Agglomeration and Its Mechanisms on Exporting Directly

Dakshina G. De Silva* Hojin Jung† Qianqian Wang‡

July 2024

Abstract

This paper investigates the effects of agglomeration on manufacturing firms' export mode decision-making. A detailed firm-level dataset combines the manufacturing and product-level transaction trade data from China. We use a dynamic multinomial logit model with random effects to analyze the effects of agglomeration on the transition probabilities of firms' exporting mode. The results indicate that the agglomeration of direct exporters positively affect firm's choice of exporting directly, and further indicate that the agglomeration effect is identified through productivity spillover. Moreover, these effects are not destination-specific, but are industry-specific.

JEL classification: F10, F14

Keywords: Agglomeration; Firms' Export Mode; Productivity Spillover; Industry-Specific

*Lancaster University, Lancaster, LA2 9QZ, UK; email: d.desilva@lancaster.ac.uk

†Jeonbuk National University, Jeonju-si, Jeollabuk-do 54896, Korea; e-mail: hojin.jung@jbnu.ac.kr

‡Corresponding author: School of Economics, Henan University; email: angelstella02@gmail.com

1 Introduction

Manufacturing firms can choose to export their products directly or indirectly (through a trade intermediary). This is known as their export mode. A significant amount of exports are done through trade intermediaries. For example, 10% of U.S. exports (Bernard *et al.*, 2010), 22% of Chinese exports (Ahn *et al.*, 2011) and 35% of Chilean imports (Blum *et al.*, 2010) are handled by intermediaries. He *et al.* (2023) conducted a recent study examining the dual role of trade intermediaries in firms' export market diversification. Their empirical findings indicate that while these intermediaries facilitate market entry and knowledge transfer, they can also impose discretionary market barriers that limit firms' direct exporting activities.

Previous research has documented that firms choose their export modes according to their productivity levels. The cost of information acquisition is a key component of the costs of entry into export markets (Roberts and Tybout, 1997). Information costs also influence export mode decision-making (Petropoulou, 2011). Ahn *et al.* (2011) discovered that the most productive firms self-select into direct exporting as they are prepared to bear the initial fixed costs associated with learning the export process. Wang and Gibson (2018) further expanded this theoretical model by considering quality heterogeneity and found that firms with the highest levels of quality-adjusted productivity are more likely to engage in direct exporting.

We observe that both Chinese direct and indirect exporters locate themselves near coastal regions. This is probably mainly due to their access to seaports with lower transportation costs. Further, we observe a relatively high transition rate from indirect to direct exporters. Cassey and Wang (2022) found that in firms choosing to export directly, spatial spillover effects are several times more important than productivity. Duranton and Puga (2004) identified sharing, matching, and learning as key mechanisms at the microeconomic level that underpin agglomeration economies. Rosenthal and Strange (2004) distinguished the empirical literature on agglomeration economies through three sources of urban increasing returns: labor market pooling, input sharing, and knowledge spillovers. In this paper, we emphasize the agglomeration and its mechanisms on exporting directly. We study how the neighboring firms' exporting mode (agglomeration by exporting type) affects a firm's export mode choice, and find the effects are industry specific, but not destination specific. However, the export spillovers observed are specific to certain industries and/or destinations in Koenig (2009) and Koenig *et al.* (2010).

Entering the foreign markets will entail high sunk costs, entry as direct exporter will entail even higher sunk costs than as indirect exporter. There is a relatively high probability of firms entering as indirect exporter and then switch to be direct exporter. Self-learning plays a crucial role in these dynamics, yet in practice, firms often gather essential information from neighboring firms prior to undertaking substantial investments (Hausman and Rodrik, 2003). Our model illustrates how insights gained from nearby firms influence the entry decisions and subsequent export mode transitions of new exporters. By following Fernandes and Tang (2014), we consider a firm's export profits in a market to be influenced by three factors: firm-specific productivity, the unique appeal of its products in a specific market, and the overall market demand. Before entering a new market, an exporter is aware of its productivity but may not know the specific demand in the country or how appealing its products will be there. By observing the export outcomes of neighboring firms in that market, a firm can refine its initial assumptions about the demand common to all firms. We show that a firm's export mode choice decision and post export mode switching depend on the firm's own prior knowledge about the market, which is reflected by the average neighbors' export growth in our model.

Zhang *et al.* (2021) empirically investigated how firm heterogeneity and agglomeration affect urban exports in Chinese cities. They found that diverse firm productivity and various types of agglomeration exert distinct impacts on the productivity and export levels across different cities. Specifically, heterogeneous productivity and specialized agglomeration positively influence exports in central regions, whereas diversified agglomeration benefits eastern and northeastern regions. Total factor productivity was also found to spread geographically from one firm to another (Ertur and Koch, 2007; Baltagi *et al.*, 2016). Greenstone *et al.* (2010), Ellison *et al.* (2010), Bloom *et al.* (2013), Faggio *et al.* (2017), Hanlon and Miscio (2017), among others, also show that the productivity of firms rises with the presence of nearby connected firms. Despite this broad evidence for productivity spillovers of agglomeration economies, little empirical evidence exists about the impacts of productivity spillovers on the firm entry to foreign markets regarding to their export mode choice and their post export mode switching. Therefore, we further investigate the agglomeration process through the lens of productivity spillover by averaging regional (district) industry-level total factor productivity.

We constructed detailed firm-level data between 2001-2007 by merging manufacturing production data with firm-specific product-level transaction trade data from China. We then defined the direct, indirect, and non-exporting firms with the merged data. In Feng *et al.* (2022), it is argued that under export capacity constraints, the most productive manufacturing firms are termed "dual-channel exporters." These firms export a portion of their products directly and the remainder through intermediaries. While we follow Bai *et al.* (2017), firms that report exports exceeding those in customs data engage in both direct and indirect exporting, and are categorized as direct exporters in this study. Ultimately, we created an export mode transition matrix. In our empirical analysis, we used a dynamic multinomial logit with random effects to analyze the effects of agglomeration and its mechanisms on the transition probabilities of firms' choice of export mode.

Our empirical results indicate that state dependence is a crucial factor, as a firm's export mode choice is significantly affected by its own prior choices. Second, a firm's choice of export mode is positively and significantly affected by its previous productivity. More importantly, we find that the neighboring firms' exporting mode (agglomeration by exporting type) affects a firm's export mode, and these effects are industry-specific, but not destination-specific. We further find that productivity spillover from neighbors strongly supports the existence of agglomeration effects on a firm's choice of export mode.[1]

2 Theoretical Motivation

Consider a firm i deciding whether to export to country j, if export, whether export directly or indirectly through intermediaries. Bai *et al.* (2017) outline that the revenues from local markets, indirect exports, and direct exports are characterized as follows:

$$\ln r_{it}^{H} = a^{H} + \ln \Phi_t^{H} + \left(1 - \sigma^{H}\right)\left(\beta_0 + \beta_k k_{it} + \beta_t D_t + \beta_s D_s + \beta_l D_l - \omega_{it}\right), \tag{1}$$

$$\ln r_{it}^{Xm} = a^{X} + \ln \Phi_t^{X} + \left(1 - \sigma^{X}\right)\left(\beta_0 + \beta_k k_{it} + \beta_t D_t + \beta_s D_s + \beta_l D_l - \omega_{it}\right) + z_{it} - d_{it}^{I}\sigma^{X}\ln\lambda, \tag{2}$$

where $a^j = \left(1 - \sigma^j\right)\ln\left(\frac{\sigma^j}{\sigma^j - 1}\right)$ and $\Phi_t^j = \frac{Y_t^j}{\left(P_t^j\right)^{1-\sigma^j}}, j = H, X$, where H denotes the home market and X the foreign market. Substitution between domestic goods is parameterized by σ^H which differs from that between foreign goods parameterized by σ^X. Firm's revenues in each market is influenced by overall market conditions, represented by Φ_t^H and Φ_t^X, along with firm-specific productivity ω_{it}, capital stock k_{it}, and specific factor prices proxied by the year dummy D_t, industry-level dummy D_s, and location dummy D_l. Additionally, revenue from foreign markets also hinges on the chosen export method and a unique demand shock in the foreign market, z_{it}. The logarithm of revenue from indirect exports is consistently lower than from direct exports by the value of $\sigma^X \ln\lambda$.

We assume the Dixit-Stiglitz model of consumer preferences and monopolistic competition, then the profits of a firm in its home market,in foreign markets with indirect and direct export method can be expressed as

$$\pi_{it}^{H} = \frac{1}{\sigma^H} r_{it}^{H}(\Phi_t^H, w_{it}, \omega_{it}) \tag{3}$$

$$\pi_{it}^{XI} = \frac{1}{\sigma^X} r_{it}^{XI}(\Phi_t^X, w_{it}, \omega_{it}, z_{it}, \lambda) \tag{4}$$

[1] Note that in our empirical analysis, the productivity spillovers are industry-specific rather than destination-specific. Indirect exporters also produce final goods; the primary distinction from direct exporters is that they export through intermediaries. These intermediary firms report to Chinese customs, and as a result, we only observe the destination countries to which the intermediary firms ship.

and

$$\pi_{it}^{XD} = \frac{1}{\sigma^X} r_{it}^{XD}(\Phi_t^X, w_{it}, \omega_{it}, z_{it}). \tag{5}$$

According to Koenig (2009), the number of nearby exporters or local workers employed by firms that already export directly to foreign markets serves as an indicator of the potential decrease in initial entry costs for firm i when entering a foreign market. The sunk and fixed costs arise from independent distributions, denoted as G^r. For instance, if firm i used intermediaries for exporting in the previous period and plans to export directly this year, it incurs a sunk cost $f^{IDS}(z_{it})$, drawn from the distribution G^{IDS}. Additionally, exporters need to cover a fixed cost to maintain their presence in the export market. These costs are represented as $f^{DF}(z_{it})$ for direct exporters, drawn from G^{DF}, and $f^{IF}(z_{it})$ for indirect exporters. In addition, both the fixed and sunk costs are affected by the firm-specific demand shock z_{it}. All this is summarized in Table 1. Therefore, the number of neighboring exporters who utilize a specific export mode and already participate in foreign markets serves as a measure of the potential decrease in the sunk costs for firm i when entering a foreign market, attributed to the existence of these exporters with that particular export mode.

INSERT TABLE 1 HERE

According to Koenig (2009), a firm begins to export to foreign market if the present value of the ensuring future profits is greater than the sunk cost for entry. By assuming no uncertainty about future profit, the probability that becoming a direct exporter for a previous indirect exporter can be written as

$$p_{it}^{ID} = \Pr\left[\left(\frac{\pi_{it}^{XD}}{\rho} > f^{IDS}(z_{it})\right) \;\&\; \frac{\pi_{it}^{XD} - f^{IDS}(z_{it})}{\rho} > \pi_{it}^{XI} - f^{IF}(z_{it})\right] \tag{6}$$

From Equation (2), we know, $r_{it}^{XI} = \frac{1}{\lambda^{\sigma^x}} r_{it}^{XD}$. Then written in logs according to Equation (4) and (5), Equation (6) yields

$$p_{it}^{ID} = \begin{array}{c} \Pr\left[-\ln\sigma^x + \ln r_{it}^{XD} > \ln\rho + \ln f^{IDS}(z_{it}) \;\; \& \right. \\ \left. \ln(\lambda\sigma^x - \rho) - \ln\sigma^x - \sigma^x \ln\lambda + \ln r_{it}^{XD} > \ln\left(f^{IDS}(z_{it}) - \rho f^{IF}(z_{it})\right)\right] \end{array} \tag{7}$$

The sunk cost and fixed cost can be written as

$$\ln f^{IDS}(z_{it}) = \gamma_0 + \gamma_1 \ln z_{it} + \varepsilon_{it} \tag{8}$$

$$\ln\left(f^{IDS}(z_{it}) - \rho f^{IF}(z_{it})\right) = \delta_0 + \delta_1 \ln z_{it} + \mu_{it} \tag{9}$$

where ε_{it} and μ_{it} includes the effects specific to firms, locations, countries, and years.

The probability of starting to directly export from indirect export is then:

$$p_{it}^{ID} = \begin{array}{c} \Pr\left[\begin{array}{c} -\ln\sigma^x + a^X - \ln\rho + \ln\Phi_t^X + \left(1-\sigma^X\right)\left(\beta_0 + \beta_k k_{it} + \beta_t D_t + \beta_s D_s + \beta_l D_l - \omega_{it}\right) \\ +z_{it} - \gamma_0 - \gamma_1 \ln z_{it} > \varepsilon_{it} \end{array}\right. \quad \& \\ \left.\begin{array}{c} \ln\left(\lambda\sigma^x - \rho\right) - \ln\sigma^x - \sigma^x\ln\lambda + \ln\Phi_t^X + \left(1-\sigma^X\right)\left(\beta_0 + \beta_k k_{it} + \beta_t D_t + \beta_s D_s + \beta_l D_l - \omega_{it}\right) \\ +z_{it} - \delta_0 - \delta_1 \ln z_{it} > \mu_{it} \end{array}\right] \end{array} \tag{10}$$

Under the assumptions that ε_{it} and μ_{it} are distributed logistically, Equation (10) can be estimated via a dynamic multinomial logit model.

3 Data

In this section we explain our data and how we construct the data used in our empirical analysis. To determine the presence of spatial externalities in the choice of export modes, we collected data at the firm level regarding the geographical distribution of exporters and the monetary value of their exports. We collected this data from two Chinese sources: 1) the National Bureau of Statistics and 2) the General Administration of Customs. Further descriptions of our data and merging technique can be found in the appendix. Merging the firm level production data with the transaction level trade data raises an important question relative to the sample of exporters because our firm data covers firms with sales at least 5 million Yuan (about 800,000 US dollars), and spillovers are assessed based on the number of exporting neighbors adjacent to the exporting firm. Nevertheless, the restriction of our sample is justified, as Koening (2010) notes that small and medium-sized firms are typically facing challenges when entering and expanding in international markets.

3.1 Location and Transition Matrix

In China, we observe that direct (Panel A of Figure 1) and indirect exporters (Panel B of Figure 1) are predominantly located in coastal areas such as Fujian, Guangdong, Jiangsu, Shandong, and Zhejiang.

INSERT FIGURE 1 HERE

According to Bai *et al.* (2017), indirect exporters are usually smaller manufacturing firms that engage in foreign trade through intermediaries. By doing so, they incur a lower fixed cost but higher marginal cost than they would if they were to directly export. By locating in coastal regions, they can lower their transportation costs. They do, however, face higher rental and labor costs than in the inner regions. This would increase its marginal costs.

INSERT TABLE 2 HERE

Table 2 presents the export modes and average transition of export status during the sample period among all the firms. The first column indicates a firm's export mode at year $t-1$, and the rest of three columns present the three possible export modes at year t.

In line with Roberts and Tybout (1997), our findings indicate that sunk costs play a crucial role in exporting, particularly in direct exporting. For instance, the high persistence of non-exporting firms (85%) indicates substantial sunk costs that hinder firms from beginning to export. Similarly, once a firm becomes a direct exporter, it is likely to continue in that mode the following year. In our empirical analysis, we take into account the continuity of export modes by incorporating the previous year's export mode as an independent variable in our model.

However, the situation is somewhat different with respect to indirect exporting. While there remains a notable probability of continuing as an indirect exporter next year, approximately 17.8% of firms switch to direct exporting. Additionally, there is a 12.8%

chance that a firm becomes a non-exporter. This is in line with firms self-selecting into various export modes on the basis of the level of their productivity. It is also possible that indirect exporters learn from intermediary firms or their direct exporting neighbors about foreign markets and enter directly with lower costs later. This high transition rate suggests that there is a significant incentive for indirect exporters to locate close to direct exporters.

Motivated by location of direct and indirect exporters in Figure 1 and the transition pattern in Table 2, we examine whether neighboring firms' exporting mode (agglomeration by exporting type) affects a firm's export mode.

4 Empirical Analysis

4.1 Econometrical Model Specification

Our study employs a dynamic multinomial logit model to determine the factors that affect decision to transition between the three export modes. We assume that an individual firm selects the export mode (non-export, indirect export, and direct export) that produces the highest utility in every time period. We define a random utility function for firm i with export mode j at time t as follows:

$$U_{ijt} = x_{it}\beta_j + y_{i,t-1}\gamma_j + \alpha_{ij} + \varepsilon_{ijt} \tag{11}$$

where j is the choice of three export modes (non-export, indirect export and direct export); x_{it} is a vector of observed characteristics in time t and $y_{i,t-1}$ contains the lagged export mode, which consists of two dummy variables that indicate the export mode in period $t-1$ with non-export as the base category. The vectors β_j and γ_j indicate alternative-specific coefficients; γ_j specifically is the effect of the previous export mode on the utility at time t and a measure of state dependence in export modes. Vector $\alpha_i = \{\alpha_{i1}, \alpha_{i2}, \alpha_{i3}\}$ describes firm specific unobserved heterogeneity. Error term ε_{ijt} is assumed to be independently type I extreme value distributed.

A standard multinomial logit model sometimes requires unrealistic and restrictive assumption on independence of irrelevant alternatives (IIA). According to Train (2003), this assumption can be relaxed by introducing random effects that must be integrated out. In addition, the integral includes denominators of the logit formula and thus cannot be cancelled out when the probability ratio of two alternatives is measured.

Including lagged export mode in the model causes an initial conditions problem. In this study, we can observe only export mode choices during the sample period while we do not observe firms' choices before the sample. Therefore, the random error component is more likely to be correlated with the initially observed export mode, which leads to inconsistent estimates. To address this endogeneity issue, we apply the method described in Wooldridge (2005). In short, our study considers unobserved heterogeneity α_{ij} as a function of the initial export mode y_{i0} and firm-specific explanatory variables x_i. We also assume that a random error term, α_{ij}, is uncorrlerated with the initial export mode. We assume α_{ij} to be normally distributed, with $\alpha_{ij} \sim N(0, \sigma^2)$. Hence, the probability that firm i with export mode choice j at time t conditional on both unobserved and observed characteristics and the export mode at $t-1$ can be expressed as

$$\Pr\left(Y_{i,t} = j | x_{it}, y_{i,t-1}, y_{i0}, \alpha_i\right) = \frac{\exp(x_{it}\beta_j + y_{i,t-1}\gamma_j + y_{i0}\delta_{1j} + x_i\delta_{2j} + \alpha_{ij})}{\sum_{k=1}^{3} \exp(x_{it}\beta_k + y_{i,t-1}\gamma_k + y_{i0}\delta_{1k} + x_i\delta_{2k} + \alpha_{ik})} \quad (12)$$

The individual firm likelihood function for choosing alternative j is then:

$$L = \int_{-\infty}^{\infty} \prod_{t=1}^{T} \frac{\exp(x_{it}\beta_j + y_{i,t-1}\gamma_j + y_{i0}\delta_{1j} + x_i\delta_{2j} + \alpha_{ij})}{\sum_{k=1}^{3} \exp(x_{it}\beta_k + y_{i,t-1}\gamma_k + y_{i0}\delta_{1k} + x_i\delta_{2k} + \alpha_{ik})} f(\alpha) d\alpha \quad (13)$$

Being non-exporter is the base category and the coefficients $\beta_1, \gamma_1, \delta_{11}, \delta_{21}$ and the unobserved heterogeneity term α_{i1} are set to 0 for identification. At the estimation stage, we employ the Maximum Simulated Likelihood (MSL) to integrate over the unobserved heterogeneity. Note that this approach uses simulated probabilities, and independent random draws from mixture distributions are generally utilized when simulating. Here, we apply

Halton Sequences as an alternative method, and we use 50 Halton draws per individual firm.

4.2 Empirical Results

The merged data described in Section 2 are used to create the firm's export mode dummy variables. A direct exporter dummy is used for direct exporters, and an indirect exporter dummy for indirect exporters. We use one period lag of direct exporter dummy, indirect exporter dummy, and non-exporter dummy to capture the state dependence. Then, we use the estimated total factor productivity (TFP) to examine the firm's self-selection effect on export mode. In order to detect learning to export from neighbors, we use the method described by Fernandes and Tang (2014) in which we first sum up all the lagged direct exporters for a given year by district, industry, and exporting country to determine the number of neighbors who are selling in foreign market 'm.', then construct neighbors' export growth proxied by the average exporters' sales in market 'm.' Note that the number of exporters and volume of exports in a market reveals information about the average future profitability of selling in the same market (Fernandes and Tang, 2014). Similar measures are constructed to determine the numbers of indirect exporters and non-exporters. We construct neighbors' weighted average total factor productivity by dividing the total number of direct exporting firms in a given industry and district to capture knowledge spillover effects on export mode choices.

INSERT TABLE 3 HERE

INSERT TABLE 4 HERE

Table 3 presents the number of entrants and continuing firms across years. The total number of existing firms in a particular year can be found by adding the number of entrants to the number of continuing firms. For example, the total number of existing firms in 2003 was $160,734(=40,526+120,208)$. The discrepancies with 196,217 in table A2 arise because the final data we use provide estimated total factor productivity (TFP), which dropped missing values when estimate TFP. The table clearly shows that new entrants experience serious difficulties surviving beyond a few years; about 20% of sampled new entrants exited the market within one year.

As shown in the table, the number of observations is 1,438,327. However, note that we drop lagged observations for new entrants when we construct our key variables, such as the firm's export mode dummies and Log of TFP_{t-1}. We are therefore left with 921,221 observations (Table 4). Finally, to indicate three possible export mode choices for each firm in our empirical analysis, we expand the data and obtained a sample of 2,763,663 observations.

INSERT TABLE 5 HERE

Table 5 summarizes relevant statistics associated with for the variables used in the empirical analysis. Note that Ellison, Glaeser, and Kerr (2010) found that mechanisms of labor market pooling, intellectual spillovers, and transport costs account for industry coagglomeration. Accordingly, our measure of neighbor export types is designed to capture possible agglomeration effects and learning to export from neighbors. The average number of manufacturing firms each district is 2.660 for direct exporters, 2.883 for indirect exporters, and 6.071 for non-exporters.

We find that, on average, the TFP and TFP within a district are 7.069 and 81.934, respectively. The latter variable in particular can be used to capture a possible channel of productivity spillover for export mode decisions. Notably, the export growth rate within a district, on average, is 0.416 during the sample period.

INSERT TABLE 6 HERE

It is appropriate to use a dynamic multinomial logit model with random effects to estimate the probability of switching exporter status in China. The IIA assumption is most likely violated when a firm makes dynamic decisions about their export mode. For example, the relative likelihood of choosing an indirect export mode over a non-exporting mode is based on whether the direct export opportunity is optional in a dynamic setting where learning-by-exporting is possible (See Bai *et al.*, 2017). Note that non-exporting mode is set as the base category in our empirical analysis. Each column in Table 6 shows the effect of each explanatory variable on the probability of switching between the three distinct modes (direct, indirect, and non-exporter). Model 1 shows the effects of a firm's status as an exporter in the previous year, Model 2 shows the agglomeration effects within a local district, while Model 3 shows a basic set of covariates. This study focuses on the last two sets of estimates obtained after controlling for all the relevant variables that could affect export mode choices. The coefficient estimates of a firm's export mode the previous year are statistically significant even at the 0.01 level. The results are robust across all model specifications considered in the study. For example, when a firm directly exported last year, the probability that firm's choice to adopt a direct export mode the following year is statistically high. Likewise, the probability of switching export modes is lower than the likelihood of retaining the pre-existing mode. This finding shows that individual firms can change export channels over time but do not tend to do so drastically.

A significant positive relationship is apparent between the firm's total factor productivity and its decision to export directly or indirectly through a trade intermediary. Specifically, the impact of a firm's productivity depends on the firm's choice of export mode last year. For example, the effect of total factor productivity on direct exporting versus non-exporting is -2.960 in Model 4; but, if the firm was a direct exporter in the previous year, the productivity effect is 0.055 (the difference of the coefficient estimates between the Log of a firm's TFP and its interaction term with a dummy variable). This empirical evidence supports the hypothesis that the productivity effect is much stronger on direct exporters than non-exporters. If a firm was a direct exporter in the last year, increasing its TFP has a greater impact on the decision to be a direct exporter again in the following year as it is easier for the firm to retain its exporting status, possibly due to the sunk cost mentioned in Tybout and Roberts (1997). The net effect of an indirect firm's TFP is $1.511(= 3.015 - 1.504)$, which indicates that the effect of productivity on the decision to switch export modes is larger for indirect exporters than non-exporters.

These empirical results clearly show the presence of local agglomeration effects: a firm's choice of export mode is significantly affected by the export modes of its neighboring firms. The positive signs suggest a greater preference for direct exporting over the other export modes. Firms are more likely to choose the same export mode as nearby firms, as firms benefit from local agglomeration economies by sharing information and skilled workers. A district containing a large number of indirect exporters increases the probability of more firms located there choosing to export indirectly. However, competition effects can overwhelm industrial agglomeration forces, as there more non-exporters in a local area. This finding provides empirical evidence that specific locations are more likely to attract firms with an exporter status, consistent with those of firms already located in the district.

Our model also reveals that the agglomeration impacts firm's export mode transition as shown in Table 7. If a firm was a direct exporter last time period, surrounded by direct exporting firms would increase its probability of being a direct exporter at current time with a net effect of $0.196(= 0.539 - 0.343)$, if a firm was an indirect exporter last time period, surrounded by direct exporting firms would increase its probability of being a direct exporter at current time with a net effect of $0.641(= 0.539 + 0.102)$. In addition, agglomerated indirect exporters would have a positive effect on a previously direct exporting firm still to be direct currently $(0.439 = 0.085 + 0.354)$, while have a negative effect on a previously indirect exporting firm transit to be direct currently $(-0.325 = 0.085 - 0.410)$. This tells that if a firm wanted to switch their export mode from indirect to direct, it is better for them to locate near direct exporters, this finding is consistent with Figure 1.

We now examine one mechanism of agglomeration – how learning to export works on a firm's export mode choice decision and post export mode switching, that is reflected by the average neighbors' export growth in our model. The export growth rates by direct exporters within a district positively impact a firm's choice to become a direct exporter instead of a non-exporter. This advantage may arise as a result of learning to export from neighboring direct exporter firms. Similarly, we also find that the export growth rates by indirect exporters within a district positively impact a firm's choice to become an indirect exporter instead of a non-exporter.

We also observe that a higher level in the TFP of direct exporter firms prompts a firm to sell in foreign markets directly or indirectly. The positive relationship suggests a mechanism whereby neighboring firms affect export mode choice through the potential channel of productivity spillover. We conclude that both learning to export and productivity spillover from neighbors strongly support the existence of agglomeration effects on a firm's export mode choice.

4.3 Robustness Check

When study the export spillovers, Koenig (2009) found that the local number of country-specific exporters positively influences a firm's decision to begin exporting to that country. In this paper, first, we observe that both direct and indirect exporters agglomerated (Figure 1), second, we do not observe the specific country that indirect exporter exports to. Therefore, instead, we examine if the local number of direct and indirect exporters without country specification affect the choice of exporting mode. In order to identify causal interpretation, we also need to account for any other variables that could create a similar correlation between the agglomeration of local exporters and the selection of an export mode. The local nature advantage and transport infrastructure plays an important role in determining agglomeration, while the inclusion of firm-specific firm fixed effects can account for the characteristics of the area where the firm locates.

INSERT TABLE 7 HERE

The results with firm fixed effects estimation are reported in Table 8. Because observations without variation in the outcome variable over time were dropped, there are 197,305 observations left in the end. The agglomeration of direct exporters would increase the probability of being direct and indirect exporters, as the coefficients of *Log number of neighboring direct exporters*$_{l,t-1}$ and *Log number of neighboring indirect exporters*$_{l,t-1}$ are 0.224 and 0.077, respectively, and also significant. The impact of learning-to-export could vary widely among firms, largely due to differences in unobservable firm-specific characteristics. In our empirical analysis, we account for these characteristics by incorporating firm fixed effects. When these effects are added, the learning-to-export channel is no longer statistically significant. This finding suggests that the perceived benefits of learning to export may be confounded by intrinsic characteristics specific to each firm.

INSERT TABLE 8 HERE

The export growth rate of direct and indirect exporting firms does not affect firm's choice of exporting directly, which is quite different from the findings in Fernandes and Tang (2014) that the learning effects are quite significant on new exporters' entry. They use Chinese transaction-level trade data which only reflects direct exporters' entry, and those new direct exporters could export through intermediaries as indirect exporters in previous year, for example, there are 10,428 new direct exporters in year 2005, but there are 4,783 out of them exported through intermediaries in year 2004, which accounts for about 46% of new direct exporters in year 2005. Therefore, the new direct exporters may enter by learning through intermediaries, not through agglomerated neighboring direct exporters. The positive agglomeration effect on exporting directly is actually identified by the productivity spillovers, as the TFP at a district level for a given industry is positively significant.

5 Conclusion

This study utilized a dynamic multinomial logit model with random effects to investigate the factors that influence firms' export mode choices in China. Chinese manufacturing firm-level data and transaction-level trade data from 2001 to 2007 were analyzed and strong evidence of local agglomeration effects were found. Our estimation results also showed that a firm's choice of export mode is significantly affected by the mode adopted by its neighboring firms; i.e. a firm tends to select an export mode similar to that of other nearby firms. More interestingly, we found strong evidence of a productivity effect on export mode decisions. This effect was higher for indirect exporters than non-exporters.

Our estimation results provide insights into various aspects of a firm's decisions concerning export modes. Our study about the determination of productivity on direct and non-exporting choice is consistent with the literature, while indirect and non-exporting is to the contrary, especially after China joined the WTO in 2001. In addition, we focused only on the decision to switch modes and did not inquire into the initial choice to locate a firm in a particular area. To expand on our study, one can investigate productivity and dynamic firm's choices on entry and export modes. We leave these interesting topics to future research.

Acknowledgements

Qianqian Wang thanks financial support from the Foundation of He'nan Educational Committee (2023-ZDJH-362). This work was also supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A3A2A02089585). All authors contributed equally to this work.

Disclosure Statement

No potential conflict of interest was reported by the authors.

References

[1] Ahn, J., A. K. Khandelwal, S. J. Wei. 2011. The role of intermediaries in facilitating trade. *Journal of International Economics*, 84(1): 73-85.

[2] Baltagi, B. H., E. H. Egger, and M. Kesina. 2016. Firm-level Productivity Spillovers in China' s Chemical Industry: A Spatial Hausman-Taylor Approach. *Journal of Applied Econometrics*, 31(1): 214-248.

[3] Bai, X., K. Krishna, and H. Ma. 2017. How you export matters: Export mode, learning and productivity in China. *Journal of International Economics*, 104: 122-137.

[4] Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott. 2010. Wholesalers and Retailers in US Trade. *American Economic Review Papers & Proceedings*, 100(2): 408-413.

[5] Blum, B. S., S. Claro, and I. Horstman. 2010. Facts and Figures on Intermediated Trade. *American Economic Review Papers & Proceedings*, 100(2): 419-423.

[6] Brandt, L., J. V. Biesebroeck, and Y. F. Zhang. 2014. Challenges of Working with the Chinese NBS firm-level data. *China Economic Review*, 30: 339-352.

[7] Cassey, A. and Q. Wang. 2022. Productivity, Agglomeration, and Exporting Directly: Evidence from the Chinese Textile Industry. *Journal of International Trade & Economic Development*, 31(8): 1149-1172.

[8] Duranton, G., D. Puga. 2004. Micro-foundations of Urban Agglomeration Economies. *Handbook of Regional and Urban Economics*, 4, 2063-2117.

[9] Ellison, G., E. Glaeser, and W. Kerr. 2010. What causes industry agglomeration? evidence from coagglomeration patterns. *American Economic Review*, 100(3): 1195-1213.

[10] Ertur, C. and W. Koch. 2007. Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics*, 22(6): 1033-1062.

[11] Feng, X., Y. J. Wang, and L. X. Zhao. 2022. Export Capacity Constraints and Distortions. *Journal of Development Economics*, 157.

[12] Fernades, A. P. and H. W. Tang. 2014. Learning to export from neighbors. *Journal of International Economics*, 94: 67-84.

[13] He, C., S. Jiang, and X. Hu. 2023. The effects of trade intermediaries on firms' export market diversification: Evidence from China. *Journal of International Trade & Economic Development*, 32(6): 973-989.

[14] Kamal, F. and A. Sundaram. 2016. Buyer-seller relationships in international trade: do your neighbors matter? *Journal of International Economics*, 102: 128-140.

[15] Koenig, P. 2009. Agglomeration and the export decisions of French firms. *Journal of Urban Economics,* 66 (3): 186-195.

[16] Koenig, P., F. Mayneris, S. Poncet. 2010. Local Export Spillovers in France. *European Economic Review*, 54(4): 622-641.

[17] Levinsohn, J., A. Petrin. 2003. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2): 317-341.

[18] Petropoulou, D. 2011. Information Costs, Networks and Intermediation in International Trade. Globalization and Monetary Policy Institute Working Paper 76, Federal Reserve Bank of Dallas.

[19] Roberts, M., J. R. Tybout. 1997. The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs. *American Economic Review*, 87(4): 545-564.

[20] Rosenthal, S., W. C. Strange. 2004. Evidence on the Nature and Sources of Agglomeration Economies. Journal of Regional and Urban Economics, 4, 2119-2171.

[21] Train, K.E. 2003. Discrete choice methods with simulation. Cambridge university press.

[22] Wang, Q., M. Gibson. 2018. Exporters in Cross-Section: Direct versus Intermediated Trade. *Review of International Economics*, 26(1): 84-95.

[23] Wooldridge, J. 2005. Simple Solutions to the Initial Conditions Problem for Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity. *Journal of Applied Econometrics*, 20: 39-54.

[24] Zhang, W., R. Li, and F. Zuo. 2021. Enterprise heterogeneity, agglomeration model, and urban exports: Evidence from Chinese cities and micro enterprises. *Journal of International Trade & Economic Development*, 30(8): 1207-1230.

Figure 1: Distribution of Direct and Indirect Exporters in China

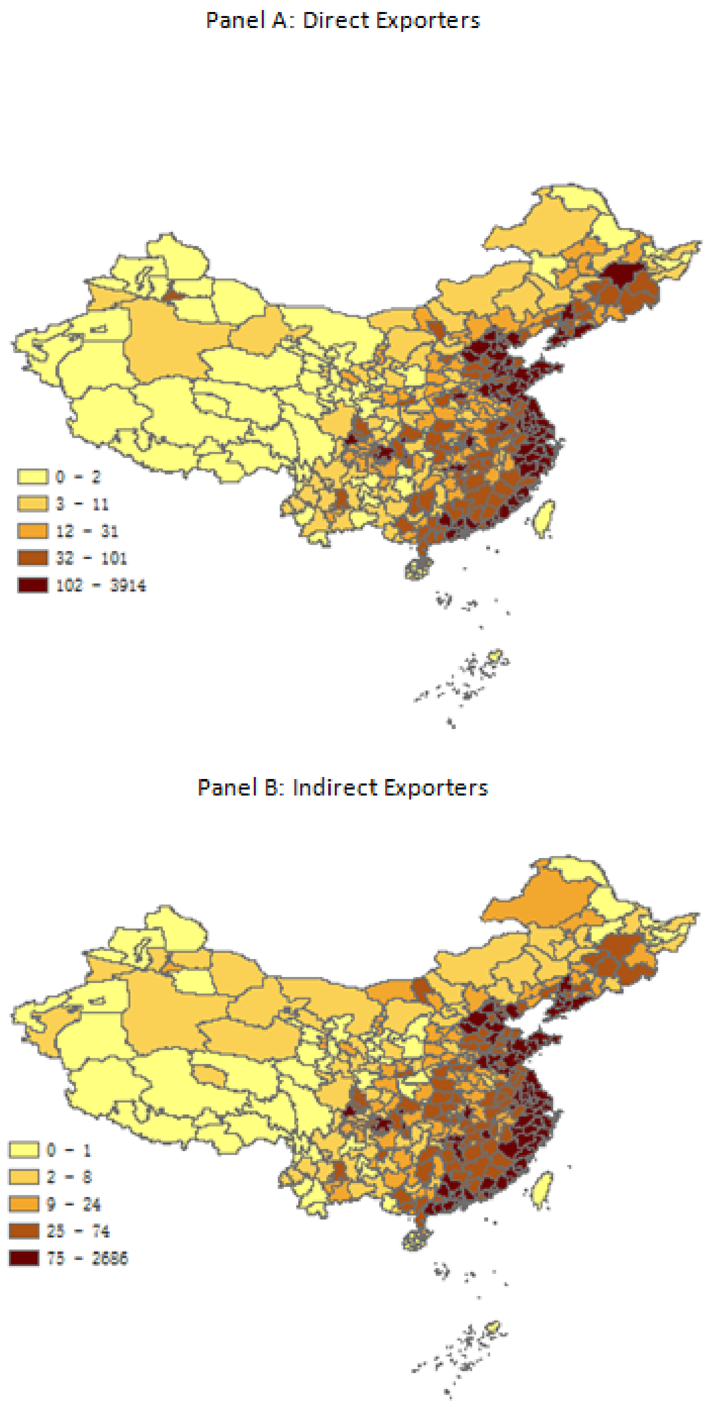

Table 1: Costs of Exporting

Export Status Time $t-1$	Time t Non-exporter	 Indirect exporter	 Direct exporter
Non-exporter	0	$f^{HIS}(z_{it})$	$f^{HDS}(z_{it})$
Indirect exporter	0	$f^{IF}(z_{it})$	$f^{IDS}(z_{it})$
Direct exporter	0	$f^{DIS}(z_{it})$	$f^{DF}(z_{it})$

Table 2: Export Modes Transition

Export status Time $(t-1)$	Time (t) Non-exporter	 Indirect exporter	 Direct exporter
Non-exporter	0.850	0.098	0.050
Indirect exporter	0.128	0.694	0.178
Direct exporter	0.022	0.039	0.938

Table 3: Firm Trend for Entrants and Continuing Firms

Entry Year	Number of Entrants	Number of Continuing Firms							
		Year							
		2001	2002	2003	2004	2005	2006	2007	Total
2000	119,161	89,781	74,877	60,850	45,798	37,835	29,041	17,780	475,123
2001	49,237		41,027	32,929	24,611	20,156	15,068	8,938	191,966
2002	33,305			26,429	19,939	16,617	12,729	7,769	116,788
2003	40,526				30,004	25,895	20,893	13,816	131,134
2004	102,077					79,768	64,624	44,061	290,530
2005	39,859						33,561	24,088	97,508
2006	45,519							36,293	81,812
2007	53,466								53,466
Total	483,150	89,781	115,904	120,208	120,352	180,271	175,916	152,745	1,438,327

Note: This table includes the number of entrants and continuing firms across years.

Table 4: Firm Trend for Continuing Firms

Entry Year	Number of Continuing Firms							
	Year							
	2001	2002	2003	2004	2005	2006	2007	Total
2000	84,481	71,617	58,694	43,625	36,970	28,567	17,704	341,658
2001		38,527	31,442	23,415	19,722	14,843	8,891	136,840
2002			25,070	18,950	16,233	12,525	7,739	80,517
2003				28,152	25,287	20,465	13,746	87,650
2004					74,973	63,327	43,805	182,105
2005						32,603	23,854	56,457
2006							35,994	35,994
Total	84,481	110,144	115,206	114,142	173,185	172,330	151,733	921,221

Note: This table includes continuing firms across years by dropping lagged observations for new entrants.

Table 5: Summary Statistics

Variable	Obs.	Mean	Std.Dev.
Direct exporter dummy$_{t-1}$	2,763,663	0.190	0.392
Indirect exporter dummy$_{t-1}$	2,763,663	0.153	0.360
Number of neighboring direct exporters$_{t-1}$	2,763,663	2.660	8.919
Number of neighboring indirect exporters$_{t-1}$	2,763,663	2.883	11.344
Number of neighboring non-exporters$_{t-1}$	2,763,663	6.071	17.396
Total factor productivity (TFP)	2,763,663	7.069	1.264
TFP within a district$_{t-1}$	2,763,663	81.934	220.376
Export growth rate within a district	2,763,663	0.416	1.914

Table 6: Estimation of Agglomeration and Channels of Agglomeration on Export Mode

	Model 1		Model 2		Model 3		Model 4		Model 5	
	DE	IDE	DE	IDE	DE	IDE	DE	IDE	DE	IDE
Direct exporter$_{i,l,t-1}$	6.495***	3.854***	6.200***	3.599***	6.496***	3.852***	12.602***	3.588***	12.265***	3.329***
	(0.014)	(0.015)	(0.014)	(0.015)	(0.014)	(0.015)	(0.207)	(0.216)	(0.205)	(0.216)
Indirect exporter$_{i,l,t-1}$	3.379***	4.560***	3.120***	4.245***	3.378***	4.555***	6.415***	5.284***	6.284***	5.117***
	(0.013)	(0.010)	(0.013)	(0.010)	(0.013)	(0.010)	(0.194)	(0.141)	(0.193)	(0.142)
Log of a firm's TFP$_{i,t-1}$	1.258***	0.746***	1.477***	0.967***	1.245***	0.745***	3.015***	0.735***	3.119***	0.916***
	(0.034)	(0.029)	(0.034)	(0.030)	(0.034)	(0.029)	(0.062)	(0.046)	(0.062)	(0.046)
Log of a firm's TFP$_{i,t-1}\times$							-2.960***	0.070	-2.880***	0.123
Direct exporter$_{i,l,t-1}$							(0.097)	(0.102)	(0.097)	(0.102)
Log of a firm's TFP$_{i,t-1}\times$							-1.504***	-0.397***	-1.517***	-0.425***
Indirect exporter$_{i,l,t-1}$							(0.092)	(0.068)	(0.091)	(0.068)
Log number of			0.658***	0.391***					0.426***	0.150***
neighboring direct exporters$_{l,t-1}$			(0.009)	(0.008)					(0.013)	(0.012)
Log number of			0.040***	0.301***					0.079***	0.345***
neighboring indirect exporters$_{l,t-1}$			(0.009)	(0.008)					(0.009)	(0.008)
Log number of			-0.361***	-0.325***					-0.389***	-0.334***
neighboring non-exporters$_{l,t-1}$			(0.006)	(0.005)					(0.006)	(0.006)
Export growth rate of					0.056***	0.043***			0.041***	-0.030***
direct exporting firms$_{l,t-1}$					(0.005)	(0.005)			(0.005)	(0.005)
Export growth rate of					-0.039***	-0.060***			0.006	0.056***
indirect exporting firms$_{l,t-1}$					(0.007)	(0.007)			(0.007)	(0.007)
Log of TFP for							0.304***	0.273***	0.190***	0.193***
direct exporting firms$_{l,t-1}$							(0.004)	(0.004)	(0.007)	(0.007)
Constant	-6.277***	-4.741***	-6.658***	-5.162***	-6.267***	-4.742***	-10.179***	-4.922***	-10.150***	-5.109***
	(0.070)	(0.060)	(0.071)	(0.062)	(0.070)	(0.060)	(0.132)	(0.095)	(0.131)	(0.095)
Number of obs.	2,763,663	2,763,663	2,763,663	2,763,663	2,763,663	2,763,663	2,763,663	2,763,663	2,763,663	2,763,663
Log likelihood	-370364.75		-363241.2		-370,222.97		-365,572.98		-361,705.14	

Notes: DE and IDE indicate direct exporter$_{i,l,t}$ and indirect exporter$_{i,l,t}$, respectively. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 7: Estimation of Agglomeration on Export Mode Transitions

	Model 1		Model 2	
	DE	IDE	DE	IDE
Direct exporter$_{i,l,t-1}$	6.495***	3.854***	12.249***	3.351***
	(0.014)	(0.015)	(0.207)	(0.217)
Indirect exporter$_{i,l,t-1}$	3.379***	4.560***	6.760***	5.529***
	(0.013)	(0.010)	(0.195)	(0.144)
Log of a firm's TFP$_{i,t-1}$	1.258***	0.746***	3.151***	0.952***
	(0.034)	(0.029)	(0.062)	(0.047)
Log of a firm's TFP$_{i,t-1}\times$			-2.801***	0.193***
Direct exporter$_{i,l,t-1}$			(0.097)	(0.102)
Log of a firm's TFP$_{i,t-1}\times$			-1.679***	-0.526***
Indirect exporter$_{i,l,t-1}$			(0.092)	(0.069)
Log number of			0.539***	-0.113***
neighboring direct exporters$_{l,t-1}$			(0.019)	(0.016)
Log number of			0.085***	0.731***
neighboring indirect exporters$_{l,t-1}$			(0.018)	(0.013)
Log number of			-0.418***	-0.328***
neighboring non-exporters$_{l,t-1}$			(0.010)	(0.008)
Log number of neighboring direct			-0.343***	0.170***
exporters$_{l,t-1}\times$ Direct exporter$_{i,l,t-1}$			(0.025)	(0.026)
Log number of neighboring direct			0.102***	0.397***
exporters$_{l,t-1}\times$ indirect exporter$_{i,l,t-1}$			(0.023)	(0.018)
Log number of neighboring indirect			0.354***	-0.153***
exporters$_{l,t-1}\times$ Direct exporter$_{i,l,t-1}$			(0.027)	(0.027)
Log number of neighboring indirect			-0.410***	-0.668***
exporters$_{l,t-1}\times$ indirect exporter$_{i,l,t-1}$			(0.022)	(0.016)
Log number of neighboring			-0.130***	-0.181***
non-exporters$_{l,t-1}\times$ Direct exporter$_{i,l,t-1}$			(0.017)	(0.018)
Log number of neighboring			0.123***	0.026**
non-exporters$_{l,t-1}\times$ indirect exporter$_{i,l,t-1}$			(0.016)	(0.012)
Export growth rate of direct exporting firms$_{l,t-1}$			0.040***	-0.035***
			(0.005)	(0.005)
Export growth rate of indirect exporting firms$_{l,t-1}$			0.011	0.065***
			(0.007)	(0.007)
Log of TFP for direct exporting firms$_{l,t-1}$			0.185***	0.185***
			(0.007)	(0.007)
Number of obs.	2,763,663	2,763,663	2,763,663	2,763,663
Log likelihood	-370,364.75		-360,220	

Table 8: Estimation of Agglomeration on Export Mode

	Model 1		Model 2		Model 3		Model 4		Model 5	
	DE	IDE	DE	IDE	DE	IDE	DE	IDE	DE	IDE
Direct exporter$_{i,l,t-1}$	0,831***	0.313***	0.809***	0.322***	0.830***	0.313***	3.216***	-2.443***	3.169***	-2.438***
	(0.020)	(0.028)	(0.020)	(0.028)	(0.020)	(0.028)	(0.310)	(0.359)	(0.310)	(0.359)
Indirect exporter$_{i,l,t-1}$	0.126***	-0.056***	0.123***	-0.040***	0.126***	-0.056***	4.655***	1.074***	4.681***	1.033***
	(0.027)	(0.015)	(0.027)	(0.016)	(0.027)	(0.015)	(0.348)	(0.234)	(0.348)	(0.234)
Log of a firm's TFP$_{i,t-1}$	1.273***	0.348***	1.161***	0.408***	1.269***	0.356***	2.536***	0.181***	2.418***	0.230***
	(0.084)	(0.073)	(0.085)	(0.074)	(0.084)	(0.073)	(0.123)	(0.092)	(0.123)	(0.093)
Log of a firm's TFP$_{i,t-1}\times$							-1.125***	1.295***	-1.114***	1.297***
Direct exporter$_{i,l,t-1}$							(0.144)	(0.168)	(0.144)	(0.168)
Log of a firm's TFP$_{i,t-1}\times$							-2.143***	-0.534***	-2.157***	-0.507***
Indirect exporter$_{i,l,t-1}$							(0.164)	(0.111)	(0.163)	(0.111)
Log number of			0.191***	0.083***					0.224***	0.077***
neighboring direct exporters$_{l,t-1}$			(0.022)	(0.020)					(0.022)	(0.021)
Log number of			-0.051***	-0.100***					-0.062***	-0.093***
neighboring indirect exporters$_{l,t-1}$			(0.019)	(0.015)					(0.020)	(0.015)
Log number of			0.050***	-0.049***					0.027*	-0.052***
neighboring non-exporters$_{l,t-1}$			(0.015)	(0.013)					(0.016)	(0.013)
Export growth rate of					0.001	-0.030***			-0.010	-0.021***
direct exporting firms$_{l,t-1}$					(0.008)	(0.007)			(0.008)	(0.008)
Export growth rate of					0.007	0.031***			0.004	0.014
indirect exporting firms$_{l,t-1}$					(0.010)	(0.010)			(0.010)	(0.010)
Log of TFP for							0.137***	0.022***	0.163***	0.047***
direct exporting firms$_{l,t-1}$							(0.011)	(0.011)	(0.013)	(0.012)
Number of obs.	197,305	197,305	197,305	197,305	197,305	197,305	197,305	197,305	197,305	197,305
Log likelihood	-78,816.43		-78674.339		-78,801.709		-78,501.94		−78,347.303	

Note: DE and IDE indicate direct exporter$_{i,l,t}$ and indirect exporter$_{i,l,t}$, respectively. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A Appendix

A.1 Firm-level production manufacturing data

Firm-level manufacturing data are the Annual Survey of Industrial Firms (ASIF) gathered from China's National Bureau of Statistics. Firm-level manufacturing data are the Annual Survey of Industrial Firms (ASIF) gathered from China's National Bureau of Statistics. The dataset covers all State-Owned Enterprises (SOEs) and non-SOEs with sales revenue above 5 million Yuan (about 800,000 US dollars). The data report the firm's age, total employment, capital stocks, and value added from 2000-2007. The data also report the total export value of a firm. Therefore, these production data comprise direct exporters (those firms that directly sell abroad), indirect exporters (firms that report export value not in a record of transaction trade data), and non-exporters.

A.2 Product-level transaction trade data

The transaction-level trade data are retrieved from China's General Administration of Customs. The dataset covers disaggregate product-level information on firm's sales price, quantity shipped, and value at the HS8 digit level. More importantly, these data provide information on whether a firm exports the output produced by itself or by other production firms. We categorize the first as a direct exporter and the second as an intermediary firm.

A.3 Merging the manufacturing and customs data

Matching two data sets is not a trivial task because the firm ID used in the two data sets are different. We first match production data with trade data by firm name (by year). Also, using the start year, we identify new firms entering the market. Next, we match two data sets using firm name, postal code, and telephone number.

Next, we identify firms as direct exporters when the transaction trade dataset matches the manufacturing data. Firms that report export totals in the manufacturing data but not in transaction-level trade data are considered indirect exporters. Further, we narrow the list by dropping businesses that assist exporters. These are the firms whose

Table A1: Exporter and Producer Statistics

Year	Trade Data		Production Data		Merged Data			
	Exporter	Value ($10^8)	Exporter	Value ($10^8)	Direct Exporter	Value ($10^8)	Indirect Exporter	Value ($10^8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2000	62,772	2,490	40,583	1,522	19,198	924	21,385	598
2001	68,487	2,910	44,770	1,692	21,874	1,040	22,896	652
2002	78,612	3,260	49,715	2,077	24,652	1,304	25,063	773
2003	95,688	4,385	56,041	2,780	28,146	1,624	27,895	1,156
2004	120,590	5,937	81,254	4,119	40,855	2,269	40,399	1,850
2005	144,030	7,567	83,978	5,142	43,610	3,026	40,368	2,116
2006	171,205	9,685	89,647	6,342	49,777	4,024	39,870	2,318
2007	193,568	12,201	91,176	7,680	52,125	4,265	39,051	3,415

name indicates they are an intermediary in the trade transaction data. As in Ahn *et al.* (2011), we look for firms identified by names that include the Chinese characters "jinchukou", "jingmao", "kemao", "maoyi", "shangmao", "waijing", and "waimao" which means trading, export, or import. We exclude intermediaries from our study as our focus is on the export modes of exporting firms, not on firms that facilitate trade.

Another issue is that, in principle, a firm might choose to export directly to one destination while opting for indirect exporting to another. Hence, we follow Bai *et al.* (2017) to define firms that report exports larger in the production data than those recorded in the trade data to export directly also as direct exporters.

The statistics for exporters and producers are presented in Table A1. After merging, the total number of exporters (column 3) equals the combined total of direct exporters (column 5) and indirect exporters (column 7). Similarly, the sum of direct exports (column 6) and indirect exports (column 8) matches the export figures reported in the production data (column 4).

To examine our matching rate, we first identify intermeidiaries by following Ahn *et al.* (2011). Then, for instance, intermediaries account for 16 % of the total exporters in 2004. This indicates that 84% of the total number of exporters are producing ex-

Table A2: Composition of Firms

Year	Production Firm	Non-Exporter	Indirect Exporter	Direct Exporter
2000	162,885	122,302	21,385	19,198
2001	171,256	126,486	22,896	21,874
2002	181,557	131,842	25,063	24,652
2003	196,217	140,176	27,895	28,146
2004	276,474	195,220	40,399	40,855
2005	271,835	187,857	40,368	43,610
2006	301,961	212,314	39,870	49,777
2007	336,768	245,591	39,051	52,125

porters, unsurveyed, unmatched, and matched in 2004. Our matching rate is about 40.4% with respect to exporters in the trade data. For example, in 2004, 40,855 out of 101,218 (=120,590*84%) firms in the trade customs data are matched. This matched sample accounts for 50% (=2269/4561(direct export value)) of trade value. One contributing factor to the presence of numerous unmatched firms is that a significant portion of these firms are small, with sales of less than 5 million RMB. A significant number of firms remain unmatched primarily because many are small, with sales under 5 million RMB. Additionally, the discrepancy arises because the trade data encompass mining and agricultural exporters, while the production data are limited to manufacturing firms.

Further, there are numerous companies with incomplete information in the trade data such as a missing name, zip code, or telephone. Hence, we are unable to match these firms. Lastly, since the trade data are recorded from customs forms, any shipment that clears customs, including very small ones like mail, is documented. We cannot match them to a firm as well.

Table A2 presents the composition of firms. In 2004, there were 276,474 unique firms in the production data. According to Brandt *et al.* (2014), the total number of firms in the above-scale sample is 82,870 (net entry) in this census year because of the economy's rapid expansion. Most firms experienced considerable growth in sales and exceeded a specific sales threshold for several years before the National Bureau of Statistics included them in the sample. Of those, 81,254 firms had positive exports, which indicate that they are either a direct or an indirect exporter. We can match 40,855 of them with records of exports in the trade data, identifying these as the direct exporters. We identify 40,399 indirect exporters in the production data who are not present in the transaction trade data. This means that slightly less than 50% of firms are direct exporters.