Quantifying the Aggregate Effects of Inter-Firm Knowledge Networks

Chong Liu*[1] and Zi Wang†[2]

[1]Peking University

[2]Shanghai University of Finance and Economics

February 3, 2019

Abstract

This paper evaluates the impacts of inter-firm knowledge networks on aggregate growth and welfare. We use data on patent citations across Chinese manufacturing firms to study the structure of inter-firm knowledge networks. We document that (i) within a narrowly-defined industry firms still occupy heterogeneous positions in patent citation networks, and (ii) firms exposed to better technologies tend to grow faster. These empirical regularities motivate and guide the development of a quantifiable general equilibrium model with firm innovation and inter-firm knowledge diffusion. Using data on patent citations, we estimate the model's key parameters via a constrained simulated method of moments, suggesting that larger firms are more likely to learn from and be learned by other firms. Under our estimated inter-firm knowledge networks, we find that (i) a firm's impacts on the aggregate economy depend crucially on its position in inter-firm knowledge networks, and (ii) the welfare-maximizing innovation subsidy rate decreases with firm size. Extending our model to an open economy, we find that domestic producers can gain from trade liberalization by absorbing knowledge from exporters, which mitigates the inter-firm reallocation effect of trade emphasized by Melitz (2003).

Keywords: *Growth*; *Innovation*; *Knowledge Spillovers*; *Firm Heterogeneity*; *Trade.*
JEL classification: *O11*; *O33*; *O38*; *F62.*

*cliu.econ@pku.edu.cn
†wang.zi@mail.shufe.edu.cn

1 Introduction

Firm innovation and technology progress are key drivers to aggregate growth. It is well-documented that firms do not develop their technologies in isolation. Instead, they learn with each other via highly selective and unevenly spread knowledge networks (Jaffe et al., 1993; Giuliani, 2007). The structure of inter-firm knowledge networks affects innovation and technology progress of individual firms as well as aggregate technology evolution. The effectiveness of government polices targeted at stimulating firm innovation hinges on better understanding of the network nature of inter-firm knowledge spillovers. However, there is currently a lack of tractable theory that integrates the inter-firm knowledge network observed in the micro data into a macro-growth model in order to evaluate its aggregate dynamic effects, a gap this paper aims to fill.[1]

We have access to a rich data set for Chinese manufacturing firms that documents inter-firm citation relationships and the firm characteristics such as sales, employment, and R&D expenditure of both citing and cited firms. This allows us to characterize the structure of inter-firm knowledge networks and connect inter-firm knowledge diffusion with firm characteristics. The descriptive results show that (i) a large fraction of patent citations occur between firms in the same industry, (ii) within a narrowly-defined industry, firms still occupy heterogeneous positions in patent citation networks; and (iii) patent citations exhibit positive matching, *i.e.* larger and more connected firms are connected to firms that are also larger and more connected. We also find reduced-form evidence that a firm tends to grow faster if it cites patents from faster-growing firms. These findings highlight the importance of *inter-firm* knowledge networks to technology diffusion and improvements.

Motivated by these empirical regularities, we develop a structural model of firm innovation and knowledge diffusion. In the model, firms vary in innovation efficiency and randomly meet with each other. The probability of meeting and learning depends flexibly on individual characteristics of both learning and learned firms. This specification allows our model to capture the structure of inter-firm knowledge networks observed in the micro patent citation data. Through these inter-firm connections, firms innovate and learn with each other, which jointly determine the technology evolution of individual firms and aggregate economy. Despite rich firm heterogeneity and flexible knowledge networks, our model remains tractable and yields simple structural equations that characterize both the steady-state and transitional dynamics of the equilibrium. We provide sufficient conditions for the existence and

[1] In their recent work, Cai and Li (2018) develop a quantitative model of firm innovation and technology diffusion. But their empirical findings and quantification focus on inter-sector knowledge networks.

uniqueness of steady-state and develop simple algorithms to compute the model's steady-state as well as transitional dynamics.

We then estimate the model's key parameters using data on patent citations across Chinese manufacturing firms. To achieve this, we develop a simulated methods of moments framework which regards equilibrium conditions as constraints. This framework is useful in estimating general equilibrium models with rich heterogeneity and interdependence across individuals. Our estimates indicate strong positive matching assortativity of inter-firm knowledge spillovers: larger and more connected firms learn from and are learned by firms that are also larger and more connected. Using the sensitivity measure of estimates with estimation moments developed by Andrews et al. (2017), we find that our estimates on parameters of inter-firm knowledge networks rely closely on the positive matching assortativity in the data of patent citations. We also find that taking inter-firm knowledge spillovers into consideration is crucial for our model to replicate the empirical firm size distribution in the Chinese firm-level data.

Armed with our estimates, we simulate the model to study how firm-level instantaneous shocks propagate via inter-firm knowledge networks and translate into aggregate dynamic effects. The key findings are as follows. First, eliminating inter-firm knowledge spillovers reduces the steady-state productivity as well as real income by more than 50%, which suggests that inter-firm knowledge networks are quantitatively important to aggregate productivity and welfare. Second, we compute the elasticity of aggregate welfare with respect to idiosyncratic technology shocks. Unlike Hulten (1978), we find that this elasticity is not proportional to firm size, but depends crucially on the firm's position in inter-firm knowledge networks.[2] Third, in the presence of inter-firm knowledge spillovers, it is optimal for the government to subsidize the innovation in all firms, with subsidy rates *decreasing* with firm size. There are two conflicting forces that determine optimal innovation subsidies: (i) larger firms diffuse more knowledge to other firms, implying that they should be subsidized more; (ii) larger firms receive more knowledge from other firms, indicating that they gain more from the subsidies received by other firms. In our quantification practice, the latter dominates the former.[3]

[2]Hulten (1978) has shown that in the efficient economy, the elasticity of real value-added with respect to labor productivity of a set of firms is equal to these firms' share in aggregate sales. Our model is not efficient due to positive externalities of innovation. However, we show that without inter-firm knowledge networks, the elasticity of aggregate welfare with technology shocks of individual firms is still proportional to their sales.

[3]Bloom et al. (2013) consider only the former force that larger firms generate more technology spillovers. So they conclude that the optimal innovation subsidies should target on large firms.

Finally, we study how inter-firm knowledge networks affect the productivity and welfare consequences of trade liberalization. To achieve this, we combine our baseline model with the trade model developed by Melitz (2003). As in Melitz (2003), trade incurs both iceberg and fixed costs, and only the most productive firms are engaged in international trade. Therefore, trade liberalization leads to *reallocation across firms*: large exporters expand, innovate more, and improve their productivities (Aghion et al., 2018), whereas small domestic producers shrink and innovate less. Interestingly, we that inter-firm knowledge networks *mitigate* this reallocation effect since the domestic producers can indirectly gain from trade by learning from exporters.

How important is this mitigation effect to welfare gains from trade? Here we separately discuss the gains from trade, i.e. the welfare changes by moving from the observed trade shares to autarky (as in Costinot and Rodrguez-Clare (2014)), and the gains from trade liberalization, i.e. the welfare effect from a decline in trade costs. First, we find that inter-firm knowledge networks *reduce* welfare gains from trade.[4] Without inter-firm knowledge networks, firms benefit from export markets only through direct exporting. In this case, the move back to autarky leads to substantial productivity losses. Inter-firm knowledge networks can reduce these losses by diffusing the exporters' technologies to domestic producers and thus reduce the gains from trade. Second, we find that inter-firm knowledge networks *magnifies* welfare gains from trade liberalization. Intuitively, inter-firm knowledge diffusion enables more firms to overcome fixed export costs, inducing more export entry under the same trade liberalization and thereby magnifying the gains from trade liberalization.

In quantifying aggregate effects of technology spillovers, this paper is most closely related to the spatial growth models developed by Desmet and Rossi-Hansberg (2014) and Desmet, Nagy, and Rossi-Hansberg (2018). In both of these models, as in this paper, the firm's innovation decision is a repeated static problem, which leads to tractable innovation process in equilibrium. However, these models focus on the geography of technology spillovers and do not consider the network feature of inter-firm knowledge diffusion, which is the focus of this paper. Moreover, Cai and Li (2018) incorporate inter-sectoral knowledge linkages into the growth model developed by Klette and Kortum (2004) and estimate the model using patent citation data. This paper, instead, focuses on the aggregate implications of inter-firm knowledge linkages. To achieve this, we allow flexible inter-firm knowledge connections to capture rich patterns in the firm-level patent citation data and develop a novel constrained simulated method of moments to recover inter-firm knowledge networks from the data.

[4] Making this argument, we compare the gains from trade implied by different models that are consistent with the same data.

In modeling the structure of knowledge diffusion, this paper builds on an extensive literature of technology spillovers. Bloom, Schankerman, and Van Reenen (2013) model and identify inter-firm technology spillovers. However, their model is stylized and cannot be used for quantification. Acemoglu, Akcigit, and Kerr (2017) depict the structure of innovation networks using 1.8 million U.S. patents and their citation properties. Moreover, Alvarez, Buera, and Lucas (2014) and Buera and Oberfield (2017) build general equilibrium models with technology spillovers via trade. This paper, so far as we are aware of, is the first attempt to incorporate inter-firm knowledge networks into a quantifiable general equilibrium framework.

In modeling networks across firms, this paper is related to the literature of inter-firm production and trade networks. Lim (2018), Oberfield (2018), Acemoglu and Azar (2018), and Tintelnot et al. (2018) build tractable models of input-output linkages across firms, arguing that firm- and sector-level shocks propagate via production networks and lead to aggregate fluctuations (see also Acemoglu et al., 2012). Moreover, Bernard, Moxnes, and Saito (2015) and Bernard, Moxnes, and Ulltveit-Moe (2015) characterize firm-to-firm trade linkages. This paper builds on the techniques that model firm-to-firm networks used in these models, in particular Lim (2018). However, we depart from these studies by focusing on inter-firm knowledge networks, which are by nature engaged in externalities and leave room for government interventions.

In addition, this paper contributes to trade models with firm heterogeneity. As in Melitz (2003), trade liberalization leads to reallocation across firms. However, since the firm innovation depends on market size, trade does not only affect the firms' market share but also directly affect the firms' productivity distribution, similar to the recent work by Aghion et al. (2018). The framework developed in this paper combines firm innovation with inter-firm knowledge diffusion and thereby allows us to study the implications of technology spillovers for gains from trade and trade liberalization.

The outline of this paper is as follows. We begin in Section 2 by documenting patterns of inter-firm knowledge networks in the data on patent citations across Chinese manufacturing firms. Then in Section 3, we develop a model of firm innovation and knowledge diffusion that characterizes how firm size, innovation, and aggregate productivity evolution depend on inter-firm knowledge networks. In Section 4, we bring our model to data on firm performance and patent citation and estimate the model's key parameters. Armed by the estimated model, we conduct counterfactual exercises to quantify the aggregate effects of inter-firm knowledge networks in Section 5. Section 6 concludes.

2 Technology Improvements via Knowledge Networks: Data and Evidence

The purposes of this section are two-folded. First, we introduce the data on patent citations across Chinese manufacturing firms and document some key features of patent citation networks that guide our model specification and structural estimation. Second, we provide reduced-form evidence showing that a firm tends to grow faster if it cites patents from faster-growing firms. In other words, the patent citation network is a good proxy for inter-firm knowledge networks.

2.1 Data Sources

The data used in this section come from two datasets. The first is Annual Survey of Chinese Manufacturers (ASCM), which collects performance data on Chinese manufacturing firms whose annual sales exceed 5 millions RMB (about 0.6 million U.S. dollars) over 1998-2013.

The second dataset is based on patent information recorded by China Intellectual Property Office, combined with patent citations from Google Patents. We match patents and patents with firms in ASCM by firm name and contact information. The details on our matching algorithm and summary statistics of this dataset are presented in our online appendix.

We construct our database as follows. First, we focus on firms' long-term average performances such as sales, employment, and productivity. We average these variables over the whole sample period 1998-2013 for each firm.[5] Second, we consider patents that were granted over 1998-2007, looking at the citations of these patents over the whole sample period 1998-2013. Doing this, we address the problem that patent citations grow with the age of patents so newly-granted patents may not draw enough attention. Within this sample, we say that firm i cites patents from firm j if firm i has cited at least one patent from firm j over 1998-2013. We document the empirical regularities in Section 2.2 and conduct structural estimation in Section 4 using this cross-sectional dataset.

[5]Sales are deflated by Producer Price Index.

2.2 Structure of Inter-Firm Patent Citation Networks

The Chinese patent citation data allows us to construct citation networks among Chinese firms, and therefore document a set of facts on heterogeneity of citing and cited firms and their relationships. We let these facts guide our model of inter-firm knowledge networks and subsequent structural estimation.

Fact 1: A large fraction of patent citations occur between firms in the same industry. Within a narrowly-defined industry, firms still occupy heterogeneous positions in patent citation networks.

# Firm pairs with patent citations	85471
In which:	
Cited and citing firms in the same 2-digit CIC industry	36939
Cited and citing firms in the same 3-digit CIC industry	22012

Table 1: Patent Citation Linkages across Chinese Manufacturing Firms: 1998-2007

Table 1 shows that about 43% of firm pairs that have patent citation relationships in our data set are within the same 2-digit CIC industry. Even if we are looking at the 3-digit CIC industries which are narrowly defined, intra-industry citation linkages still account for about a quarter of total inter-firm citation linkages.

Figure 1 visualizes inter-firm patent citation networks within CIC industry 395, the manufacturing of household appliances (typical products are refrigerator and air conditioner). A node represents one firm, with the node size proportional to the firm's long-term average sales. An arrow represents a citation linkage from one firm to another.[6] It shows that within a narrowly-defined industry, firms still occupy heterogeneous positions in inter-firm knowledge networks, with large firms in the center and small firms on the periphery. All these patterns highlight that it is important to understand inter-firm, instead of merely inter-industry, knowledge networks.

Fact 2: The distributions of patent citations are characterized by many firms with few connections and a few firms with many connections.

We plot the number of firms each firm cites, namely *in-degree*, against the fraction of firms citing at least that many firms (Panel (a) of Figure 2). We find that the distribution is largely consistent with a Pareto distribution as the CDF is close to linear, except in the tail, with many firms citing very few firms and a few firms citing many firms.

[6] A citation linkage means that A firm has cited at least one patent from B firm over 1998-2013.

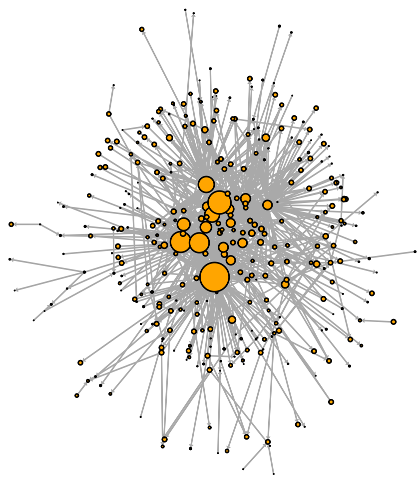
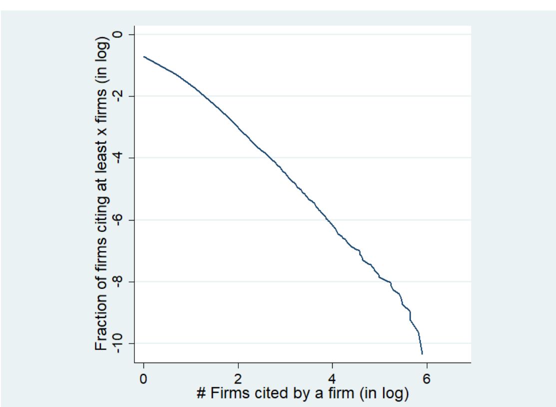
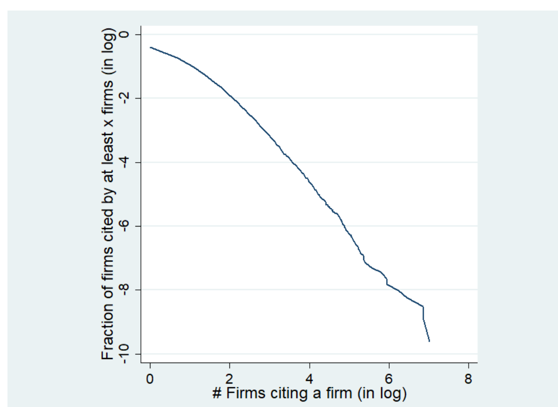


Figure 1: Inter-firm Patent Citation Networks, CIC 395 (Household Appliances)

(Notes: 1998-2013 data (averaged). Each node represents a firm in CIC 395, the manufacturing of household appliances, with the node size proportional to long-term average firm sales. Each arrow represents that firm i cites patents from firm j.)



(a) In-Degree



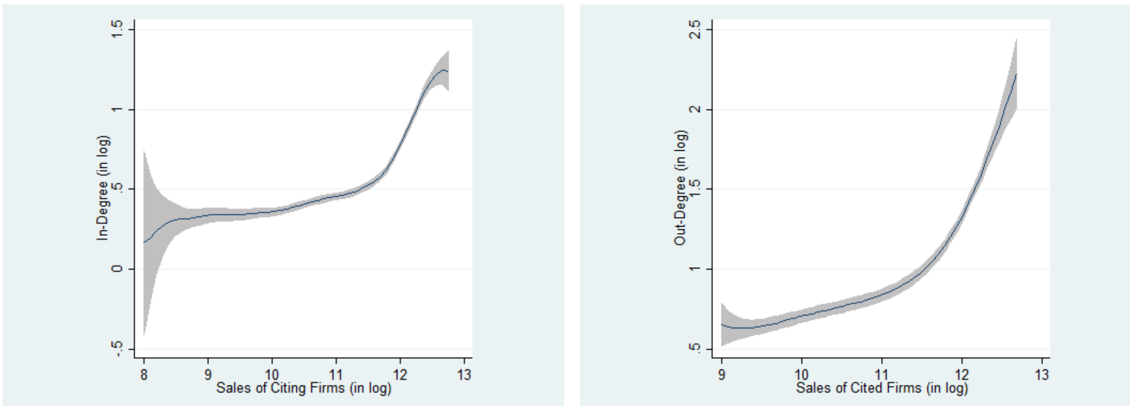
(b) Out-Degree

Figure 2: Distributions of In- and Out-Degree

(Notes: 1998-2013 data (averaged). In-degree is defined as the number of firms from which a firm cites. Out-degree is defined as the number of firms citing a firm. Panel (a) plots in-degree x (in log) against $1 - CDF(x)$ (in log). Panel (b) does it for out-degree.)

We also plot the number of firms citing each firm, namely *out-degree*, against the fraction of firms cited by at least that many firms (Panel (b) of Figure 2). Again the distribution is approximated by Pareto, except in the tail, with many firms cited by very few firms and a few firms cited by many firms.

Fact 3: Larger firms cite more firms and are cited by more firms.



(a) In-Degree

(b) Out-Degree

Figure 3: Firm Sales and In-/Out-Degree

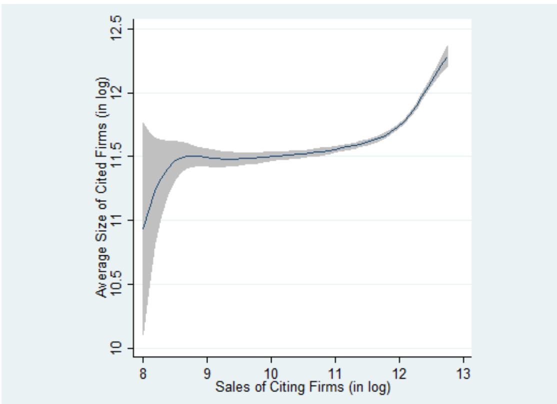
(Notes: 1998-2013 data (averaged). In-degree is defined as the number of firms from which a firm cites. Out-degree is defined as the number of firms citing a firm. The figure plots a firm's long-term average sales (in log) against its in- and out-degree (in log). The solid line is the fit from a kernel-weighted local polynomial regression, and the gray area is the 99 percent confidence interval.)

Figure 3 plots the relationship between a firm's in-/out-degree and its long-term average sales. The solid line is the fit from a kernel-weighted local polynomial regression, and the gray area is the 99 percent confidence interval. The results suggest that larger firms cite from more firms and are also cited by more firms. Intuitively, larger and more productive firms tend to create better patents and be more aware of other firms' patents, which posit them at the center of the patent citation network.

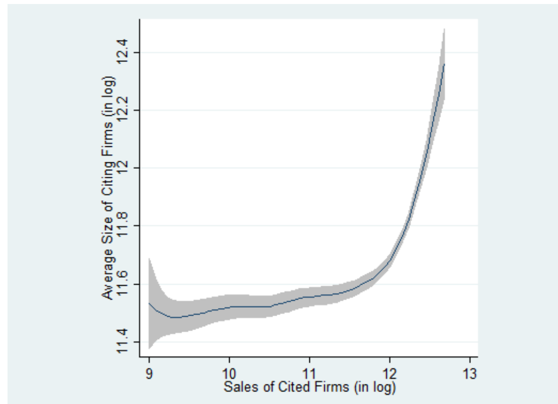
Fact 4: Larger and more connected firms are connected to firms that are also larger and more connected.

Figure 4 depicts the matching assortativity between firms in the patent citation network. Panel (a) of Figure 4 plots a firm's sales against the average sales of firms it cites from. Again, the solid line is the fit from a kernel-weighted local polynomial regression, and the gray area is the 99 percent confidence interval. Likewise, Panel (b) of Figure 4 plots a firm's sales against the average sales of firms that cite its patents. The results suggest that larger firms tend to cite patents from larger firms and also be cited by larger firms.

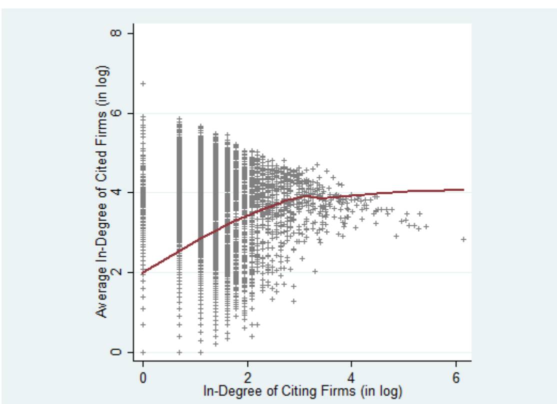
Moreover, Panel (c) of Figure 4 plots a firm's in-degree against the average in-degree



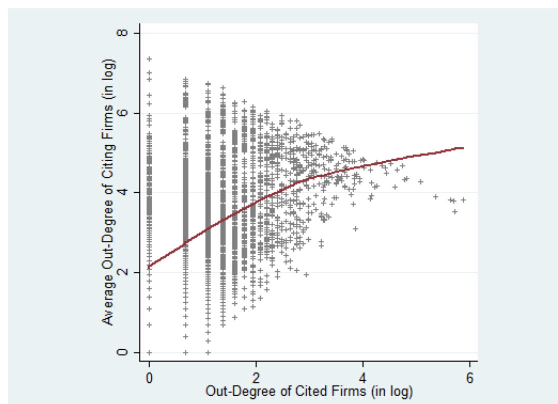
(a) Sales Assortativity (citing firms)



(b) Sales Assortativity (cited firms)



(c) In-Degree Assortativity



(d) Out-Degree Assortativity

Figure 4: Matching Assortativity (Sales and Degree)

(Notes: 1998-2013 data (averaged). In-degree is defined as the number of firms from which a firm cites. Out-degree is defined as the number of firms citing a firm. Panel (a) and (b) plot a firm's long-term average sales (in log) against the average long-term sales of firms it cites and are cited by (in log). The solid line is the fit from a kernel-weighted local polynomial regression, and the gray area is the 99 percent confidence interval. Panel (c) and (d) plot of a firm's in-degree (out-degree) against the average in-degree (out-degree) of firms it cites (is cited). The red line is the fit from a locally weighted smoothing regression.)

of firms it cites from. The red line is the fit from a locally weighted smoothing regression. It shows that a firm that cites from more firms tends to cite patents from firms that also cite from more firms. Similarly, Panel (d) of Figure 4 plots a firm's out-degree against the average out-degree of firms that cite its patents. The result shows that a firm cited by more firms tends to be cited by firms that are also cited by more firms. In sum, measured both by sales and connectedness, the firm matching in the patent citation network exhibits positive assortativity.

2.3 Technology Progress along Citation Networks

Empirical regularities in Section 2.2 showed that the inter-firm patent citation network is unevenly spread and highly structural. However, it is still unclear that whether, and to what extent, this patent citation network approximates the inter-firm knowledge network on which this paper focuses. To investigate these questions, we examine how a firm's technology improvements depend on the performances of firms whose patents it has cited, controlling for its individual shocks as well as sectoral and regional shocks.

To be more precise, we consider the following regression:

$$\Delta \log \phi_{it} = \rho^{\mathrm{NX}} \Delta \log \phi_{it}^{\mathrm{NX}} + \rho^{C} \log X_{it} + fe_t + fe_s + fe_p + \epsilon_{it}, \tag{1}$$

where i refers to individual firm i, t refers to year between 1998 and 2013, and Δ denotes the change in the variable from year $t-1$ to t. ϕ_{it} is the productivity measure for firm i at period t. We consider various commonly-used productivity measures and the results are robust. X_{it} is a vector of firm-level controls. We also control for year, 3-digit CIC industry, and provincial fixed effects to exclude the effects of common shocks.

Our key explanatory variable, $\Delta \log \phi_{it}^{\mathrm{NX}}$, is constructed as follows:

$$\Delta \log \phi_{it}^{\mathrm{NX}} = \log \left(\sum_{j \in \mathbf{C}_i} w_{jt} \phi_{jt} \right) - \log \left(\sum_{j \in \mathbf{C}_i} w_{jt-1} \phi_{jt-1} \right), \tag{2}$$

where $\mathbf{C}_i$ is the set of firms that was cited by firm i over the whole sample period 1998-2013. We intentionally fix the patent citation network over this period, examining to what extent technology improvements can diffuse from firm j to firm i via this fixed network. w_{jt} is the weight of the cited firm j. We consider various weights and the results are robust. This idea of fixed network is in line with the shift-share instrument developed by Autor, Dorn, and

Hanson (2013).[7]

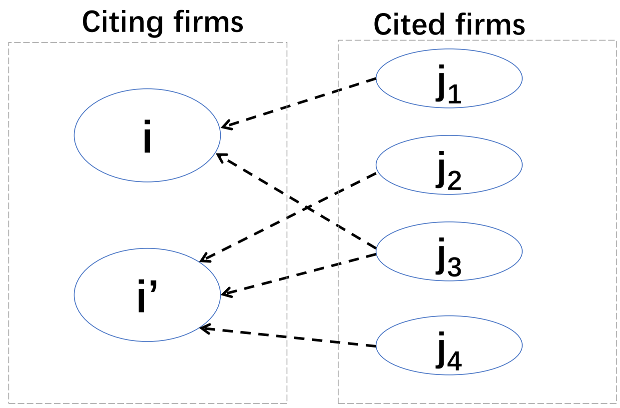


Figure 5: Identifying Knowledge Diffusion via Citation Networks

Figure 5 illustrates our identification of ρ^{NX} in Equation (1). We compare two firms, i and i', that have cited from different sets of firms. If the patent citation is a good indicator for knowledge diffusion, then we expect that firm i tends to experience a faster productivity growth than firm i' if on average firm j_1 and firm j_3 grow faster than firm j_2, j_3, and j_4. Indeed, we are testing whether firm i and i' are affected by idiosyncratic productivity shocks to firms in their citation sets.

One concern for our identification strategy is that the technology progress of cited firms may be correlated with unobserved factors that affect the technology progress of citing firms. To address this issue, we instrument the technology progress of cited firms by export shocks which are constructed following Hummels et al. (2014). Suppose that firm j belongs to the 3-digit CIC industry k. Then the export shocks to firm j at period t is constructed as

$$\mathrm{ES}_{jt} = \sum_{n=1}^{N} \tilde{w}_{kn,2000} \Delta \log \left(\mathrm{GDP}_{nt}\right), \tag{3}$$

where $\tilde{w}_{kn,1998}$ is the share of exports of industry k to country n in the year 2000, and $\Delta \log \left(\mathrm{GDP}_{nt}\right)$ is the GDP growth rate of country n at period t. The exclusive restriction requires that factors that affect fundamentals of industry k in China do not affect the aggregate GDP growth of destination country n. Moreover, to avoid the endogeneity of weight $\tilde{w}_{kn,2000}$, we let $t \geq 2003$.[8]

[7] See the recent work Borusyak et al. (2018) for detailed technical discussions of this method.

[8] The result is robust if we let $t \geq 2004$.

Then we construct the instrument to $\Delta \log \phi_{it}^{\text{NX}}$ as follows:

$$\text{ES}_{it}^{\text{NX}} = \sum_{j \in \mathbf{C}_i} w_{jt} \text{ES}_{jt}. \tag{4}$$

	$\Delta \log \phi_{it}^{\text{NX}}$ (OP)	$\Delta \log \phi_{it}^{\text{NX}}$ (LP)	$\Delta \log \phi_{it}^{\text{NX}}$ (VA)
$\text{ES}_{it}^{\text{NX}}$	.774**	1.046***	1.264***
	(.316)	(.321)	(.268)
Firm age (log)	-.00665	-.00463	.00277
	(.00992)	(.0101)	(.00835)
Asset (log)	.00298	.00557	.00285
	(.00531)	(.00539)	(.00443)
Export dummy	.0129	-.000118	.00667
	(.0171)	(.0174)	(.0144)
SOE dummy	.0690**	.0798***	.0440*
	(.0300)	(.0305)	(.0248)
# Obs.	11,198	11,181	11,552

Table 2: The First-Stage Results of IV Regression

(Note: Standard errors in parentheses.)

Table 2 shows the first-stage results of our IV regression. The "network export shock", $\text{ES}_{it}^{\text{NX}}$, is strongly positively correlated with our key explanatory variable, $\Delta \log \phi_{it}^{\text{NX}}$.

	OLS			IV		
	$\Delta \log \phi_{it}^{\text{OP}}$	$\Delta \log \phi_{it}^{\text{LP}}$	$\Delta \log \phi_{it}^{\text{VA}}$	$\Delta \log \phi_{it}^{\text{OP}}$	$\Delta \log \phi_{it}^{\text{LP}}$	$\Delta \log \phi_{it}^{\text{VA}}$
$\Delta \log \phi_{it}^{\text{NX}}$	.045***	.052***	.0535***	.822	.743*	.621**
	(.011)	(.013)	(.012)	(.545)	(.385)	(.256)
Firm age (log)	-.047***	-.104***	-.06***	-.0502***	-.106***	-.0655***
	(.01)	(.010)	(.007)	(.0138)	(.0128)	(.0101)
Asset (log)	.009	.02***	.002	.00454	.0113	-.0013
	(.006)	(.006)	(.004)	(.0073)	(.0072)	(.0054)
Export dummy	.005	.017	.001	-.0059	.011	-.011
	(.015)	(.016)	(.01)	(.024)	(.022)	(.018)
SOE dummy	-.04	-.043*	.002	-.069	-.081*	-.022
	(.025)	(.026)	(.02)	(.054)	(.049)	(.032)
Prov. f.e.	✓	✓	✓	✓	✓	✓
3-digit industry f.e.	✓	✓	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓	✓	✓
# Obs.	12,783	12,765	13,177	11,198	11,181	11,552
Adjusted R^2	.024	.029	.03	-	-	-
Cragg-Doald Wald F				6	10.63	22.23

Table 3: Technology Progress along Citation Networks: Reduced-form Evidence

(Note: Standard errors in parentheses. Standard errors are clustered at 3-digit industry level.)

Table 3 shows the reduced-form evidence. It suggests that technology improvements do diffuse via patent citation networks: a firm has faster TFP growth if the firms whose patents it cites on average experience faster TFP growth. This does not come from sectoral or local shocks that are common across cited and citing firms since we have controlled for industry and provincial fixed effects. Our IV estimators confirm the findings of OLS estimators.

The reduced-form results are robust with different empirical specifications. The robustness exercises are presented in our online appendix.

Put together, the results in this section suggest that inter-industry knowledge networks are not sufficient to understand knowledge diffusion across firms. The technology improvements of a firm is significantly affected by idiosyncratic shocks to firms whose patents it cites. This finding is consistent with Jaffe et al. (2000), who show that patent citations do provide an indication of inter-firm knowledge spillovers. Motivated by this evidence, we proceed by developing a model of firm innovation and knowledge networks.

3 A Model of Firm Innovation and Knowledge Networks

This section sets up the model and describes firms' innovation decisions and the specification of inter-firm knowledge networks.

Time is discrete and goes to infinity. The economy consists of a measure L of workers who are infinitely lived and a unit mass of firms. Each worker supplies her one unit of labor inelastically in each period and has constant-elasticity-of-substitution (CES) preferences over a continuum of varieties, with elasticity of substitution $\sigma > 1$. Each firm produces a differentiated variety of goods using labor. The market structure for all firm sales is assumed to be monopolistic competition.

3.1 Firms' Production and Innovation

Each firm is owned by a family of entrepreneurs. Each entrepreneur lives for one-period and has the same preference with workers. At the beginning of period t, the entrepreneur of firm ω born at period t inherits labor productivity $\phi_t(\omega)$ from her ancestors and makes the innovation decision $\kappa_t(\omega)$. The effective productivity of firm ω at period t is then $\kappa_t(\omega)\phi_t(\omega)$. The initial productivity distribution $\{\phi_0(\omega)\}$ is exogenous.

Firms are heterogeneous in innovation efficiency. To achieve innovation $\kappa_t(\omega)$, firm ω has to employ $\frac{\kappa_t(\omega)^\alpha}{z(\omega)}$ additional units of labor where $z(\omega) > 0$ denotes the innovation efficiency of firm ω. We regard $z(\omega)$ as the *fundamental characteristics* of firm ω so that $z(\omega)$ is exogenous and time-invariant. Therefore, we index firm ω by its innovation efficiency z and denote the cumulative distribution function of z as $G(z)$. Moreover, we take labor wage as the numeraire.

We assume that the entrepreneur dies at the end of period t and does not internalize the benefits of innovation received by her successors. Therefore, the innovation decision $\kappa_t(z)$ is solved by the following one-period profit-maximization problem:

$$\max_{\kappa_t(z)} \tilde{\sigma} \left[\kappa_t(z)\phi_t(z)\right]^{\sigma-1} D_t - \frac{\kappa_t(z)^\alpha}{z}, \quad D_t = P_t^{\sigma-1} X_t, \quad \alpha > \sigma - 1, \tag{5}$$

where $\tilde{\sigma} = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}$, P_t is the aggregate price index, and X_t is the total expenditure.

Notably, our setting of one-period-lived entrepreneurs is isomorphic to the model of firm innovation developed by Desmet, Nagy, and Rossi-Hansberg (2018). They assume that firms face local competition for land. Firms innovate in order to maximize their bid for land and obtain zero profits after covering their innovation costs. This way they transform a dynamic innovation problem into a sequence of static innovation decisions that maximize one-period profits. Our model abstracts from spatial economy and land competition. To keep innovation decisions tractable, we further simplify the model in Desmet, Nagy, and Rossi-Hansberg (2018) by assuming that entrepreneurs are one-period-lived. Similar techniques has been used in Desmet, and Rossi-Hansberg (2014).

The first-order condition of Problem 5 implies that

$$\kappa_t^*(z) = \tilde{D}_t \left[z\phi_t(z)^{\sigma-1}\right]^{\frac{1}{\alpha-(\sigma-1)}}, \quad \tilde{D}_t := \left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha} D_t\right]^{\frac{1}{\alpha-(\sigma-1)}}. \tag{6}$$

Equation (6) links the firm's innovation investment to its exogenous innovation efficiency z and its productivity $\phi_t(z)$ which is determined by previous innovation decisions. It is consistent with the empirical evidence suggesting that larger and more productive firms do more R&D (See, for example, Cohen, 2010).

3.2 Inter-firm Knowledge Networks

In this subsection, we specify the formation of inter-firm knowledge networks. To achieve this, we assume that the productivity at the beginning of period $t+1$, $\phi_{t+1}(z)$, does not only depend on the firm's own innovation outcomes at period t, but also on other firms' innovation outcomes. More specifically, we assume that productivity $\{\phi_t(z)\}$ is evolved as follows:

$$\phi_{t+1}(z) = \left[\kappa_t^*(z)\phi_t(z) + \delta \int_{S_z} m(z', z)\kappa_t^*(z')\phi_t(z')\, dG(z')\right]^\beta, \quad \beta \leq 1, \quad \delta > 0. \tag{7}$$

Several issues are worth further discussing. First, firm z is only able to receive knowledge from firm z' with probability $m(z', z) \in [0, 1]$. Given that there exists a continuum of firms of every state z, $m(z', z)$ is also equal to the fraction of z'-firms that diffuse their knowledge to a given z-firm, as well as the fraction of z-firms that receive knowledge from a given z'-firm. As a result, the structure of inter-firm knowledge networks can be fully characterized by this matching function $m(z', z)$. Notably, Lim (2018) has utilized a similar matching function to characterize production networks across firms.

Second, in our baseline specification, we assume that the matching function $m(z', z)$ is time-invariant and depends only on firms' *fundamental characteristics* z. In this paper, we focus on the implications of stable and time-invariant characteristics of inter-firm knowledge networks. Moreover, as shown below, this specification leads to simple conditions that ensure the uniqueness of steady-state equilibrium and transparent estimates of the model's key parameters. We extend our model to allow dynamic knowledge network formation and find that most of our quantitative results hold in this extension. The details of this extension are presented in our online appendix.

Third, we do not specify a search and matching process to rationalize our matching function $m(z', z)$. The general form of $m(z', z)$ makes our model sufficiently flexible to capture rich patterns of inter-firm knowledge spillovers observed in the patent citation data. We leave the micro-foundation of matching function to future work.

Finally, it is straightforward to incorporate other firm characteristics than innovation efficiency z into our matching function. In this paper, we concentrate on the single-dimensional innovation efficiency because most of the firm characteristics are strongly correlated with firm size. In some context, firm characteristics other than size may be in special interest. For example, if we are interested in how foreign ownership affects inter-firm knowledge spillovers, we have to incorporate ownership status into our matching function. Moreover, variables that characterize physical and technological distances across firms can also be included. We leave these concerns to future work.

3.3 General Equilibrium

Now we close the model by aggregating the firms' decisions into market clearing conditions. First, the aggregate price index can be expressed as

$$P_t = \frac{\sigma}{\sigma - 1} \left[\int_{S_z} \left[\kappa_t^*(z) \phi_t(z) \right]^{\sigma - 1} dG(z) \right]^{\frac{1}{1-\sigma}} . \tag{8}$$

The sales of firm z can be expressed as

$$\begin{aligned} X_t(z) &= \sigma\tilde{\sigma}\left[\kappa_t(z)\phi_t(z)\right]^{\sigma-1} D_t \\ &= \sigma\tilde{\sigma}\left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\right]^{\frac{\sigma-1}{\alpha-(\sigma-1)}} z^{\frac{\sigma-1}{\alpha-(\sigma-1)}}\phi_t(z)^{\frac{\alpha(\sigma-1)}{\alpha-(\sigma-1)}} D_t^{\frac{\alpha}{\alpha-(\sigma-1)}}. \end{aligned} \tag{9}$$

It is straightforward to show that for each firm z, the innovation expenditure has a fraction $\frac{1}{\sigma}\frac{\sigma-1}{\alpha}$ in $X_t(z)$. Therefore, the net profit as a share of total sales is $\frac{1}{\sigma}\left(1-\frac{\sigma-1}{\alpha}\right)$.

The total expenditure is the sum of wage income and net profits. Since we take wage as numeraire, the total expenditure is

$$X_t = \frac{1}{1-\frac{1}{\sigma}\left(1-\frac{\sigma-1}{\alpha}\right)} L. \tag{10}$$

The instantaneous welfare of the economy can be measured by the real income $W_t = \frac{X_t}{P_t}$.

Definition 1 (Dynamic Equilibrium) *Given L, $G(z)$, $m(z',z)$, and $\{\phi_0(z)\}$, the dynamic equilibrium of our model consists of $\{\kappa_t^*(z), \phi_t(z), P_t, X_t\}$ such that*

1. *Consumers maximize their utility.*
2. *Given $\{\phi_t(z), P_t, X_t\}$, each entrepreneur decides $\kappa_t^*(z)$ as in Equation (6).*
3. *Given $\{\phi_0(z)\}$, productivities evolve as in Equation (7).*
4. *Aggregate price index P_t is given by Equation (8).*
5. *Total expenditure is given by Equation (10).*

3.4 Steady-State

In this subsection, we characterize the steady-state of our dynamic equilibrium. In the steady-state, firms' productivities are time-invariant, i.e. $\phi_t(z) = \phi(z)$ for all t. As a result, X_t, P_t and $\kappa_t^*(z)$ are all time-invariant in the steady-state.

First, it is straightforward to derive the steady-state productivity if there is no inter-firm knowledge spillovers, i.e. $m(z',z) = 0$ for all (z',z).

Lemma 2 (Steady-State without Knowledge Spillover) *Suppose that $m(z',z) = 0$ for all (z',z). Suppose that $1-\frac{\alpha\beta}{\alpha-(\sigma-1)} > 0$. Then the steady-state productivity can be given by*

$$\phi(z) = \tilde{D}^{\frac{\beta}{1-\frac{\alpha\beta}{\alpha-(\sigma-1)}}} z^{\frac{\beta}{\alpha-(\sigma-1)}\frac{1}{1-\frac{\alpha\beta}{\alpha-(\sigma-1)}}}. \tag{11}$$

In this case, the distribution of steady-state productivity can be derived directly from the distribution of innovation efficiency. It is trivial to show the existence and uniqueness of the steady-state.

Does the steady-state exist under general $m(z', z)$? The following results establish the existence and uniqueness of the steady-state productivity under general $m(z', z)$.

Proposition 3 (Steady-State with Arbitrary Matching Function) *Suppose that* $1 - \frac{\alpha\beta}{\alpha-(\sigma-1)} > 0$. *Then there exists a unique distribution of steady-state productivity for any matching function* $m(z', z)$. *Moreover, given the demand shifter* $\tilde{D}$, *the steady-state* $\{\phi(z)\}$ *can be computed by iterating the following system of equations:*

$$\phi(z) = \tilde{D}^{\beta} \left[z^{\frac{1}{\alpha-(\sigma-1)}} \phi(z)^{\frac{\alpha}{\alpha-(\sigma-1)}} + \delta \int_{S_z} m(z', z) \, (z')^{\frac{1}{\alpha-(\sigma-1)}} \, \phi(z')^{\frac{\alpha}{\alpha-(\sigma-1)}} dG(z') \right]^{\beta}, \quad \forall z \in S_z. \tag{12}$$

Proposition 3 suggests that as long as $m(z', z)$ is exogenous, the uniqueness of our steady-state equilibrium does not rely on the structure of inter-firm knowledge networks. Instead, the sufficient conditions for uniqueness include only three parameters, (α, β, σ). This result greatly simplifies our equilibrium characterization as well as structural estimation.

4 Structural Estimation

In this section, we recover the structure of inter-firm knowledge networks using data on patent citations across Chinese manufacturing firms. We first specify our parametric assumptions on the matching function $m(z', z)$. Then we discuss the estimation procedures. We then conduct our estimation in two steps and finally elaborate the model fit to targeted and untargeted moments.

4.1 Parametric Assumptions

To proceed with the structural estimation, we propose parametric assumptions for the distribution of innovation efficiency $G(z)$ and the matching function $m(z', z)$. First, we assume that the innovation efficiency z are log-normally distributed, with the mean μ_z and variance σ_z^2. Since the model is scale invariant, we normalize $\mu_z = 0$. This log-normality assumption is consistent with the literature suggesting that the major part of firm size distribution, except for the tail, can be well-approximated by the log-normal distribution.

Since our empirical regularities have shown that most of the firms have very few knowledge connections, our model will yield a log-normal firm size distribution, except for the tail.

Second, we parameterize $m(z', z)$ as

$$m(z', z) = \frac{\gamma}{1 + \exp\{-[\xi_1 \log z' + \xi_2 \log z + \rho \log z' \log z]\}}, \quad \gamma \in [0, 1]. \tag{13}$$

There are four parameters in this matching function. First, $\gamma \in [0, 1]$ characterizes the average matching rate across firms. Second, ξ_1 and ξ_2 characterize how matching rates vary, respectively, with respect to innovation efficiency of the cited firm z' and the citing firm z. Finally, ρ characterizes the assortativity of inter-firm knowledge networks, i.e. whether larger and more connected firms are connected with firms that are also larger and more connected. Notably, we have

$$\frac{\partial^2 \log m(z', z)}{\partial \log z \partial \log z'} = \left[\rho + \frac{(\xi_1 + \rho \log z)(\xi_2 + \rho \log z')}{m(z', z)/\gamma}\right] \frac{\exp\{-[\xi_1 \log z' + \xi_2 \log z + \rho \log z' \log z]\}}{m(z', z)/\gamma}. \tag{14}$$

Therefore, $m(z', z)$ is log-supermodular if $\rho > 0$ and $\xi_1, \xi_2 > 0$.

4.2 Estimation Procedure

The model has following parameters: (1) the elasticity of substitution, σ; (2) the curvature of innovation cost, α; (3) the magnitude of knowledge transfer, δ; (4) the curvature of productivity evolution, β; (5) the total labor, L; (6) the variance of innovation efficiency, σ_z^2; and (7) the parameters of matching function, $(\gamma, \xi_1, \xi_2, \rho)$.

We first discuss the parameters that are not estimated from data. First, we set the value of the elasticity of substitution σ to 4, which is close to the estimates in the literature. Second, since the model is scale invariant, we normalize L so that $\tilde{D} = 1$. Third, we set $\alpha = 15$ so that the net profit share is equal to $\frac{1}{\sigma}\left(1 - \frac{\sigma-1}{\alpha}\right) = 0.2$. Finally, we set the curvature of productivity evolution $\beta = 0.72$. Without knowledge diffusion, this leads to $\frac{\partial \log(\phi_{t+1}(z))}{\partial \log(\phi_t(z))} = \frac{\alpha\beta}{\alpha-(\sigma-1)} = 0.9$, which is close to the estimates in the literature.[9]

The remaining parameters of the model, $(\gamma, \xi_1, \xi_2, \rho)$, and (δ, σ_z^2), are then estimated using data on inter-firm citation linkages and firm size. In the following subsections, we first estimate (δ, σ_z^2) from the observed firm size distribution and citation linkages by a maximum likelihood estimator. Then given the estimates on (δ, σ_z^2), we estimate parameters of matching function, $(\gamma, \xi_1, \xi_2, \rho)$, from citation linkages by constrained simulated method

[9] See, for example, Roberts, et al. (2011).

of moments.

4.3 Maximum Likelihood Estimator on (δ, σ_z^2)

In this subsection, we estimate (δ, σ_z^2) using data on firms' long-term average sales and citation linkages $\{\mathbf{1}\,[\text{firm j cites from firm i}]\}$. Through the lens of our model, the firm's long-term average sales can be expressed as $x_i = \left(z_i^{\frac{1}{\alpha-(\sigma-1)}} \phi_i^{\frac{\alpha}{\alpha-(\sigma-1)}}\right)^{\sigma-1}$. Then we can recover the innovation efficiency of firm j directly by combining the sales equation with Equation (12):

$$z_j(\delta) = \frac{\left(x_j^{\frac{1}{\sigma-1}}\right)^{\alpha-(\sigma-1)}}{\left[x_j^{\frac{1}{\sigma-1}} + \delta \sum_{i=1}^{R} m_{ij} x_i^{\frac{1}{\sigma-1}}\right]^{\alpha\beta}}, \tag{15}$$

where the empirical matching rate is constructed by $m_{ij} := \frac{\mathbf{1}[\text{firm j cites from firm i}]}{R}$.

Equation (15) provides identification for δ. If $\delta = 0$, then $\{x_j\}$ should be log-normally distributed, as we have assumed for $\{z_j\}$. The extent to which the observed $\{x_j\}$ deviate from log-normal distribution identifies the magnitude of inter-firm knowledge spillover, δ. More specifically, let $K_j(\delta) = \log z_j(\delta) - \frac{1}{S}\sum_{i=1}^{S} \log z_i(\delta)$. Under the assumption that z is log-normally distributed, $K_j(\delta) \sim N(0, \sigma_z^2)$. Therefore, (δ, σ_z^2) can be estimated by the maximum likelihood estimator (MLE):

$$\max_{(\delta,\sigma_z^2)} \ell(\delta, \sigma_z^2; \{x_i, m_{ij}\}) = -\frac{S}{2}\log\left(\sigma_z^2\right) - \frac{1}{2\sigma_z^2}\sum_{j=1}^{S} K_j(\delta)^2. \tag{16}$$

Parameter		Value	Standard Error
Magnitude of spillover	δ	3.118	.204
Variance of z	σ_z^2	.130	.001

Table 4: Estimates on (δ, σ_z^2)

(Notes: the standard errors are estimated based on the asymptotics of extreme estimator.)

The MLE estimates on (δ, σ_z^2) are shown in Table 4. The estimate on δ is sizable and significantly positive. This result indicates that the observed distribution of long-term sales $\{x_j\}$ substantially deviate from log-normal distribution. In other words, although the innovation efficiency z_j is assumed to be log-normal, the resulting productivity and sales deviate from log-normality because of the heterogeneous inter-firm knowledge linkages.

4.4 Constrained Simulated Method of Moments on $(\gamma, \xi_1, \xi_2, \rho)$

In this subsection, we estimate parameters of matching function, $(\gamma, \xi_1, \xi_2, \rho)$, by the following constrained simulated method of moments. We draw N observations $\{U_i\}_{i=1}^N$ independently from $N(0, \sigma_z^2)$ and compute $z_i = \exp\{U_i\}$. Then the simulated sales $\{x_i\}_{i=1}^N$ can be computed by

$$x_j^{\frac{1}{\sigma-1}} = z_j^{\frac{1}{\alpha-(\sigma-1)}} \left[x_j^{\frac{1}{\sigma-1}} + \delta \sum_{i=1}^{N} m(z_i, z_j; \gamma, \xi_1, \xi_2, \rho) x_i^{\frac{1}{\sigma-1}} \right]^{\frac{\alpha\beta}{\alpha-(\sigma-1)}}. \tag{17}$$

As described above, we observe firm j citing patents from firm i. We thereby can compute firms' in-degree and out-degree. Our simulation can generate the corresponding statistics as:

$$tm_j = \sum_{i=1}^{N} m(z_i, z_j), \quad tm_i = \sum_{j=1}^{N} m(z_i, z_j). \tag{18}$$

Moreover, we can compute the average sales of firms from which firm j cites and the average sales of firms citing firm i. The corresponding simulated statistics can be computed by

$$am_j = \frac{\sum_{i=1}^{N} m(z_i, z_j) x_i}{\sum_{i=1}^{N} m(z_i, z_j)}, \quad am_i = \frac{\sum_{j=1}^{N} m(z_i, z_j) x_j}{\sum_{j=1}^{N} m(z_i, z_j)}. \tag{19}$$

	Simulated Moment	Data Moments	Simulation Result
	(1)	(2)	(3)
$sm(.)$	(i) Slope of regressing $\log(tm_j)$ on $\log(x_j)$	.223	.2234
	(ii) Slope of regressing $\log(tm_i)$ on $\log(x_i)$	.311	.3075
	(iii) Slope of regressing $\log(am_j)$ on $\log(x_j)$	.180	.1542
	(iv) Slope of regressing $\log(am_i)$ on $\log(x_i)$	.140	.1645
	$p75(\log x_i)/p50(\log x_i)$	1.0472	1.0472

Table 5: Simulated Moments

Our targeted moments are summarized in Column (1) of Table 5. Column (2) of Table 5 presents the values of our targeted moments in the data. The first four moments in $sm(.)$ are used to identify (ξ_1, ξ_2, ρ). In particular, the first and second moments characterize, respectively, how in- and out-degree vary with respect to firm size, which aim at identifying

(ξ_1, ξ_2). The third and fourth moments, instead, characterize to what extent large firms cite and are cited by large firms. These two moments are set to identify ρ.

The fifth moment in Table 5 is used to identify γ, since if $\gamma = 0$ then $\log x_i$ is normally distributed and thereby $p75(\log x_i)/p50(\log x_i) = \infty$.

Given the simulated $\{z_i\}_{i=1}^N$, the constrained simulated method of moments can be expressed as

$$\min_{\gamma,\xi_1,\xi_2,\rho,\{x_i\}_{i=1}^N} sm(\gamma, \xi_1, \xi_2, \rho, \{x_i\})'\Omega sm(\gamma, \xi_1, \xi_2, \rho, \{x_i\}),$$

s.t.

$$x_j^{\frac{1}{\sigma-1}} = z_j^{\frac{1}{\alpha-(\sigma-1)}} \left[x_j^{\frac{1}{\sigma-1}} + \delta \sum_{i=1}^{N} m(z_i, z_j; \gamma, \xi_1, \xi_2, \rho) x_i^{\frac{1}{\sigma-1}} \right]^{\frac{\alpha\beta}{\alpha-(\sigma-1)}}, \quad \forall j = 1, \ldots, N, \tag{20}$$

$$\frac{p75(\log x_i)}{p50(\log x_i)} = 1.0472,$$

where Ω is a positive definite weighting matrix.

We set $N = 100$ and use identity matrix as the weighting matrix. Let G be the Jacobian matrix of $sm(.)$ with respect to parameters and V_m be the variance-covariance matrix of the moments. Then by the property of extreme estimator, the variance-covariance matrix of the estimated parameters can be given by $\frac{1}{B}(G'G)^{-1}G'V_mG(G'G)^{-1}$ where V_m is computed by bootstrapping and B is the number of repetitions for bootstrapping.

Parameter		Value	Standard Error
Level of Matching Rate	γ	.2204	.103
Marginal effect of z'	ξ_1	1.2965	.1188
Marginal effect of z	ξ_2	.9711	.0879
Cross effect	ρ	4.2067	.1441

Table 6: Estimates on the Matching Function $m(z', z)$

Table 6 shows the estimation results for $(\gamma, \xi_1, \xi_2, \rho)$. Both ξ_1 and ξ_2 are significantly positive, confirming the empirical regularity that larger firms cite more and are cited by more firms. Moreover, ρ is positive and sizable, suggesting strong positive matching across firms. Based on Equation (14), our estimates on (ξ_1, ξ_2, ρ) suggest that large firms are well-connected with each other and lie at the center of inter-firm knowledge networks, whereas small firms can hardly get connected.

How do our estimates depend on our estimation moments? Andrews et al. (2017) develop a sensitivity matrix that measures the dependence of estimates on moments. Following their methodology we first compute the Jacobian matrix of our simulated moments $sm(.)$ in Equation (20) with parameters (ξ_1, ξ_2, ρ), denoted as $\tilde{G}$. Then our estimates $\hat{\theta} := (\hat{\xi}_1, \hat{\xi}_2, \hat{\rho})$ has first-order asymptotic bias:

$$E(\tilde{\theta}) = \tilde{\Lambda} E(sm), \quad \tilde{\Lambda} = -\left(\tilde{G}'\Omega\tilde{G}\right)^{-1} \tilde{G}'\Omega. \tag{21}$$

	Moments			
	(i)	(ii)	(iii)	(iv)
$\hat{\xi}_1$	6.68	3.78	1.83	2.37
$\hat{\xi}_2$	8.88	-0.007	1.63	1.61
$\hat{\rho}$	12.56	1.99	14.96	16.23

Table 7: Sensitivity Matrix of Estimates with Simulated Moments

The results are presented in Table 7. It suggests that our estimates of $\hat{\xi}_1$ and $\hat{\xi}_2$ strongly positively relate to data moments (i) and (ii) in Table 5, whereas $\hat{\rho}$ is sensitive to data moments (iii) and (iv) in Table 5. These results confirm that the positive matching assortativity in the patent citation data is crucial for identifying parameters of inter-firm knowledge networks.

4.5 Model Fit

We have estimated four of our model's key parameters targeting on five moments listed in Table 5. As shown in Table 5, our model matches the targeted moments quite well. Our simulation generates strong positive correlation between firms' sales and their in-/out-degrees, which approximates the data tightly. Our model also generates strong positive correlation between firms' sales and the average sales of firms they cite from/are cited.

Moreover, our simulation exactly replicates the ratio of the 75th percentile of log sales over the median of log sales in the data. This result provides further evidence suggesting that although the exogenous innovation efficiency is assumed to be log-normal, the resulting firm sales distribution substantially deviates from log-normality because firms occupy heterogeneous positions in knowledge networks. Figure 6 shows the model's fit of the firm sales distribution. The model generates reasonable good approximation to the empirical firm size distribution.

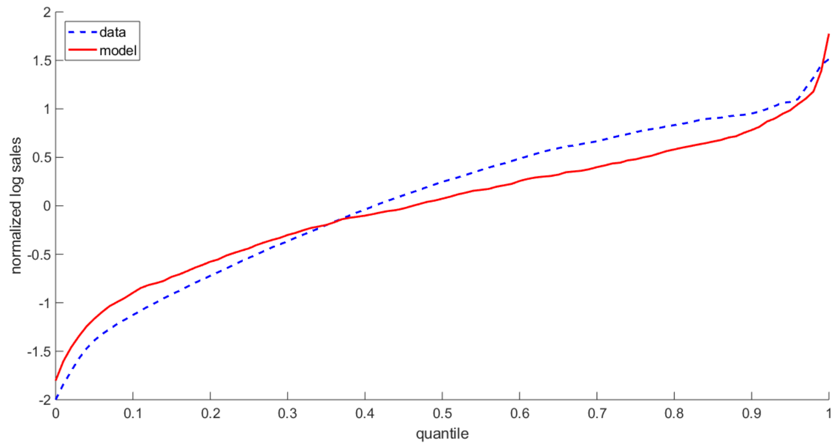


Figure 6: Firm Sales Distribution

(Notes: log sales are normalized by subtracting their mean and then dividing their standard deviation.)

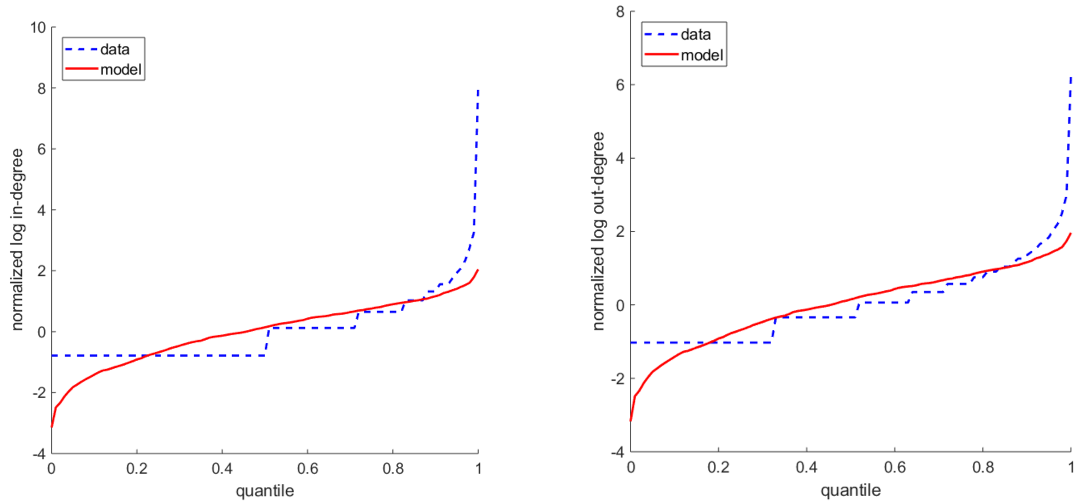
Figure 7 shows that our model matches well the shape of normalized log degree distributions, although it underpredicts the connectivity of the firms that are most connected in the knowledge networks.

Figure 8 shows the model's fit of the correlation between firm sales and degree. While the model over-predicts the connectivity of the largest firms, it is nonetheless consistent with the empirical pattern that larger firms tend to cite from and be cited by more firms.

Finally, Figure 9 and 10 illustrate the model's fit of the matching assortativity, which characterizes whether larger and more connected firms are connected to firms that are also larger and more connected (positive matching), or to firms that are smaller and less connected (negative matching). Figure 9 shows that larger firms indeed cite from larger firms and are cited by larger firms. Figure 10 shows that firms that cite from more firms tend to cite from firms that cite from more firms themselves. Similarly, firms that are cited by more firms tend to be cited by firms that are cited by more firms themselves. Therefore, the model replicates positive matching, both in terms of sales and degree, in the data.

5 Counterfactual Exercises

Armed by the estimates on our model's key parameters, we conduct counterfactual exercises to study how firm-level instantaneous shocks propagate via inter-firm knowledge networks and translate into aggregate dynamic effects. We first explore the welfare implications of inter-firm knowledge networks by eliminating inter-firm knowledge spillovers in

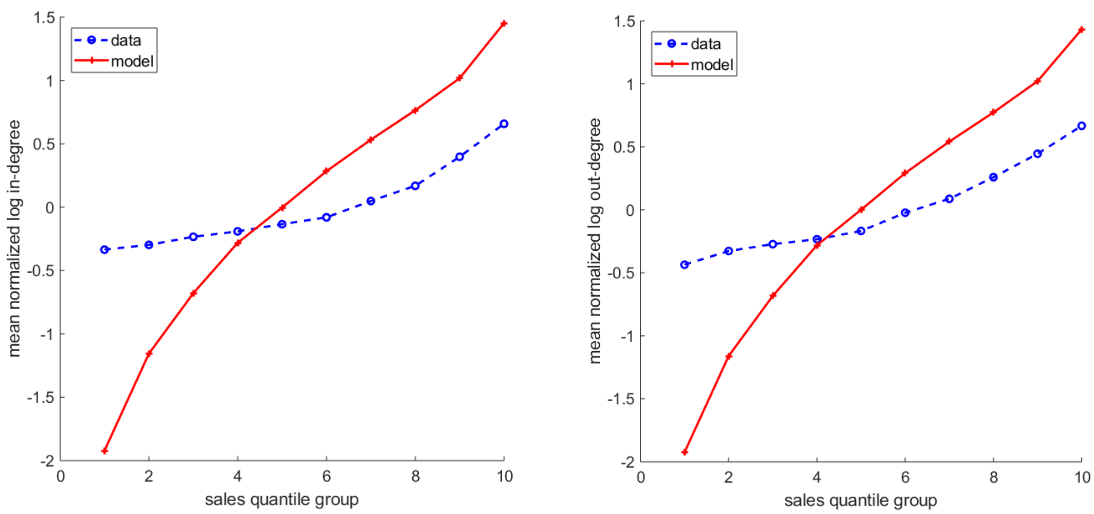


(a) In-degree

(b) Out-degree

Figure 7: Firm Degree Distribution

(Notes: log in-degrees and out-degrees are normalized by subtracting their mean and then dividing their standard deviation.)



(a) In-degree

(b) Out-degree

Figure 8: Firm Sales and Degree

(Notes: log in-degrees and out-degrees are normalized by subtracting their mean and then dividing their standard deviation. Sales quantile group 1 refers to sales between the 0th and 10th percentile. Similarly, Sales quantile group 10 refers to sales between the 90th and 100th percentile.)

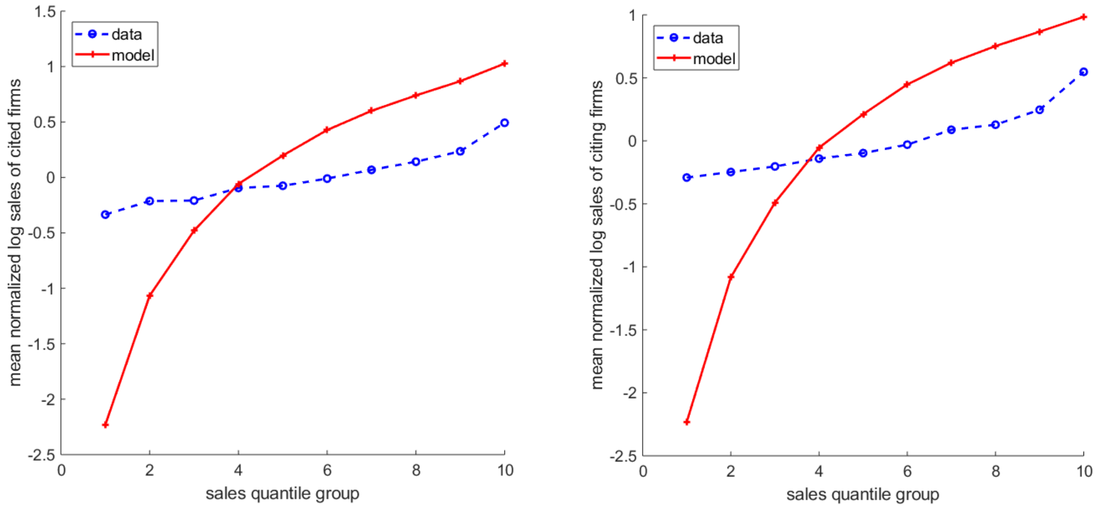


Figure 9: Firm Matching Assortativity (Sales)

(Notes: log sales of cited and citing firms are normalized by subtracting their mean and then dividing their standard deviation. Sales quantile group 1 refers to sales between the 0th and 10th percentile. Similarly, Sales quantile group 10 refers to sales between the 90th and 100th percentile.)

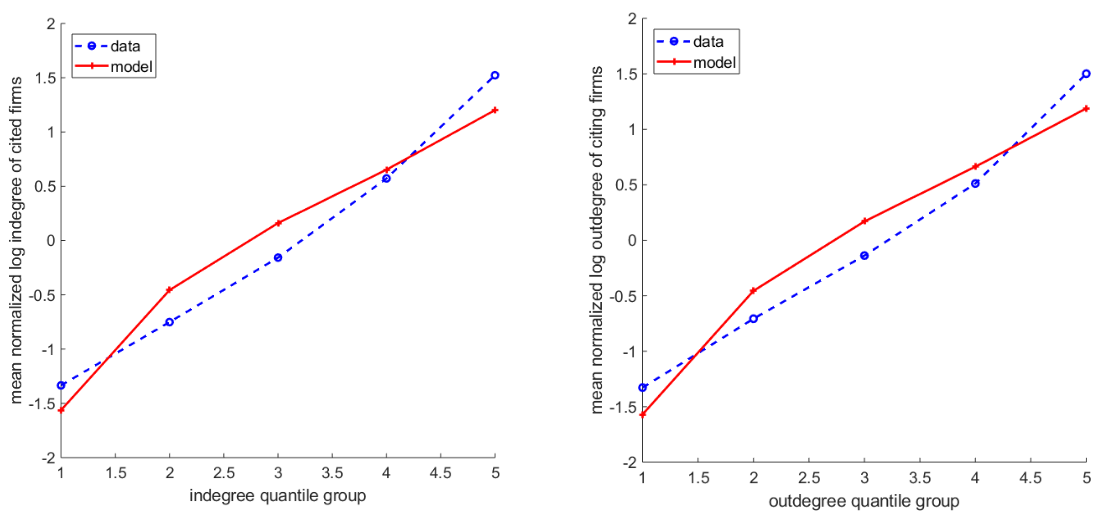


Figure 10: Firm Matching Assortativity (Degree)

(Notes: log in-degrees and out-degrees are normalized by subtracting their mean and then dividing their standard deviation. Sales quantile group 1 refers to sales between the 0th and 20th percentile. Similarly, Sales quantile group 5 refers to sales between the 80th and 100th percentile.)

the model and computing the counterfactual steady-state welfare. Second, we study the aggregate effects of negative demand shocks incurred by Chinese manufacturing firms in the 2008 financial crisis. Third, we characterize the optimal innovation policy in the presence of inter-firm knowledge networks. Finally, we study the implications of inter-firm knowledge spillovers for productivity and welfare effects of trade liberalization.

Given parameters $\{\alpha, \beta, \sigma, \sigma_z^2, \delta, \gamma, \xi_1, \xi_2, \rho\}$, we simulate the model and compute the steady-state via the following algorithm:

Algorithm 4 (Computing the Steady-State) *We compute the steady-state equilibrium as follows:*

1. *Generate a random vector $\{z_i\}_{i=1}^N$ where $\log(z_i)$ is i.i.d. drawn from $N(0, \sigma_z^2)$.*

2. *Initial guess for the aggregate demand shifter $\tilde{D} > 0$.*

3. *Given the guess for $\tilde{D}$, compute $\{\phi(z)\}_{i=1}^N$ by simple iteration using Equation (12).*

4. *Compute the steady-state price index by Equation (8) and update $\tilde{D}$ by Equation (6) and (10).*

5. *Repeat until $\tilde{D}$ converges.*

Starting from an initial state $\{\phi_0(z)\}$, we can also compute the transitional dynamics of our equilibrium under some exogenous shocks. The algorithm is similar to Algorithm 4 and described in detail in the appendix.

5.1 The Value of Inter-Firm Knowledge Networks

In this subsection, we employ the model to study the value of inter-firm knowledge spillovers. Starting from the steady-state of the model corresponding to the parameter values estimated above, we eliminate the inter-firm knowledge networks by setting $\gamma = 0$ and compute the counterfactual equilibrium outcomes using Algorithm 4.

Let W^{baseline} be the steady-state welfare under our baseline estimation. Let $W^{\text{no spillover}}$ be the steady-state welfare under $\gamma = 0$. Our simulation shows that $\frac{W^{\text{baseline}}}{W^{\text{no spillover}}} = 2.13$, which suggests that inter-firm knowledge networks have substantial impacts on the aggregate productivities and welfare in the long run.

Moreover, we start from the baseline steady-state at period 0 and permanently eliminate inter-firm knowledge networks by setting $\gamma = 0$ at period 1. Figure 11 shows that the

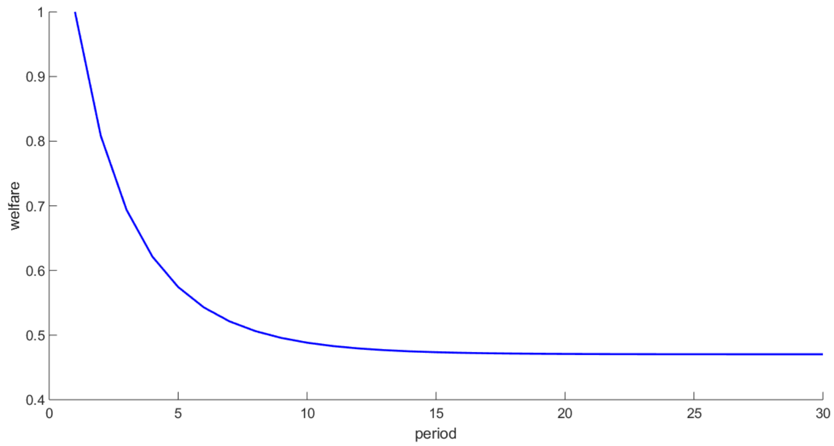


Figure 11: Permanent Elimination of Inter-Firm Knowledge Networks

(Notes: Starting from the observed steady-state at period 0, we set $\gamma = 0$ at period 1 and then on. The welfare at period 0 is normalized into 1.)

aggregate welfare declines gradually after the elimination of inter-firm knowledge networks and converges to the new steady-state after about 20 periods.

Inter-firm knowledge spillovers have heterogeneous effects across firms since firms vary in the size and occupy heterogeneous positions in knowledge networks. We compute the steady-state productivity of firm z under our baseline estimation, $\phi(z)^{\text{baseline}}$, as well as under $\gamma = 0$, $\phi(z)^{\text{no spillover}}$. Figure 12 illustrates $\frac{\phi(z)^{\text{baseline}}}{\phi(z)^{\text{no spillover}}}$ for each percentile of z. There are two forces shaping the distribution of productivity effects of inter-firm knowledge networks: (1) larger firms tend to learn from more firms so that they gain more from inter-firm knowledge spillovers; and (2) smaller firms can learn from firms with much more advanced technologies. Figure 12 suggests that the latter force dominants the former one: the inter-firm knowledge networks are more important to smaller firms since they will lose their access to advanced technologies once the inter-firm knowledge spillovers are absent.

5.2 Welfare Implications of Idiosyncratic Shocks

In this subsection, we utilize our model to study how firm-level technology shocks affect aggregate welfare in the presence of inter-firm knowledge networks. To this end, we conduct two exercises.

First, following the spirit of Hulten (1978), we compute the elasticity of steady-state welfare, W, with respect to firm-level innovation efficiency, z. Hulten (1978) suggests that in an efficient economy, the elasticity of aggregate output with respect to an individual

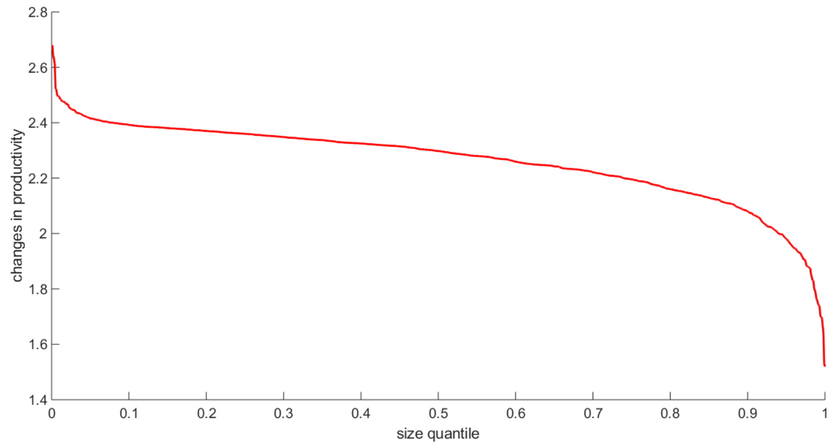


Figure 12: Inter-Firm Knowledge Networks on Productivity Distribution

(Notes: changes in productivity refers to $\frac{\phi(z)^{\text{baseline}}}{\phi(z)^{\text{no spillover}}}$.)

firm's productivity is proportional to the firm's sales share. In other words, in an efficient economy, the firm's sales share is a sufficient statistic for its first-order impact on aggregate output, regardless of how firms are connected with each other. However, our model departs from Hulten's theorem by considering an economy with positive externalities. Importantly, the degree and structure of externalities depend on the structural of inter-firm knowledge networks. Therefore, inter-firm knowledge networks could be relevant in understanding the impacts of individual productivity shocks on aggregate welfare.

Based on our simulated $\{z_i\}_{i=1}^N$, we compute the steady-state elasticity $\frac{\partial \log(W)}{\partial \log(z_i)}$ and compare $\frac{\partial \log(W)}{\partial \log(z_i)}$ with its steady-state sales share $\frac{X_i}{\sum_{k=1}^N X_k}$. To highlight the role of inter-firm knowledge networks, we compare the result in our baseline model with the one in the model without knowledge spillovers, i.e. $\delta = 0$.

The results are illustrated by Figure 13. In the model without knowledge spillovers, although the economy is not of the first best, the impact of individual innovation efficiency on aggregate welfare is still proportional to the firm's sales share. However, in the presence of inter-firm knowledge networks, the firm's sales share is no longer a sufficient statistic for its first-order impact on aggregate welfare. The relationship between welfare elasticity of innovation efficiency and sales share becomes highly non-linear. This result highlights the importance of inter-firm knowledge networks in understanding the first-order impacts of individual firms on aggregate outputs.

Second, we examine the propagation of firm-level negative shocks via inter-firm knowledge networks during the Great Recession. This exercise is aimed to understand to what

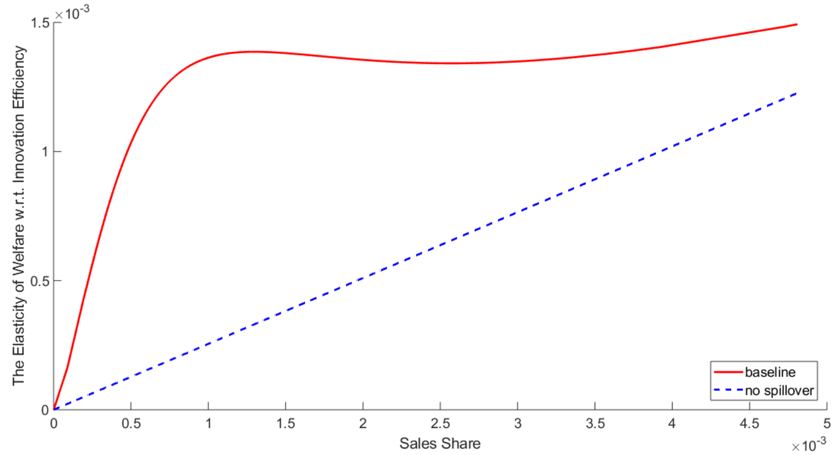


Figure 13: Welfare Elasticity of Firm-Level Innovation Efficiency

extent firm-level instantaneous shocks can translate into aggregate dynamic effects in a real-world background. We start from the steady-state under our baseline estimation at period 0. Following the real business cycle literature, we assume that firm z incurs a one-period productivity shock $\eta(z)$ at period 1. So the sales of firm z at period 1 can be given by

$$X_1(z) = \sigma\tilde{\sigma}\left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\right]^{\frac{\sigma-1}{\alpha-(\sigma-1)}} z^{\frac{\sigma-1}{\alpha-(\sigma-1)}}\left[\eta(z)\phi_1(z)\right]^{\frac{\alpha(\sigma-1)}{\alpha-(\sigma-1)}} D_1^{\frac{\alpha}{\alpha-(\sigma-1)}}. \tag{22}$$

We calibrate $\eta(z)$ to replicate the sales changes of Chinese manufacturing firms over 2007-2008. Then we employ our model to study the contagion and propagation of firm-level negative demand shocks via inter-firm knowledge networks.

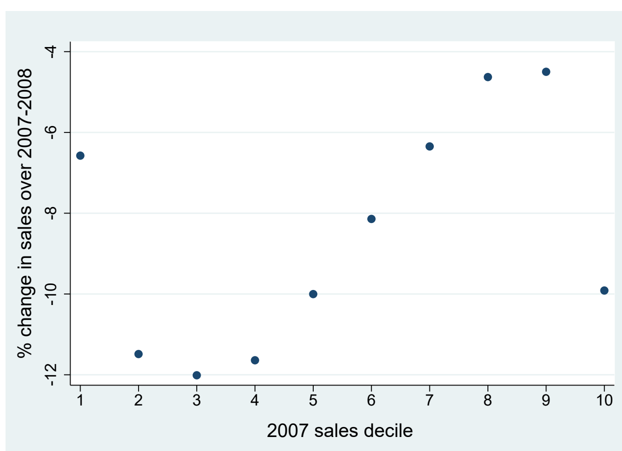


Figure 14: Sales Decline over 2007-2008

(Notes: we collect the median of sales in each decile group and compute their percentage changes over 2007-2008. Sales are deflated by Producer Price Index (PPI) in China.)

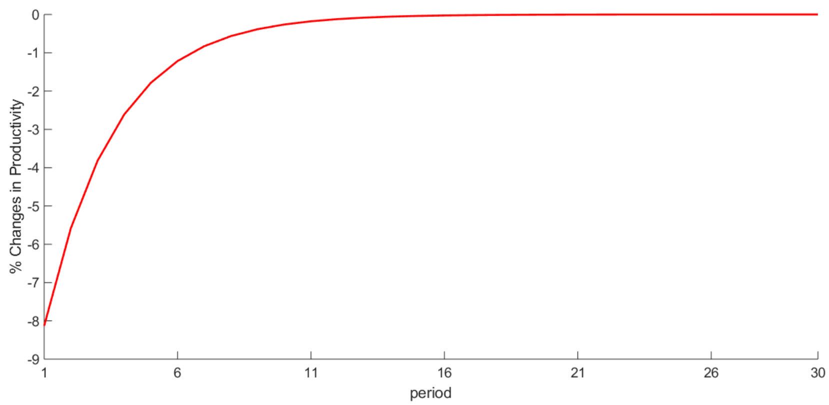
Figure 14 summarizes the sales decline of Chinese manufacturing firms over 2007-2008. For most of the decile groups, the decline is substantial: the sales of 8 out of 10 groups have declined by more than 5%. The decline varies across decile groups, with the largest about 12% and the smallest just above 4%. Based on Equation (22), we translate sales decline into negative productivity shocks.

Decile	1	2	3	4	5	6	7	8	9	10
$\eta(z)$: baseline	.927	.915	.914	.915	.919	.923	.928	.932	.932	.919
$\eta(z)$: no spillover	.927	.914	.913	.914	.918	.922	.927	.931	.931	.918

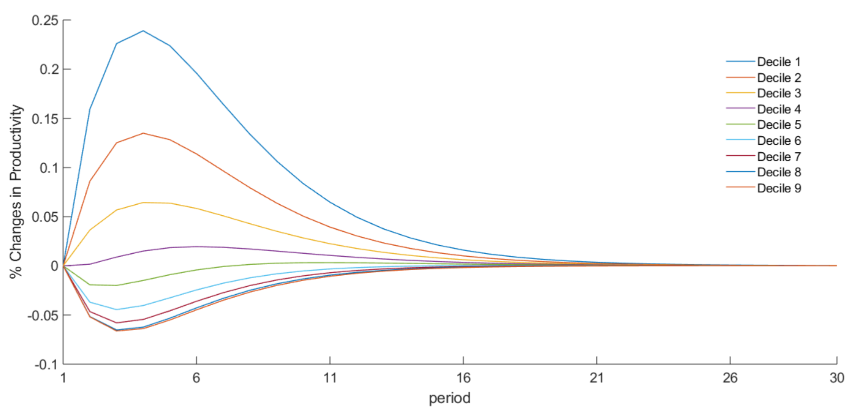
Table 8: Calibrating the One-Period Productivity Shocks $\eta(z)$

Now we focus on the productivity effects of negative productivity shocks of the top 10% firms. We start from the steady-state under our baseline estimation. The one-period productivity shocks of the top 10% firms occur at $t = 1$. The productivity effects of this one-period negative shock are illustrated by Figure 15. The results suggest that the productivities of top 10% firms decline substantially initially but recover rapidly afterwards. Moreover, the productivity effects of other firms vary dramatically across their size groups, with the smallest firms benefiting and the medium and large firms losing from this negative shock. This differentiated effect highlights the importance of inter-firm knowledge networks to the propagation of firm-level shocks: the smallest firms have very limited knowledge connections with the largest firms, so that they gain from the negative shocks on the top 10% firms due to the ease of competition. In contrast, the medium and large firms rely heavily on the knowledge spillovers from the top 10% firms and thereby suffer from the shrinkage of the largest firms.

To further clarify the role of inter-firm knowledge networks in propagating firm-level shocks, we start from the steady-state under $\gamma = 0$. The one-period productivity shocks on the top 10% firms again occur at $t = 1$. Figure 16 differs from Figure 15 in two aspects. First, without inter-firm knowledge networks, all firms other than the top 10% firms gain from the negative productivity shocks on the top 10% firms. This is simply due to the ease of competition, similar to the productivity effects for the smallest firms in Figure 15. Second, without inter-firm knowledge networks, the recovery of the top 10% firms is much slower than the one in the baseline case. This is because without inter-firm knowledge networks, the top 10% firms cannot gain from the productivity improvements of other firms. In sum, inter-firm knowledge networks matter for the distribution and dynamics of productivity effects of firm-level shocks.

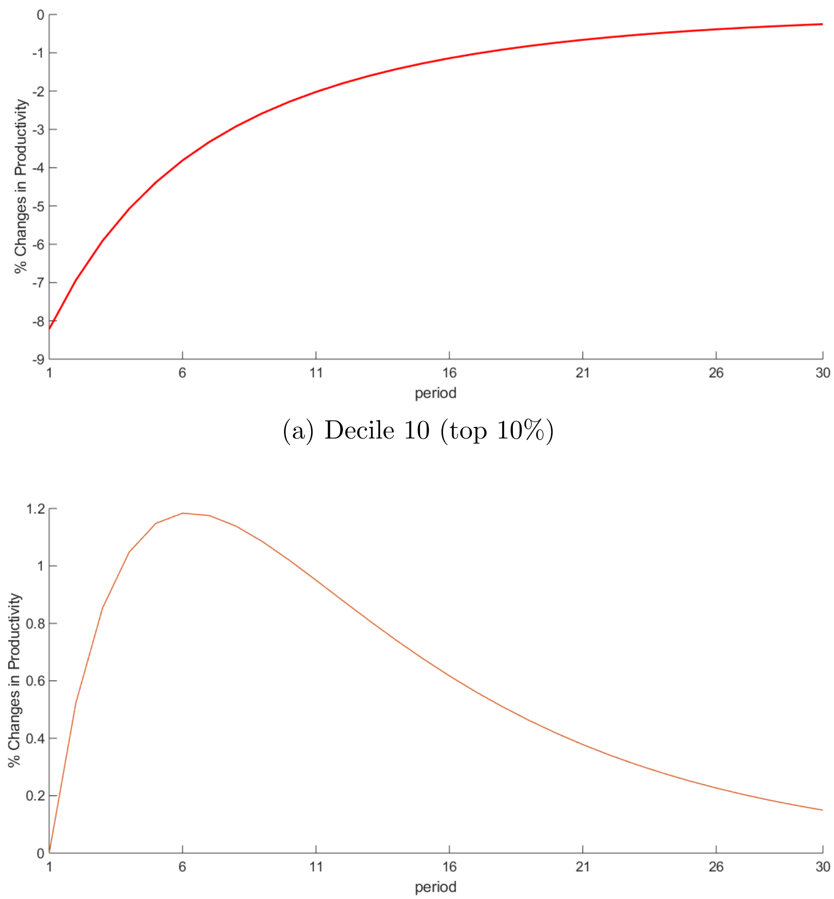


(a) Decile 10 (top 10%)



(b) Decile 1 to 9

Figure 15: Productivity Effects of the Productivity Shocks of the Top 10% Firms: Baseline



(a) Decile 10 (top 10%)

(b) Decile 1 to 9

Figure 16: Productivity Effects of the Productivity Shocks of the Top 10% Firms: No Spillover

5.3 Optimal Innovation Policy

The literature has long argued that the social benefits of innovation exceed the private benefits. This rationalizes the governments' subsidies on firm innovation. Our model deviates from the first-best world due to two externalities: First, entrepreneurs do not internalize the benefits of innovation for their successors. We call this intergeneration externality of innovation. Second, firms do not internalize the benefits of innovation for other firms via inter-firm knowledge networks. We call this network externality of innovation. Innovation subsidies can potentially correct the inefficiency of these two externalities and push our equilibrium towards efficiency.

In particular, we consider that the government subsidizes firm z's profit gross of innovation costs: firm z receives profits $\tilde{\sigma}s(z)\left[\kappa(z)\phi(z)\right]^{\sigma-1}D$ gross of innovation costs where $s(z)>1$ implies positive innovation subsidies and $s(z)\in(0,1)$ implies innovation taxes. Then given the simulated $\{z_i\}_{i=1}^N$, the government solves

$$\max_{\{s_i,\phi_i\},X,P}\frac{X}{P}$$

s.t.

$$\begin{aligned}
&\phi_i=\tilde{D}^{\beta}\left[(s_iz_i)^{\frac{1}{\alpha-(\sigma-1)}}\,\phi_i^{\frac{\alpha}{\alpha-(\sigma-1)}}+\delta\sum_{j=1}^{N}m(z_j,z_i)\,(s_jz_j)^{\frac{1}{\alpha-(\sigma-1)}}\,\phi_j^{\frac{\alpha}{\alpha-(\sigma-1)}}\right]^{\beta},\\
&\tilde{D}:=\left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}P^{\sigma-1}X\right]^{\frac{1}{\alpha-(\sigma-1)}},\\
&P=\frac{\sigma}{\sigma-1}\left\{\sum_{i=1}^{N}\left[\tilde{D}(s_iz_i)^{\frac{1}{\alpha-(\sigma-1)}}\phi_i^{\frac{\alpha}{\alpha-(\sigma-1)}}\right]^{\sigma-1}\right\}^{\frac{1}{1-\sigma}},\\
&X_i=\sigma\tilde{\sigma}\left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\right]^{\frac{\sigma-1}{\alpha-(\sigma-1)}}(s_iz_i)^{\frac{\sigma-1}{\alpha-(\sigma-1)}}\,\phi_i^{\frac{\alpha(\sigma-1)}{\alpha-(\sigma-1)}}\left(P^{\sigma-1}X\right)^{\frac{\alpha}{\alpha-(\sigma-1)}},\\
&X=L+\frac{1}{\sigma}\left(1-\frac{\sigma-1}{\alpha}\right)\sum_{i=1}^{N}s_iX_i-\frac{1}{\sigma}\sum_{i=1}^{N}(s_i-1)X_i.
\end{aligned}\tag{23}$$

We first characterize the optimal innovation subsidies without inter-firm knowledge networks. This exercise isolates the intergeneration externality of innovation in shaping optimal innovation subsidies since the network externality of innovation is absent in this case. In this special case, we can characterize the optimal innovation policy analytically as follows:

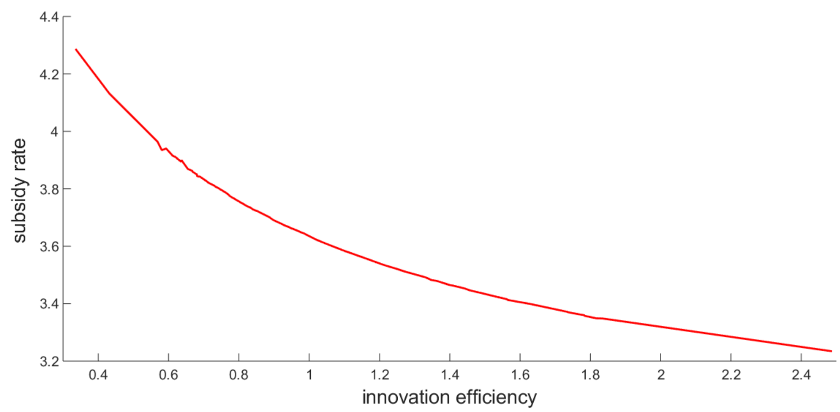
Proposition 5 (Optimal Innovation Subsidy without Spillovers) *Suppose that $\gamma=0$, i.e. inter-firm knowledge networks do not exist. Then the welfare-maximizing innovation* *subsidies* $\{s^*(z)\}$ *satisfy*

1. $s^*(z) = s^*$ *for all* z.

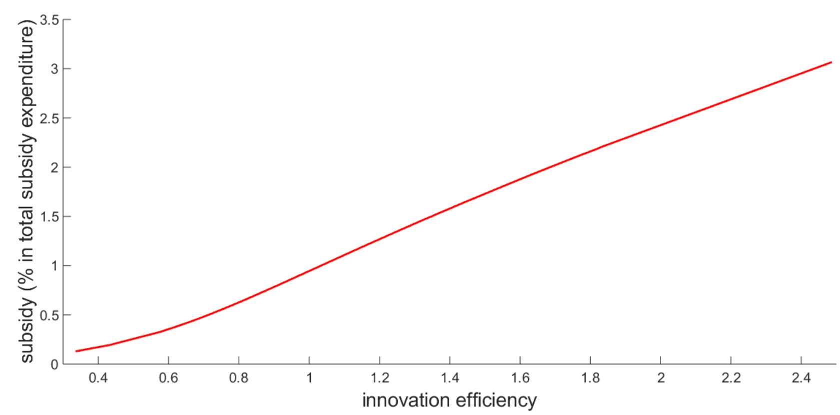
2. $s^* > 1$.

Proposition 5 suggests that without inter-firm knowledge networks the innovation subsidy rates should be uniform across firms. This is because the intergeneration externality of innovation is identical to all firms.

In the presence of inter-firm knowledge networks, the optimal innovation subsidy rates are no longer uniform. Panel (a) of Figure 17 suggests that optimal subsidy rates in our baseline case should decrease with respect to firm size. Our specification of the matching function $m(.,.)$ in Equation (13) and the estimates of parameters in Table 6 indicate that there are two forces of network externalities shaping optimal innovation policies.



(a) Subsidy Rate

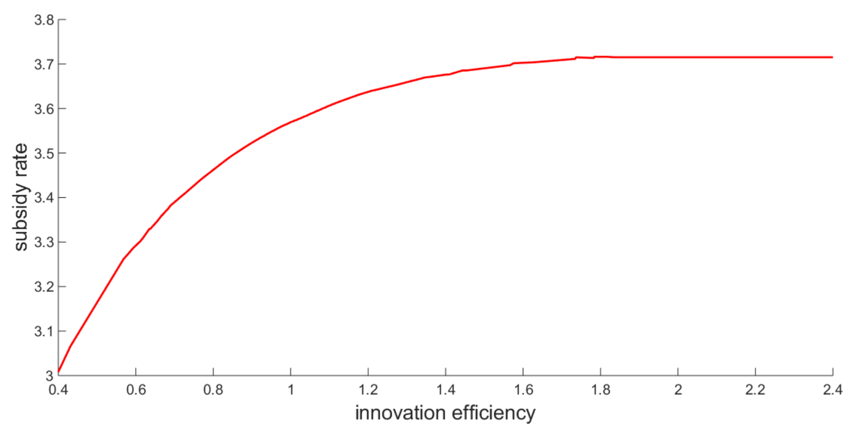


(b) Subsidy Expenditure

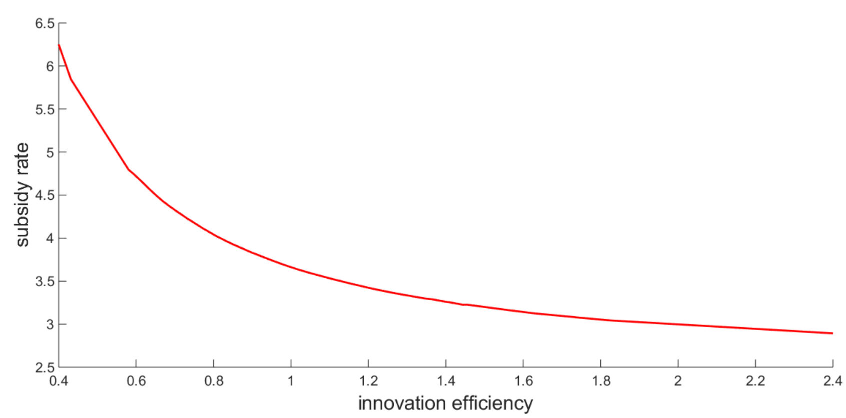
Figure 17: Optimal Innovation Policy: Subsidy Rates and Expenditures

First, since $\xi_1 > 0$ and $\rho > 0$, larger firms are more likely to be learned by other firms. Therefore, subsidizing these large firms can indirectly benefit other firms, which implies that the optimal innovation subsidies should increase with firm size. To isolate this channel, we let $\xi_2 = 0$ and $\rho = 0$, i.e. the matching rate varies only with the size of firms to be learned. In this scenario, as shown in Panel (a) of Figure 18, the optimal innovation subsidies monotonically increase with innovation efficiency. This result is consistent with Bloom et al. (2013).

Second, since $\xi_2 > 0$ and $\rho > 0$, larger firms also learn more form other firms. Therefore, these large firms can gain more from innovation subsidies received by other firms, which implies that the optimal subsidies should decrease with firm size. To see this effect, we let $\xi_1 = 0$ and $\rho = 0$, i.e. the matching rate varies only with the size of learning firms. In this scenario, as shown in Panel (a) of Figure 18, the optimal innovation subsidies monotonically decrease with innovation efficiency. This result is ignored by Bloom et al. (2013).



(a) $\rho = 0$ and $\xi_2 = 0$



(b) $\rho = 0$ and $\xi_1 = 0$

Figure 18: Optimal Innovation Policy: the Role of Network Externalities

In our baseline quantification, as shown in Figure 15, the optimal innovation subsidy rate is strictly decreasing with firm innovation efficiency, which implies that the second force dominates the first one above. Moreover, the optimal subsidy expenditure is still higher for larger firms since they spend more on innovation, as shown in Panel (b) of Figure 15.

In sum, our quantification of the optimal innovation subsidies justifies the policy practices aiming at subsidizing innovation in small and medium sized firms. The rationale is that these small and medium sized firms tend to be isolated in inter-firm knowledge networks; so they cannot benefit much from innovation subsidies received by other firms but rely more on direct subsidies.

5.4 International Trade

Recent progress of trade theories emphasizes the role of firms in shaping aggregate effects of trade liberalization (see Melitz (2003)). In particular, the literature has argued that (1) only a very small fraction of firms are engaged in international trade, and (2) exporting firms are larger and more productive than non-exporting firms. Therefore, the implications of international trade tend to vary systematically with firms' sizes and their positions in inter-firm knowledge networks. In this subsection, we incorporate international trade into the baseline model to understand the impacts of inter-firm knowledge networks on gains from trade liberalization. To compare our model with Melitz (2003), we consider the world with two symmetric countries. Still, we take the wage as a numeraire.

Exporting at period t incurs a fixed cost $f_t^X > 0$ in terms of labor and a standard iceberg cost $\tau_t \geq 1$. Then firm z exports at period t if and only if

$$\tilde{\Lambda}\left[z^{\frac{1}{\alpha-(\sigma-1)}}\phi_t(z)^{\frac{\alpha}{\alpha-(\sigma-1)}}\right]^{\sigma-1}\left[\left(1+\tau_t^{1-\sigma}\right)^{\frac{\alpha}{\alpha-(\sigma-1)}}-1\right]D_t^{\frac{\alpha}{\alpha-(\sigma-1)}} \geq f_t^X, \tag{24}$$

where the constant $\tilde{\Lambda} = \left[1-\frac{\sigma-1}{\alpha}\right]\tilde{\sigma}\left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\right]^{\frac{\sigma-1}{\alpha-(\sigma-1)}}$.

The marginal exporter $\tilde{z}_t$ satisfies

$$\tilde{\Lambda}\left[\tilde{z}_t^{\frac{1}{\alpha-(\sigma-1)}}\phi_t(\tilde{z}_t)^{\frac{\alpha}{\alpha-(\sigma-1)}}\right]^{\sigma-1}\left[\left(1+\tau_t^{1-\sigma}\right)^{\frac{\alpha}{\alpha-(\sigma-1)}}-1\right]D_t^{\frac{\alpha}{\alpha-(\sigma-1)}} = f_t^X. \tag{25}$$

The equilibrium innovation thereby can be expressed as

$$\kappa_t^*(z) = \begin{cases} \left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\left(1+\tau_t^{1-\sigma}\right)D_t\right]^{\frac{1}{\alpha-(\sigma-1)}}\left[z\phi_t(z)^{\sigma-1}\right]^{\frac{1}{\alpha-(\sigma-1)}}, & \text{if } z \geq \tilde{z}_t. \\ \left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}D_t\right]^{\frac{1}{\alpha-(\sigma-1)}}\left[z\phi_t(z)^{\sigma-1}\right]^{\frac{1}{\alpha-(\sigma-1)}}, & \text{if } z < \tilde{z}_t. \end{cases} \tag{26}$$

Then the productivity $\phi_{t+1}(z)$ is given by Equation (7). The aggregate price index can be expressed as

$$P_t = \frac{\sigma}{\sigma-1}\left[\int_{S_z} \left[\kappa_t^*(z)\phi_t(z)\right]^{\sigma-1} dG(z) + \int_{\tilde{z}_t}^{\infty} \tau_t^{1-\sigma}\left[\kappa_t^*(z)\phi_t(z)\right]^{\sigma-1} dG(z)\right]^{\frac{1}{1-\sigma}}. \tag{27}$$

Total expenditure satisfies

$$X_t = L + \frac{1}{\sigma}\left(1 - \frac{\sigma-1}{\alpha}\right) X_t - \left[1 - G(\tilde{z}_t)\right] f_t^X. \tag{28}$$

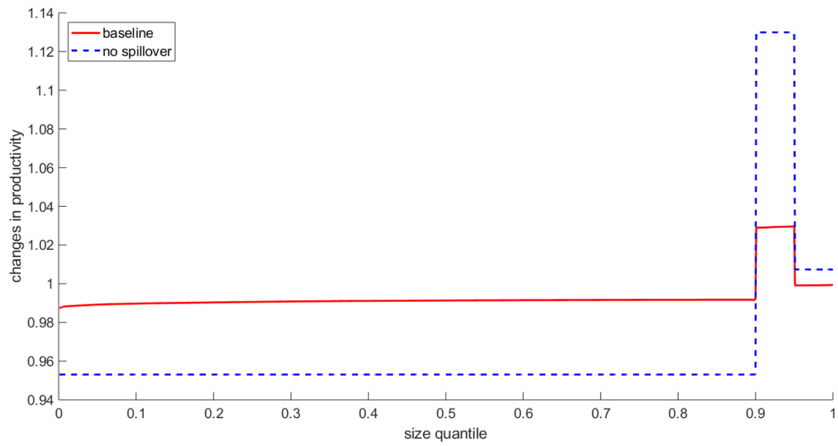
5.4.1 Productivity Effects of Trade Liberalization

We set parameters $\{\alpha, \beta, \sigma, \sigma_z^2, \delta, \gamma, \xi_1, \xi_2, \rho\}$ as estimated above and then study trade liberalization by reducing the iceberg trade cost τ_t or the fixed trade cost f_t^X. First, we keep the fixed trade cost $f_t^X = 1.4$ and reduce the iceberg trade cost permanently so that the exporter share in each country increases from 5% to 10%. The same exercise is conducted without inter-firm knowledge networks, i.e. $\gamma = 0$. The productivity effects of trade liberalization under our baseline model and the model without knowledge spillover are illustrated in Panel (a) of Figure 19. As expected, trade liberalization benefits the firm between 90 and 95 size quantiles most and leads to productivity losses for smaller firms due to tougher competition. This result is analogous to the reallocation effect of trade liberalization in Melitz (2003). Comparing our baseline model to the model without knowledge spillovers, we find that inter-firm knowledge networks *mitigate* this reallocation effects of trade liberalization by diffusing the exporters' knowledge to domestic producers. In other words, domestic producers indirectly benefit from trade liberalization via inter-firm knowledge networks.

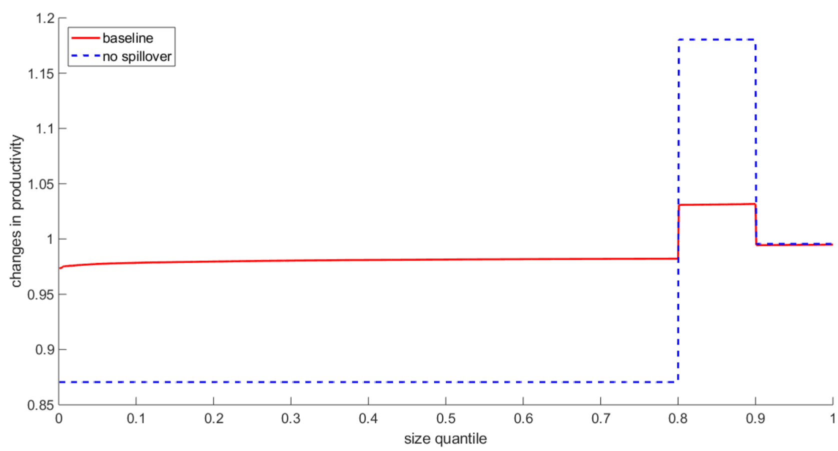
Panel (b) of Figure 19 illustrates the productivity effects of trade liberalization that increases exporter share in each country from 10% to 20%. The effects of inter-firm knowledge networks are substantial in this case: in our baseline model, trade liberalization reduces the productivities of domestic producers by 2.5%, whereas in the model without inter-firm knowledge networks, trade liberalization reduces the productivities of domestic producers by 13%. In sum, the knowledge diffusion from exporters to domestic producers is quantitatively important to understanding the productivity effects of trade liberalization.

5.4.2 Welfare Gains from Trade and Trade Liberalization

In this section, we discuss welfare gains from trade and trade liberalization. First, we study the welfare gains from trade in our model, asking how much the real income will in-



(a) Exporter Share increases from 5% to 10%



(b) Exporter Share increases from 10% to 20%

Figure 19: Productivity Effects of Trade Liberalization: the Decline in Iceberg Trade Costs

(Note: the fixed trade cost is set as $f_t^X = 1.4$.)

crease from autarky to an economy with certain level of import share. This exercise follows Arkolakis, Costinot, and Rodriguez-Clare (2011) in which the import share is a sufficient statistics for welfare gains from trade. To understand the role of inter-firm knowledge networks, we compute gains from trade both in our baseline model and in the model without knowledge spillovers. Figure 20 plots the real income against import share, with the autarkic real income normalized into 1. The results show that inter-firm knowledge networks tend to *reduce* welfare gains from trade.

Why do inter-firm knowledge networks *reduce* welfare gains from trade? Without inter-firm knowledge spillovers, a firm can benefit from export markets only by directly exporting. In this case, a move back to autarky will lead to a substantial decline in productivity. In contrast, in our baseline model, a firm can indirectly benefit from export markets by learning from exporters. So inter-firm knowledge networks tend to reduce welfare gains from trade.

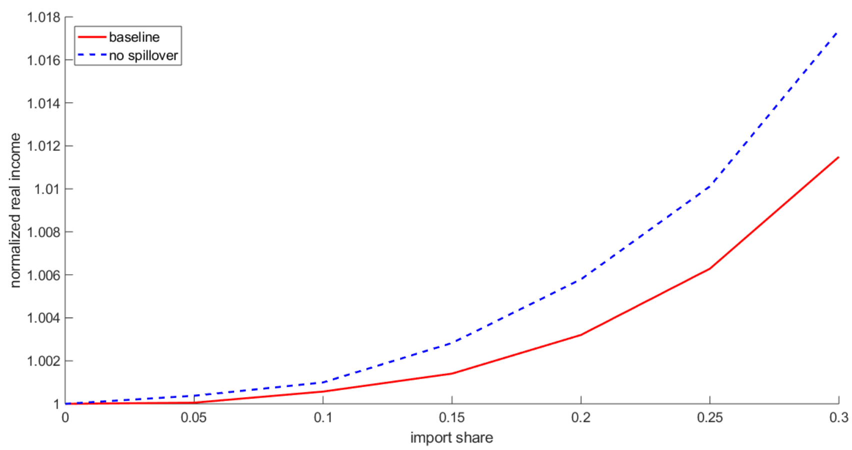


Figure 20: Welfare Gains from Trade

(Note: the autarkic real income is normalized into 1. We fix $f_t^X = 1.4$ and adjust τ_t.)

Second, we explore welfare gains from trade liberalization. To achieve this, we start from an economy in which the import share is 0.2, decreasing iceberg trade costs and computing changes in welfare accordingly. Again, we conduct this exercise both in our baseline model and in the model without inter-firm knowledge networks. Figure 21 plots changes in welfare against the decline in τ. It shows that inter-firm knowledge networks tend to magnify welfare gains from trade liberalization.

Why do inter-firm knowledge networks have different implications for gains from trade and gains from trade liberalization? Notably, gains from trade are conditional on import shares both before and after change, whereas gains from trade liberalization are not conditional on the import share after change. In the presence of inter-firm knowledge networks, the same trade liberalization tends to yield more export entry due to the mitigation of firm reallocation discussed above.

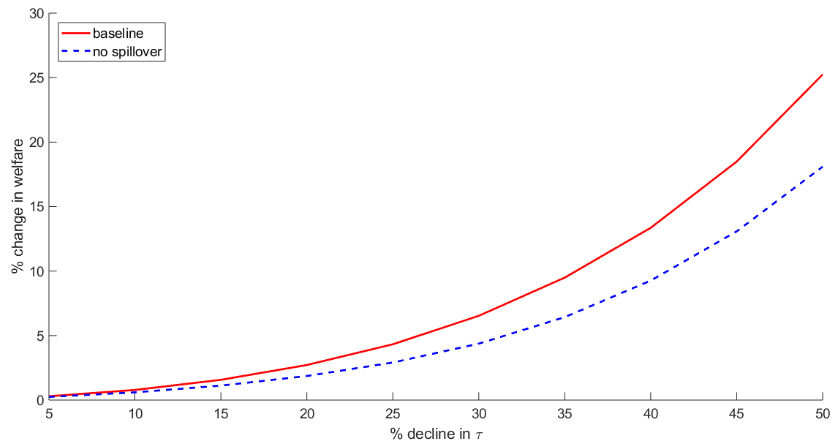


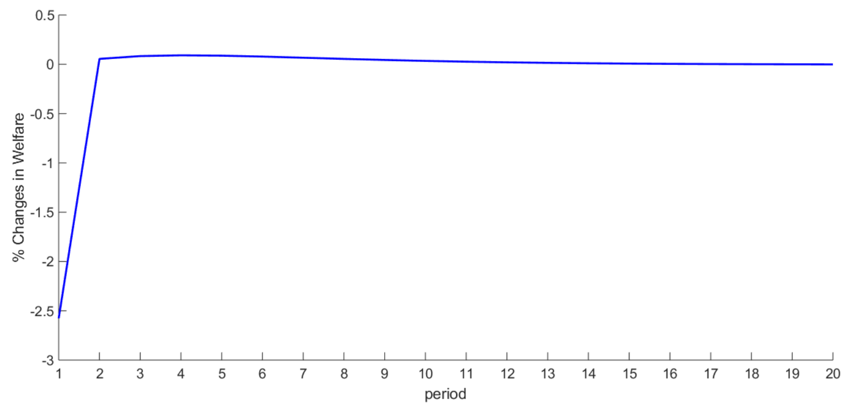
Figure 21: Welfare Gains from Trade Liberalization
(Note: the benchmark is an economy with import share 0.2. We fix $f_t^X = 1.4$.)

5.4.3 Trade Dynamics

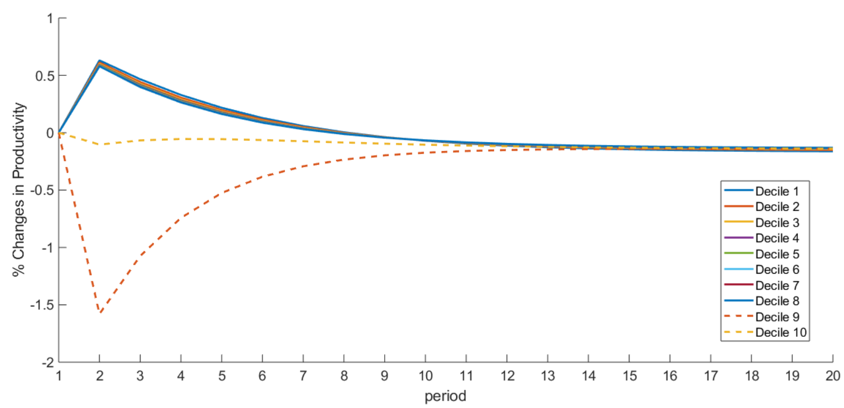
Our model can also be used to study the dynamics after trade collapse, as the one during Great Recession. To do this, we fix $f_t^X = 2$. Then we start from an economy with import share 0.3 and increase iceberg trade costs temporarily in period 1 so that the import share decreases to 0.25.

Panel (a) of Figure 22 suggest that the dynamic effects of trade collapse is not persistent. A temporary trade collapse will decrease the exporters' innovation for one period. But this effect is quantitatively small and partially offset by the increase in the domestic producers' innovation.

Panel (b) of Figure 22 suggests that the productivity dynamics after trade collapse vary across firms. Firms in decile 9 (size quantile between 80 and 90) incur the largest decline in productivities since the trade collapse turns them into domestic producers. Interestingly, the top 10% firms also experience the decline in productivities. This is because (1) trade collapse reduces their export sales, and (2) they learn less from firms in decile 9. All other firms experience productivity improvements due to the ease of competition.



(a) Welfare



(b) Productivity

Figure 22: Dynamics after a Temporary Trade Collapse

(Note: we start from an economy with import share 0.3 and increase iceberg trade costs temporarily in period 1 so that the import share decreases to 0.25. We fix $f_t^X = 2$.)

6 Conclusion

This paper quantifies the impacts of inter-firm knowledge networks on aggregate growth and welfare. We first characterize the structure of inter-firm knowledge networks using data on patent citations across Chinese manufacturing firms. We find that within a narrowly-defined industry firms still occupy heterogeneous positions in patent citation networks, and that a firm tends to grow faster if it cites patents from faster-growing firms. These findings highlight the importance of inter-firm knowledge networks to technology diffusion and improvements.

Motivated and guided by our empirical findings, we develop a tractable model that links the micro structure of firm innovation and knowledge diffusion with the macro technology progress. Despite rich firm heterogeneity and flexible knowledge networks, our model remains tractable and yields simple structural equations that characterize both the steady-state and transitional dynamics of the equilibrium. We provide sufficient conditions for the existence and uniqueness of steady-state and develop simple algorithms to compute the model's steady-state as well as transitional dynamics.

We estimate key model parameters using data on patent citations by a simulated method of moments regarding equilibrium conditions as constraints. The estimated model fits the data well both in the targeted and untargeted moments. Armed by the estimated model, we conduct counterfactual exercises and find that (i) eliminating inter-firm knowledge networks will reduce the aggregate welfare by more than 50%; (ii) a firm's impacts on the aggregate economy depend crucially on its position in inter-firm knowledge networks; (iii) the welfare-maximizing innovation subsidies are decreasing with firm size; and (iv) inter-firm knowledge networks can mitigate the reallocation effect of trade liberalization in Melitz (2003) by diffusing the exporters' technologies to domestic producers.

We leave several interesting extensions to future explorations. First, it is interesting to rationalize our exogenous matching function by search and matching mechanisms that lead to endogenous network formation. Though it is challenging for these endogenous networks to fit the inter-firm knowledge networks observed in the data. Second, we can incorporate more firm characteristics such as foreign ownership and export status, and more bilateral relationships such as physical distance, into our matching function. This extension enables us to discuss important topics such as the geography of knowledge diffusion and the knowledge spillovers from FDI. Third, our model can be extended to quantify the effects of innovation policies in reality, such as China's tax reduction for high-tech firms.

References

[1] Acemoglu, D., Akcigit, U., Kerr, W. (2016) "Innovation Network" *Proceedings of the National Academy of Sciences*, 113(41).

[2] Acemoglu, D., and Azar, P. (2018) "Endogenous Production Networks" *mimeo.*

[3] Acemoglu, D., Carvalho, V., Ozdaglar, A., Tahbaz-Salehi, A. (2012) "The Network Origins of Aggregate Fluctuations" *Econometrica* 80(5).

[4] Aghion, P., Bergeaud, A., Lequien, M., Melitz, M. (2018) "The Impact of Exports on Innovation: Theory and Evidence" *mimeo.*

[5] Alvarez, F., Buera, F., Lucas, R. (2014) "Idea Flows, Economic Growth, and Trade" *mimeo.*

[6] Andrews, I., Gentzkow, M., and Shapiro, JM. (2017). "Measuring the Sensitivity of Parameter Estimates to Estimation Moments" *Quarterly Journal of Economics*, Vol. 132, no. 4, 1553-1592.

[7] Arkolakis, C., Costinot, A., and Rodriguez-Clare, A. (2011). "New Trade Models, Same Old Gains?" *American Economic Review*, Vol. 102, no. 1, 94-130.

[8] Autor, D., Dorn, D., Hanson, G. (2013) "The China syndrome: Local labor market impacts of import competition in the united states" *American Economic Review* 103(6).

[9] Bernard, A., Moxnes, A., Saito, Y. (2015) "Production Networks, Geography and Firm Performance" *NBER working paper.*

[10] Bernard, A., Moxnes, A., Ulltveit-Moe, K. (2016) "Two-sided Heterogeneity and Trade" *mimeo.*

[11] Bloom, N., Schankerman, M., Reenen, J. (2013) "Identifying Technology Spillovers and Product Market Rivalry" *Econometrica*, 81(4).

[12] Borusyak, K., Hull, P., Jaravel, X. (2018) "Quasi-experimental shift-share research designs" *mimeo.*

[13] Buera, F., Oberfield, E. (2017) "The Global Diffusion of Ideas" *mimeo.*

[14] Cai, J., Li, N. (2018) "Growth Through Inter-sectoral Knowledge Linkages" *Review of Economic Studies*, *forthcoming.*

[15] Cohen, W., Klepper, S. (1996) "Firm Size and the Nature of Innovation within Industries: the Case of Process and Product R&D" *Review of Economics and Statistics* 78(2).

[16] Costinot, A., Rodriguez-Clare, A. (2014) "Trade Theory with Numbers: Quantifying the Consequences of Globalization" *Handbook of International Economics.*

[17] Desmet, K., Nagy, D., Rossi-hansberg, E. (2018) "The Geography of Development" *Journal of Political Economy*, 126(3).

[18] Desmet, K., Rossi-hansberg, E. (2014) "Spatial Development" *American Economic Review*, 104(4).

[19] Giuliani, E (2007) "The selective nature of knowledge networks in clusters: Evidence from the wine industry" *Journal of Economic Geography*, 7(2).

[20] Hulten, C.R. (1978) "Growth accounting with intermediate inputs" *The Review of Economic Studies*, 511-518.

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

[22] Jaffe, A., Trajtenberg, M., Fogarty, M. (2000) "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors" *American Economic Review, Papers and Proceedings* 90(2).

[23] Jaffe, A. B., Trajtenberg, M., Henderson, R. (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations" *The Quarterly Journal of Economics*, 108(3).

[24] Klette, J., Kortum, S. (2004) "Innovating Firms and Aggregate Innovation" *Journal of Political Economy*, 112(5).

[25] Lim, K. (2017) "Firm-to-firm Trade in Sticky Production Networks" *mimeo.*

[26] Melitz, M. (2003). "The impact of trade on intra-industry reallocations and aggregate industry productivity". *Econometrica*, 71(6).

[27] Oberfield, E. (2018) "A Theory of Input-Output Architecture" *Econometrica* 86(2).

[28] Tintelnot, F., Ken, A., Mogstad, M., Dhyne, E. (2017) "Trade and Domestic Production Networks" *mimeo*.

Appendix A Theories

A.1 Proof to Proposition 3

Proof. In the steady-state, we have $\phi_t(z) = \phi(z)$ for all t. Inserting Equation (6) into Equation (7), we have

$$\phi(z) = \tilde{D}^{\beta}\left[z^{\frac{1}{\alpha-(\sigma-1)}}\phi(z)^{\frac{\alpha}{\alpha-(\sigma-1)}} + \delta\int_{S_z} m(z',z)\,(z')^{\frac{1}{\alpha-(\sigma-1)}}\,\phi(z')^{\frac{\alpha}{\alpha-(\sigma-1)}}dG(z')\right]^{\beta}, \quad \forall z \in S_z. \tag{29}$$

Rearranging Equation (29), we have

$$\phi(z)^{\frac{1}{\beta}} = \int_{S_z}\tilde{D}\left[\mathbf{1}(z'=z) + \delta m(z',z)\right](z')^{\frac{1}{\alpha-(\sigma-1)}}\,\phi(z')^{\frac{\alpha}{\alpha-(\sigma-1)}}dG(z'). \tag{30}$$

Let $\tilde{\phi}(z) = \left[\tilde{D}^{\frac{1}{\frac{\alpha\beta}{\alpha-(\sigma-1)}-1}}\phi(z)\right]^{\frac{1}{\beta}}$.

$$\tilde{\phi}(z) = \int_{S_z}\left[\mathbf{1}(z'=z) + \delta m(z',z)\right](z')^{\frac{1}{\alpha-(\sigma-1)}}\,\tilde{\phi}(z')^{\frac{\alpha\beta}{\alpha-(\sigma-1)}}dG(z'). \tag{31}$$

Since $1 - \frac{\alpha\beta}{\alpha-(\sigma-1)} > 0$, by Theorem 1 of Allen, Arkolakis, and Li (2017), there exists a unique solution to Equation (31) and the solution can be computed by a simple iteration procedure. Notice that price index P can be computed by Equation (8) and the total expenditure X is exogenous. Given the unique $\{\tilde{\phi}(z)\}$, the price index P is unique and so is $\{\phi(z)\}$. Therefore, our steady-state equilibrium is unique.

■

A.2 Proof to Proposition 5

Proof. Without spillovers, we have

$$\phi_i = \tilde{D}^{\frac{\beta}{1-\frac{\alpha\beta}{\alpha-(\sigma-1)}}}\,(s_i z_i)^{\frac{1}{\alpha-(\sigma-1)}\frac{\beta}{1-\frac{\alpha\beta}{\alpha-(\sigma-1)}}}, \tag{32}$$

where

$$\tilde{D} = \left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\right]^{\frac{1}{\alpha-(\sigma-1)}} D^{\frac{1}{\alpha-(\sigma-1)}}, \quad D = P^{\sigma-1}X. \tag{33}$$

Therefore, we have

$$\begin{aligned}
&P = \frac{\sigma}{\sigma-1}\left\{\sum_{i=1}^{N}\left[\tilde{D}(s_i z_i)^{\frac{1}{\alpha-(\sigma-1)}}\phi_i^{\frac{\alpha}{\alpha-(\sigma-1)}}\right]^{\sigma-1}\right\}^{\frac{1}{1-\sigma}} \\
&\Rightarrow P = \frac{\sigma}{\sigma-1}\tilde{D}^{-\frac{\alpha-(\sigma-1)}{\alpha-(\sigma-1)-\alpha\beta}}\left\{\sum_{i=1}^{N}\left[(s_i z_i)^{\frac{1}{\alpha-(\sigma-1)-\alpha\beta}}\right]^{\sigma-1}\right\}^{\frac{1}{1-\sigma}} \\
&\Rightarrow P = \frac{\sigma}{\sigma-1}\left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\right]^{-\frac{1}{\alpha-(\sigma-1)-\alpha\beta}}\left(P^{\sigma-1}X\right)^{-\frac{1}{\alpha-(\sigma-1)-\alpha\beta}}\left\{\sum_{i=1}^{N}\left[(s_i z_i)^{\frac{1}{\alpha-(\sigma-1)-\alpha\beta}}\right]^{\sigma-1}\right\}^{\frac{1}{1-\sigma}} \\
&\Rightarrow P = \Lambda_1 X^{-\frac{1}{\alpha(1-\beta)}}\left\{\sum_{i=1}^{N}\left[(s_i z_i)^{\frac{1}{\alpha-(\sigma-1)-\alpha\beta}}\right]^{\sigma-1}\right\}^{\frac{1}{1-\sigma}\frac{\alpha-(\sigma-1)-\alpha\beta}{\alpha(1-\beta)}}.
\end{aligned} \tag{34}$$

On the other hand, we have

$$X = \frac{1}{1-\frac{1}{\sigma}\left(1-\frac{\sigma-1}{\alpha}\right)}L - \frac{\frac{1}{\sigma}\frac{\sigma-1}{\alpha}}{1-\frac{1}{\sigma}\left(1-\frac{\sigma-1}{\alpha}\right)}\sum_{i=1}^{N}(s_i-1)X_i, \tag{35}$$

where

$$\begin{aligned}
&X_i = \sigma\tilde{\sigma}\left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\right]^{\frac{\sigma-1}{\alpha-(\sigma-1)}}(s_i z_i)^{\frac{\sigma-1}{\alpha-(\sigma-1)}}\phi_i^{\frac{\alpha(\sigma-1)}{\alpha-(\sigma-1)}}\left(P^{\sigma-1}X\right)^{\frac{\alpha}{\alpha-(\sigma-1)}} \\
&\Rightarrow X_i = \sigma\tilde{\sigma}\left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\right]^{\frac{\sigma-1}{\alpha-(\sigma-1)-\alpha\beta}}(s_i z_i)^{\frac{\sigma-1}{\alpha-(\sigma-1)-\alpha\beta}}\left(P^{\sigma-1}X\right)^{\frac{\alpha(1-\beta)}{\alpha-(\sigma-1)-\alpha\beta}}.
\end{aligned} \tag{36}$$

Then the first order condition of the optimal subsidy for firm i can be expressed as

$$\begin{aligned}
&\frac{\partial}{\partial s_i}(\log X - \log P) = 0 \\
&\Rightarrow \left(1+\frac{1}{\alpha(1-\beta)}\right)\frac{\frac{1}{\sigma}\frac{\sigma-1}{\alpha}}{1-\frac{1}{\sigma}\left(1-\frac{\sigma-1}{\alpha}\right)}\frac{1}{X}\left(X_i + s_i\frac{\partial X_i}{\partial s_i}\right) \\
&= \frac{1}{1-\sigma}\frac{\alpha-(\sigma-1)-\alpha\beta}{\alpha(1-\beta)}\frac{\frac{\sigma-1}{\alpha-(\sigma-1)-\alpha\beta}(s_i z_i)^{\frac{\sigma-1}{\alpha-(\sigma-1)-\alpha\beta}}/s_i}{\sum_{i=1}^{N}\left[(s_i z_i)^{\frac{1}{\alpha-(\sigma-1)-\alpha\beta}}\right]^{\sigma-1}}.
\end{aligned} \tag{37}$$

Notice that $X_i + s_i\frac{\partial X_i}{\partial s_i} = \left(1+\frac{\sigma-1}{\alpha-(\sigma-1)-\alpha\beta}\right)X_i$. We also have

$$\frac{X_i}{z_i^{\frac{\sigma-1}{\alpha-(\sigma-1)-\alpha\beta}}} = \sigma\tilde{\sigma}\left[\frac{(\sigma-1)\tilde{\sigma}}{\alpha}\right]^{\frac{\sigma-1}{\alpha-(\sigma-1)-\alpha\beta}} s_i^{\frac{\sigma-1}{\alpha-(\sigma-1)-\alpha\beta}}\left(P^{\sigma-1}X\right)^{\frac{\alpha(1-\beta)}{\alpha-(\sigma-1)-\alpha\beta}}. \tag{38}$$

So the optimal $s_i^* = s^*$ for any i.

Hence, we have

$$X = \frac{1}{1 - \frac{1}{\sigma} + \frac{1}{\sigma}\frac{\sigma-1}{\alpha}s}L, \tag{39}$$

and

$$P = \Lambda_1 X^{-\frac{1}{\alpha(1-\beta)}} \left\{ \sum_{i=1}^{N} \left[(z_i)^{\frac{1}{\alpha-(\sigma-1)-\alpha\beta}} \right]^{\sigma-1} \right\}^{\frac{1}{1-\sigma}\frac{\alpha-(\sigma-1)-\alpha\beta}{\alpha(1-\beta)}} s^{-\frac{1}{\alpha(1-\beta)}}. \tag{40}$$

It is straightforward to verify that

$$\frac{\partial\left(\log X - \log P\right)}{\partial s}|_{s=1} > 0, \tag{41}$$

and

$$\frac{\partial^2\left(\log X - \log P\right)}{\partial s^2} < 0. \tag{42}$$

So we have $s^* > 1$. ■

Appendix B Empirics and Quantification

B.1 Technology Progress along Citation Networks: Robustness

One concern for identification is that many citation linkages are two-way (see Figure 1). Therefore, $\Delta \log \phi_{it}$ could reversely affect $\Delta \log \phi_{it}^{\mathrm{NX}}$ through citation networks. To mitigate this concern, we regress $\log \phi_{jt}$ on $\log \phi_{jt}^{\mathrm{NX}} := \log \sum_k w_{kt}\phi_{kt}$, denote the residual of this auxiliary regression as $\hat{\phi}_{jt}$, and construct the explanatory variable by

$$\Delta \log \phi_{it}^{\mathrm{NX}} = \log \sum_j w_{jt}\hat{\phi}_{jt} - \log \sum_j w_{jt-1}\hat{\phi}_{jt-1}. \tag{43}$$

Doing this, we eliminate the technologies of firm j that comes from the firms whose patents it cites and focus on the diffusion of firm j's own technologies to firm i.

B.2 Computation of transitional dynamics

Algorithm 6 *Starting from an initial state $\{\phi_0(z)\}$, the transitional dynamics $\{\phi_t(z)\}$ can be computed as*

1. *Guess $\tilde{D}_0$.*

2. *Compute $\kappa_0^*(z)$ by Equation (6) and thereby P_0 by Equation (8).*

3. *Update $\tilde{D}_0$ by Equation (6). Iterate until converge.*

4. *Compute $\{\phi_1(z)\}$ by Equation (7).*

5. *Solve $\tilde{D}_1$ by iteration. Compute $\{\phi_t\}_{t=2}^{\infty}$ successively.*

B.3 Welfare Dynamics under Firm-Level Shocks

In this subsection, we discuss the welfare effects of firm-level shocks. We compare the welfare dynamics of the negative productivity shock on the top 10% firms to the one on the bottom 10% firms. The results are shown in Figure 23. As expected, the shocks on the top 10% firms account for about 25% of the aggregate welfare decline, while the shocks on the bottom 10% firms account for merely 2.5%. This is intuitive since the largest firms do not only have the best fundamental characteristics (innovation efficiencies), but also a large number of knowledge connections with other firms. The negative shocks on the top 10% firms propagate via inter-firm knowledge networks, as illustrated in Figure 15, leading to sizable aggregate welfare losses.

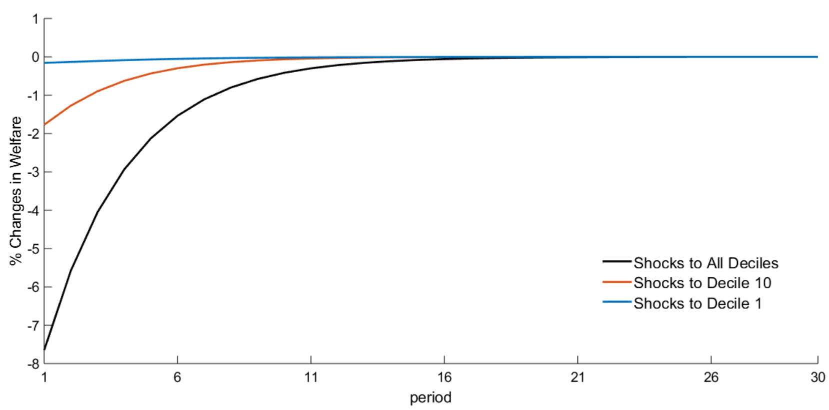
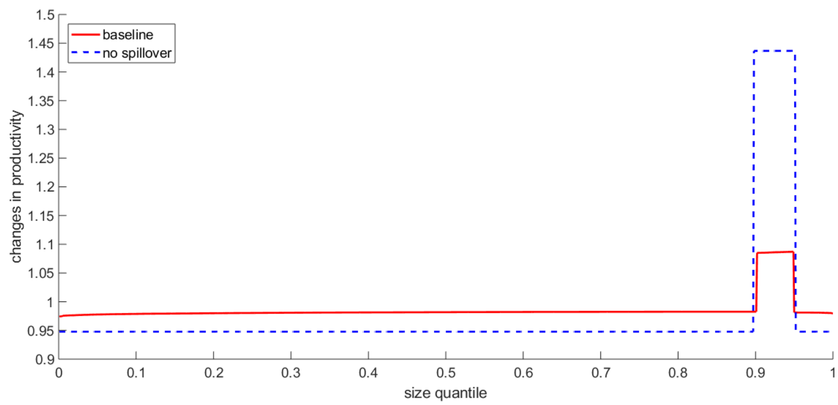
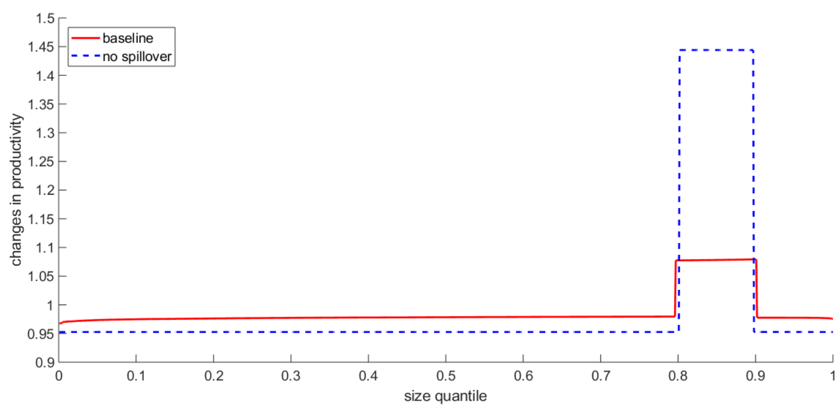


Figure 23: Welfare Effects of the Productivity Shocks

B.4 The Decline in Fixed Trade Costs



(a) Exporter Share increases from 5% to 10%



(b) Exporter Share increases from 10% to 20%

Figure 24: Productivity Effects of Trade Liberalization: the Decline in Fixed Trade Costs