^{-1}} + \alpha^{\sigma^{-1}} \int_{S_\chi} m(\chi, \chi') \Delta(\chi') dF_\chi(\chi') \]  
(A.2.8)

where \( \Phi(\chi) \equiv \eta(\chi)^{1-\sigma} \) and \( \Delta(\chi) \equiv \frac{1}{\Delta_H} X(\chi) \eta(\chi)^{\sigma} \).

Note that equations (A.2.4)-(A.2.8) are identical to equations (1.3.2), (1.3.7), (1.3.8), (1.3.19), and (1.3.20) respectively only when \( \mu = 1 \). This tells us that the static market equilibrium allocation is identical to the planner’s allocation if and only if the markups charged by all firms are equal to one. With a finite elasticity of substitution \( \sigma \), the static market equilibrium is therefore inefficient relative to the planner’s allocation because of the monopoly markup distortion.
A.2.2 Dynamic efficiency

To study the efficiency properties of the dynamic market equilibrium, we consider the problem of a social planner that chooses the set of relationships to activate and terminate at each date so as to maximize the present discounted value of household welfare, subject to the same dynamic frictions faced by firms in the market equilibrium. From the results in section A.2.1, we know that given the matching function \( m_t \) and the total mass of labor used to pay relationship costs \( L_{f,t} \), household utility at date \( t \) under the planner’s optimal allocation can be written as:

\[
U_t = (L - L_{f,t}) C_t
\]  
(A.2.9)

where \( C_t \) measures the total connectivity of the static production network:

\[
C_t \equiv \left[ \int_{S_\chi} \int_{S_\chi} \left[ \sum_{d=0}^{\infty} \alpha^d \delta \phi \left( \delta \phi' \right)^{\sigma-1} m_t^d \left( \chi, \chi' \right) F_\chi \left( \chi \right) dF_\chi \left( \chi' \right) \right] \right]^{\frac{1}{\sigma-1}}
= \left[ \int_{S_\chi} \Phi_t \left( \chi \right) \delta \phi^{\sigma-1} dF_\chi \left( \chi \right) \right]^{\frac{1}{\sigma-1}}
= \left[ \int_{S_\chi} \Delta_t \left( \chi \right) \phi^{\sigma-1} dF_\chi \left( \chi \right) \right]^{\frac{1}{\sigma-1}}
\]  
(A.2.10)

and \( \Phi_t \) and \( \Delta_t \) are given by the date \( t \) equivalents of equations (A.2.7) and (A.2.8) respectively.

To study the planner’s dynamic optimization problem, let \( V_t (m_{t-1}) \) denote the present value of discounted household utility at date \( t \) under the planner’s optimal dynamic allocation when the matching function in the previous period is given by \( m_{t-1} \). At each date \( t \), the planner’s choice about which relationships to activate and terminate is equivalent to a choice over the values \( \{ \xi_{max,t} (\chi, \chi') \} \in S_\chi^2 \), where \( \xi_{max,t} (\chi, \chi') \) specifies the maximum value of the idiosyncratic relationship cost shock component for which \( \chi - \chi' \) firm pair relationships
are accepted. The Bellman equation for the planner’s problem can therefore be written as:

\[ V_t(m_{t-1}) = \max_{\{\xi_{\max,t}(\chi,\chi')\}} \left[ U_t + \beta V_{t+1}(m_t) \right] \]  \hspace{1cm} (A.2.11)

where the maximization is subject to \( \xi_{\max,t}(\chi,\chi') \geq 0 \) for all \( t \) and \( (\chi,\chi') \in S^2_\chi \), as well as the following constraints:

\[ U_t = (L - L_{f,t}) C_t \]  \hspace{1cm} (A.2.12)

\[ C_t = \left[ \int_{S_\chi} \Phi_t(\chi) \delta^{\sigma-1} dF_\chi(\chi) \right]^{\frac{1}{\sigma-1}} \]  \hspace{1cm} (A.2.13)

\[ \Phi_t(\chi) = \phi^{\sigma-1} + \alpha^{\sigma-1} \int_{S_\chi} m_t(\chi,\chi') \Phi_t(\chi') dF_\chi(\chi') \]  \hspace{1cm} (A.2.14)

\[ L_{f,t} = f \int_{S_\chi} \int_{S_\chi} \nu m_{t-1}(\chi,\chi') dF_\chi(\chi) dF_\chi(\chi') \]  \hspace{1cm} (A.2.15)

\[ + f (1 - \nu) \int_{S_\chi} \int_{S_\chi} \int_{0}^{\xi_{\max,t}(\chi,\chi')} \xi dF_\xi(\xi) dF_\chi(\chi) dF_\chi(\chi') \]

\[ m_t(\chi,\chi') = \nu m_{t-1}(\chi,\chi') + (1 - \nu) F_\xi \left[ \xi_{\max,t}(\chi,\chi') \right] \]  \hspace{1cm} (A.2.16)

For brevity, denote \( \xi_{\max,t}^* \equiv \xi_{\max,t}(\chi^*,\chi'^*) \) and \( m_t^* \equiv m_t(\chi^*,\chi'^*) \) for a given firm pair \((\chi^*,\chi'^*)\). The first step in solving the dynamic planner’s problem is to find an expression for the derivative of \( U_t \) with respect to \( \xi_{\max,t}^* \). First, we differentiate (A.2.15) with respect to \( \xi_{\max,t}^* \) to get:

\[ \frac{dL_{f,t}}{d\xi_{\max,t}^*} = (1 - \nu) H(\chi^*,\chi'^*,\xi_{\max,t}^*) f_{\xi_{\max,t}} \]  \hspace{1cm} (A.2.17)

where \( H(\chi,\chi',\xi) \equiv f_\chi(\chi) f_\chi(\chi') f_\xi(\xi) \) is the product of three probability densities. Next, differentiating (A.2.16) for \((\chi,\chi') = (\chi^*,\chi'^*)\) with respect to \( \xi_{\max,t}^* \) gives:

\[ \frac{dm_t^*}{d\xi_{\max,t}^*} = (1 - \nu) f_\xi(\xi_{\max,t}^*) \]  \hspace{1cm} (A.2.18)
Differentiating the functional equation (A.2.8) with respect to $\xi^*_{\text{max},t}$, we then obtain:

\[
\frac{d\Phi_t (\chi)}{d\xi^*_t} = \frac{d\Phi_t (\chi) \ dm^*_t}{dm^*_t \ d\xi^*_t} = (1 - \nu) f_\xi (\xi^*_t, t) \ a^{\sigma-1} \Phi_t (\chi^{'}) \ 1_{\chi^*} (\chi) \tag{A.2.19}
\]

\[
+ (1 - \nu) f_\xi (\xi^*_t, t) \ a^{\sigma-1} \int_{\chi^*} m_t (\chi, \chi^{'}) \ d\xi^*_t \ dF_t (\chi^{'})
\]

\[
= (1 - \nu) H (\chi^*, \chi^{'}, \xi^*_t, t) \left[ \sum_{d=0}^{\infty} a^{d(\sigma-1)} m_t^{(d)} (\chi, \chi^{'}) \right] a^{\sigma-1} \Phi (\chi^{'}) \tag{A.2.20}
\]

where $1_{\chi^*} (\chi)$ is the indicator function that equals 1 if $\chi = \chi^*$ and 0 otherwise. (Note that equation (A.2.21) summarizes the effect of a change in the mass of connections between $\chi^* - \chi^*$ firm pairs on the network productivities of all firms that are downstream of $\chi^*$ firms.)

Differentiating equation (A.2.12) with respect to $\xi^*_{\text{max},t}$ and using (A.2.17) and (A.2.21), we then get:

\[
\frac{dU_t}{d\xi^*_t} = (1 - \nu) H (\chi^*, \chi^{'}, \xi^*_t, t) C_t \left[ \tilde{\pi}_t (\chi^*, \chi^{'}) - f^*_t \right] \tag{A.2.22}
\]

where we have defined:

\[
\tilde{\pi}_t (\chi^*, \chi^{'}) \equiv \frac{a^{\sigma-1}}{\sigma - 1} \Delta_{H,t} \Delta_t (\chi^*) \Phi_t (\chi^{'}) \tag{A.2.23}
\]

Note that conditional on the network characteristic functions, $\tilde{\pi}_t$ differs from the profit function $\pi_t$ in the dynamic market equilibrium (given by equation (1.4.4)) only by a constant fraction $\mu^{-\sigma}$.

The next step in solving the planner’s problem is to derive an expression for the derivative of the continuation value $V_{t+1} (m_t)$ with respect to $\xi^*_{\text{max},t}$. First, we note that:

\[
\frac{dV_{t+1}}{d\xi^*_t} = (1 - \nu) f_\xi (\xi^*_t, t) \frac{dV_{t+1}}{dm^*_t} \tag{A.2.24}
\]

The envelope condition then gives us:

\[
\frac{dV_{t+1}}{dm^*_t} = \frac{dU_{t+1}}{dm^*_t} + \beta \nu \frac{dV_{t+2}}{dm^*_t} \tag{A.2.25}
\]
Using the same approach as in solving for \( \frac{dU_t}{d\xi_{\text{max},t}} \), it is straightforward to show that:

\[
\frac{dU_{t+1}}{d\nu \tilde{m}_t} = \nu f(\chi^\ast) f(\chi'^\ast) C_{t+1} \left[ \tilde{\pi}_{t+1} \left( \chi^\ast, \chi'^\ast \right) - f \right]
\] (A.2.26)

Combining (A.2.24), (A.2.25) and (A.2.26), we then obtain:

\[
\frac{dV_{t+1}}{d\xi_{\text{max},t}} = \nu (1 - \nu) H \left( \chi^\ast, \chi'^\ast, \xi_{\text{max},t}^\ast \right) \sum_{s=0}^{\infty} (\beta \nu)^s C_{t+1+s} \left[ \tilde{\pi}_{t+1+s} \left( \chi^\ast, \chi'^\ast \right) - f \right]
\] (A.2.27)

Piecing together equations (A.2.22) and (A.2.27), we can finally write the first-order condition with respect to \( \xi_{\text{max},t} \left( \chi, \chi' \right) \) in the planner’s problem as:

\[
\xi_{\text{max},t} \left( \chi, \chi' \right) = \max \left\{ \frac{\tilde{\pi}_t (\chi, \chi')}{f} + \sum_{s=1}^{\infty} (\beta \nu)^s \left( \frac{C_{t+s}}{C_t} \right) \left[ \frac{\tilde{\pi}_{t+s} (\chi, \chi')}{f} - 1 \right], 0 \right\}
\] (A.2.28)

### A.3 Model Extensions

#### A.3.1 Multiple industries

To introduce multiple industries into the model, we can partition the set of firms \( \Omega \) into \( N \) subsets of equal mass and allow the input suitability parameter \( \alpha \) to vary across industry pairs. This variation in input suitability captures how “upstream” or “downstream” one industry is relative to another, and allows the model to match industry-level input-output tables. Assuming that the distribution of fundamental firm characteristics is the same in all industries and denoting by \( \alpha_{uv} \) the suitability of inputs from industry \( v \) for use in producing goods in industry \( u \), the analogs of equations (1.3.19) and (1.3.20) in steady-state are then:

\[
\Phi_u (\chi) = \phi^{-1} + \frac{1}{N} \sum_{v=1}^{N} \left( \frac{\alpha_{uv}}{\mu} \right)^{-1} \int_{S_{\chi}} m_{uv} (\chi, \chi') \Phi_v (\chi') dF_{\chi} (\chi')
\] (A.3.1)

\[
\Delta_u (\chi) = \mu^{-1} \delta^{-1} + \frac{1}{N} \sum_{v=1}^{N} \mu^{-\sigma} \alpha_{vu}^{-1} \int_{S_{\chi}} m_{vu} (\chi', \chi) \Delta_v (\chi') dF_{\chi} (\chi')
\] (A.3.2)
where now the network productivity and quality functions $\Phi_u$ and $\Delta_u$ are industry-specific, and the matching function $m_{uv}$ is industry-pair-specific. The matching function for each industry pair can in turn be computed using the corresponding version of equation (1.4.11).

Given the network characteristic functions for each industry and the matching function for each industry pair, we can then use equation (1.3.32) to calculate input-output shares. The share of industry $u$’s inputs that are sourced from industry $v$, for example, is given by:

$$S^{I}_{uv} = \frac{\alpha_{uv}^{-1} \int_{S_x} \int_{S_{\chi'}} m_{uv}(\chi,\chi') \Delta_u(\chi) \Phi_v(\chi') dF_{\chi}(\chi) dF_{\chi}(\chi)}{\sum_{w=1}^{N} \alpha_{uw}^{-1} \int_{S_x} \int_{S_{\chi'}} m_{uw}(\chi,\chi') \Delta_u(\chi) \Phi_w(\chi') dF_{\chi}(\chi) dF_{\chi}(\chi)}$$ (A.3.3)

while the share of industry $u$’s intermediate sales that accounted for by customers in industry $v$ is:

$$S^{O}_{uv} = \frac{\alpha_{uv}^{-1} \int_{S_x} \int_{S_{\chi'}} m_{vu}(\chi,\chi') \Delta_u(\chi) \Phi_u(\chi') dF_{\chi}(\chi) dF_{\chi}(\chi)}{\sum_{w=1}^{N} \alpha_{uw}^{-1} \int_{S_x} \int_{S_{\chi'}} m_{wu}(\chi,\chi') \Delta_u(\chi) \Phi_u(\chi') dF_{\chi}(\chi) dF_{\chi}(\chi)}$$ (A.3.4)

### A.3.2 Customer-supplier bargaining and cost-sharing

In this section, I discuss how the model’s assumptions can be modified to allow for a more general split of both the relationship surplus and the relationship fixed cost between the buying and selling firm.

First, note that without loss of generality, we can write the prices charged by a $\chi$-firm to the household and to a potential $\chi'$-buyer as markups $\mu_H(\chi)$ and $\mu(\chi,\chi')$ respectively over the seller’s marginal cost $\eta(\chi)$. The system of equations defining the network productivity and quality functions in the static market equilibrium can then be written as:

$$\Phi(\chi) = \phi^{\sigma-1} + \alpha^{\sigma-1} \int_{S_{\chi'}} \mu(\chi,\chi')^{1-\sigma} m(\chi,\chi') \Phi(\chi') dF_{\chi}(\chi')$$ (A.3.5)

$$\Delta(\chi) = \mu_H(\chi)^{-\sigma} \delta^{\sigma-1} + \alpha^{\sigma-1} \int_{S_{\chi'}} \mu(\chi',\chi)^{-\sigma} m(\chi',\chi) \Delta(\chi') dF_{\chi}(\chi')$$ (A.3.6)
while the profit that a $\chi$-firm makes from its sales to a $\chi'$-firm is given by:

$$\pi(\chi, \chi') = \mu(\chi, \chi')^{-\sigma} \left[ \mu(\chi, \chi') - 1 \right] \alpha^{\sigma-1} \Delta_H \Delta(\chi) \Phi(\chi')$$  \hspace{1cm} (A.3.7)

Note also that the total profit of a $\chi$-firm can be written as:

$$\pi(\chi) = \Delta_H \tilde{\Delta}(\chi) \Phi(\chi)$$  \hspace{1cm} (A.3.8)

where

$$\tilde{\Delta}(\chi) \equiv \mu_H(\chi)^{-\sigma} \left[ \mu_H(\chi) - 1 \right] \delta^{\sigma-1}$$  \hspace{1cm} (A.3.9)

$$+ \alpha^{\sigma-1} \int_{S_\chi} \mu(\chi', \chi)^{-\sigma} \left[ \mu(\chi', \chi) - 1 \right] \Delta(\chi') dF_\chi(\chi')$$

depends only on variables relating to the $\chi$-firm’s customers.

Now suppose that instead of assuming a market structure characterized by monopolistic competition, we assume that firms take the markups charged by all other firms as given, and that the markup $\mu(\chi, \chi')$ is chosen to maximize the product $[v^C(\chi, \chi')]^\theta [v^S(\chi, \chi')]^{1-\theta}$. In other words, buyers and sellers engage in bilateral Nash bargaining (which we will soon see is equivalent to multilateral Nash bargaining in the static model), with $v^C(\chi, \chi')$ and $v^S(\chi, \chi')$ denoting the surplus to the customer and supplier respectively of the relationship between a $\chi$-buyer and a $\chi'$-seller. The parameter $\theta \in [0, 1]$ measures the bargaining power of the customer relative to the supplier.

From (A.3.5), (A.3.7), and (A.3.8), the surplus values can be written as:

$$v^C(\chi, \chi') = \mu(\chi, \chi')^{1-\sigma} \alpha^{\sigma-1} \tilde{\Delta}(\chi) \Phi(\chi')$$  \hspace{1cm} (A.3.10)

$$v^S(\chi, \chi') = \mu(\chi, \chi')^{-\sigma} \left[ \mu(\chi, \chi') - 1 \right] \alpha^{\sigma-1} \Delta(\chi) \Phi(\chi')$$  \hspace{1cm} (A.3.11)

Note that $\tilde{\Delta}(\chi)$ and $\Delta(\chi)$ depend only on interactions between the $\chi$-buyer and its own
customers, while $\Phi(\chi')$ depends only on interactions between the $\chi'$-seller and its own suppliers. In other words, because of the CES structure of the production function, the surplus of the relationship between a $\chi$-buyer and a $\chi'$-seller is independent of the interactions between the buying firm and its other suppliers, and is also independent of the interactions between the selling firm and its other customers. As a result, bilateral Nash bargaining is equivalent in the model to the multilateral generalization of Nash bargaining proposed in Stole and Zwiebel (1996).

From equations (A.3.10) and (A.3.11), it is then straightforward to verify that firms again charge a constant markup over marginal cost, but that this markup is now given by:

$$
\mu = \frac{\sigma - \theta}{\sigma - 1} \quad (A.3.12)
$$

Note that when all bargaining power resides with the supplier ($\theta = 0$), the markup charged is the same as that under monopolistic competition, whereas when all bargaining power resides with the buyer ($\theta = 1$), the markup is the same as that under perfect competition. In general, we have $\mu \in [1, \frac{\sigma}{\sigma-1}]$. Furthermore, if we assume that firms sell to households indirectly via a unit continuum of retailers that produce differentiated varieties of a retail good, and that sales between producers and retailers are characterized by the same bargaining process, then the same analysis as above can be used to rationalize markups for final sales that are also constant and given by (A.3.12).

We can also allow for a more general split of relationship costs between buyers and sellers by assuming that the buying firm pays a constant fraction $b$ of the fixed cost in each relationship. In this case, whether a potential relationship is mutually desired by both buyer and seller depends on how the respective cost shares compare to the surplus values (A.3.10) and (A.3.11). Supposing that firms' pricing decisions remain characterized by constant markups equal to $\mu$, it is straightforward to verify that a relationship is mutually desirable if and only if profits from that relationship are at least greater than an effective fixed cost.
given by:

\[ f_{\text{eff}} \equiv f \max \{b\mu, 1 - b\} \]  \hspace{1cm} (A.3.13)

Note that the effective fixed cost is minimized when \( b = \frac{1}{\mu + \tau} \). This implies that relationships are more likely to form if selling firms pay a larger share of relationship costs whenever the markups that they charge are also higher.

Through these additional assumptions, the model therefore allows for richer variation in inter-firm markups and effective relationship costs. It is important to point out, however, that these assumptions about bargaining and cost-sharing become much more restrictive once embedded in the dynamic model with endogenous network formation. For example, the characterization of the dynamic model discussed in the main text remains valid with buyer-supplier Nash bargaining only if we rule out repeated bargaining between potential buyer-supplier pairs. The possibility of transfers between buyers and sellers also needs to be ruled out once the fixed relationship cost is taken into account. Furthermore, as discussed in the main text, once the buying firm pays a positive share of the relationship cost, constant markup pricing is not necessarily optimal for all firms in the dynamic model. For these reasons, I retain monopolistic competition as the assumed market structure and set \( b = 0 \) in the main model, and leave development of richer models of bargaining and cost-sharing under the setting of sticky relationships for future work.
Appendix B

Appendix to Chapter 2

B.1 Proofs of Propositions

B.1.1 Proof of Proposition 4

Consider an active relationship with state \((\epsilon, a)\). The buyer-seller pair in this relationship chooses prices \(\{p^D, p^U\}\), quantities \(\{x^D, x^U\}\), labor inputs \(\{l^D, l^U\}\), and the producer’s share of the fixed operating cost \(s\), so as to maximize the (log) generalized Nash product:

\[
\max_{p^D, p^U, x^D, x^U, l^D, l^U, s} \left\{ \nu \log R^D + (1 - \nu) \log R^U \right\}
\]  

(B.1.1)

where we have omitted time subscripts and dependence on the state \((\epsilon, a)\) for brevity, and imposed the free-entry condition (2.3.12). This maximization is subject to the final demand function (2.3.1):

\[x^D = A (p^D)^{-\sigma},\]  

(B.1.2)

the upstream production technology (2.3.2):

\[x^U = \epsilon \kappa (a) l^U,\]  

(B.1.3)

and the downstream production technology (2.3.4):

\[x^D = \left( \frac{l^D}{\alpha} \right)^\alpha \left( \frac{x^U}{1 - \alpha} \right)^{1 - \alpha} \frac{1}{\tau^{1 - \alpha}}\]  

(B.1.4)

where we have used the market clearing condition for intermediates (2.3.5) to substitute \(\frac{1}{\tau} x^U\) for \(x^M\).
Furthermore, note from (2.3.8) and (2.3.9) that the value of the relationship to the downstream and upstream firms can respectively be written as:

\[ R_D = \pi_D + C^D \tag{B.1.5} \]
\[ R_U = \pi_U + C^U \tag{B.1.6} \]

where \( C^D \) and \( C^U \) are continuation values that do not depend on the choice variables for the maximization in (B.1.1). Recall also that the flow payoffs are given by:

\[ \pi_D = p^D x^D - w l^D - p^U x^U - sf \tag{B.1.7} \]
\[ \pi_U = p^U x^U - w l^U - (1 - s) f \tag{B.1.8} \]

To obtain the solution to the firms’ Nash bargaining problem, first define the effective transfer from the downstream to the upstream firm as:

\[ T \equiv p^U x^U + sf \tag{B.1.9} \]

Now, note that the first-order condition with respect to \( T \) is:

\[ \frac{\nu}{R^D} \frac{\partial \pi_D}{\partial T} + \frac{1 - \nu}{R^U} \frac{\partial \pi_U}{\partial T} = 0 \tag{B.1.10} \]

and furthermore:

\[ \frac{\partial \pi_D}{\partial T} + \frac{\partial \pi_U}{\partial T} = 0 \tag{B.1.11} \]

Equation (B.1.11) can be interpreted as follows: the choices of \( p^U \) and \( s \) are equivalent means of transferring profit between the downstream and upstream firms, and thus the choice of the effective transfer \( T \) is a zero-sum problem. Combining equations (B.1.10) and (B.1.11)
then gives us:

\[
\frac{\nu}{R^D} = 1 - \frac{\nu}{R^U}
\]  

(B.1.12)

and since the relationship surplus under free entry is given by \( S = R^D + R^U \), we then have:

\[
R^D = \nu S
\]  

(B.1.13)

\[
R^U = (1 - \nu) S
\]  

(B.1.14)

In other words, \( p^U \) and \( s \) are chosen such that the downstream producer and upstream supplier receive constant fractions \( \nu \) and \( 1 - \nu \) respectively of the total relationship surplus, with the specific values of \( p^U \) and \( s \) undetermined.

Given (B.1.12), the derivative of the objective function in (B.1.1) with respect to each remaining variable \( \chi \in \{ p^D, x^D, x^U, l^D, l^U \} \) can simply be expressed as \( \frac{\partial \pi}{\partial \chi} + \frac{\partial \pi^U}{\partial \chi} \). Denoting the Lagrange multipliers on constraints (B.1.2), (B.1.3), and (B.1.4) by \( \mathcal{L}^F \), \( \mathcal{L}^U \), and \( \mathcal{L}^D \) respectively, the remaining first-order conditions can then be written as:

\[
x^D = \sigma \frac{x^D}{p^D} \mathcal{L}^F
\]  

(B.1.15)

\[
p^D = \mathcal{L}^F + \mathcal{L}^D
\]  

(B.1.16)

\[
\mathcal{L}^U = (1 - \alpha) \frac{x^D}{x^U} \mathcal{L}^D
\]  

(B.1.17)

\[
w = \alpha \frac{x^D}{l^D} \mathcal{L}^D
\]  

(B.1.18)

\[
w = \epsilon \kappa (a) \mathcal{L}^U
\]  

(B.1.19)

Now from (B.1.15) and (B.1.16), we obtain:

\[
p^D = \mu \mathcal{L}^D
\]  

(B.1.20)
where $\mu \equiv \frac{\sigma}{\sigma - 1}$. Using (B.1.18) to substitute for $L^D$, we have:

$$p^D = \frac{\mu w l^D}{\alpha x^D}$$  \hspace{1cm} (B.1.21)

and using the downstream production function (B.1.4) to substitute for $x^D$ gives:

$$p^D = \mu w \tau^{1-\alpha} \left[ \frac{(1 - \alpha) l^D \gamma}{\alpha x^U} \right]^{1-\alpha}$$  \hspace{1cm} (B.1.22)

Next, using (B.1.19) to substitute for $L^U$ and the upstream production function (B.1.3) to substitute for $x^U$ in (B.1.17) gives:

$$L^D = \frac{1}{1 - \alpha} \left( \frac{w^U}{x^D} \right)$$  \hspace{1cm} (B.1.23)

Combining (B.1.23) with (B.1.18), we then obtain:

$$\alpha l^U = (1 - \alpha) l^D$$  \hspace{1cm} (B.1.24)

Substituting (B.1.24) into (B.1.22) and using the upstream production function (B.1.3) finally gives the optimal price for final goods sales as:

$$p^D = \mu w \left[ \frac{\tau}{\epsilon K(a)} \right]^{1-\alpha}$$  \hspace{1cm} (B.1.25)

Having solved for $p^D$, we can easily recover the remaining variables. Output of the downstream firm is given by (B.1.2) as:

$$x^D = \mu^{-\sigma} Aw^{-\sigma} \left[ \frac{\epsilon K(a)}{\tau} \right]^\sigma(1-\alpha)$$  \hspace{1cm} (B.1.26)
and total revenue from final sales is then:

\[ \pi^D = p^D x^D = \mu^{1-\sigma} Aw^{1-\sigma} \left[ \frac{\epsilon \kappa (a)}{\tau} \right]^{(\sigma-1)(1-\alpha)} \] (B.1.27)

Labor demand by the downstream firm is given by (B.1.21) as:

\[ l^D = \alpha \mu^{-\sigma} Aw^{-\sigma} \left[ \frac{\epsilon \kappa (a)}{\tau} \right]^{(\sigma-1)(1-\alpha)} \] (B.1.28)

and labor demand by the upstream firm follows from (B.1.24):

\[ l^U = (1 - \alpha) \mu^{-\sigma} Aw^{-\sigma} \left[ \frac{\epsilon \kappa (a)}{\tau} \right]^{(\sigma-1)(1-\alpha)} \] (B.1.29)

The total amount of profit generated by the relationship can then be derived from (B.1.7) and (B.1.8) as:

\[ \bar{\pi} (\epsilon, a) \equiv \pi^D + \pi^U = B \left[ \frac{\epsilon \kappa (a)}{\tau} \right]^b - f \] (B.1.30)

where \( B \equiv \frac{1}{\sigma} \mu^{1-\sigma} Aw^{1-\sigma} \) and \( b \equiv (\sigma - 1)(1 - \alpha) \).

Furthermore, with free entry and constant surplus shares, equations (2.3.8) and (2.3.9) allow us to write the value of the relationship to the downstream and upstream firms respectively as:

\[ R^D_t (\epsilon, a) = \pi^D_t (\epsilon, a) + \nu \beta (1 - \delta) \int_{\xi}^{\infty} \max \{ S_{t+1} (\epsilon', a + 1), 0 \} dF (\epsilon') \] (B.1.31)

\[ R^U_t (\epsilon, a) = \pi^U_t (\epsilon, a) + (1 - \nu) \beta (1 - \delta) \int_{\xi}^{\infty} \max \{ S_{t+1} (\epsilon', a + 1), 0 \} dF (\epsilon') \] (B.1.32)

Adding these values together consequently implies that the surplus of the relationship satisfies the following Bellman equation:

\[ S_t (\epsilon, a) = \bar{\pi}_t (\epsilon, a) + \beta (1 - \delta) \int_{\xi}^{\infty} \max \{ S_{t+1} (\epsilon', a + 1), 0 \} dF (\epsilon') \] (B.1.33)
and the split of surplus shares then tells us that profit within the relationship is also split in proportion to each firm’s bargaining power:

\[
\pi^D = \nu \bar{\pi} \quad \text{(B.1.34)}
\]

\[
\pi^U = (1 - \nu) \bar{\pi} \quad \text{(B.1.35)}
\]

This allows us to solve for the effective transfer value as:

\[
T = (b + 1 - \nu) B \left[ \frac{e^\kappa(a)}{\tau} \right]^b + \nu f \quad \text{(B.1.36)}
\]

Now, note that the total amount of revenue earned by the upstream firm from intermediate sales within the relationship is equal to:

\[
r^U = \bar{T} - sf \quad \text{(B.1.37)}
\]

\[
= (b + 1 - \nu) B \left[ \frac{e^\kappa(a)}{\tau} \right]^b + (\nu - s) f \quad \text{(B.1.38)}
\]

Since the value of \( s \) is indeterminate, the size of the upstream firm is also undetermined without further assumptions. However, from (B.1.38), we see that upstream firm size is proportional to downstream firm size if and only if \( s = \nu \). Therefore, under assumption 7, upstream firm size is given by:

\[
r^U = (b + 1 - \nu) B \left[ \frac{e^\kappa(a)}{\tau} \right]^b \quad \text{(B.1.39)}
\]

and the implied price charged by the upstream firm is a constant markup over its marginal cost:

\[
p^U = \left( \frac{b + 1 - \nu}{b} \right) \left[ \frac{w}{e^\kappa(a)} \right] \quad \text{(B.1.40)}
\]
B.1.2 Proof of Proposition 5

We will prove existence and uniqueness of the steady-state equilibrium by solving for it.

In a steady-state equilibrium, the surplus function satisfies the following Bellman equation:

\[
S(\epsilon, a) = B \left[ \frac{\epsilon \kappa(a)}{\tau} \right]^b - f + (1 - \delta) \int_{\xi}^{\xi} \max \{ S(\epsilon', a + 1), 0 \} dQ(\epsilon'|\epsilon) \quad (B.1.41)
\]

For ages \( a \geq \bar{a} \), there is no further growth in relationship capital, and therefore the surplus function no longer varies with age. We then have \( S(\epsilon, a) = \bar{S}(\epsilon) \) for all \( a \geq \bar{a} \), where \( \bar{S} \) satisfies:

\[
\bar{S}(\epsilon) = B \left( \frac{\epsilon \bar{\kappa}}{\tau} \right)^b - f + (1 - \delta) \int_{\xi}^{\xi} \max \{ \bar{S}(\epsilon'), 0 \} dQ(\epsilon'|\epsilon) \quad (B.1.42)
\]

Given the properties of the Markov efficiency process described in Assumption 3, the functional equation (B.1.42) is a contraction mapping. Therefore, given a value of \( B \), there exists a unique solution for \( \bar{S} \). Denote this solution by \( \bar{S}(\cdot|B) \).

Now given the solution for \( \bar{S} \), the surplus functions for all ages \( a < \bar{a} \) can then be computed by iterating backwards using (B.1.41). This yields in particular a solution for the surplus function at age 0, denoted by \( S(\cdot, 0|B) \). In any equilibrium, however, the expected surplus of a newly-formed relationship is pinned down by the free entry of downstream and upstream firms. A guess of the mass of downstream searching firms, \( d \), therefore implies a value of \( B \), denoted by \( B(d) \). This value must satisfy:

\[
\int_{\xi}^{\xi} \max \{ S(\epsilon', 0|B(d)), 0 \} dF_0(\epsilon') = \frac{\gamma^D(d)}{v\bar{m}} \quad (B.1.43)
\]

It is clear from (B.1.41) that the expected surplus at any age is negative as \( B \to 0 \) and approaches \(+\infty\) as \( B \to +\infty \). Therefore, for any value of \( d \), there exists a unique value \( B(d) \) that satisfies (B.1.43), and furthermore, \( B'(d) > 0 \).

Given \( d \), the unique solution for \( B(d) \) pins down the solution for the efficiency cutoff.
$\epsilon^* (a|d)$ at each age $a$. If the cutoff lies in the interior of the support of $\epsilon$, then it satisfies:

$$S [\epsilon^* (a|d), a|B (d)] = 0$$

(B.1.44)

and since the relationship surplus is strictly increasing in the stock of relationship capital, the efficiency cutoffs must be weakly decreasing in age. Furthermore, since the surplus functions are strictly increasing in $B$, the efficiency cutoffs are weakly increasing in $B$ and therefore weakly decreasing with the guess of $d$.

Finally, it remains to show that the masses of downstream and upstream searching firms, $d$ and $u$, are uniquely determined. Market tightness $\theta \equiv \frac{\mu}{\sigma}$ is completely determined by (2.3.31). Therefore, we only need to solve for $d$. Now define $\hat{n} (a|d) \equiv n (a)/d$. Then from (2.3.40), we have for all $a \geq 0$:

$$\hat{n} (a + 1|d) = \hat{n} (a|d) \left[ 1 - \Delta^{out} (a + 1|d) \right] (1 - \delta)$$

(B.1.45)

and from (2.3.41):

$$\hat{n} (0|d) = \bar{n} \left[ 1 - \Delta^{out} (0|d) \right]$$

(B.1.46)

where $\Delta^{out} (a|d)$ is the endogenous termination probability implied by the guess of $d$. The labor market clearing condition can then be rewritten as:

$$\frac{L}{\bar{d}} = \frac{\gamma^D (d)}{\nu} = \frac{\sigma}{\mu} B (d) \left[ \frac{\hat{Z} (d)}{\tau} \right]^b + f \hat{N} (d)$$

(B.1.47)
where:

\[ \hat{Z} (d) \equiv \left[ \sum_{a=0}^{\infty} \hat{n} (a|d) [\bar{\epsilon} (a|d) \kappa (a)]^b \right]^{\frac{1}{b}} \tag{B.1.48} \]

\[ \hat{N} (d) \equiv \sum_{a=0}^{\infty} \hat{n} (a|d) \tag{B.1.49} \]

and \( \bar{\epsilon} (a|d) \) is the average efficiency of relationships of age \( a \) implied by the guess of \( d \).

Now, since \( B' (d) > 0 \), then evidently \( \Delta^\text{out} (a|d) \) is declining in \( d \) for all \( a \), so that \( \hat{n} (a|d) \) is increasing in \( d \) for all \( a \). Therefore, \( \hat{N'} (d) \geq 0 \). Furthermore, note that

\[ \hat{n} (a|d) \bar{\epsilon} (a|d)^b = \bar{m} (1 - \delta)^a \prod_{s=0}^{a-1} \left[ 1 - \Delta^\text{out} (s|d) \right] \int_{\epsilon^*}^{\infty} \epsilon^b dF_a (\epsilon) \tag{B.1.50} \]

Since the efficiency cutoffs are also declining with \( d \), this implies that \( \hat{n} (a|d) \bar{\epsilon} (a|d)^b \) is increasing in \( d \) for any \( a \), and therefore \( \hat{Z'} (d) \geq 0 \). Existence and uniqueness of the steady-state equilibrium then follows from the observation that the right-hand side of (B.1.47) is weakly increasing in \( d \), while the left-hand side is strictly decreasing in \( d \) from \( +\infty \) to \( -\infty \).
Appendix C

Appendix to Chapter 3

C.1 Derivation of the Aggregate Price Index for the Case with General CES Preferences

When household preferences take the form specified in (3.3.38), demand for non-imitated and imitated varieties in country \( i \) can be written as:

\[
q^M_i(\omega) = \mu L p^M_i(\omega)^{-\sigma} (P^M_i)^{1-\sigma} \\
q^I_i(\omega) = \mu L p^I_i(\omega)^{-\sigma} (P^I_i)^{1-\sigma}
\]  
(C.1.1)  
(C.1.2)

where \( p^M_i(\omega) \) and \( p^I_i(\omega) \) are the prices charged by individual firms, and \( P^M_i \) and \( P^I_i \) are price indices for the non-imitated and imitated sectors that are defined as follows:

\[
P^M_i = \left[ \int_{\omega \in \Omega^M} p^M_i(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}} \\
P^I_i = \left[ \int_{\omega \in \Omega^I} p^I_i(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}}
\]  
(C.1.3)  
(C.1.4)

Per-capita utility is then given by:

\[
V_i \equiv \frac{U_i}{L} = P_i^{-1}
\]  
(C.1.5)

where \( P_i \) is an aggregate price index defined by:

\[
P_i = (P^M_i)^{\mu} (P^I_i)^{1-\mu}
\]  
(C.1.6)
Now, when monopolists choose prices to maximize profits subject to (C.1.1) and firms engaging in FDI are more productive than exporting firms, the free-entry condition for entering firms in country \(i\) can be written as:

\[
\gamma f^D \int_{s_i}^{\infty} \left[ \left( \frac{\varphi}{\varphi_i} \right)^{\gamma - 1} - 1 \right] dG(\varphi) + \gamma \gamma_j \left[ \int_{s_j}^{\infty} \left[ \left( \frac{\varphi}{\varphi_j} \right)^{\gamma - 1} - 1 \right] dG(\varphi) + f^P \int_{s_j}^{\infty} \left[ \left( \frac{\varphi}{\varphi_j} \right)^{\gamma - 1} - 1 \right] dG(\varphi) = f^E \quad \text{(C.1.7)}
\]

where \(\varphi_i\), \(\varphi_j\), and \(\varphi_j^X\) are the productivity cutoff values for firms from \(i\) for domestic production in \(i\), exporting to \(j\), and engaging in FDI in \(j\) respectively. These cutoff values are themselves given by:

\[
\begin{align*}
\varphi_i^* &= (zf^D)\frac{1}{\sigma - 1} (P_i^M)^{-1} \\
\varphi_j^X &= (zf^X)\frac{1}{\sigma - 1} \tau (P_j^M)^{-1} \\
\varphi_j^F &= (zf^F)\frac{1}{\sigma - 1} \left( \frac{1 - f^X/f^F}{h(\gamma_j, T_j)\sigma^{-1} - \tau^{1 - \sigma}} \right)^{\frac{1}{\sigma - 1}} (P_j^M)^{-1} \\
\end{align*}
\]

Clearly, for it to be true that exporting firms are more productive than non-exporting firms in the symmetric equilibrium, we require the first part of Assumption 12 to hold. For it to be true that there are both firms engaging in exporting and FDI with the latter being more productive than the former (that is, \(\varphi_i^* > \varphi_j^X\) and \(\varphi_j^* > \varphi_j^X\)), we require that \(1 < h(\gamma, T) \tau < \left( \frac{f^E}{f^P} \right)^{\frac{1}{\sigma - 1}}\) in equilibrium in both countries, which is possible only under the second part of Assumption 12. Then, substituting (C.1.8)-(C.1.10) into (C.1.7), and doing the same for the counterparts of equations (C.1.7)-(C.1.10) for country \(j\), we obtain a system of two equations in \(P_i^M\) and \(P_j^M\). Solving this system for \(P_i^M\), we obtain equation (3.3.41), where \(H_i^0\) is a function of \((\gamma_i, T_i)\) defined as follows:

\[
H_i^0 = (f^D)^{\frac{1}{\sigma - 1}} \gamma^{-\alpha} + (f^P)^{\frac{1}{\sigma - 1}} \left[ \frac{f^X}{f^P} \frac{1}{\sigma - 1} \frac{1}{\frac{1}{\gamma_j(T_j)\sigma^{-1} - \tau^{1 - \sigma}}} \right]^{\frac{1}{\sigma - 1}} \quad \text{(C.1.11)}
\]

To solve for the imitator price index \(P_i^I\), note that the definitions of the price indices and our assumptions about the marginal costs of successful imitators imply the following relation between \(P_i^M\) and \(P_i^I\):

\[
\left( \frac{P_i^I}{P_i^M} \right)^{1 - \sigma} = \frac{\gamma - \gamma_i M_i f^P E_i \int_{s_i}^{\infty} \varphi_i^{\sigma - 1} dG(\varphi) + (1 - \gamma_i) \gamma_j M_j f^P E_j \int_{s_j}^{\infty} \varphi_j^{\sigma - 1} dG(\varphi)}{\gamma M_i E_i \int_{s_i}^{\infty} (\rho_i \varphi)^{\sigma - 1} dG(\varphi) + \gamma_j M_j f^P E_j \int_{s_j}^{\infty} \varphi_j^{\sigma - 1} dG(\varphi) + (1 - \gamma_j) \gamma_i M_i f^P E_i \int_{s_i}^{\infty} \varphi_i^{\sigma - 1} dG(\varphi) + \gamma_j M_j f^P E_j \int_{s_j}^{\infty} \varphi_j^{\sigma - 1} dG(\varphi)}
\]

Substituting (C.1.9) and (C.1.10) into (C.1.12) and simplifying, we obtain equation (3.3.42),
where $H_1^i$ and $H_2^i$ are functions of $(\gamma_i, T_i)$ defined as follows:

\begin{align}
H_1^i &= \left(\frac{f^D}{f^F}\right)^{\frac{r-1}{\sigma}} r^{-\alpha} + \left(\frac{f^D}{f^F}\right)^{\frac{r-1}{\sigma}} \left[h(\gamma_i, T_i)^{\sigma-1} - r^{1-\sigma}\right]^{\frac{r}{\sigma}} \\
H_2^i &= \left(\frac{f^D}{f^F}\right)^{\frac{r-1}{\sigma}} (r^{(\sigma-1)-\alpha} + \left(\frac{f^D}{f^F}\right)^{\frac{r-1}{\sigma}} (\gamma_i^{\sigma-1} - 1) \left[\frac{1 - f^X}{f^F} h(\gamma_i, T_i)^{\sigma-1} - r^{1-\sigma}\right]^{\frac{r}{\sigma}}
\end{align}

(C.1.13)

(C.1.14)

To find the ratio of the mass of entrants $M_R$, we make use of the following relation between the monopolist price indices in $i$ and $j$:

\begin{equation}
\left(\frac{p_{RI}^M}{p_{RJ}^M}\right)^{1-\sigma} = \frac{\gamma_i M_{E}^i \int_{\rho \varphi}^{\infty} (\rho \varphi^{\sigma-1}) dG(\varphi) + \gamma_i \gamma_j M_{E}^i \left[f_{\gamma_i}^{\rho \varphi} \left(\frac{\varphi}{\rho}\right)^{\sigma-1} dG(\varphi) + \int_{\rho \varphi}^{\infty} [\varphi h(\gamma_i, T_i) \varphi]^{\sigma-1} dG(\varphi)\right]}{\gamma_j M_{E}^j \int_{\rho \varphi}^{\infty} (\rho \varphi^{\sigma-1}) dG(\varphi) + \gamma_i \gamma_j M_{E}^i \left[f_{\gamma_j}^{\rho \varphi} \left(\frac{\varphi}{\rho}\right)^{\sigma-1} dG(\varphi) + \int_{\rho \varphi}^{\infty} [\varphi h(\gamma_j, T_j) \varphi]^{\sigma-1} dG(\varphi)\right]}
\end{equation}

(C.1.15)

Substituting (C.1.8)-(C.1.10) and (3.3.41) as well as their counterparts for country $j$ into (C.1.15), we can solve for $M_R$. Doing so, we obtain equation (3.3.43).
Bibliography


Bernard, Andrew, Andreas Mo xnes, and Y ukiko Saito. 2015. “Production Networks, Geog-

Bernard, Andrew, Andreas Mo xnes, and Karen Helene Ulltveit-Moe. 2015. “Two-sided Hetero-
geneity and Trade.” Mimeo: Dartmouth University.

Bernard, Andrew, Stephen J Redding, and Peter Schott. 2007. “Comparative Advantage and


