Export by Cohort*

Zhang Chen[†] Qing Huang[‡]

August 1, 2022

First draft, preliminary

Abstract

Using the Chinese customs data, we find a cohort effect among new exporters in the same destination market: firms entering later tend to perform better at the same age. Better performance of later entrants indicates improvement in firm fundamentals. To identify such changes, we rely on structural models of new exporter dynamics, among which two competing theories stand out: demand learning and customer accumulation. We show that the relationship between sales lifecycle and cohort is informative about the main driver of post-entry exporter growth. A major class of demand learning models à la Jovanovic (1982) predict flatter lifecycles over cohorts, which are inconsistent with the parallel lifecycles seen in the data. On the other hand, we show analytically that customer accumulation models can generate parallel lifecycles. Guided by the qualitative analysis, we build a tractable customer base accumulation model with advertising, estimate it structurally and validate its capability to replicate the empirical cross-cohort exporter lifecycles. The model estimates suggest that the cohort effect is a combination of productivity and demand effects: exporters entering one cohort later on average gain 0.2% in measured productivity and start with a 6.7% larger customer base.

Keywords: Export Pioneers, Cohort Effect, Exporter Dynamics, Demand Learning, Customer Base Accumulation
JEL Classification: D22, D25, D83, F14, F17, F61, L25, O12

*Acknowledgement to be added.
[†]Department of Economics, Princeton University, zhangc@princeton.edu
[‡]National School of Development, Peking University, qhuang2017@nsd.pku.edu.cn

1 Introduction

There is a long-lasting conviction that export pioneers can generate positive spillovers to other producers in their home countries. It seems reasonable that export pioneers may spread helpful information about the foreign markets, diffuse new technologies, or build an international reputation. Hausmann and Rodrik (2003) then famously argued that there is a coordination problem between export pioneers and followers since export discoveries made by the former have greater social returns than private returns. However, surprisingly little empirical work has been done to systematically verify the existence of a pioneering effect and investigate how it works.[1]

In this paper, we study the export behaviors of pioneers and followers. Since the new millennium, the world has witnessed a tremendous increase in Chinese export. A marked feature, as we will show later, is the great expansion in its width: China has been exporting new products or to new destinations at an amazing speed. The sheer number of new export activities makes China ideal for our study. We use the Chinese customs data from 2000 to 2011, which covers the fastest growing period of Chinese export.

The first part of the paper generalizes the notion of pioneers. We index firms by cohort, i.e., the relative timing of their entries into a foreign market. Specifically, an exporter is said to be in the first cohort of a market–a product-destination pair–if it is one of the first domestic firms ever exporting that product to the destination country. Other cohorts are then defined inductively. Furthermore, we focus on the new markets, in which China only begins to export during the sample period. In this way, early cohorts are authentic export pioneers.

We examine export pioneering activities at both market and firm levels. First, new markets are both challenges and opportunities for exporters. Compared to the old markets, markets other than the new markets, fewer firms export in the new markets. Exporters in the new markets exit more frequently, and the total export activities vanish more quickly. However, there is, on average, faster export growth in the new markets than in old markets with similar lengths of export spells. This is consistent with Hausmann and Rodrik's description of an export discovery. Second, we find a cohort effect among exporters in the same market: later cohorts on average ship more goods and make more revenue than their predecessors at the same age. They are more likely to survive as well. The identification leverages the rich within-firm market variations in the custom data and that cohorts are market-specific. That is, we obtain the cohort effect from the cross-destination cohort variations within a firm-product-year triplet, controlling for market-specific factors and age. In addition, we show that the cohort effect cannot be explained by tariff changes, selection, re-entrants, types of trade, and other confounders. Lastly, we study the removal of quotas on the Chinese textile export in 2005. Since it was decided long before China's WTO accession in 2001, the quota removal is

[1]The best evidence we have been aware of comes from a research project conducted by the Inter-American Development Bank. Nine research teams in seven Latin American countries identify and document leading examples of new export activities in great detail, for example, blueberries and TV format in Argentina, GM soybeans and wood furniture in Brazil, fresh cut flowers in Colombia, and so on. These case studies support the above conjecture that export pioneers reveal information about the foreign markets, promote new technologies, and contribute to a national brand. They have been carefully recorded in the IADB publication Sabel et al. (2012).

an exogenous entry shock to many Chinese exporters, alleviating concerns about self-selection into cohorts. Among these firms, we still find that they perform better in markets with more previous export cohorts.

Given the cohort effect, it is not hard to understand the differences between the new and old markets, which are compositions of export cohorts. Then, it suffices to understand the cohort effect–what makes later cohorts more advantageous in doing oversea business. A direct approach is to estimate the production function and obtain cohort-specifc changes in firms' characteristics. Since cohort effect is identified from market variations within a firm, production function estimation will require input data at the firm-market level. The data problem becomes worse if we consider intangible inputs, such as advertising and marketing, which receive lots of attentions in the literature as important margins of the growth of the new firms. In light of these difficulties, we choose an indirect approach to back out changes in unobserved firms' characteristics from observed firm dynamics using structural models. Nevertheless, as we only observe exporters' lifecycles in sales, quantity, and continuation, models with multiple mechanisms run the risk of being unidentified since these mechanisms may have similar lifecycle implications. Consequently, a careful selection on the mechanisms becomes necessary.

Our model selection takes three steps. First, we review the literature to search for candidate mechanisms of exporter dynamics. A recent survey by Alessandria, Arkolakis and Ruhl (2021) on firm dynamics and trade simplifies this task tremendously. Evaluating all exporter dynamics models on their list, our shortlist consists of only two candidates: demand learning and customer base accumulation. In demand learning models, firms gradually update their beliefs on the market demand through their sales experience. Alternatively, firms gradually build their customer bases to expand their sales in customer base accumulation models. Other models are excluded because they are either supply-side mechanisms, which do not account for market variations, or isomorphic to these two. Second, we compare the qualitative properties of these two models. It is well known that firms' growth are negatively correlated with their sizes in models in which firm dynamics are driven by selection. This property is intuitive since larger firms are less selected and then display less growth conditional on survival. Learning models à la Jovanovic (1982) are selection models with age-dependent exit thresholds and possess this property. To test this property empirically, we consider a particular demand learning model based on Arkolakis, Papageorgiou and Timoshenko (2018), which is the template of nearly all demand learning models we have seen in the exporter dynamics literature. We further prove that a cohort with larger initial sales must have lower sales growth in all ages. This theoretical prediction has an exact empirical interpretation: regressing log sales on cohort and age, the product of any cohort intercept and cohort-age interaction coefficient should be negative. More explicitly, it predicts that a later cohort has a flatter lifecycle in sales. Visualizing regressions coefficients into a figure of lifecycles by cohort, it is straightforward that the lifecycles are parallel to each other with later cohorts on the top. Complementing the eyeball check, we test all possible combinations of coefficients statistically. Most of them are insignificantly different from zero, and the rest are, if anything, significantly different from being negative.

Whereas the demand learning model makes inconsistent predictions, we show analytically that parallel lifecycles across cohorts can hold in a simple model of customer base accumulation. This is not surprising if we think later cohorts face lower unit costs and have larger customer bases at entry. A cost advantage compliments the customer base and incentivizes firms to invest more in the customer base. In contrast, decreasing returns to the size of the customer base discourages investment. Two forces counteract each other exactly such that the shape of the lifecycle remains unchanged, yet the initial sales increase. In sum, our model selection test is based on the correlation between firms' growth and sizes across models. It does not rely on the details of the specific variants of demand learning or customer base accumulation models we use. Instead, it applies more generally to differentiate between growth mechanisms driven by selection or investment. The test results are robust with alternative interpretations of the cohort effect, alternative variants of demand learning or customer base accumulation models, and some general classes of selection and investment models. Our test should be a meaningful alternative to a popular approach in the literature that uses price-age relationships to differentiate between demand learning and customer base accumulation models.

The only candidate that passes the first two rounds is customer base accumulation. The third and last step is to verify whether it can quantitatively explain all observed exporter dynamics. We extend the simple model in the second step. In the model, each firm faces a residual demand which shifts outward with the expansion of the customer base. Firms do advertising to attract new customers into their customer bases. In each period, firms decide whether to stay or start exporting, given realized fixed cost shocks. If they export, they further decide on the prices and the amount of advertising. The customer base depreciates over time and is wiped out after an exit. We show that the extended model still inherits the key analytical properties.

We model the cohort effect as the joint outcome of cohort-specific productivities and initial customer bases. There are cohort-specific components in firms' productivities and initial customer bases. We estimate the model by solving the firm's decision problem and matching the implied model moments to the data. The model can precisely match firms' lifecycles in both sales and survival rates as well as initial relative sales. The structural estimates of the cohort-specific components render a transparent interpretation of the cohort effect. Later cohorts exhibit higher measured productivities and begin with larger customer bases. For Chinese exporters in the first six cohorts, entering one cohort later implies a 0.2% gain in measured productivity and a 6.7% larger initial customer base. Moreover, the increase in initial customer base explains over 90% of the gains in initial sales. This suggests the spillovers from export pioneers are mainly on the demand side: their business activities create a national reputation that raises the foreign demand for domestic productions. The cost advantage of the followers seems to play a secondary role.

The rest of the paper is organized as follows. In section 2, we summarize our data and present the cohort effect with reduced form evidence. We detail our strategy to model cohort effect in section 3. Section 4 introduces the customer base accumulation model for estimation and discusses its analytical properties. In section 5, we estimate the model structurally and interpret the cohort effect accordingly. Finally, section 6 concludes.

1.1 Related Literature

This paper is related to a strand of literature on economic development that studies the externalities among exporters. Early theoretical models by Hoff (1997) and Hausmann and Rodrik (2003) argue that market failure due to informational externality is why many developing countries fail to realize their potential comparative advantages in exports. Segura-Cayuela and Vilarrubia (2008) provides a micro-foundation for the "cost discovery" in Hausmann and Rodrik (2003) through the sequential entries of firms. More recently, a few papers have made empirical investigations on export pioneers. Freund and Pierola (2010), Artopoulos, Friel and Hallak (2013) and Iacovone and Javorcik (2010) present descriptive evidence on export pioneers and export discoveries. Wei, Wei and Xu (2021) quantify the aforementioned market failure in a model in which a sunk discovery cost is payable only to the pioneers. In contrast to our focus on the intensive margin, they consider the impact of export pioneers only on the extensive margin. Similar to us, Wagner and Zahler (2015) and Haidar (2020) also find that export followers tend to make more sales than pioneers. However, neither of them obtain this result controlling for both firm and market heterogeneities, nor do they explore the sources of this late mover advantage.[2] Therefore, they cannot conclude that the pioneering effect is more than selections over firms or markets.

Our research methodology builds on an extensive literature on the exporter dynamics. Roberts and Tybout (1997) and Das, Roberts and Tybout (2007) pioneer the use of dynamic discrete choice models to study the entry and exit of exporters. Later, Ruhl and Willis (2017) show that standard sunk cost models cannot generate the post-entry growth in exporters' sales and survival rates. It spurs a large number of following papers to study various plausible engines of the intensive margin export growth.[3] Alessandria, Arkolakis and Ruhl (2021) is a comprehensive and up-to-date survey on this topic. In addition, many papers show that the growth patterns still hold using only within-firm market variations, indicating demand-side mechanisms. Demand learning and demand accumulation models stand out to lead the discussion. Both Berman, Rebeyrol and Vicard (2019) and Fitzgerald, Haller and Yedid-Levi (2022) make serious attempts to compare these two mechanisms but conclude differently. The former favors learning because demand accumulation models have great difficulties explaining the large share of negative export growth or the declining prices over the lifecycles observed among French exporters. However, the latter argues that a

[2] The dependent variable in Wagner and Zahler (2015) is at firm-product-year level, and they do not control for firm fixed effect. The dependent variable in Haidar (2020) is at firm-product-destination-year level. He includes separate firm and product fixed effects but has no destination-specific controls.

[3] Plausible mechanisms include selection (Arkolakis, 2016), financial frictions (Kohn, Leibovici and Szkup, 2016), labor market frictions (Fajgelbaum, 2020) , capital adjustment (Rho and Rodrigue, 2016), learning (Arkolakis, Papageorgiou and Timoshenko, 2018; Berman, Rebeyrol and Vicard, 2019), customer base accumulation (Piveteau, 2021; Fitzgerald, Haller and Yedid-Levi, 2022) and experimentation (Fanelli and Hallak, 2021), among others. Chaney (2014), Morales, Sheu and Zahler (2019) and Alfaro-Ureña et al. (2022) study exporters' expansion into new destinations with complementarity across markets. There are also papers studying new exporters in other aspects, e.g., partial years (Bernard et al., 2017), spin-offs (Blum et al., 2020) and multinationals (Gumpert et al., 2020).

learning model provides a very poor fit to the Irish customs data.[4] We revisit and comment on most of their arguments throughout the paper. Besides, we contribute to this discussion by introducing new empirical moments for model selection.

Finally, this paper is related to the study of industry lifecycles, e.g. Gort and Klepper (1982), Jovanovic and MacDonald (1994), Klepper (1996) and Dinlersoz and MacDonald (2009). A consensus in this literature is that an industry's lifecycle can be divided into three stages: it begins with rapid growth, experiences shakeouts, and eventually stabilizes. We investigate the early stages of Chinese exports and compare firms' dynamics by their entry cohorts. Our analysis shows that later entrants are exposed to greater demand and indicates that the industry's early growth is likely the consequence of rising demand for the product.

2 Stylized Facts

2.1 Data

The primary data used in our empirical analysis is the *Chinese Customs Transactions Database*. It contains information on the values and quantities of all international trade transactions completed by Chinese firms between 2000 and 2011. Each transaction is recorded by the firm, country, and 8-digit product category of the Harmonized System (HS8) at a yearly frequency. For concordance reasons, the unit of analysis in our sample is at the firm-product (6-digit)-country-year level.[5] Specifically, a *market* in this paper is a pair of HS6 products and destination countries. Throughout this paper, firms are indexed by i, products by j, destination countries by d, and years by t. Consequently, markets are indexed by jd.

Cohorts are specific to markets. For a firm that simultaneously exports into two distinct markets for the first time, it might be in the second cohort of one market and the third cohort of the other, given that there are respectively one and two cohorts of firms exporting to those markets prior to its entry. Therefore, it is essential to locate the first cohort, i.e., the group of firms that export to markets to which no firms from the home country have ever exported. The left truncation of our data in 2000 is an obvious challenge to this knowledge. For instance, we will have little idea about the cohort of a new exporter i in market jd if market jd has consecutive export records spanning the entire sample period. To resolve this problem, we introduce the notion of *new market*: a market is a new market if there is no custom record between 1997 and 2001 but at least one record from 2002 to 2011. Since it acceded to the WTO in 2001, China has witnessed substantial export growth, partly driven by the rapid growth in the extensive margins. The following figure 1 shows trends in the number of Chinese exporters and the number of destination markets between 2000 and 2011. Both surges began after 2002, so our sampling period (2002-2011) will capture them well. On the

[4] Earlier versions of their paper contain more discussions on the poor fit of learning models. In Fitzgerald, Haller and Yedid-Levi (2019), they argue that the poor fit is expected since learning models with quantity choices cannot generate either the positive correlation between durations and initial quantities or the flat lifecycle in prices in the data.

[5] See appendix A.1 for the details of our concordance procedure.

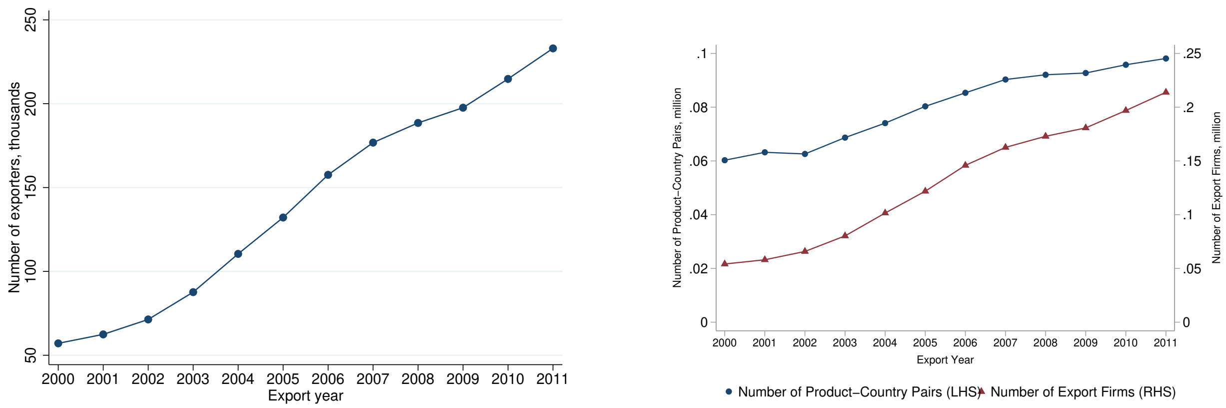
other hand, we obtain aggregate trade flow data from the GTA database of IHS Markit to cover the Chinese trade flow by destination market between 1996 and 1999. A potential threat to the selection of new markets is a type II error, namely, the probability that a selected new market has export records before the window period. In appendix A.2, we show that the window period between 1997 and 2001 is a great choice to balance the type II error with the number of new markets. Eventually, our final sample comprises the universe of export records to the new markets.

We now describe the construction of our key variables *age* and *cohort*. We define the age of firm i in market jd and year t as the number of years firm i has been consecutively exporting to market jd by year t. Firm i is of age 1 in market jd and year t if it exports to market jd in year t but not in year $t-1$. A firm-market export spell is then the episode in which a firm makes positive export to a market every consecutive year. The first and last year of the spell is determined respectively as such that there is no export to that market from the firm in the year before the first year and after the last year. A firm may have multiple export spells in the same market if it takes breaks in exporting to that market. Lastly, cohorts are defined inductively as well. Firm i would be in the n-th cohort of market jd at year t if $n-1$ cohorts of firms exported to that market by the first year of its current export spell. In particular, firm i is in the first cohort of market jd at year t if no firms had exported to market jd by year t. In the rest of this paper, we use the n-th cohort and cohort n interchangeably. As cohorts are labeled based on the export spells, a firm will be counted in a new cohort if it exits and re-enters. Table A2 in appendix A provides an additional example to illustrate this case and all the notions. The above construction closely tracks the canonical version of age, year, and cohort decomposition such that the third variable will be pinned down once the other two variables are known.[6]

2.2 Summary Statistics

Table 1 provides a simple but direct comparison between the new and old markets during 2002-2011. Several things are easy to notice. First, **new markets are marked with significantly fewer export activities.** The average number of exporters per year in the new markets is only 3.5, which is much fewer than in the old markets (27.4). The median number of exporters is 1 and 6 in new and old markets. Similarly, both average and median annual export values in the old markets are almost eight times as large as those in the new markets. Second, new markets are more volatile than the old markets. We construct three measures to unveil this point: the total active years, market spells, and market survival rates. The total active years of a market are the sum of years in which there are export records. Unsurprisingly, new markets have witnessed much more retreats of the Chinese exporters than old markets. Nevertheless, this might be mechanical since export histories in the new markets are likely to start more recently. To overcome this bias, we compare all markets with positive exports in 2002. The median active year of those new markets

[6]Note that there is a slight difference between our construction and the usual equality "year = cohort + age". As it is not guaranteed that a new cohort will enter every year, that equality does not necessarily hold literally. However, if we replace cohort by "cohort year" in table A2, the perfect linearity will hold exactly, i.e., year = cohort year + age - 1. It then validates our construction as decomposition, and the usual identification problem will still be present.



(a) Number of exporters

(b) Number of product-country pairs

Figure 1: Growth in the Extensive Margins: Firms and Markets

Notes: Figure 1 is calculated by authors. The product is at the HS 6-digit level.
Data source: Customs Data 2000-2011.

is only 6, while that of the old markets is equal to the length of the whole sample period. We define a *market spell* in the same fashion as the firm-market spell. It is the longest period when Chinese firms consecutively export to that market. Likewise, markets can have multiple spells. While the average number of spells is comparable between both markets, the average length of spells is much smaller in the new markets. In addition, we look into the difference between the old and new markets in market survival rates, i.e., the probability that Chinese firms will export next year conditional on them exporting this year. The market survival rate is a measly 67% in the new market compared to the 90% in the old market. Statistics on firm-market spells also echo this point. The median length of a market spell is 4 years for exporters in the old markets, yet it is only 1 year for new markets.

The above evidence aligns with the conventional wisdom that new markets are very uncertain. A blessing in disguise is that with greater risks comes greater opportunity. Indeed, we find that **there is stronger export growth in the new markets conditional on them being successful.** Figure 2 exhibits the comparison between the median annual growth rates of the old and new markets by market spells. Particularly, the length of a spell is informative about the market fundamental, e.g., a very long spell might result from strong demand for Chinese goods. This raises concerns on selection as previous statistics show that new and old markets differ substantially in the composition of spells. To address this, we consider market spells at least 2, 4, 6, and 8 years long separately. We restrict market spells for comparison on those which begin in 2002. The median annual growth rates in the number of firms and total export value increase slightly with the spell length in the old markets. However, they increase significantly with spell length in the new markets, suggesting the important role of export pioneers in selecting promising destination markets. It is also worth noting that the growth rate is higher in the new markets in almost all configurations of spell length and export activity measures. The only exception is that the median

Table 1: Summary Statistics: New Markets vs. Old Markets

	New Markets 2002-2011			Old Markets 2002-2011		
Statistics	N	Mean	Median	N	Mean	Median
Market level						
# markets	185,233			240,498		
total active years	185,233	3.43	3	240,498	7.32	9
total active years starting from 2002	21,822	6.00	6	148,517	8.92	10
# market spells	185,233	1.49	1	240,498	1.41	1
Market-spell level						
length of market spells	275,951	2.30	1	339,165	5.19	4
Market-year level						
# firms	635,688	3.50	1	1,761,160	27.44	6
export value	635,688	440,063	15,949	1,761,160	3,508,707	127,266
survival rate	539,047	0.67		1,573,755	0.90	

Notes: Table 1 is calculated by authors. The total active years of a market are the sum of years in which there are export records during the 2002-2011. Market spell is the longest period during which Chinese firms consecutively exports to that market. The unit of export value is CNY.
Data source: Customs Data 2002-2011.

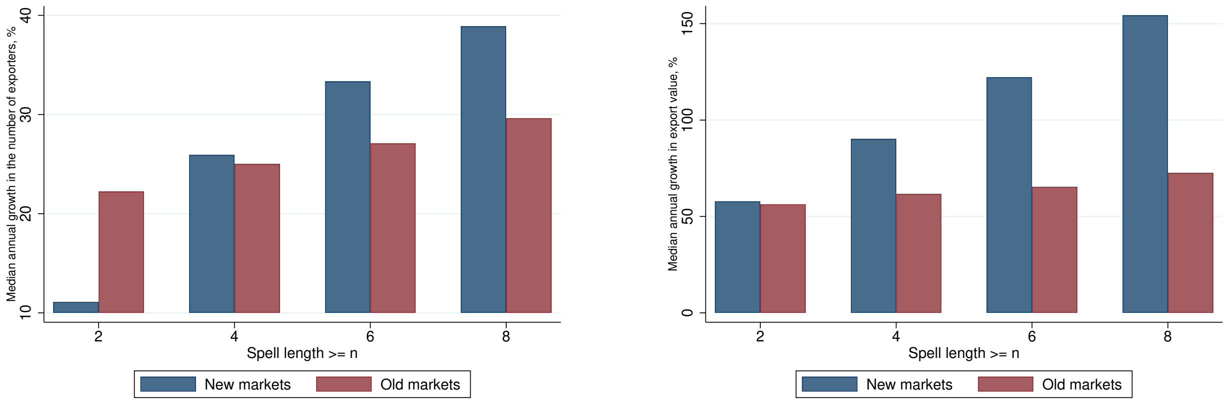
growth rate in the number of firms is higher in the old markets when all market spells with at least 2 years are included. This seems very plausible because short spells signal unsatisfactory export environments, which discourage both export volume and entries of followers. Given that the new markets are relatively mysterious to the Chinese firms, it is more likely to encounter unsuitable markets in the new markets, resulting in short spells. Then, the median growth rate in the new markets shrinks considerably by including a large number of short spells. To conclude, new markets are both challenges and opportunities. Export discoveries made by the pioneers select promising markets and lead to profound subsequent export growth.

2.3 Specification

The above section indicates remarkable differences between the old and new markets. In this section, we take one step further to explore whether and how exporters behave differently in the new and old markets. Note that the number of cohorts is a precise quantification of market age and then a basis of the division between the "old" and "new." Therefore, a more general version of the above questions is whether and how exporters of different cohorts behave differently. The conceptual equivalence is clear: early exporters must export to "new" markets because very few cohorts had exported there before them.[7] We then empirically investigate the relationship between firms' performances and their cohorts through the following specification:

$$y_{ijdt} = \beta_a' \mathbf{a}_{ijdt} + \beta_c' \mathbf{c}_{ijdt} + \mathbf{x}_{jdt} + \eta_{ijt} + \varepsilon_{ijdt}. \tag{1}$$

[7] This shall not be confused with our restrictions on the new markets, which is necessary in order to have well-defined cohorts. In the end, we are still comparing "new" and "old" markets within these markets.



(a) Number of exporters (b) Total export value

Figure 2: Median Export Growth Between Old and New Markets

Notes: Figure 2 is calculated by authors. Market spell is a period during which Chinese firms consecutively exports to that market. See table A3 in appendix A for detailed statistics.
Data source: Customs Data 2002-2011.

y_{ijdt} denotes three dependent variables of interest: export value and quantity in logarithm and survival. Survival is given by a dummy variable which equals to one if a firm that exports this year continues to export in the next year. $\mathbf{a}_{ijdt}$ is a vector of dummies which characterizes the age of firm i in market jd and year t. Similarly, $\mathbf{c}_{ijdt}$ is a vector of dummies which characterizes the cohort of firm i in market jd and year t.[8] We topcode both cohort and age at 7 years.[9] η_{ijt} is the firm-HS6 product-year fixed effect and $\mathbf{x}_{jdt}$ is a set of controls at the product-destination-year level, which include the HS4 product-destination-year fixed effect γ_{4dt}, log import value of market jd in year t and effectively applied tariff rates. We obtain the import value data from the CEPII database and tariff data from WITS-TRAINS.

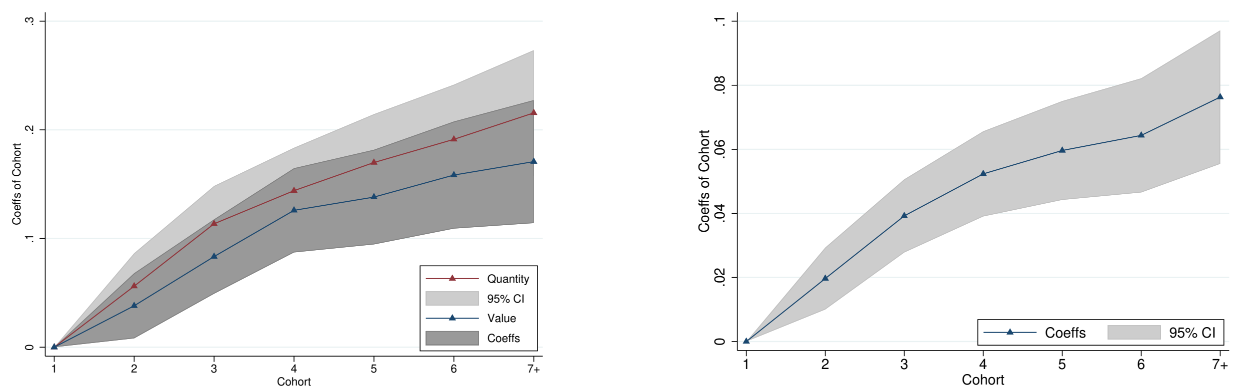
The identification of cohort coefficients β_c' comes from the cross-destination cohort variations within a firm-product-year triplet, controlling for market-specific factors and age. First, the firm-product-year fixed effect η_{ijt} isolates the variations in firms' performances from variations in the supply side, e.g., firms' product-specific marginal costs. Hence, these coefficients are driven by demand-side factors only. Second, the controls $\mathbf{x}_{jdt}$ aim to absorb market-specific factors that are common to all firms. This is important so that the results are not driven by the composition effect of market heterogeneities. In principle, we should replicate the previous fixed effect strategy to include the HS6 product-destination-year fixed effect γ_{jdt}, i.e., the year effect. However, such inclusion will cause the well-known identification problem that the age, year, and cohort variables are perfectly collinear. Note that age, year, and cohort variables are only approximations of the underlying unobservables, which are not themselves linearly dependent. In comparison, the approximation is so crude that it creates a problem of its own.[10] In our case, the year effect is only a proxy for the

[8] Cohort has a time subscript t because it depends on the current firm-market export spell.
[9] The largest cohort and age in our sample are 10. Our results are robust to non-topcoding.
[10] See Heckman and Robb (1985) for more on this point.

market-specific demand, so a simple fix is to find alternative proxies. In table A4 of the appendix, we show that our control bundle does equally well as γ_{jdt}. Namely, we estimate the post-entry sales-age profile of a representative exporter with respective controls. The resulting age coefficients are very close, and the adjusted R^2 are identical.

2.4 Results

Our baseline regression uses the HS4 product-destination-year fixed effect and the log import of the destination market as market-specific controls. We visualize the cohort effect in figure 3 by plotting the cohort coefficients β_c against the number of cohorts. Three connected lines correspond to three regressions in (1), whose dependent variables are export value, quantity, and survival rate, respectively. The coefficient of the first cohort is normalized to zero. The upward trends in all figures present the cohort effect very straightforwardly: **later cohorts tend to sell more and are more likely to survive at the same age.** In particular, the first cohort could on average sell 4% more in value and 5.6% more in quantity per annum had they entered the second cohort. The per annum survival rate is also 2% higher in the second cohort than the first cohort, holding anything else constant. All advantages become increasingly more salient in later cohorts. The 7th and later cohort could make 17% more sales and ship 21.6% more goods than their first cohort counterpart at the same age. They are also 7.6% more likely to survive an additional year.



(a) Log Value & Quantity

(b) Survival Rate

Figure 3: Cohort Effect: Baseline Estimation

Notes: Figure 3 shows the estimated cohort effect. Different lines correpond to coefficients of cohort when using different dependent variables. All the regressions control for firms' experience, log market size, firm-product (HS6)-year FEs and product (HS4)-country-year FEs. Cohort and experience are topcoding at 7. Standard errors are robust clustered. See Table 2 for detailed results.

Table 2 provides more details of the empirical results. First, changes in tariffs seem to be particularly relevant since such changes may induce export discoveries over new markets. We observe that tariff heterogeneities between 6-digit products within a 4-digit category are very limited. Adding tariffs as controls not only costs over half of the observations due to data coverage limitations but

Table 2: Cohort Effect: Baseline Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Export Value		Log Export Quantity		Survival Rate	
Cohort						
2	0.038**	0.034	0.056***	0.062**	0.020***	0.023***
	(0.015)	(0.024)	(0.015)	(0.024)	(0.005)	(0.008)
3	0.083***	0.092***	0.114***	0.136***	0.039***	0.043***
	(0.017)	(0.027)	(0.018)	(0.028)	(0.006)	(0.009)
4	0.126***	0.141***	0.144***	0.178***	0.052***	0.054***
	(0.020)	(0.031)	(0.020)	(0.032)	(0.007)	(0.011)
5	0.138***	0.146***	0.170***	0.195***	0.060***	0.060***
	(0.022)	(0.035)	(0.023)	(0.036)	(0.008)	(0.013)
6	0.158***	0.167***	0.191***	0.227***	0.064***	0.068***
	(0.025)	(0.040)	(0.026)	(0.041)	(0.009)	(0.015)
7+	0.171***	0.185***	0.216***	0.263***	0.076***	0.089***
	(0.029)	(0.047)	(0.029)	(0.047)	(0.011)	(0.018)
Age						
2	0.658***	0.698***	0.686***	0.733***	0.135***	0.140***
	(0.010)	(0.014)	(0.010)	(0.014)	(0.003)	(0.005)
3	1.025***	1.089***	1.073***	1.146***	0.218***	0.218***
	(0.015)	(0.021)	(0.016)	(0.022)	(0.005)	(0.008)
4	1.323***	1.372***	1.377***	1.443***	0.275***	0.281***
	(0.021)	(0.030)	(0.022)	(0.031)	(0.008)	(0.011)
5	1.542***	1.563***	1.606***	1.653***	0.305***	0.300***
	(0.029)	(0.040)	(0.030)	(0.041)	(0.011)	(0.015)
6	1.693***	1.734***	1.764***	1.821***	0.333***	0.329***
	(0.041)	(0.055)	(0.042)	(0.057)	(0.015)	(0.021)
7+	1.952***	1.964***	2.042***	2.076***	0.356***	0.364***
	(0.049)	(0.065)	(0.051)	(0.067)	(0.020)	(0.027)
Log Market Size	0.196***	0.186***	0.192***	0.189***	0.019***	0.017***
	(0.005)	(0.010)	(0.005)	(0.010)	(0.002)	(0.003)
Tariff rate		0.004		0.006**		-0.001
		(0.003)		(0.003)		(0.001)
Fixed Effect	Firm-product (HS6)-year, Product (HS4)-country-year					
N	486,598	226,071	484,439	224,841	359,517	164,240
adj. R2	0.672	0.667	0.870	0.876	0.388	0.354

Notes: Table 2 reports main coefficients of firm's age and cohort. The observation is at the firm-product(HS6)-country-year level. Market size is measured by total value imported by destination country. Tariff rate is measured by weigted average tariff rate implemented by destination country. Market size and tariff rate are at the product (HS6)-country-year level. Standard errors in parentheses are robust. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.
Source: Customs Data 2002-2011, market size from CEPII, and tariff rate from WITS-TRAINS.

barely impacts the results. Indeed, the tariff coefficients are insignificant and very close to zero with the presence of the HS4 product-destination-year fixed effect. This is why our baseline regression excludes tariffs as controls. Second, we still observe a strong age effect with the cohort effect. This complements a large existing literature on exporter dynamics, which documents that export volume increases with exporters' experience in the destination market. Lastly, we can compare the relative size of the age and cohort effects. It makes sense to conjecture that the age effect is more influential on firms' sales as it captures firm-specific changes directly. In contrast, the cohort effect captures only market-specific changes and only indirectly affects firms' sales. The estimation results confirm that the age effect on sales is an order of magnitude larger than the cohort effect. The average sales growth is 66% in the first year of export. In contrast, a firm in the second cohort on average makes merely 4% more sales than its first cohort counterpart at the same age. The relative size of the age and cohort effects also holds with quantities.

2.5 Robustness

We discuss in this section a set of robustness checks to increase the credibility of the main results. Relevant results are stored in the appendix B.1.

Selection Exporter dynamics at the market level are marked by high-frequency entry and exit. Firms that export continuously for many years can differ from those with occasional export records. This is known as a threat to a regression on firms' age and has received extensive attention in the literature. In addition to selection on ages, interpretation of the cohort coefficients might also be confounded by selection on cohorts. Chinese firms might continuously export to some markets for many years and stop exporting after a few cohorts. A longer export spell could then be an indicator of better market fundamentals. Consequently, firms' better sales performances in older markets, measured by the number of past cohorts, might capture merely the selection effect that older markets are better markets. We run the same regression on various subsamples to partially account for selection on age and cohort and document the results in table A5. In the first column, the subsample includes only firms with export spells of at least two years in the destination markets. This exercise is meaningful since we confirm huge attritions: only a third of our baseline sample satisfies this restriction. Verifying that the cohort effect remains strong among surviving firms is reassuring. In the second to fourth columns, the respective subsamples include only markets with export spells of at least 2/4/6 cohorts. That is, we compare firms' performance in markets with increasing similarities. The results suggest selection on markets is unlikely to be the main driver of the observed cohort effect. The cohort effects change little among different subsamples of markets.

Learning the relationship between firm size and their entry behaviors helps to understand better the direction of selection bias. Figure A1 shows that exporters with larger yearly export values or quantities tend to enter earlier cohorts. We also confirm the same relationship in Figure A2 on a subsample of exporters, in which we have information about their annual domestic sales and employment.[11] In other words, exporters with larger domestic sales or employment tend to enter earlier cohorts. Thus, our estimates of the cohort effect are likely to be a lower bound, given that larger firms enter first. This finding is consistent with the convention wisdom of selection into trade and provides direct evidence against the conjecture in Wagner and Zahler (2015).[12] They also find that export followers make more sales than pioneers, but they interpret it as indirect evidence that larger firms enter later and conclude it to be at odds with the standard Melitz model.

Re-entry Our construction of the cohort is based on current export spells. Firms that pause exporting for one year and resume afterward will be counted in the new cohort following their re-entries. This may introduce a bias that later cohorts have a larger share of re-entrants. It seems plausible that re-entrants have more knowledge about the markets and perform relatively better than the new entrants. Fanelli and Hallak (2021) find that conditional on age, the average survival rate is significantly higher in re-entrants than in new entrants among Peruvian exporters. Hence, it is important to show that our results are not driven by the composition effect of re-entrants. We introduce an additional dummy variable into the baseline regression to account for re-entry. Results in table A6 verify our conjecture that re-entrants outperform new entrants. Conditional on age and cohort, re-entrants on average export 20.2% more in sales, 21.5% more in quantity, and have a 5.5 percentage point higher probability of survival. Despite the superior performances of re-entrants, we still observe that later entrants tend to perform better at the same age.

Processing Trade Chinese export during our sample period features a substantial fraction of processing trade. Dai, Maitra and Yu (2016) find that 14% of Chinese firms accounting for 17% of export value engaged purely in processing trade from 2000 to 2006. Another 23% of the firms engaged in both processing and non-processing trade, accounting for around 60% of export value. Processing trade is heavily dependent on imported intermediate inputs, raising the concern that its business model differs from a representative exporter's. In other words, exporters of processing trade may experience faster exporter growth due to a quick demand accumulation process. Once they have proved their credibility as suppliers, they receive more and larger orders from their old buyers and are less incentivized to search for new buyers. On the other hand, there are also concerns that the share of processing trade might be correlated with cohorts. In a processing relationship, the exporter is mainly in charge of the manufacturing process, and the foreign buyer is usually responsible for the marketing and distribution of final products. Early cohorts are reasonably more likely to deal with simple tasks like assembly, whereas later cohorts can handle more sophisticated procedures, take charge of the marketing and distribution, and increase the value-added content. Therefore, we consider excluding processing trade to check if the same results still hold. We run the baseline regression on a subsample of excluding processing transactions and a subsample of

[11] We obtained this information from the Annual Survey of Industrial Firms produced by the National Bureau of Statistics of China.

[12] Empirically, it is consistent with the findings in Freund and Pierola (2010), who find pioneers are likely to be larger and more experienced exporters in Peru.

excluding processing firms.[13] As shown in table A7, the cohort effect remains strong among non-processing transactions or non-processing exporters.

Firm Ownership Chinese export firms consist of different ownership such as private firms, foreign-owned firms, and state-owned firms. State-owned firms are more likely to receive export subsidies and may not follow market forces. Foreign-owned firms may hold more information about foreign markets or are more recognized by foreign customers. Pooling these firms together in the regression might confound the interpretation of the cohort effect. So we run a robustness check including only private firms. The cohort effect remains when considering only private exporters and is slightly larger than the baseline estimation. Results can be found in table A8.

Partial Year Effect Bernard et al. (2017) show firms export in different months during the first calendar year of their export spells. With the first calendar year of firms' exports being a partial year, there is an upward bias in the age effect of the first year in both sales and quantities. Therefore, there are similar concerns about whether partial years bias the estimation of the cohort effect. We exploit a subsample of the customs data from 2000 to 2006 with monthly transaction records.[14] We examine whether the pattern of firms' market entries by month varies systematically with their cohorts. Figure A3 displays the distribution of cohorts across the months of their entries. The share of each cohort is stable across all months, suggesting that the distribution of firms' entry months varies very little across cohorts.

2.6 Case Study: MFA quota removal

The Multifiber Arrangement (MFA) and its successor, the Agreement on Textile and Clothing (ATC), imposed quotas on the exports of clothing and textile products from developing countries to the United States, European Union, Canada, and Turkey. It was not abolished until the Uruguay Round in 1995 and was then replaced by a new agreement to eliminate the quotas over four phases. 1 Jan 2005 was the start date of the final phase, in which countries were required to remove all remaining quotas representing 49 percent of their 1990 import volumes. China was not eligible for the quota removal until its accession to the WTO at the end of 2001. As 1 Jan 2002 was the start date of phase three, quotas faced by Chinese exporters between 2002 and 2004 were only those to be removed in the final phase. Khandelwal, Schott and Wei (2013) provide additional institutional background of the event and analyzes the impact of quota removal on Chinese export with an emphasis on allocative efficiency.[15]

[13] The Chinese customs data provide information on whether a transaction is a processing one or not. In particular, it records five categories of processing trade: compensation trade, processing and assembly trade with supplied materials, processing trade with imported materials, processing and assembling imported equipment, and export goods of foreign contracted projects. We define the processing transactions as the firm-product(HS8)-country-year level records involving processing trade. The processing firms are exporters that purely engage in processing transactions.

[14] As far as we are concerned, this is the longest monthly data series from the Chinese customs that are available for public access.

[15] Another useful reference on the same event is Roberts et al. (2018), in which they structurally estimate the impact of MFA quota removal on the Chinese footwear industry. Defever, Heid and Larch (2015) also study the same

Quota removal is an entry shock to Chinese firms. Unsurprisingly, the annual growth in the number of exporting firms in quota-bound regions from 2004 to 2005 is 96 percent, much higher than the 24 percent annual growth from 2000 to 2004. It is also orthogonal to the 2005 demand and supply factors since the decision of quota removal was made a decade ago and was not specific to China. Then, it seems reasonable to consider the entries of new exporters to quota-bound regions in 2005 as exogenous since they are driven by quota removal. More specifically, we consider the following firms to be compilers of the quota removal: firms that would have exported to quota-bound regions before 2005 if they had export licenses. The underlying assumption is that these firms had been well prepared for exporting into quota-bound regions, and their entries were delayed only because of the quota. Hence, their entry decisions in 2005 are unlikely to be affected by the number of export cohorts in the prospective markets, reducing concerns about self-selections into cohorts.[16]

We then replicate the baseline regression (1) on the compilers in 2005. We follow Khandelwal, Schott and Wei (2013) to focus on the US, EU, and Canada as destination countries. Empirically, a firm-product pair is a compiler if it satisfies two conditions: 1) this firm had exported the product before 2005 but only to quota-free regions, and 2) exported the product to quota-bound regions in 2005. Furthermore, we treat cohort as a continuous variable to combat the significant loss of observations. Table 3 contains the results of this case study. In line with our baseline results, firms tend to sell more in value and quantity in markets with more previous cohorts. The correlation between cohort and survival rate is positive, albeit insignificant.

3 Model Cohort Effect

Evidence from the last section indicates the advantages of late entrants in doing business overseas. A natural follow-up question is what makes them more competitive. Nevertheless, this is not an easy one. On the one hand, many mechanisms can potentially account for this late mover advantage. For example, by analyzing the sales records of their predecessors, late entrants may develop more appropriate sales strategies to better cater to the foreign population. It might also be that the successes of export pioneers establish a good reputation as "made in China", which makes foreign customers more easily accept new import varieties from China. Alternatively, a growing export transaction may facilitate the supply of non-production ancillary services, e.g., translation, which will decrease the total cost embodied in each shipment.

On the other hand, data availability limits the instruments to investigate this question. First, we do not observe firms' expenditure by domestic and export use or by export destination markets.[17]

event but with a different emphasis on the path dependence on the export destinations.

[16] Khandelwal, Schott and Wei (2013) show that quota licenses are awarded based on political connections rather than business performances. Firms with quota licenses before the removal are mostly state-owned or tightly connected to the government. In contrast, entrants following the quota removal are mostly privately owned and are more productive.

[17] Standard census datasets on firms usually do not provide this information. In China, the firm census and the annual survey of industrial firms do not provide this information.

Table 3: Cohort Effect: MFA Quota Removal

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Export Value		Log Export Quantity		Survival Rate	
Cohort	0.895**	1.564***	0.787*	1.595***	0.070	0.192
	(0.384)	(0.376)	(0.420)	(0.381)	(0.071)	(0.148)
Log Market Size	2.562***	1.505***	2.845***	2.467***	-0.255**	-0.056
	(0.609)	(0.245)	(0.570)	(0.297)	(0.110)	(0.144)
Tariffs	No	Yes	No	Yes	No	Yes
Fixed Effects	Firm-product(HS6), Product(HS4)-country					
N	93	50	93	50	93	50

Notes: Table 3 reports main coefficients of firm's cohort. The observation is at the firm-product (HS6)-country level. The sample includes new entrants of quota-removal regions in 2005. We select the sample to firms which have ever exported these products to any of other markets before 2005, and to new markets which have no export record for at least four years starting from 2000. Data on quota-removal regions comes from Khandelwal, Schott and Wei (2013). Market size is measured by total value imported by destination country. Tariff rate is measured by weighted average tariff rate implemented by destination country. Market size and tariff rate are at the product-market-year level. Standard errors in parentheses are clustered at the firm level. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.
Source: Customs Data 2005, market size from CEPII, and tariff rate from WITS-TRAINS.

It is then infeasible to test the relationship between cohorts and factors in reduced form, e.g., a direct test on whether the unit cost decreases with the cohort. A production function estimation is also off the table for the same reason. Second, many candidate mechanisms involve intangible factors, say, beliefs and customer capital, which are only measurable through the lens of models. The complexity of the model is, however, capped by the number of useful data moments. As we only observe exporters' lifecycles in sales, quantity, and continuation, multiple mechanisms with similar lifecycle implications cannot be identified simultaneously. We need a parsimonious yet sufficiently rich model to capture key economic forces.

We develop the following strategy to approach this question and overcome these difficulties. The first step is to abstract key elements from candidate mechanisms for evaluation. In section 3.1, we review the existing literature on new exporter dynamics. Among various attempts to explain exporters' lifecycle, we conclude that two mechanisms–demand learning and customer base accumulation–are most relevant for our purpose. The second step is a qualitative analysis of the fitness of the mechanisms. The importance of a mechanism is seriously questionable if it makes qualitatively different predictions from the data. We then leave out mechanisms that fail the qualitative test. In section 3.2, we present two models of demand learning and customer base accumulation. We show in section 3.3 that the demand learning model predicts inconsistently with the data on the sales lifecycles across cohorts. However, this is not a problem for the customer base accumulation model, as explained in section 3.4. We conclude this step in section 3.5 by discussing the generality of this qualitative approach and comparing this model selection method to the price dynamics approach in section 3.6. The third and last step is to confirm that surviving mechanisms provide good quantitative fits. The following two sections show that an extended customer base accumulation model perfectly matches the cross-cohort exporter lifecycles. Therefore, interpreting the model estimates answers what makes late entrants more competitive.

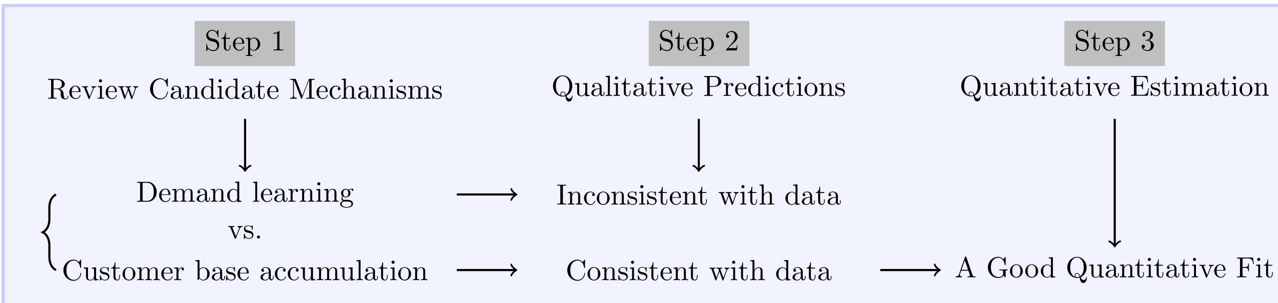


Figure 4: A Flow Diagram on Model Selections

3.1 How new exporters grow

In that cohort effects are time-invariant (up to export spells), different cohorts are isomorphic to firms that start their current export spells with different initial conditions. Then, understanding the constituents of the cohort effect is equivalent to understanding what initial conditions are there and how they have changed. This is deemed to be challenging since initial conditions are unobserved. Nevertheless, if the growth mechanism behind firms' lifecycles is known, variations in lifecycles across different cohorts can identify the changes in their initial conditions. It then suffices to learn the growth mechanisms of exporters, namely, the new exporter dynamics.

The literature on the new exporter dynamics is relatively recent. Alessandria, Arkolakis and Ruhl (2021) summarize in table A.2 of their appendix eight mechanisms that have sought to explain the post-entry growth in both sales and survival rates of new exporters. Our search for potential growth mechanisms begins from there. First, the identified cohort and age effects in our specification are purged from firm-product-specific supply-side factors. This rules out mechanisms that affect firm characteristics common to all markets. Based on this observation, we focus on and discuss mechanisms that have natural firm-market-level interpretations: (1) demand a function of time in market, (2) customer accumulation, (3) learning, and (4) search. The first mechanism. Demand a function of time in market, is studied in Ruhl and Willis (2017) as a fix to standard sunk cost models. The authors point out that standard sunk cost models cannot generate intensive margin growth in new exporters, but the problem can be fixed by adding an exogenous age-dependent demand shifter. Therefore, it is best viewed as an accounting device rather than an economic mechanism. The latter three are follow-up attempts to rationalize the age effect with micro-foundations. Eaton et al. (2021) study the search mechanism in conjunction with learning. They aim to explain the new exporter dynamics through the lens of firm-to-firm trade. In their model, exporters search for potential foreign buyers, form business relationships, and sell their products. Meanwhile, they gradually learn the profitability of their products in the foreign markets and decide on search intensity accordingly. Hence, they study both (2) customer accumulation and (3) learning based on buyer-seller relationship formation.[18] We choose not to follow their approach for two reasons: 1) we do not have information at firm-to-firm transaction level,[19] and 2) we intend to focus on basic mechanisms that are mutually exclusive. Our shortlist consists of only (2) customer accumulation and (3) learning.

Capital adjustment dynamics Financial friction Selection (on productivity) Importing and exporting	} × Supply side, don't vary across markets
Demand a function of time in market Search	} × Isomorphic to customer accumulation
Customer accumulation **Learning**	} ✓ Candidate mechanisms

Figure 5: An Illustration of Mechanisms Selection

The word "learning" is polysemous in economics. In this paper, it refers particularly to learning about uncertainty. Quite a few papers have studied exporters' learning behaviors in an uncertain environment. Most of them follow the tradition of Jovanovic (1982) and generate firms' lifecycles through their Bayesian learning on unknown demand components.[20] [21] In these papers, uncertainty about foreign demand for their products makes new exporters start small. Firms then update their beliefs on foreign demand based on realized transactions and quit exporting if realized demand is weaker than expected. Therefore, selection implies positive growth in sales conditional on continuation. Older exporters are more likely to continue exporting since new information is unlikely to alter their entrenched beliefs. Despite other margins, all papers cited above adopt the same structural learning model as in Arkolakis, Papageorgiou and Timoshenko (2018). Firms hold normal priors over the permanent component of their demand shocks. In each period, they choose quantities before the realization of demand shocks and infer current demand shocks from realized prices. Firms then update their beliefs with new demand shocks and decide whether to exit. We formally replicate this demand learning model in the next section and derive testable implications.

[18] The other paper cited by Alessandria, Arkolakis and Ruhl (2021) under the "search" category is Lu, Mariscal and Mejıa (2016). However, they study import switching at the firm level, which is a supply-side mechanism.

[19] Also, that level of granularity is not necessary at our targeted level of aggregation. We are interested in the dynamics of a firm's total sales but much less in the dynamics of different margins of its sales.

[20] As far as we have concerned, this list goes as follows: Timoshenko (2015), Cebreros (2016), Arkolakis, Papageorgiou and Timoshenko (2018), Berman, Rebeyrol and Vicard (2019), Li (2018), Bastos, Dias and Timoshenko (2018), and Chen et al. (2020).

[21] Trade papers with other learning modes usually concentrate on stylized facts different from the age effects on sales and survival rates. Albornoz et al. (2012) and Nguyen (2012) focus on experimentation across markets to explain low first-year survival rate. Fanelli and Hallak (2021) model learning as a jump process on productivity to explain the better performance of re-entrants. Monarch and Schmidt-Eisenlohr (2016) study buyers' Bayesian learning on the types of their suppliers to explain trade volume growth within a buyer-seller relationship. Learning in all these models are still learning about uncertainty and are passive.

The prevalence of this learning model in the literature is not an attestation that this is the only or correct variant of learning models. We leave the discussion on other variants of learning models and the generality of the mechanism into section 3.5.

In recent years demand side factors have received increasingly more attention in firm dynamics. As argued in Foster, Haltiwanger and Syverson (2016), a good motivation for customer accumulation is that differences in marginal cost explain little variations in sizes between new and established businesses. The idea is that, like physical capital, it takes time to accumulate customers, which shifts firms' residual demand curves to the right. Models of customer accumulation root deeply in traditional factor accumulation models, e.g., the neoclassical growth model. The customer base often depreciates and is decreasing return to scale. Firms then make "investments" to increase their customer base and reach steady states. However, the particular investment channel differs among variants of customer base accumulation models. One important variant features non-pricing schemes, mostly advertising and marketing.(Arkolakis, 2010; Fitzgerald, Haller and Yedid-Levi, 2022; Drozd and Nosal, 2012) In these models, firms "purchase" new customers through advertising instead of price discrimination. Correspondingly, another variant concerns pricing schemes, in which firms' customer base grows with past sales. (Piveteau, 2021; Fitzgerald, Haller and Yedid-Levi, 2022; Gourio and Rudanko, 2014; Rodrigue and Tan, 2019) Hence, young firms would give price discounts to increase their future customer base. Price dynamics is often used to test the implications of these two variants. In that we do not focus on price, we choose a minimal customer base accumulation model with advertising as a representative and argue that it is a good qualitative fit. We leave the discussion on other variants of customer base accumulation models, the generality of such mechanism, and the testability of price dynamics into section 3.5.

3.2 A demand learning model of new exporter dynamics

We present two firm's decision problems featuring demand learning and customer accumulation respectively. For a clean comparison, both models share the same economic environment except for necessary departures. The demand learning model borrows heavily from Arkolakis, Papageorgiou and Timoshenko (2018) and represents a standard treatment in the literature. We postpone the presentation of a customer base accumulation model to section 3.4. Now consider a single-product firm i which sells a differentiated good to a foreign market. The residual demand function facing it in a particular foreign market m at time t is

$$Q_{imt} = Y_{mt} D_{imt}^{\alpha} P_{imt}^{-\sigma} Z_{imt}, \quad \sigma > 1 \tag{2}$$

where Y_{mt} is the market demand that are common to all firms, Q_{imt} the quantity of firm's product, and P_{imt} its price. σ is then the demand elasticity. D_{imt} and Z_{imt} compose idiosyncratic demand. D_{imt} captures customer accumulation and is deterministic. It is usually interpreted as a type of intangible capital such as customer base. The parameter α then measures the return of customer base and is in $[0, 1]$. In contrast, Z_{imt} captures the randomness in idiosyncratic demand. Input choices are static and production function is constant return to scale. Hence, it is without loss of generality to assume that the production function takes the form $Q = \varphi L$, in which φ is the firm-specific productivity and L is labor. Given labor wage w_t and the residual demand function, firms maximize their profits. The resulting profit function is standard:

$$\pi_{imt} = \frac{1}{\sigma} R_{imt} = \frac{1}{\sigma} \Big(\frac{\sigma}{\sigma - 1} \frac{w_t}{\varphi_{it}}\Big)^{1-\sigma} Y_{imt} D_{imt}^{\alpha} Z_{imt} = A_{imt} D_{imt}^{\alpha} Z_{imt}, \tag{3}$$

where R_{imt} is revenue and $A_{imt} \equiv \frac{1}{\sigma}\big(\frac{\sigma}{\sigma-1}\frac{w_t}{\varphi_{it}}\big)^{1-\sigma} Y_{mt}$. Since w_t and Y_{mt} are common to all exporters in market m, A_{imt} are their measured productivities in that market. In line with our regression specifications, we assume constant marginal cost and market size. That we focus on a firm's decision problem allows us to suppress subscripts i, m and t, and we take this advantage whenever applicable. Two models diverge henceforth, and we describe first the learning model.

Demand Learning We set $\alpha = 0$ to shut down customer accumulation, so the residual demand function is $Q = Y P^{-\sigma} Z$. We use lowercase letter to denote the logarithm of the corresponding variable. Age is denoted by n and indexed by 0, 1, 2, $\cdots$, in which age 0 marks the entry of an exporter into the market. Log idiosyncratic demand z_n comprises a permanent component θ and a temporary component ϵ_n, namely,

$$z_n = \theta + \epsilon_n,$$

Each firm draws the permanent component θ at entry. θ is constant within the export spell and follows a normal distribution with mean zero and variance σ_θ^2. The temporary component ε_n, however, is drawn each period. It is an i.i.d. normal variable with mean zero and variance σ_ϵ^2.[22] We use τ to denote the precision of a distribution. Then, $\tau_\epsilon = 1/\sigma_\epsilon^2$, and $\tau_\theta = 1/\sigma_\theta^2$. The distribution of each component is known to all firms, whereas the realization is not.

The timing of each firm's decisions is as follows. At the beginning of each period, firms decide on quantities before the realization of idiosyncratic demand shock z_n. With the presence of fixed cost, they also decide on whether to stay exporting. Then, all idiosyncratic demand shocks z_n realize. Firms use the inverse demand function to infer z_n from the realized prices, which clear the good markets. At the end of each period, firms update their beliefs over θ with the new signal z_n following the Bayes' rule. The diagram in figure 6 illustrates this process graphically.

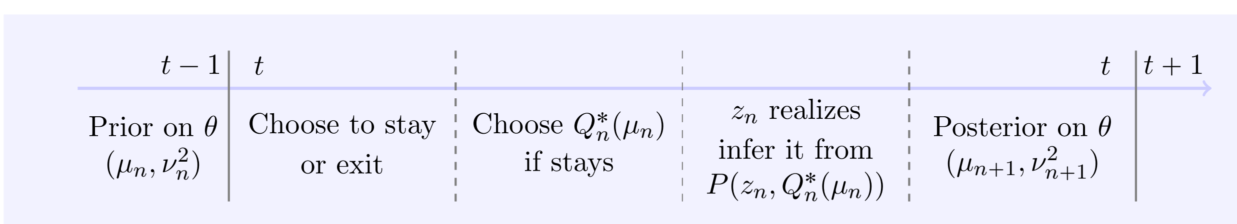


Figure 6: The Sequence of Actions

[22]The zero-mean assumption is innocuous since a change in the mean is identical to a change in the measured productivity A.

At age 0, or entry, firms have the same prior over θ, which is given by $\mathcal{N}(0, \sigma_\theta^2)$. The normality assumption implies that firms' posteriors at all ages are normal as well. Let μ_n and ν_n be the respective mean and variance of a firm's prior distribution over θ at the beginning of age n. For $n \geqslant 1$, standard results in Bayesian learning apply to characterize μ_n and ν_n^2 in a recursive form:

$$\mu_n = \mu_{n-1} + \frac{z_{n-1} - \mu_{n-1}}{g_n}, \qquad \nu_n^2 = \frac{\nu_{n-1}^2 \sigma_\epsilon^2}{\nu_{n-1}^2 + \sigma_\epsilon^2}, \tag{4}$$

in which $g_n = \tau_\theta/\tau_\epsilon + n$. Accordingly, firm's prior on the idiosyncratic demand z_n at age n is given by $\mathcal{N}(\mu_n, \nu_n^2 + \sigma_\epsilon^2)$. Conditional on its productivity φ, prior mean μ_n and age n, each firm chooses quantity to maximize the expected static profit.

$$\begin{aligned} \max_Q \quad & \pi_n(\mu_n) = \mathbb{E}[P(z,Q)Q|n, \mu_n] - \frac{w}{\varphi}Q - F \\ \text{s.t. } & P(z,Q) = Y^{\frac{1}{\sigma}} Q^{-\frac{1}{\sigma}} \exp\left(\frac{z}{\sigma}\right), \end{aligned} \tag{5}$$

in which $P(z, Q)$ is the inverse demand function and F the fixed cost. Solving (5), the optimal quantity is

$$Q_n^*(\mu_n) = Y\left(\frac{\sigma}{\sigma-1}\frac{w}{\varphi}\right)^{-\sigma} b_n(\mu_n)^\sigma, \quad \text{in which } b_n(\mu_n) = \exp\left(\frac{\mu_n}{\sigma} + \frac{1}{2}\left(\frac{\nu_n^2 + \sigma_\epsilon^2}{\sigma^2}\right)\right). \tag{6}$$

Given the optimal quantity choice, we could rewrite firm's expected static profit

$$\pi_n(\mu_n) = Ab_n(\mu_n)^\sigma - F, \tag{7}$$

where A is defined as in (3). As before, A is the measured productivity and a model primitive. It is usually suppressed unless we do comparative statics on it. Lastly, the following Bellman equation characterizes firms' dynamic decisions on whether to exit with value zero:

$$V_n(\mu_n) = \max\{Ab_n(\mu_n)^\sigma - F + \beta\,\mathbb{E}[V_{n+1}(\mu_{n+1})|n, \mu_n], 0\}. \tag{8}$$

The following lemma shows that firm's exit decision follows a threshold rule. All proofs in this paper are contained in the appendix.

Lemma 1. *In this demand learning model, V_n exists and increases on prior mean μ_n. Then, there exits a threshold μ_n^* such that a firm of age n continues exporting if and only if $\mu_n \geqslant \mu_n^*$. Moreover, μ_n^* decreases on the measured productivity A and increases on τ_θ.*

3.3 A test on the learning model

To closely mimic our empirical specification, we assume away market fluctuations, which are "year effects". That is, all firms in this model, regardless of their cohorts, face demand shocks with the same distributional structures, i.e., $(\sigma_\theta^2, \sigma_\epsilon^2)$ is constant. It allows us to recast the lifecycles of different cohorts in the model as firm's lifecycles under different model primitives. Several notations are very useful to aide a concise display of our comparative statics results. Let $q_n(\mu_n)$ be the log quantity conditional on prior mean μ_n and age n. Normalizing it by the quantity sold at entry, $\tilde{q}_n$ denotes the relative mean log quantity to initial log quantity at age n, i.e., $\tilde{q}_n = \mathbb{E}_0[q_n(\mu_n)] - q_0$. Then,

$$\tilde{q}_n = \mathbb{E}_0[\mu_n] + \frac{1}{2\sigma}\left(\nu_n^2 - \sigma_\theta^2\right),$$

in which we use $\mu_0 = 0$ and $\nu_0^2 = \sigma_\theta^2$. We then present the following proposition.

Proposition 1. *In this learning model, $\tilde{q}_n$ decreases strictly on A.*

The intuition behind this proposition is straightforward: that more productive firms can survive with lower demand reduces the mean of the permanent demand component. However, this is not trivial when selection takes place gradually. Since truncation does not preserve the first order stochastic dominance, fiercer selection in multiple periods does not guarantee a higher mean $E_0[\mu_n]$. We establish this proposition utilizing the Markov property of Bayesian learning with a normal prior. This proposition extends the local comparative statics results on firm growth in Arkolakis et al. (2018) to the whole lifecycle.[23] In contrast, it does not condition on size (μ_n) yet takes into account selection by exit (μ_n^*). The best part of this proposition is that it allows an exact mapping to realistic data, which are hugely attrited. A non-parametric test on the learning model then becomes possible. The last step to such a test is to find a proxy for A, which happens to be the average initial sales. Note that conditional on μ_n, the mean log sales is given by

$$r_n(\mu_n) \equiv \mathbb{E}[\log P(z, Q_n^*(\mu_n))Q_n^*(\mu_n)|\mu_n, n] = \log(\sigma A) + \mu_n + \frac{\sigma-1}{2\sigma^2}\left(\nu_n^2 + \sigma_\epsilon^2\right).$$

So $r_0(\mu_0)$ is a monotonic transformation of A. We can define analogously $\tilde{r}_n$ as the relative mean log sales to initial mean log sales at age n. Then,

$$\tilde{r}_n = \mathbb{E}_0[\mu_n] + \frac{\sigma-1}{2\sigma^2}\left(\nu_n^2 - \sigma_\theta^2\right).$$

A direct application of proposition 1 gives us the following testable corollary.

Corollary 1. *If a cohort has a larger initial sales, then it must have lower growth in sales at all ages, i.e., $\tilde{r}_n$ decreases strictly on r_0.*

To test it, we enrich our baseline empirical specification (1) with interaction terms between age and cohort. This enables a complete account of firms' lifecycles in relative log sales by cohort. Specifically,

$$y_{ijdt}^r = \boldsymbol{\beta}_a'\mathbf{a}_{ijdt} + \boldsymbol{\beta}_c'\mathbf{c}_{ijdt} + \boldsymbol{\beta}_i'\mathbf{c}_{ijdt} \otimes \mathbf{a}_{ijdt} + \mathbf{x}_{jdt} + \eta_{ijt} + \varepsilon_{ijdt}, \tag{9}$$

in which y^r denotes the log sales and $\otimes$ the Kronecker product. Let $\mathbb{E}[y^r|c, a]$ denote the mean log sales at cohort c and age a. $\mathbb{E}[y^r|1, 1]$ is set to be the base level and absorbed in the constant term.

[23] Proposition 2 in Arkolakis et al. (2018) conditions on μ_n and concerns only growth between consecutive periods.

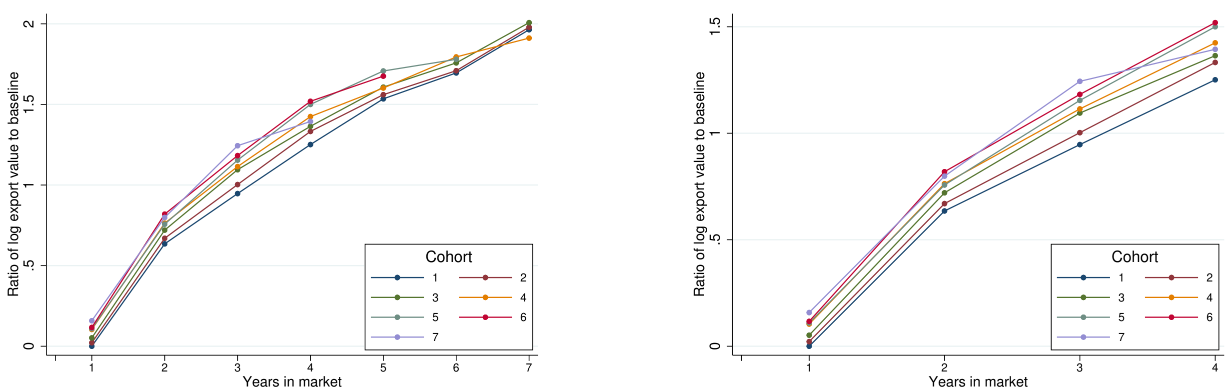
Then, the age and cohort intercepts are $\beta_a = \mathbb{E}[y^r|1,a] - \mathbb{E}[y^r|1,1]$ and $\beta_c = \mathbb{E}[y^r|c,1] - \mathbb{E}[y^r|1,1]$ respectively. Note that for $c \geqslant 2$ and $a \geqslant 2$,

$$\begin{aligned} & \mathbb{E}[y^r|c,a] = \mathbb{E}[y^r|1,1] + \beta_a + \beta_c + \beta_i^{c,a} \\ \Leftrightarrow\, & \beta_i^{c,a} = (\mathbb{E}[y^r|c,a] - \mathbb{E}[y^r|c,1]) - (\mathbb{E}[y^r|1,a] - \mathbb{E}[y^r|1,1])\,. \end{aligned}$$

We extend the previous notations to include cohort as an argument. Therefore, $\beta_c = r_0(c) - r_0(1)$ and $\beta_i^{c,a} = \tilde{r}_a(c) - \tilde{r}_a(1)$. The corollary then makes a precise prediction on the coefficients in (9).

Claim. *For all c and a, $\beta_c \beta_i^{c,a} < 0$.*

We visualize the regression coefficients in figure 7 which plots firms' lifecycles in log sales. It is clear that later cohorts have parallel trajectories above early cohorts. The parallel lifecycles are confirmed by the insignificance in almost all coefficients $\beta_i^{c,a}$. If anything, nearly all of them are positive. Consistent with our baseline results, the cohort effect materializes itself in the graph as the elevation in the initial sales, which is captured by significantly positive β_c. The details of all coefficient estimates are stored in the regression table A9. In sum, this learning model fails to match the data qualitatively.



(a) All Years

(b) The First Four Years

Figure 7: Firms' Lifecycles in Log Sales

Notes: Figure 7 shows the evolution of log export value at the firm-product-market level with market experience, allowing trajectories to differ by cohort. The regression controls for log market size, firm-product (HS6)-year FEs and product (HS4)-market-year FEs. Cohort and experience are topcoding at 7. Standard errors are robust clustered.

We attempt several alternative checks to consolidate the test results. First, we conduct sign tests on $\beta_c \beta_i^{c,a}$ for all combinations of cohort and age and document the results in table A10. Unsurprisingly, we cannot reject the null hypothesis that $\beta_c \beta_i^{c,a} = 0$ for most cases, validating the invariance of lifecycles across cohorts. Next, we replace the dependent variable to log quantity and re-run our regression (9). That initial sales increases on the cohort in the sales regression indicates improvements in the measured productivity A over cohorts. Hence, we could test directly on proposition 1 using quantities. As before, we find insignificant and positive coefficients $\beta_i^{c,a}$ for the interactions of cohort and age and significantly positive cohort coefficients β_c. Table A9 in the appendix stores those results. Furthermore, we consider direct shifters of A and test proposition 1. Since A comprises market demand and firm-specific marginal costs, we consider both market demand shifters and firms' marginal cost shifters. In particular, we use the total import value by the market and the total number of cohorts in the market as proxies for market demand. Firms' total export value to all markets and the length of export spells are proxies of firms' marginal costs. We examine whether the age profile of exporters varies in shifter A by incorporating the interaction term of the shifter and firms' age in the regression. Table A11 contains the details of our treatment. The interactions between age and various shifters are all significantly positive. The results again contradict the prediction that firms with higher measured productivities grow more slowly.

3.4 A simple customer base accumulation base model

We follow the same economic environment as described in section 3.2. We set $Z_t \equiv 1$ to shut down learning and other exogenous random shocks so that we can focus entirely on customer base accumulation. The residual demand function is then $Q = YP^{-\sigma}D^{\alpha}$. Correspondingly, the profit function is $\pi(A, D) = AD^{\alpha}$. We restrict α to $(0, 1)$ to assume decreasing return to scale. Importantly, F is set be zero to eliminate endogenous exit. This extreme treatment gets rid of any selection forces so that the growth process is driven completely by factor accumulation. Firms do advertising to expand their customer base D, which depreciates at a constant rate δ. Let I denote the size of new customers, then the next period customer base $D' = I + (1-\delta)D$. That the number of new customers cannot be negative imposes an irreversibility constraint $D' \geqslant (1-\delta)D$. The total advertising cost to acquire I new customers is given by $c(D, D')$. We assume for simplicity that $c(D, D')$ is a standard combination of a linear advertising cost and a quadratic adjustment cost, i.e.,

$$c(D, D') = D' - (1-\delta)D + \phi\frac{(D' - (1-\delta)D)^2}{D}. \tag{10}$$

Firm's investment problem is then given by the following Bellman equation

$$V(A, D) = \max_{D' \geqslant (1-\delta)D} \pi(A, D) - c(D, D') + \beta V(A, D').$$

$D^*(A, D)$ denotes the associated policy. It is straightforward to show a steady state customer base exists, and we denote it by D_{ss}.[24] Let $\tilde{D}_{ss}$ be the steady state capital level when $A = 1$, then we can rewrite $D_{ss} = \tilde{D}_{ss}A^{1/(1-\alpha)}$. Let $d = D/D_{ss}$ and $v(A, d) = V(A, dD_{ss})/D_{ss}$. The Bellman

[24] The steady state customer base

$$D_{ss} = \Big(\frac{\beta\alpha A}{(1-\beta(1-\delta))(1+2\phi\delta) - \beta\phi\delta^2}\Big)^{\frac{1}{1-\alpha}}. \tag{11}$$

equation of normalized value function has the following form:

$$v(A,d) = \max_{d'} \tilde{D}_{ss}^{\alpha-1} d^{\alpha} - (d' - (1-\delta)d) - \phi \frac{(d' - (1-\delta)d)^2}{d} + \beta v(A, d'). \tag{12}$$

The value function v is then independent of productivity A, i.e., $v(A,d) = v(d)$. The normalized customer base d becomes a sufficient statistics of firms' lifecycles in quantity and sales. We will have as well the associated policy function $d^*(A,d) = d^*(d)$. Let d_0 denote the initial normalized customer base, and d_n the customer base at age n, namely, $d_n = d^*(d_{n-1})$. Since sales and quantity have the same growth, it is without loss to consider sales growth only. We also construct g_n, the ratio between sales at age n and initial sales, as follows,

$$g_n(d_0) = \frac{R(A, D_n)}{R(A, D_0)} = \left(\frac{d_n}{d_0}\right)^{\alpha}.$$

The following proposition is an analogous corollary 1 in the customer accumulation model.

Proposition 2. *In this customer accumulation model, $g_n(d_0)$ decreases strictly on d_0.*

Hence, it is immediate that variation in initial sales is barely informative on changes in firms' lifecycles. A larger initial sales can accompany a steeper, flatter, or even identical lifecycle as long as the normalized initial customer base is respectively smaller, larger or identical. This is possible since d_0 is proportional to the ratio between D and $A^{1/(1-\alpha)}$, both of which yet increase the sales. All it needs to match the parallel empirical lifecycles seen in figure X is to have both D and $A^{1/(1-\alpha)}$ increase at the same rates over cohort. It also makes sense to conjecture that customers accrued by export pioneers will also be interested in the products of the followers. At least qualitatively, customer base accumulation models demonstrate great potential.

3.5 Discussion

Despite that both models may seem stylized, we argue that these results have profound generality. We discuss alternative interpretations of the cohort effect and other modeling choices. More importantly, our comparison method shall not confine itself to demand learning and customer base accumulation models, which belong to respective families of selection and investment mechanisms. Based on the general properties of these two mechanisms, it extends to compare selection and investment growth in general.

3.5.1 More on learning models

First of all, we point out that this demand learning model is essentially a selection model. Recall that

$$\tilde{r}_n = \underbrace{\mathbb{E}_0[\mu_n] - \mu_0}_{\text{Selection}} + \underbrace{\frac{\sigma-1}{2\sigma^2}\left(\nu_n^2 - \nu_0^2\right)}_{\text{Variance}}.$$

The first term captures the effect of cumulative selection by firms' exits over age and is positive. The second term is a Jensen term, which comes from the log normality assumption and is negative. We show in proposition A1 of the appendix that the selection term increases over age while the variance term decreases over age. Hence, selection is the real engine of firms' growth. Learning about uncertainty merely prolongs the selection process as firms take time to learn their demand. In contrast, the variance term is a by-product of modeling convenience without real economic significance.[25] Besides, our previous comparative statics results are independent of the variance term. With that regard, we shall focus on the selection term. We discuss in below several deviations from our earlier treatment.

Permanent demand shock In section 3.3, we make the assumption that demand shocks are independent of cohort. This is because demand shocks are time-varying, yet our focus is on the time-invariant cohort effect. A plausible relaxation could be that the variance of permanent demand component decreases on cohort. One might understand it as a form of consumer learning: user experience with products of the pioneers gives consumers more ideas on similar products from the followers. Note that $\mu_{n+1} - \mu_n \sim \mathcal{N}(0, s_{n+1}^2)$ with $s_{n+1}^2 = \frac{\nu_n^2 + \sigma_\epsilon^2}{g_{n+1}^2} = \frac{1}{\tau_\epsilon g_n g_{n+1}}$. The selection growth conditional on current sales and age has a closed form:

$$\mathbb{E}[\mu_{n+1} - \mu_n | \mu_{n+1} \geqslant \mu_{n+1}^*, \mu_n] = s_{n+1} \lambda \left(\frac{\mu_{n+1}^* - \mu_n}{s_{n+1}} \right),$$

in which λ is the hazard function of a standard normal distribution. We show in the appendix (lemma A1) that this growth rate increases on μ_{n+1}^* and s_{n+1}. Thus, an increase in τ_θ decreases s_{n+1} and then reduces selection growth. An increase in the measured productivity is still necessary to have more initial sales. It decreases cutoff μ_{n+1}^* and reduces selection growth. A caveat is that by lemma 1, an increase in τ_θ also increases μ_{n+1}^*, which counteracts some downward pressure.[26]

The intuition is even more general. Consider the extreme case in which the permanent demand shock is observable. Then, there is no growth, and the model degenerates into the Melitz model. In a nutshell, a reduction in uncertainty dwarfs the role of further selection and leaves less space for selection growth.

[25] Despite the popularity of this model, we do not notice any empirical support on the negative unconditional growth in log quantity. In our data, we still observe positive log quantity growth in highly selected samples, e.g., firms with very longer spell or very large markets. On the other hand, it makes some sense if we consider instead the quantity growth. The unconditional quantity growth is given by

$$\mathbb{E}\left[\frac{Q_{n+1}^*(\mu_{n+1})}{Q_n^*(\mu_n)} | \mu_n, n \right] = \exp\left(\frac{\sigma - 1}{2\sigma} \frac{\tau_\epsilon}{(\tau_\theta + n\tau_\epsilon)(\tau_\theta + (n+1)\tau_\epsilon)} \right), \tag{13}$$

which is positive. Nevertheless, there is no good support or need on the opposite signs of these two growth rates.

[26] An increase in τ_θ has opposite impact on the unconditional growth in log quantity and quantity. For log quantity, firms adjust less downward with higher precision, so the variance term increases. The quantity growth, however, decreases on τ_θ as shown in (13). This is another reason for excluding a discussion on the variance term that it is less desirable to have the results depend on specific distributional assumptions.

Fixed cost One might suggest that fixed cost could also be cohort specific, especially under long-term contract. Given parallel lifecycles in the data, one approach to fit them with the demand learning model would be to keep the ratio between fixed cost and measured productivity constant over cohort. Consequently, the lifecycle of one cohort is exactly a scaled version of another. However, the drawback of this approach is also apparent. With scaling in both measured productivity and fixed cost, the exit decision will be unchanged. Hence, we would expect lifecycles on firms' survival to be invariant on cohort, which obviously contradicts the cohort effect on firms' survival. This thought experiment unveils a major feature of selection growth: higher growth is built at the cost of fiercer selections. With that regard, selection cannot be a dominant force to generate both higher growth and more survival.

Price setting In section 3.2, firms choose quantity before the realization of demand shock. An alternative modeling choice would be to choose price. Then, firm will always charge the same markup, $(\sigma - 1)/\sigma$, as in the standard case. The relative log sales to initials log sales is equal to the relative log quantity to initial log quantity, which is given by $z_n - z_0$. The relative mean log sales to initial mean log sales at age n, $\tilde{r}_n$, satisfies

$$\tilde{r}_n = \mathbb{E}_0[z_n] - \mathbb{E}_0[z_0] = \mathbb{E}_0[\mu_n].$$

The second equality uses the fact that $\mathbb{E}_0[z_n] = \mathbb{E}_0[\mathbb{E}[z_n|\mu_n]] = \mathbb{E}_0[\mu_n]$, and $\mathbb{E}_0[z_0] = 0$. Hence, corollary 1 applies directly, and our test results are robust to the price setting scenario. Also, it worths noting that with price setting, there is no variance term in the growth rate. It echoes with our earlier point that demand learning is a selection mechanism in essence.

Active learning In our model, it takes one period to obtain one new signal for all firms. It seems also plausible that firms may do market research by themselves or pay for consulting services. That is to say, firms may pay to acquire additional signals. Adapting it into our setting, the precision updating will satisfy $\tau_{n+1} = \tau_n + (1 + k)\tau_\epsilon$, in which k is the amount of signals obtained by active learning. We extend the model with active learning in appendix D.2 and show that the optimal choice $k^* = 0$. Higher precision lowers the present value since the profit function π_n decreases on the precision and is convex on the prior mean, Then, the marginal benefit of precision is negative, and firms do not pay for any additional signals. All our results extend to this case instantly.

In general, it is much harder to comment on how active learning would alter the relationship between sales growth and sales. Whether high type firms are more incentivized to acquire information is highly model dependent. Nevertheless, the intuition that learning is a prolonged selection process goes through regardless of the exact flow of signals. Since true θ reveals in the limit, firms stay permanently if the expected static profit, $A \exp\left(\theta + \frac{\sigma_\epsilon^2}{2\sigma}\right) - F$, is positive. The threshold θ^* then decreases on A. The mean of θ is expected to decrease on A with an increasing share of low θ. Using our earlier language, that $\tilde{r}_\infty$ decreases on r_0 roughly holds. It is reassuring that we do not find data support favoring this pattern. In other words, we do not observe either $\beta_i^{c,a}$ are more likely to be negative in older ages, or $\beta_i^{c,a}$ decreases on age.

3.5.2 Other customer base accumulation models

We present in appendix D.3 a variant of customer base accumulation model with strategic pricing. Two models differ in the mechanics of customer base accumulation. In section 3.4, firms accumulate customer base by advertising, so $D' = (1-\delta)D + I$. Contrastingly, customer base is build upon past sales in the other model, i.e., $D' = (1-\delta)D + PQ$. That makes pricing decision a dynamic one since firms have to trade off between current and future profits. Despite differences in pricing decisions, we show that firms' lifecycles in customer base are the same in both models. This can be expected since what matters is the complimentarity between the measured productivity and customer base. An increase in measured productivity increases the return of customer base and boosts investment, which materializes in either more advertising or more discount. The decreasing return to scale in customer base implies that a larger customer base discourages investment. The relative magnitude of these two forces, the normalized customer base in our case, then determines whether sales growth increases on sales. As long as both forces are present, the details of the model do not restrict the relationship between firms' growth and sales.

3.5.3 General mechanisms

In the demand learning model, growth is driven by selection. Whereas in the customer base accumulation model, growth is the consequence of investment. Let type be productivity, demand, or other persistent firm heterogeneities. Given that cohort effect is a shock on types, what differentiates these two models is the response of sales growth to a type shock. Intuitively, a high type firm usually sees a higher return in investment while is less influenced by selection. It will grow faster if growth is mostly driven by investment or slower if growth is mostly due to selection.

We formalize this insight in appendix D.4 where we present a general model of firm dynamics. Under fairly standard assumptions, we show selection growth decreases on high types while investment growth increase on them. Thus, the correlation between sales growth and firms' types is a useful identification test between general classes of selection and investment mechanisms. It is particularly convenient since most demand or cost shifters qualify as shocks on firm types. The wide availability of proxies for type shocks also improves the quality of test results. That is, one can validate the test results using various type shocks. Some of our robustness checks at the end of section 3.3 are good examples of this practice. All in all, this test merits more attention for its generality and stability.

In the end, it is important to highlight the scope of this test on the underlying selection and investment mechanisms. For selection mechanisms, the variable subject to selection has to be unidimensional. Otherwise more selections do not always increase growth. For investment mechanisms, firm's type has to be complimentary to capital, the object of investment, so that high type firms are more incentivized to invest. Proposition A4 and A5 provide the relevant technical details.

3.6 On the price dynamics approach

In the existing literature, it is a popular approach to distinguish between demand learning and customer base accumulation models by price dynamics. Namely, models have different predictions on the trend of prices on age. In the demand learning model of section 3.2, the inverse demand function is fixed given the realizations of initial permanent demand shocks. A positive trend in quantities over age implies a negative price trend. In the customer base accumulation model of section 3.4, there are no dynamics in prices given pricing with constant markup. Alternatively, a positive price trend is expected in a customer base accumulation model with strategic pricing, as shown in D.2. Thus, the empirical pattern of the price dynamics becomes crucial. Berman, Rebeyrol and Vicard (2019) (BRV in short) find a negative price trend and then favor the demand learning model. Contrastingly, Fitzgerald, Haller and Yedid-Levi (2022) (FHY in short) document no price dynamics and recommend the customer base accumulation model with advertising.

We discuss issues related to the price dynamics approach. First, there are huge controversies on the empirical relationship between prices and ages. In stark contrast to the well-known increasing age profiles on sales or quantities, the literature on firm dynamics has seen all sorts of age profiles on prices. Table A12 summarizes papers which document firms' age profiles on prices. All of these papers run similar regressions on ages using log prices. The diversity of these results may come from variations in countries, periods, types of firms, or fixed effects. Narrowing our search to the price dynamics at the exporter-market level does not help reduce the discrepancies: BRV, Piveteau (2021), and FHY present respectively negative trend, no trend, and positive trend in prices.[27]

Next, the scope of price dynamics as a valid test is somewhat limited. It is demanding on the details of the learning or accumulation models and particularly on the details of pricing. As we discussed before, many models may be isomorphic regarding age profiles in sales or quantity but differ in pricing mechanisms. In other words, both demand learning and customer base accumulation have model variants that can generate distinct price trends. For demand learning, there will be no price trend if firms choose to set prices rather than quantities. The trend will be positive if firms, in addition, face oligopolistic competitions rather than monopolistic competitions.[28] As for customer base models, a negative price trend could be modeled as firms' investment in reducing the marginal cost. For example, a reduction in search cost for potential buyers could be due to the network effect of a growing customer base, as suggested in Chaney (2014). This possibility is entailed in the model presented in section 3.4 as part of the investment return. Therefore, one has to take strong stances on candidate models to invoke price dynamics for model differentiation.

Lastly, the stability of price dynamics as a selection test is questionable. This can be seen partly from the vast array of results in the previous table. Furthermore, we conduct the following exercise to show price trends' sensitivity to regression specifications or sample restrictions. Using the same 2002-2011 Chinese customs data, Zhao (2018) concludes a positive price trend. She regresses log price on age controlling for firm-product-country fixed effects and two aggregate variables at the

[27] See Sec. 5.4.4 of FHY for more discussions on the comparison among the three.

[28] See Appendix F.2 of BRV for a variant of learning model with price setting and oligopolistic competition.

product-destination-year level: average log price and total export quantity. When we implement her empirical specification, we find the same increasing relationship between prices and exporters' tenure. However, once the controls are changed to two fixed effects, firm-product-year fixed effects and product-destination-year fixed effects, we obtain a negative price trend in both the sample with only new markets and the full sample. Contrastingly, export value or quantity growth remains positive under different specifications. While it is hard to argue which specification is unanimously better, these discrepancies reveal difficulties in measuring price dynamics. The results can be found in table A13 and figure A4 and A5 of the appendix. On the other hand, Rodrigue and Tan (2019) also find a positive price trend using the Chinese customs data. They study only privately owned exporters who engage in ordinary trade between 2000-2006 and adopt the same empirical specification as FHY. We replicate their analysis and present the results in table A14 of the appendix. Again, we could reproduce the findings on export value and quantity, but not on price.

In sum, we suggest caution in using price dynamics for model selections because of its limited scope and instability. Given the prevalence of selection and investment mechanisms in economics, developing more robust and accurate selection tests is encouraged for future research.

4 A customer base accumulation model with exit

We extend the customer base accumulation model in section 3.4 with endogenous exit. Ruhl and Willis (2017) argue that a stochastic fixed cost is essential to generating the upward-sloping age profile in survival rates among new exporters. They show that a baseline model with productivity shocks and constant fixed costs will generate a counterfactual declining trend in the survival rate. This is intuitive since entrants are selected to be those with "lucky" draws of productivity shock, which will later revert to the mean. In light of their arguments, we assume a random fixed cost with a continuous distribution and exclude any randomness that affects the revenue. Further, it is without loss to assume away sunk cost since we focus on the intensive margin. We describe in the following the firm decision problem and show that the extended version still inherits the analytical properties of the simple model.

As before, each firm is characterized by its measured productivity A and customer base D. The decision process is almost identical to the simple model, except that firms now have to decide whether to quit exporting. Each firm observes the realization of the fixed cost F at the beginning of each period and decides whether to export the next period. If it chooses not to export next period, it will become a new entrant. Otherwise, it will be an existing exporter. The only difference between a new entrant and an existing exporter is the customer base D facing them next period. While new entrants start with common customer bases D_0, existing exporters continue with customer bases D accumulated from their previous export activities. Given that D is included in the state variables, we do not additionally distinguish between new entrants and existing exporters. Conditional on export participation, firms make sales and do advertising as they do in the simple model. Similarly, they respect an irreversibility constraint: they cannot sell their customers for profit. The following Bellman equation summarizes the whole decision process:

$$V(A,D,F) = \max\left\{\max_{D'\geqslant(1-\delta)D} \pi(A,D)-c(D,D')-F+\beta\,\mathbb{E}[V(A,D',F')],\beta\,\mathbb{E}[V(A,D_0,F')]\right\}, \quad (14)$$

in which F is drawn independently each period from an identical distribution G. Let G denote the CDF of this distribution. We impose the following regularity coditions.

Assumption 1. *The profit function, cost function, and fixed cost distribution satisfy the following conditions:*

(i) $\pi \in \mathcal{C}^2, \pi_1 > 0, \pi_2 > 0, \pi_{22} < 0, \pi_{12} > 0$*, and* π_2 *satisfies the Inada condition;*

(ii) $c \in \mathcal{C}^2$*,* $c_2 > 0, c_{22} > 0, c_1 < 0, c_{11} > 0, c_{12} < 0$*, and* c *is homogeneous of degree one;*

(iii) $G \in \mathcal{C}^1$*, has a finite mean and has support* $[0,\infty)$*.*

With assumption 1, standard dynamic programming arguments apply to obtain the existence and uniqueness of value function V. Let $F(D,A)$ be the required fixed cost with which an exporter would be indifferent between continuation and exit. Namely,

$$\beta U + F(D,A) = \max_{D'\geqslant(1-\delta)D} \pi(A,D) - c(D,D') + \beta\,\mathbb{E}[V(A,D',F')], \quad (15)$$

where $U \equiv \mathbb{E}[V(A,D_0,F')]$. Then, $V(A,D,F)$ can be rewritten in two parts:

$$V(A,D,F) = \begin{cases} \beta U + F(D,A) - F, & \text{if} \quad F \leqslant F(D,A), \\ \beta U, & \text{if} \quad F > F(D,A) \end{cases} \quad (16)$$

Plugging (16) back into (15), the following Bellman equation characterizes the required fixed cost:

$$F(D,A) = \max_{D'\geqslant(1-\delta)D} \pi(A,D) - c(D,D') + \beta\int_{F(D_0,A)}^{F(D',A)} G(F)\,dF. \quad (17)$$

We leave a detailed derivation to the appendix. $F(D,A)$ is well defined given that G has a finite mean. It then suffices to solve for the value function $F(D,A)$ instead. Denoting the optimal policy of next period capital by s, s is the policy set of both Bellman equations (17) and (14). The following proposition states that the policy set is still increasing with the presence of exit.

Proposition 3 (Monotonicity)**.** *If* $D' > D$*,* $D_1' \in s(D')$*, and* $D_1 \in s(D)$*, then* $D_1' > D'$*.*

The monotonicity of policy function is a standard result in most growth models. Nevertheless, one should be aware that exit, which essentially clears the return of capital whenever realized, reduces the incentive of investment. In the simple model, more investment unanimously reduces the capital return with decreasing return to scale in the capital. However, investment may now increase the (expected) capital return by reducing the exit hazard. This second force complicates the relationship between current investment and future capital return. Namely, first order conditions would be insufficient to tell whether the policy function is monotonic. Proposition 3 then relies heavily on the supermodularity of the return function.[29]

Next, we define the set of steady states $\tilde{D}_{ss}$ as follows, for all $\tilde{D} \in \tilde{D}_{ss}$,

$$\tilde{D} \in \underset{D' \geqslant (1-\delta)\tilde{D}}{\arg\max}\ \pi(A, \tilde{D}) - c(\tilde{D}, D') + \beta \int_{F(D_0,A)}^{F(D',A)} G(F)\, dF. \tag{18}$$

The following proposition states properties on the convergence of optimal path. Let D_{ss} be the steady state level of capital in the absence of fixed costs, which is guaranteed with the above conditions.[30] Assume henceforth that $D_0 < D_{ss}$, the following proposition states that any path of customer base accumulation converges.

Proposition 4 (Convergence)**.** *Denote by $\{D_t\}_{t=0}^{\infty}$ the optimal path of customer base accumulation. Then, $\{D_t\}_{t=0}^{\infty}$ either increase or decrease to a finite constant. In addition, any steady state in $\tilde{D}_{ss}$ is smaller than D_{ss}.*

Proposition 4 points out that negative growth is possible in a customer base accumulation model. BRV argues that demand accumulation model has limited capacity for generating heterogeneous post-entry dynamics in sales. Specifically, they argue that demand accumulation models are unlikely to generate negative sales growth and increasing variances over age. While their arguments seem reasonable in the simple model, exit breaks the restrictions on customer base accumulation models. First, low productivity firms may retreat slowly from the market due to irreversibility constraints, which exhibit negative growth. Second, as entrants may either grow or contract, it is fairly intuitive to see why the variance of sales in a cohort increases over age. Consider the extreme case that one cohort consists of only two firms, one grows, and the other contracts after entry. The variance then increases as both firms age.

Proposition 5 (Supermodularity)**.** *If $A' > A$, $D'_0 = D_0$ and $D_1 > D_0$, then $D'_t > D_t$ for all t.*

Proposition 5 is the analogy of proposition 2 in this customer base accumulation model with exit. It states similarly that more productive firms, conditional on the initial customer base, will have steeper age profiles in sales. This result is weaker than proposition 2 since the presence of exit muddles the impact of investment on capital return. Nonetheless, the basic intuition that productivity increases capital return still holds. With decreasing return to scale, a higher customer base will eventually decrease the capital return and slows down the growth rate. Therefore, a parallel trend is still plausible with increased measured productivity and initial customer base, like in the simple model. We will see more of this in the next section. Since sales $R = \sigma\pi$ and the

[29]It is frequently used in growth models with non-concave production functions. See, for example, Dechert and Nishimura (1983), Amir, Mirman and Perkins (1991), and Kamihigashi and Roy (2007).

[30]It is straightforward to verify that absent fixed cost, the exit option won't change the steady state of Bellman equation (14) whenever $D_0 < D_{ss}$.

conditional survival rate $\phi = G(F(D, A))$, the following proposition is a translation of the previous results in the context of firm dynamics. It depicts the lifecycle of an entrant with positive sales growth at entry, which is congruent with that in the simple model. We include it here without proof for easy reference.

Proposition 6. *Given positive growth upon entry, both sales and conditional survival rates are increasing to a steady state level. Conditional upon survival, more productive firms would make more sales and have higher conditional survival rate at all time.*

5 Structural Estimation

5.1 Baseline Multi-cohort Model

In this section, we parameterize and estimate the customer base accumulation model with exits for multiple cohorts. Note that all our regression results are interpreted as the behavior of an average firm in an average market. The estimated model is then tailored to describe a firm decision problem faced by a representative firm of each cohort in a representative market. The cohort effect embodies itself in the cohort-specific structural parameters.[31] Reading these estimates off the chart, we will have a transparent idea of what constitutes the cohort effect and by how much.

Time is discrete and ordered by 1, 2, 3, etc. Each domestic firm has a chance to be eligible for export in each period. Once it becomes eligible, the firm is eligible for export forever. That is, it can export whenever it finds exporting profitable. Cohorts are also ordered by 1, 2, 3, etc. A firm will be in cohort c if by the time it becomes eligible for export, $c-1$ cohorts of firms had already exported. Firms in the same cohort receive common cohort-specific productivity shocks whenever they export. Similarly, their initial customer bases are also cohort-specific. Let A^c denote the productivity shock and D_0^c the initial customer base faced by firms of cohort c. In each period after they become export eligible, firms decide on export participation, pricing, and advertising as described in the last section. The following Bellman equation summarizes the decision problem of a firm of cohort c:

$$V(A_c, D, F) = \max\Big\{ \max_{D' \geqslant (1-\delta)D} \pi(A_c, D) - c(D, D') - F + \beta\,\mathbb{E}[V(A_c, D', F')], \\ \beta\,\mathbb{E}[V(A_c, D_0^c, F')]\Big\}. \tag{19}$$

It is worth noting that there will be a new cohort in each period. With the possibility of the fixed cost being arbitrarily close to zero, there are always some newly eligible firms exporting on their first day.

We now present our choices of functional forms. Following our discussion in section 3.2, the profit function has the form $\pi(A, D) = AD^\alpha$. That $\alpha \in (0, 1)$ captures the decreasing return on the

[31] We do not model the interactions between firms of different cohorts that may micro-found the cohort effect. We leave it for future research and share our thoughts in conclusion.

customer base. We also copy the quadratic adjustment cost function in (10) as the cost function:

$$c(D, D') = D' - (1-\delta)D + \phi\frac{(D' - (1-\delta)D)^2}{D}.$$

ϕ captures the influence of adjustment cost, and δ is the depreciation rate. The fixed cost distribution is a composition of two independent shocks. There is a death shock with probability $1-\gamma$. If a firm survives the death shock, it draws from the distribution G, which is a type II Pareto distribution or a Lomax distribution. Explicitly, the CDF

$$G(F) = 1 - \left(1 + \frac{F}{\kappa}\right)^{-\theta},$$

in which κ is the scale parameter and θ the shape parameter. In addition, $\kappa > 0$, and $\theta > 1$. Compared to a standard Pareto distribution, a Lomax distribution has full support on $[0, +\infty)$ for all parameters. The equation below summarizes the distributional assumptions on the fixed cost:

$$F \begin{cases} \sim G(F), & w.p. \quad \gamma, \\ = \infty, & w.p. \quad 1-\gamma. \end{cases}$$

The set of structural parameters for estimation is then $\Omega = \{\beta, \alpha, \phi, \delta, \kappa, \theta, \gamma, \{A_c\}, \{D_0^c\}\}$.

5.2 Identification

We discuss the identification of all structural parameters. Identifying the discount factor in discrete choice models is a known challenge.[32]. Therefore, we follow the usual practice to preset β. The rest of paramters can be divided into two categories: lifecylce parameters $\{\alpha, \phi, \delta, \kappa, \theta, \gamma\}$ and cohort effect parameters $\{A_c, D_0^c\}$. In standard investment models, adjustment cost parameters are usually identified using capital data. As mentioned before, the "capital" in our model, the customer base, is both intangible and market-specific. That precludes the use of any capital-related moments. Hence, capital will be completely latent, and capital accumulation manifests itself in sales growth. Growth in sales by age then uncovers the underlying lifecycle parameters which control capital accumulation. On the other hand, recall that firms in different cohorts differ only in their initial conditions. The relative initial sales by cohort are then informative about changes in initial conditions across cohorts and identify the cohort effect parameters. We illustrate both points using the simple model in 3.4.

As shown in (12), the initial normalized customer base d_0 pins down the trajectory of a firm's sales over time, given a set of lifecycle parameters. d_0^c then denotes the initial normalized customer base of cohort c. We define the relative inital sales of cohort c as the ratio between initial sales of the c-th and the first cohort. In addition, we normalize the measured productivity of the first cohort, i.e., $A_1 = 1$. Then, A_c and D_0^c can be recovered from relative initial sales given d_0^c and a

[32]See Abbring and Daljord (2020) for a thorough investigation.

set of lifecycle paramters. Denoting the relative initial sales of cohort c by n_c,

$$n_c = \frac{A_c(D_0^c)^\alpha}{A_1(D_0^1)^\alpha} = \frac{A_c(A_c^{1/(1-\alpha)}\tilde{D}_{ss}d_0^c)^\alpha}{A_1(A_1^{1/(1-\alpha)}\tilde{D}_{ss}d_0^1)^\alpha} = \frac{A_c^{\frac{1}{1-\alpha}}(d_0^c)^\alpha}{(d_0^1)^\alpha} \Rightarrow A_c = n_c^{1-\alpha}\Big(\frac{d_0^1}{d_0^c}\Big)^{\alpha(1-\alpha)}. \tag{20}$$

Accordingly, the initial customer base of cohort c is obtained from $D_0^c = A_c^{1/(1-\alpha)}\tilde{D}_{ss}d_0^c$.

Since there is no fixed cost in the simple model, it now suffices to discuss the identification of $\{k_0^c\}$ and lifecycle parameters $\{\alpha, \phi, \delta\}$ using sales growth by age. Let y_t^c be the sales growth at age t of cohort c. Then, $y_t = \big(d_{t+1}^c/d_t^c\big)^\alpha$. By the first order conditions, y_t satisfies the following difference equation:

$$B + 2(y_t^c)^{\frac{1}{\alpha}} = (B - \beta + 2)(d_0^c)^{\alpha-1}\prod_{s=0}^{t}(y_s^c)^{\frac{\alpha-1}{\alpha}} + \beta(y_{t+1}^c)^{\frac{2}{\alpha}}, \tag{21}$$

in which $B = \frac{1}{\phi}(1-\beta(1-\delta)) + 2\delta(1-\beta) + \beta\delta^2 + \beta - 2$. For each cohort, sales growth of the first three periods, or sales of the first four periods, is in principle enough to identify the three parameters B, α, and d_0^c. Noticing that B and α are common to all cohorts, sales growth of the first two periods in each cohort are thus sufficient for identification if there are at least two cohorts.

A caveat in the simple case is that ϕ and δ can only be identified up to a function of them, which is the above B. This raises concerns of using dynamic moments for identification. Fortunately, this problem is resolved in the extended model with exit. Since similar normalization can be conducted on the extended model, the relative initial sales identify the cohort effect parameters in the exact same way. We only sketch here the identificatin of lifecycle parameters in the extended model and leave the details to the appendix. An Euler equation analogous to (21) can be obtained as follow:

$$B_t^c + 2(y_t^c)^{\frac{1}{\alpha}} = p_t^c(B - \beta + 2)(d_0^c)^{\alpha-1}\prod_{s=0}^{t}(y_s^c)^{\frac{\alpha-1}{\alpha}} + \beta p_t^c(y_{t+1}^c)^{\frac{2}{\alpha}}, \tag{22}$$

in which $B_t^c = \frac{1}{\phi}\Big[1 - \beta p_t^c(1-\delta)\Big] + 2\delta(1-\beta p_t^c) + \beta p_t^c\delta^2 + \beta p_t^c - 2$. p_t^c is the conditional survival rate at age t of cohort c, so p_t^c is observable. In the simple model, $B_t^c = B = \beta\alpha\tilde{D}_{ss}^{\alpha-1}/\phi + \beta - 2$ since $p_t^c = 1$. As long as ϕ and δ satisfy B, B_t^c is satisfied as well. That there is only one equation with two unknowns renders ϕ and δ unidentified. In the extended model, however,

$$B_t^c = p_t^c B + (\frac{1}{\phi} + 2\delta - 2)(1 - p_t^c)$$

with $p_t^c < 1$. Then, B is not a sufficient statistic of B_t^c, and ϕ and δ can be pinned down using both B_t^c and B. Therefore, lifecycle moments $\{\alpha, \phi, \delta, d_0^c\}$ can be identified using firm dynamics moments, namely, the conditional survival rates p_t^c and sales growth y_t^c. With $\{\alpha, \phi, \delta, d_0^c\}$ in hand, the path of customer base $\{d_t^c\}$ becomes known. As $p_t^c = G_c(F(d_{t+1}^c))$[33], conditional survival rates

[33] See the appendix for the definition of G_c and other derivations.

p_t^c then identify the fixed cost shock parameters $\{\kappa, \theta, \gamma\}$.

To conclude this section, we compare our model setup to a close counterpart in FHY from the perspective of identification. Both of us consider a firm's export decision problem and generate sales growth using customer base accumulation through advertising. Meanwhile, both of us treat the customer base as a latent variable and identify its accumulation by conditional sales or quantity growth. The difference takes place in the approaches to model firms' exit. While we assume a stochastic fixed cost with a continuous distribution, they rely on a combination of an AR(1) process of demand shock and discrete random fixed cost to generate exit. Although a stochastic demand shock is common in firm dynamics models, adding it here will largely increase the risk of un-identification. Demand shocks contribute directly to firms' sales and lead to sales growth by selection. Since both demand shocks and customer base are unobserved, it becomes increasingly risky to identify two latent variables using only one variable, namely, sales. Even if demand shocks affect firms' survival, it remains challenging to identify three sets of parameters on investment, demand shocks, and fixed costs using only two sets of moments, i.e., sales growth and survival rates. On the other hand, our empirical observations make it difficult to utilize other standard moments. For example, it is tough to construct the variance of firms' sales in a representative market, which is a standard moment to identify demand shock parameters. In sum, we avoid additional stochastic components for sharper identification.

5.3 Estimation

We fix the discount rate to match an annual interest rate of 5%, then $\beta = 1.05^{-1}$. Consistent with the reduced form exercise, we focus on the first six cohorts. With $A_1 = 1$, there are 17 parameters for estimation, i.e., $\{\alpha, \phi, \delta, \kappa, \theta, \gamma, \{A_c\}_{c=2}^6, \{D_0^c\}_{c=1}^6\}$. In practice, we estimate the normalized initial customer base d_0^c instead of D_0^c for better numerical stability, i.e., $d_0^c = D_0^c / D_{ss}$. We use a classical minimum distance estimator (CMD) to obtain these estimates. Formally, we minimize the criterion function

$$(\hat{m} - m(\Omega))' \, W \, (\hat{m} - m(\Omega)) \, ,$$

in which $\hat{m}$ are data moments and $m(\Omega)$ model moments with parameter choice Ω.[34] Guided by the identification strategy, we match sales and survival moments of the first six cohorts and the first six years of their lifecycles. Namely, we match the following 65 moments: conditional growth rates in sales by cohort and age y_t^c (6×5), conditional survival rates by cohort and age $p_t^c (6 \times 5)$, and relative initial sales of each cohort n_c (5). All of them are readily obtained from our baseline regression (1). At last, we choose the identity matrix as the weighting matrix W.

The merit of i.i.d. fixed cost shocks with Lomax distribution is that it dramatically simplifies

[34] The asympoptic variance matrix is standard (see e.g. Cameron and Trivedi (2005, Ch 6.7)):

$$(G'WG)^{-1}(G'WV[\hat{m}]WG)(G'WG)^{-1},$$

in which $G = \partial m(\theta)/\partial \theta'$ and $V[\hat{m}]$ is the variance-covariance matrix of moment $\hat{m}$. We obtain an estimate of $V[\hat{m}]$ through bootstrap.

the estimation. With this functional assumption, the integral term in (17) has an exact closed form, so there is no need for numerical integration. Besides, we have shown in the previous section that all model moments have simple closed forms, so simulation is not necessary for estimation. Our estimation then matches data with exact model moments instead of simulated moments. In terms of computation, we adopt a particular policy function iteration scheme to improve efficiency. It is well known that value functions are non-concave and non-smooth in discrete choice models. Value function iteration is a popular choice due to its robustness, despite its reputation for being notoriously slow. Though faster, policy function iteration is particularly problematic with non-concave and non-smooth value functions. Fella (2014) modifies the standard endogenous grid method (EGM) for policy function iteration and applies it to discrete choice models with continuous controls. The idea is that with non-concave value functions, first order conditions are insufficient for optimality. Fella (2014) then imposes an optimality check through grid search to rule out suboptimal first order condition solutions obtained from usual EGM. It shows substantial improvement in both accuracy and speed. This approach is beneficial since we need to solve the firm's decision problem multiple times for different cohorts for each guess of parameters. This way, model moments $m(\Omega)$ can be solved rather efficiently. We then use standard optimization routines to solve for the optimal parameter Ω.

5.4 Results

Table 4 and 5 report the estimated lifecycle and cohort effect parameters respectively. We find a strong decreasing return to scale ($\alpha = 0.42$) with little depreciation ($\delta = 0.004$). The normalized initial customer bases vary very little across the cohort. This is expected since, in the simple model, parallel growth trajectories imply constant normalized initial customer bases. Moreover, there is a clear upward trend in the measured productivity A_c over the cohort. This echoes the theoretical prediction that nondecreasing sales growth implies measured productivity improvement. Without productivity increases, larger initial sales indicate larger initial customer bases, which will depress the sales growth with decreasing return to scale. The model estimates confirm real productivity gains.

Table 4: Lifecylce Parameters

α	ϕ	δ	κ	θ	γ
0.4169	0.0310	0.0039	0.8158	2.0000	0.8497
(0.0009)	(0.0034)	(0.0006)	(0.0017)	(0.0018)	(0.0021)

Model Fit Figure 8, 9 and 10 compare the model-based firm lifecycle moments to the data moments. Given that the model is over-identified, that all moments are well fitted attests to the validity of the model. It is also evident that the model is able to replicate the parallel cross-cohort

Table 5: Cohort Effect Parameters

A^1	A^2	A^3	A^4	A^5	A^6
1	1.0033	1.0067	1.0093	1.0098	1.0104
n.a.	(0.0002)	(0.0004)	(0.0010)	(0.0014)	(0.0005)
d_0^1	d_0^2	d_0^3	d_0^4	d_0^5	d_0^6
0.0015	0.0016	0.0018	0.0019	0.0020	0.0020
(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0003)

sales trajectories seen in the data. We then complete the last step of our model selection and are confident to use this model for understanding the underlying factors of the cohort effect.

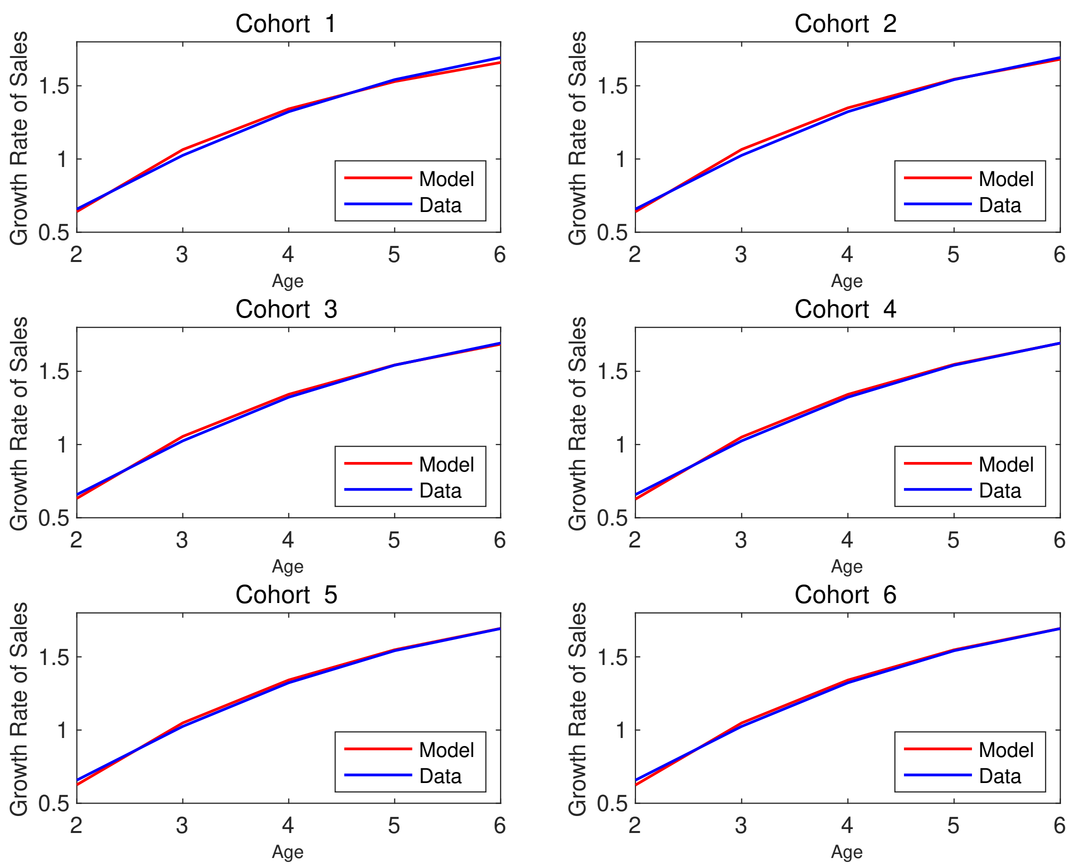


Figure 8: Model Fit - Growth Rate Relative to First Year Sales

Decomposing the cohort effect Finally, we are able to answer the question of what makes the cohort effect. First, we recover the initial customer base D_0^c from d_0^c and compute its changes over the cohort in percentage. Noticing that $D_0^c = d_0^c \tilde{D}_{ss} A^{1/(1-\alpha)}$,

$$\Delta \log D_0^c = \Delta \log d_0^c + \frac{1}{1-\alpha} \Delta \log A^c.$$

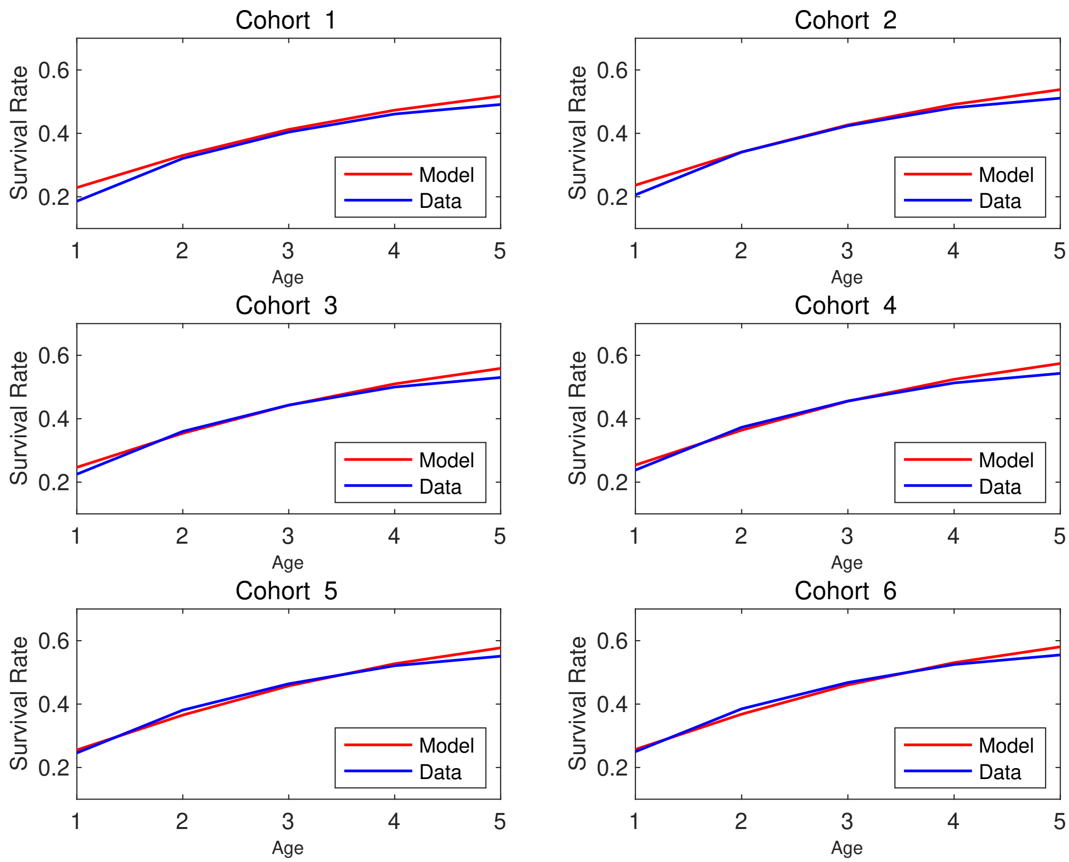


Figure 9: Model Fit - Conditional Survival Rates

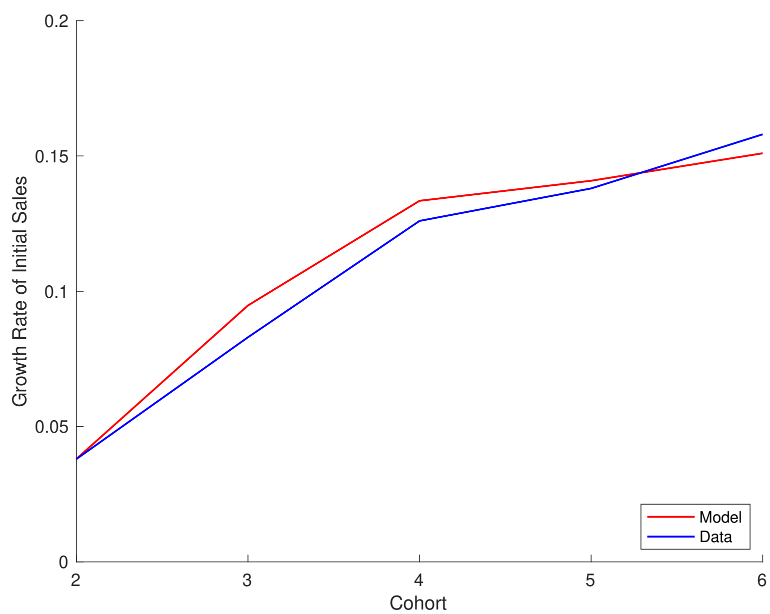


Figure 10: Model Fit - Growth Rate of Initial Sales

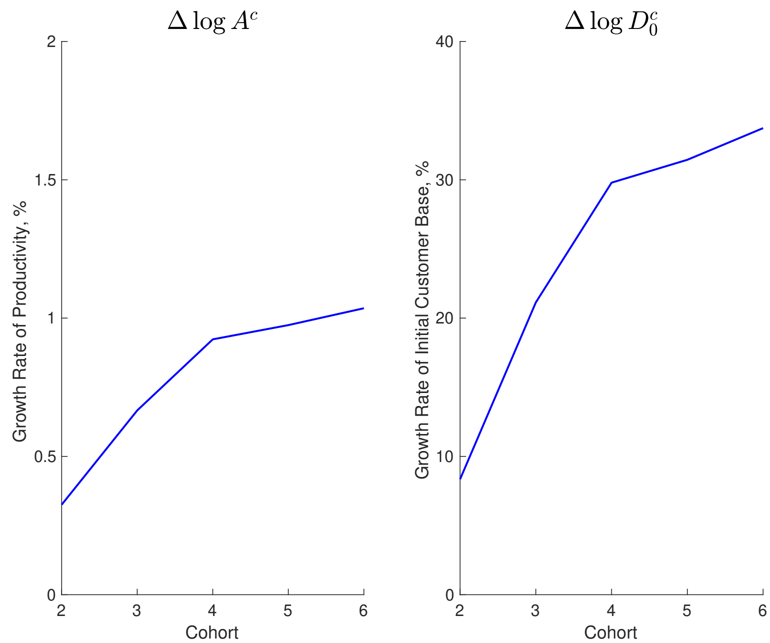


Figure 11: Growth Rate of Productivity and Initial Customer Base

Figure 11 conveniently visualizes the trend of A_c and D_0^c over the cohort in percentage changes. Comparing to A_c, D_0^c grows tremendously over the cohort. This suggests an important demand effect. Business activities of export pioneers create a public good–a national reputation–that raises the foreign demand for all domestic productions. New entrants ship more goods and make more revenue since they start with a larger pool of potential customers.[35]

In the end, we quantify the relative magnitude of the productivity and demand effects. The growth in initial sales can be decomposed as follows:

$$\Delta \log R_0^c = \Delta \log A^c + \alpha \Delta \log D_0^c.$$

Let ξ_A^c and ξ_D^c be the respective contributions to the relative initial sales of cohort c by productivity and initial customer base, i.e.,

$$\xi_A^c = \frac{\Delta \log A^c}{\Delta \log R_0^c}, \qquad \xi_D^c = \frac{\alpha \Delta \log D_0^c}{\Delta \log R_0^c}.$$

Table 6 summarizes the relative contribution of productivity effect and demand effect. The demand effect is dominant in the cohort effect in the sense that it contributes to over 90% of the sales advantages. It alone increases the initial sales of the 6th cohort by 15% relative to the first cohort. This finding is consistent with competitive trade theory that countries export accoding to their comparative advantages. With comparative advantage, or a cost advantage, it is unlikely that the

[35] A number of papers have documented peer effects in international trade decisions and emphasized knowledge diffusion, e.g., Fernandes and Tang (2014), Mion, Opromolla and Sforza (2017) and Bisztray, Koren and Szeidl (2018). More relevantly, Zhao (2018) studies a reputation problem among Chinese electronics exporters. She finds a negative externality from low-quality firms on high-quality firms.

short run export growth is driven by further lowering unit cost.[36]

Table 6: Decomposition of Cohort Effect

cohort	$\Delta \log A^c$	$\Delta \log D_0^c$	$\Delta \log R_0^c$	ξ_A^c	ξ_D^c
2	0.31	8.35	3.80	8.55	91.45
3	0.67	21.13	9.48	7.03	92.97
4	0.92	29.79	13.34	6.92	93.08
5	0.97	31.45	14.08	6.92	93.08
6	1.04	33.74	15.10	6.86	93.14

Notes: All numbers are in precent form.

5.5 An Extension with Strategic Entry

In the baseline model, a firm's cohort status is determined by the time it becomes export eligible and is invariant in its lifetime. In contrast, the cohort is defined by the first year of its current export spells. The discrepancy between the two is whether a firm can choose strategically to enter or re-enter in a later cohort to utilize the cohort effect. We study here an extension in which firms strategically choose their cohort statuses. With strategic entry, time matters for firms' decisions.

The following Bellman equation summarizes the decision problem faced by a current exporter of cohort c in period t:

$$V_t(A_c, D, F) = \max \Big\{ \max_{D' \geqslant (1-\delta)D} \pi(A_c, D) - c(D, D') - F + \beta\, \mathbb{E}[V_{t+1}(A_c, D', F')], \\ \beta\, \mathbb{E}[V_{t+1}(A_{t+1}, D_0^{t+1}, F')] \Big\}. \tag{23}$$

The only deviation from the baseline Bellman equation (19) is the exit option value, which is now time-varying. The strategic entry is most evident with eligible non-exporters in the current period. They will be in cohort t if they enter, so they solve the above Bellman equation with individual states (A_t, D_0^t, F). In other words, their entry decisions are based on not only the current realization of fixed cost but the cohort effect as well.

We assume that the cohort effect stabilizes after many years. That is, there exists a threshold period t^* such that $A_c = A_{t*}$, and $D_0^c = D_0^{t^*}$ for $c \geqslant t_0^*$. The computation is more straightforward in this extension since we only need to solve the recursive problem once for V_{t*} and then obtain other V_t by backward induction. In contrast, we have to solve the recursive problem separately for the value function of each cohort.

[Results are forthcoming in appendix F.2.]

[36]The development of Chilean wine industry discussed in Sabel et al. (2012) is also an example of this view. Chine had a long history of wine production but only exported wine in very small volumes and basically to other Latin American countries. It took the success of a foreign producer–Miguel Torres–to have foreigners recognize Chilean wines.

6 Conclusion

Among Chinese exporters, we find supportive evidence that export pioneers have positive spillovers on their followers. The parallel lifecycles across cohorts suggest that demand learning is unlikely the driver of post-entry sales growth. We account for the cohort effect in a model of customer base accumulation. Through the lens of the model, the cohort effect is a combination of productivity and demand effects: later cohorts have higher measured productivities and larger initial customer bases. The rising demand from customer base expansion explains most of the sales advantage of later cohorts.

We position this paper as an exploration on the existence and composition of the cohort effect. Another important question is about the formation of the cohort effect. One way to understand it is the preferential attachment in Chaney (2014). Suppose domestic producers can search for new customers randomly or through other exporters. The latter way of search is more cost-effective conditional on knowing an exporter, but few producers will know exporters if there are very few in the country. Thus, the overall search cost will decrease if more exporters are in the country. That is to say, followers have both lower unit costs and larger customer bases because pioneers can introduce them customers.[37] We believe it will be a fruitful avenue for future research since a complete understanding of the cohort effect is crucial to the design of export promotion policies for developing countries.

[37] A growing strand of papers have studied international buyer-seller relationships, e.g., Bernard, Moxnes and Ulltveit-Moe (2018) and Eaton et al. (2021, 2022).

References

Abbring, Jaap H., and Øystein Daljord. 2020. “Identifying the discount factor in dynamic discrete choice models.” *Quantitative Economics*, 11(2): 471–501. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE1352.

Albornoz, Facundo, Héctor F. Calvo Pardo, Gregory Corcos, and Emanuel Ornelas. 2012. “Sequential exporting.” *Journal of International Economics*, 88(1): 17–31.

Alessandria, George, Costas Arkolakis, and Kim J. Ruhl. 2021. “Firm Dynamics and Trade.” *Annual Review of Economics*, 13(1): 253–280.

Alfaro-Ureña, Alonso, Juan Manuel Castro-Vincenzi, Sebastián Fanelli, and Eduardo Morales. 2022. “Firm Export Dynamics in Interdependent Markets.” *Mimeo.*

Amir, Rabah, Leonard J. Mirman, and William R. Perkins. 1991. “One-Sector Nonclassical Optimal Growth: Optimality Conditions and Comparative Dynamics.” *International Economic Review*, 32(3): 625–644. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University].

Arkolakis, Costas. 2010. “Market Penetration Costs and the New Consumers Margin in International Trade.” *Journal of Political Economy*, 118(6): 1151–1199. Publisher: The University of Chicago Press.

Arkolakis, Costas. 2016. “A Unified Theory of Firm Selection and Growth.” *The Quarterly Journal of Economics*, 131(1): 89–155.

Arkolakis, Costas, Theodore Papageorgiou, and Olga A. Timoshenko. 2018. “Firm learning and growth.” *Review of Economic Dynamics*, 27: 146–168.

Artopoulos, Alejandro, Daniel Friel, and Juan Carlos Hallak. 2013. “Export emergence of differentiated goods from developing countries: Export pioneers and business practices in Argentina.” *Journal of Development Economics*, 105: 19–35.

Bastos, Paulo, Daniel A. Dias, and Olga A. Timoshenko. 2018. “Learning, prices and firm dynamics.” *Canadian Journal of Economics/Revue canadienne d’économique*, 51(4): 1257–1311. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/caje.12361.

Berman, Nicolas, Vincent Rebeyrol, and Vincent Vicard. 2019. “Demand Learning and Firm Dynamics: Evidence from Exporters.” *The Review of Economics and Statistics*, 101(1): 91–106.

Bernard, Andrew B., Andreas Moxnes, and Karen Helene Ulltveit-Moe. 2018. “Two-Sided Heterogeneity and Trade.” *The Review of Economics and Statistics*, 100(3): 424–439.

Bernard, Andrew B., Esther Ann Boler, Renzo Massari, Jose-Daniel Reyes, and Daria Taglioni. 2017. “Exporter Dynamics and Partial-Year Effects.” *American Economic Review*, 107(10): 3211–3228.

Bisztray, Márta, Miklós Koren, and Adam Szeidl. 2018. “Learning to import from your peers.” *Journal of International Economics*, 115: 242–258.

Blum, Bernardo S., Sebastian Claro, Ignatius Horstmann, and Trevor Tombe. 2020. "The DNA of New Exporters: Spin-offs and FDI at the Extensive Margin of Trade." *American Economic Review: Insights*, 2(3): 397–408.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* Cambridge University Press.

Cebreros, Alfonso. 2016. "The Rewards of Self-Discovery: Learning and Firm Exporter Dynamics." Banco de México 2016-08.

Chaney, Thomas. 2014. "The Network Structure of International Trade." *American Economic Review*, 104(11): 3600–3634.

Chen, Cheng, Tatsuro Senga, Chang Sun, and Hongyong Zhang. 2020. "Uncertainty, Imperfect Information, and Expectation Formation over the Firm's Life Cycle." *CESifo Working Paper*, No. 8468.

Dai, Mi, Madhura Maitra, and Miaojie Yu. 2016. "Unexceptional exporter performance in China? The role of processing trade." *Journal of Development Economics*, 121: 177–189.

Das, Sanghamitra, Mark J. Roberts, and James R. Tybout. 2007. "Market Entry Costs, Producer Heterogeneity, and Export Dynamics." *Econometrica*, 75(3): 837–873. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0262.2007.00769.x.

Dechert, W. Davis, and Kazuo Nishimura. 1983. "A complete characterization of optimal growth paths in an aggregated model with a non-concave production function." *Journal of Economic Theory*, 31(2): 332–354.

Defever, Fabrice, Benedikt Heid, and Mario Larch. 2015. "Spatial exporters." *Journal of International Economics*, 95(1): 145–156.

Dinlersoz, Emin M., and Glenn MacDonald. 2009. "The industry life-cycle of the size distribution of firms." *Review of Economic Dynamics*, 12(4): 648–667.

Drozd, Lukasz A., and Jaromir B. Nosal. 2012. "Understanding International Prices: Customers as Capital." *The American Economic Review*, 102(1): 364–395. Publisher: American Economic Association.

Eaton, Jonathan, David Jinkins, James R. Tybout, and Daniel Xu. 2022. "Two-Sided Search in International Markets." *NBER Working Paper*, No. 29684.

Eaton, Jonathan, Marcela Eslava, David Jinkins, C. J. Krizan, and James R. Tybout. 2021. "A Search and Learning Model of Export Dynamics." *NBER Working Paper*, No. 29100.

Fajgelbaum, Pablo D. 2020. "Labour Market Frictions, Firm Growth, and International Trade." *The Review of Economic Studies*, 87(3): 1213–1260.

Fanelli, Sebastián, and Juan Hallak. 2021. "Export Survival with Uncertainty and Experimentation." Red Nacional de Investigadores en Economía (RedNIE) Working Paper 87.

Fella, Giulio. 2014. "A generalized endogenous grid method for non-smooth and non-concave problems." *Review of Economic Dynamics*, 17(2): 329–344.

Fernandes, Ana P., and Heiwai Tang. 2014. "Learning to export from neighbors." *Journal of International Economics*, 94(1): 67–84.

Fitzgerald, Doireann, Stefanie Haller, and Yaniv Yedid-Levi. 2019. "How Exporters Grow." *CESifo Working Paper*, No. 8077.

Fitzgerald, Doireann, Stefanie Haller, and Yaniv Yedid-Levi. 2022. "How Exporters Grow." *Working Paper, Federal Reserve Bank of Minneapolis.*

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2016. "The Slow Growth of New Plants: Learning about Demand?" *Economica*, 83(329): 91–129.

Freund, Caroline, and Martha Denisse Pierola. 2010. "Export Entrepreneurs : Evidence from Peru." *World Bank Policy Research working paper*, no. WPS 5407. Accepted: 2012-03-19T18:41:41Z.

Gort, Michael, and Steven Klepper. 1982. "Time Paths in the Diffusion of Product Innovations." *The Economic Journal*, 92(367): 630–653. Publisher: [Royal Economic Society, Wiley].

Gourio, François, and Leena Rudanko. 2014. "Customer Capital." *Review of Economic Studies*, 81(3): 1102–1136.

Gumpert, Anna, Haishi Li, Andreas Moxnes, Natalia Ramondo, and Felix Tintelnot. 2020. "The life-cycle dynamics of exporters and multinational firms." *Journal of International Economics*, 126: 103343.

Haidar, Jamal Ibrahim. 2020. "Late-movers outperform first-movers in export markets." *Economics Letters*, 196: 109565.

Hausmann, Ricardo, and Dani Rodrik. 2003. "Economic development as self-discovery." *Journal of Development Economics*, 72(2): 603–633.

Heckman, James, and Richard Robb. 1985. "Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations." In *Cohort Analysis in Social Research: Beyond the Identification Problem.* , ed. William M. Mason and Stephen E. Fienberg, 137–150. New York, NY:Springer.

Hoff, Karla. 1997. "Bayesian learning in an infant industry model." *Journal of International Economics*, 43(3): 409–436.

Iacovone, Leonardo, and Beata S. Javorcik. 2010. "Multi-Product Exporters: Product Churning, Uncertainty and Export Discoveries." *The Economic Journal*, 120(544): 481–499.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica*, 50(3): 649–670.

Jovanovic, Boyan, and Glenn M. MacDonald. 1994. "The Life Cycle of a Competitive Industry." *Journal of Political Economy*, 102(2): 322–347.

Kamihigashi, Takashi, and Santanu Roy. 2007. "A nonsmooth, nonconvex model of optimal growth." *Journal of Economic Theory*, 132(1): 435–460.

Khandelwal, Amit K., Peter K. Schott, and Shang-Jin Wei. 2013. “Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters.” *American Economic Review*, 103(6): 2169–2195.

Klepper, Steven. 1996. “Entry, Exit, Growth, and Innovation over the Product Life Cycle.” *The American Economic Review*, 86(3): 562–583.

Kohn, David, Fernando Leibovici, and Michal Szkup. 2016. “Financial Frictions and New Exporter Dynamics.” *International Economic Review*, 57(2): 453–486.

Li, Shengyu. 2018. “A structural model of productivity, uncertain demand, and export dynamics.” *Journal of International Economics*, 115: 1–15.

Lu, Dan, Asier Mariscal, and Luis-Fernando Mejıa. 2016. “How Firms Accumulate Inputs: Evidence from Import Switching.” *Working Paper*.

Mion, Giordano, Luca David Opromolla, and Alessandro Sforza. 2017. “The Diffusion of Knowledge Via Managers’ Mobility.” *SSRN Electronic Journal*.

Monarch, Ryan, and Tim Schmidt-Eisenlohr. 2016. “Learning and the Value of Relationships in International Trade.” U.S. Census Bureau, Center for Economic Studies Working Paper.

Morales, Eduardo, Gloria Sheu, and Andrés Zahler. 2019. “Extended Gravity.” *The Review of Economic Studies*, 86(6): 2668–2712.

Nguyen, Daniel X. 2012. “Demand uncertainty: Exporting delays and exporting failures.” *Journal of International Economics*, 86(2): 336–344.

Piveteau, Paul. 2021. “An Empirical Dynamic Model of Trade with Consumer Accumulation.” *American Economic Journal: Microeconomics*, 13(4): 23–63.

Rho, Youngwoo, and Joel Rodrigue. 2016. “Firm-Level Investment and Export Dynamics*.” *International Economic Review*, 57(1): 271–304.

Roberts, Mark J., and James R. Tybout. 1997. “The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs.” *The American Economic Review*, 87(4): 545–564.

Roberts, Mark J, Daniel Yi Xu, Xiaoyan Fan, and Shengxing Zhang. 2018. “The Role of Firm Factors in Demand, Cost, and Export Market Selection for Chinese Footwear Producers.” *The Review of Economic Studies*, 85(4): 2429–2461.

Rodrigue, Joel, and Yong Tan. 2019. “Price, Product Quality, and Exporter Dynamics: Evidence from China.” *International Economic Review*, 60(4): 1911–1955.

Ruhl, Kim J., and Jonathan L. Willis. 2017. “New Exporter Dynamics.” *International Economic Review*, 58(3): 703–726.

Sabel, Charles, Gabriel Sánchez, Inés Butler, Ricardo Rozemberg, Hernán Ruffo, María Angélica Arbeláez, Marcela Meléndez Arjona, Nicolás León, Angela Da Rocha, Beatriz Kury, Joana Monteiro, Manuel R. Agosin, Claudio Bravo Ortega, Edgar Aragón, Alexandre Darzé, Regis Bonelli, Armando Castelar Pinheiro, Michele Snoeck, Lucía Pittaluga, Alejandro Artopoulos, Daniel Friel, and Juan Carlos Hallak. 2012. *Export Pioneers in Latin America.* Inter-American Development Bank.

Segura-Cayuela, Rubén, and Josep M. Vilarrubia. 2008. "Uncertainty and entry into export markets." Banco de España 0811.

Timoshenko, Olga A. 2015. "Learning versus sunk costs explanations of export persistence." *European Economic Review*, 79: 113–128.

Wagner, Rodrigo, and Andrés Zahler. 2015. "New exports from emerging markets: Do followers benefit from pioneers?" *Journal of Development Economics*, 114: 203–223.

Wei, Shang-Jin, Ziru Wei, and Jianhuan Xu. 2021. "On the market failure of "missing pioneers"." *Journal of Development Economics*, 152: 102705.

Zhao, Yingyan. 2018. "Your (Country's) Reputation Precedes You: Information Asymmetry, Externalities and the Quality of Exports." *Working Paper. George Washington University.*

Appendix A Data

A.1 Concordance

The data we use in this paper, the Chinese Customs Transactions Database, has the full coverage of export records of Chinese firms. The raw data is recorded by the firm, country, and HS 8-digit product category at a monthly frequency between 2000 and 2006, and at a yearly frequency between 2007 and 2011. We drop all the transactions exported to tax-avoiding destinations, including Liechtenstein, Andorra, Monaco, Channel Islands, Isle of Man, Cayman Islands, Bermuda, Bahamas, Netherlands Antilles, British Virgin Islands, Cook Islands, Solomon Islands, Society Islands, Marshall Islands, Canary Islands, and Cook Islands. Then we aggregate the data at the level of firm, country, HS 6-digit product category, and year.

The HS 6-digit classification system changes every five year. During our sample period 1996-2011, it is slightly revised in 1996, 2002, and 2007. With each change, some HS codes are removed, others have changes to their definitions – with some definitions expanding to include more goods and some definitions being grouped into more larger category. Some reconfigurations have been made to provide better representation for new technologies or new products. We concord the product-level data over time according to the correspondence tables made availble by UN Trade Statistics, and keep the HS six-digit product categories which can be one-one mapped in the HS classification system between years 1996, 2002 and 2007. [38] Specifically, there are respectively 5113, 5224, and 5051 HS six-digit product categories in 1996, 2002, and 2007. The number of HS 6-digit product categories with one-one mapping is 4744, 4080, and 4471 for 1996-2002, 1996-2007, and 2002-2007. Finally, we get a list with 4080 categories of HS 6-digit products.

A.2 Sampling

To seperate new markets from old markets, we combine product-level trade flow data from China to its trade parters during 1996 to 1999, and firm-level trade data during 2000 to 2011. We first set a window period before 2001, and label the markets exported by Chinese firms during the window period as old markets. The new markets are then defined as the markets exported by Chinese firms after 2002, and not exported during the window period. We experiment different window periods. Table A1 presents some summary statistics in terms of definition of new market. In particular, the last column, ratio of false labelled, indicates the probability of a type II error, namely, the probability that a selected new market has export records prior to the window period. We find that when using the window period of 1997-2001, the number of false-defined markets is only around 2% of total number of new markets. This ratio is much less than the one using shorter window periods, suggesting it is reasonable to use the 5-years window period to define new markets.

[38] The data is available in the website https://unstats.un.org/unsd/trade/classifications/correspondence-tables.asp.

Table A1: Definition of New Market

Window Period	Length	# of Products	# of Markets	# of Product-Markets	Ratio of False Labelled
1997-2001	5 Years	4,031	192	219,842	4,574 (2.08%)
1998-2001	4 Years	4,031	192	225,672	10,404 (4.61%)
1999-2001	3 Years	4,031	192	233,885	18,617 (7.96%)
2000-2001	2 Years	4,031	192	246,669	31,401 (12.73%)

Notes: Define false labelled as: exported in at least one year from 1996 to the year before the window period, and exported again in at least one year during 2002 to 2011.
Data source: Customs Data 2000-2011, and Trade Flow Data 1996-1999.

Table A2: Sample Export Records of A New Market

Firm ID	Year	Destination	Product	Cohort	Age	Cohort Year
1	2002	340	844711	1	1	2002
1	2003	340	844711	1	2	2002
1	2004	340	844711	1	3	2002
2	2004	340	844711	2	1	2004
3	2004	340	844711	2	1	2004
2	2005	340	844711	2	2	2004
4	2005	340	844711	3	1	2005
1	2006	340	844711	4	1	2006
2	2006	340	844711	2	3	2004
4	2006	340	844711	3	2	2005
5	2006	340	844711	4	1	2006

Table A3: Median Export Growth Between Old and New Markets

	Export growth: number of firms		Export growth: total export value	
	New markets	Old markets	New markets	Old markets
Market spell length >=2	0.111	0.222	0.579	0.563
Market spell length >=4	0.259	0.25	0.902	0.616
Market spell length >=6	0.333	0.271	1.222	0.654
Market spell length >=8	0.389	0.296	1.543	0.726

Notes: Table A3 is calculated by authors. The market spell is the longest period during which Chinese firms consecutively exports to that market.
Data source: Customs Data 2002-2011.

A.3 Control Test

In an attempt to provide further support on the assumption that our control bundle γ_{4dt} plus $\mathbf{x}_{jdt}$ could capture most of yearly demand shocks, we do the following control test. Note that concerns for yearly demand shock are also important in the more widely studied area of exporter dynamics, where cohort effect is assumed absent. A common treatment is then to control for HS6 product×destination×year fixed effects γ_{jdt} directly. Since we assume that γ_{4dt} would be an approximate sufficient statistics for γ_{jdt}, controlling for γ_{4dt} would in principle change very little of the results. In our sample, we run the standard regressions of export value on firm's age in a market, controlling all for η_{ijt} while controlling for γ_{4dt} and γ_{jdt} in separate regressions. Table A4 displays the regression results. It is reassuring to see coefficients of age are very similar in all regressions. Reading from the adjusted R^2, we also lose very little explanatory power replacing γ_{jdt} by γ_{4dt}. The regression controlling for γ_{4dt} and market size in column (5) even has the equal adjusted R^2 to the setting controling for γ_{jdt} in column (7). Then, we include γ_{4dt} and $\mathbf{x}_{jdt}$ to control yearly demand shock instead of γ_{jdt} in the rest of this paper.

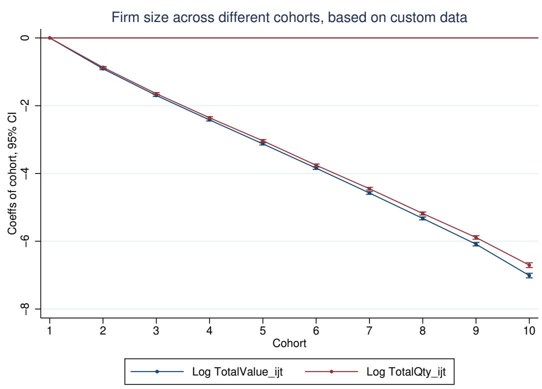
Table A4: Control Test in Different Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log Export Value						
Age							
2	0.769***	0.677***	0.717***	0.648***	0.640***	0.678***	0.644***
	(0.007)	(0.007)	(0.010)	(0.009)	(0.009)	(0.013)	(0.010)
3	1.191***	1.048***	1.111***	1.001***	0.988***	1.049***	0.992***
	(0.011)	(0.011)	(0.014)	(0.014)	(0.014)	(0.018)	(0.015)
4	1.509***	1.325***	1.383***	1.280***	1.265***	1.308***	1.260***
	(0.015)	(0.015)	(0.019)	(0.019)	(0.019)	(0.025)	(0.021)
5	1.739***	1.523***	1.566***	1.478***	1.464***	1.476***	1.430***
	(0.022)	(0.021)	(0.027)	(0.026)	(0.026)	(0.033)	(0.028)
6	1.917***	1.672***	1.751***	1.610***	1.591***	1.624***	1.573***
	(0.032)	(0.031)	(0.039)	(0.038)	(0.038)	(0.047)	(0.041)
7+	2.170***	1.902***	1.948***	1.833***	1.816***	1.813***	1.799***
	(0.040)	(0.038)	(0.047)	(0.044)	(0.044)	(0.055)	(0.048)
Log Market Size		0.201***	0.210***		0.200***	0.190***	
		(0.001)	(0.002)		(0.005)	(0.010)	
Tariff Rate			0.005***			0.004	
			(0.000)			(0.003)	
Fixed Effect	Firm-product (HS6)-year			Firm-product(HS6)-year Product (HS4)-country-year			Firm-product(HS6)-year Product (HS6)-country-year
N	623,735	623,735	288,889	486,598	486,598	226,071	407,421
Adj. R^2	0.619	0.639	0.639	0.670	0.672	0.667	0.672

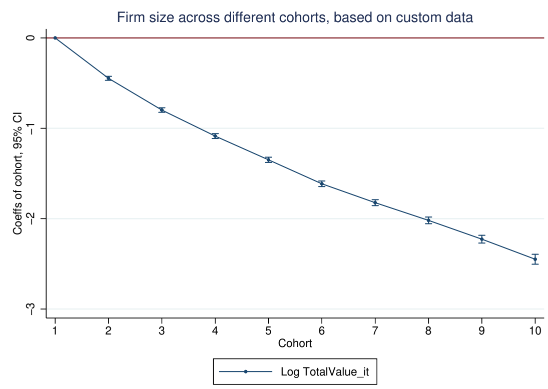
Notes: Table A4 reports coefficients of firm's age. The observation is at the firm-product (HS6)-country-year level. Omitted category is age equal to one. Market size is measured by total value imported by destination country. Tariff rate is measured by weigted average tariff rate implemented by destination country. Market size and tariff rate are at the product (HS6)-country-year level. Standard errors in parentheses are robust. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.
Source: Customs Data 2002-2011, market size from CEPII, and tariff rate from WITS-TRAINS.

Appendix B Additional Reduced Form Results

B.1 Robustness Check



(a) Firm-Product-Year



(b) Firm-Year

Figure A1: Larger Exporters Enter in Earlier Cohorts

Notes: Figure A1 shows the associations between firms' cohort number and firm size. The estimation equation is

$$Y = \beta_c' \mathbf{c}_{ijdt} + \lambda_{jdt} + \varepsilon_{ijdt}. \tag{A.1}$$

The observation is at the level of firm i, product j, destination country d, and year t. The dependent variables in Panel A are at the firm-product-year level, Y_{ijt}, which are firm i's log total export value/quantity across all destinations (old and new) selling product j at year t. The dependent variable in Panel B is at the firm-year level, Y_{it}, which is firm i's log total export value across all products and destinations (old and new) at year t. $\mathbf{c}_{ijdt}$ is a vector of dummies which characterizes the cohort number of firm i in market jd and year t. λ_{jdt} is the HS6 product-country -year fixed effect. Standard errors are robust clustered.
Source: Customs Data 2002-2011.

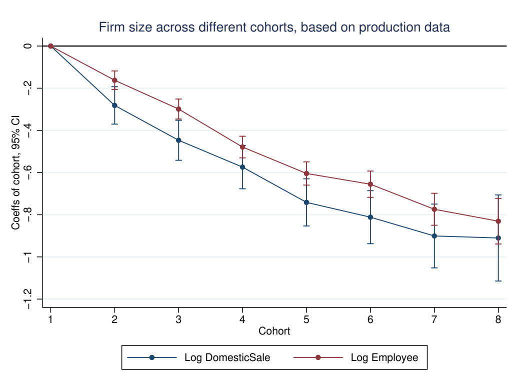


Figure A2: Larger Exporters Enter in Earlier Cohorts: Based on Production Data

Notes: Figure A2 shows the associations between firms' cohort number and firm size. The estimation equation is the same as the one of figure A1. Firm size measures are based on production data, including domestic sales and employment at the firm-year level.
Source: Annual Survey of Industrial Firms 2002-2009, from National Bureau of Statistics, China

Table A5: Robustness Check on Cohort Effect: Selection

	(1) Markets with firms exporting spell >= 2		(2) Markets with # cohort >=2		(3) Markets with # cohort >= 4		(4) Markets with # cohort >= 6	
Panel A Dependent Variable = Log Export Value								
Cohort 2	0.005	(0.026)	0.030*	(0.017)	0.020	(0.021)	0.040	(0.030)
3	0.048	(0.030)	0.073***	(0.019)	0.054**	(0.024)	0.041	(0.034)
4	0.085**	(0.036)	0.117***	(0.022)	0.101***	(0.027)	0.079**	(0.039)
5	0.105**	(0.042)	0.126***	(0.024)	0.107***	(0.030)	0.091**	(0.045)
6	0.121**	(0.049)	0.147***	(0.027)	0.131***	(0.034)	0.121**	(0.050)
7+	0.115**	(0.058)	0.159***	(0.031)	0.142***	(0.038)	0.140**	(0.057)
Age 2	0.461***	(0.019)	0.658***	(0.010)	0.663***	(0.011)	0.668***	(0.013)
3	0.838***	(0.026)	1.025***	(0.015)	1.031***	(0.017)	1.041***	(0.021)
4	1.124***	(0.035)	1.318***	(0.022)	1.328***	(0.024)	1.330***	(0.030)
5	1.318***	(0.045)	1.535***	(0.030)	1.534***	(0.032)	1.523***	(0.041)
6	1.486***	(0.060)	1.689***	(0.042)	1.678***	(0.045)	1.650***	(0.055)
7+	1.759***	(0.072)	1.936***	(0.050)	1.952***	(0.054)	1.947***	(0.066)
Log Market Size	0.250***	(0.012)	0.196***	(0.006)	0.188***	(0.007)	0.190***	(0.010)
Fixed Effect	Firm-product (HS6)-year, Product (HS4)-country-year							
N	142,432		471,674		416,484		323,238	
adj. R2	0.664		0.670		0.663		0.651	
Panel B Dependent Variable = Log Export Quantity								
Cohort 2	0.032	(0.026)	0.049***	(0.017)	0.043**	(0.022)	0.071**	(0.031)
3	0.081***	(0.031)	0.103***	(0.020)	0.087***	(0.025)	0.074**	(0.035)
4	0.107***	(0.037)	0.136***	(0.022)	0.119***	(0.028)	0.097**	(0.040)
5	0.144***	(0.043)	0.157***	(0.025)	0.136***	(0.031)	0.113**	(0.046)
6	0.158***	(0.050)	0.180***	(0.028)	0.158***	(0.035)	0.135***	(0.052)
7+	0.161***	(0.060)	0.203***	(0.032)	0.177***	(0.039)	0.160***	(0.059)
Age 2	0.485***	(0.019)	0.686***	(0.010)	0.687***	(0.011)	0.694***	(0.013)
3	0.877***	(0.027)	1.073***	(0.016)	1.071***	(0.018)	1.078***	(0.022)
4	1.176***	(0.036)	1.372***	(0.022)	1.376***	(0.025)	1.374***	(0.031)
5	1.376***	(0.047)	1.600***	(0.031)	1.590***	(0.034)	1.567***	(0.042)
6	1.543***	(0.062)	1.760***	(0.043)	1.737***	(0.046)	1.699***	(0.057)
7+	1.816***	(0.075)	2.027***	(0.051)	2.028***	(0.056)	2.009***	(0.069)
Log Market Size	0.247***	(0.012)	0.193***	(0.006)	0.185***	(0.007)	0.184***	(0.010)
Fixed Effect	Firm-product (HS6)-year, Product (HS4)-country-year							
N	141,625		469,647		414,832		322,016	
adj. R2	0.879		0.869		0.865		0.856	
Panel C Dependent Variable = Survival Rate								
Cohort 2	0.015**	(0.006)	0.022***	(0.005)	0.021***	(0.007)	0.019**	(0.009)
3	0.030***	(0.008)	0.041***	(0.006)	0.044***	(0.008)	0.042***	(0.011)
4	0.033***	(0.009)	0.054***	(0.007)	0.056***	(0.009)	0.049***	(0.013)
5	0.029***	(0.011)	0.061***	(0.008)	0.063***	(0.010)	0.056***	(0.015)
6	0.031**	(0.013)	0.066***	(0.010)	0.068***	(0.012)	0.064***	(0.017)
7+	0.040***	(0.015)	0.078***	(0.011)	0.080***	(0.014)	0.078***	(0.020)
Age 2			0.135***	(0.004)	0.135***	(0.004)	0.134***	(0.005)
3	0.081***	(0.006)	0.219***	(0.006)	0.218***	(0.006)	0.218***	(0.008)
4	0.134***	(0.008)	0.274***	(0.008)	0.278***	(0.009)	0.278***	(0.011)
5	0.160***	(0.011)	0.305***	(0.011)	0.303***	(0.012)	0.297***	(0.015)
6	0.179***	(0.016)	0.332***	(0.015)	0.332***	(0.016)	0.320***	(0.020)
7+	0.185***	(0.021)	0.360***	(0.020)	0.367***	(0.022)	0.357***	(0.026)
Log Market Size	0.020***	(0.003)	0.019***	(0.002)	0.019***	(0.002)	0.016***	(0.003)
Fixed Effect	Firm-product (HS6)-year, Product (HS4)-country-year							
N	119,646		349,045		309,699		242,348	
adj. R2	0.439		0.389		0.393		0.405	

Notes: Table A5 reports main coefficients of firm's age and cohort. The observation is at the firm-product(HS6)-country-year level. Market size is measured by total value imported by destination country, and at the product (HS6)-country-year level. Standard errors in parentheses are robust. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.
Source: Customs Data 2002-2011, market size from CEPII.

Table A6: Robustness Check on Cohort Effect: Re-entrant

	(1) Log Export Value	(2) Log Export Quantity	(3) Survival Rate
Cohort			
2	0.032**	0.050***	0.018***
	(0.015)	(0.015)	(0.005)
3	0.072***	0.101***	0.036***
	(0.017)	(0.018)	(0.006)
4	0.109***	0.126***	0.048***
	(0.020)	(0.020)	(0.007)
5	0.116***	0.147***	0.054***
	(0.022)	(0.023)	(0.008)
6	0.133***	0.164***	0.058***
	(0.025)	(0.026)	(0.009)
7+	0.139***	0.182***	0.068***
	(0.029)	(0.029)	(0.011)
Age			
2	0.664***	0.693***	0.136***
	(0.010)	(0.010)	(0.003)
3	1.048***	1.097***	0.224***
	(0.015)	(0.016)	(0.005)
4	1.366***	1.423***	0.286***
	(0.021)	(0.022)	(0.008)
5	1.603***	1.670***	0.321***
	(0.029)	(0.030)	(0.011)
6	1.770***	1.845***	0.352***
	(0.042)	(0.043)	(0.015)
7+	2.046***	2.142***	0.380***
	(0.049)	(0.051)	(0.020)
Log Market Size	0.195***	0.191***	0.019***
	(0.005)	(0.005)	(0.002)
Dummy For Re-entrant	0.202***	0.215***	0.055***
	(0.013)	(0.013)	(0.005)
Fixed Effect	Firm-product (HS6)-year, Product (HS4)-country-year		
N	486,598	484,439	359,517
adj. R2	0.673	0.870	0.389

Notes: Table A6 reports main coefficients of firm's age and cohort. The observation is at the firm-product(HS6)-country-year level. Dummy for re-entrant is a dummy variable at the firm-product(HS6)-country level, equal to 0 indicating the period of firm's first export spell and 1 for the period of firm's later export spells. Market size is measured by total value imported by destination country, and is at the product (HS6)-country-year level. Standard errors in parentheses are robust. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.
Source: Customs Data 2002-2011, and market size from CEPII.

Table A7: Robustness Check on Cohort Effect: Exclude Processing Trade

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Export Value		Log Export Quantity		Survival Rate	
	Drop processing transactions	Drop processing firms	Drop processing transactions	Drop processing firms	Drop processing transactions	Drop processing firms
Cohort						
2	0.047**	0.039**	0.068***	0.057***	0.020***	0.020***
	(0.020)	(0.015)	(0.021)	(0.015)	(0.007)	(0.005)
3	0.112***	0.084***	0.135***	0.115***	0.032***	0.040***
	(0.024)	(0.017)	(0.024)	(0.018)	(0.008)	(0.006)
4	0.171***	0.127***	0.198***	0.145***	0.047***	0.053***
	(0.027)	(0.020)	(0.028)	(0.020)	(0.009)	(0.007)
5	0.190***	0.139***	0.225***	0.171***	0.064***	0.060***
	(0.031)	(0.022)	(0.032)	(0.023)	(0.011)	(0.008)
6	0.220***	0.159***	0.260***	0.192***	0.075***	0.065***
	(0.035)	(0.025)	(0.036)	(0.026)	(0.013)	(0.009)
7+	0.299***	0.172***	0.344***	0.217***	0.078***	0.077***
	(0.040)	(0.029)	(0.041)	(0.029)	(0.015)	(0.011)
Age						
2	0.580***	0.658***	0.603***	0.686***	0.090***	0.135***
	(0.013)	(0.010)	(0.014)	(0.010)	(0.005)	(0.003)
3	0.878***	1.025***	0.934***	1.073***	0.148***	0.218***
	(0.022)	(0.015)	(0.023)	(0.016)	(0.008)	(0.005)
4	1.125***	1.324***	1.182***	1.378***	0.178***	0.275***
	(0.034)	(0.021)	(0.035)	(0.022)	(0.013)	(0.008)
5	1.296***	1.543***	1.347***	1.607***	0.205***	0.306***
	(0.050)	(0.029)	(0.052)	(0.030)	(0.019)	(0.011)
6	1.420***	1.693***	1.480***	1.765***	0.201***	0.333***
	(0.080)	(0.041)	(0.082)	(0.042)	(0.032)	(0.015)
7+	1.528***	1.953***	1.536***	2.043***	0.193***	0.357***
	(0.101)	(0.049)	(0.104)	(0.051)	(0.046)	(0.020)
Log Market Size	0.204***	0.196***	0.202***	0.192***	0.017***	0.019***
	(0.007)	(0.005)	(0.007)	(0.005)	(0.003)	(0.002)
Fixed Effect	Firm-product (HS6)-year, Product (HS4)-country-year					
N	301,217	485,856	299,869	483,697	230,047	358,791
adj. R2	0.660	0.672	0.870	0.870	0.257	0.388

Notes: Table A7 reports main coefficients of firm's age and cohort. The observation is at the firm-product(HS6)-country-year level. In Column (1)(3)(5), we drop all firm-product(HS8)-country-year level transactions involving processing trade (referred to as "processing transactions"), and then aggregate the rest of transactions to firm-product(HS6)-country-year level. In Column (2)(4)(6), we drop firms that only engage in processing transactions (referred to as "processing firms"). Market size is measured by total value imported by destination country, and is at the product (HS6)-country-year level. Standard errors in parentheses are robust. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.
Source: Customs Data 2002-2011, and market size from CEPII.

Table A8: Robustness Check on Cohort Effect: Private Firm

	(1) Log Export Value	(2) Log Export Quantity	(3) Survival Rate
Cohort			
2	0.010	0.031	0.017
	(0.036)	(0.037)	(0.012)
3	0.118***	0.130***	0.041***
	(0.042)	(0.043)	(0.014)
4	0.194***	0.196***	0.065***
	(0.048)	(0.049)	(0.016)
5	0.168***	0.182***	0.062***
	(0.055)	(0.057)	(0.020)
6	0.216***	0.224***	0.076***
	(0.064)	(0.065)	(0.023)
7+	0.227***	0.256***	0.098***
	(0.075)	(0.077)	(0.028)
Age			
2	0.646***	0.676***	0.134***
	(0.024)	(0.025)	(0.008)
3	1.008***	1.045***	0.208***
	(0.037)	(0.038)	(0.013)
4	1.249***	1.318***	0.267***
	(0.050)	(0.052)	(0.018)
5	1.468***	1.527***	0.313***
	(0.067)	(0.069)	(0.025)
6	1.677***	1.766***	0.333***
	(0.091)	(0.093)	(0.034)
7+	1.861***	1.960***	0.334***
	(0.112)	(0.115)	(0.048)
Log Market Size	0.192***	0.194***	0.028***
	(0.014)	(0.014)	(0.005)
Fixed Effect	Firm-product (HS6)-year, Product (HS4)-country-year		
N	84,667	84,335	68,133
adj. R2	0.644	0.877	0.354

Notes: Table A8 reports main coefficients of firm's age and cohort. The samples only include private firms. The observation is at the firm-product(HS6)-country-year level. Market size is measured by total value imported by destination country, and is at the product (HS6)-country-year level. Standard errors in parentheses are robust. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.
Source: Customs Data 2002-2011, and market size from CEPII.

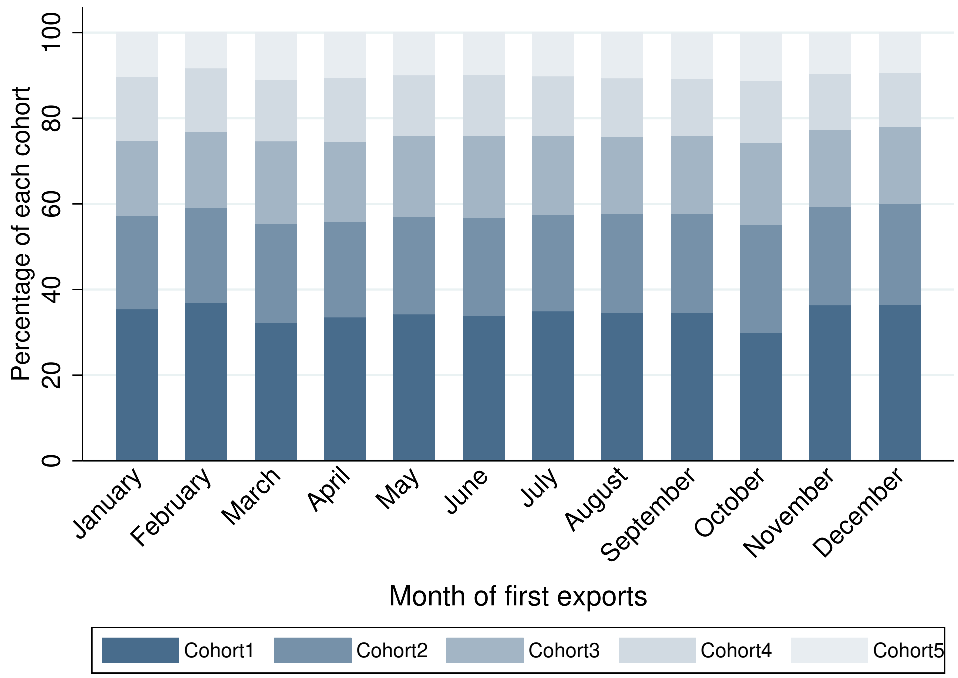


Figure A3: Partial Year Effect: The Distribution of Cohort Across Months of First Exports

Notes: The month of first export indicates the initial month of export for a firm-product(HS6)-country-cohort quadruplet. The number of firm-product(HS6)-country triplets of cohort 1 is respectively 11318, 8348, 10437, 12265, 11346, 12677, 13582, 13832, 14049, 10481, 14477, 15182 for which starts exporting in January-December.
Source: Customs Data 2002-2006 at monthly level.

Table A9: Firms' Lifecycles in Sales, Quantity, and Survival Rate

		(1)		(2)		(3)	
		Log Export Value		Log Export Quantity		Survival Rate	
Cohort				Age intercept			
	2	0.022	(0.023)	0.034	(0.024)	0.015*	(0.008)
	3	0.052**	(0.025)	0.073***	(0.025)	0.027***	(0.008)
	4	0.104***	(0.027)	0.115***	(0.027)	0.043***	(0.009)
	5	0.109***	(0.028)	0.132***	(0.029)	0.059***	(0.010)
	6	0.117***	(0.030)	0.146***	(0.031)	0.057***	(0.011)
	7+	0.158***	(0.032)	0.200***	(0.033)	0.068***	(0.012)
Age				Cohort intercept			
	2	0.635***	(0.031)	0.659***	(0.031)	0.123***	(0.010)
	3	0.947***	(0.040)	1.001***	(0.040)	0.203***	(0.013)
	4	1.251***	(0.049)	1.283***	(0.050)	0.258***	(0.015)
	5	1.535***	(0.060)	1.572***	(0.061)	0.306***	(0.018)
	6	1.695***	(0.075)	1.746***	(0.076)	0.326***	(0.024)
	7+	1.963***	(0.071)	2.024***	(0.073)	0.359***	(0.026)
Age				Cohort = 2			
	2	0.013	(0.041)	0.017	(0.042)	0.001	(0.013)
	3	0.034	(0.050)	0.023	(0.051)	0.012	(0.016)
	4	0.060	(0.062)	0.071	(0.063)	0.027	(0.019)
	5	0.004	(0.078)	0.034	(0.080)	-0.004	(0.024)
	6	-0.009	(0.099)	0.013	(0.101)	-0.007	(0.031)
	7+	-0.006	(0.099)	0.067	(0.102)	-0.021	(0.039)
Age				Cohort = 3			
	2	0.033	(0.039)	0.042	(0.039)	0.018	(0.013)
	3	0.096*	(0.052)	0.095*	(0.052)	0.018	(0.017)
	4	0.061	(0.063)	0.088	(0.065)	0.034*	(0.020)
	5	0.021	(0.078)	0.071	(0.080)	0.012	(0.025)
	6	0.010	(0.104)	0.016	(0.106)	0.044	(0.035)
	7+	-0.009	(0.115)	0.005	(0.117)	0.001	(0.051)
Age				Cohort = 4			
	2	0.024	(0.039)	0.021	(0.040)	0.021	(0.013)
	3	0.063	(0.051)	0.077	(0.052)	0.022	(0.017)
	4	0.069	(0.066)	0.104	(0.067)	0.028	(0.022)
	5	-0.037	(0.083)	-0.038	(0.085)	-0.010	(0.028)
	6	-0.006	(0.116)	0.027	(0.120)	-0.019	(0.047)
	7+	-0.156	(0.179)	-0.105	(0.183)		
Age				Cohort = 5			
	2	0.013	(0.040)	0.025	(0.041)	0.002	(0.013)
	3	0.098*	(0.053)	0.081	(0.054)	0.009	(0.018)
	4	0.140**	(0.069)	0.169**	(0.070)	-0.012	(0.024)
	5	0.064	(0.091)	0.125	(0.094)	-0.011	(0.035)
	6	-0.026	(0.147)	0.081	(0.148)		
Age				Cohort = 6			
	2	0.068	(0.041)	0.073*	(0.042)	0.013	(0.014)

	(1)		(2)		(3)	
	Log Export Value		Log Export Quantity		Survival Rate	
3	0.119**	(0.056)	0.110*	(0.058)	0.028	(0.020)
4	0.152**	(0.076)	0.189**	(0.078)	-0.019	(0.031)
5	0.024	(0.123)	0.038	(0.128)		
Age			Cohort = 7+			
2	0.005	(0.037)	0.008	(0.037)	0.025*	(0.014)
3	0.139**	(0.055)	0.119**	(0.057)	0.011	(0.026)
4	-0.016	(0.104)	0.004	(0.107)		
Log Market Size	0.196***	(0.005)	0.192***	(0.005)	0.019***	(0.002)
Fixed Effect	Firm-product (HS6)-year, Product (HS4)-country-year					
N	486,598		484,439		359,517	
adj. R2	0.672		0.870		0.388	

Notes: Table A9 reports main coefficients of firm's age, cohort, and their interactions. The observation is at the firm-product(HS6)-country-year level. Market size is measured by total value imported by destination country, and is at the product (HS6)-country-year level. Standard errors in parentheses are robust. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.

Source: Customs Data 2002-2011, and market size from CEPII.

Table A10: Sign Test: p-Value for H_0: $\beta_c \beta_i^{c,a} = 0$

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Dep. Var. = Log Export Value						
	Cohort = 2	Cohort = 3	Cohort = 4	Cohort = 5	Cohort = 6	Cohort = 7+
Age = 2	0.708	0.316	0.513	0.745	0.071	0.884
Age = 3	0.460	0.055	0.174	0.044	0.027	0.011
Age = 4	0.397	0.297	0.257	0.036	0.044	0.879
Age = 5	0.957	0.784	0.666	0.467	0.843	
Age = 6	0.933	0.920	0.962	0.862		
Age = 7+	0.951	0.939	0.409			
Panel B: Dep. Var. = Log Export Quantity						
	Cohort = 2	Cohort = 3	Cohort = 4	Cohort = 5	Cohort = 6	Cohort = 7+
Age = 2	0.629	0.222	0.579	0.509	0.056	0.829
Age = 3	0.609	0.044	0.107	0.102	0.042	0.030
Age = 4	0.265	0.146	0.096	0.013	0.015	0.968
Age = 5	0.654	0.350	0.668	0.167	0.764	
Age = 6	0.900	0.880	0.818	0.578		
Age = 7+	0.508	0.969	0.579			
Panel C: Dep. Var. = Survival Rate						
	Cohort = 2	Cohort = 3	Cohort = 4	Cohort = 5	Cohort = 6	Cohort = 7+
Age = 2	0.930	0.100	0.063	0.890	0.347	0.055
Age = 3	0.402	0.224	0.158	0.620	0.144	0.674
Age = 4	0.154	0.069	0.163	0.637	0.555	
Age = 5	0.881	0.632	0.723	0.752		
Age = 6	0.824	0.194	0.687			
Age = 7+	0.624	0.979				

Notes: Table A10 reports p values where the H_0 assumption is $\beta_c \beta_i^{c,a} = 0$. All the p values are based on the delta method, calculated via *testnl*. All coefficients are estimated in table A9.

Table A11: Productivity and Sales Growth: A Direct Test

	(1) Log Value	(2) Log Quantity	(3) Log Value	(4) Log Quantity	(5) Log Value	(6) Log Quantity	(7) Log Value	(8) Log Quantity
Age								
2	0.005	0.021	0.493***	0.516***	-0.091*	-0.073	0.138***	0.180***
	(0.042)	(0.043)	(0.030)	(0.031)	(0.048)	(0.050)	(0.018)	(0.018)
3	0.067	0.108	0.694***	0.728***	0.155**	0.146*	0.096***	0.107***
	(0.064)	(0.066)	(0.047)	(0.048)	(0.075)	(0.078)	(0.033)	(0.035)
4	0.109	0.210**	0.928***	0.945***	0.092	0.103	-0.147**	-0.110*
	(0.090)	(0.091)	(0.068)	(0.070)	(0.107)	(0.111)	(0.057)	(0.059)
5	0.108	0.195	0.927***	1.045***	-0.080	-0.058	-0.451***	-0.389***
	(0.125)	(0.128)	(0.100)	(0.103)	(0.147)	(0.153)	(0.098)	(0.101)
6	0.299	0.400**	0.910***	0.937***	-0.091	-0.043	-1.087***	-0.966***
	(0.182)	(0.187)	(0.155)	(0.160)	(0.215)	(0.225)	(0.178)	(0.182)
7+	0.115	0.268	0.816***	0.860***	0.025	0.260	-0.967***	-0.760***
	(0.213)	(0.218)	(0.178)	(0.180)	(0.232)	(0.244)	(0.220)	(0.224)
Age interacted with	Log total import value		total # cohorts		Log total export value to all destinations		length of export spell	
2	0.077***	0.077***	0.027***	0.026***	0.058***	0.058***	0.072***	0.063***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.004)	(0.007)	(0.007)
3	0.110***	0.108***	0.052***	0.051***	0.068***	0.071***	0.079***	0.083***
	(0.008)	(0.008)	(0.006)	(0.006)	(0.005)	(0.005)	(0.009)	(0.009)
4	0.135***	0.127***	0.056***	0.059***	0.091***	0.093***	0.103***	0.103***
	(0.011)	(0.011)	(0.008)	(0.008)	(0.007)	(0.008)	(0.012)	(0.012)
5	0.153***	0.148***	0.081***	0.072***	0.116***	0.117***	0.123***	0.118***
	(0.014)	(0.015)	(0.012)	(0.012)	(0.010)	(0.010)	(0.016)	(0.017)
6	0.147***	0.141***	0.097***	0.100***	0.126***	0.126***	0.174***	0.163***
	(0.021)	(0.022)	(0.018)	(0.018)	(0.014)	(0.015)	(0.026)	(0.026)
7+	0.194***	0.184***	0.139***	0.140***	0.138***	0.126***	0.107***	0.087***
	(0.025)	(0.025)	(0.020)	(0.020)	(0.015)	(0.016)	(0.026)	(0.027)
Length of export spell							0.300***	0.302***
							(0.004)	(0.005)
FE	Firm-product (HS6)-year, Product (HS6)-country-year							
N	407,421	405,520	675,680	672,257	675,680	672,257	675,680	672,257
adj. R2	0.673	0.869	0.638	0.825	0.638	0.825	0.651	0.831

Notes: Table A11 reports main coefficients of firm's age, and the interaction between firm's age and market/firm shifters. The samples only include new markets. The observation is at the firm-product(HS6)-country-year level. The estimation equation is

$$LnY_{ijdt} = \beta'_{a1}\mathbf{a}_{ijdt} + \beta'_{a2}\mathbf{a}_{ijdt} \times Shifter + \lambda_{jdt} + \eta_{ijt} + \varepsilon_{ijdt}. \tag{A.2}$$

where the dependent variable LnY_{ijdt} is export value or export quantity in logarithm. $\mathbf{a}_{ijdt}$ is a vector of dummies which characterizes the age of firm i in market jd and year t. $Shifter$ is the proxy for market demand or firm's marginal cost. We topcode the age at 7 years. λ_{jdt} is the HS6 product-country -year fixed effect, and η_{ijt} is the firm-HS6 product-year fixed effect. Standard errors are robust clustered.

Total import value in Column (1)-(2) is measured by total value imported by destination country, and is at the product (HS6)-country-year level. Total number of cohorts in Column (3)-(4) is at the product (HS6)-country level. Total export value to all destinations in Column (5)-(6) is export value to all destination countries (including new markets and old markets) of product j by firm i in year t. The lenght of export spell in Column (7)-(8) is at the firm-product (HS6)-country level. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.

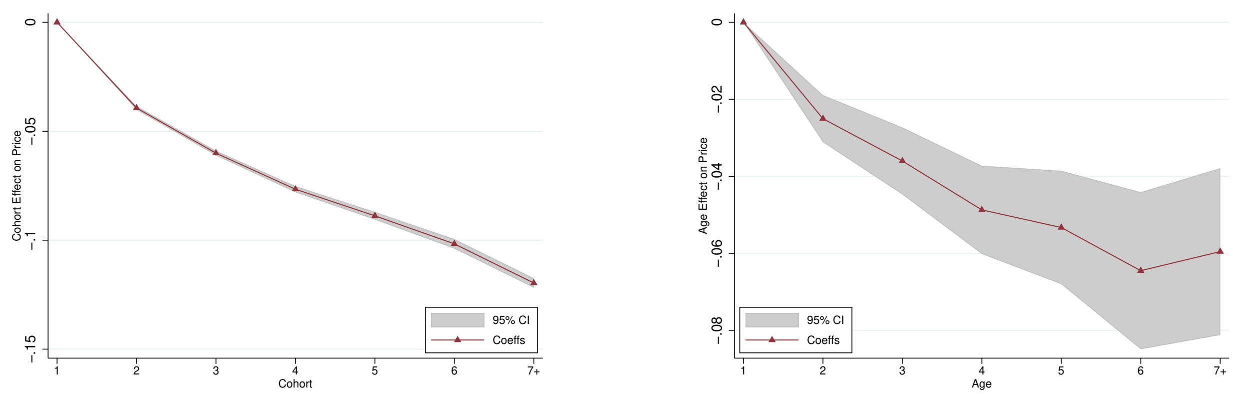
Source: Customs Data 2002-2011, and market size from CEPII.

B.2 Price Dynamics

Table A12: A Summary of Price Dynamics Papers

Paper	Country	Period	Exporter (Domestic)	Firm FE	Market FE	Trend
Berman et al. (2019)	France	1994 - 2005	Exporters	Yes	Yes	Decreasing
Piveteau (2021)	France	1995 - 2010	Exporters	Yes	Yes	Increasing
Fitzgerald et al. (2022)	Ireland	1996 - 2014	Exporters	Yes	Yes	No
Argente et al. (2021)	US	2006 - 2017	Domestic Firms	Yes	Yes	No
Rodrigue and Tan (2019)	China	2000-2006	Exporters	Yes	Yes	Increasing
Bastos et al. (2018)	Portugal	2005 - 2009	Exporters	Yes	Yes	Decreasing
Zhao (2018)	China	2002 - 2011	Exporters	Yes	Yes	Increasing
Foster et al. (2008)	US	1977 - 1997	Domestic Firms	No	No	Increasing

Notes: We tag the existence of firm fixed effect if the regression includes firm dummies, despite whether they are interacted with other variables/dummies. The existence of market fixed effect is tagged if the regression includes market dummies (product-destination country for exporters, or designated market area for domestic firms), despite whether they are interacted with other variables/dummies.



(a) All Markets

(b) New Markets

Figure A4: Age Effect on Price

Notes: Figure A4 shows estimated age effect on price. The observation is at the firm-product(HS6)-country-year level. The estimation equation is

$$LnPrice_{ijdt} = \beta_a' \mathbf{a}_{ijdt} + \lambda_{jdt} + \eta_{ijt} + \varepsilon_{ijdt}. \tag{A.3}$$

The dependent variable $LnPrice_{ijdt}$ is unit price in logarithm. The unit price is the ratio of export value to export quantity. $\mathbf{a}_{ijdt}$ is a vector of dummies which characterizes the age of firm i in market jd and year t. We topcode the age at 7 years. λ_{jdt} is the HS6 product-country -year fixed effect, and η_{ijt} is the firm-HS6 product-year fixed effect. Standard errors are robust clustered. Panel A includes both new markets and old markets, while Panel B includes only new markets.
Source: Customs Data 2002-2011.

Table A13: Robustness Check on Price Dynamics: Compared to Zhao (2018)

	(1) Log Unit Price	(2) Log Export Quantity	(3) Log Export Value	(4) Log Unit Price	(5) Log Export Quantity	(6) Log Export Value
Age						
2	-0.043***	0.699***	0.656***	0.003***	0.301***	0.303***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
3	-0.067***	1.175***	1.109***	0.015***	0.357***	0.372***
	(0.001)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
4	-0.090***	1.558***	1.469***	0.026***	0.370***	0.394***
	(0.001)	(0.004)	(0.004)	(0.001)	(0.002)	(0.002)
5	-0.109***	1.898***	1.792***	0.039***	0.354***	0.391***
	(0.002)	(0.005)	(0.004)	(0.001)	(0.003)	(0.003)
6	-0.130***	2.230***	2.103***	0.050***	0.343***	0.391***
	(0.002)	(0.006)	(0.006)	(0.002)	(0.004)	(0.004)
7	-0.151***	2.570***	2.423***	0.067***	0.328***	0.394***
	(0.003)	(0.008)	(0.008)	(0.002)	(0.005)	(0.005)
8	-0.165***	2.861***	2.699***	0.089***	0.301***	0.388***
	(0.004)	(0.011)	(0.011)	(0.003)	(0.008)	(0.007)
9	-0.186***	3.130***	2.947***	0.107***	0.294***	0.398***
	(0.006)	(0.019)	(0.018)	(0.005)	(0.013)	(0.012)
10	-0.216***	3.492***	3.280***	0.126***	0.304***	0.428***
	(0.010)	(0.035)	(0.033)	(0.008)	(0.024)	(0.023)
Controls	Firm-product (HS6)-year FEs, Product (HS6)-country-year FEs			Firm-product (HS6)-country FEs, average log price and total export quantity at product (HS6)-country-year level		
N	8,089,408	8,089,408	8,102,447	6,735,305	6,735,305	6,744,832
adj. R2	0.885	0.652	0.551	0.926	0.774	0.678

Notes: Table A13 reports main coefficients of firm's age. The sample replicates the one used in Zhao (2018), i.e., only domestic private firms, and includes all markets. The observation is at the firm-product(HS6)-country-year level. Standard errors in parentheses are robust. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively.
Source: Customs Data 2002-2011.

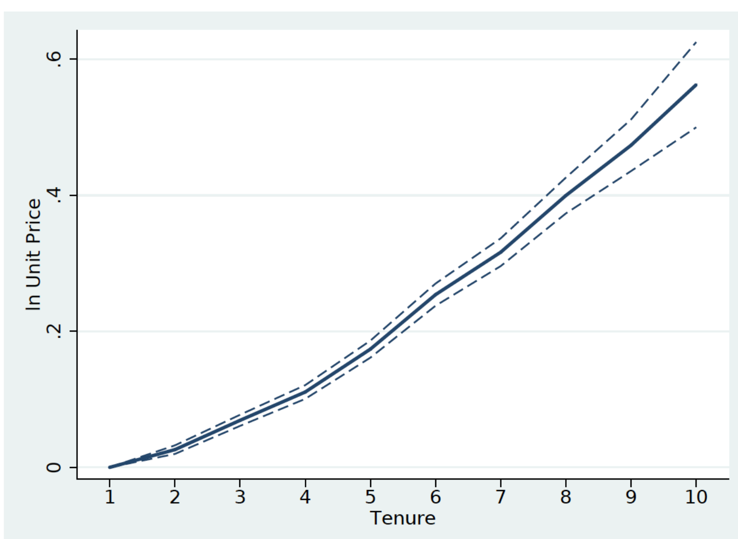


Figure A5: Age Effect on Price: Result of Zhao (2018)

Notes: Figure A5 is from figure 1 of Zhao (2018), which regresses log unit price on firm's tenure in export market, after controlling for firm-product HS6-country fixed effects.

Table A14: Robustness Check on Price Dynamics: Compared to Rodrigue and Tan (2019)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Replication Results						Raw Results from Rodrigue and Tan (2019)		
		Log Value	Log Quantity	Log Price	Log Value	Log Quantity	Log Price	Log Value	Log Quantity	Log Price
Spell length										
	2	0.468 (0.010) ***	0.479 (0.010) ***	-0.010 (0.004) **	0.581 (0.008) ***	0.592 (0.009) ***	-0.011 (0.004) ***	1.99 (0.07)***	1.24 (0.08) ***	0.66 (0.05) ***
	3	0.708 (0.024) ***	0.711 (0.026) ***	-0.002 (0.010)	0.897 (0.019) ***	0.902 (0.020) ***	-0.00 (0.008)	2.27 (0.07) ***	1.69 (0.08) ***	0.70 (0.05) ***
	4	0.853 (0.054) ***	0.890 (0.056) ***	-0.036 (0.022) *	1.044 (0.039) ***	1.044 (0.042) ***	0.000 (0.015)	2.41 (0.07) ***	1.77 (0.08) ***	0.67 (0.05) ***
	5	1.085 (0.126) ***	1.149 (0.129) ***	-0.064 (0.045)	1.149 (0.087) ***	1.151 (0.093) ***	-0.002 (0.036)	2.59 (0.07) ***	1.91 (0.08) ***	0.72 (0.05) ***
	6+	1.114 (0.114) ***	1.301 (0.124) ***	-0.187 (0.043) ***	1.345 (0.067) ***	1.387 (0.074) ***	-0.042 (0.029)	2.60 (0.07) ***	1.91 (0.09) ***	0.71 (0.05) ***
cens		0.143 (0.076) *	0.117 (0.080)	0.026 (0.025)	0.118 (0.056) **	0.100 (0.059) *	0.018 (0.020)	2.85 (0.05) ***	2.00 (0.07) ***	0.98 (0.03) ***
Age		2-year spell								
	2	0.053 (0.013) ***	0.088 (0.013) ***	-0.036 (0.006) ***	0.075 (0.011) ***	0.122 (0.011) ***	-0.048 (0.005) ***	0.03 (0.01) ***	-0.00 (0.01)	0.03 (0.00) ***
Age		3-year spell								
	2	0.533 (0.031) ***	0.579 (0.034) ***	-0.046 (0.013) ***	0.535 (0.025) ***	0.597 (0.027) ***	-0.062 (0.010) ***	0.10 (0.02) ***	0.06 (0.02) ***	0.01 (0.00) **
	3	0.172 (0.032) ***	0.228 (0.035) ***	-0.056 (0.013) ***	0.151 (0.026) ***	0.226 (0.028) ***	-0.075 (0.011) ***	0.36 (0.34) ***	0.34 (0.01) ***	0.04 (0.01) ***
Age		4-year spell								
	2	0.694 (0.069) ***	0.673 (0.072) ***	0.021 (0.027)	0.734 (0.053) ***	0.772 (0.056) ***	-0.037 (0.020) *	0.15 (0.03) ***	0.11 (0.04) ***	0.01 (0.01)
	3	0.690 (0.070) ***	0.724 (0.073) ***	-0.034 (0.027)	0.734 (0.054) ***	0.828 (0.058) ***	-0.094 (0.020) ***	0.54 (0.03) ***	0.52 (0.03) ***	0.03 (0.01) **
	4	0.234 (0.070) ***	0.224 (0.073) ***	0.010 (0.027)	0.223 (0.055) ***	0.284 (0.059) ***	-0.060 (0.021) ***	0.49 (0.02) ***	0.47 (0.02) ***	0.04 (0.01) ***
Age		5-year spell								
	2	0.571 (0.159) ***	0.602 (0.164) ***	-0.032 (0.058)	0.739 (0.114) ***	0.804 (0.121) ***	-0.065 (0.046)	0.31 (0.07) ***	0.29 (0.07) ***	0.01 (0.00) ***
	3	0.757 (0.163) ***	0.713 (0.169) ***	0.043 (0.056)	0.918 (0.118) ***	0.965 (0.127) ***	-0.046 (0.045)	0.78 (0.06) ***	0.77 (0.06) ***	0.04 (0.00) ***
	4	0.518 (0.158) ***	0.477 (0.164) ***	0.040 (0.057)	0.690 (0.119) ***	0.726 (0.128) ***	-0.036 (0.047)	0.87 (0.05) ***	0.87 (0.05) ***	0.07 (0.00) ***
	5	0.034 (0.164)	-0.006 (0.171)	0.040 (0.059)	0.092 (0.122)	0.134 (0.131)	-0.042 (0.048)	0.61 (0.04) ***	0.61 (0.04) ***	0.12 (0.00) ***
Age		6+ years spell								
	2	0.767 (0.137) ***	0.671 (0.147) ***	0.096 (0.050) *	0.721 (0.087) ***	0.744 (0.094) ***	-0.023 (0.035)	0.23 (0.14) *	0.13 (0.14)	0.01 (0.01)
	3	0.972 (0.134) ***	0.826 (0.144) ***	0.146 (0.050) ***	1.008 (0.088) ***	1.020 (0.095) ***	-0.012 (0.035)	0.85 (0.12) ***	0.80 (0.14) ***	0.02 (0.01) ***
	4	1.048 (0.133) ***	0.901 (0.144) ***	0.147 (0.050) ***	1.024 (0.089) ***	1.074 (0.097) ***	-0.051 (0.036)	1.03 (0.11) ***	1.00 (0.12) ***	0.04 (0.01) ***
	5	0.946 (0.137) ***	0.777 (0.147) ***	0.168 (0.052) ***	0.858 (0.093) ***	0.891 (0.100) ***	-0.034 (0.037)	0.95 (0.09) ***	0.93 (0.10) ***	0.07 (0.01) ***
	6+	0.750 (0.176) ***	0.574 (0.189) ***	0.177 (0.066) ***	0.587 (0.130) ***	0.621 (0.140) ***	-0.034 (0.051)	0.63 (0.07) ***	0.62 (0.08) ***	0.13 (0.01) ***
Controls		Firm-product (HS6)-year FEs, Product (HS6)-country-year Fes			Firm-HS6 product-year FEs, Country FEs			Firm-HS6 product-year FEs, Country FEs		
N		859,976	859,976	859,976	1,007,581	1,007,581	1,007,581	1,396,461	1,396,461	1,396,461
adj. R2		0.545	0.605	0.826	0.517	0.599	0.842	0.49	0.49	0.60

Notes: The sample replicates the one used in table 17 of Rodrigue and Tan (2019), i.e., only domestic private non-importing exporters, and includes all markets. The observation is at the firm i-product (HS6) j-country d-year t level. Column (1)-(6) are estimated using the following equation

$$y_{ijdt} = \delta_d + c_{ijt} + \beta' \left(\boldsymbol{a}_{ijdt} \otimes \boldsymbol{s}_{ijdt}\right) + cens_{ijd} + \varepsilon_{ijdt} \tag{A.4}$$

where y_{ijdt} represents dependent variable, i.e, log export value, log export quantity, and log unit prices respectively. $\mathbf{a}_{ijdt}$ is a vector of dummies which characterizes the age of firm i in market jd and year t. $\mathbf{s}_{ijdt}$ is a vector of indicators for the length of the relevant spell. We top-code both firm's age and spell length at 6 years. $cens_{ijd}$ is a separate indicator for spells that are both left- and right-censored. δ_d is country fixed effect, and c_{ijt} is the firm-HS6 product-year fixed effect. Standard errors in parentheses are robust. ***, **, * denotes statistical significance at the 1%, 5%, 10% levels, respectively. Column (7)-(9) are directly from column (1)-(3) of table 17 in Rodrigue and Tan (2019).
Source: Customs Data 2000-2006.

Appendix C Omitted Proofs in Section 3

C.1 Proof of lemma 1

Proof. (4) implies that $\mu_n \xrightarrow{p} \theta$ and $\nu_n \xrightarrow{p} 0$ when $n \to \infty$. $\pi_n(\mu_n)$ then converges in probability to a finite constant. The principle of optimality then applies, so it suffices to consider instead the sequence problem of (8). We define W^m for any pair (μ_n, n) as follows:

$$W^m(\mu_n, n) = \max_{\{X_t^m\}} \mathbb{E}\left[\sum_{t=0}^{m} \beta^t S_t^m \pi_{n+t}(\mu_{n+t}) | \mu_n, n\right],$$

in which X_t^m is the binary exit decision t periods from now and $S_t^m = \Pi_{i=0}^t X_i^m$. It is straightforward that W^m increases in m and is bounded. Hence, $W = \lim_{m\to\infty} W^m$ exists for all μ_n and n. Note from (7) that π_n, or equivalently W^0, increases strictly on μ_n and A and decreases strictly on τ_θ. Moreover, π_n is convex in μ_n. Then, the associated policy function of W^0 is a threshold rule. There exists $\mu_n^{*,0}$ such that $X_0^0 = 1$ if and only if $\mu_n \geqslant \mu_n^{*,0}$. Easily, $\mu_n^{*,0}$ decreases on A and increases on τ_θ.

We prove that W^m have all these properties by induction. Assume that W^{m-1} increases strictly on μ_n and A, decreases strictly on τ_θ, is convex in μ_n, and admits a threshold rule. Furthermore,

$$W^m(\mu_n, n) = \max_{X_0^m} X_0^m \left[\pi_n(\mu_n) + \beta\, \mathbb{E}\left[W^{m-1}(\mu_{n+1}, n+1)|\mu_n, n\right]\right]$$

(4) implies that $\mu_{n+1}|\mu_n$ increases on μ_n in the first order stochastic dominance sense. Therefore, the term in the square bracket increases strictly in μ_n and A, and so does W^m. W^m then admits a exit threshold $\mu_n^{*,m}$ which decreases on A. To see that W^m decreases on τ_θ, consider $\tau_\theta' > \tau_\theta$. Then, $\mu_{n+1}|(\mu_n; \tau_\theta)$ is a mean preserving spread of $\mu_{n+1}|(\mu_n; \tau_\theta')$. Hence,

$$\begin{aligned}\mathbb{E}\left[W^{m-1}(\mu_{n+1}, n+1; \tau_\theta')|\mu_n, n; \tau_\theta'\right] &< \mathbb{E}\left[W^{m-1}(\mu_{n+1}, n+1; \tau_\theta)|\mu_n, n; \tau_\theta'\right] \\ &< \mathbb{E}\left[W^{m-1}(\mu_{n+1}, n+1; \tau_\theta)|\mu_n, n; \tau_\theta\right].\end{aligned}$$

The first inequality comes from that W^{m-1} decreases strictly on τ_θ, and the second from that W^{m-1} is convex in μ_n. Therefore, $\mathbb{E}\left[W^{m-1}(\mu_{n+1}, n+1)|\mu_n, n\right]$ decreases on τ_θ, and so does W^m. It further confirms that the exit threshold $\mu_n^{*,m}$ increases in τ_θ. Lastly, we show that W^m is also convex in μ_n. It suffices to show that $\xi(\mu_n, n) = \mathbb{E}\left[W^{m-1}(\mu_{n+1}, n+1)|\mu_n, n\right]$ is convex in μ_n in that convexity is preserved under maximum. Note that $\mu_{n+1} - \mu_n \sim \mathcal{N}(0, s_{n+1}^2)$ with $s_{n+1}^2 = \frac{\nu_n^2 + \sigma_\epsilon^2}{g_{n+1}^2}$. Then, we can rewrite

$$\begin{aligned}\xi(\mu_n, n) &= \int_{-\infty}^{\infty} W^{m-1}(x, n+1) \frac{1}{s_{n+1}} \phi\left(\frac{x - \mu_n}{s_{n+1}}\right) dx \\ &= \int_{-\infty}^{\infty} W^{m-1}(y + \mu_n, n+1) \frac{1}{s_{n+1}} \phi\left(\frac{y}{s_{n+1}}\right) dy.\end{aligned}$$

By supposition, $W^{m-1}(y + \mu_n, n+1)$ is convex in μ_n for all y, then the integral is also convex in μ_n.

Then, we conclude that the limiting function W has all the desirable properties, and so does V. It is easy to know the strict monotonicity of V in μ_n, A, and τ_θ given the strict monotonicity of π. So V admits a threshold rule, where the threshold decreases on A and increases in τ_θ. The proof is then complete. ∎

C.2 Proof of proposition 1

Proof. Let $\varphi_n(\mu|\theta; A)$ be the density function of μ at age n conditional on permanent demand component θ. Notice that

$$\tilde{q}_n(A) = \int \xi(\theta) \int \mu_n \varphi_n(\mu_n|\theta; A)\, d\mu_n d\theta - \mu_0 + \frac{1}{2\sigma}(\nu_n^2 - \nu_0^2), \tag{A.5}$$

in which ξ is the distribution over θ.Then, the statement holds if $\varphi_n(\cdot|\theta; A)$ stochastically decreases on A for all θ. Explicitly,

$$\varphi_n(\mu_n|\theta; A) = \frac{\int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} f_n(\mu_n, x_{n-1}, \cdots, x_1|\theta)\, dx_1 \cdots dx_{n-1}}{\int_{\mu_n^*}^{\infty} \int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} f_n(x_n, x_{n-1}, \cdots, x_1|\theta)\, dx_1 \cdots dx_{n-1}\, dx_n}. \tag{A.6}$$

We use boldface $\mathbf{x}_n = (x_n, x_{n-1}, \cdots, x_1)^T$ to denote the vector of belief means at age n. Then, $f_n(\mathbf{x}_n|\theta)$ is the density of the joint distribution over $\mathbf{x}_n$ conditional on θ. It is independent of productivity A since temporary demand shocks facing by all firms follow the same data generating process. That makes $\varphi_n(\cdot|\theta; A)$ the marginal distribution on μ_n of the joint distribution $f_n(\mathbf{x}_n|\theta)$ truncated at $\boldsymbol{\mu}_n^*(A)$. Lemma 1 shows that $\mu_n^*(A)$ decreases on A for all n. We show in two steps that $\varphi_n(\mu_n|\theta; A)$ decreases on A in the sense of first order stochastic dominance (FOSD). Let $\underline{\mathbf{x}}_i = (x_n, \cdots x_{n-i+2}, 0, x_{n-i} \cdots, x_1)^T$ and $\mathbf{e}_i = (0, \cdots, 1, \cdots, 0)^T$, in which the $(n+1-i)$-th component is one.

STEP 1: We show that

$$\frac{f_n(b\mathbf{e}_i + \underline{\mathbf{x}}_i|\theta)}{f_n(b'\mathbf{e}_i + \underline{\mathbf{x}}_i|\theta)} = \begin{cases} \exp\left(-\frac{b^2-b'^2}{2\sigma_\epsilon^2 m_n^2} + \frac{b-b'}{\sigma_\epsilon^2}\frac{x_{n-1}}{m_n m_{n-1}} + \frac{(b-b')\theta}{\sigma_\epsilon^2 m_n}\right) & \text{if } i = 1, \\ \exp\left(-\frac{b^2-b'^2}{\sigma_\epsilon^2 m_{n-i+1}^2} + \frac{b-b'}{\sigma_\epsilon^2}\frac{x_{n-i+2}}{m_{n-i+2}m_{n-i+1}} + \frac{b-b'}{\sigma_\epsilon^2}\frac{x_{n-i}}{m_{n-i+1}m_{n-i}}\right) & \text{if } i = 2, \cdots, n-1, \\ \exp\left(-\frac{b^2-b'^2}{2\sigma_\epsilon^2 m_1^2} + \frac{b-b'}{\sigma_\epsilon^2}\frac{x_2}{m_2 m_1}\right) & \text{if } i = n. \end{cases} \tag{A.7}$$

Proof. To see that, note first that $\mathbf{z}_n|\theta \sim \mathcal{N}(\theta\boldsymbol{\iota}_n, \sigma_\varepsilon^2\mathbf{I}_n)$, in which $\boldsymbol{\iota}_n = (1, 1, \cdots, 1)^T$ and $\mathbf{I}_n$ is the identity matrix of dimension n. From (4), we have $\mathbf{x}_n|\theta \sim \mathcal{N}(\theta\mathbf{M}_n\boldsymbol{\iota}_n, \sigma_\varepsilon^2\mathbf{M}_n\mathbf{M}_n^T)$, in which

$$\boldsymbol{\iota}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathbf{M}_n = \begin{bmatrix} m_n & m_n & \cdots & m_n \\ & m_{n-1} & \cdots & m_{n-1} \\ & & \ddots & \vdots \\ & & & m_1 \end{bmatrix},$$

and $m_n = \tau_\epsilon/(\tau_\theta + n\tau_\epsilon)$. Let $\tilde{\boldsymbol{\theta}}_n = \theta\mathbf{M}_n\boldsymbol{\iota}_n$ and $\Sigma_n^{-1} = \frac{1}{\sigma_\epsilon^2}(\mathbf{M}_n^T)^{-1}\mathbf{M}_n^{-1}$. We have

$$\mathbf{M}_n^{-1} = \begin{bmatrix} \frac{1}{m_n} & -\frac{1}{m_{n-1}} & & \\ & \frac{1}{m_{n-1}} & -\frac{1}{m_{n-2}} & \\ & & \ddots & -\frac{1}{m_1} \\ & & & \frac{1}{m_1} \end{bmatrix}, \quad \Sigma_n^{-1} = \frac{1}{\sigma_\epsilon^2}\begin{bmatrix} \frac{1}{m_n^2} & -\frac{1}{m_n m_{n-1}} & & \\ -\frac{1}{m_n m_{n-1}} & \frac{2}{m_{n-1}^2} & -\frac{1}{m_{n-1}m_{n-2}} & \\ & & \ddots & -\frac{1}{m_2 m_1} \\ & & -\frac{1}{m_2 m_1} & \frac{1}{m_1^2} \end{bmatrix}.$$

Then,

$$\frac{f_n(b\mathbf{e}_i + \underline{\mathbf{x}}_i|\theta)}{f_n(b'\mathbf{e}_i + \underline{\mathbf{x}}_i|\theta)} = \exp\left(-\frac{1}{2}\left[(b\mathbf{e}_i + \underline{\mathbf{x}}_i - \tilde{\boldsymbol{\theta}}_n)^T\Sigma_n^{-1}(b\mathbf{e}_i + \underline{\mathbf{x}}_i - \tilde{\boldsymbol{\theta}}_n) - (b'\mathbf{e}_i + \underline{\mathbf{x}}_i - \tilde{\boldsymbol{\theta}}_n)^T\Sigma_n^{-1}(b'\mathbf{e}_i + \underline{\mathbf{x}}_i - \tilde{\boldsymbol{\theta}}_n)\right]\right)$$

Furthermore,

$$\begin{aligned}
&(b\mathbf{e}_i + \underline{\mathbf{x}}_i - \tilde{\boldsymbol{\theta}}_n)^T\Sigma_n^{-1}(b\mathbf{e}_i + \underline{\mathbf{x}}_i - \tilde{\boldsymbol{\theta}}_n) - (b'\mathbf{e}_i + \underline{\mathbf{x}}_i - \tilde{\boldsymbol{\theta}}_n)^T\Sigma_n^{-1}(b'\mathbf{e}_i + \underline{\mathbf{x}}_i - \tilde{\boldsymbol{\theta}}_n) \\
=&(b\mathbf{e}_i + \underline{\mathbf{x}}_i)^T\Sigma_n^{-1}(b\mathbf{e}_i + \underline{\mathbf{x}}_i) - 2\tilde{\boldsymbol{\theta}}_n^T\Sigma_n^{-1}(b\mathbf{e}_i + \underline{\mathbf{x}}_i) - (b'\mathbf{e}_i + \underline{\mathbf{x}}_i)^T\Sigma_n^{-1}(b'\mathbf{e}_i + \underline{\mathbf{x}}_i) + 2\tilde{\boldsymbol{\theta}}_n^T\Sigma_n^{-1}(b'\mathbf{e}_i + \underline{\mathbf{x}}_i) \\
=&(b^2 - b'^2)\mathbf{e}_i^T\Sigma_n^{-1}\mathbf{e}_i + 2(b-b')\mathbf{e}_i^T\Sigma_n^{-1}\underline{\mathbf{x}}_i - 2(b-b')\tilde{\boldsymbol{\theta}}_n^T\Sigma_n^{-1}\mathbf{e}_i.
\end{aligned}$$

We compute these three terms as follows. The first term

$$(b^2 - b'^2)\mathbf{e}_i^T\Sigma_n^{-1}\mathbf{e}_i = \frac{b^2 - b'^2}{\sigma_\epsilon^2}\mathbf{e}_i^T(\mathbf{M}_n^T)^{-1}\mathbf{M}_n^{-1}\mathbf{e}_i = \begin{cases} \frac{b^2-b'^2}{\sigma_\epsilon^2 m_{n-i+1}^2}, & \text{if } i = 1, n, \\ \frac{2(b^2-b'^2)}{\sigma_\epsilon^2 m_{n-i+1}^2}, & \text{if } i = 2, \cdots, n-1. \end{cases}$$

Note that

$$\mathbf{e}_i^T\Sigma_n^{-1} = \begin{cases} \frac{1}{\sigma_\epsilon^2}[\frac{1}{m_n^2}, -\frac{1}{m_n m_{n-1}}, \cdots], & \text{if } i = 1, \\ \frac{1}{\sigma_\epsilon^2}[\cdots, -\underbrace{\frac{1}{m_{n-i+2}m_{n-i+1}}}_{(i-1)\text{-th}}, \underbrace{\frac{2}{m_{n-i+1}^2}}_{i\text{-th}}, -\underbrace{\frac{1}{m_{n-i+1}m_{n-i}}}_{(i+1)\text{-th}}, \cdots], & \text{if } i = 2, \cdots, n-1, \\ \frac{1}{\sigma_\epsilon^2}[\cdots, -\frac{1}{m_2 m_1}, \frac{1}{m_1^2}], & \text{if } i = n. \end{cases}$$

Then, the second term

$$2(b-b')\mathbf{e}_i^T\Sigma_n^{-1}\underline{\mathbf{x}}_i = \begin{cases} -\frac{2(b-b')}{\sigma_\epsilon^2}\frac{x_{n-1}}{m_n m_{n-1}} & \text{if } i = 1, \\ -\frac{2(b-b')}{\sigma_\epsilon^2}\frac{x_{n-i+2}}{m_n m_{n-i+1}} - \frac{2(b-b')}{\sigma_\epsilon^2}\frac{x_{n-i}}{m_{n-i+1}m_{n-i}} & \text{if } i = 2, \cdots, n-1, \\ -\frac{2(b-b')}{\sigma_\epsilon^2}\frac{x_2}{m_2 m_n} & \text{if } i = n. \end{cases}$$

Lastly, the third term

$$\begin{aligned}
-2(b-b')\tilde{\boldsymbol{\theta}}_n^T\Sigma_n^{-1}\mathbf{e}_i &= -2(b-b')\theta\boldsymbol{\iota}^T\mathbf{M}_n^T\Sigma_n^{-1}\mathbf{e}_i \\
&= -2\frac{b-b'}{\sigma_\epsilon^2}\theta\boldsymbol{\iota}_n^T\mathbf{M}_n^T(\mathbf{M}_n^T)^{-1}\mathbf{M}_n^{-1}\mathbf{e}_i \\
&= -2\frac{b-b'}{\sigma_\epsilon^2}\theta\boldsymbol{\iota}_n^T\mathbf{M}_n^{-1}\mathbf{e}_i \\
&= -2\frac{b-b'}{\sigma_\epsilon^2}\theta[\frac{1}{m_n}, \cdots]\mathbf{e}_i \\
&= -2\frac{(b-b')\theta}{\sigma_\epsilon^2 m_n}\mathbb{1}\{i = 1\}
\end{aligned}$$

Putting these three items into the square bracket, we obtain equation (A.7). ■

STEP 2: Let $\boldsymbol{\mu}_i = (\mu_n, \cdots, \mu_{n-i+1})$ with $\boldsymbol{\mu}_0 = \{\}$. Also, let $f_n(x_{n-i}|\boldsymbol{\mu}_i, \theta; A)$ be the marginal density of x_{n-i} conditional on $\boldsymbol{\mu}_i$ and θ with measured productivity A. Then,

$$f_n(x_{n-i}|\boldsymbol{\mu}_i, \theta; A) = \frac{\int_{\mu^*_{n-i-1}}^{\infty} \cdots \int_{\mu^*_1}^{\infty} f_n(\boldsymbol{\mu}_i, x_{n-i}, x_{n-i-1}, \cdots, x_1)\, dx_1 \cdots dx_{n-i-1}}{\int_{\mu^*_{n-i}}^{\infty} \int_{\mu^*_{n-i-1}}^{\infty} \cdots \int_{\mu^*_1}^{\infty} f_n(\boldsymbol{\mu}_i, x_{n-i}, x_{n-i-1}, \cdots, x_1)\, dx_1 \cdots dx_{n-i-1}\, dx_{n-i}},$$

in which μ^*_k is short for $\mu^*_k(A)$. We show that for all $i = 0, \cdots, n-2$, the marginal distribution satisfies the following monotone likelihood ratio property (MLRP). Namely, for $A \geqslant A'$ and $x'_{n-i} \geqslant x_{n-i}$,

$$\frac{f_n(x'_{n-i}|\boldsymbol{\mu}_i, \theta; A)}{f_n(x_{n-i}|\boldsymbol{\mu}_i, \theta; A)} \leqslant \frac{f_n(x'_{n-i}|\boldsymbol{\mu}_i, \theta; A')}{f_n(x_{n-i}|\boldsymbol{\mu}_i, \theta; A')}$$

Proof. Note that when $i = n-1$,

$$f_n(x_1|\boldsymbol{\mu}_{n-1}, \theta; A) = \frac{f_n(\boldsymbol{\mu}_{n-1}, x_1|\theta)}{\int_{\mu^*_1}^{\infty} f_n(\boldsymbol{\mu}_{n-1}, x_1|\theta)\, dx_1}$$

Then, for $x > \mu^*_1$,

$$1 - F_n(x_1|\boldsymbol{\mu}_{n-1}, \theta; A) = \frac{\int_x^{\infty} f_n(\boldsymbol{\mu}_{n-1}, x_1|\theta)}{\int_{\mu^*_1}^{\infty} f_n(\boldsymbol{\mu}_{n-1}, x_1|\theta)\, dx_1}.$$

Since μ^*_1 decreases strictly on A, then $1 - F_n(x_1|\boldsymbol{\mu}_{n-1}, \theta; A)$ decreases strictly on A. Therefore, $f_n(x_1|\boldsymbol{\mu}_{n-1}, \theta; A)$ decreases on A in the sense of first order stochastic dominance (FOSD). Now consider $i = n-2$,

$$f_n(x_2|\boldsymbol{\mu}_{n-2}, \theta; A) = \frac{\int_{\mu^*_1}^{\infty} f_n(\boldsymbol{\mu}_{n-2}, x_2, x_1|\theta)\, dx_1}{\int_{\mu^*_2}^{\infty} \int_{\mu^*_1}^{\infty} f_n(\boldsymbol{\mu}_{n-2}, x_2, x_1|\theta)\, dx_1\, dx_2}.$$

Then,

$$\begin{aligned}
\frac{f_n(x'_2|\boldsymbol{\mu}_{n-2}, \theta; A)}{f_n(x_2|\boldsymbol{\mu}_{n-2}, \theta; A)} &= \frac{\int_{\mu^*_1}^{\infty} f_n(\boldsymbol{\mu}_{n-2}, x'_2, x_1|\theta)\, dx_1}{\int_{\mu^*_1}^{\infty} f_n(\boldsymbol{\mu}_{n-2}, x_2, x_1|\theta)\, dx_1} \\
&= \frac{\int_{\mu^*_1}^{\infty} w(\boldsymbol{\mu}_{n-2}, x_2, x'_2, x_1, \theta) f_n(\boldsymbol{\mu}_{n-2}, x_2, x_1|\theta)\, dx_1}{\int_{\mu^*_1}^{\infty} f_n(\boldsymbol{\mu}_{n-2}, x_2, x_1|\theta)\, dx_1},
\end{aligned}$$

in which by STEP 1,

$$\begin{aligned}
w(\boldsymbol{\mu}_{n-2}, x_2, x'_2, x_1, \theta) &= \frac{f_n(\boldsymbol{\mu}_{n-2}, x'_2, x_1|\theta)}{f_n(\boldsymbol{\mu}_{n-2}, x_2, x_1|\theta)} \\
&= \exp\left(-\frac{(x'_2)^2 - (x_2)'^2}{\sigma^2_\epsilon m^2_2} + \frac{x'_2 - x_2}{\sigma^2_\epsilon} \frac{\mu_3}{m_3 m_2} + \frac{x'_2 - x_2}{\sigma^2_\epsilon} \frac{x_1}{m_2 m_1} \right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{f_n(x_2'|\boldsymbol{\mu}_{n-2},\theta;A)}{f_n(x_2|\boldsymbol{\mu}_{n-2},\theta;A)} \\
&= \exp\left(-\frac{(x_2')^2-(x_2)'^2}{\sigma_\epsilon^2 m_2^2}+\frac{x_2'-x_2}{\sigma_\epsilon^2}\frac{\mu_3}{m_3m_2}\right)\int_{-\infty}^{\infty}\exp\left(\frac{x_2'-x_2}{\sigma_\epsilon^2}\frac{x_1}{m_2m_1}\right)f_n(x_1|\boldsymbol{\mu}_{n-2},x_2,\theta;A)\,dx_1,
\end{aligned}$$

in which we use $f_n(x_1|\boldsymbol{\mu}_{n-2},x_2,\theta;A)=\frac{f_n(\boldsymbol{\mu}_{n-2},x_2,x_1|\theta)}{\int_{\mu_1^*}^{\infty}f_n(\boldsymbol{\mu}_{n-2},x_2,x_1|\theta)\,dx_1}$. Since $x_2'\geqslant x_2$, $\exp(\frac{x_2'-x_2}{\sigma_\epsilon^2}\frac{x_1}{m_2m_1})$ increase on x_1. Given that $f_n(x_1|\boldsymbol{\mu}_{n-2},x_2,\theta;A)$ stochastically decreases on A, we have the ratio decreases on A as well. It is well known that MLRP implies first order stochastic dominance[39], so $1-F_n(x_2'|\boldsymbol{\mu}_{n-2},\theta;A)$ also decreases on A.

Now suppose the statement holds for $i=m+1$ with $m\leqslant n-3$. We show that it also holds for $i=m$. Following exactly the same procedure, we can write

$$\frac{f_n(x_{n-m}'|\boldsymbol{\mu}_m,\theta;A)}{f_n(x_{n-m}|\boldsymbol{\mu}_m,\theta;A)}=\int_{-\infty}^{\infty}w(\boldsymbol{\mu}_m,x_{n-m},x_{n-m}',\mathbf{x}_{n-m-1},\theta)f_n(x_{n-m-1}|\boldsymbol{\mu}_m,x_{n-m},\theta;A)\,dx_{n-m-1}.$$

By STEP 1, we know that for $m>0$,

$$\begin{aligned}
&w(\boldsymbol{\mu}_m,x_{n-m},x_{n-m}',\mathbf{x}_{n-m-1},\theta) \\
&=\exp\left(-\frac{(x_{n-m}')^2-(x_{n-m})^2}{\sigma_\epsilon^2m_{n-m}^2}+\frac{x_{n-m}'-x_{n-m}}{\sigma_\epsilon^2}\frac{\mu_{n-m+1}}{m_{n-m+1}m_{n-m}}+\frac{x_{n-m}'-x_{n-m}}{\sigma_\epsilon^2}\frac{x_{n-m-1}}{m_{n-m}m_{n-m-1}}\right);
\end{aligned}$$

for $m=0$,

$$w(x_n,x_n',\mathbf{x}_{n-1},\theta)=\exp\left(-\frac{(x_n')^2-(x_n)^2}{2\sigma_\epsilon^2m_n^2}+\frac{x_n'-x_n}{\sigma_\epsilon^2}\frac{x_{n-1}}{m_nm_{n-1}}+\frac{(x_n'-x_n)\theta}{\sigma_\epsilon^2m_n}\right).$$

Hence, for $x_{n-m}'\geqslant x_{n-m}$, $w(\boldsymbol{\mu}_m,x_{n-m},x_{n-m}',\mathbf{x}_{n-m-1},\theta)$ is a function of only x_{n-m-1} instead of $\mathbf{x}_{n-m-1}$. It also increases in x_{n-m-1}. Since $f_n(x_{n-m-1}|\boldsymbol{\mu}_m,x_{n-m},\theta;A)$ satisfies the MLRP and then decreases stochastically in A, the integral decreases in A as well. This completes the induction step and the whole proof. ■

STEP 3: It is instant from (A.6) that $\varphi_n(\mu_n|\theta;A)$ is $f_n(x_{n-i}|\boldsymbol{\mu}_i,\theta;A)$ in STEP 2 with $i=0$. Then, $\varphi_n(\mu_n|\theta;A)$ satisfies the above MLRP in A, and $\bar{\Phi}_n(\mu_n|\theta;A)\equiv 1-\Phi_n(\mu_n|\theta;A)$ decreases in A. Using integration by parts, we obtain

$$\int_{\mu_n^*}^{\infty}\varphi_n(x|\theta;A)\,dx=\mu_n^*+\int_{\mu_n^*}^{\infty}\bar{\Phi}_n(x|\theta;A)\,dx,$$

[39]See, for example, https://en.wikipedia.org/wiki/Monotone_likelihood_ratio

in which we use $\lim_{x\to\infty} x\bar{\Phi}_n(x|\theta; A) = 0$. For $A > A'$, we have $\mu_n^*(A) < \mu_n^*(A^*)$. Then,

$$
\begin{aligned}
&\int_{\mu_n^*(A)}^{\infty} \varphi_n(x|\theta; A)\, dx - \int_{\mu_n^*(A')}^{\infty} \varphi_n(x|\theta; A')\, dx \\
&= \mu_n^*(A) - \mu_n^*(A') + \int_{\mu_n^*(A)}^{\mu_n^*(A')} \bar{\Phi}_n(x|\theta; A)\, dx + \int_{\mu_n^*(A')}^{\infty} \left(\bar{\Phi}_n(x|\theta; A) - \bar{\Phi}_n(x|\theta; A')\right)\, dx \\
&< \mu_n^*(A) - \mu_n^*(A') + \mu_n^*(A') - \mu_n^*(A) + \int_{\mu_n^*(A')}^{\infty} \left(\bar{\Phi}_n(x|\theta; A) - \bar{\Phi}_n(x|\theta; A')\right)\, dx \\
&= \int_{\mu_n^*(A')}^{\infty} \left(\bar{\Phi}_n(x|\theta; A) - \bar{\Phi}_n(x|\theta; A')\right)\, dx \leqslant 0
\end{aligned}
$$

The first inequality uses that $\bar{\Phi}_n(x|\theta; A) < 1$, and the second inequality comes from that $\bar{\Phi}_n(\mu_n|\theta; A)$ is decreasing in A. Therefore, the mean is strictly decreasing. The proof is then complete.

■

Remark. Since the variance on μ_n exists, we can obtain $\lim_{x\to\infty} x\bar{\Phi}_n(x|\theta; A) = 0$ from the Chebyshev's inequality. An alternative way is to show a version of Mill's inequality with truncated multivariate normal distribution. Here we suppress θ, A and n for notational brevity. From (A.6), it is easy to show that

$$
\bar{\Phi}(\mu) = B \int_{\mu}^{\infty} \int_{\boldsymbol{\mu}_{n-1}^*} \exp\left(-\frac{1}{2}\left[A(x_n - p(\boldsymbol{x}_{n-1}))^2 + q(\mathbf{x}_{n-1})\right]\right)\, d\mathbf{x}_{n-1}\, dx_n.
$$

A, a diagonal element, is positive since the covariance matrix is positive definite. B is a constant, and p and q are functions of $\mathbf{x}_{n-1}$. Note that

$$
\begin{aligned}
\mu\bar{\Phi}(\mu) &< B \int_{\boldsymbol{\mu}_{n-1}^*} \int_{\mu}^{\infty} x \exp\left(-\frac{1}{2}\left[A(x_n - p(\boldsymbol{x}_{n-1}))^2 + q(\mathbf{x}_{n-1})\right]\right)\, dx_n\, d\mathbf{x}_{n-1} \\
&= B \int_{\boldsymbol{\mu}_{n-1}^*} \int_{\mu}^{\infty} (x_n - p(\mathbf{x}_{n-1})) \exp\left(-\frac{1}{2}\left[A(x_n - p(\boldsymbol{x}_{n-1}))^2 + q(\mathbf{x}_{n-1})\right]\right)\, dx_n\, d\mathbf{x}_{n-1} \\
&\quad + B \int_{\boldsymbol{\mu}_{n-1}^*} \int_{\mu}^{\infty} p(\mathbf{x}_{n-1}) \exp\left(-\frac{1}{2}\left[A(x_n - p(\boldsymbol{x}_{n-1}))^2 + q(\mathbf{x}_{n-1})\right]\right)\, dx_n\, d\mathbf{x}_{n-1} \\
&= \frac{B}{A} \int_{\boldsymbol{\mu}_{n-1}^*} \exp\left(-\frac{1}{2}\left[A(\mu - p(\boldsymbol{x}_{n-1}))^2 + q(\mathbf{x}_{n-1})\right]\right)\, d\mathbf{x}_{n-1} \\
&\quad + B \int_{\boldsymbol{\mu}_{n-1}^*} p(\mathbf{x}_{n-1}) \exp(-\frac{1}{2} q(\mathbf{x}_{n-1})) \int_{\mu}^{\infty} \exp\left(-\frac{1}{2}\left[A(x_n - p(\boldsymbol{x}_{n-1}))^2)\right]\right)\, dx_n\, d\mathbf{x}_{n-1}
\end{aligned}
$$

When $\mu \to \infty$,

$$
\begin{aligned}
\exp\left(-\frac{1}{2}\left[A(\mu - p(\boldsymbol{x}_{n-1}))^2 + q(\mathbf{x}_{n-1})\right]\right) &\to 0, \\
\int_{\mu}^{\infty} \exp\left(-\frac{1}{2}\left[A(x_n - p(\boldsymbol{x}_{n-1}))^2)\right]\right)\, dx_n &\to 0.
\end{aligned}
$$

Thus, both integrals approach to zero.

C.3 Proof of proposition 2

Proof. We first show that $d^*(d)$ increases on d while $d^*(d)/d$ decreases on d. Note that standard dynamic programming arguments apply to show the strict concavity of v if the instantaneous return function is strictly concave. It then suffices to have $f(d', d) = -\phi(d' - (1-\delta)d)^2 d^{-1}$ concave. This can be easily verified, and we have shown a more general version of it in the proof of proposition D.3. On the other hand, the unconstrained first order condition on d' is

$$1 + 2\phi\left(\frac{d'}{d} - (1-\delta)\right) = \beta v'(d').$$

Since an increase in d lowers the LHS, the RHS has to decrease accordingly. The new optimal d' is then strictly higher. A smaller RHS in turn implies a smaller ratio d'/d. The statements hold trivially when irreversibility constraint binds.

Now we consider $d'_0 > d_0$. Applying the above results, we obtain instantly that $d'_n > d_n$, and $d'_n/d'_{n-1} < d_n/d_{n-1}$ for all n. Therefore,

$$g_n(d_0) = \left(\frac{d_n}{d_0}\right)^\alpha = \left(\prod_{i=0}^{n-1}\frac{d_{i+1}}{d_i}\right)^\alpha > \left(\prod_{i=0}^{n-1}\frac{d'_{i+1}}{d'_i}\right)^\alpha = g_n(d'_0).$$

The proof is then complete. ■

Appendix D Additional Results on Model Selections

D.1 Selection Growth

Proposition A1. *The selection term increases over ages while the variance term decreases.*

Proof. Let $\varphi_n(\mu_n)$ be the unconditional density of μ_n at age n, i.e., $\varphi_n(\mu_n) = \int \varphi_n(\mu_n|\theta)\xi(\theta)\,d\theta$. Then we can suppress A and rewrite (A.5) as follows:

$$\tilde{q}_n = \underbrace{\int \mu_n \varphi_n(\mu_n)\,d\mu_n - \mu_0}_{\text{Selection}} + \underbrace{\frac{1}{2\sigma}(\nu_n^2 - \nu_0^2)}_{\text{Variance}}$$

The second part is straightforward since ν_n^2 decreases on n. To see the first part, we notice that

$$\varphi_n(\mu_n) = \frac{\int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} f_n(\mu_n, x_{n-1}, \cdots, x_1)\,dx_1 \cdots dx_{n-1}}{\int_{\mu_n^*}^{\infty} \int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} f_n(x_n, x_{n-1}, \cdots, x_1)\,dx_1 \cdots dx_{n-1}\,dx_n}.$$

Then, the unconditional density of μ_n, $\tilde{\varphi}_n(\mu_n)$, can be defined as

$$\tilde{\varphi}_n(\mu_n) = \frac{\int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} f_n(\mu_n, x_{n-1}, \cdots, x_1)\,dx_1 \cdots dx_{n-1}}{\int_{-\infty}^{\infty} \int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} f_n(x_n, x_{n-1}, \cdots, x_1)\,dx_1 \cdots dx_{n-1}\,dx_n}.$$

Let $\tilde{\Phi}$ denote its cdf, then

$$\varphi_n(\mu_n) = \begin{cases} \frac{\tilde{\varphi}_n(\mu_n)}{1-\Phi(\mu_n^*)} & \text{if } \mu_n \geqslant \mu_n^*, \\ 0 & \text{if } \mu_n < \mu_n^*. \end{cases}$$

That φ_n is a truncated distribution of $\tilde{\varphi}_n$ implies $\varphi_n \geqslant \tilde{\varphi}_n$ in the sense of FOSD. On the other hand,

$$\begin{aligned}
\int_{-\infty}^{\infty} \mu_n \tilde{\varphi}_n(\mu_n)\,d\mu_n &= \frac{\int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} \int_{-\infty}^{\infty} \mu_n f_n(\mu_n, \mathbf{x}_{n-1})\,d\mu_n\,dx_1 \cdots dx_{n-1}}{\int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} \int_{-\infty}^{\infty} f_n(x_n, \mathbf{x}_{n-1})\,dx_n\,dx_1 \cdots dx_{n-1}} \\
&= \frac{\int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} \int_{-\infty}^{\infty} \mu_n f_n(\mu_n|\mathbf{x}_{n-1})\,d\mu_n f_{n-1}(\mathbf{x_{n-1}})\,dx_1 \cdots dx_{n-1}}{\int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} \int_{-\infty}^{\infty} f_n(x_n|\mathbf{x}_{n-1})\,dx_n f_{n-1}(\mathbf{x}_{n-1})\,dx_1 \cdots dx_{n-1}} \\
&= \frac{\int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} x_{n-1} f_{n-1}(\mathbf{x_{n-1}})\,dx_1 \cdots dx_{n-1}}{\int_{\mu_{n-1}^*}^{\infty} \cdots \int_{\mu_1^*}^{\infty} f_{n-1}(\mathbf{x}_{n-1})\,dx_1 \cdots dx_{n-1}} \\
&= \int \mu_{n-1} \varphi_n(\mu_{n-1})\,d\mu_{n-1}.
\end{aligned}$$

We obtain the third equation by $\mathbb{E}[\mu_n|\mathbf{x}_{n-1}] = x_{n-1}$, which is instantly verified from the updating formula $\mu_n - \mu_{n-1} = \frac{z_{n-1}-\mu_{n-1}}{g_{n+1}}$. Hence, we obtain the first part, i.e.,

$$\int \mu_n \varphi_n(\mu_n)\,d\mu_n > \int \mu_{n-1} \varphi_n(\mu_{n-1})\,d\mu_{n-1}.$$

■

Lemma A1. *Let $f(s,\mu) = s\lambda\left(\frac{\mu}{s}\right)$, in which λ is the hazard function of a standard normal distribution. Then, f increases on s and μ.*

Proof. Notice that $\lambda(x) = \phi(x)/(1-\Phi(x))$, in which ϕ and Φ are the respective pdf and cdf of a standard normal distribution. Then, $\lambda'(x) = \lambda(x)\left(\lambda(x) - x\right)$. Lemma 2 in Chapter VII of Feller (1957, Vol I) shows $x < \lambda(x) < \frac{x^3}{x^2-1}$. Then, $\lambda(x)$ increases in x. Mogens Fosgerau shows a tighter upper bound $x + \frac{1}{x}$.[40]

Therefore,

$$\begin{aligned}
\frac{\partial f(s,\mu)}{\partial s} &= \lambda\left(\frac{\mu}{s}\right) - \frac{\mu}{s}\lambda'\left(\frac{\mu}{s}\right) = \lambda\left(\frac{\mu}{s}\right)\left[1 - \frac{\mu}{s}\left(\lambda\left(\frac{\mu}{s}\right) - \frac{\mu}{s}\right)\right] > 0, \\
\frac{\partial f(s,\mu)}{\partial \mu} &= \lambda'\left(\frac{\mu}{s}\right) > 0.
\end{aligned}$$

■

D.2 A demand learning model with active learning

We extend our model in section 3.2 with active learning. Namely, in each period, firms can do market research to obtain additional signals about their permanent demand components. Let $S_0 = \tau_\theta/\tau_\epsilon$, and it without loss of generality to assume $\tau_\epsilon = 1$. Then, a pair (μ, S) depicts a firm's prior on θ, i.e., $\theta|(\mu, S) \sim \mathcal{N}(\mu, \frac{1}{S})$. It is easy to verify that the initial prior is given by $(0, S_0)$. With the new notations, the static profit function in (7) can be written as follows:

$$\pi(\mu, S) = Ab(\mu, S)^\sigma - F = A\exp\left(\mu + \frac{1}{2\sigma}\left(\frac{1}{S} + 1\right)\right) - F.$$

With active learning, firms can pay for k additional signals at a cost $c(k, S)$, and $c_1(k, S) > 0$. Each additional signal will be the same as the one obtained from price realization, that is, $z^a = \theta + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, 1)$. Information acquisition takes place after the quantity choice. Therefore, at the end of a period, firms will have $k + 1$ signals if they have paid for k additional signals. As before, we rewrite the Bellman equation (8) as

$$\begin{aligned}
V(\mu, S) &= \max\{\max_{k\geqslant 0} \pi(\mu, S) - c(k, S) + \beta\,\mathbb{E}[V(\mu', S')|\mu, S'], 0\}, \\
\text{s.t. } \mu' &= \mu + \frac{(k+1)(\bar{z}_k - \mu)}{S + k + 1}, \quad S' = S + k + 1.
\end{aligned}$$

Let k^* be the associated policy function of active learning. We show in the following proposition that $k^*(\mu, S) = 0$.

Proposition A2. *Firms do not acquire any additional signals, i.e., $k^*(\mu, S) = 0$.*

Proof. As before, we define $\xi(\mu, S') \equiv \mathbb{E}[V(\mu', S')|\mu, S']$. The marginal benefit of an additional signal is

$$\frac{d}{dk}\left\{-c(k, S) + \beta\xi(\mu, S + k + 1)\right\} = -c_1(k, S) + \beta\frac{d\xi(\mu, S + k + 1)}{dk}.$$

[40]See https://math.stackexchange.com/questions/1326879/limit-of-normal-hazard-rate.

Let $\nu^2 = 1/S'$, then

$$\xi(\mu, S') = \int_{-\infty}^{\infty} V(x, S') \frac{1}{\nu} \phi\left(\frac{x-\mu}{\nu}\right) dx = \int_{-\infty}^{\infty} V(\nu x + \mu, S')\phi(x)\, dx,$$

in which ϕ is the pdf of a standard normal distribution. Hence,

$$\begin{aligned} \frac{d\xi(\mu, S+k+1)}{dk} &= \int_{-\infty}^{\infty} \left[V_1(\nu x + \mu, S') x \frac{d\nu}{dk} + V_2(\nu x + \mu, S') \frac{dS'}{dk} \right] \phi(x)\, dx. \\ &= -\frac{1}{2\sqrt{S'}} \int_{-\infty}^{\infty} V_1(\nu x + \mu, S') x\phi(x) dx + \int_{-\infty}^{\infty} V_2(\nu x + \mu, S')\phi(x)\, dx \end{aligned}$$

Following the same argument in the proof of lemma 1, we can show that $V_1 > 0$, $V_2 < 0$, and V_1 is increasing.[41] Notice that

$$\begin{aligned} \int_{-\infty}^{\infty} V_1(\nu x + \mu, S') x\phi(x) dx &= \int_{0}^{\infty} V_1(\nu x + \mu, S') x\phi(x) dx + \int_{-\infty}^{0} V_1(\nu x + \mu, S') x\phi(x) dx \\ &= \int_{0}^{\infty} \left[V_1(\nu x + \mu, S') - V_1(-\nu x + \mu, S') \right] x\phi(x) dx > 0. \end{aligned}$$

The second equality uses the substitution $t = -x$ and that $\phi(x) = \phi(-x)$. In sum, $\frac{d\xi(\mu, S+k+1)}{dk} < 0$. Then, the marginal benefit of an additional signal is negative. The optimal level of active learning is zero.

■

D.3 A customer base accumulation model with strategic pricing

We introduce an alternative customer base accumulation model in which firms expand their customer base through strategic pricing. Our model follows closely to the Customer Markets model in Fitzgerald et al. (2022). In the end, we show that similar to the advertising model, this strategic pricing model also puts no restriction on the relationship between sales and its growth rate. Likewise, we write this model in the style of a deterministic growth model for better illustration. That is, we assume away any random shocks and fixed cost. Exit is exogenous and subsumed in the discount factor.

Firms are characterized by their (constant) marginal cost C and existing customer base D. A larger customer base moves a firm's residual demand curve outward and leads to more revenue. Firms compete monopolistically in the market and accumulate customer base by current sales. Namely, a larger current sales contributes to a larger customer base next period. Therefore, firms trade off between current and future profits. In each period, they set prices to maximize the present value of profit flow V. The following Bellman equation characterizes firms' decision problem.

$$\begin{aligned} V(D, C) = \max_{P \geqslant 0} PQ - CQ + \beta V(D', C) \\ \text{s.t. } Q = D^{\alpha} P^{-\sigma}, \quad D' = (1-\delta)D + PQ, \end{aligned}$$

in which $\sigma > 1$, and $\alpha, \delta \in (0, 1)$. All notations are the same as that in the advertising model. We

[41] The induction step to show V decreases on S relies on the result $k^* = 0$. To deal with that, one can follow the same argument here to include $k^* = 0$ in the induction.

could rewrite this Bellman equation as follows:

$$V(D,C)=\max_{D'\geqslant(1-\delta)D} D'-(1-\delta)D-\left(\frac{D'-(1-\delta)D}{D^{\alpha}}\right)^{-\frac{\sigma}{1-\sigma}} D^{\alpha}C+\beta V(D',C).$$

Now consider $\tilde{D}=C^{\frac{1-\sigma}{1-\alpha}}$, and let $d=D/\tilde{D}$, $v(d,C)=V(D,C)/\tilde{D}\equiv v(d)$. To see this, $v(d)$ is given as follows

$$v(d)=\max_{d'\geqslant(1-\delta)d} d'-(1-\delta)d-\left(d'-(1-\delta)d\right)^{-\frac{\sigma}{1-\sigma}} d^{\frac{\alpha}{1-\sigma}}+\beta v(d'). \tag{A.8}$$

Let d^* be the associated policy function. Since the marginal cost is constant, the sales growth rate of a firm with normalized customer base d_0 is given by

$$\frac{P'Q'}{PQ}=\frac{D''-(1-\delta)D'}{D'-(1-\delta)D}=\frac{d^*(d^*(d_0))-(1-\delta)d^*(d_0)}{d^*(d_0)-(1-\delta)d_0}.$$

Then, firms with the same normalized customer base have the same growth rates in sales. However, they can differ in actual sales. Given the same normalized customer base, firms make more sales if they face smaller marginal costs. Hence, there is no necessary relationship between firms' sales and their growth rates. We can replicate the same strategy used in the advertising model. That is, we can increase initial customer base D_0 and decrease marginal cost C to keep the normalized customer base constant. This ensures that different cohorts have the same lifecycles up to initial sales (or quantity).

For completeness, we prove the following proposition and contrast it with proposition 2.

Proposition A3. *v defined in equation (A.8) is strictly concave. Then, $d^*(d)$ increases on d and $d^*(d)/d$ decreases on d. Moreover, markup P/C increases on d.*

Proof. It suffices to show that $f(d',d)=-\left(d'-(1-\delta)d\right)^{-\frac{\sigma}{1-\sigma}} d^{\frac{\alpha}{1-\sigma}}$ is strictly concave. We consider first that $g(x,y)=-x^a y^b$, in which $a=\frac{\sigma}{\sigma-1}>1$ and $b=\frac{\alpha}{1-\sigma}<0$. Note that for $x,y>0$,

$$f_{11}=-a(a-1)x^{a-2}y^b<0,\quad f_{12}=f_{21}=-abx^{a-1}y^{b-1}>0,\quad f_{22}=-b(b-1)x^a y^{b-2}<0,$$

and

$$f_{11}f_{22}-f_{12}^2=ab(-b-a+1)x^{2a-2}y^{2b-2}=ab\frac{\alpha-1}{\sigma-1}x^{2a-2}y^{2b-2}>0.$$

The last inequality holds because $a>0$, $b<0$, and $\alpha<1$. Then, $g(x,y)$ is strictly concave. Note that $f(d',d)=g(d'-(1-\delta)d,d)$, then for any (d_1',d_1), $(d_2'.d_2)$, and $\theta\in(0,1)$,

$$\begin{aligned}
&f(\theta d_1'+(1-\theta)d_2',\theta d_1+(1-\theta)d_2)\\
=&g(\theta d_1'+(1-\theta)d_2'-(1-\delta)(\theta d_1+(1-\theta)d_2),\theta d_1+(1-\theta)d_2)\\
=&g(\theta(d_1'-(1-\delta)d_1)+(1-\theta)(d_2'-(1-\delta)d_2),\theta d_1+(1-\theta)d_2)\\
>&\theta g(d_1'-(1-\delta)d_1,d_1)+(1-\theta)g(d_2'-(1-\delta)d_2,d_2)\\
=&\theta f(d_1',d_1)+(1-\theta)f(d_2',d_2).
\end{aligned}$$

This completes the first part of the proposition. To see the second part, the first order condition gives

$$1+\beta v'(d') = \frac{\sigma}{\sigma-1} d^{\frac{-\alpha}{\sigma-1}} \left(d'-(1-\delta)d\right)^{\frac{1}{\sigma-1}}.$$

An increase in d reduces the right hand side. Then, d' must increase as v' is a decreasing function. Hence, $d^*(d)$ increases on d. On the other hand, notice that

$$1+\beta v'(d') = \frac{\sigma}{\sigma-1} d^{\frac{1-\alpha}{\sigma-1}} \left(\frac{d'}{d}-(1-\delta)\right)^{\frac{1}{\sigma-1}}.$$

An increase in d decreases the LHS, so d'/d must also increase.

As for the markup, the first order condition to the Bellman equation before normalization is

$$Q+P\frac{dQ}{dP} - C\frac{dQ}{dP} + \beta V_1(D',C)\frac{dPQ}{P} = 0 \Longrightarrow \frac{P}{C} = \frac{\sigma}{\sigma-1}\frac{1}{1+\beta V_1(D',C)} = \frac{\sigma}{\sigma-1}\frac{1}{1+\beta v'(d')}.$$

Since d' increases on d and v is concave, then P/C increases on d. This completes the proof.

■

In both models of customer base accumulation, the growth rate in customer base decreases on the normalized customer base. Therefore, an increase in the measured productivity increases the growth in customer base. The difference falls on the implementation of the growth. With advertising, proposition 2 implies that more productive firms will spend more on advertising. Yet with strategic pricing, proposition shows that more productive firms will charge a lower markup (a lower d).

Whereas the growth rate in sales also decreases on the customer base in the advertising model, it does not necessarily hold in the customer market model. This is because the markup growth might decrease on the customer base, which reduces the sales loss from moving upward along the demand curve and increases the total sales growth. Formally,

$$\frac{P'Q'}{PQ} = \frac{D''-(1-\delta)D'}{D'-(1-\delta)D} = \underbrace{\left(\frac{D''}{D'}-(1-\delta)\right)}_{\frac{D''}{D'}\downarrow\Rightarrow\downarrow} \underbrace{\left(1+\frac{1-\delta}{\frac{D'}{D}-(1-\delta)}\right)}_{\frac{D'}{D}\downarrow\Rightarrow\uparrow}.$$

The first item of the RHS decreases while the second increases, so the overall effect is ambiguous. Same argument can be applied to quantity growth.

D.4 Selection and Investment mechanisms

We consider a firm's decision problem which generates dynamics in its sales and market participation. Time is discrete and denoted by t. $\mathbf{z}_t \equiv (z_t^i)$ is a vector that represents all exogenous states, and $\mathbf{s}_t \equiv (s_t^j)$ is a vector that represents all endogenous states. We use $\mathbf{x}_t = (\mathbf{s}_t, \mathbf{z}_t)$ to denote the joint state. Assume constant discount factor β and death shock γ, the following Bellman equation summarizes the maximization problem of firm's present value:

$$v(\mathbf{s}_t,\mathbf{z}_t) = \max\left\{\max_{\mathbf{s}_{t+1}\in\Gamma(\mathbf{s}_t,\mathbf{z}_t)} \pi(\mathbf{s}_t,\mathbf{s}_{t+1},\mathbf{z}_t) + \beta\gamma\,\mathbb{E}_{\mathbf{z}}\left[v(\mathbf{s}_{t+1},\mathbf{z}_{t+1})|\mathbf{s}_t,\mathbf{z}_t\right], u\right\}, \tag{A.9}$$

in which π is the instantenous profit function and u the value of exit. Each firm chooses next period endogenous state $\mathbf{s}_{t+1}$ and decides on whether to exit. The law of motion on $\mathbf{z}_t$ is governed by the equation

$$\mathbf{z}_{t+1} = g(\mathbf{z}_t, \boldsymbol{\varepsilon}_t),$$

where $\boldsymbol{\varepsilon}_t$ are some random shocks independent over time. It is without loss to have g increase on all of its arguments. We adopt the standard assumption that a firm's exit decision follows a threshold rule. Denoting by C the set of all states $\mathbf{x}$ in which firm chooses to continue its operation, a threshold rule indicates that if $\mathbf{x}' \geqslant \mathbf{x}$ and $\mathbf{x} \in C$, then $\mathbf{x}' \in C$. Lastly, let q be a firm size measure in log, and $q(\mathbf{x})$ increases in $\mathbf{x}$. Then, $h(\mathbf{x}_t, \mathbf{x}_{t+1}) = q(\mathbf{x}_{t+1}) - q(\mathbf{x}_t)$ is our measure of realized firm size growth.

We consider mechanisms as distinct elements in a partition of model ingredients. In other words, a mechanism is a special case of the model with all other mechanisms shut down. We consider selection and investment mechanisms in particular and argue that they make distinct predictions on the relationship between size and growth.

Selection We refer selection growth to the increase in average sales due to exit of less capable firms. Hence, we shut down the changes in endogenous state $\mathbf{s}_t$, and $\mathbf{x}_t = \mathbf{z}_t$. Furthermore, we consider ε to be a scalar.[42] Then we could rewrite h as $h(\mathbf{z}_t, \varepsilon_t) = q(g(\mathbf{z}_t, \varepsilon_t)) - q(\mathbf{z}_t)$. Let $\tilde{h}(\mathbf{z}_t)$ be the expected growth conditional on current size and survival.

Proposition A4. *Suppose $h(\cdot, \varepsilon)$ is decreasing on $\mathbf{z}$ for all ε. Then, $\tilde{h}(\mathbf{z})$ also decreases on $\mathbf{z}$.*

Proof. Note that $\tilde{h}(\mathbf{z}) = \mathbb{E}[h(\mathbf{z}, \varepsilon) | g(\mathbf{z}, \varepsilon) \in C]$. Consider $\mathbf{z} \geqslant \mathbf{z}'$, then

$$\mathbb{E}[h(\mathbf{z}, \varepsilon) | g(\mathbf{z}, \varepsilon) \in C] \leqslant \mathbb{E}[h(\mathbf{z}', \varepsilon) | g(\mathbf{z}, \varepsilon) \in C] \leqslant \mathbb{E}[h(\mathbf{z}', \varepsilon) | g(\mathbf{z}', \varepsilon) \in C].$$

The first inequality comes from the supposition that $h(\cdot, \boldsymbol{\varepsilon})$ is decreasing. To see the second inequality, let f denote the pdf of ε, and let $e(\mathbf{z}) = \inf\{\varepsilon : g(\mathbf{z}, \varepsilon) \in C\}$. Furthermore, let

$$H(\mathbf{z}', \mathbf{z}) \equiv E[h(\mathbf{z}', \boldsymbol{\varepsilon}) | g(\mathbf{z}, \boldsymbol{\varepsilon}) \in C] = \frac{\int_{e(\mathbf{z})}^{\infty} h(\mathbf{z}', \varepsilon) f(\varepsilon) \, d\varepsilon}{\int_{e(\mathbf{z})}^{\infty} f(\varepsilon) \, d\varepsilon}$$

Since $h(\mathbf{z}', \varepsilon)$ increases on ε and $e(\mathbf{z})$ decreases on $\mathbf{z}$, $H(\mathbf{z}', \cdot)$ is decreasing for all $\mathbf{z}'$.

■

Proposition A4 is a statement on how selection growth decreases on high types. The assumption of a decreasing $h(\cdot, \varepsilon)$ is widespread in economics. In productivity shock models à la Hopenhayn (1992) or Luttmer (2007), z_t follows a AR(1) process with $\rho \leqslant 1$, i.e., $z_t = \alpha + \rho z_{t-1} + \varepsilon_t$. In learning models à la Jovanovic (1982), z_t has the form that $z_t = \alpha_t + \rho_t z_{t-1} + \varepsilon_t$. For example, in our case, (4) implies that $\rho_t = 1 - 1/g_{t+1}$ and $\alpha_t = \theta / g_{t+1}$.

Investment To avoid the interference of selections, we shut down dynamics in exogenous state $\mathbf{z}_t$ so that $\mathbf{z}$ are equivalent to constant firm heterogeneities. Bellman equation (A.9) can be reformulated as follows.

$$v(\mathbf{s}_t, \mathbf{z}) = \max \left\{ \max_{\mathbf{s}_{t+1} \in \Gamma(\mathbf{s}_t, \mathbf{z})} \pi(\mathbf{s}_t, \mathbf{s}_{t+1}, \mathbf{z}) + \beta \gamma v(\mathbf{s}_{t+1}, \mathbf{z}), u \right\}. \tag{A.10}$$

[42] The following result does not necessarily hold for multivariate $\boldsymbol{\varepsilon}$. This is because a left truncation of multivariate distribution needs not improve its mean.

Assume in addition that $\pi(\mathbf{s}_t, \mathbf{s}_{t+1}, \mathbf{z})$ is supermodular in $(\mathbf{s}_{t+1}, \mathbf{z})$ and correspondence $\Gamma(\mathbf{s}, \cdot)$ is increasing for all $\mathbf{s}$.[43] Denote by $s^*(\mathbf{s}_t, \mathbf{z})$ the associated policy. Correspondingly, h takes the form $h(\mathbf{s}_t, \mathbf{z}) = q(s^*(\mathbf{s}_t, \mathbf{z}), \mathbf{z}) - q(\mathbf{s}_t, \mathbf{z})$.

Proposition A5. *Suppose v is supermodular in $(\mathbf{z}, \mathbf{s})$ and s^* is singled-valued. If $q(\mathbf{s}, \mathbf{z})$ is separable in $\mathbf{z}$ and $\mathbf{s}$, i.e., $q(\mathbf{s}, \mathbf{z}) = q_s(\mathbf{s}) + q_z(\mathbf{z})$, $h(\mathbf{s}_t, \mathbf{z})$ then increases on $\mathbf{z}$.*

Proof. It is easy to verify that the RHS of equation (A.10) is supermodular in $(\mathbf{s}_{t+1}, \mathbf{z})$, so a direct application of Topkis' monotonicity theorem implies that $s^*(\mathbf{s}_t, \mathbf{z}) \geqslant s^*(\mathbf{s}_t, \mathbf{z}')$ if $\mathbf{z} \geqslant \mathbf{z}'$.[44] With the separability of q, $h(\mathbf{s}_t, \mathbf{z}) = q_s(s^*(\mathbf{s}_t, \mathbf{z})) - q_s(\mathbf{s}_t)$. Then, $h(\mathbf{s}_t, \mathbf{z}) \geqslant h(\mathbf{s}_t, \mathbf{z}')$ for $\mathbf{z} \geqslant \mathbf{z}'$ in that q_s increases on all of its arguments. ■

It needs no introduction on the prevalence of supermodularity in economic models. Both Neoclassical growth models and investment models with adjustment cost are glorified examples. The log separability of firm heterogeneity is also a common feature of many production functions. Then, proposition A5 is a summary of how investment growth increases with types. Since supermodularity implies that the marginal return of capital increases with types, high type firms are more incentivized to expand their capital stock and grow faster.

[43] As both $\mathbf{s}$ and $\mathbf{z}$ are in the Euclidian space, we adopt the standard definitions of join and meet, i.e., $\mathbf{x} \vee \mathbf{x}' = (\max\{x_i, x_i'\})_i$ and $\mathbf{x} \wedge \mathbf{x}' = (\min\{x_i, x_i'\})_i$. Supermodularity is then defined accordingly. For two sets A and B, the set order $\geqslant_X$ is defined as $A \geqslant_X B$ if $a \vee b \in A, a \wedge b \in B$ for all $a \in A$ and $b \in B$. Then correspondence $\Gamma(\mathbf{s}, \cdot)$ is increasing if $\Gamma(\mathbf{s}, \mathbf{z}) \geqslant_X \Gamma(\mathbf{s}, \mathbf{z}')$ for all $\mathbf{z} \geqslant \mathbf{z}'$.

[44] See, for example, Milgrom and Roberts (1990) for an application of Topkis' monotonicity theorem in economics.

Appendix E Omitted Proofs in Section 4

E.1 Derivation of equation (17)

Evaluating (16) at D_0 gives us

$$U = \mathbb{E}[V(A, D_0, F)] = \beta U + \int_0^{F(D_0,A)} (F(D_0, A) - F)\, dG(F)$$
$$\Rightarrow (1-\beta)U = \int_0^{F(D_0,A)} (F(D_0, A) - F)\, dG(F).$$

Plugging (16) into the expectation term of (15),

$$\beta U + F(D, A) = \max_{D' \geqslant (1-\delta)D} \pi(A, D) - c(D, D') + \beta \left\{ \beta U + \int_0^{F(D',A)} \Big(F(D', A) - F\Big)\, dG(F) \right\},$$
$$\Rightarrow F(D, A) = \max_{D' \geqslant (1-\delta)D} \pi(A, D) - c(D, D') + \beta \left\{ \int_0^{F(D',A)} \Big(F(D', A) - F\Big)\, dG(F) - (1-\beta)U \right\}.$$

On the other hand,

$$\begin{aligned}
&\int_0^{F(D',A)} \Big(F(D', A) - F\Big)\, dG(F) - (1-\beta)U, \\
=&\int_0^{F(D',A)} \Big(F(D', A) - F\Big)\, dG(F) - \int_0^{F(D_0,A)} (F(D_0, A) - F)\, dG(F), \\
=&\int_0^{F(D',A)} G(F)\, dF - \int_0^{F(D_0,A)} G(F)\, dF = \int_{F(D_0,A)}^{F(D',A)} G(F)\, dF,
\end{aligned}$$

in which the second to last equality is from integration by parts. Combining the two, we obtain equation (17).

E.2 Proof to proposition 3

Proof. Suppose that for $D' > D$, there exists $D_1' \in s(D')$ and $D_1 \in s(D)$ such that $D_1 \geqslant D_1' \geqslant (1-\delta)D' > (1-\delta)D$. Then, D_1 is feasible at D' and D_1' feasible at D.[45] The optimality of D_1 and D_1' implies that

$$\pi(A, D) - c(D, D_1) + \beta\tilde{V}(A, D_1) \geqslant \pi(A, D) - c(D, D_1') + \beta\tilde{V}(A, D_1'),$$
$$\pi(A, D') - c(D', D_1') + \beta\tilde{V}(A, D_1') \geqslant \pi(A, D') - c(D', D_1) + \beta\tilde{V}(A, D_1),$$

in which $\tilde{V}(A, D) = \mathbb{E}[V(A, D, F)]$. Summing up these two inequalities, we obtain

$$-c(D, D_1) - c(D', D_1') \geqslant -c(D, D_1') - c(D', D_1). \tag{A.11}$$

if $D_1' < D_1$, the above inequality violates the strict supermodularity of $-c$.

Now consider $D_1' = D_1$. The first order condition implies $-c_2(D, D_1) + \beta\tilde{V}_2(A, D_1) + \mu = 0$, in which μ is the Langrange multiplier of the irreversibility constraint and $\mu \geqslant 0$. If $D_1' > (1-\delta)D'$,

[45] The following argument follows closely from Dechert and Nishimura (1983), in which they discuss a one-sector growth model with non-concave production function.

constraints are non-binding at both D and D'. Then, FOC would imply $D = D'$, contrary to our supposition. Now consider the binding case where $D'_1 = (1-\delta)D' > (1-\delta)D$. We obtain $\beta\tilde{V}_2(A, D_1) = c_2(D, D_1)$ from FOC at D. Plugging it into the LHS of FOC at D', we obtain the following inequality

$$-c_2(D', D_1) + c_2(D, D_1) + \mu > 0,$$

since $c_2(D, D')$ decreases strictly in D. This violates the FOC. ■

E.3 Proof to proposition 4

Proof. As a direct result of proposition 3, sequence $\{D_t\}_{t=0}^{\infty}$ is monotonic if $D_0 \notin D_{ss}(A)$. Convergence is trivial if $\{D_t\}$ has only finitely many terms. $\{D_t\}$ is increasing if $D_1 > D_0$ while decreasing if $D_1 < D_0$. Note that all D_t are bounded from below by zero. Hence, when $D_1 < D_0$, $\{D_t\}$ is convergent. It then suffices to find an upper bound for $\{D_t\}$ starting from all D_0 and $D_1 > D_0$. Let $V^*(A, D)$ be the value function in the simple case, where there are no fixed costs and exit. Simiarly, let s^* be the associated policy on D' and D_{ss} the steady state. It is a textbook exercise to show that D_{ss} exists and is unique, and s^* is a well-defined singled-valued function. We first prove the following lemma.

Lemma A2. *For all D,* $\sup s(D) \leqslant s^*(D)$.

Proof. It is straightforward that the Bellman operator defined in (14) satisfies Blackwell's sufficient conditions and then a contraction. Value function $V(A, D, F)$ and exit value U exist and are unique. Given U, let T be a Bellman operator defined as follows

$$Tf(A, D, F) = \max\Big\{ \max_{D' \geqslant (1-\delta)D} \pi(A, D) - c(D, D') - F + \beta\, \mathbb{E}[f(A, D', F')], \beta U \Big\}.$$

Then, we have $V = \lim_{n\to\infty} T^n W$ for any initial function W in a proper complete metric space. Now consider $W(A, D, F) = V^*(A, D)$ as an initial point and denote $V^n = T^n W$. Noticing that for each n, we obtain threshold function F_n in the same way as that in (15). V^n then has the same truncated form as (16). For $n > 1$, we could update F^n as follows

$$F^n(D, A) = \max_{D' \geqslant (1-\delta)D} \pi(A, D) - c(D, D') + \beta \int_0^{F^{n-1}(D', A)} G(F)\, dF - \beta(1-\beta)U. \tag{A.12}$$

The first order condition implies that

$$-c_2(D, D') + \beta F_1^{n-1}(D', A) G(F^{n-1}(D', A)) + \mu = 0,$$

in which μ is the Lagrange multiplier of the irreversibility constraint. The envelope theorem implies that

$$F_1^n(D, A) = \pi_2(A, D) - c_1(D, D') - (1-\delta)\mu.$$

We show inductively that $F_1^n(D, A) \leqslant V_2^*(A, D)$. It is straightforward to verify that this property holds at $n = 1$. Suppose it holds for F^{n-1} and let D'_n solve optimization problem (A.12). If D'_n is an interior solution, then the above first order condition implies

$$c_2(D, D'_n) = \beta F_1^{n-1}(D'_n, A) G(F^{n-1}(D'_n, A)) < \beta F_1^{n-1}(D'_n, A) \leqslant \beta V_2^*(A, D'_n).$$

Let $D^* = s^*(D)$, then $c_2(D, D^*) = \beta V_2^*(A, D^*)$. We obtain $D^* > D'_n$ since c is convex in D' and

V^* concave in D. Hence,

$$F_1^n(D,A) = \pi_2(A,D) - c_1(D,D'_n) < \pi_2(A,D) - c_1(D,D^*) = V_2^*(A,D).$$

On the other hand, if D'_n is a corner solution while D^* is not,

$$F_1^n(D,A) = \pi_2(A,D) - c_1(D,(1-\delta)D) - (1-\delta)\mu < \pi_2(A,D) - c_1(D,D^*) = V_2^*(A,D).$$

The inequality holds since $D^* > (1-\delta)D$ and $\mu > 0$. Lastly, if both $D'_n = D^* = (1-\delta)D$,

$$\begin{aligned}\mu &= c_2(D,(1-\delta)D) - \beta F_1^{n-1}((1-\delta)D,A)G(F^{n-1}((1-\delta)D,A)) \\ &> c_2(D,(1-\delta)D) - \beta V_1^*((1-\delta)D,A) = \mu^*.\end{aligned}$$

Then,

$$F_1^n(D,A) = \pi_2(A,D) - c_1(D,(1-\delta)D) - (1-\delta)\mu < \pi_2(A,D) - c_1(D,(1-\delta)D) - (1-\delta)\mu^* = V_2^*(A,D).$$

The proof of the induction step is then complete.

Given $F_1^n(D,A) \leqslant V_2^*(A,D)$ and $D' > D^*$,

$$\begin{aligned}& - c_2(D,D') + \beta F_1^{n-1}(D',A)G(F^{n-1}(D',A)) \\ \leqslant & - c_2(D,D') + \beta V_2^*(A,D') < -c_2(D,D^*) + \beta V_2^*(A,D^*) \leqslant 0\end{aligned}$$

The first inequality is due to $F_2^{n-1}(D',A) \leqslant V_2^*(A,D')$ and that G is a distribution function. The second inequality comes from $c_{22} > 0$ and $V_{22}^* < 0$, whereas the last one comes from the FOC associated with V^*. Therefore,

$$\begin{aligned}& \max_{D' \in [(1-\delta)D, D^*]} \pi(A,D) - c(D,D') + \beta \int_0^{F^{n-1}(D',A)} G(F)\,dF - \beta(1-\beta)U. \\ \geqslant & \max_{D' \geqslant D^*} \pi(A,D) - c(D,D') + \beta \int_0^{F^{n-1}(D',A)} G(F)\,dF - \beta(1-\beta)U. \\ \Longrightarrow \quad & \max_{D' \in [(1-\delta)D, D^*]} \pi(A,D) - c(D,D') + \beta \int_0^{F(D',A)} G(F)\,dF - \beta(1-\beta)U. \\ \geqslant & \max_{D' \geqslant D^*} \pi(A,D) - c(D,D') + \beta \int_0^{F(D',A)} G(F)\,dF - \beta(1-\beta)U.\end{aligned}$$

We obtain the second line since the associated max operator is continuous[46], and F^n converges to F uniformly. Whereas, the first line comes from integration. In sum, any optimal choice D' must be no larger than D^*. The proof of the lemma is then complete. ■

Now consider a optimal path in the simple model $\{D_t^*\}_{t=0}^{\infty}$. It is a textbook exercise to show that $D_t^* \leqslant D_{ss}^*$ if $D_0^* \leqslant D_{ss}^*$ and $D_t^* \leqslant D_0^*$ if $D_0^* > D_{ss}^*$. Hence, any optimal path of the extended model is bounded by the upper bound of $\{D_t^*\}_{t=0}^{\infty}$. We then proved that all optimal paths $\{D_t\}_{t=0}^{\infty}$ are convergent.

In the end, let $\hat{D} = \lim_{t\to\infty} D_t$. The continuity of F implies that $\hat{D}$ satisfies (17) and then a

[46]It is actually a contraction on F because it satisfies the Blackwell's sufficient conditions.

steady state. Furthermore, for all steady state $\tilde{D} \in \tilde{D}_{ss}$, it satisfies the first order condition

$$\begin{aligned} & c_2(\tilde{D}, \tilde{D}) = \beta G(F(\tilde{D}, A))(\pi_1(A, \tilde{D}) + c_1(\tilde{D}, \tilde{D})) < \beta(\pi_1(A, \tilde{D}) + c_1(\tilde{D}, \tilde{D})) \\ \Rightarrow \quad & c_2(\tilde{D}, \tilde{D}) - \beta c_1(\tilde{D}, \tilde{D}) < \beta \pi_1(A, \tilde{D}). \end{aligned}$$

That c is HOD 1 implies that both of its partial derivatives are HOD 1, i.e., $c_D(D, D)$ and $c_{D'}(D, D)$ are constant over D. Since $c_2(D_{ss}, D_{ss}) - \beta c_1(D_{ss}, D_{ss}) = \beta \pi_2(A, D_{ss})$, $D_{ss} > \tilde{D}$. The proof is then complete.

■

E.4 Proof to proposition 5

Proof. We first prove the following lemma that F is supermodular in (A, D).

Lemma A3. *F is supermodular in A and D.*

Proof. Consider the construction V^n in the proof of lemma A2. Since $\lim_{n\to\infty} V^n = V$, F^n also converges to F by (15). It then suffices to show that F^n are supermodular for all n. Before that, it is fairly straightforward that $V^*(A, D)$ is strictly supermodular. One way to see this is from the sequence problem, in which $V^*(A, D)$ is the limit of a N-period sequence problem. A similar induction as below can obtain that a N-period value function is supermodular.[47] Therefore, V^* is supermodular. Further, strict supermodularity comes from the recursive formulation of V^* and strict supermodularity of the profit function.

As before, let D'_n denote the policy associated with F^n. $D'_1(A, D)$ is then increasing in A for all D. To see this, suppose there exists $D'_1(A_1, D) < D'_1(A_2, D)$ for some D and $A_1 > A_2$. Let $\hat{D}_i = D'_1(A_i, D)$ for notational brevity. The optimality of $\hat{D}'_i$ then implies

$$\begin{aligned} \pi(A_1, D) - c(D, \hat{D}_1) + \beta V^*(A_1, \hat{D}_1) &\geqslant \pi(A_1, D) - c(D, \hat{D}_2) + \beta V^*(A_1, \hat{D}_2), \\ \pi(A_2, D) - c(D, \hat{D}_2) + \beta V^*(A_2, \hat{D}_2) &\geqslant \pi(A_2, D) - c(D, \hat{D}_1) + \beta V^*(A_2, \hat{D}_1). \end{aligned}$$

Summing up these two inequalities,

$$\beta V^*(A_1, \hat{D}_1) + \beta V^*(A_2, \hat{D}_2) \geqslant \beta V^*(A_1, \hat{D}_2) + \beta V^*(A_1, \hat{D}_2),$$

which clearly violates the strict supermodularity of V^*. Next, we consider the first order derivative of F^1,

$$F^1_1(D, A) = \pi_2(A, D) - c_1(D, D'_1) - (1 - \delta)\mu.$$

When $\mu = 0$, it is easy to see that $-c_1(D, D'_1)$ increases in A since D'_1 increases in A. Provided that $\pi_{12} > 0$, F^1_1 increases strictly in A. When $\mu > 0$, $\mu = \beta c_2(D, (1-\delta)D) - \beta V^*_2(A, (1-\delta)D)$. An increase in A strictly decreases μ, so F^1_1 increases strictly in A. Hence, F^1 is strictly supermodular in A and D.

Lastly, suppose that F^{n-1} is strictly supermodular. Note that

$$\frac{\partial}{\partial D} \int_0^{F^{n-1}(D,A)} G(F)\, dF = G(F^{n-1}(D, A)) F^{n-1}_2(D, A),$$

which increases strictly in A.[48] Since U is a function of only A, the integral term (expected net

[47] This is much simpler than below since we can use the first order conditions and that V^n is strictly concave in D.

[48] Similar induction argument can establish that $F^{n-1}(D, A)$ strictly increases in A.

of exit value) in (A.12) is strictly supermodular, and D'_n increases in A. Replicating the same argument in the base case, we establish that F^n is also strictly supermodular. The induction argument is then complete. ■

Observe that

$$\frac{\partial}{\partial D}\int_{F(D_0,A)}^{F(D,A)} G(F)\,dF = G(F(D,A))F_1(D,A).$$

It increases in A strictly since $F(D,A)$ increases strictly in A and G has full support on $\mathbb{R}^+$. Hence, the expected value of next period is strictly supermodular. We can then apply the same argument to conclude $D'_1 \geqslant D_1$. To see that the inequality is strict, the non-binding first order condition

$$c_2(D_0, D_1) = G(F(D_1, A))F_1(D_1, A)$$

never holds for two different A and A' since the RHS increases strictly in A.

In the end, we show that a more productive firm would always have a larger capital stock conditional on survival. The above analysis indicates that $D'_1 > D_1$. Suppose there exists t such that $D'_t > D_t$ but $D'_{t+1} < D_{t+1}$. The optimality implies that

$$\begin{aligned}
&- c(D'_t, D'_{t+1}) + \beta\tilde{V}(A', D'_{t+1}) - c(D_t, D_{t+1}) + \beta\tilde{V}(A, D_{t+1}) \geqslant \\
&\qquad - c(D'_t, D_{t+1}) + \beta\tilde{V}(A', D_{t+1}) - c(D_t, D'_{t+1}) + \beta\tilde{V}(A, D'_{t+1}) \qquad \text{(A.13)}
\end{aligned}$$

This proves to be a contradiction since $-c(D, D')$ is strictly supermodular in (D, D') and $\tilde{V}(A, D)$ in (A, D). Also, we use first order condition to show that $D'_{t+1} = D_{t+1}$ is impossible:

$$c_2(D_t, D_{t+1}) = G(F(D_{t+1}, A))F_1(D_{t+1}, A),$$

in which the LHS strictly decreases in D_t yet the RHS strictly increases in A.

■

Appendix F Additional Results of Structural Estimation

F.1 Identification details

Extended Model Identifying structural parameters in the extended model follows a similar strategy. Since the fixed cost shock is independent of firm productivity, the resulting value function $v(A, d, f)$ from previous normalization would still be dependent on A. For the sake of clarity, we consider a fixed cohort and then suppress superscript c. We obtain the following Bellman equation after normalizing (14) by D_{ss} in (11):

$$v(A,d,f) = \max\left\{ \max_{d' \geqslant (1-\delta)d} \tilde{D}_{ss}^{\alpha-1} d^{\alpha} - (d' - (1-\delta)d) - \phi \frac{(d' - (1-\delta)d)^2}{d} - f + \beta\,\mathbb{E}[v(A,d',f')], \right.$$
$$\left. \beta\,\mathbb{E}[v(A,d_0,f')] \right\} \tag{A.14}$$

in which $f \sim \text{Lomax}(\kappa_c, \theta)$ conditional on surviving the death shock and $\kappa_c = \kappa / D_{ss}(A_c)$. The distribution of f depends on A_c and is denoted by G_c. As before, for each cohort, the identification of $\{\alpha, \phi, \delta, d_0\}$ comes from the Euler equation, albeit that it will require not only conditional sales growth but also conditional survival rates, which is given by $p_t = G_c(F(d_{t+1}))$. Explicitly, the Euler equation is as follows[49]

$$B_t + 2y_t^{\frac{1}{\alpha}} = \frac{\beta G_c(F(d_{t+1}))\alpha \tilde{D}_{ss}^{\alpha-1}}{\phi} d_0^{\alpha-1} \prod_{s=0}^{t} y_s^{\frac{\alpha-1}{\alpha}} + \beta G_c(F(d_{t+1})) y_{t+1}^{\frac{2}{\alpha}}, \tag{A.15}$$

in which $B_t = \frac{1}{\phi}\Big[1 - \beta G_c(F(d_{t+1}))(1-\delta)\Big] + 2\delta(1 - \beta G_c(F(d_{t+1}))) + \beta G_c(F(d_{t+1}))\delta^2 + \beta G_c(F(d_{t+1})) - 2$. Equation (22) is obtained from the above by inserting $p_t = G_c(F(d_{t+1}))$ and $B + 2 - \beta = \beta\alpha\tilde{D}_{ss}^{\alpha-1}/\phi$. In the simple model, $B_t = B = \beta\alpha\tilde{D}_{ss}^{\alpha-1}/\phi + \beta - 2$ since $p_t = 1$. As long as ϕ and δ satisfy B, B_t is satisfied as well. That there is only one equation with two unknowns renders ϕ and δ unidentified. In the extended model, however,

$$B_t = p_t B + (\frac{1}{\phi} + 2\delta - 2)(1 - p_t)$$

with $p_t < 1$. Then, B is not a sufficient statistic of B_t, and ϕ and δ can be pinned down using both B_t^c and B. Therefore, lifecycle moments $\{\alpha, \phi, \delta, d_0\}$ can be identified using firm dynamics moments, namely, the conditional survival rates p_t and sales growth y_t.

With $\{\alpha, \phi, \delta, d_0\}$ in hand, the threshold function F is then a function of κ_c, θ and γ, and so is the survival rate function $G_c(F(\cdot))$. Matching it to conditional survival rates in the data gives the estimates of κ_c, θ and γ. For estimation, we use moment conditions of all cohorts since the lifecycle parameters are common across cohorts. In the end, another way to obtain cohort effect A_c is to recover them from κ_c with normalization $\kappa = \kappa_1$. Initial capital D_0 would also be straightforward given A_c, normalized initial capital d_0 and other parameters.[50] Alternatively, one could still follow

[49]Due to the irreversibility constraint, we will, in principle, need to replace the Euler equation by a complementary slackness condition such that either (22) holds or $d_{t+1} = (1-\delta)d_t$. However, this won't change any of the following analysis. In addition, that the post-entry export value is growing at all ages in the data implies that the irreversibility constraint would never bind, given the model is right. We then continue using the Euler equation.

[50]One may aware that in the extended model, cohort effect A^c could be identified without using relative initial sales. This is because productivity and capital affects the distribution of fixed cost differently. A change in productivity shifts the distribution while a change in capital only moves along the distribution. With data on survival rate, we

the same procedure outlined in the simple model using relative initial sales and normalized initial capital d_0. It should be emphasized that the identification of d_0 from (22) is independent of A_c. The only place in the Euler equation where A_c plays a role is $G_c(F(d_{t+1}))$, which is data. Hence, A_c could be inverted exactly as in (20). In the actual estimation, we include relative initial sales as moment conditions.

Derivation of (21) From the first order condition to (12)

$$\begin{aligned}
&1+2\phi\Big(\frac{d_{t+1}}{d_t}-(1-\delta)\Big)=\beta\Big[\tilde{D}_{ss}^{\alpha-1}\alpha d_{t+1}^{\alpha-1}+(1-\delta)\Big(1+2\phi\Big(\frac{d_{t+2}}{d_{t+1}}-(1-\delta)\Big)\Big)+\phi\Big(\frac{d_{t+2}}{d_{t+1}}-(1-\delta)\Big)^2\Big]\\
\Leftrightarrow &1-2\phi(1-\delta)+2\phi\frac{d_{t+1}}{d_t}=\Big[(1-\beta(1-\delta))(1+2\phi\delta)-\beta\phi\delta^2\Big]d_{t+1}^{\alpha-1}+\beta\Big[(1-\delta)-\phi(1-\delta)^2+\phi\Big(\frac{d_{t+2}}{d_{t+1}}\Big)^2\Big]\\
\Leftrightarrow &\frac{1}{\phi}(1-\beta(1-\delta))-2(1-\delta)+\beta(1-\delta)^2+2\frac{d_{t+1}}{d_t}=\Big[\frac{1}{\phi}(1-\beta(1-\delta))+(1-\beta(1-\delta))2\delta-\beta\delta^2\Big]d_{t+1}^{\alpha-1}+\beta\Big(\frac{d_{t+2}}{d_{t+1}}\Big)^2\\
\Leftrightarrow &\Big(\frac{1}{\phi}(1-\beta(1-\delta))+2\delta(1-\beta)+\beta\delta^2-2+\beta\Big)+2\frac{d_{t+1}}{d_t}=\Big[\frac{1}{\phi}(1-\beta(1-\delta))+2\delta(1-\beta)+\beta\delta^2\Big]d_{t+1}^{\alpha-1}+\beta\Big(\frac{d_{t+2}}{d_{t+1}}\Big)^2\\
\Leftrightarrow &B+2\frac{d_{t+1}}{d_t}=(B+2-\beta)d_{t+1}^{\alpha-1}+\beta\Big(\frac{d_{t+2}}{d_{t+1}}\Big)^2.
\end{aligned}$$

Plug in $d_t = d_0\prod_{s=0}^{t} y_s^{\frac{1}{\alpha}}$, we obtain (21). Also, $B+2-\beta=\beta\alpha\tilde{D}_{ss}^{\alpha-1}/\phi$ follows directly from (11).

Derivation of (A.15) From the first order condition to (A.14)

$$\begin{aligned}
&1+2\phi\Big(\frac{d_{t+1}}{d_t}-(1-\delta)\Big)=\beta G_c(F(d_{t+1}))\Big[\tilde{D}_{ss}^{\alpha-1}\alpha d_{t+1}^{\alpha-1}+(1-\delta)\Big(1+2\phi\Big(\frac{d_{t+2}}{d_{t+1}}-(1-\delta)\Big)\Big)+\phi\Big(\frac{d_{t+2}}{d_{t+1}}-(1-\delta)\Big)^2\Big]\\
\Leftrightarrow &1-2\phi(1-\delta)+2\phi\frac{d_{t+1}}{d_t}=\beta G_c(F(d_{t+1}))\alpha\tilde{D}_{ss}^{\alpha-1}d_{t+1}^{\alpha-1}+\beta G(F(d_{t+1}))\Big[(1-\delta)-\phi(1-\delta)^2+\phi\Big(\frac{d_{t+2}}{d_{t+1}}\Big)^2\Big]\\
\Leftrightarrow &1-2\phi(1-\delta)-\beta G_c(F(d_{t+1}))\Big[(1-\delta)-\phi(1-\delta)^2\Big]+2\phi\frac{d_{t+1}}{d_t}=\beta G_c(F(d_{t+1}))\alpha\tilde{D}_{ss}^{\alpha-1}d_{t+1}^{\alpha-1}+\beta G_c(F(d_{t+1}))\phi\Big(\frac{d_{t+2}}{d_{t+1}}\Big)^2\\
\Leftrightarrow &\frac{1}{\phi}\Big[1-\beta G_c(F(d_{t+1}))(1-\delta)\Big]+2\delta(1-\beta G_c(F(d_{t+1})))+\beta G_c(F(d_{t+1}))\delta^2+\beta G_c(F(d_{t+1}))-2+2\frac{d_{t+1}}{d_t}\\
&=\frac{\beta G_c(F(d_{t+1}))\alpha\tilde{D}_{ss}^{\alpha-1}}{\phi}d_{t+1}^{\alpha-1}+\beta G_c(F(d_{t+1}))\Big(\frac{d_{t+2}}{d_{t+1}}\Big)^2\\
\Leftrightarrow &B_t+2\frac{d_{t+1}}{d_t}=\frac{\beta G_c(F(d_{t+1}))\alpha\tilde{D}_{ss}^{\alpha-1}}{\phi}d_{t+1}^{\alpha-1}+\beta G_c(F(d_{t+1}))\Big(\frac{d_{t+2}}{d_{t+1}}\Big)^2
\end{aligned}$$

F.2 Estimation Results with Strategic Entry

can no longer completely counteract the impact of productivity using capital alone, as we did in the simple model.