北京大学中国经济研究中心
China Center for Economic Research
讨论稿系列
Working Paper Series

E2020001 2020-01-03

Firm Quasi-Dynamics in the Chinese Manufacturing Industries

Lixing Li [a] Qihong Liu [b] Changyu Ren [a] [1]

[a] National School of Development, Peking University
[b] Department of Economics, University of Oklahoma

This version: December 31, 2019

Abstract

We explore the firm quasi-dynamics (entry/exit and growth) in the Chinese manufacturing industries and investigate how these dynamics vary across regions. Our results show that relative to provinces with less developed economies, in provinces with more developed economies: (1) There are higher shares of new firms; (2) New firms are smaller and more labor-intensive; (3) Firms exit at a quicker rate and surviving firms grow faster. These results point toward cross-region differences in market efficiency in terms of how much it costs a firm to enter or exit the market. Our findings shed light on how firms should adapt their strategies across regions and how the government should create sound policies on industrial upgrading and relocation.

Keywords: firm dynamics, entry and exit, industrial shift, market thickness, labor cost

JEL Classification: D24, L25, O14

[1] The authors are listed in alphabetical order with equal contribution. We thank participants at the 2017 Chinese Economists Society Conference in Nanjing, China for helpful comments and suggestions. Li acknowledges financial support from the National Natural Science Foundation of China (#71773004). Liu thanks the visiting scholars program of the China Center for Economic Research, National School of Development, Peking University.

1. Introduction

After decades of economic reform, Chinese manufacturing industries have experienced remarkable growth and have reached a stage that calls for dramatic industrial upgrading and relocation (due to rising labor costs). An extensive literature has been devoted to analyzing the Chinese manufacturing industries (see Brandt et al., 2012 and the papers cited there) Most of the existing studies focus on established firms, facilitated by the availability of quality data such as the Annual Survey of Industrial Firms (ASIF) which only includes above-scale industrial firms. In contrast, data on small Chinese firms are scarce, and as a result, small firms are rarely the subject of academic research.

We aim to fill the gap in this paper and shed light on small Chinese firms and their behaviors. Using two waves of Chinese economic census data (in 2004 and 2008 respectively), we explore the firm quasi-dynamics (entry/exit and growth) in the Chinese manufacturing industries and investigate how these dynamics vary across regions. We characterize regions (provinces) by their migration status, which closely mimic the level of their economic development. That is, provinces with more developed economies, in general, attract workers and immigrants from other provinces. If a province has a higher ratio of net immigrants, then we say the province has a more developed economy. We are particularly interested in cross-region comparison of the numbers and characteristics of new firms, and their "exit" and "growth" patterns over time.

We start by analyzing firm entry (new firms). Our results show that provinces with more developed economies see higher numbers of new firms, which are also smaller and more labor-intensive relative to their counterparts at provinces with less developed economies. It is rather counter-intuitive that new firms are more labor-intensive in regions with more developed economies. After all, labor is more expensive in more developed provinces, and one would thus expect new firms there to be less labor-intensive. Next, we proceed to analyze firm exit and growth. We find that in provinces with more developed economies, firms exit at a quicker rate, and surviving firms grow faster (using several measures of firm size). As supporting evidence, we also conduct cross-sectional comparison firms by comparing the 2004 and 2008 data, and by examining the time-series change of firms' capital intensity. We find that while firms have generally become more capital intensive from 2004 to 2008, the capital intensity of firms has decreased in more developed provinces relative to firms in less developed regions.

These results seem to point toward cross-region differences in market efficiency in terms of how much it costs a firm to enter or exit the market. That is, entry/exit costs are lower in provinces with more developed economies. With lower entry costs, more developed provinces would see higher numbers of new firms, which are also smaller. Smaller firms tend to be more labor-intensive, which more than offset the cross-region differences on factor prices, leading to the finding that new firms are more labor-intensive in more developed provinces. Similarly, lower exit cost means that a non-performing firm can easily exit the market in a more developed province. Entry and exit costs can feed into each other, and having the option to exit the market easily likely induces more firms to enter the market in the first place.

Understanding our results and market efficiency is necessary for firms to adapt their strategies across regions and for the government to create sound policies on industrial upgrading and relocation. For example, the cross-region difference in market efficiency may arise because of bureaucracy and red tape. Historically, regions with less developed economies are perceived to have more red tape, increasing the cost of doing business and, in turn, entry/exit cost for firms. But other factors may also be at play, in particular, clustering or market thickness. For example, a new firm operating in certain business may require specific types of labor which are clustered in provinces in more developed economies. In a thick market, a new firm can easily recruit such labor to start operations, and if it fails, it can easily lay off workers – they can easily find replacement jobs in a thick market. Different sources of market efficiency heterogeneity, for example, bureaucracy or market thickness, call for different policies to address potential market inefficiency.

The main contribution of our paper is using census data to look at small firms and explore industry dynamics in China. This has rarely been done in the literature.[2] Nevertheless, small firms play a critical role in the economy, and in helping us understand the industrial upgrading and relocation patterns for the Chinese economy. Next, we discuss several important roles which small firms have been shown to play in developed economies where small firms are more frequently studied.

First, small firms are frequently quoted as a major driver for job growth in both academic studies and popular media.[3] This is not surprising. Small firms tend to be labor-intensive and become more capital intensive as they grow.[4] Through job creation, taxes, and various other channels, small firms also contribute significantly to the local community and government.

Second, small firms are important for innovation. Successful innovation often leads to the creation of new firms, which, while small, have great potential (this is best seen in tech startups). Small firms have to rely on substantial innovation to survive and grow. The situation can be quite different for established firms. For example, an established firm may choose to focus on process innovation to reduce the cost of producing existing products, or to push existing successful products to new areas (such as new geographic markets or new product lines).

Third, small firms are significant because of industry dynamics. When new opportunities arise, facilitated by either new technologies or new ways of doing business, new firms enter the market. These new firms are almost always small firms. Based on studies of industry dynamics (e.g., Dunne et al., 1988; Dunne et al., 1989), the majority of the small firms will not survive, but they are critical because they bring new opportunities. The surviving firms tend to grow rapidly

---

[2] An exception is Ju & Liu (2019).

[3] See, for example, Birch (1979) and Davis et al., (1996), and two media reports. “Small Business Are Having A Bigger Impact On Job Creation Than Large Corporations, *Forbes*, May 5, 2019; “Small Businesses Drive Job Growth in the U.S.”, U.S. Small Business Administration, April 25, 2018.

[4] Consistent with this notion, in our data, older and larger firms are more capital-intensive relative to newer and smaller firms.

and make a significant contribution to the economy, and oftentimes, a subsequent round of new firms are established around those surviving/successful firms.

Our paper is closely related to the literature on firms' location choice and cross-region investment, given our focus of cross-region comparison of firm dynamics. Existing studies have examined the role of agglomeration (Lu & Tao, 2009; Li & Lu, 2011), environmental regulation (Wu et al., 2017), transportation infrastructure (Yang, 2018), and political connection (Shi et al., 2018). Other studies have documented industrial convergence in China (Deng & Jefferson, 2011; Lemoine et al., 2015). Our results suggest that there are systematical cross-region differences in market efficiency (in terms of entry and exit cost), which lead to different firm entry, exit and growth patterns. It is likely to translate into significant cross-region differences of established firms as well, which will persist over time. This is in a similar spirit to findings in studies using the "flying geese model" (Kojima, 2000; Qu et al., 2012). There, the argument is that as factor prices go up in a developed country, labor-intensive industries tend to move to other, less expensive places, while the developed country retains only those capital-intensive high–value-added industries.

Our paper also provides a new angle to look at the industrial policy. While we emphasize that regional differences in industrial development could be a result of market reasons, the firm dynamics patterns observed in our study could also be driven by industrial policies. Various types of industrial policies have been shown to play an important role in firm growth in China (e.g., Aghion et al., 2015; Chen et al., 2019; Barwick et al., 2019). Our findings suggest that successful industrial policies should encourage the entry and exit of small firms by reducing market frictions. Efficient entry and exit of firms are a pre-condition for regional industrial policies to work.

Our paper is also related to the literature on regulatory entry barriers in developing countries. While regulatory barriers are widely regarded as an obstacle for firm entry around the world (Klapper et al., 2004) and particularly important in China (Bai et al., 2004; Brandt et al., 2018), it affects firm entry not just because of policy factors such as red tape, but also through market forces such as factor market mobility and clustering. We discuss how both the market and policy factors can lead to regional differences in observed industry dynamics, yet they have very different policy implications. A direct application is an industrial transition across regions in China, and our research provides guidance to governments on how they may facilitate such a transition. We also provide some ideas on how one can distinguish these two factors in future work.

The remaining of the paper is organized as follows. Section 2 discusses data and variables. Sections examines firm entry. Section 4 analyzes the exit/growth behavior of firms. Section 5 provides additional evidence by linking the 2004 and 2008 datasets. We provide further discussions of our findings, draw policy implications, and propose future research topics in Section 6.

2. Data and variables

2.1 Description of the data

The data we use mainly come from population censuses and economic censuses. The population census data comes from China's population censuses of 2000 and 2010, and the 2005 inter-census population survey.

Our firm-level data comes from China's 2004 and 2008 economic censuses carried out by the National Bureau of Statistics (NBS) of China. The economic censuses cover all firms in all sectors that were engaged in economic activities at the end of the corresponding year. Our study focuses on the manufacturing industries. Universal coverage is the major advantage of this dataset (see, e.g., Long & Zhang, 2011; Ju & Liu, 2019; Wang, 2019). In particular, compared with the Annual Survey of Industrial Firms (ASIF) that are widely used in the literature, the economic census data include not only large firms but also small firms with annual sales below 5 million RMB, which are of major interest to us. For each firm, the dataset provides a wide range of firm characteristics, including location, industry, ownership, year of establishment, and various financial information. For example, the 2004 data includes annual output value, business income, assets, number of employees, and payable wage (part of the full sample) and the value of output. The 2008 data reports similar information (except output value, which is reported in 2004 data only).

The data cleaning process is discussed in the Appendix.

2.2 Construction of key variables

1) Age: A firm's age is defined as Age = Census Year – Establish Year + 1. For example, in the 2004 census data, a firm established in 2000 will have Age=5. Similarly, Age=1 means the firm is established in 2004 and is thus counted as a new firm.

2) $Migration\_Pct$: We use the population census data in 2000, 2005 and 2010, which include information on number of residents and migrants. In each of these three years, for province *i*, we obtain the following numbers: permanent residents ($Residents_i$); immigrants living in province *i* with *hukou* in another province ($Immigrants_i$); and emigrants living in other provinces with hukou in province *i* ($Emmigrants_i$). We then calculate

$$Migration\_Pct_i = \frac{Immigrants_i - Emigrants_i}{Residents_i}$$

This number may be positive or negative. A positive (negative) migrant ratio implies that the province experiences a net population inflow (outflow). Also, the number is in decimals rather than percentage points. That is, if province *i* has a 5% net population inflow, then $Migration\_Pct_i = 0.05$. We calculate provincial-level $Migration\_Pct_i$ for the year 2000, 2005 and 2010 respectively, and report the average as $Migration\_pct_i$.

3) $LogRatio_j$: We first calculate the firm level capital-labor ratio ($Ratio$), defined as total assets divided by total number of employees. We then take log to obtain firm-level log capital-labor ratio (logratio), and calculate the simple average across all firms in each industry j ($\mathrm{LogCLR_j}$). Finally, we normalize the industry level log capital-labor ratio across

industries by using the z-scores, $LogRatio_j = \frac{LogCLR_j - Avg\ (LogCLR_j)}{Std\ (LogCLR_j)}$.[5] This is easier to interpret – an industry has above (below) average capital intensity if its $LogRatio_j > 0$ (<0). This is similar to $Migration\_Pct_i$, where a positive (negative) $Migration\_Pct_i$ implies that the province has more immigration (emigration).

4) *LogNum* and *Share*. We first count the number of firms operating in province i, industry j and age t ($Number_{ijt}$), using the 2004 and 2008 data, respectively. Taking into account the frequent cases of $Number_{ijt}$= 0, we define $LogNumber_{ijt} = \log(Number_{ijt} + 1)$ and introduce a dummy variable $Num0_{ijt}$ which takes value 1 when $Number_{ijt}$= 0, and 0 otherwise.

We also characterize the distribution of firms across provinces. Adding up the numbers over all provinces, we can obtain ($Number_{jt}$). $Share_{ijt} = \frac{Number_{ijt}}{Number_{jt}}$ captures the share of firms in industry *j* with age *t* which operate in province *i*. There are also multiple firm characteristics variables directly coming from the data. They include total assets (log_asset), total employment (log_employee), total output value (log_output, in 2004 data only) and business income (log_income).

2.3 Summary Statistics

Table 1 lists the summary statistics of key variables for 2004 and 2008 respectively.

Table 1. Summary Statistics

	obs	mean	sd	median	min	max
Migration_Pct	31	0.01	0.09	0.00	-0.10	0.27
2004 Data: N = 1156358						
LogRatio	30	0.00	5.48	-1.12	-6.80	21.12

Age = 1, province-industry level

[5] For an industry with an average $LogCLR_j$, the normalization gives $LogRatio_j$ = 0, which in turn implies the interaction term $Migration\_Pct_i$ * $LogRatio_j$ = 0. Then the impact of $Migration\_Pct_i$ is picked up by the term $Migration\_Pct_i$ alone, without the need to consider the interaction term.

Num0	930	0.08	0.27	0	0	1
LogNum	930	3.57	1.86	3.78	0	8.05
Share	930	0.03	0.05	0.01	0	0.43
All ages, firm level						
Age	1141887	6.33	5.23	5	1	27
logratio	1156353	1.85	1.11	1.88	-5.52	10.19
log_output	1156270	7.33	2.14	7.62	0.00	16.11
log_employee	1156358	3.20	1.15	3.04	0.00	10.39
log_asset	1156357	4.97	1.61	4.81	-2.30	13.81
log_income	933969	6.72	1.81	7.17	0.00	13.55
			2008 Data: N= 1692813			
LogRatio	30	0.00	5.48	-1.69	-7.69	11.28
Age = 1, province-industry level						
Num0	930	0.07	0.26	0	0	1
LogNum	930	3.53	1.81	3.69	0	7.93
Share	930	0.03	0.05	0.02	0	0.38
All ages, firm level						
Age	1686820	6.89	5.18	6	1	31
logratio	1692813	2.32	1.13	2.33	-9.52	9.48
log_employee	1692813	3.11	1.13	3.04	0.69	12.08
log_asset	1692813	5.35	1.63	5.21	-2.30	16.21
log_income	1692813	5.61	1.89	5.77	0.00	16.77

3. Firm entry: Cross-sectional comparison of new firms

3.1 The number of new firms

We start by analyzing the entry of new firms, defined as all firms with *Age* = 1. We do so using the 2004 and 2008 dataset, respectively. New firms have been an important source of economic growth since the beginning of economic reforms in China. On the extensive margin, they draw in new resources (labor and capital) to the economy; on the intensive margin, they help raise total factor productivity in the economy (Brandt et al., 2012).

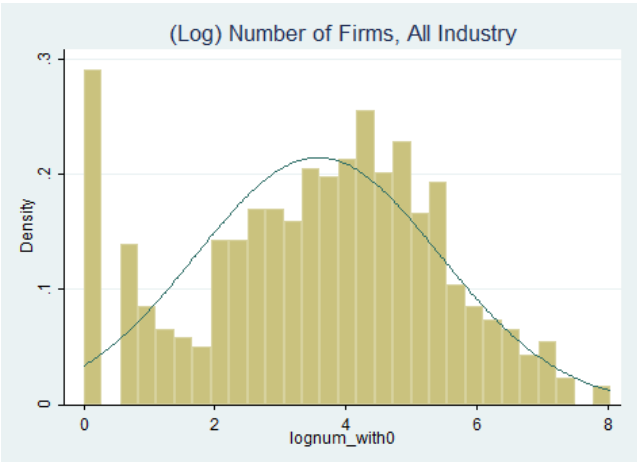
We use the following econometric model to examine the impact of migrant ratio on firm entry:

$$LogNum_{ij} = a_0 + a_1\ Migration\_Pct_i + a_2\ LogRatio_j + a_3\ Migration\_Pct_i \times LogRatio_j + \varepsilon_{ij} \quad (1)$$

where $LogNum_{ij}$ is the log number of new firms at the province-industry (*ij*) level, $Migration\_Pct_i$ is the province-level ratio of net immigration and $LogRatio_j$ is the industry-level log capital-labor ratio.

Naturally the number of new firms in a province-industry cell ($Number_{ij}$) is censored from below at zero. One common solution to such censoring is to run a Tobit regression. However, the problem in our setting goes beyond censoring. This can be seen from the density plot in Figure 1, where we plot the distribution of new firms across provinces and industries. The horizontal axis is $LogNum_{ij}$, while the vertical axis is the density. Note that $LogNum_{ij} = \log(Number_{ij} + 1) \geq 0$, and it takes value 0 if and only if $Number_{ij} = 0$. We try to fit the density plot to a symmetric normal distribution, with mean/peak at $LogNum_{ij} = \mu$. It is straightforward to see that there is substantial clustering at $LogNum_{ij} = 0$, in particular, $prob(LogNum_{ij} \leq 0)$ far outweighs $prob(LogNum_{ij} \geq 2\mu)$. This suggests that, beyond censoring, additional factors are responsible for having $LogNum_{ij} = 0$. In practice, province *i* may simply lack essential inputs to enter into industry *j,* and the problem is not just censoring. To take this into account, we add to equation (1) a dummy variable $Num0_{ij}$, which takes value 1 if and only if the province-industry combination has no new firm ($Number_{ij} = 0$).

Figure 1: Number of new firms: Density plot



The estimation results of equation (1) are reported in Table 2. Columns (1) - (3) use the 2004 data. We can see that the coefficient of *Migration_Pct* is positive and significant in all columns. Consider an industry with average capital intensity ($LogRatio_j = 0$). From Column (1), if $Migration\_Pct_i$ increases by 1 percentage point (e.g., from 0.02 to 0.03), then the number of new firms in province *i* will increase by about 2.1%. Moving onto $LogRatio_j$, the results suggest that fewer new firms exist in more capital-intensive industries. This may be because capital intensive industries pose a higher entry barrier so that fewer new firms can overcome that hurdle. The estimate for the dummy variable $Num0$ is significant and fairly large in magnitude, consistent with our earlier conjecture that unobserved factors play a significant role in the cases of $LogNum_{ij} = 0$. Column (2) introduces the interaction term $Migration\_Pct_i * LogRatio_j$. The estimate is small in magnitude and statistically insignificant. In Column (3), we control for industry fixed effects. $LogRatio_j$ is automatically dropped since it does not vary within an industry. We can see that the results are quite similar to those in Column (2). Columns (4) - (6) use 2008 data. The estimates for $Migration\_Pct$ remain positive and significant, but the magnitude goes down slightly (from 2.1 down to about 1.3).

Table 2. Impact of migrant ratio on firm entry

	(1)	(2)	(3)	(4)	(5)	(6)
	2004	2004	2004	2008	2008	2008
Migration_Pct	2.104***	2.104***	2.101***	1.271**	1.271**	1.231**
	[0.541]	[0.541]	[0.487]	[0.536]	[0.537]	[0.480]
LogRatio	-0.070***	-0.070***		-0.073***	-0.073***	
	[0.010]	[0.010]		[0.009]	[0.009]	
Migration_Pct * LogRatio		-0.045	-0.045		-0.054	-0.062
		[0.101]	[0.090]		[0.100]	[0.089]
Num0	-3.280***	-3.279***	-3.221***	-3.559***	-3.556***	-3.030***
	[0.199]	[0.199]	[0.192]	[0.187]	[0.188]	[0.196]
Industry FE			Yes			Yes
R-squared	0.369	0.369	0.505	0.349	0.349	0.495
Observations	930	930	930	930	930	930

Note: Regression is at the province-industry level. New firms are defined as age = 1. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

We also replace the dependent variable $LogNum_{ij}$ by $Share_{ij}$ — the percentage of new firms in industry j that are located in province i. We then run regressions similar to equation (1). The results are presented in Table 3, where 2004 (2008) data are used in columns (1)-(3) ((4)-(6)). We can see that they are mostly consistent with the results in Table 2. In particular, provinces with a larger migrant ratio see more (a higher share) of new firms.

Table 3. Impact of migrant ratio on share of new firms

	(1)	(2)	(3)	(4)	(5)	(6)
	2004	2004	2004	2008	2008	2008
Migration_Pct	0.117***	0.117***	0.117***	0.106***	0.106***	0.107***
	[0.018]	[0.018]	[0.018]	[0.017]	[0.017]	[0.017]
LogRatio	0.001***	0.001***		0.000	0.000	
	[0.000]	[0.000]		[0.000]	[0.000]	
Migration_Pct *		-0.004	-0.004		-0.002	-0.002
LogRatio		[0.003]	[0.003]		[0.003]	[0.003]
Num0	-0.045***	-0.044***	-0.051***	-0.037***	-0.037***	-0.050***
	[0.007]	[0.007]	[0.007]	[0.006]	[0.006]	[0.007]
Industry FE			Yes			Yes
R-squared	0.086	0.088	0.095	0.076	0.076	0.090
Observations	930	930	930	930	930	930

Note: Regression is at the province-industry level. New firms are defined as age = 1. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

3.2 New firm size

Previous results show that provinces with higher ratios of immigrants see a higher number/share of new firms. Next, we explore how the new firms in different provinces compare with each other using firm-level data. Our focus is on firm size, captured by a variety of measures including (annual) output value, business income, number of employees, and the amount of assets, all in natural logs. The regression model is similar to equation (1), except that the data is at the firm level, and the dependent variables are firm size measures.

Results using the 2004 data are presented in Table 4. For easier interpretation, think of an industry with average capital intensity so $LogRatio_j = 0$ (recall that $LogRatio_j$ is normalized).

The estimate for $Migration\_Pct_i$ is negative and significant. That is, in provinces with larger migrant ratios (i.e., more developed economies), new firms are significantly smaller. In Column (1), the dependent variable is the log of annual output value. The result suggests that when $Migration\_Pct_i$ increases by 1 percentage point (e.g., from 0.3 to 0.31), annual output value on average will decrease by about 3.2%. For an industry with above-average capital intensity, the same change in $Migration\_Pct_i$ would lead to a larger drop in new firm size, because the estimate of the interaction term $Migration\_Pct_i * LogRatio_j$ is negative and significant. In Column (2), we control for industry fixed effects, and the results are qualitatively the same. Columns (3) and (4) use log of annual business income as the dependent variable, and results are comparable to those in (1). The estimates of $Migration\_Pct_i$ are smaller when number of employees or total assets is used to measure firm size. For example, an increase of 1 percentage point in $Migration\_Pct_i$ leads to a less than 1% decrease in total number of employees.

Table 4. Impact of migrant ratio on new firm size - 2004

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	log_output	log_output	log_income	log_income	log_employee	log_employee	log_asset	log_asset
Migration_Pct	-3.229***	-2.595***	-2.721***	-2.389***	-0.860***	-0.956***	-1.281***	-0.869***
	[0.078]	[0.084]	[0.072]	[0.077]	[0.039]	[0.040]	[0.055]	[0.059]
LogRatio	0.042***		0.003		0.009***		0.092***	
	[0.002]		[0.002]		[0.001]		[0.002]	
Migration_Pct	-0.191***	-0.091***	-0.093***	-0.019	-0.003	-0.024*	-0.207***	-0.174***
* LogRatio	[0.024]	[0.026]	[0.022]	[0.024]	[0.012]	[0.013]	[0.017]	[0.019]
Industry FE		Yes		Yes		Yes		Yes
R-squared	0.021	0.044	0.019	0.036	0.007	0.070	0.029	0.043
Observations	129,317	129,317	116,876	116,876	129,322	129,322	129,322	129,322

Note: Regression is at the firm level. The sample is restricted to new firms with age = 1. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

We also use 2008 data and run similar regressions. The total output value is not reported in the 2008 data, but the other three firm size measures are all reported. The results are presented in Table 5. We can see that the results are in line with those in Table 4, but the estimates for $Migration\_Pct_i$ are larger in magnitude. Since $Migration\_Pct_i$ is held fixed in 2004 and 2008, this suggests that the provincial differences in terms of new firm sizes have increased from 2004 to 2008.

Table 5. Impact of migrant ratio on new firm size – 2008

	(1)	(2)	(3)	(4)	(5)	(6)
	log_income	log_income	log_employee	log_employee	log_asset	log_asset
Migration_Pct	-4.384***	-4.066***	-1.628***	-1.594***	-3.468***	-2.970***
	[0.072]	[0.076]	[0.035]	[0.036]	[0.055]	[0.059]
LogRatio	-0.010***		-0.008***		0.032***	
	[0.002]		[0.001]		[0.001]	
Migration_Pct *	-0.079***	-0.029*	0.037***	0.027***	-0.127***	-0.105***
LogRatio	[0.015]	[0.017]	[0.008]	[0.008]	[0.012]	[0.013]
Industry FE		Yes		Yes		Yes
R-squared	0.037	0.050	0.026	0.059	0.043	0.057
Observations	121,942	121,942	121,942	121,942	121,942	121,942

Note: Regression is at the firm level. The sample is restricted to new firms with age = 1. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

3.3 The capital intensity of new firms

Beyond firm size, we examine an additional measure of firm characteristics: capital intensity. We use firm-level $logratio$ as the dependent variable and see how it varies across provinces and industries. The regression model is similar to equation (1). We first use the 2004 data, and the results are presented in Table 6. Focusing on $Migration\_Pct_i$, one can see that the estimates vary a lot across columns (1)-(3). It is insignificant in column (1) but becomes significant and takes different signs in columns (2) and (3). A natural question is why. We believe that results in columns (1) and (2) are prone to two selection biases. In the first selection bias, new firms in more developed provinces seem more likely to select into labor-intensive industries. As a result, if this selection is not controlled for (via industry fixed effects), its impact will be picked up by the variable $Migration\_Pct_i$. The impact of this selection is so large that it overturns the actual positive impact of $Migration\_Pct_i$, which based on column (3) shows that new firms in more developed provinces tend to be more capital intensive.

Table 6. Impact of migrant ratio on new firm's capital intensity – 2004 data

	(1)	(2)	(3)	(4)	(5)	(6)
log_income				0.061***	0.061***	0.066***
				[0.002]	[0.002]	[0.002]
Migration_Pct	0.041	-0.351***	0.163***	0.285***	-0.114***	0.308***
	[0.033]	[0.042]	[0.043]	[0.032]	[0.041]	[0.042]
LogRatio	0.076***	0.083***		0.065***	0.072***	
	[0.001]	[0.001]		[0.001]	[0.001]	
Migration_Pct * LogRatio		-0.206***	-0.151***		-0.200***	-0.148***
		[0.013]	[0.014]		[0.012]	[0.013]
Industry FE			Yes			Yes
Observations	129,322	129,322	129,322	116,876	116,876	116,876
R-squared	0.035	0.037	0.090	0.036	0.038	0.093

Note: Regression is at the firm level. The sample is restricted to new firms with age = 1. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

The second selection relates to firm size. As we have shown earlier, new firms in more developed provinces tend to be smaller. To take this into account, we control for firm size in Columns (4)-(6) by adding the log of business income as an explanatory variable.[6] Once the firm size is controlled for, the provincial differences in terms of new firm capital intensity increase. In particular, the estimate for $Migration\_Pct_i$ almost doubles, increasing from 0.163 to 0.308.[7] The results suggest that the estimate for $Migration\_Pct_i$ in Column (3) is biased downward, because new firms in more developed provinces tend to be smaller, and smaller firms are more labor-intensive. Based on Column (6), a one percentage point increase in $Migration\_Pct_i$ leads to a 0.3% increase in the capital-labor ratio of new firms.

The coefficient of $LogRatio_j$ is positive and significant. This is obvious. Industries with more capital-intensive firms (new and old) are likely to have capital intensive new firms as well. Coefficients of the interaction terms are negative and significant, suggesting that the provincial differences shrink (increase) for more (less) capital intensive industries.

[6] We also use annual output value to measure firm size and the results are qualitatively the same.

[7] Tests show that the coefficients of Columns (3), (5), and (6) are significantly different from each other.

We also run similar regressions using the 2008 data, and the results are reported in Table 7. While the results for $LogRatio_j$ and $Migration\_Pct_i * LogRatio_j$ are comparable to those in Table 6, the estimates for $Migration\_Pct_i$ are reversed. This suggests an important change from 2004 to 2008. Consider an industry with average capital intensity ($LogRatio_j = 0$). Controlling for industry fixed effects and firm size, new firms from more developed provinces are more capital intensive relative to new firms from less developed provinces in 2004. This pattern has been reversed in 2008.

Table 7. Impact of migrant ratio on new firm's capital intensity – 2008 data

	(1)	(2)	(3)	(4)	(5)	(6)
log_income				0.053***	0.053***	0.054***
				[0.002]	[0.002]	[0.002]
Migration_Pct	-1.334***	-1.653***	-1.197***	-1.109***	-1.422***	-0.977***
	[0.037]	[0.040]	[0.041]	[0.037]	[0.040]	[0.042]
LogRatio	0.034***	0.042***		0.034***	0.042***	
	[0.001]	[0.001]		[0.001]	[0.001]	
Migration_Pct *		-0.172***	-0.139***		-0.168***	-0.138***
LogRatio		[0.008]	[0.009]		[0.008]	[0.009]
Industry FE			Yes			Yes
R-squared	0.022	0.025	0.056	0.030	0.033	0.064
Observations	121,942	121,942	121,942	121,942	121,942	121,942

Note: Regression is at the firm level. The sample is restricted to new firms with age = 1. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Taken together, results reported in section 3 paint an interesting contrast of new firm characteristics across provinces. Provinces with more developed economies see more firms entering, but their newly entered firms tend to be smaller in size relative to new firms in less developed provinces. The results are mixed on new firm capital intensity. Using 2004 data, new firms in more developed provinces tend to be more capital intensive, after controlling for firm selection in terms of sizes and industries, but the result is reversed when using the 2008 data.

Overall, these findings suggest that there is less friction in terms of firm entry (and possibly exit) in provinces with more developed economies. One explanation is that the more developed provinces have thicker markets or clustering. Thus, it is easier to secure capital and matching skill sets of labor to start a business there. Moreover, when a business fails, it may also be easier to liquidate (e.g., laid off workers can easily find other similar jobs in the same market). Alternatively, it may be due to differences in the policy environment. For example, there may be less red tape in more developed provinces, so that it is easier to start a new firm.[8] The same idea applies when a business fails and needs to be shut down. If there is less red tape in more developed provinces, we would expect more and smaller firms to enter and take their chances. In turn, one would also expect a higher exit rate of firms in these provinces, which will be explored and confirmed in the next section.

4. Firm quasi-dynamics

4.1 Firm exit

We now include firms of all ages into our regression analysis. Without a panel data, we cannot truly explore firms' exit patterns. Instead, we compare the number of firms at each age cohort and conduct a "quasi-dynamic" type of analysis on firm exit. For example, suppose the number of firms with Age = 3 is 10% less than the number of firms with Age = 2. Assuming that these two cohorts started with the same number of firms (year-over-year change of firm entry is likely to be small relative to firm exit), then we say that roughly 90% of the firms survived from Age = 2 to Age = 3 (or equivalently 10% exit rate).

To be more specific, we estimate the following model at the province-industry level, using the 2004 and 2008 data separately. In each year's data, Age is defined as the firm's age as of that year (2004 or 2008).

$$LogNum_{ij} = \alpha_0 + \alpha_1 Age + \alpha_2 Age^2 + \alpha_3 Age_1 + \beta_1 Age \times Migration\_Pct_i + \beta_2 Age \times LogRatio_j + \beta_3 Age \times Migration\_Pct_i \times LogRatio_j + \varepsilon_{ij}, (2)$$

[8] This fits into the general category of business friendliness which is frequently explored in cross-country analysis. A common measure is the number of calendar days needed to complete the procedures to legally operate a business.

where the dependent variable $LogNum_{ij}$ is the log number of firms within province i and industry j. Age_1 is a dummy variable which takes the value of 1 if the firm was new (Age = 1), and 0 otherwise.[9]

The results are presented in Table 8. It is not surprising that the number of firms decreases with Age at a decreasing rate ($\alpha_1 < 0, \alpha_2 > 0$). Our key variable of interest is the interaction term $Age \times Migration\_Pct_i$, which captures the regional differences in age-dynamics. The coefficient (β_1) is negative and significant, suggesting that the number of firms decreases faster in provinces with a higher migrant ratio (i.e., provinces with more developed economies). This is consistent with our earlier conjecture that entry and exit costs are lower in provinces with more developed economies. We also run similar regressions using the 2008 data, and the results (in Columns (3)-(4)) are similar to those using the 2004 data. We can see that the coefficient of Age (α_1) goes up slightly, but the coefficient of Age^2 (α_2) is only half of the size of the coefficient estimated using the 2004 data.[10] The coefficient for $Age \times Migration\_Pct_i$ (β_1) also changed significantly. Comparing Column (3) to Column (1), the magnitude of the coefficient is reduced to about one-third of the previous level, suggesting that the provincial differences in age dynamics have become much less pronounced.

Table 8. Firm quasi-dynamics: Firm exit

	(1)	(2)	(3)	(4)
	2004	2004	2008	2008
Age	-0.187***	-0.188***	-0.172***	-0.172***
	[0.003]	[0.003]	[0.002]	[0.002]

[9] We are concerned that not all firms started in 2004 made it to the data, possibly because those started in late 2004 did not have enough time to report in the Economic Census. Thus we include the Age_1 dummy to control for this problem. In Table 8, we can see that the coefficient of Age_1 is negative and significant, confirming prevalent under-reporting of newly established firms.

[10] It remains to be seen how this is driven by the following changes from 2004 to 2008 data. First, there are new firms entering in between 2004 and 2008 and their dynamics may be different from firms in the 2004 data. Second, some firms in the 2004 data did not survive to show up in the 2008 data, and the attrition process is likely to be non-random. Third, all firms in 2004 which survived to 2008 will have their age increased by 4, so they will be on different part of the age distribution.

Age^2	0.002***	0.002***	0.001***	0.001***
	[0.000]	[0.000]	[0.000]	[0.000]
Age_1	-0.447***	-0.447***	-0.853***	-0.853***
	[0.027]	[0.027]	[0.028]	[0.028]
$Age * Migration\_Pct$	-0.117***	-0.112***	-0.044***	-0.041***
	[0.007]	[0.007]	[0.006]	[0.006]
$Age * LogRatio$	0.001***	0.001***	0.001***	0.001***
	[0.000]	[0.000]	[0.000]	[0.000]
$Age * LogRatio$		0.003***		0.003***
$* Migration\_Pct$		[0.001]		[0.001]
Province FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
R-squared	0.817	0.817	0.824	0.824
Observations	19,677	19,677	22,106	22,106

Note: Regression is at the province-industry-age level. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

4.2 Firm "growth"

We then use firm-level data to estimate the impact of migrant ratio on firm "growth". We adopt an empirical specification similar to equation (2). However, the observations are now at the firm level (rather than the province-industry level), and the dependent variables are various measures of firm size. The estimation results using the 2004 data are reported in Table 9. From Column (1), we can see that the coefficient of Age is positive and significant. Ignoring Age^2, this suggests that firm size, measured in the log of total output, increases with age. Our estimate seems to suggest an annual "growth" rate of about 6% (note that the dependent variable is in natural log). However, given that firms exit over time and smaller firms are more likely to exit, the average size of surviving firms should increase even without true "growth".[11] The coefficient of Age^2 is always negative and significant, suggesting that firm size increases with age at a decreasing rate.[12]

Table 9. Firm quasi-dynamics: firm growth using the 2004 data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)

[11] For simplicity, we state that firm size increases with age, but also acknowledge that this "growth" process is a mixture of organic growth and selective exit.

[12] Based on the results in Column (1), firm size may eventually decrease with age. The cutoff age can be calculated as follows. 0.06 + 2 (-0.003) * Age = 0, suggesting that Age = 10.

	log_output	log_output	log_income	log_income	log_employee	log_employee	log_asset	log_asset
Age	0.060***	0.060***	0.002	0.002	0.030***	0.030***	0.069***	0.071***
	[0.005]	[0.005]	[0.003]	[0.003]	[0.003]	[0.003]	[0.005]	[0.005]
Age^2	-0.003***	-0.003***	-0.001***	-0.001***	-0.001***	-0.001***	-0.002***	-0.002***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Age_1	-0.611***	-0.611***	-0.490***	-0.491***	-0.116***	-0.116***	-0.228***	-0.227***
	[0.023]	[0.023]	[0.020]	[0.020]	[0.008]	[0.008]	[0.015]	[0.016]
Age	0.210***	0.210***	0.114***	0.120***	0.107***	0.104***	0.193***	0.159***
$*$ Migration_Pct	[0.015]	[0.021]	[0.011]	[0.014]	[0.010]	[0.014]	[0.015]	[0.021]
Age $*$ LogRatio	0.002***	0.002***	-0.002***	-0.002***	0.001	0.001*	0.003***	0.004***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.001]	[0.001]
Age $*$ LogRatio		-0.000		0.003		-0.002		-0.016***
$*$ Migration_Pct		[0.005]		[0.003]		[0.004]		[0.005]
Province FE	YES	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	0.084	0.084	0.086	0.086	0.091	0.091	0.089	0.089
Observations	1,141,818	1,141,818	919,562	919,562	1,141,887	1,141,887	1,141,886	1,141,886

Note: Regression is at the firm level. Standard errors are clustered at the province-industry level. *** p<0.01, ** p<0.05, * p<0.1.

Our key variable of interest is the interaction term $Age \times Migration\_Pct$. Result in column (1) suggests that a 1 percentage point increase in $Migration\_Pct$ will further raise the firm growth rate by 0.21 percentage point, from 6% to 6.21%. We also use other measures of firm size (all taking natural logs), including business income, number of employees, and total assets. The results, presented in Columns (3) - (8), are quite similar. We also control of industry capital intensity, including the triple interaction term. However, whether we control for the triple interaction term does not seem to make much difference.

We then repeat the estimation using the 2008 data. The results are presented in Table 10, which are similar to Table 9. The age-size pattern of firms remains. As for the key variable we are interested in, $Age \times Migration\_Pct$, the coefficient remains positive and significant.

Table 10. Firm quasi-dynamics: firm growth using the 2008 data

	(1)	(2)	(3)	(4)	(5)	(6)
	log_income	log_income	log_employee	log_employee	log_asset	log_asset
Age	0.101***	0.101***	0.053***	0.053***	0.091***	0.092***

	[0.004]	[0.004]	[0.003]	[0.003]	[0.005]	[0.005]
Age^2	-0.003***	-0.003***	-0.001***	-0.001***	-0.003***	-0.003***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Age_1	-0.754***	-0.753***	-0.128***	-0.128***	-0.219***	-0.219***
	[0.030]	[0.030]	[0.011]	[0.011]	[0.017]	[0.017]
Age	0.239***	0.243***	0.144***	0.142***	0.275***	0.264***
$*$ Migration_Pct	[0.015]	[0.018]	[0.012]	[0.015]	[0.017]	[0.020]
Age $*$ LogRatio	0.003***	0.003***	0.001***	0.001***	0.003***	0.003***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Age $*$ LogRatio		0.002		-0.001		-0.005*
$*$ Migration_Pct		[0.003]		[0.003]		[0.003]
Province FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
R-squared	0.090	0.090	0.091	0.091	0.074	0.075
Observations	1,686,820	1,686,820	1,686,820	1,686,820	1,686,820	1,686,820

Note: Regression is at the firm level. Standard errors are clustered at the province-industry level. *** p<0.01, ** p<0.05, * p<0.1.

In sum, the quasi-dynamics analyses show that firm size increases faster in provinces with a higher migrant ratio (i.e., provinces with more developed economies). This is consistent with our earlier finding that more and smaller firms enter in such provinces, and non-performing firms would exit at a faster rate. Putting together, these results suggest that the overall business environment and factor mobility are quite different across provinces.

5. Linking the 2004 and 2008 datasets

So far, we have utilized data from 2004 and 2008 separately. Next, we link these two datasets together. We do so in two ways. First, we compare firms of the same age (e.g., $Age = t$) in the two data sets. Note that these are not the same firms since a firm with $Age = t$ in 2004 would have $Age = t + 4$ in 2008 conditional on survival. Even though these are different firms, such a comparison tells us how the firms in the same stage of their life cycles from the two datasets compare with each other. Second, we compare firms with $Age = t$ in 2004 to firms with $Age = t + 4$ in 2008. The latter should be a subset of the former.

5.1 Compare firms in the same age group

We introduce a dummy variable $Year\_2008$ which takes value 1 if the data comes from 2008 and 0 otherwise. We first calculate two variables are the ijt-level, for province i, industry j and age t. $LogNum_{ijt}$ is the log number of firms, while $LogRatio_{ijt}$ is the log capital-labor ratio. In order to reflect the different stages of a firm's life cycle, we divide firms into 4 groups based on their age (age = 1, age between 2 and 5, age between 6 and 10, and age > 10) and introduce three age group dummies $Age1$, $Age2\_5$, and $Age6\_10$. The reference group then consists of all firms with $Age > 10$. We regress $LogNum_{ijt}$ and $LogRatio_{ijt}$ on a set of explanatory variables, including the age group dummies, the year dummy, and their interactions. The results are presented in Table 11 Columns (1)-(2).

Table 11. Comparing firms in the 2004 and 2008 datasets with the same age

	(1)	(2)	(3)	(4)	(5)
	LogNum	LogRatio	log_income	log_employee	log_asset
Year_2008	0.387***	0.444***	-0.903***	-0.060***	0.450***
	[0.015]	[0.020]	[0.032]	[0.010]	[0.017]
Age1	-0.494***	-0.346***	-0.436***	-0.436***	-0.673***
	[0.025]	[0.023]	[0.029]	[0.022]	[0.044]
Age2_5	0.870***	-0.091***	0.084***	-0.250***	-0.307***
	[0.022]	[0.021]	[0.018]	[0.014]	[0.025]
Age6_10	0.304***	0.037*	0.103***	-0.129***	-0.080***
	[0.016]	[0.022]	[0.014]	[0.012]	[0.015]
Age1 * Year_2008	-0.442***	0.117***	-0.937***	-0.158***	-0.206***
	[0.023]	[0.029]	[0.049]	[0.020]	[0.026]
Age2_5 * Year_2008	-0.113***	0.059**	-0.500***	-0.100***	-0.159***
	[0.021]	[0.024]	[0.029]	[0.013]	[0.018]
Age6_10 * Year_2008	0.199***	-0.033	-0.185***	-0.024**	-0.077***
	[0.020]	[0.028]	[0.019]	[0.012]	[0.016]
Province FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES
R-squared	0.865	0.543	0.160	0.084	0.077
Observations	7,055	7,043	2,606,382	2,828,707	2,828,706

Note: Regression is at the province-industry-age-cohort level for Columns (1) – (2) and at the firm level for Columns (3) – (5). Standard errors are clustered at the province-industry level. *** p<0.01, ** p<0.05, * p<0.1.

For Column (1), the coefficient of $Year\_2008$ is 0.39, suggesting that the reference group (firms with $Age > 10$) has 39% more firms in 2008 relative to 2004. For other age groups, we need to add up the coefficients of $Year\_2008$ and the interaction term. For the age between 2 - 5 and age between 6 - 10 groups, there are about 28% (= 0.39 - 0.11) and 59% (0.39 + 0.20) more firms in 2008, respectively. Moving onto Column (2), we can see that capital intensity has significantly increased from 2004 to 2008 for all age groups. This finding is also consistent with results later on from Columns (4) and (5), which show that firms with the same age use more capital and employ fewer workers. In other words, firms are generally more capital intensive in 2008 than in 2004.

We also run regressions at the firm level, using firm size variables such as *log_income*, *log_employee,* and *log_asset*. The results are presented in Table 11, Columns (3)-(5). We can see that compared to firms in the same age group in 2004, firms in 2008 earn significantly less income, use more capital (higher assets) and hire fewer employees. This is true for every age group (whether we look at the coefficient of $Year\_2008$ alone or add it to the coefficient of the corresponding interaction term). Note that, as firm age goes up, the magnitude of the interaction term coefficient goes down. This suggests that the gap between firms in the 2004 and 2008 data shrinks when age increases. Take Column (4) for example. For firms over the age of 10, the number of employees was 6% less on average in 2008. In contrast, the number of employees was 21.8%/16%/8.4% less on for firms in *Age1*, *Age2_5*, and *Age6_10* groups, respectively.

5.2 Compare the same group of firms before and after 4 years

Next, we compare firms with $Age = t$ in the 2004 dataset with firms with $Age = t + 4$ in the 2008 dataset. The latter, due to firm exit, should be a subset of the former. Note that the age dummies are defined differently than in Section 5.1. Let us see this through an example. Consider a firm with $Age = 3$ in 2004 which survives to 2008. In Section 5.1, this firm would have $Age2\_5 = 1$ in 2004 but $Age6\_10 = 1$ in 2008. In contrast, here, we use firm age in 2004 to define the age dummies in both 2004 and 2008. As a result, this firm continues to have $Age2\_5 = 1$ in 2008, which means that this firm's age back in 2004 was between 2 and 5.

Similar to Section 5.1, we compare firms in the two datasets in terms of firm number, capital intensity, and firm size. This exercise allows us to examine true industry dynamics, even though we only observe the same firm twice (in 2004 and 2008 respectively). Table 12 reports the results. From Column (1), one can see that the sum of the coefficients of $Year\_2008$ and the corresponding interaction term is always negative. That is, the number of firms goes down in each age group, which is not surprising. We do not look at firms with $Age = 1$ due to under-reporting. Column (2) suggests that firms surviving from 2004 to 2008 become more capital intensive. This may be because firms become more capital intensive as they age, or because less capital-intensive firms are more likely to exit. As for the number of workers, the magnitude of the coefficient is fairly small as shown in Columns (4). However, Column (5) shows that firms are using significantly more capital (higher assets) after 4 years. It is clear that firms have become more capital intensive, which is consistent with the findings of previous tables.

Table 12. Comparing firms with age t in 2004 and firms with age t + 4 in 2008

	(1)	(2)	(3)	(4)	(5)
	LogNum	LogRatio	log_income	log_employee	log_asset
Year_2008	-0.403***	0.332***	-0.911***	-0.002	0.471***
	[0.012]	[0.022]	[0.033]	[0.008]	[0.015]
Age1	-0.550***	-0.424***	-0.443***	-0.438***	-0.678***
	[0.023]	[0.023]	[0.030]	[0.023]	[0.044]
Age2_5	0.762***	-0.169***	0.077***	-0.252***	-0.314***
	[0.019]	[0.021]	[0.018]	[0.014]	[0.024]
Age6_10	0.209***	-0.041*	0.101***	-0.129***	-0.082***
	[0.014]	[0.022]	[0.014]	[0.011]	[0.015]
Age1 * Year_2008	0.574***	0.402***	0.259***	0.145***	0.384***
	[0.017]	[0.030]	[0.026]	[0.012]	[0.025]
Age2_5 * Year_2008	0.307***	0.212***	-0.162***	0.035***	0.124***
	[0.015]	[0.025]	[0.021]	[0.007]	[0.016]
Age6_10 * Year_2008	0.199***	0.094***	-0.101***	0.011*	0.019**
	[0.013]	[0.027]	[0.017]	[0.006]	[0.009]

Province FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES
R-squared	0.863	0.553	0.125	0.080	0.096
Observations	7,002	7,002	1,943,095	2,165,420	2,165,419

Note: Standard errors are clustered at the province-industry level. *** p<0.01, ** p<0.05, * p<0.1.

6. Discussions and policy implications

The key findings of this paper are as follows. In provinces with higher migrant ratios (i.e., provinces with more developed economies), it is easier for a prospective firm to enter the market and for a non-performing firm to exit. As a result, these provinces see more firm entry, and their new firms tend to be smaller and more labor-intensive. Over time, partly because non-performing firms exit at faster rates in these provinces, remaining firms grow faster relative to firms in provinces with lower migrant ratios. The whole scene is more dynamic in more developed provinces, driven by lower entry/exit costs there.

Why may entry/exit costs be lower in more developed economies? One reason may be due to market factors such as thick markets or clustering. Let us consider skilled labor as an example. A firm entering into an industry may need specific types of skilled labor. In a thick market, it is easy for the firm to hire the exact types of labor it needs. Similarly, if the firm fails, its workers can easily find other employment opportunities, so the firm can easily exit. Less developed provinces have smaller scales of economies and naturally will have thinner markets of essential resources. We see this happening across countries as well. China, with its thick/clustered market of various inputs needed for manufacturing industries, enjoy significant advantage over many other developing countries which, despite lower costs (wages and other resources), lack thick/clustered markets of inputs. Another reason relates to policy factors such as red tape (or even corruption). More developed provinces are likely to have better institutions, more transparent rules etc. This would reduce the cost of doing business, including entering/exit a market.

Distinguishing and isolating the market and policy factors have important policy implications. If provincial differences in terms of industry dynamics are mainly due to policy factors (e.g., red tape), then the government may be able to influence or even control them. However, if the driving force is market factors such as thick market or clustering, then it will be more difficult for the government to intervene. One possibility is for less developed provinces to specialize in a subset of industries. This would allow them to cultivate a thick market of resources (specialized capital and skilled labor, management, etc.) in specific industries. Then it will be easier for a firm to enter into the market and for a non-performing one to exit. However, concentrating on specific markets can be quite risky, and ex-post wrong decisions can put the government in tough spots. In addition, the government may be poorly informed or equipped to find out what specific industries to focus on. Counterintuitively, when the thick market is essential but not met in less developed provinces, the second-best for the government sometimes may be to focus on capital intensive firms/industries, possibly through state-owned enterprises.

Our analysis directly touches on migration and industrial transition across regions in China. First, presumably, a region is competitive if it attracts entrepreneurs to create jobs. This is the goal of many Chinese cities in implementing a variety of talent-attracting programs. Our study shows that migration-attracting regions indeed enjoy more firm entry. Second, the relocation of industries from the east coast to the west hinterland is an inevitable phenomenon, expected to continue for a while. Our study suggests that, in addition to factor prices, the clustering effect induced by the economy of scale/scope is also important. Complementarity within industries and among related industries could make factor markets thicker, creating dynamic markets for firm entry and growth.

An important question for future research is how to isolate the role of market and policy factors. If we are willing to accept that state-owned enterprises (SOEs) face comparable red-tape across regions, then provincial differences of the SOEs would allow us to isolate the role of market factors. One may also directly explore the role of market factors. For example, some industries may only require general labor, so even in less developed provinces, the market for such labor is thick enough. Thus, firm entry/exit is pretty easy from labor's perspective. Ignoring the role of capital, we should see smaller provincial differences in terms of industry dynamics (entry, growth, and exit) for these industries. In fact, such industries can easily thrive in less developed provinces, and relatedly, we would see few such firms in more developed provinces due to higher labor costs.

Another topic for future research is to link the cross-region differences in firm dynamics with firm R&D and innovation. How do small firms' R&D input and output compare with larger, established firms? Are there significant cross-region differences in R&D and innovation for similar-sized firms? If so, what are the driving forces? Understanding these questions is necessary to help with the industrial upgrading of the Chinese economy.

Appendix

We clean the firm level data as follows.

1) Drop observations that are reported to be established before the year of 1978.
2) Drop stated-owned firms.
3) Drop firms in natural resources and mining. Since firm quasi-dynamics is the primary issue we want to shed light on, we focus on sufficiently mobile industries. Firms in natural resources and mining (e.g., a coal mine), by nature, are likely to be immobile.
4) Having winsorization procedure to 2004 and 2008 data, respectively. Firstly we calculate the capital-labor ratio for each firm, which is defined as the total assets of the firm divided by total number of employees. Secondly, we group all the firms in to "province-industry (2-digit)" level, so the data will be separated into (31*30) cells. We then drop off the top and bottom firms at each cell, which we suspect are likely due to misreporting or punch errors. The exact cutoffs depend on the size of the province-industry cell. If the cell has more than 200 firms, we drop the top 2% and the bottom 2% observations. If the cell has between 50 and 200 firms, then we drop the top and the bottom 1% observations; If the cell has fewer than 50 firms, then we do not drop any firm. After winsorizing, we have a total number of 2,849,389 observations:

1,156,416 in 2004, and 1,692,973 in 2008 (For contrast, before winsorizing, the numbers are 2,998,059/ 1,209,655/ 1,788,404).

5) We also excluded firms with capital = 0 (less than 0.01%).

References

Aghion, Philippe, Cai Jing, Mathias Dewatripont, Du Luosha, Ann Harrison, Patrick Legros. 2015. "Industrial Policy and Competition." *American Economic Journal: Macroeconomics*, 7(4), 1-32.

Bai, C., Y. Du, Z. Tao, and S. Tong. 2004. "Protectionism and Regional Specialization: Evidence from China's Industries." *Journal of International Economics* 63, 397-417.

Barwick, Panle Jia, Myrto Kalouptsidi, Nahim Bin Zahur. 2019. "China's Industrial Policy: an Empirical Evaluation." NBER Working Paper No. 26075.

Brandt, Loren, Johannes Van Biesebroeck, Yifan Zhang. 2012. "Creative Accounting or Creative Destruction: Firm Level Productivity Growth in Chinese Manufacturing." *Journal of Development Economics* 97(2), 339–351.

Brandt, Loren, Gueorgui Kambourov, Kjetil Storesletten, 2018. "Barriers to Entry and Regional Economic Growth in China." Manuscript, University of Toronto.

Birch, David. 1979. *The Job Generation Process*. MIT Program on Neighborhood and Regional Change, Vol. 302, pp. 1979. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1510007

Chen, Zhao, Zhikuo Liu, Juan Carlos Suárez Serrato, Daniel Yi Xu. 2019. "Notching R&D Investment with Corporate Income Tax Cuts in China." NBER Working Paper No. 24749.

Davis, Steven, John Haltiwanger, and Scott Schuh, 1996. Job *Creation and Destruction*. Cambridge, MA: MIT Press.

Deng, P., & Jefferson, G., 2011. "Explaining spatial convergence of China's industrial productivity." *Oxford Bulletin of Economics and Statistics*, 73(6), 818 – 832.

Dunne, T., M. Roberts, L. Samuelson, 1988. "Patterns of firm entry and exit in US manufacturing industries." *Rand Journal of Economics*, 495-515.

Dunne, T., M. Roberts, L. Samuelson, 1989. "The growth and failure of US manufacturing plants." *Quarterly Journal of Economics* 104 (4), 671-698.

Ju, Jiandong, Zhengwen Liu, 2019. “Too Big to Exit: Does the Market Select More Productive Firms.” Manuscript, Tsinghua University.

Kojima, Kiyoshi. 2000. “The “flying geese” model of Asian economic development: origin, theoretical extensions, and regional policy implications” *Journal of Asian Economics*, 11, 375-401.

Lemoine, Françoise, Sandra Poncet, Deniz Ünal. 2015. “Spatial rebalancing and industrial convergence in China” *China Economic Review* 34, 39-63.

Li, Dongya and Yi Lu. 2011. “Industrial agglomeration and firm size: Evidence from China” *Regional Science and Urban Economics*.

Long, Cheryl, and Xiaobo Zhang. 2011. “Cluster-based Industrialization in China: Financing and Performance.” *Journal of International Economics* 84, 112-123.

Lu, Jiangyong, and Zhigang Tao. 2009. "Trends and Determinants of China’s Industrial Agglomeration." *Journal of Urban Economics* 65, 167-180.

Qu, Yue, Fang Cai and Xiaobo Zhang. 2012. “Has the ‘Flying Geese’ Phenomenon in Industrial Transformation Occurred in China?” in *Rebalancing and Sustaining Growth in China*, Huw McKay and Ligang Song (eds), Australian National University Press.

Shi, Xiangyu, Tianyang Xi, Xiaobo Zhang, Yifan Zhang, 2018. “Moving ‘Umbrella’: Bureaucratic Transfers, Collusion, and Rent-seeking in China.” CCER working paper E2018011, National School of Development, Peking University.

Klapper, Leora, Luc Laeven, and Raghuram Rajan. 2004. “Business Environment and Firm Entry: Evidence from International Data.” Manuscript, University of Chicago.

Wang, Yang. 2019. “Skill Complementarity in Teams: Matching, Sorting and Agglomeration in China.” Manuscript, University of California San Diego.

Wu, Haoyi, Huanxiu Guo, Bing Zhang, Maoliang Bu. 2017. “Westward movement of new polluting firms in China: Pollution reduction mandates and location choice.” *Journal of Comparative Economics* 45, 119-138.

Yang, Yang. 2018 “Transport Infrastructure, City Productivity Growth and Industry Reallocation: Evidence from China.” IMF Working Paper 18-276.