What Determines Entrepreneurial Outcomes in Emerging Markets? The Role of Initial Conditions

Meghana Ayyagari **Asli Demirguc-Kunt** **Vojislav Maksimovic**

February 2016

What determines a firm's performance in its initial years? We ask how firm characteristics and the institutional environment predict a firm's size and growth over its early lifecycle. Using census data from India, we find that initial firm size is remarkably persistent. Large and small entrants do not grow at different rates across states with different institutions or industries with differing reliance on external finance. We do find that greater financial development is associated with higher entry rates and on average smaller sized entrants. However, there is little evidence that these smaller entrants grow relatively faster than larger entrants.

---

*Ayyagari: School of Business, George Washington University, ayyagari@gwu.edu, Ph: 202-994-1292; Demirgüç-Kunt: World Bank, ademirguckunt@worldbank.org, Ph: 202-473-7479; Maksimovic: Robert H. Smith School of Business at the University of Maryland, vmaksimovic@rhsmith.umd.edu, Ph: 301-405-2125. We would like to thank Paolo Bastos, Miriam Bruhn, Thomas Chemmanur, Roberto Fattal, Todd Gormley, David McKenzie, and seminar participants at the Cass Business School, London, the Robert H. Smith School of Business at the University of Maryland, University of Bristol, University of Exeter, Purdue University, CEMP-CIEPS-HKUST IEMS Workshop, and the participants at the 2014 Finance, Organization, and Markets Research Group Symposium at USC, and the Fourth Symposium on Emerging Financial Markets: China and Beyond for their comments and suggestions. We also thank Xiaoyuan Hu for outstanding research assistance. This paper's findings, interpretations, and conclusions are entirely those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

I. Introduction

What determines a firm's performance in its initial years? Do successful firms possess certain characteristics that distinguish them when they start up or do they owe their success mostly to the institutional and regulatory environment that enables their productivity and growth? There is a large literature that establishes that firms, particularly small firms, are likely to be more productive and grow faster in developed institutional environments with easier access to finance, stronger legal protections, and lack of corruption (e.g. Beck, Demirguc-Kunt, and Maksimovic (2005), Demirguc-Kunt and Maksimovic (1998) and Rajan and Zingales (1998)). More recently, another strand of the literature has emphasized the importance of factors intrinsic to the firm – such as managerial vision and comparable best practices - in influencing firm growth and productivity across countries (Bloom and Van Reenen (2007, 2010), Bruhn, Karlan and Schoar (2010) and Bloom et al. (2013)). However, very little is known about the corporation as an entity during its founding years and how institutions and initial firm characteristics influence entrepreneurial outcomes in these crucial early years.

In this paper, we investigate whether the founding conditions of a firm predict success over its first eight years in different financing and institutional environments and for different industries. Specifically, we first examine which of the following initial conditions – startup size (number of employees), startup productivity, and legal form (public limited or private limited company) – predict an entrant's growth trajectory relative to other entrants. Next, we examine whether there is heterogeneity in the relations between these founding conditions and growth across different industry technologies, financing needs, and different local financial and labor market institutions. Finally, we investigate the role of institutions in influencing the characteristics of startups in different industries.

We use census data from the Annual Survey of Industries (ASI), which is the primary source of data on manufacturing firms in the formal sector in India, for the period 2001-2010. We follow firms through eight years of their early life cycle and all the initial conditions are defined when the firm is one year old. The novel component of this empirical design is that since we are observing firms right from their entry, we can consider the initial conditions to be truly exogenous with respect to subsequent outcomes.

India offers an ideal laboratory for testing the role of institutions on firm lifecycle given the large "persistent" differences in institutions, business environment, and income across different regions in India (Ahluwalia (2002)). There is substantial and well researched heterogeneity in financial and labor institutions across the different states in India (e.g. Burgess and Pande (2005), Besley and Burgess (2004), Hasan, Mitra, and Ramaswamy (2007)). Given the short time period of our sample, we focus on the cross-sectional variation in institutions across Indian states. At the same time, comprehensive Census data at the firm level are available to researchers, thereby sidestepping many of the concerns arising from data comparability in cross-country studies.

We uncover several surprising findings. First, we find that initial size at start up is a strong predictor of persistence in firm size over early life-cycle. Thus firms born large (small) remain relatively large (small) over early lifecycle. Start-up size is more important than initial productivity and whether the firm is organized as a public or private limited company, in explaining the evolution in average size over the early lifecycle. This persistence also implies that there is no significant difference in the growth rates of small and large entrants over the first eight years.

Second, the size differential and the similarity in growth rates across firm sizes are remarkably stable and robust to institutional differences across states in India. They are unaffected by the level of credit provided by each state's banking system, the strictness of labor regulations, the quality of local business regulations and by a general indicator often used for the quality of business conditions, income per capita. The size differential and the similarity in growth rates are also unaffected by industry dependence on external finance as defined by Rajan and Zingales (1998), industry production structure (labor versus capital intensive) and industry growth rates. Moreover, there is no evidence that the growth rates of more productive small entrants relative to those of less productive small entrants differ across states with the development of the local financing system. Thus, relative size ranking is not affected by industries or institutions. Entry size serves as the blueprint for the typical firm's size during its early lifecycle.

We also find that entrants differ along other characteristics. Large entrants have more complex production structures, which is suggestive of them having more skilled or efficient managers. Entrants with high initial productivity have higher future productivity and profits suggesting that increased efficiency of entrants translates into better financial performance.

Our results are robust to alternate definitions of large versus small entrants including using a continuous measure of firm size. We also do not find threshold effects in entry size where firms above (or below) a certain threshold grow faster (or slower).[1]

One of the concerns with our finding above could be "selection of the largest firms" induced by the exit of small firms that are unable to survive the competition, rendering the

[1]Ayyagari, Demirguc-Kunt, and Maksimovic (2015) use cross-sectional data across a number of countries and show that the size at start up explains the most variation in size, growth, and productivity across 40 years in over 100 developing countries. The focus in that paper is not on early life-cycle or in exploring the persistence of initial conditions.

growth rates of surviving small entrants and large entrants similar. To address this, we follow the structural approach in Combes, Duranton, Gobillon, Puga, and Roux (2012) to parameterize the strength of the selection forces operating in our data. Strong selection forces should result in fewer small firms surviving leading to a left truncation of the size distribution. We find no evidence that the size distribution at the end of the early lifecycle is a truncated version of the initial size distribution. Overall, we are confident that our results on initial size are not being driven by our inability to track small firms over time, either due to their inability to survive as they get older or attrition from the census count. [2]

Finally, when we examine the entry process further, we find that average start-up size is strongly affected by the quality of the financial system and labor regulations. In Indian states with stronger credit availability, there is more entry and the average entrant is also smaller.[3] By contrast, average initial productivity of entrants is not affected by the availability of credit, but by business and labor regulations. These results suggest that firms that enter the formal sector when institutions are poorly developed are on average larger and have higher productivity, presumably to be able to overcome financing and regulatory obstacles and still be viable.

Overall, our results show that the initial rate of entry and entry size are sensitive to local institutions. Upon entry, however, the initial conditions of the entrants are remarkably persistent. Small and large firms grow at the same rate across different industries and institutions. There is little evidence of more productive small firms entering small and increasing their relative size over time, as would be expected if they were relying on retained earnings to make up for failures of the banking system to finance expansion.

[2] Our results are also robust to sensitivity analyses where we correct for exit by varying the growth rate of exiting firms between 0 and -100%. More details are in section III.B.

[3] Below we investigate the extensive margin and find that this is the result of greater entry overall, with a larger increase in the entry rate by smaller entrants.

The findings in our paper on the persistence of initial size relate to the literature documenting long-run persistence in firm specific outcomes such as capital structure choices and the underlying mechanisms driving this persistence. For instance, Lemmon, Roberts, and Zender (2008) show that leverage ratios in the US are stable over time so that firms with relatively high (low) leverage maintain relatively high (low) leverage for over 20 years and this persistence is driven by an unobserved time-invariant firm effect, which they do not explore.[4] Our paper posits that founding conditions of firms, which are determined by local institutions, could be a potential source of persistence in firm performance. Other studies such as Malmendier and Nagel (2001) and Malmendier, Tate, and Yan (2011) attribute the firm fixed effect to managerial traits such as overconfidence and effect of past experiences that have long run impacts on corporate policies. This is part of a growing literature in corporate finance emphasizing the importance of "managerial fixed effects" for a firm's decisions and performance.[5] Others have emphasized the importance of managerial capital (or lack thereof) in developing countries (Bloom and Van Reenen (2007, 2010), Bruhn, Karlan, and Schoar (2010)) and in particular in India (Bloom et. al. (2013)). These papers document the persistence of dysfunctional managerial styles in firms and posit that the variations in management practices have broader implications for firm growth and productivity differences across countries. All these papers focus on managerial characteristics of continuing firms and not on characteristics at the time of startup. Our results suggest that measurable initial factors intrinsic to the firm such as managerial capital may be more important in explaining growth patterns rather than institutions across different countries and industries. We also find that firms with larger entry size are also those with more complex production

[4]DeAngelo and Roll (2015) question the stability of capital structures and show that leverage cross-sections more than a few years apart differ markedly.

[5] See, for instance, Bertrand and Schoar (2003), PérezGonzález (2006), Bennedsen, Nielsen, Pérez-González, and Wolfenzon (2007), Kaplan, Klebanov, and Sorensen, (2012), Cronqvist, Makhika and Yonker (2012), Graham, Li, and Qiu (2012), and Benmelech and Frydman (2015) among others.

processes, suggesting that initial size may be capturing the managerial capacity of the entrepreneur. [6]

Second, a large literature has focused on the importance of institutions for economic growth. One stream of papers provide cross-country evidence on how well developed financial institutions (measured by banking sector depth, number of listed firms per capita, and overall size of the stock market) lead to greater economic growth (e.g. La Porta, Lopez-de-Silanes, and Shleifer (2002), La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1997))[7] , better firm performance and governance (e.g. Demirguc-Kunt and Maksimovic (1998), Beck et al.(2005), Dyck and Zingales (2004), Doidge, Karolyi, and Stulz (2007)) among others) and higher entry rates (Guiso, Sapienza, and Zingales (2004), Michelacci and Silva (2007), and Klapper, Laeven, and Rajan (2006)). A second related stream of papers focuses on institutional change, documenting the positive impact of structural banking reforms in the US, France and other countries on survival and better performance of more efficient firms ((Black and Strahan (2002), Cetorelli and Strahan (2006), Bertrand, Schoar and Thesmar (2007), Kerr and Nanda (2009)).

The focus of our paper is complementary to the institutions literature, but is different in two key dimensions. First, unlike most of the papers above, we limit our analysis to firm entry and to the early years of a firm's lifecycle. Thus, while the prior literature has established the primacy of institutions for the growth rates of more mature firms, our paper suggests that for entrants, institutions matter but only in the selection of firms. Second, given the short period and relative stability in the financial sector, we are focused on the cross-sectional variation in financial development across Indian states rather than an institutional change. Importantly, we

[6] See Lucas (1978) or Rauch (1991) for models of the relation between managerial talent and initial conditions.
[7] See Levine (2005) for a review of this literature.

show that at least over the first eight years of firms' lifecycle, initial starting conditions dominate the effect of institutions in influencing the growth trajectory.

Our findings on the selection process at entry reflect the extensive and intensive margin forces at play. We find greater financial development is associated with higher entry rates (extensive margin) but much of this entry is smaller sized entry. Thus, when institutions are poor, firms have to be larger to survive. Other studies have reported similar results in the US (e.g. Kerr and Nanda (2009, 2010), Cetorelli (2004)) but unlike our paper, their focus is on institutional change. They find that US banking reform led to increased entry of small firms and thus reduced average entry size. Similarly, Branstetter, Lima, Taylor, and Venancio (2014) look at the effect of entry de-regulation reform in Portugal and find that while the reform increased the number of business start-ups, the start-ups established after the program were smaller and more "marginal firms" . The authors interpret these results as suggesting that the barriers prior to deregulation were less of an impediment to larger, high-quality firms. Doidge, Karolyi, and Stulz (2015) focus on new listings on US capital markets and find that over the period 1997-2012 during which there has been a tightening of regulatory and listing standards, the US has seen fewer fewer small listings. Other papers have also emphasized the importance of strong ex-ante fundamentals though they do not focus on entry characteristics per se. For instance, Cetorelli and Strahan (2006) find that within manufacturing average firm size declines after US banking reforms. Zingales (1998) shows that after an important piece of deregulation in the trucking industry, firms with the best chances of survival were those that were more efficient but also those with ex ante stronger financial fundamentals (lower leverage).

Our paper also contributes to the entrepreneurship and venture capital literature that focuses on, for the most part, US entrepreneurs,[8] (see, for example, Bhide (2000), Hellmann and Puri (2000), Kaplan, Sensoy, and Stromberg (2009), Chemmanur and Fulghieri (2014)). From these studies we know that young firms face specific challenges and that a venture capital industry has arisen to assist a small proportion of firms judged to be most likely to be successful. Our paper, on the other hand focuses on what determines the outcomes of a broader subsection of young firms in developing countries which do not have specialized institutions such as a well-developed venture capital market to cater to young firms as is the case in India.

II. Data

A. Indian Manufacturing Census

We use panel data for the period 2001-2010 on formal manufacturing plants in India from the Annual Survey of Industries (ASI), which is conducted by the Indian Ministry of Statistics and Program Implementation.[9] The ASI sampling frame consists of all registered factories employing 10 or more workers using power or 20 or more workers without using power.[10] The sampling frame consists of the "Census" sector which are surveyed every year (typically plants

[8] Recently, a few studies have started looking at international investments in venture capital such as Chemmanur, Hull, and Krishnan (2010).

[9] The ASI also contains some establishments outside of manufacturing. Thus, while the primary unit of enumeration in the survey is a factory in the case of manufacturing industries, it could also be a workshop (for repair services), an undertaking or a licensee (electricity, gas & water supply undertakings) or an establishment (bidi & cigar industries). According to the Ministry of Statistics, "the owner of two or more establishments located in the same State and pertaining to the same industry group and belonging to census scheme is, however, permitted to furnish a single consolidated return. Such consolidated returns are common feature mostly in the case of bidi and cigar establishments, electricity and certain public sector undertakings"

[10] As seen in the summary statistics, we have a number of firms that report < 10 employees - these are firms that do not need to be registered but are nevertheless registered. Several papers such as Bedi and Banerjee (2007), Harrison, Martin, and Nataraj (2012), and Chatterjee and Kanbur (2013) have noted this phenomenon and proposed several explanations including administrative hurdles in the way of de-registration, temporary squeezes in employment, voluntary disclosures to signal plans for future growth that that do not affect the interpretation of our results. Our results are robust to excluding these firms.

having 100 or more workers) and the "Sample" sector where plants are sampled randomly and unit multipliers are provided to take into account sampling probabilities. [11]

The specific ASI variables we use are described below: **Firm Age** is defined as the year of the census - year of initial production reported by the firms. **Firm Size** is the total number of workers which includes workers employed directly, workers employed through contractors, supervisory and managerial staff, other employees, working proprietors, unpaid family workers, and unpaid working members if it is a cooperative factory.

We define initial conditions (initial size, initial TFP, and initial legal form) when the firm is 1 year old. There is no standard definition in the literature on identifying new firms. For instance, Klapper, Laeven, and Rajan (2006) define new firms as all firms below the age of 2; Acs, Desai, and Klapper (2008) look at newly registered firms less than 1 year. Ghani, Kerr, and O'Connell (2014) define entrepreneurs as all firms less than three years old. Our results are robust to defining initial conditions at age 0, or age 2 or an average size over ages 0, 1, and 2.

We define **Small Entrants** as all firms in the bottom two quintiles of the size distribution of all entrants (i.e. aged 1) across all years in the sample and **Large Entrants** as all firms in the top 3 quintiles of the size distribution of all entrants across all years. The number of employees at age 1 ranges from 2 to 16 employees for Small Entrants and 17-848 employees for Large Entrants.[12] Our results are robust to alternate definitions of Small vs. Large entrants including defining Small vs. Large entrants depending on the distribution each year rather than across all

[11] This is similar to census data from other countries such as the Annual Survey of Manufacturers in the U.S. While the definitions of the Census and Sample sectors have changed over the years, for our entire sample period, the Census sector covered all units having 100 or more workers

[12] Our focus is on the comparison of outcomes of small entrants across institutional regimes and industry characteristics. In these tests large entrants are a control group for small entrants. Thus, the interpretation of the results is not affected if large entrants are disproportionately parts of business groups or some other advantaged group of entrants with access to internal capital markets.

years, using median as the cut-off point and looking at the tails of the distributions (i.e. defining Small as the bottom two deciles and Large as the top two deciles). We also find similar results using a continuous measure of size at age 1.

We define **Low Initial TFP** as all firms in the bottom two quintiles of the productivity distribution of all entrants across all years in the sample and **High Initial TFP** as all firms in the top three quintiles of the productivity distribution of all entrants across all years. TFP is Log (Revenue Productivity) measured by the product of physical productivity and a firm's output price (Hsieh and Klenow (2009)). Once again, our results are robust to defining initial conditions based on the productivity distribution each year, looking at alternate ways of classifying Low vs. High using median as the cut-off or using the tails of the distribution, as well as to using continuous measure of initial TFP.

Public Limited Company takes the value 1 if the initial legal form is a public limited company and 0 if the company is organized as a private limited company or proprietorship.[13] The most important distinction between public limited companies and private limited companies relates to their ability to raise funds from the public. While both public and private limited companies are incorporated and registered, private limited companies are not allowed to issue share capital whereas public limited companies have an unrestricted right to issue share capital so only public limited companies are eligible to be listed on a stock exchange. A similar distinction exists in many countries including the United Kingdom, as discussed by Brav (2009).

We look at the following performance metrics. **Employment Growth** is the annual growth in total number of workers. **Profits** are defined as the ratio of Profits to Total Assets. We

[13] The private limited company/proprietorship category consists of wholly privately owned firms organized as individual proprietorships, joint Hindu family business, partnerships, private limited companies, co-operative society, a corporation established by special Act of Parliament or State Legislature, and others including trusts, etc.

also examine whether entrants differ in their production structures, specifically, in more value-creating combinations of inputs by defining **Complexity of Production** as the ratio of Excise Taxes paid/Sales following Siegel and Choudhury (2012). Excise Tax is an indirect tax levied on the act of production or manufacture of goods paid by the manufacturer. Thus a lower tax (scaled by size) would imply that the value added from the manufacturing process is lower.

We follow the firms during the first eight years of their lifecycle.[14] Our results are robust to following firms up to 10 years, the maximum number of years we can follow a firm from its entry since we have data from 2001-2010. The confidence intervals are much wider due to lower sample sizes beyond eight years and hence we are more comfortable with restricting our early lifecycle analysis to the first eight years. To deal with outliers, within each age bin we winsorize the bottom and top 0.5% of all plant-level variables. We further winsorize top and bottom 0.5% of the ratios of variables. We winsorize within each age bin so as to not introduce systematic bias in our estimations such as that which would be created by winsorizing only the values for old firms. All our results are robust to winsorizing only the initial conditions (regressors). We also drop clear data errors where the year of initial production is given to be after the year of the survey.[15]

[14] As in the US Census' Annual Survey of Manufactures, probabilistic sampling in the Indian Census makes us unable to capture all exits. Thus, in the main tables we track surviving firms using the sampling weights provided by the census. We address additional sample selection issues in section III.B using the Combes et. al. (2012) methodology.

[15] The panel data was provided to us as individual annual files with establishment identifiers. There were some inconsistencies and missing values in the year of initial production reported from one year to next for the same establishment. We replaced the missing (19% of sample) and zero values (3% of sample) with the non-zero value reported in the subsequent year. To deal with the inconsistencies for each firm, we replaced all values of year of initial production with the mode, provided that there are only less than half of the observations different from the mode. If there are at least half of the observations that are different from the mode, we replace all observations with the value reported in the first year. Our results are robust to restricting the sample to years for which we have no inconsistency in the year of initial production. Several papers including Bollard, Klenow, and Sharma (2013); Dougherty, Frisancho Robles and Krishna (2011), and Harrison, Martin, and Nataraj (2012) have identified the presence of significant outliers in the Indian panel data and use algorithms similar to ours to ensure consistency across years.

The data also provide National Industry Classification (NIC) codes that map onto different revisions of the International Standard Industry Classification (ISIC) codes. Using this we construct three-digit NIC industry dummies that are consistent across all census-years and restrict the data to only the manufacturing sector. [16]

For confidentiality purposes, the ASI data do not provide firm identifiers. However the firms also report the total number of units the company has, which allows us to restrict all our analysis to factories that report that the company is not a multi-establishment firm (we take values 0 and 1 to be single establishment firms). Around 86% of the observations in our sample were single-establishment firms.

B. Industry Variables

We wish to explore if there are consistent differences in firm lifecycles across different types of industries. In particular, we look at whether growth and productivity over early lifecycle is a function of external financing needs, industry growth opportunities, and the type of production structure (capital intensive vs. labor intensive).

As an estimate of the external financing needs of the firm, we use US industries' dependence on external financing from Rajan and Zingales (1998) (RZ index). The RZ index is based on the assumption that since U.S. financial markets are developed, sophisticated, have fewer market imperfections and relatively open they should allow US firms to achieve their desired financial structure. Thus assuming that there are technological reasons why some

[16] The 2001/02 census uses NIC-98 which maps onto ISIC-Revision 3 at the 3-digit level; the 2002/03 and 2003/04 censuses use NIC-98 which maps onto ISIC-Revision 3.1 at the 3-digit level; the 2004/05, 2005/06 and 2007/08 censuses use NIC-04 which maps onto ISIC-Revision 3.1 at the 3-digit level; and the 2008/09, 2009/10 and 2010/11 censuses use NIC-08 which maps onto ISIC-Revision 4 at the 3-digit level. We drop recycling from the manufacturing sector since it is not included under manufacturing in the ISIC classification.

industries depend more on external finance than others, the RZ index offers an exogenous way to identify the extent of external dependence of an industry anywhere in the world. The methodology does not require that the US markets are perfect but rather that market imperfections in the US do not distort the ranking of industries in terms of their technological dependence on external financing.

The RZ index is at the 3-digit ISIC level that maps onto the Indian NIC classification. We construct **EFD,** a dummy variable that takes the value 1 for an industry if its dependence on external finance is greater than or equal to the median value of dependence on external finance across industries and 0 if it is less than the median value across industries.

Second, we create **Growing Industries** which is a dummy variable that takes the value 1 for an industry if its growth in employment over the period 2001-2010 is greater than (or equal to) the median industry growth over this period and 0 if the industry's growth in employment over this period is less than the median. Third, we follow Hasan and Jandoc (2012) in constructing **Labor Intensive Industries** which is a dummy variable that takes the value 1 for labor intensive industries and 0 for capital intensive industries.[17]

C. Local Institutions

To take into account institutional differences that may affect firms' lifecycle, we focus on the income level, level of financial development and the stringency of labor regulations across different states of India. Each of the measures is described below.

[17] Hasan and Jandoc (2012) classify the following industries in India to be capital intensive industries: Machinery, Electrical Machinery, Transport, Metals and Alloys, Rubber/Plastic/Petroleum/Coal and Paper/Paper Products. The labor-intensive industries are: Beverages and Tobacco, Textile Products, Wood/Wood Products, Leather/Leather Products and Non-Metallic Products. The remaining industries are not as clearly distinguishable and include: Food Products, Textiles, Basic Chemicals, Metal Products and Other Manufacturing.

For each year of the sample, depending on the value of state GDP/capita, we classify states into **Rich States** (≥ median) and **Poor States** (< median)**.**

There is a large variation across India's states in level of financial development. Bajpai and Sachs (1999) note that there has been a wide variation in the adoption of economic reforms with states like Maharashtra being very reform oriented while others, especially the poorest BIMARU states (Bihar, Madhya Pradesh, Rajasthan, and Uttar Pradesh) being slower to adopt. Aghion et al. (2008) also note the reforms in the 1990s to be associated with increasing cross-state inequality in industrial performance. We measure financial development by the ratio of total Commercial Bank Credit outstanding to the Net State Domestic Product (SDP) in each year and gauges the depth of financial development. The data is sourced from Burgess and Pande (2005) with updates from the Reserve Bank of India (http://dbie.rbi.org.in). We only have data on financial development across 15 Indian states but these are the major states of India accounting for 95% of India's population and 90% of India's GDP in 2004/05. Based on Credit/SDP, we construct a dummy variable, **FD**, which takes the value 1 for a particular state in a particular year if that state is at or above the median value of financial development in that year across states and 0 for states that are below the median value of financial development. As robustness we also construct FD based on the initial value of financial development in 1995 (before our sample period). Our results are also robust to using financial breadth (bank branches per capita) rather than depth as a measure of financial development.

Several papers suggest that India's labor regulations are responsible for the stagnant share of manufacturing outputs in India's GDP because of the impediments placed on hiring and firing workers (e.g. Dougherty (2009) and Hasan, Mitra, and Ramaswamy (2007)). A large literature has evolved quantifying labor market regulations across different states of India (e.g. Besley and

Burgess (2004), Hasan, Mitra, and Ramaswamy, 2007; Dougherty (2009), Gupta et al. (2008)).[18] Following Gupta et al.'s (2008) composite classification, we create a **Flexible State** dummy that takes the value 1 for states with flexible labor regulation (Andhra Pradesh, Karnataka, Rajasthan, Tamil Nadu, and Uttar Pradesh) and 0 for states with rigid (Maharashtra, Orissa, and West Bengal) or neutral labor regulations (Assam, Bihar, Gujarat, Haryana, Kerala, Madhya Pradesh, and Punjab). [19] We make one change to the Gupta et. al. classification of using the pre-2000 state boundaries in classifying states. So flexible states include Uttaranchal (which split from Uttar Pradesh in 2000) and inflexible states include Jharkhand and Chattisgarh (which were formerly part of Bihar and Madhya Pradesh, respectively).

As a measure of the overall business environment, we use the World Bank's Doing Business Indicators for 2009 that ranks 17 Indian cities (in 17 states) by the quality of doing business. Subnational Doing Business indicators are being widely used by the World Bank and governments to understand local differences in business regulations and implementation of national laws. The ease of doing business index, **DB Rank**, averages each city's percentile ranking along seven dimensions – Starting a Business, Dealing with Construction Permits, Registering Property, Paying Taxes, Trading across Borders, Enforcing Contracts, and Resolving Insolvency, and ranges from 1 (for Punjab) to 17 (for West Bengal) with higher values corresponding to states with worse doing business environments. We construct **Good Doing**

[18] Hasan et al. (2007) argue that the scores on cumulative amendments between 1980 and 1997 do not vary much over time within states, with eight of the states showing no amendment activity since 1980. Dougherty (2009) further reports that only 8 amendments (in 3 states) have been recorded since 1990, and only one amendment passed in 2004 appears to be of material importance to labor market outcomes. Gupta et al. (2008) build a composite index based on a simple majority rule across the indicators proposed in Besley and Burgess (2004), Bhattacharjea (2006), and Dougherty (2009).

[19] The labor market regulation index is not available for the following states and union territories: Jammu & Kashmir, Chandigarh, Nagaland, Manipur, Tripura, Meghalaya, Daman & Diu, Dadra & Nagar Haveli, Pondicherry, Lakshadweep, and Andaman & Nicobar Islands.

Business, a dummy variable that takes the value 1 for states with good doing business environments (DB Rank ≤9) and 0 for states with poor doing business (DB Rank >9).

We also control for human capital using **Literacy Rate** which is the proportion of persons who can both read and write with understanding in any language among population aged 7 years and above. Table A1 of the web appendix presents summary statistics and pairwise correlations of our main variables and shows that none of the correlation coefficients are very high to suggest multi-collinearity. Panel C of Table A1 shows that large entrants appear to be systematically different from small entrants. Large entrants have smaller cash/asset and loans/asset ratios but more complex production structures and higher profitability and productivity than small entrants.

III. Initial Conditions and Early Firm Lifecycle in India

A. Initial Conditions and Size and Growth over the Lifecycle

In this section we first investigate which of the following founding conditions – size, TFP, and legal form – have the largest explanatory power in determining average size and growth over the first eight years of a firm's lifecycle. In particular, we want to compare the explanatory power of initial conditions to that of state dummies to understand what role initial conditions play in determining size and growth over the early lifecycle. To do this, we estimate:

$$y_{ijst} = \mu + \beta_1 Initial\ Size(or\ TFP\ or\ Legal\ Form)_{ijs0} + \delta_j + \pi_s + \upsilon_{ijst} \quad (1)$$

where the dependent variable y_{ijst} is size or growth of firm i in industry j, in state s, and year t, μ is the average response across all firms, δ_j are industry effects, π_s are state effects and the υ_{ijst} are random disturbances. We look at three initial conditions - L_{i0} (dummy for Large Entrant), TFP_{i0}

(dummy for High Initial TFP entrant) and F_{i0} (dummy for initial legal form). The regression is estimated using ordinary least squares with sampling weights taken into consideration.

Insert Table 1 here

Using the full panel of firms, we present a variance decomposition analysis in Table 1 to compare the relative importance of different initial conditions in explaining firm size and growth. We begin with a benchmark specification with state dummies to model institutional variation at the state level. This specification provides us with the upper bound for the variation that can be explained at the state level. In subsequent specifications, we calculate the increment to adjusted R-square with industry effects followed by one of the initial conditions.

Col. 1 of panel A of Table 1 shows that the adjusted R-square when we regress Establishment size on state dummies is 3%. This also means that any state-level institutional variable that we might want to substitute the state effects with can explain a maximum of 3% in the variation in establishment size. When we add Industry Dummies we explain an additional 4% and when we add Large Entrant dummy to this regression, we explain an additional 7.9%. Thus initial size has a larger explanatory power than institutions or industry effects in determining average size over the first eight years. By comparison, Panels B and C of Col. 1 show that High Initial TFP dummy and Public Limited Company do not add as much explanatory power to the baseline specifications with state and industry dummies as Large Entrant dummy. In Col. 2 we keep the sample size constant across the three panels and again find that the Large Entrant dummy has the highest explanatory power in explaining average size over the early lifecycle.

Cols 3 and 4 of Table 1 show that none of the initial condition variables has any explanatory power in explaining the variation in average firm growth over the early lifecycle.

We also do not find state and industry dummies to explain any variation in firm growth rates over the early lifecycle.

Overall, Table 1 shows that size at start up has the highest explanatory power in predicting size over the first eight years of a firm's lifecycle. While initial TFP has no explanatory power in determining size, legal organization of the firm, that is, whether a firm is organized as a public limited company or as a private limited company or proprietorship/partnership, explains 3.4% of the variation in size, which is about half that explained by initial size. Since initial legal form explains much less than initial size, and both are likely to be correlated, in the rest of the paper, we focus mostly on the role of initial size.

Insert Table 2 here

Next we examine how the effect of initial size varies over the early lifecycle of the firm. Thus, we run variations of the specification:

$$y_{iat} = \mu + \sum_{a=2}^{T} \delta_a + \beta_1 Initial\ Size_{i1.} + \beta_2 Initial\ TFP_{i1.} + \sum_{a=2}^{T} \delta_a \times Initial\ Size_{i1.} + \gamma_t + \delta_j + \pi_s + \upsilon_{ijst} \quad (2)$$

where the dependent variable y_{iat} is a relevant characteristic of firm i aged a years at time $t > 0$, δ_a are dummy variables that take on the value 1 when the firm is a years old and zero otherwise, Initial Size and Initial TFP are defined when the firm is one year old, μ is the average response across all firms, δ_j are industry effects, π_s are state effects, γ_t are year fixed effects and the υ_{ijst} are random disturbances. We use robust standard errors in all specifications but our results also hold if we were to double-cluster standard errors on both firm and time following the procedure in Cameron, Gelbach, and Miller (2001).

In Table 2, we regress Size and Employment Growth on age dummies, Large Entrant dummy, and their interactions. Since we follow firms right from age 1, we can consider the Large Entrant dummy to be exogenous to the system. In all estimations we use weighted regressions with sampling weights. We also control for unobserved heterogeneity at the state, industry, and year level using dummies as well as the initial TFP of the firm.

Col. 1 of Table 2 presents regressions without any interaction terms. All the age coefficients are positive and significant, suggesting that firms on average are larger as they age, and firms that are born large are on average larger than firms that are born small. For instance, the coefficient of Age 8 dummy shows that the average firm that is 8 years old has 56 employees more than the average firm that is one year old. The coefficient of the Large Entrant dummy shows that a firm that is classified as a large entrant has 68 employees more than a firm that is classified as a small entrant. We also see that controlling for initial size, firms with high initial TFP are on average smaller (6 fewer employees) than firms with low initial TFP. Thus, to the extent that firms' initial size is constrained by market imperfections, there is little evidence that high initial TFP allows firms to relax those constraints.

In col. 2 of Table 2, we interact Large Entrant with age dummies and find the interaction term to be positive and significant, showing that firms that are born large are larger at all points during the lifecycle.[20] These effects are also economically significant. A calculation of the predictive margins reveals that the average firm that is classified as a large entrant has 67 employees at age 1 and a firm that is classified as a small entrant has 16 employees at age 1. By age 8, the average large entrant has 139 employees and the average small entrant has 41

[20] In unreported estimations, High Initial TFP is not significant when we don't control for initial size in Col. 1. Furthermore, even after controlling for initial size, the interactions of High initial TFP and age dummies are not significant in Col.2. We do not report these estimations since Table 1 shows that High Initial TFP has very little explanatory power in determining size.

employees. We plot the complete set of economic magnitudes of all the interaction coefficients at different ages, with confidence intervals in Panel A of Figure 1. The figure clearly shows the persistence of initial size over the early lifecycle.

In cols. 3-4 we repeat the specifications in cols. 1-2 but using annual Employment growth as the dependent variable. In col. 3 when we do not include any interaction terms, we find that neither the Large Entrant dummy nor the High Initial TFP dummy is significant. In col. 4, the interactions of Large Entrant dummy and age dummies are insignificant, suggesting that there is no evidence that large entrants grow differently than small entrants during the early lifecycle. With the presence of interaction terms, the negative and significant coefficient on the Large Entrant dummy only suggests that the average small entrant at age 2 is growing 20% faster than the average large entrant at age 2 (omitted age category). Panel B of Figure 1 plots the complete set of predictive margins of the interaction coefficients in col. 4 with confidence intervals.[21]The figure confirms that the growth rates of large versus small entrants are not significantly different during the early lifecycle.

In unreported regressions, we estimate two sets of interaction terms – interaction of the Large Entrant dummy and age dummies and interaction of the High Initial TFP dummy and age dummies. Figure 2 plots the predictive margins of these interaction effects and shows that initial size dominates initial productivity in predicting size over the early lifecycle.

[21] In this case, predictive margins provide the average growth rate for entrants keeping everything else constant. For instance, keeping all the covariates as they happen to be, the average growth rate at age 2 of small entrants is 0.59 whereas the average growth rate of large entrants is 0.39. At age 3, average growth rate of small entrants is 0.26 and that of large entrants is 0.60 and so on. None of these differences are statistically significant once we add in confidence intervals.

Overall, Table 2 and associated figures show that the relative size at entry matters for how large a firm is going to get over its early lifecycle and the difference in growth rates between large and small entrants is not economically significant.

We subject our results to a battery of robustness tests. First, we obtain very similar results when we replace the Large Entrant dummy with a continuous measure of initial size. Second, our results on persistence are robust to the following alternate definitions of Large vs. Small entrants - using the median entry size as a cut-off for Large (>=median) and Small (<median); focusing on the tails of the distribution and classifying the first two deciles of the entry size distribution as Small Entrants and the top two deciles of the entry size distribution as Large Entrants; defining Large vs. Small entrants looking at the size distribution each year; and defining Small and Large entrants using total assets rather than total number of employees. Appendix Figures A1 and A2 provide the economic effects (predictive margins) for the size and growth regressions using these alternate definitions and confirm our findings.

Second, we examine if initial leverage (High Debt dummy) accounts for the persistence in size. The High Debt dummy takes the value 1 if the value of Total Loans to Assets at age 1 was in the top three quintiles and 0 if the value of Total Loans to Assets at age 1 was in the bottom two quintiles of the leverage distribution. Total Loans includes both outstanding loans and overdraft facilities. We begin by first using initial leverage as a control variable in the size regression as shown in Col. 1 of Web Appendix Table A2 and find our results on initial size unchanged. In Col. 2, we estimate a regression of Size on Large Entrant dummy, Age dummies, High Debt dummy and a triple interaction of Large Entrant dummy, Age dummies, and High Debt dummy including pair-wise interaction terms. The economic effects of these interaction terms are presented in the predictive margin graphs in panel A of Figure 3. The predictive

margins show that it is initial size rather than initial leverage which has significant effects on size over the early lifecycle. In unreported tests, we find no difference in the growth rates of large versus small entrants irrespective of initial debt ratios.

In Col. 3 of Web Appendix Table A2, we estimate a regression of Size on Large Entrant dummy, Age, Public Limited Company and a triple interaction of Large Entrant dummy, Age dummies, and Public Limited Company including the pair-wise interaction terms. As before, we focus on the economic magnitudes by looking at the predictive margins of the triple interaction effects in Panel B of Figure 3. The figure shows that large entrants that are public limited companies are the largest in size at all ages followed by large entrants that are either private limited companies or proprietorships/partnerships. There is no difference in size between the small entrants that are public limited companies and small entrants that are private limited companies/proprietorships/partnerships.

Overall this section shows that size at start up is remarkably persistent over the early lifecycle and the growth rates of large and small entrants are not significantly different from each other over the first eight years.

B. Sample Selection and Attrition Issues

One of the concerns in investigating lifecycle effects is the role of selection. In particular, an alternative explanation to our findings above might be that small firms exit the market at a larger rate than large firms. Thus, the increase in average size with age could result from a Schumpeterian selection of firms rather than a monotonic growth pattern over the early lifecycle.

To address this, we follow the approach in Combes et al. (2012) to parameterize the extent of selection bias. The method relies on comparing the quantiles of the two distributions

under very mild distributional assumptions. The two distributions may differ for several reasons. Firms may grow on average over the six year period between ages 2 and 8 which causes a positive shift in the distribution of eight-year-old firms relative to the distribution of two-year-old firms (shift parameter A). Second, firms' performance may diverge, as firms that started at similar sizes grow differentially. This process will cause the distribution of eight-year-old firms to be a dilated transformation (dilation parameter D) of the distribution of two-year-old firms. Third, selection effects caused by under-sampling of small firms or by a higher rate of exit of small firms could result in the distribution of eight-year-old firms being a truncated version (truncation parameter, S) of the distribution of two-year-old firms.

Insert Table 3 here

Table 3 presents the values of the shift parameter A, dilation parameter D, and truncation parameter S from this empirical estimation together with bootstrapped standard errors. In panel A, when we estimate all three parameters, we find that the value of A is positive but not significant, showing that there is no significant right shift of the distribution at age 8. The estimate of D is above 1 and statistically significant, suggesting that size distribution of older firms is more dilated than that of younger firms. The value of S is positive but not significant. Taken together, these estimates of A, D, and S provide strong evidence that there are no differences between younger and older firms in the truncation of distribution of firm sizes. The pseudo-R^2 measures how much of the mean-squared quantile difference between the size distribution of younger and older firms is explained by the three parameters and is above 0.9 suggesting that the fit is very good. In panels B-E, we compare the baseline results in A with constrained specifications to explore how important it is to estimate all three parameters. In Panel B when we impose the restriction of no selection, we find that A is positive and significant

and D >1 and significant and the fit is equally good. In panel C when we assume only shift and truncation and no dilation, we find A<0 and S to be positive and significant but the fit to be much poorer (R^2=0.627). These estimates are biased as they attempt to approximate a dilation and we tend to overestimate truncation and underestimate shift. Similarly in panels D and E when we only assume shift and truncation, respectively, the shift is very poor. Together, panels A to E suggest that the best fit is achieved when we assume no selection.

Overall, Table 3 suggests that selection does not play a major role in explaining the size distribution of older firms vis-à-vis younger firms in our sample. Instead there is evidence that size distribution at age 8 is right-shifted and dilated relative to the distribution at age 2.[22]

A second concern with the findings in section A is whether they are driven by panel attrition. While attrition reduces the sample size, a more serious concern is attrition bias where firms that drop out of the panel differ systematically from those who remain in the panel. The specific concern is that our results may be driven by small firms dropping out of the panel because of the sampling scheme.

To address this, following Wooldridge (2002) we estimate an attrition probability function based on initial size and obtain predicted attrition probabilities for each observation. That is, we create a dummy variable that takes the value 1 if the firm is in the panel at age 2 and 0 if the firm is not in the panel at age 2 and estimate a probit attrition model by regressing this

[22] In additional sensitivity analyses we find that correcting for firm exits does not change our results. In 5 of the less industrially developed states in India, all industrial units were surveyed each year. We compute the exit rates of small and large entrants at each age across the five states. We then use this exit rate and an exit correction to the growth rate (0 or -50% or -100%) to compute an average expected growth rate for small and large entrants in the rest of the states. The difference between the observed growth rate and the expected growth rate corrected for exit gives us a discount factor for small versus large entrants at different ages. We then adjust the growth rate of every establishment by this discount factor and re-run our regressions and find no material difference to our results.

variable on firm size at entry. The predicted probabilities from this regression provide the attrition probabilities at age 2. We repeat the process eight times to estimate the attrition probabilities at each age. We then adjust the sampling weights by the inverse of these attrition probabilities to obtain an overall weight. We then re-estimate our tables using this new weight instead of the sampling weight. Overall we find no material difference to our results when we account for panel attrition. Appendix Figure A3 shows that even after accounting for panel attrition, initial size is persistent over the early lifecycle and the growth rates are not different. That is, small and large firms grow at the same rate over the early lifecycle.

C. Initial Conditions and Institutions

In Table 4, we look at the early lifecycle effects in size and growth across different institutional environments. In all regressions we control for industry and year fixed effects.

Insert Table 4 here

In panel A of Table 4, we focus on the Establishment size regressions looking at the interaction of Large Entrant dummy with each institutional variable. For the sake of brevity and ease of interpretation we only present results with the double interaction of the Large Entrant dummy and the institutional variable but discuss the results of the triple interaction of Large Entrant x Institutional variable x Age dummies in the text below. Cols. 1 to 4 of Panel A show the main effects without any interaction terms in the regression and cols. 5 to 8 show the regressions with the interaction term.

Cols. 1 to 4 show that firms are on average larger in richer states and states with rigid labor regulations. Firms in rich states have on average 11 more employees than firms in poor

states and firms in flexible labor states have 11 fewer employees than firms in states with rigid labor regulations. Interestingly, the coefficients on Financially Developed dummy and the Good Doing Business dummy are not significant.

Cols. 5 to 8 show that the interaction of Large Entrant dummy and the institutional variables are statistically significant. The interaction terms suggest that large entrants are on average larger in rich states, financially underdeveloped states, states with poor doing business environments and states with rigid labor regulations. To see whether these differences are economically significant we look at the predictive margins (unreported). The predictive margins for Col. 5 show that the difference in size between large entrants and small entrants is 77 employees in financially developed states and 70 employees in financially under-developed states. Small entrants in financially developed states are larger by 3 employees (statistically significant at 5% level) than small entrants in financially under-developed states. Large entrants in financially developed states are smaller by 4 employees than large entrants in financially under-developed states though this is not statistically significant. The complete set of economic effects of the interaction term (Financial Developed dummy x Large Entrant dummy) at different ages is plotted in the first panel in Figure 4. The figure shows that initial size dominates the effect of financial institutions on average size over early lifecycle. [23] The difference in differences tests show that there are no economically significant differences between large and small entrants across states with developed and under-developed financial institutions. The figures in the other panels of Figure 4 confirm that this holds across other types of institutions - initial size dominates the effect of labor market regulations, income, and doing business environment on average size over early lifecycle.

[23] We obtain similar results using an alternate measure of financial development – the total number of operating bank branches per million persons in each state.

In Panel B of Table 4, we repeat the specifications of Panel A but with Employment Growth as the dependent variables. Cols. 1 to 4 show that none of the institutional variables are shown to predict growth and neither is Large Entrant dummy. The interaction terms of Large Entrant x Institutional Variable in cols. 5 to 8 are also mostly insignificant suggesting that there is no difference in annual employment growth rates of small versus large entrants across different types of institutions including financial development, income, labor regulations, and ease of doing business. The complete set of economic effects of the interaction term (Financial Developed dummy x Large Entrant dummy) at different ages is plotted in the first panel in Figure 5. The figure confirms the absence of differences in growth rates of different types of entrants across different levels of financial development. The figures in the other panels of Figure 5 confirm that the growth rates of small and large entrants are not different across states with different income levels, labor regulations or doing business environments.

We perform a number of robustness tests that are available on request. First, our results on persistence of size across different institutions hold when we restrict the sample size in panel A of Table 4 to the subset of observations in panel B of Table 4. Second, we obtain very similar results if we were to replace Large Entrant dummy with a continuous measure of initial size in panels A and B as shown in Appendix Table A3. Third, we estimate a regression of Growth Rates on a triple interaction term – Large Entrant dummy x High Initial TFP dummy x Age Dummies – and the corresponding main effects and double interactions while controlling for industry and year dummies. We find no evidence that productive small entrants grow fast and increase their relative size over time in states with good financial institutions.

Finally, one of the concerns with treating initial conditions as being exogenous for future outcomes may be the endogeneity of firm location. That is, selection of firms into states may

differ by both firm size and growth potential. For example, growth oriented firms may also choose to locate in a state with less stringent labor regulations, even if such a move were costly. If so, we would expect to see that in states with flexible labor regulations small firms grow relatively faster than large firms in such states, and also relatively faster than small firms in states with less flexible labor regulations. However we find no evidence that institutions affect the performance of subsets of firms post entry. Thus our results are robust to the possibility of pre-entry movements of entrepreneurs across states.

Overall, Table 4 and the associated figures show that institutions do not play a major role in influencing the growth rates of large versus small entrants in India over the early lifecycle. This finding differs from the existing literature on institutional change (e.g. Bertand, Schoar and Thesmar (2007), Cetorelli and Strahan (2006), Cetorelli (2014), and Kerr and Nanda (2009)) from which it is natural to assume that different institutions would have a differential impact on the growth rates of entrants with differing initial conditions. However, these papers examine changes in firm growth following a change in institutions. By contrast, we examine the relative growth of firms across a range of institutions during the first eight years of a firm's early lifecycle. In contrast to the prior literature which examines firms' adjustments to changes in institutions, the firms we examine are adapted to their institutional environment. In the following sections, we examine if we observe differences across industries and if institutions perhaps have a more significant role in the selection of firms that are entering.

D. Initial Conditions and Industry Classifications

In this section we study the lifecycle effects across different industry classifications. We examine whether industry differences matter for the persistence of initial conditions established in Tables 1 and 2. We look at three different classifications of industries – labor intensive vs. capital intensive, industries with high dependence on external finance vs. industries with low dependence on external finance and high growth industries vs. declining industries.

In Panel A of Table 5 we focus on initial size and average size over the first eight years and in panel B we focus on initial size and employment growth over the first eight years. In both panels, in cols. 1-3 we do not look at any interaction effects and focus on the industry main effects and effect of initial size. In cols. 4-6 we study the interaction of Initial Size x Industry. In all regressions we control for state and year fixed effects.

Cols. 1 to 3 of Panel A show that large entrants and firms in labor intensive industries are on average larger. The coefficient of labor intensive industry in col.1 shows that the average firm in labor intensive industry has 9 more employees than the average firm in a capital intensive industry. However there is no evidence that the external finance dependence of an industry or industry growth is related to average size over the early lifecycle. In cols. 4 to 6, we see that the only interaction term that is significant is Labor Intensive x Large Entrant suggesting that on average, large entrants in labor intensive industries are larger than large entrants in capital intensive industries. The economic effects can be seen from the unreported predictive margins of these interaction terms. The predictive margins for Col. 4 show that the difference in size between large entrants and small entrants is 70 employees in capital intensive industries and 85 employees in labor intensive industries. While small entrants in labor intensive industries are not

significantly different from small entrants in capital intensive industries, we find that large entrants in labor intensive industries are larger by 14 employees (statistically significant at 1% level) than large entrants in capital intensive industries. The complete set of economic effects of the interaction term (Labor Intensive Industry x Large Entrant dummy) at different ages is plotted in the first panel in Figure 6. The two other panels of Figure 6 plot the complete set of economic effects of High Dependence on External Finance x Large Entrant Dummy and Growing Industry x Large Entrant Dummy respectively. The panels all show that initial size dominates the effect of industry classification in explaining size over the early lifecycle.

In panel B of Table 5, we repeat the same specifications as in cols. 1-6 of panel A but with growth as the dependent variable.[24] None of the industry main effects are significant in cols 1-3 and neither are the Large Entrant x Industry interaction effects in cols. 4-6. Thus there is no evidence that industry classification has a differential effect on the growth rate of large versus small entrants. In unreported tests we find consistent results when we use triple interactions of Large Entrant Dummy x Age Dummies x Industry Characteristic. Appendix Figure A4 plots the predictive margins of these triple interaction terms and shows that there is no difference in growth rates of large versus small entrants across different industry classifications.

In unreported robustness, we examine if perhaps certain combinations of industries and institutions matter for growth rates. So we examine if firms in industries that are dependent on external finance grow faster in states with better developed financial institutions. Appendix Figure A5 presents the predictive margins of these interaction effects and shows that there is no evidence that entrants in financially developed states in industries that are highly dependent on

[24] In unreported regressions of growth on interaction of Large entrant dummy with industry dummies, a joint significance test of the interaction effects equal to zero is not rejected.

external finance grow faster than entrants in financially under-developed states in industries that are not highly dependent on external finance.

Overall, Table 5 shows that initial firm characteristics, specifically founding size is a key determinant of size over the early firm lifecycle and this is robust to different industry classifications. We do not find any differences in growth rates of large and small entrants in the overall sample or across different industry classifications.

E. Initial Conditions and Other Performance Metrics over the Lifecycle

In this section, we explore whether entrants differ in other performance metrics over the early lifecycle. We look at three different performance metrics. In cols. 1-3 of Table 6 we look at productivity, in cols. 4-6 we examine if certain types of entrants engage in more value-added manufacturing than other entrants, and in cols. 7-9 we look at profitability ratios over the first eight years. In addition to looking at large vs. small entrants we also look at entrants with high vs. low initial TFP.[25]

Col. 1 of Table 6 shows that on average large entrants and entrants with high initial TFP have higher productivity over the early lifecycle. In col. 2 we look at interactions of Large Entrant dummy and age dummies and in col. 3 we look at interactions of High Initial TFP dummy and age dummies. None of the interaction terms are significant in col. 2 whereas the interaction of High Initial TFP and age dummies are negative and significant in col. 3. Figure 7 plots the predictive margins of the interaction effects in cols. 2 and 3. Figure 7 shows that there are no significant differences in the productivity of large and small entrants over the early lifecycle. However, entrants with high initial productivity continue to have high initial

[25] We don't focus on initial form here since initial form and initial size are likely to be highly correlated and their effects are not easily distinguishable.

productivity over the first eight years and firms with low initial productivity continue to have low initial productivity over the first eight years. Thus productivity is also persistent over the early lifecycle but the figure also shows that there is a mean convergence effect with the productivity of low initial TFP entrants increasing with age and that of high initial TFP entrants declining with age. In unreported regressions, we look at a triple interaction term – Large Entrant Dummy x High Initial TFP x Age dummies – and the predictive margins of the interaction effects show that initial productivity dominates initial size in predicting TFP over the early lifecycle.

In cols. 4-6 of Table 6, we examine whether certain entrants engage in more value-creating combinations of inputs as proxied by the ratio of Excise Taxes paid/Sales following Siegel and Choudhury (2012). Excise Tax is an indirect tax levied on the act of production or manufacture of goods paid by the manufacturer. Thus a lower tax (scaled by size) would imply that the value added from the manufacturing process is lower. The estimation in Col. 4 of Table 6 has no interaction effects and shows that on average large entrants have more complex production structures and there is no association of high initial TFP with complex production structures. The interaction effects of Large Entrant x Age dummies are mostly positive and significant in col. 5 whereas the interaction effects of High Initial TFP x Age dummies are mostly insignificant. Figure 8 presents the predictive margins of the interaction effects from cols. 4 and 5 and shows that large entrants have more complex production structures i.e. engage in more value-added manufacturing, than small entrants whereas entrants with high and low initial TFP do not look very different in the extent of value added manufacturing they undertake.

In cols. 7-9 of Table 6 we look at profitability ratios. The Large Entrant dummy is insignificant in the profits regression in col. 7 whereas the High Initial TFP dummy is positive

and significant. Col. 8 presents interactions of Large Entrant dummy with Age dummies and col. 9 presents interactions of High Initial TFP dummy with Age dummies. The interaction effects of High Initial TFP dummy and age dummies are consistently negative and significant in col. 9. Figure 9 plots the predictive margins of the interaction effects from cols. 8 and 9 and shows that while large and small entrants do not look very different in their profit ratios, entrants with high initial TFP have higher profits ratios than entrants with low initial TFP. [26]We also see that entrants with low initial productivity ramp up their profit ratios as they get older. In unreported tests we also find that entrants with high initial TFP have higher operating cash flow volatility (defined as the standard deviation of Operating Cash flow to Total Assets).

Overall this section shows that large and small entrants are fundamentally different from each other in the extent of value added manufacturing they undertake. While we do not have direct evidence of managerial skill in these firms, this finding suggests that large entrants have potentially higher skilled workers to perform the value added manufacturing. We also see that the higher efficiency of the high initial TFP entrants is persistent and translates into higher profit/asset ratios over the early lifecycle.

IV. *Role of Institutions on Initial Entry*

Given the importance of initial conditions – initial size for size and complexity of production and initial TFP for productivity and profit ratios over the early lifecycle - established in the previous sections, we now explore if institutions have an impact on the selection of firms at entry.

[26] When we don't winsorize the profitability ratios, the difference between high and low initial TFP entrants is less stark.

We begin by first presenting summary statistics on the entry process in the population of firms in Table 7. When we look across time, we see that the percentage of entrants increases from 2003 to 2007 and thereafter drops, potentially due to repercussions from the global financial crisis. We then examine the size distribution of the entrants in each year by looking at the following size bins – 1-5 employees, 6-20 employees, 21-50 employees, 51-100 employees, and 100+ employees. Each year we see that the largest share of entry is in the 6-20 employees category followed by the 21-50 employees category. The average entry size shows an increasing trend over the years ranging from 42.70 employees in 2001 to 47.13 in 2010.

In panel A of Table 8 we examine the relation between firm initial conditions and institutions in a multivariate setting by estimating equation (2). We regress the Initial Size at entry and Initial TFP at entry on Credit/SDP controlling for the following: income level of the state (Rich State dummy), strength of labor regulation (Flexible state dummy), overall doing business environment (DB Rank), literacy rate, industry and year dummies. The Flexible State dummy and DB Rank are time invariant so we do not include state fixed effects in our regressions. Cols. 1-3 of Table 8 show that on average, the size at entry is lower in states with better developed financial institutions and states with flexible labor regulations. We also find entry size to be larger in richer states. The literacy rate seems to be positively associated with larger entry size, but this is significant only when we do not control for state income.

Cols. 4-6 of Table 8 show that financial development does not seem to be associated with initial TFP at entry. However, we find initial TFP at entry to be larger in states with worse doing business environments and states with rigid labor regulations.

Overall, Table 8 shows that that poor institutions are on average associated with larger and higher initial productivity of entrants, presumably to be able to overcome financing and regulatory obstacles. In an unreported tabulation, we show that much of the net effect on firm size appears to be due to a relative decrease in the proportion of entrants with 6-20 employees in less developed states. This size pattern is consistent with the "missing middle" observation in the development literature which notes that when comparing the employment in developing and developed countries, in developing countries there appears to be a shortfall in the proportion of employment in firms in the "middle" range of firm sizes compared to larger (over 50 employees) and very small firms (e.g. Tybout, 2000). The findings on smaller-sized entry with financial development also show that financial development affects firm entry in a developing economy analogously to banking deregulation in the U.S. as described by Kerr and Nanda (2010).

In panel B of Table 8, we perform robustness tests by estimating regressions at the state-year level. In addition to examining the association between financial development and average size of entrants and the average productivity of entrants in each state-year, aggregating up to the state-year level allows us to examine the extensive margin effects (percentage of entrants) of financial development. Col. 1 shows that there is greater percentage entry in financially developed states, poorer states, and states with flexible labor regulations. Col. 2 shows that average entry size is smaller in financial developed states. Col. 3 shows no relation between financial development and average productivity of entrants.

Overall panel B of Table 8 shows a significant impact of financial institutions on both the extensive margin (rate of entry) and intensive margins (size at entry). Greater access to external finance is associated with greater entry but also smaller size entry. This is consistent with studies

like Kerr and Nanda (2010) who show that in the US, banking deregulations brought in exceptional entry but the greatest increase in entry was among the very small start-ups.

The results on the effect of institutions on entry size in Table 8 are also consistent with the findings in Table 4 that institutions predict the average size of firms in the first eight years of their lifecycle but not their comparative growth rates. Taken together, our results show that the channel through which institutions affect the relative outcomes of young firms is through the initial distribution of firm characteristics at entry rather than their effect on the relative performance of the firms post entry.

V: Conclusion

In this paper, we ask how firm characteristics and the institutional environment predict a firm's success over its early lifecycle. Using data on the formal manufacturing sector in India, we find that initial firm size is remarkably persistent. Small and large entrants have similar growth rates, so that small firms tend to stay relatively small throughout the first 8 years of their lifecycle period. The size differential and growth rate similarity across firm sizes also appear to be unaffected by industry production structure (labor versus capital intensive), industry growth rates, and industry dependence on external finance. We find that large entrants engage in more complex production than small entrants.

Conditional on initial size, we find that institutional differences do not make a large difference to firm growth over early life cycle. We do find however, that local institutions make a great deal of difference both to the level and composition of entry. There is more entry in

regions with more access to external finance, and more entry by smaller firms. However, there is little evidence that these smaller entrants subsequently grow relatively faster than larger entrants.

Our findings point to the importance of institutions in selecting the composition of firms in the economy primarily through their effect on the level of entry and initial conditions. Our results suggest that policies facilitating entry may have high payoffs. But our results also show that firm-specific factors dominate which firms grow over the early lifecycle. The impact of better access to finance on the subsequent growth of entrants seems to be weak, suggesting that creating the right environment for entrepreneurship may be more important than trying to support the average small entrant or young firm directly.

Our results should not be interpreted as suggesting that improvements in institutions do not promote the growth of incumbent firms. Rather, for a given set of institutions there is an equilibrium level of entry of firms of different sizes and characteristics. Different firms will be affected differentially by specific institutional failures and entry will occur until the net present value of entry for marginal firms is driven to zero. For those entrants, we find that on balance the place of firms in the size distribution is persistent and there are minor productivity differences. However, this does not imply that all or a subset of the firms are not constrained along their growth path. Subsequent changes in institutions that remove regulatory obstacles to growth or increase access to capital may increase the value and growth of some or all incumbent firms that were subject to those constraints. Thus, for example, U.S. and French banking deregulation likely had that effect. Given the previous findings on the obstacles faced by firms in developing countries, it is likely that there is high value from such changes.

Overall, our paper's findings highlight the importance of initial size in forming the blueprint for firms' relative size positions over the first decade of their existence. Furthermore,

the paper highlights the effect of access to external finance on the intensive margin at initial entry, which serves as the firm's destiny for future size evolution because the firm is unable to affect its relative rank or grow differentially faster thereafter irrespective of the institutions.

References

Acs, Z.J., Desai, S. and Klapper, L. 2008. "What Does Entrepreneurship Data Really Show?" *Small Business Economics*, 31(3), 265-281.

Aghion, P., R. Burgess, S. Redding, and F Zilibotti. 2008. "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India." *American Economic Review* 98(4), 1397-1412.

Ahluwahlia, M. (2002) State Level Performance under Economic Reforms in India. In "*Economic Policy Reforms and the Indian Economy"* Ed: Anne O. Krueger, University of Chicago Press, 1 edition, Chicago, USA.

Asker, J. Farre-Mensa, J. and Ljungqvist, A. 2015. "Corporate Investment and Stock Market Listing: A Puzzle" *Review of Financial Studies* 28(2), 342-390.

Aterido, R. Hallward-Driemeier, M. and Pages, C. 2011. "Big Constraints to Small Firms' Growth? Business Environment and Employment Growth across Firms." *Economic Development and Cultural Change* 59(3), 609-647.

Ayyagari, M., A. Demirguc-kunt, and V. Maksimovic. 2015. "Are large firms born or made? Evidence from Developing Countries" *Policy Research Working Paper Series No. 7406.* The World Bank.

Ayyagari, M., A. Demirguc-kunt, and V. Maksimovic. 2014. "Does local financial development matter for firm lifecycle? Evidence from India" *Policy Research Working Paper Series No. 7008.* The World Bank.

Ayyagari, M., A. Demirguc-kunt, and V. Maksimovic. 2008. "How important are financing constraints? The role of finance in the business environment." *World Bank Economic Review 22(3), 483-516.*

Bajpai, N. and Sachs, J. D., 1999. The Progress of Policy Reform and Variations in Performance at the Sub-National Level in India. *Harvard Institute of International Development Discussion Paper No. 730*

Bargeron, L. L., Schlingemann, F. P., Stulz, R. M., and C. J. Zutter, 2008. "Why do private acquirers pay so little compared to public acquirers?" *Journal of Financial Economics* 89(3), 375-390.

Bas, M. and A. Berthou. 2012. "The Unequal Effects of Financial Development on Firms' Growth in India." *CEPII Working Paper Series* 2012-22.

Beck, T., A. Demirguc-kunt, and V. Maksimovic. 2005. "Financial and Legal Constraints to Growth: Does Firm Size Matter?" *Journal of Finance* 60(1), 137-177

Bedi, J. S. and P. K. Banerjee. 2007. “Discrepancies and Validation of Indian Manufacturing Data.” *Economic and Political Weekly*, 883-891.

Benmelech, E. and C. Frydman. 2015. “Military CEOs.” *Journal of Financial Economics* 117, 43-59.

Bennedsen, M., K. M. Nielsen, F. Perez-Gonzalez, and D. Wolfenzon. 2007. “Inside the Family Firm: The Role of Families in Succession Decisions and Performance.” *Quarterly Journal of Economics* 122: 647-691

Bertrand, M. and A. Schoar. 2003. “Managing with Style: The Effect of Managers on Firm Policies,” *The Quarterly Journal of Economics* 118, 1169-1208.

Bertrand, M., A. Schoar, and D. Thesmar. 2007. “Banking Deregulation and Industry Structure: Evidence from the French Banking Reforms of 1985.” *Journal of Finance* 62(2), 597-628.

Besley, T. and R. Burgess. 2004. Can Labor Regulation Hinder Economic Performance? Evidence from India. The Quarterly Journal of Economics 119(1), 91-134.

Bhattacharjea, A.,. 2006. Labour market regulation and industrial performance in India: Acritical review of the empirical evidence. *The Indian Journal of Labour Economics* 49(2), 211-32.

Bhide, A. 2000. *The Origin and Evolution of New Businesses*. Oxford University Press, Inc., New York, New York.

Black, S. and P. E. Strahan. 2002. “Entrepreneurship and Bank Credit Availability.” *Journal of Finance* 57, 2807-2833.

Bloom, Nicholas, and John Van Reenen. 2007. “Measuring and Explaining Management Practices across Firms and Countries.” *Quarterly Journal of Economics 122 (4), 1351-1408.*

Bloom, Nicholas, and John Van Reenen. 2010. “Why do management practices differ across firms and countries?” *Journal of Economic Perspectives*, 24(1).

Bloom, Nicholas, Benn Eifert, David McKenzie, Aprajit Mahajan, and John Roberts. 2013. “Does management matter: evidence from India.” *Quarterly Journal of Economics 128(1), 1-51.*

Bollard, A., P. Klenow, and G. Sharma, 2013. “India’s Mysterious Manufacturing Miracle.” *Review of Economic Dynamics 16(1), 59-85.*

Bramstetter, L., F. Lima, L. J. Taylor, and A. Venancio. 2014. “Do Entry Regulations Deter Entrepreneurship and Job Creation? Evidence from Recent Reforms in Portugal” *The Economic Journal 124(577),* 805-832.

Brav, O. 2009. “Access to Capital, Capital Structure, and the Funding of the Firm” *Journal of Finance 64(1), 263-308.*

Bruhn, M., D. Karlan, and A. Schoar. 2010. What Capital is Missing in Developing Countries? *American Economic Review 100(2),* 629-633.

Burgess, R. and R. Pande. 2005. "Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment." *American Economic Review 95(2),* 780-795.

Cetorelli, N. 2004. "Real effects of bank competition." *Journal of Money, Credit, and Banking* 36, 543–558

Cetorelli, N. 2014. "Surviving Credit market Competition." *Economic Inquiry* 52(1), 320-340.

Cetorelli, N. and P. E. Strahan. 2006. "Finance as a barrier to entry: Bank competition and industry structure in local U.S. markets." *Journal of Finance*, 61(1), 437-461.

Chatterjee, U. and R. Kanbur. 2013. "Regulation and Non-Compliance: Magnitudes and Patterns for India's Factories Act" *Policy Research Working Paper No. 6755.* The World Bank.

Chemmanur, T. J., and P. Fulghieri. 2014. Entrepreneurial Finance and Innovation: An Introduction and Agenda for Future Research. *Review of Financial Studies* 27(1), 1-19.

Chemmanur, T. J., T. Hull, and K. Krishnan. 2010. Do local and international venture capitalists play well together? A study of international venture capital investments. *SSRN Working Paper Series.*

Combes, P., G, Duranton, L. Gobillon, D. Puga, and S. Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. *Econometrica* 80(6), 2543-2594.

Cronqvist, H., A. Makhija, and S. E. Yonker. 2012. "Behavioral consistency in corporate finance: CEO personal and corporate leverage." *Journal of Financial Economics* 103(1), 20-40.

Demirguc-kunt, A. and V. Maksimovic. 1998. "Law, Finance and Firm Growth" *Journal of Finance* 53(6), 2107-2137.

Doidge, C., A. Karolyi, and R. Stulz. 2007. "Why do countries matter so much for corporate governance? *Journal of Financial Economics* 86(1), 1-39.

Doidge, C., A. Karolyi, and R. Stulz. 2015. "The US Listing Gap" *NBER Working Paper 21181.*

Dougherty, S. 2009. "Labor Regulation and Employment Dynamics at the State Level in India." *Review of Market Integration* 1(3), 295-337.

Dougherty, S., Frisancho Robles, V. and Krishna, K. 2011. "Employment Protection Legislation and Plant-Level Productivity in India." *NBER Working Papers 17693.*

Dyck, A. and L. Zingales. 2004. “Private Benefits of Control: An International Comparison.” *Journal of Finance* 59(2), 537-600.

Farre-Mensa, J., 2012. Comparing the cash policies of public and private firms. Unpublished working paper. Harvard University

Forbes, K. J., 2007. “One cost of the Chilean capital controls: Increased financial constraints for smaller traded firms,” *Journal of International Economics*, 71(2), 294-323.

Gao, H., Harford, J. and Li, K. 2013. “Determinants of corporate cash policy: Insights from private firms” *Journal of Financial Economics* 109, 623-639.

Ghani, E., W. Kerr, and S. O’Connell. 2014. "Spatial Determinants of Entrepreneurship in India." Special Issue on Entrepreneurship in a Regional Context. *Regional Studies* 48(6), 1071-1089.

Graham, J. R., S. Li, and J. Qiu. 2012. “Managerial Attributes and Executive Compensation.” *Review of Financial Studies* 25(1), 144-186.

Guiso, L., P. Sapienza, and L. Zingales. 2004. “Does Local Financial Development Matter?” *Quarterly Journal of Economics* 119(3), 929-969.

Gupta, P., Hasan, R., and Kumar, U., 2008, What constrains Indian manufacturing? *Macroeconomics Working Papers* 22162, East Asian Bureau of Economic Research.

Hadlock, C. J. and Pierce, J. R. 2010. “New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index” *Review of Financial Studies* 23(5), 1909-1940.

Haltiwanger, J., R. S. Jarmin, and J. Miranda. 2013. “Who Creates Jobs? Small versus Large versus Young” *Review of Economics and Statistics* 95(2), 347-361.

Harrison, A. E., L. A. Martin, and S. Nataraj. 2012. “Learning versus Stealing: How Important are Market-Share Reallocations to India’s Productivity Growth?” *World Bank Economic Review* 27(2), 202-228.

Hasan, R. and Jandoc, K. R. L., 2012, Labor Regulations and the Firm Size Distribution in Indian Manufacturing. *Columbia Program on Indian Economic Policies Working Paper No. 2012-3.*

Hasan, R., Mitra, D., and K. V. Ramaswamy.2007. Trade Reforms, Labor Regulations, and Labor-Demand Elasticities: Empirical Evidence from India. *Review of Economics and Statistics* 89 (3), 466-481

Hellmann, and M. Puri. 2000. “The interaction between product market and financing strategy: the role of venture capital.” *Review of Financial Studies* 13(4), 959-984.

Hsieh, C. T. and P. Klenow. 2009. “Misallocation and Manufacturing TFP in China and India.” *Quarterly Journal of Economics* 124, 1403-1448.

Kaplan, S., M. M. Klebanov, and M. Sorensen. 2012. “Which CEO Characteristics and Abilities Matter?” *Journal of Finance* 67, 973-1007.

Kaplan, S., B. Sensoy, and P. Stromberg. 2009. “Should investors bet on the jockey or the horse? Evidence from the evolution of firms from early business plans to public companies.” *Journal of Finance* 64, 75-115.

Kerr, W., and R. Nanda. 2010. “Banking Deregulations, Financing Constraints and Firm Entry Size.” *Journal of the European Economic Association* 8(2-3), 582-592.

Kerr, W., and R.Nanda. 2009. “Democratizing Entry: Banking Deregulations, Financing Constraints, and Entrepreneurship.” *Journal of Financial Economics* 94(1), 124-149.

Klapper, L., L. Laeven, and R. Rajan. 2006. “Entry regulation as a barrier to entrepreneurship.” *Journal of Financial Economics* 82, 591-629.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. W. Vishny. 1997. “Legal Determinants of External Finance,” *Journal of Finance*, 52(3), 1131-1150.

La Porta, R., F. Lopez-de-Silanes, F., and A. Shleifer. 2002. “Government ownership of banks,” *Journal of Finance* 57, 265-301.

Lemmon, M. L., M. R. Roberts, and J. Zender. 2008. “Back to the Beginning: Persistence and the Cross-Section of Corporate Capital Structure” *Journal of Finance* 63(4), 1575-1608.

Levine, R. 2005. Finance and Growth: Theory and Evidence. In *Handbook of Economic Growth*, ed. Philippe Aghion and Steven N. Durlauf, 865–934. Amsterdam: Elsevier.

Lucas, R. E. 1978. “On the size distribution of business firms.” *Bell Journal of Economics* 9, 508-523.

Maksimovic, V., G. Phillips, and L. Yang. 2013. “Private and Public Merger Waves.” *Journal of Finance* 2177-2217.

Malmendier, U. and S. Nagel. 2011. “Depression babies: Do Macroeconomic Experiences Affect RiskTaking?” *Quarterly Journal of Economics*, Vol. 126(1), 373-416.

Malmendier, U., G. Tate, and J. Yan. 2011. “Overconfidence and Early-life Experiences: The Effect of Managerial Traits on Corporate Financial Policies,” *Journal of Finance*, 66, 1687-1733.

Michelacci, C. and O. Silva. 2007. "Why So Many Local Entrepreneurs?," *The Review of Economics and Statistics*, 89(4), 615-633.

Michaely, R. and M. R. Roberts. 2012. "Corporate Dividend Policies: Lessons from Private Firms" *Review of Financial Studies* 25(3).

Perez-Gonzalez, F. 2006. "Inherited Control and Firm Performance." *American Economic Review* 96, 1559-1588.

Rajan, R. G., and L. Zingales. 1998. "Financial dependence and growth". *American Economic Review* 88, 559–587.

Rauch, J. E. 1991. "Modelling the informal sector informally." *Journal of Development Economics* 35, 33-47.

Sheen, A. 2009. "Do Public and Private Firms Behave Differently? An Examination of Investment in the Chemical Industry" *Working Paper* UCLA.

Siegel, J. and P. Choudhury. 2012. "A Reexamination of Tunneling and Business Groups: New Data and New Methods." *Review of Financial Studies* 25(6), 1763-1798.

Sutton, J. 1997. "Gibrat's legacy." *Journal of Economic Literature* 35, 40–59.

Tybout, James. 2000. Manufacturing Firms in Developing Countries: How Well Do They Do, and Why? *Journal of Economic Literature* Vol. 37, 11-44.

Wooldridge, J. 2002. "Inverse Probability Weighted M-Estimators for Sample Selection, Attrition, and Stratification." *Portuguese Economic Journal* 1, 117-139.

Zingales, L. 1998. "Survival of the Fittest of the Fattest? Exit and Financing in the Trucking Industry." *Journal of Finance* 53(3), 905-938.

Figure 1: Size and Growth over Early Lifecycle: Large Vs. Small Entrants

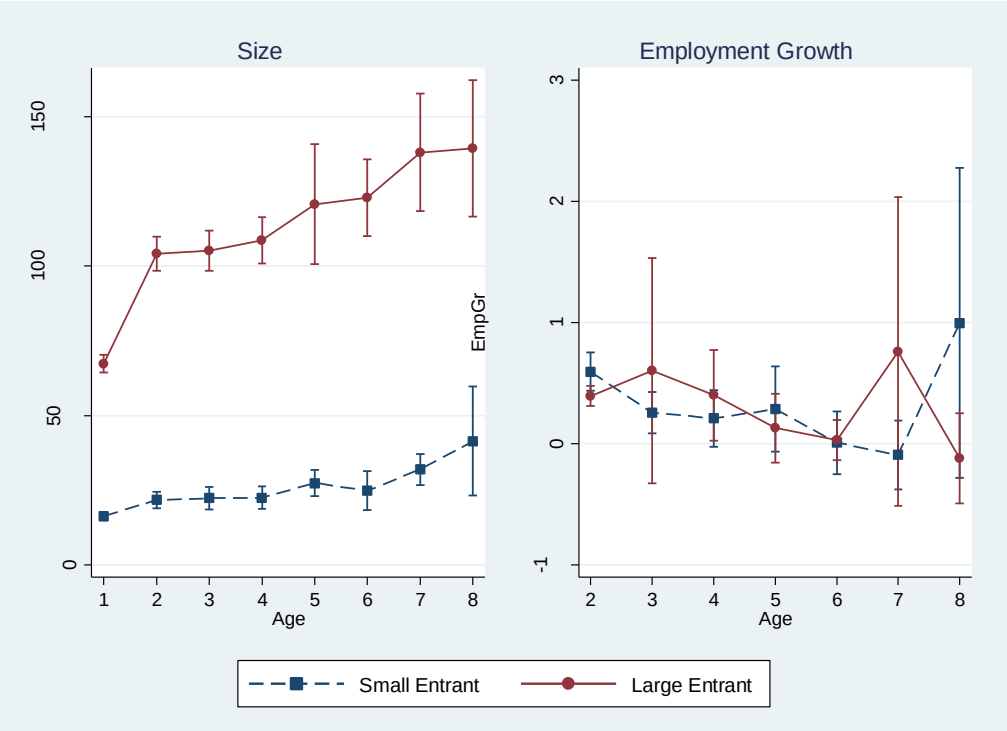


Figure 2: Persistence in Size – Initial Size vs. Initial TFP

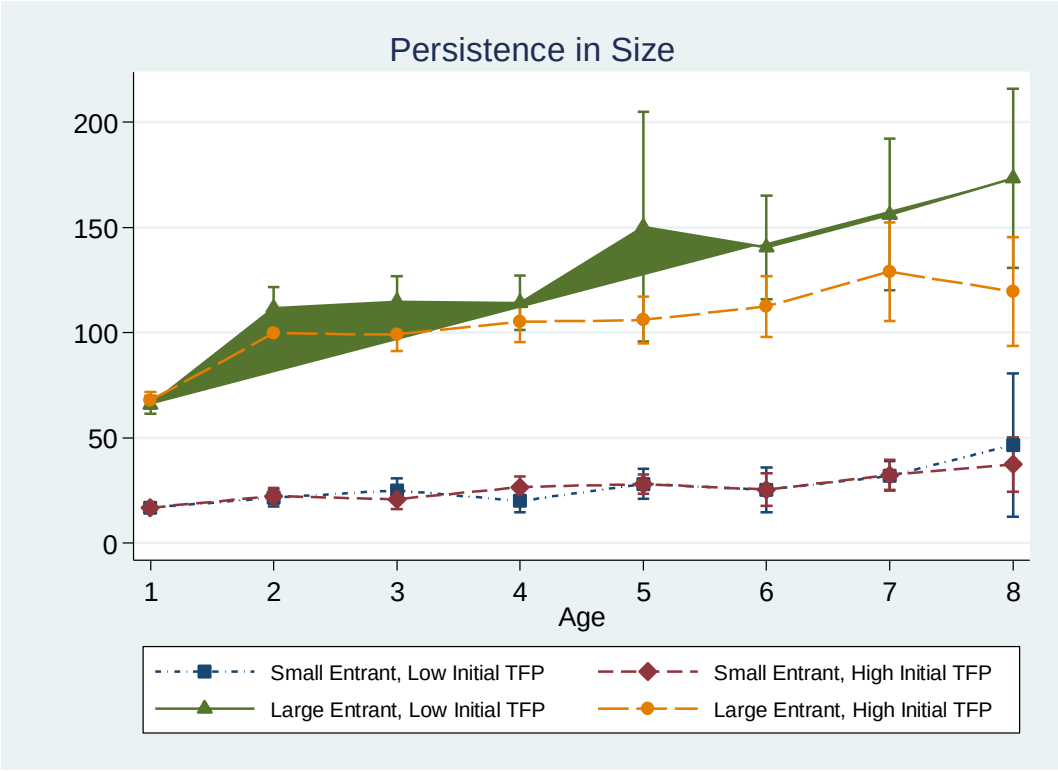


Figure 3: Persistence in Size – Initial Size vs. Initial Leverage or Legal Form

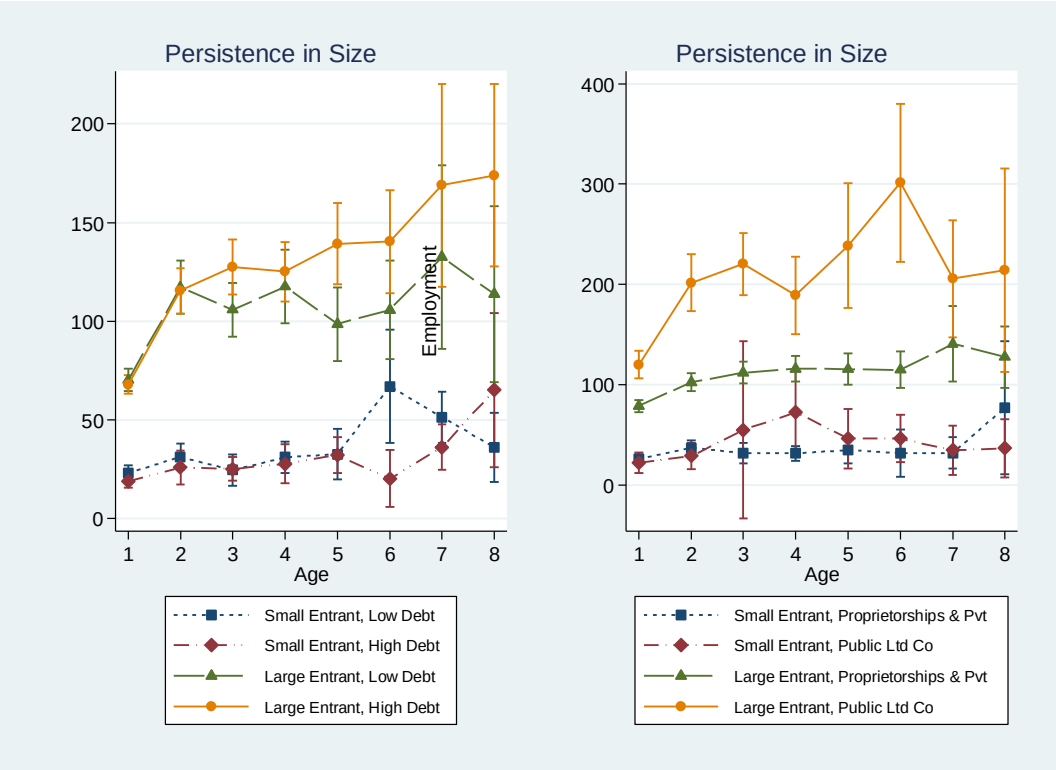
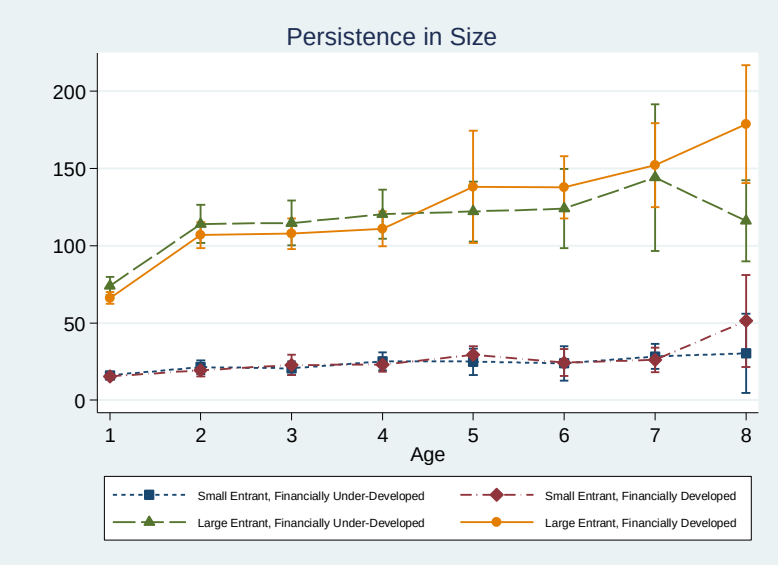
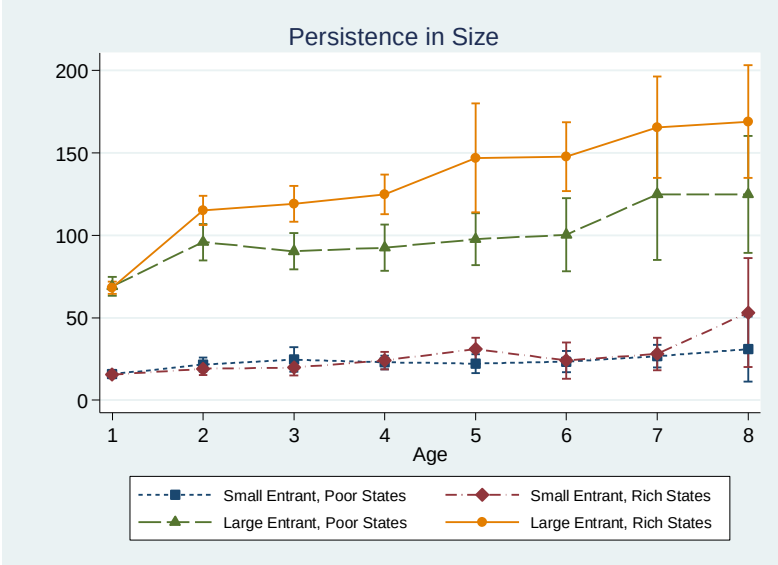
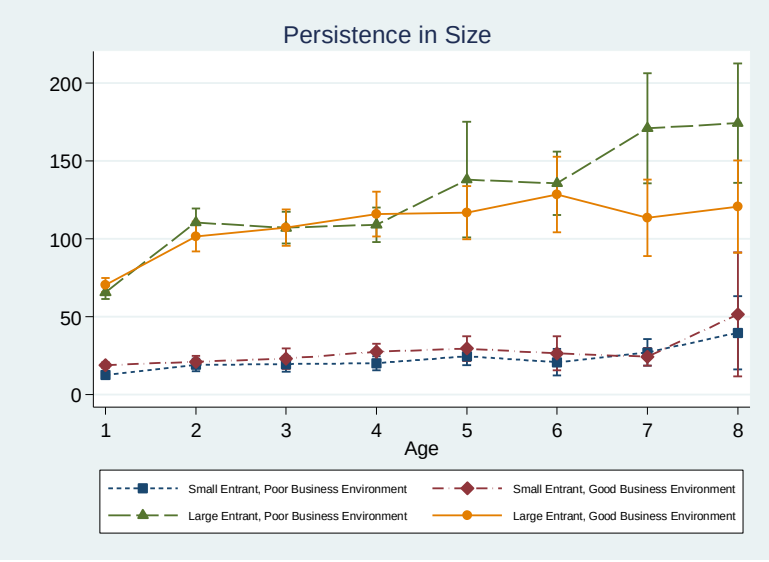


Figure 4: Size over Early Lifecycle – Large Entrant x Institutions







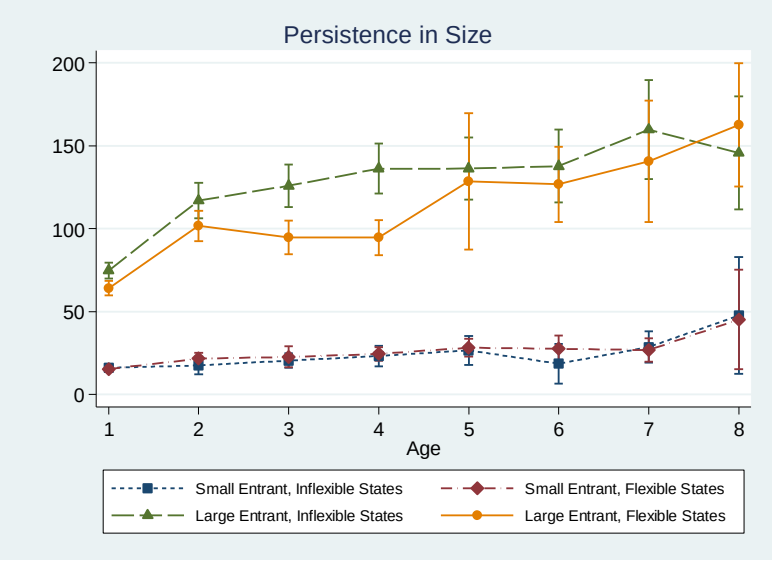


Figure 5: Employment Growth over Early Lifecycle – Large Entrant x Institutions

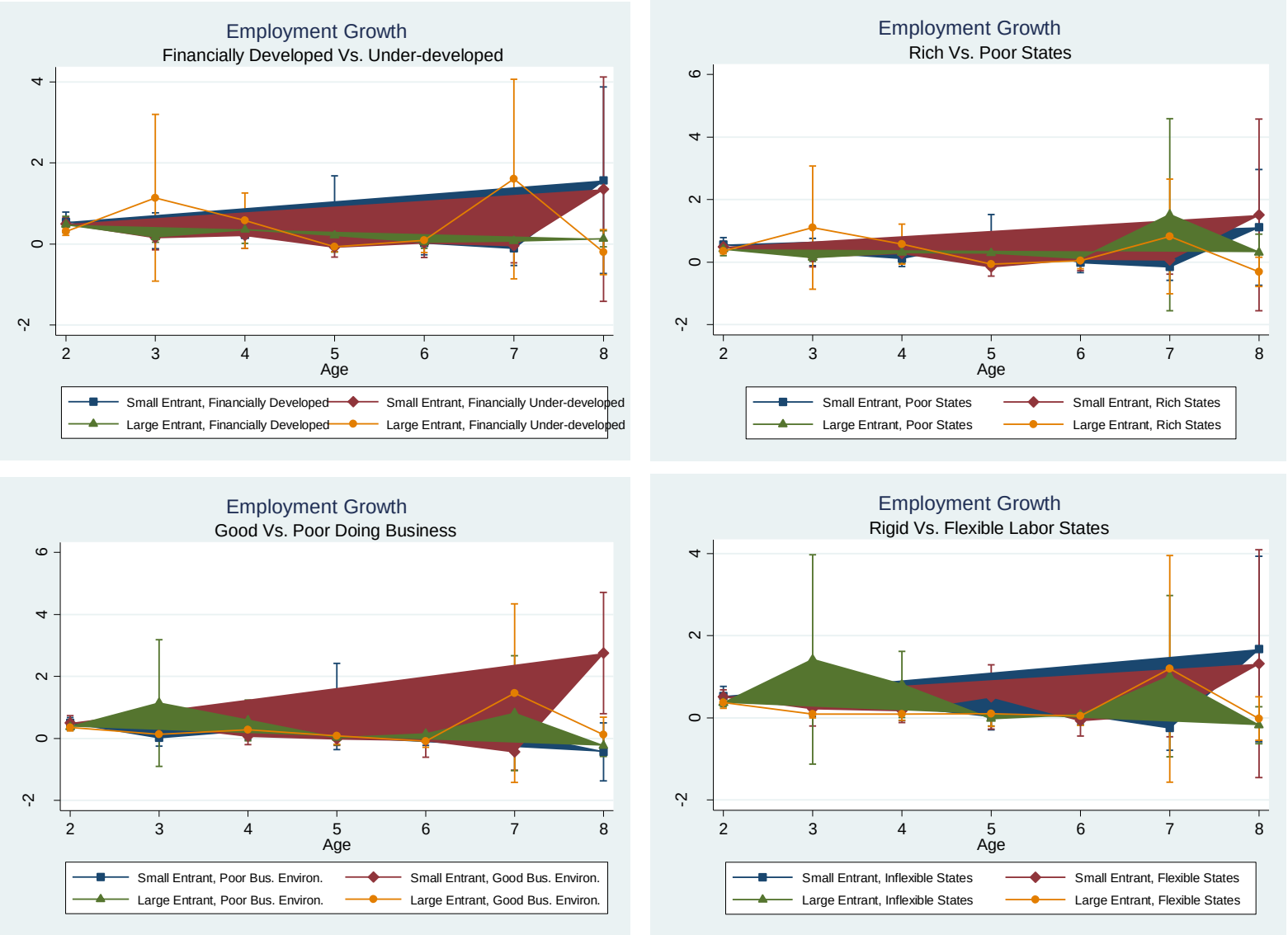
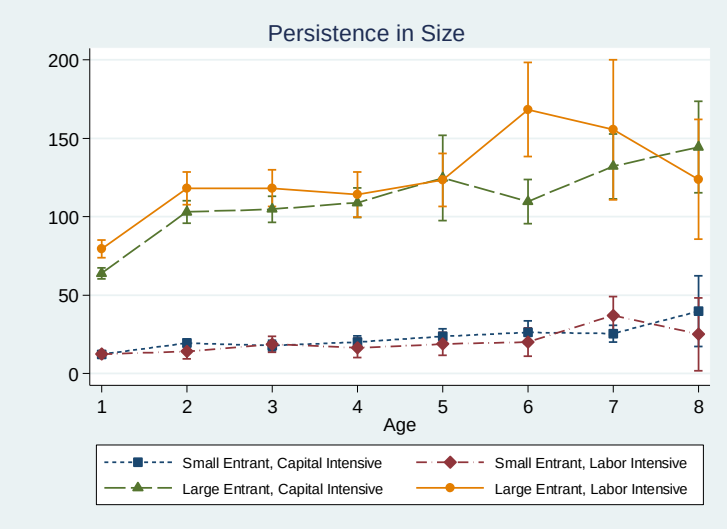
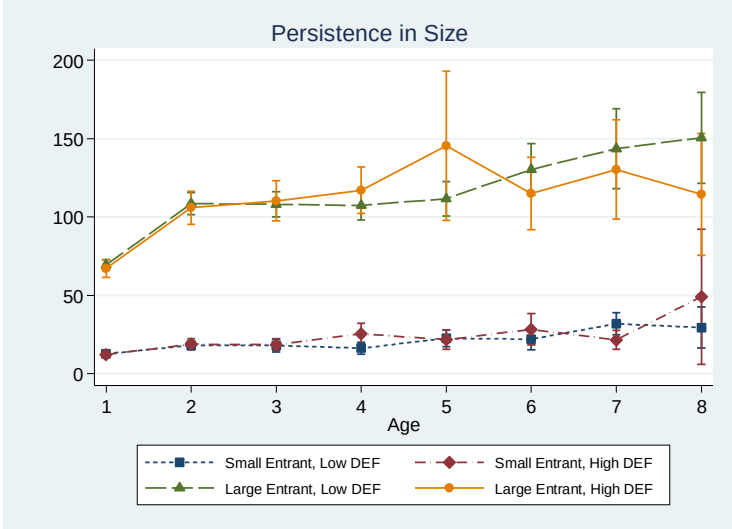


Figure 6: Persistence in Size – Large Entrant x Industry Characteristic





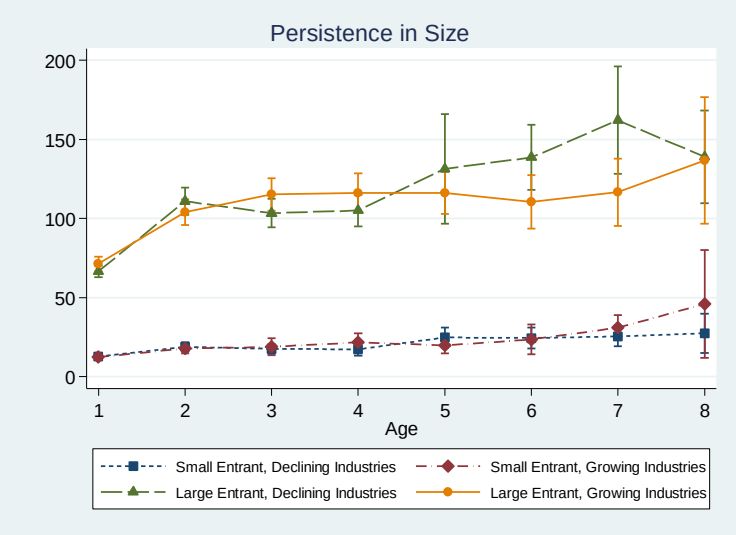


Figure 7: Initial Conditions and TFP

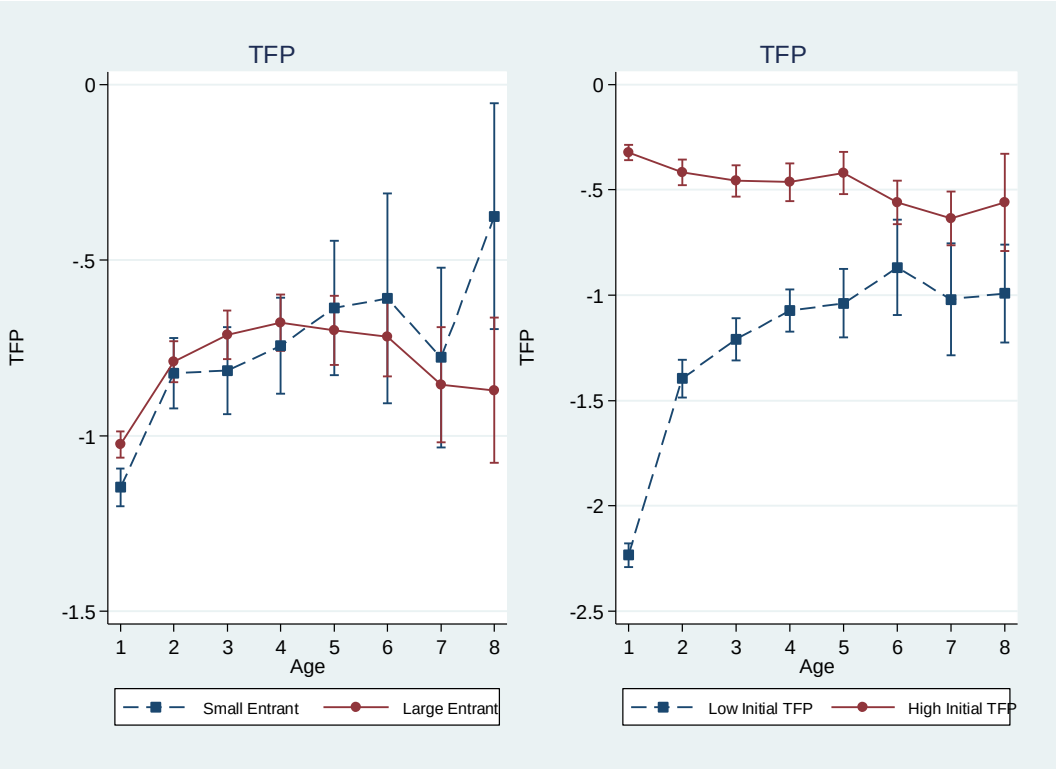


Figure 8: Initial Conditions and Complexity of Production Structure

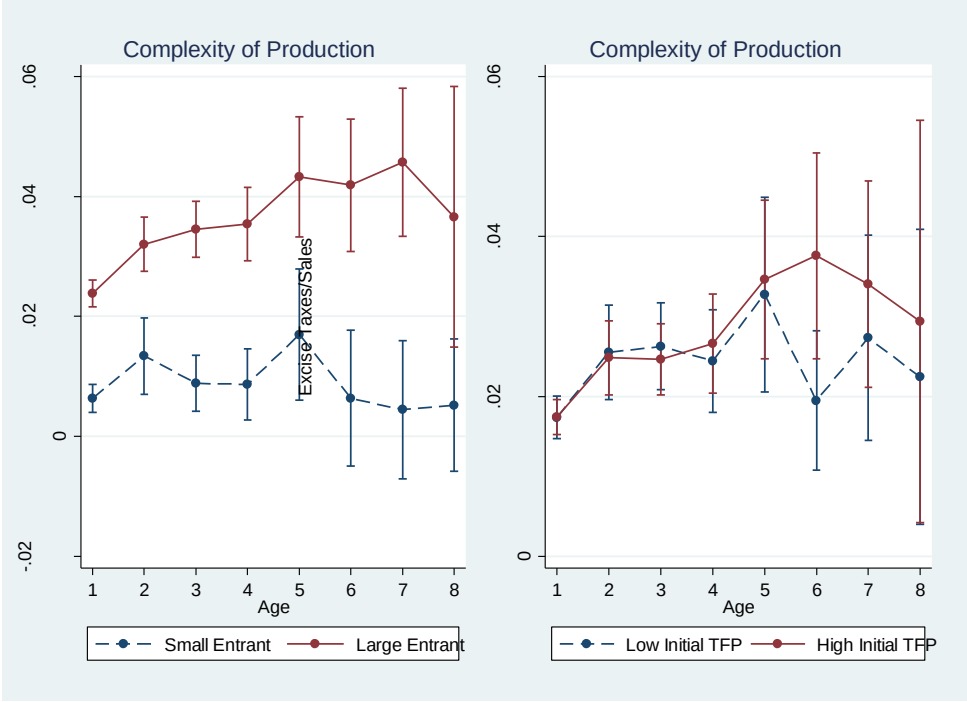


Figure 9: Initial Conditions and Profit

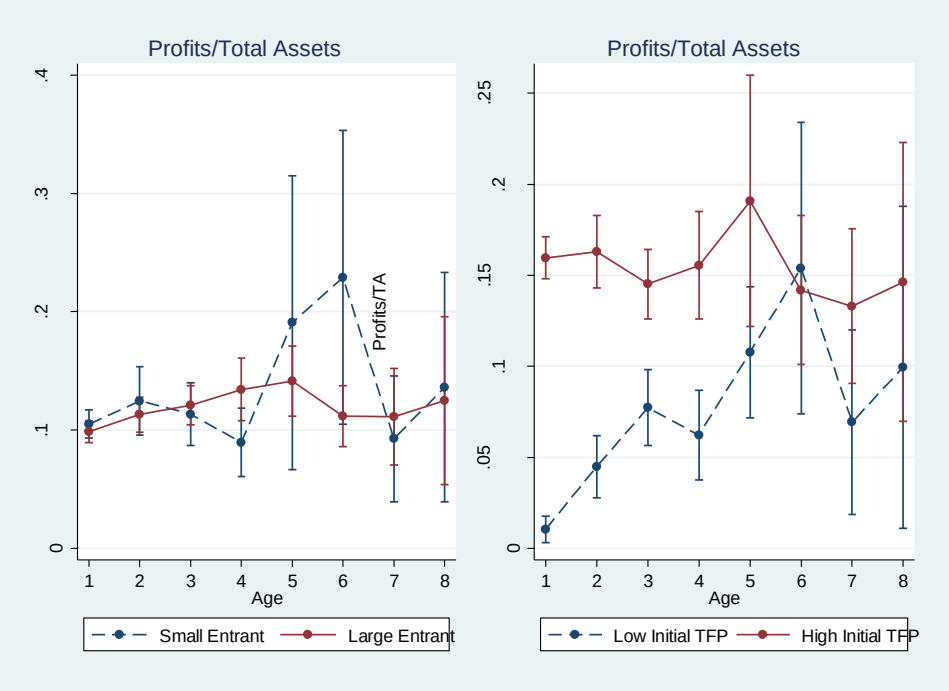


Table 1: Role of Initial Conditions – contribution to adjusted R-square

The table documents how initial conditions contribute to the adjusted R-square of the following regression models when they are entered one at a time: Establishment Size/Employment Growth = $\alpha + \beta_1$ Initial Condition + β_2State Dummies + β_3Industry Dummies + e. Establishment Size is the total number of workers which includes workers employed directly, workers employed through contractors, supervisory and managerial staff, other employees, working proprietors, unpaid family workers, and unpaid working members if cooperative factory. Employment Growth is the annual growth rate in the total number of workers. Initial Condition is one of three variables - Large Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the size distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the size distribution of all entrants over the sample period. High Initial TFP Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the TFP distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the TFP distribution of all entrants over the sample period. Public Limited Co. takes the value 1 if the firm is organized as a public limited company at age 1 or as a private limited company or proprietorship at age 1. All regressions are estimated using sampling weights. The numbers in each row present the incremental contribution to adjusted R-square. Definitions and sources of all variables are provided in the Appendix.

Panel A: Initial Condition = Size at Start up (Large Entrant Dummy)

	1	2	3	4
Dependent Variable	**Establishment Size**	**Establishment Size**	**Employment Growth**	**Employment Growth**
Adj R-sq when we use State Dummies	0.030	0.034	0	-0.001
ΔAdj R-sq when we add Industry Dummies	0.040	0.044	-0.002	-0.002
ΔAdj R-sq when we add Large Entrant Dummy	0.079	0.076	0	0
Total Adj R-sq with State Dummies + Industry Dummies + Large Entrant Dummy	*0.149*	*0.154*	*-0.002*	*-0.003*
N	22476	18030	10079	8092

Panel B: Initial Condition = TFP at Start up

	1	2	3	4
Dependent Variable	**Establishment Size**	**Establishment Size**	**Employment Growth**	**Employment Growth**
Adj R-sq when we use State Dummies	0.033	0.034	0	-0.001
ΔAdj R-sq when we add Industry Dummies	0.042	0.044	-0.003	-0.002
ΔAdj R-sq when we add High Initial TFP	0	0	0	0
Total Adj R-sq with State Dummies + Industry Dummies + High Initial TFP	*0.075*	*0.078*	*-0.003*	*-0.003*
N	18273	18030	8239	8092

Panel C: Initial Condition = Legal Form at Start up

	1	2	3	4
Dependent Variable	**Establishment Size**	**Establishment Size**	**Employment Growth**	**Employment Growth**
Adj R-sq when we use State Dummies	0.031	0.034	0	-0.001
ΔAdj R-sq when we add Industry Dummies	0.044	0.044	-0.002	-0.002
ΔAdj R-sq when we add Public Limited Co.	0.034	0.034	0	0
Total Adj R-sq with State Dummies + Industry Dummies + Legal Form at Start up	*0.109*	*0.112*	*-0.002*	*-0.003*
N	22239	18030	9923	8092

Table 2: Size and Growth over Early Firm Lifecycle

This table shows results from the following regression: Establishment Size/Employment Growth = $\alpha + \beta_1$ Age Dummies + β_2 Large Entrant + β_3 High Initial TFP + β_4 Large Entrant x Age Dummies + β_5 State Dummies + β_6Year Dummies + β_7Industry Dummies + e. Establishment Size is the total number of workers which includes workers employed directly, workers employed through contractors, supervisory and managerial staff, other employees, working proprietors, unpaid family workers, and unpaid working members if cooperative factory. Employment Growth is the annual growth rate in the total number of workers. Age Dummies are based on establishment age which is defined as the year of the census - year of initial production reported by the firms. Large Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the size distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the size distribution of all entrants over the sample period. High Initial TFP Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the TFP distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the TFP distribution of all entrants over the sample period. Robust standard errors are reported in the parentheses. All regressions are estimated using sampling weights. Definitions and sources of all variables are provided in the Appendix.

	(1)	(2)	(3)	(4)
	Establishment Size	Establishment Size	Employment Growth	Employment Growth
2 years	25.391***	5.409***		
	(2.309)	(1.487)		
3 years	26.197***	5.979***	0.060	-0.337***
	(2.631)	(1.947)	(0.328)	(0.106)
4 years	28.347***	6.126***	-0.096	-0.385***
	(2.923)	(2.096)	(0.141)	(0.101)
5 years	38.811***	11.064***	-0.291**	-0.308*
	(7.168)	(2.318)	(0.128)	(0.176)
6 years	40.217***	8.453**	-0.421***	-0.585***
	(4.910)	(3.461)	(0.113)	(0.165)
7 years	52.115***	15.625***	0.135	-0.686***
	(6.913)	(2.773)	(0.515)	(0.182)
8 years	56.093***	25.008***	-0.346	0.403
	(8.447)	(9.354)	(0.242)	(0.660)
Large Entrant	68.543***	50.877***	-0.011	-0.201***
	(1.489)	(1.514)	(0.080)	(0.075)
High Initial TFP	-6.038**	-6.237**	-0.182	-0.181
	(2.500)	(2.512)	(0.163)	(0.162)
2 years x Large Entrant		31.434***		
		(3.527)		
3 years x Large Entrant		31.908***		0.545
		(4.148)		(0.502)
4 years x Large Entrant		35.244***		0.392
		(4.573)		(0.241)
5 years x Large Entrant		42.380***		0.044
		(10.689)		(0.231)
6 years x Large Entrant		47.105***		0.224
		(7.452)		(0.151)
7 years x Large Entrant		55.102***		1.054
		(10.468)		(0.674)
8 years x Large Entrant		47.142***		-0.914
		(14.890)		(0.668)
Constant	-62.051***	-42.754***	0.669***	0.359**
	(7.227)	(7.309)	(0.192)	(0.141)
Fixed Effects	| ---------- Industry, State, Year ---------- |			
N	18273	18273	8239	8239
Adj. R-sq	0.173	0.178	-0.005	-0.005

*, **, and *** represent significance at 10%, 5%, and 1% levels respectively.

Table 3: Size and Growth over Early Firm Lifecycle – Extent of Selection Bias

This table presents estimates from the comparison of the size distribution of firms at age 2 with that at age 8 following the quantile methodology in Combes et al. (2012). Bootstrapped Standard errors for the shift, dilation and truncation parameters are reported in parentheses.

Right-shift Parameter, A	Dilation Parameter, D	Left-Truncation (or Selection) Parameter, S	R^2	Obs.
Panel A: All three parameters estimated				
10.897	1.256	0.054	0.96	4627
(16.104)	(0.206)	(0.167)		
Panel B: Only Shift and Dilation Estimated				
15.267	1.277	-	0.96	4627
(6.656)	(0.158)			
Panel C: Only Shift and Truncation Estimated				
-20.05		0.392	0.63	4627
(8.930)	-	(0.108)		
Panel D: Only Shift Estimated				
15.221	-	-	0.24	4627
(13.538)				
Panel E: Only Truncation Estimated				
-	-	0.253	0.55	4627
		(0.063)		

Table 4: Size and Growth over Early Firm Lifecycle – Initial Conditions vs. Local Institutions

This table shows results from the following regression: Establishment Size/Employment Growth = $\alpha + \beta_1$ Age Dummies + β_2 Large Entrant + β_3 High Initial TFP + β_4 Institution + β_5 Large Entrant x Institution + β_6 Industry Dummies + β_7 Year Dummies + e. Establishment Size is the total number of workers which includes workers employed directly, workers employed through contractors, supervisory and managerial staff, other employees, working proprietors, unpaid family workers, and unpaid working members if cooperative factory. Employment Growth is the annual growth rate in the total number of workers. Age Dummies are based on establishment age which is defined as the year of the census - year of initial production reported by the firms. Large Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the size distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the size distribution of all entrants over the sample period. High Initial TFP Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the TFP distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the TFP distribution of all entrants over the sample period. Institution is one of the following four variables - Rich state dummy takes the value 1 for a particular state in a particular year if that state's GDP/capita is ≥ median value of state GDP/capita in that year across states and 0 for states that are < median value of state GDP/capita; Financially Developed dummy takes the value 1 for a particular state in a particular year if that state is ≥ the median value of financial development in that year across states and 0 for states that are < the median value of financial development; Good Doing Business is a dummy variable that takes the value 1 for states with good doing business environments (DB Rank ≤9) and 0 for states with poor doing business (DB Rank >9); Flexible State dummy that takes the value 1 for states with flexible labor regulation and 0 for states with rigid or neutral labor regulations following Gupta et. al. (2008). Robust standard errors are reported in the parentheses. All regressions are estimated using sampling weights. Definitions and sources of all variables are provided in the Appendix.

Panel A: Establishment Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size
2 years	27.627***	28.524***	26.443***	26.994***	27.725***	28.291***	26.509***	26.469***
	(2.757)	(2.780)	(2.620)	(2.757)	(2.761)	(2.786)	(2.617)	(2.751)
3 years	28.432***	29.478***	26.971***	27.730***	28.565***	29.310***	27.104***	27.417***
	(3.153)	(3.178)	(2.982)	(3.138)	(3.159)	(3.179)	(2.980)	(3.130)
4 years	31.592***	32.893***	30.541***	30.828***	31.649***	32.627***	30.630***	30.938***
	(3.347)	(3.371)	(3.199)	(3.330)	(3.346)	(3.371)	(3.200)	(3.321)
5 years	45.770***	46.830***	44.160***	45.302***	45.715***	46.527***	44.232***	45.115***
	(8.496)	(8.452)	(8.067)	(8.517)	(8.512)	(8.423)	(8.064)	(8.555)
6 years	46.375***	47.713***	46.658***	44.879***	46.328***	47.342***	46.800***	44.264***
	(5.936)	(5.938)	(5.795)	(5.936)	(5.938)	(5.930)	(5.794)	(5.928)
7 years	58.095***	60.851***	56.209***	57.897***	58.350***	60.319***	56.263***	57.491***
	(8.467)	(8.516)	(8.012)	(8.462)	(8.470)	(8.546)	(8.008)	(8.461)
8 years	67.723***	68.855***	65.338***	67.257***	67.561***	68.753***	65.447***	66.806***
	(9.785)	(9.747)	(9.555)	(9.827)	(9.791)	(9.758)	(9.540)	(9.862)
Large Entrant	72.596***	71.948***	71.400***	71.838***	77.021***	63.714***	75.169***	83.484***
	(1.720)	(1.717)	(1.655)	(1.778)	(2.571)	(2.470)	(2.381)	(2.344)
High Initial TFP	-5.802*	-5.630*	-6.635**	-6.209**	-5.903**	-5.480*	-6.602**	-6.420**
	(2.967)	(2.977)	(2.887)	(2.966)	(2.956)	(2.969)	(2.880)	(2.959)
Financially Developed	-1.435				2.723**			
	(2.077)				(1.254)			
Rich State		11.901***				4.417***		
		(1.889)				(1.347)		
Good Doing Business			0.230				5.661***	
			(1.874)				(1.223)	
Flexible Labor State				-11.142***				1.265
				(2.022)				(1.342)
Large Entrant x Financially Developed					-6.660*			

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size
					(3.445)			
Large Entrant x Rich State						12.217***		
						(3.252)		
Large Entrant x Good Doing Business							-8.587***	
							(3.055)	
Large Entrant x Flexible Labor State								-19.243***
								(3.297)
Constant	2.016	-7.646*	1.961	8.657**	-0.641	-2.742	-0.918	0.382
	(4.073)	(3.962)	(3.695)	(4.129)	(3.832)	(3.816)	(3.724)	(3.817)
Fixed Effects	| ----------Industry, Year Fixed Effects----------|							
N	13656	13656	14476	13656	13656	13656	14476	13656
adj. R-sq	0.166	0.167	0.163	0.167	0.166	0.168	0.164	0.169

Panel B: Growth

	(1)	(2)	(3)	(4)	(5)	(6)`	(7)	(8)
	Employment Growth							
3 years	0.242	0.241	0.199	0.230	0.240	0.240	0.196	0.230
	(0.488)	(0.489)	(0.453)	(0.478)	(0.487)	(0.487)	(0.452)	(0.477)
4 years	0.021	0.019	-0.016	-0.001	0.022	0.018	-0.013	0.002
	(0.182)	(0.182)	(0.164)	(0.174)	(0.183)	(0.181)	(0.165)	(0.175)
5 years	-0.331***	-0.333***	-0.320***	-0.342***	-0.332***	-0.339***	-0.332***	-0.341***
	(0.107)	(0.106)	(0.102)	(0.105)	(0.106)	(0.105)	(0.102)	(0.105)
6 years	-0.347***	-0.348***	-0.367***	-0.374***	-0.342***	-0.357***	-0.368***	-0.376***
	(0.106)	(0.105)	(0.114)	(0.124)	(0.104)	(0.113)	(0.114)	(0.125)
7 years	0.469	0.474	0.453	0.435	0.462	0.469	0.453	0.437
	(0.672)	(0.669)	(0.663)	(0.675)	(0.672)	(0.670)	(0.663)	(0.674)
8 years	-0.320	-0.314	-0.304	-0.322	-0.319	-0.314	-0.306	-0.324
	(0.305)	(0.299)	(0.297)	(0.309)	(0.306)	(0.300)	(0.296)	(0.309)
Large Entrant	0.047	0.046	0.055	0.027	-0.096	-0.055	0.188	0.252
	(0.142)	(0.139)	(0.148)	(0.131)	(0.107)	(0.124)	(0.204)	(0.287)
High Initial TFP	-0.192	-0.196	-0.200	-0.214	-0.189	-0.196	-0.200	-0.219
	(0.182)	(0.186)	(0.179)	(0.195)	(0.181)	(0.186)	(0.179)	(0.198)
Financially Developed	0.111				-0.068			
	(0.131)				(0.101)			
Rich State		0.083				-0.039		
		(0.146)				(0.104)		
Good Doing Business			-0.165				0.047	
			(0.172)				(0.120)	

	(1)	(2)	(3)	(4)	(5)	(6)`	(7)	(8)
					Employment Growth			
Flexible Labor State				-0.262				0.026
				(0.217)				(0.097)
Large Entrant x Financially Developed					0.249			
					(0.203)			
Large Entrant x Rich State						0.174		
						(0.262)		
Large Entrant x Good Doing Business							-0.291*	
							(0.171)	
Large Entrant x Flexible Labor State								-0.391
								(0.311)
Constant	1.256***	1.282***	1.398***	1.493**	1.349***	1.349***	1.289***	1.306***
	(0.400)	(0.405)	(0.520)	(0.590)	(0.458)	(0.478)	(0.474)	(0.471)
Fixed Effects					Industry, Year			
N	5835	5835	6207	5835	5835	5835	6207	5835
adj. R-sq	-0.006	-0.006	-0.005	-0.005	-0.006	-0.006	-0.005	-0.006

Table 5: Size and Growth over Early Firm Lifecycle – Industry Heterogeneity

This table shows results from the following regression: Employment Growth/TFP = $\alpha + \beta_1$ Age Dummies + β_2 Large Entrant + β_3 High Initial TFP + β_4 Large Entrant x Age Dummies + β_5 High Initial TFP x Age Dummies + β_6 State Dummies + β_7 Year Dummies + e. Employment Growth is the annual growth rate in the total number of workers. TFP is the logarithm of revenue productivity defined as the product of physical productivity and a firm's output price following Hsieh and Klenow (2009). Age Dummies are based on establishment age which is defined as the year of the census - year of initial production reported by the firms. Large Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the size distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the size distribution of all entrants over the sample period. High Initial TFP Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the TFP distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the TFP distribution of all entrants over the sample period. DEF is based on the Rajan and Zingales (1998) index and is a dummy variable that takes the value 1 if industry's dependence on external finance is ≥ median value of dependence on external finance across industries and 0 if it was < the median across industries. Growing Industry dummy is a dummy variable that takes the value 1 if the industry's growth in employment over the period 2001-2010 was ≥ the median industry growth over this period and 0 if the industry's growth in employment over this period was < than the median. Labor Intensity dummy is a dummy variable that takes the value 1 for labor intensive industries and 0 for capital intensive industries following Hasan and Jandoc (2012). Robust standard errors are reported in the parentheses. All regressions are estimated using sampling weights. Definitions and sources of all variables are provided in the Appendix.

Panel A: Establishment Size

	(1)	(2)	(3)	(4)	(5)	(6)
	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size
2 years	26.987***	26.884***	26.882***	26.902***	26.884***	26.882***
	(2.352)	(2.354)	(2.354)	(2.348)	(2.354)	(2.354)
3 years	27.609***	27.483***	27.467***	27.524***	27.482***	27.468***
	(2.624)	(2.627)	(2.625)	(2.624)	(2.627)	(2.625)
4 years	28.988***	28.730***	28.726***	28.840***	28.730***	28.727***
	(2.956)	(2.956)	(2.955)	(2.958)	(2.956)	(2.955)
5 years	39.589***	39.582***	39.554***	39.657***	39.584***	39.559***
	(7.070)	(7.023)	(7.047)	(7.068)	(7.004)	(7.038)
6 years	42.028***	41.953***	41.940***	42.363***	41.953***	41.946***
	(5.074)	(5.101)	(5.097)	(5.052)	(5.101)	(5.095)
7 years	52.280***	52.031***	52.039***	51.916***	52.028***	52.035***
	(7.149)	(7.164)	(7.150)	(7.146)	(7.163)	(7.155)
8 years	54.677***	54.038***	54.057***	54.517***	54.038***	54.075***
	(8.768)	(8.758)	(8.757)	(8.748)	(8.758)	(8.755)
Large Entrant	74.197***	74.569***	74.559***	70.192***	74.624***	74.383***
	(1.418)	(1.421)	(1.395)	(1.678)	(1.533)	(1.970)
High Initial TFP	-5.269***	-4.866**	-4.786**	-5.588***	-4.868**	-4.792**
	(2.006)	(2.133)	(1.972)	(2.003)	(2.125)	(1.963)
Labor Intensive	9.039***			-0.864		
	(2.024)			(1.343)		
High Dependence on External Finance		0.681			0.779	
		(2.173)			(1.205)	

	(1)	(2)	(3)	(4)	(5)	(6)
	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size
Growing Industry			-0.075			-0.322
			(1.735)			(1.046)
Labor Intensive x Large Entrant				15.071***		
				(2.960)		
High Dependence on External Finance x Large Entrant					-0.157	
					(3.175)	
Growing Industry x Large Entrant						0.390
						(2.766)
Constant	-17.807***	-19.857***	-19.141***	-13.613***	-19.852***	-19.104***
	(4.155)	(4.493)	(4.529)	(4.227)	(4.531)	(4.435)
Fixed Effects	|---------- State, Year ----------|					
N	18273	18266	18273	18273	18266	18273
adj. R-sq	0.149	0.148	0.148	0.150	0.148	0.148

Panel B: Employment Growth

	(1)	(2)	(3)	(4)	(5)	(6)
	Employment Growth	Employment Growth	Employment Growth	Employment Growth	Employment Growth	Employment Growth
3 years	0.063	0.063	0.062	0.062	0.063	0.061
	(0.327)	(0.328)	(0.326)	(0.327)	(0.328)	(0.326)
4 years	-0.095	-0.095	-0.096	-0.095	-0.095	-0.096
	(0.139)	(0.139)	(0.139)	(0.139)	(0.139)	(0.139)
5 years	-0.274**	-0.273**	-0.272**	-0.272**	-0.273**	-0.272**
	(0.125)	(0.125)	(0.125)	(0.125)	(0.125)	(0.125)
6 years	-0.418***	-0.417***	-0.419***	-0.419***	-0.417***	-0.419***
	(0.110)	(0.112)	(0.113)	(0.110)	(0.112)	(0.113)
7 years	0.143	0.143	0.141	0.143	0.144	0.139
	(0.508)	(0.511)	(0.510)	(0.509)	(0.511)	(0.512)

	(1)	(2)	(3)	(4)	(5)	(6)
	Employment Growth	Employment Growth	Employment Growth	Employment Growth	Employment Growth	Employment Growth
8 years	-0.323	-0.323	-0.327	-0.324	-0.323	-0.328
	(0.233)	(0.232)	(0.234)	(0.232)	(0.232)	(0.235)
Large Entrant	0.019	0.021	0.020	-0.012	0.016	0.050
	(0.085)	(0.082)	(0.082)	(0.076)	(0.104)	(0.103)
High Initial TFP	-0.204	-0.203	-0.199	-0.206	-0.203	-0.198
	(0.165)	(0.168)	(0.165)	(0.165)	(0.168)	(0.164)
Labor Intensive	0.023			-0.074		
	(0.076)			(0.140)		
High Dependence on External Finance		0.004			-0.006	
		(0.058)			(0.084)	
Growing Industry			-0.032			0.014
			(0.068)			(0.082)
Labor Intensive x Large Entrant				0.129		
				(0.132)		
High Dependence on External Finance x Large Entrant					0.014	
					(0.115)	
Growing Industry x Large Entrant						-0.064
						(0.110)
Constant	0.463***	0.466***	0.476***	0.494***	0.469***	0.456***
	(0.154)	(0.151)	(0.151)	(0.152)	(0.163)	(0.156)
Fixed Effects	| ---------- State, Year ----------|					
N	8239	8233	8239	8239	8233	8239
Adj. R-sq	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002

Table 6: How are large vs. small entrants, high vs. low initial TFP entrants different?

This table shows results from the following regression: Complexity of Production Structure/Profits/TFP $= \alpha + \beta_1$ Age Dummies + β_2 Large Entrant + β_3 Large Entrant x Age Dummies + β_4 State Dummies + β_5 Year Dummies + β_6 Industry Dummies + e. Complexity of Production Structure is defined as the ratio of Excise Taxes paid/Sales following Siegel and Choudhury (2012). Profits is defined as the ratio of Profits to Total Assets; TFP is the logarithm of revenue productivity defined as the product of physical productivity and a firm's output price following Hsieh and Klenow (2009); Age Dummies are based on establishment age which is defined as the year of the census - year of initial production reported by the firms. Large Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the size distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the size distribution of all entrants (i.e. firms aged 1) over the sample period. High Initial TFP Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the TFP distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the TFP distribution of all entrants over the sample period. Robust standard errors are reported in the parentheses. All regressions are estimated using sampling weights. Definitions and sources of all variables are provided in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	TFP	TFP	TFP	Complexity of Production Structure	Complexity of Production Structure	Complexity of Production Structure	Profits	Profits	Profits
2 years	0.268***	0.325***	0.838***	0.008***	0.007**	0.008**	0.016**	0.019	0.034***
	(0.031)	(0.058)	(0.055)	(0.002)	(0.004)	(0.003)	(0.008)	(0.016)	(0.009)
3 years	0.322***	0.332***	1.024***	0.008***	0.002	0.009***	0.018**	0.008	0.067***
	(0.036)	(0.069)	(0.059)	(0.002)	(0.003)	(0.003)	(0.008)	(0.015)	(0.011)
4 years	0.369***	0.403***	1.160***	0.008***	0.002	0.007**	0.018	-0.016	0.052***
	(0.040)	(0.075)	(0.059)	(0.003)	(0.003)	(0.004)	(0.011)	(0.016)	(0.013)
5 years	0.387***	0.511***	1.195***	0.016***	0.011*	0.015**	0.056**	0.086	0.097***
	(0.050)	(0.101)	(0.089)	(0.004)	(0.006)	(0.006)	(0.024)	(0.063)	(0.019)
6 years	0.379***	0.538***	1.364***	0.013***	-0.000	0.002	0.045**	0.124*	0.143***
	(0.064)	(0.155)	(0.119)	(0.005)	(0.006)	(0.005)	(0.021)	(0.064)	(0.041)
7 years	0.237***	0.370***	1.213***	0.014***	-0.002	0.010	0.005	-0.012	0.059**
	(0.074)	(0.133)	(0.139)	(0.005)	(0.006)	(0.007)	(0.017)	(0.028)	(0.026)
8 years	0.356***	0.772***	1.242***	0.009	-0.001	0.005	0.028	0.031	0.089**
	(0.095)	(0.167)	(0.121)	(0.008)	(0.006)	(0.009)	(0.030)	(0.050)	(0.045)
Large Entrant	0.070***	0.122***	0.070***	0.020***	0.017***	0.020***	-0.007	-0.007	-0.007
	(0.025)	(0.033)	(0.024)	(0.001)	(0.002)	(0.001)	(0.008)	(0.008)	(0.008)
High Initial TFP	1.314***	1.316***	1.912***	0.001	0.001	0.000	0.114***	0.114***	0.149***
	(0.028)	(0.028)	(0.036)	(0.001)	(0.001)	(0.002)	(0.006)	(0.006)	(0.007)
2 years x Large Entrant		-0.089			0.001			-0.005	
		(0.067)			(0.004)			(0.018)	
3 years x Large Entrant		-0.020			0.008**			0.014	
		(0.079)			(0.004)			(0.018)	
4 years x Large Entrant		-0.057			0.009**			0.051**	
		(0.086)			(0.005)			(0.021)	
5 years x Large Entrant		-0.186			0.009			-0.043	

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	TFP	TFP	TFP	Complexity of Production Structure	Complexity of Production Structure	Complexity of Production Structure	Profits	Profits	Profits
		(0.113)			(0.008)			(0.064)	
6 years x Large Entrant		-0.232			0.018**			-0.111*	
		(0.165)			(0.008)			(0.065)	
7 years x Large Entrant		-0.200			0.024***			0.025	
		(0.157)			(0.009)			(0.035)	
8 years x Large Entrant		-0.618***			0.014			-0.005	
			-0.934***		(0.012)			(0.062)	
2 years x High Initial TFP			(0.064)			-0.001			-0.031**
			-1.160***			(0.004)			(0.015)
3 years x High Initial TFP			(0.071)			-0.002			-0.081***
			-1.302***			(0.004)			(0.016)
4 years x High Initial TFP			(0.075)			0.002			-0.056***
			-1.293***			(0.005)			(0.020)
5 years x High Initial TFP			(0.102)			0.002			-0.066*
			-1.601***			(0.008)			(0.040)
6 years x High Initial TFP			(0.130)			0.018**			-0.161***
			-1.527***			(0.008)			(0.046)
7 years x High Initial TFP			(0.153)			0.007			-0.085**
			-1.480***			(0.009)			(0.034)
8 years x High Initial TFP			(0.169)			0.007			-0.102*
						(0.016)			(0.060)
Constant	-1.570***	-1.625***	-2.201***	-0.006	-0.003	-0.005	-0.113***	-0.113***	-0.151***
	(0.103)	(0.104)	(0.099)	(0.005)	(0.005)	(0.005)	(0.026)	(0.026)	(0.026)
Fixed Effects	| ---------- Industry, Year ---------- |								
N	17652	17652	17652	11434	11434	11434	18188	18188	18188
Adj. R-sq	0.511	0.512	0.554	0.177	0.178	0.177	0.144	0.146	0.149

Table 7: Summary Statistics on Entry

The variables are defined as follows: An entrant is a firm at age 1. Average Size of Entrant is the establishment size at age 1 where establishment size is defined as the total number of workers which includes workers employed directly, workers employed through contractors, supervisory and managerial staff, other employees, working proprietors, unpaid family workers, and unpaid working members if cooperative factory. Definitions and sources of all variables are provided in the Appendix.

Year	# of Entrants	% of Entrants	Size Distribution of Entrants (%)					Average Size of Entrant
		Full Sample	1-5 employees	6-20 employees	21-50 employees	51-100 employees	100+ employees	
2001	2422	3.77	4.60	43.91	30.50	11.88	9.10	42.70
2002	1674	2.96	5.14	46.12	27.46	14.66	6.62	37.78
2003	2004	3.50	9.03	48.76	22.12	10.79	9.30	37.99
2004	1889	3.35	7.03	50.04	25.20	10.06	7.67	38.09
2005	2948	4.99	4.78	49.24	26.51	12.31	7.16	39.12
2006	3529	5.79	4.59	45.87	27.56	11.87	10.12	46.64
2007	3343	5.32	4.21	44.10	29.24	12.69	9.76	45.32
2008	3025	4.90	4.91	45.67	26.86	12.82	9.73	49.60
2009	3113	4.76	5.49	44.72	25.17	12.85	11.78	52.62
2010	2365	3.64	9.32	40.65	27.46	11.67	10.90	47.13

Table 8: Initial Conditions and Role of Institutions

The regression estimated is: Initial Size/Initial TFP = $\alpha + \beta_1$ Credit/SDP + β_2 Rich State + β_3 DB Rank + β_4Literacy Rate+ β_5 Flexible State + e. The variables are defined as follows: Initial Size is the total number of workers at age 1 which includes workers employed directly, workers employed through contractors, supervisory and managerial staff, other employees, working proprietors, unpaid family workers, and unpaid working members if cooperative factory. Initial TFP is the logarithm of revenue productivity at age 1, defined as the product of physical productivity and a firm's output price following Hsieh and Klenow (2009). Credit/SDP is the ratio of total Commercial Bank Credit outstanding to the Net State Domestic Product (SDP) in each census year and gauges the depth of financial development. DB Rank is the easy of doing business rank for states and ranges from 1 (good) to 17 (poor) with higher values corresponding to states with worse overall doing business environments. Flexible State dummy that takes the value 1 for states with flexible labor regulation and 0 for states with rigid or neutral labor regulations following Gupta et. al. (2008). Literacy Rate is the proportion of persons who can both read and write with understanding in any language among population aged 7 years and above. Robust standard errors are reported in the parentheses. All regressions are estimated using sampling weights.

Panel A:

	(1)	(2)	(3)	(4)	(5)	(6)
	Initial Size	Initial Size	Initial Size	Initial TFP	Initial TFP	Initial TFP
Rich State Dummy	0.664		7.307**	-0.025		-0.083
	(3.278)		(3.661)	(0.068)		(0.076)
Credit/SDP		-15.645***	-19.001***		0.134	0.173
		(3.725)	(4.127)		(0.097)	(0.108)
DB Rank	0.164	-0.022	0.406	0.014**	0.016***	0.012*
	(0.305)	(0.217)	(0.313)	(0.006)	(0.005)	(0.006)
Flexible State Dummy	-5.918**	-4.796**	-7.886***	-0.178***	-0.197***	-0.163***
	(2.626)	(2.198)	(2.724)	(0.054)	(0.052)	(0.053)
Literacy Rate	-0.022	0.504**	0.116	-0.003	-0.009*	-0.004
	(0.289)	(0.210)	(0.292)	(0.006)	(0.005)	(0.006)
Constant	36.957**	24.711*	48.179**	-1.796***	-1.362***	-1.628***
	(17.039)	(14.225)	(18.772)	(0.358)	(0.360)	(0.389)
Fixed Effects		Industry, Year			Industry, Year	
N	7250	7250	7250	5819	5819	5819
Adj. R-sq	0.054	0.056	0.057	0.329	0.330	0.330

*, **, and *** represent significance at 10%, 5%, and 1% levels respectively.

Panel B: Cross-State/Year Regressions

	(1)	(2)	(3)
	Percentage of Entrants	Average Entrant Size	Average Entrant TFP
	OLS	OLS	OLS
Rich State dummy	-0.022***	17.048	0.016
	(0.008)	(10.332)	(0.073)
Credit/SDP	0.018***	-27.262***	0.126
	(0.006)	(9.940)	(0.087)
DB Rank	-0.001	-0.555	-0.005
	(0.001)	(0.641)	(0.005)
Flexible State dummy	0.012**	-5.837	0.037
	(0.006)	(4.703)	(0.045)
Literacy Rate	0.000	0.329	0.002
	(0.001)	(0.472)	(0.005)
Constant	0.018***	-27.262***	0.126
	(0.006)	(9.940)	(0.087)
N	150	150	150
Adj. R-sq	0.105	0.086	0.023

Web Appendix
Table A1: Summary Statistics and Correlations

Panel A presents summary statistics, panel B presents correlations, and panel C presents test of means between small and large entrants. The variables are defined as follows: Establishment Size is the total number of workers which includes workers employed directly, workers employed through contractors, supervisory and managerial staff, other employees, working proprietors, unpaid family workers, and unpaid working members if cooperative factory. Employment Growth is the annual growth rate in the total number of workers. Age is defined as the year of the census - year of initial production reported by the firms. Large Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the size distribution of all entrants (i.e. firms aged 1) over the sample period and 0 if it is in the bottom two quintiles of the size distribution of all entrants over the sample period. High Initial TFP Entrant is a dummy variable that takes value 1 if the establishment is in the top 3 quintiles of the TFP distribution of all entrants over the sample period and 0 if it is in the bottom two quintiles of the TFP distribution of all entrants over the sample period. DEF is based on the Rajan and Zingales (1998) index and is a dummy variable that takes the value 1 if industry's dependence on external finance is ≥ median value of dependence on external finance across industries and 0 if it is < the median across industries. Growing Industry Dummy is a dummy variable that takes the value 1 if the industry's growth in employment over the period 2001-2010 is ≥ the median industry growth over this period and 0 if the industry's growth in employment over this period is < than the median. Labor Intensity Dummy is a dummy variable that takes the value 1 for labor intensive industries and 0 for capital intensive industries following Hasan and Jandoc (2012). Rich state is a dummy variable that takes the value 1 for a particular state in a particular year if that state's GDP/capita is ≥ median value of state GDP/capita in that year across states and 0 for states that are < median value of state GDP/capita in that year. Financially Developed is a dummy variable that takes the value 1 for a particular state in a particular year if that state is ≥ the median value of financial development in that year across states and 0 for states that are < the median value of financial development. DB Rank is the easy of doing business rank for states and ranges from 1 (good) to 17 (poor) with higher values corresponding to states with worse overall doing business environments. Flexible State is a dummy variable that takes the value 1 for states with flexible labor regulation and 0 for states with rigid or neutral labor regulations following Gupta et. al. (2008).

Panel A: Summary Statistics

	N	Mean	SD	Min	Max
Establishment Size	22476	104.05	165.50	1	1285
Employment Growth	10079	0.493	11.32	-0.996	1056
Large Entrant	9965	0.676	0.468	0	1
High Initial TFP	8064	0.617	0.486	0	1
Age	22854	2.477	1.798	1	8
Labor Intensity dummy	22854	0.296	0.457	0	1
Growing Industry dummy	22853	0.458	0.498	0	1
DEF	22800	0.339	0.473	0	1
Rich State	17295	0.669	0.470	0	1
Financially Developed	17295	0.642	0.480	0	1
DB Rank	18332	9.325	4.605	1	17
Flexible State	17295	0.522	0.499	0	1

Panel B: Correlations

	Establishment Size	Employment Growth	Large Entrant	High Initial TFP	Age	Labor Intensity dummy	Growing Industry dummy	DEF	Rich State	Financially Developed	DB Rank
Employment Growth	0.051***										
Large Entrant	0.331***	-0.003									
High Initial TFP	0.002	-0.018*	0.075***								
Age	0.230***	-0.007	0.079***	0.008							
Labor Intensity dummy	0.047***	-0.006	0.068***	0.093***	0.018***						
Growing Industry dummy	-0.026***	-0.008	0.019***	0.115***	-0.008	0.177***					
DEF	-0.019***	-0.011	-0.048***	0.143***	-0.023***	-0.199***	-0.008				
Rich State	0.122***	0.010	0.108***	0.005	-0.021**	-0.025***	-0.03***	0.028***			
Financially Developed	0.037***	0.014	0.070***	-0.053***	-0.010	-0.062***	-0.021***	-0.009	0.689***		
DB Rank	0.043***	0.022*	0.058***	0.031***	0.034***	0.020**	-0.061***	0.004	-0.117***	0.104***	
Flexible State	-0.061***	-0.014	-0.073***	0.013	-0.054***	0.119***	-0.014**	-0.032**	-0.064***	0.129***	0.099***

*, **, and *** represent significance at 10%, 5%, and 1% levels respectively.

Panel C: Large vs. Small Entrants

	Small Entrants	**Large Entrants**
Size	10.26	90.99***
Cash/Total Assets	0.053	0.041***
Total Loans/Total Assets	0.578	0.548***
Profits/Total Assets	0.045	0.067***
Complexity of Production Structure	0.013	0.032***
Productivity	-1.27	-0.97***

Web Appendix
Table A2: Regressions underlying Figure 3.

	(1)	(2)	(3)
	Establishment Size	Establishment Size	Establishment Size
2 years	7.983**	8.187**	4.634***
	(3.131)	(3.720)	(1.525)
3 years	4.436	1.542	3.884**
	(2.754)	(4.325)	(1.786)
4 years	9.143**	8.018*	4.373**
	(3.634)	(4.388)	(2.035)
5 years	12.350***	9.611	8.906***
	(4.088)	(6.816)	(2.382)
6 years	7.294	43.949***	5.935*
	(7.489)	(14.669)	(3.479)
7 years	22.783***	28.274***	13.045***
	(4.934)	(6.731)	(2.752)
8 years	37.972**	13.156	22.316**
	(16.288)	(9.167)	(9.879)
Large Entrant	48.783***	47.152***	45.807***
	(2.096)	(3.318)	(1.523)
High Initial Debt	5.284**	-4.187*	
	(2.510)	(2.225)	
High Initial TFP	-7.283**	-7.064**	-3.854
	(2.866)	(2.862)	(2.530)
2 years x Large Entrant	39.019***	38.816***	25.700***
	(5.654)	(8.177)	(3.435)
3 years x Large Entrant	45.030***	34.133***	26.544***
	(5.953)	(8.550)	(3.933)
4 years x Large Entrant	44.004***	39.318***	31.137***
	(7.078)	(10.712)	(4.443)
5 years x Large Entrant	39.039***	18.725	37.762***
	(8.264)	(11.750)	(11.401)
6 years x Large Entrant	49.435***	-8.437	39.650***
	(12.002)	(19.458)	(7.023)
7 years x Large Entrant	61.678***	34.021	48.011***
	(18.921)	(24.773)	(10.940)
8 years x Large Entrant	42.192*	30.397	40.243***
	(23.218)	(24.555)	(15.222)
2 years x High Initial Debt		-1.088	
		(5.744)	
3 years x High Initial Debt		4.770	
		(5.369)	
4 years x High Initial Debt		0.897	
		(6.628)	

	(1)	(2)	(3)
	Establishment Size	Establishment Size	Establishment Size
5 years x High Initial Debt		3.650	
		(8.271)	
6 years x High Initial Debt		-42.577***	
		(16.473)	
7 years x High Initial Debt		-10.922	
		(8.888)	
8 years x High Initial Debt		33.100	
		(21.922)	
Large Entrant x High Initial Debt		1.871	
		(4.309)	
2 years x Large Entrant x High Initial Debt		1.539	
		(11.183)	
3 years x Large Entrant x High Initial Debt		19.257	
		(11.873)	
4 years x Large Entrant x High Initial Debt		8.930	
		(14.391)	
5 years x Large Entrant x High Initial Debt		39.316**	
		(16.662)	
6 years x Large Entrant x High Initial Debt		79.510***	
		(24.819)	
7 years x Large Entrant x High Initial Debt		49.650	
		(36.868)	
8 years x Large Entrant x High Initial Debt		29.385	
		(39.521)	
Public Limited Co.			-0.822
			(4.635)
2 years x Public Ltd. Co.			6.196
			(7.648)
3 years x Public Ltd. Co.			41.729
			(44.297)
4 years x Public Ltd. Co.			62.933***
			(18.666)
5 years x Public Ltd. Co.			13.263
			(12.811)
6 years x Public Ltd. Co.			15.287**
			(7.588)
7 years x Public Ltd. Co.			-3.452
			(8.745)
8 years x Public Ltd. Co.			3.005
			(14.788)
Large Entant x Public Ltd. Co.			57.515***
			(8.366)

	(1)	(2)	(3)
	Establishment Size	Establishment Size	Establishment Size
2 years x Large Entrant x Public Ltd. Co.			43.125**
			(18.673)
3 years x Large Entrant x Public Ltd. Co.			32.639
			(47.750)
4 years x Large Entrant x Public Ltd. Co.			-29.225
			(28.273)
5 years x Large Entrant x Public Ltd. Co.			60.818
			(37.438)
6 years x Large Entrant x Public Ltd. Co.			129.642***
			(43.353)
7 years x Large Entrant x Public Ltd. Co.			30.474
			(34.043)
8 years x Large Entrant x Public Ltd. Co.			28.371
			(55.905)
Constant	-50.639***	-42.668***	-45.388***
	(11.824)	(11.717)	(7.304)
Fixed Effects		State, Industry, Year	
N	8678	8678	18030
adj. R-sq	0.179	0.181	0.210

Web Appendix

Table A3: Robustness of Table 4 with Continuous Measure of Initial Size

In this table we repeat the specification in panels A and B of Table 3 but replace Large Entrant with a continuous measure of firm size at age 1. All the regressions have the full set of control variables (Age dummies, High initial TFP dummy, industry and year fixed effects) as in Table 4 but are not shown here for the sake of brevity.

Panel A: Establishment Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size	Establishment Size
Initial Size	1.079***	1.077***	1.080***	1.078***	1.125***	1.073***	1.064***	1.112***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.023)	(0.023)	(0.015)	(0.021)
Financially Developed	1.211				5.337***			
	(1.424)				(1.631)			
Rich State		5.457***				5.194***		
		(1.261)				(1.512)		
Good Doing Business			-0.607				-2.607*	
			(1.293)				(1.468)	
Flexible Labor State				-3.456**				0.281
				(1.393)				(1.581)
Initial Size x Financially Developed					-0.075***			
					(0.028)			
Initial Size x Rich State						0.005		
						(0.028)		
Initial Size x Good Doing Business							0.037	
							(0.028)	
Initial Size x Flexible Labor State								-0.070**
								(0.027)
Constant	-4.361***	-7.475***	-2.943*	-1.120	-6.658***	-7.303***	-1.819	-3.839**
	(1.595)	(1.554)	(1.579)	(1.655)	(1.714)	(1.600)	(1.462)	(1.785)
N	13656	13656	14476	13656	13656	13656	14476	13656
Adj. R-sq	0.674	0.675	0.674	0.674	0.675	0.675	0.675	0.675

Panel B: Employment Growth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Employment Growth	Employment Growth	Employment Growth	Employment Growth	Employment Growth	Employment Growth	Employment Growth	Employment Growth
Initial Size	0.000	0.000	0.000	0.000	-0.001**	-0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Financially Developed	0.113				0.017			
	(0.136)				(0.088)			
Rich State		0.085				0.039		
		(0.141)				(0.107)		
Good Doing Business			-0.166				-0.090	
			(0.173)				(0.115)	
Flexible Labor State				-0.264				-0.118
				(0.218)				(0.141)
Initial Size x Financially Developed					0.001			
					(0.001)			
Initial Size x Rich State						0.001		
						(0.001)		
Initial Size x Good Doing Business							-0.001	
							(0.001)	
Initial Size x Flexible Labor State								-0.002
								(0.001)
Constant	1.277***	1.306***	1.424***	1.508**	1.325***	1.332***	1.377***	1.415**
	(0.417)	(0.431)	(0.546)	(0.608)	(0.447)	(0.451)	(0.507)	(0.558)
N	5835	5835	6207	5835	5835	5835	6207	5835
Adj. R-sq	-0.006	-0.006	-0.005	-0.005	-0.006	-0.006	-0.005	-0.005

Figure A1: Size - Alternate Definitions of Large and Small Entrants

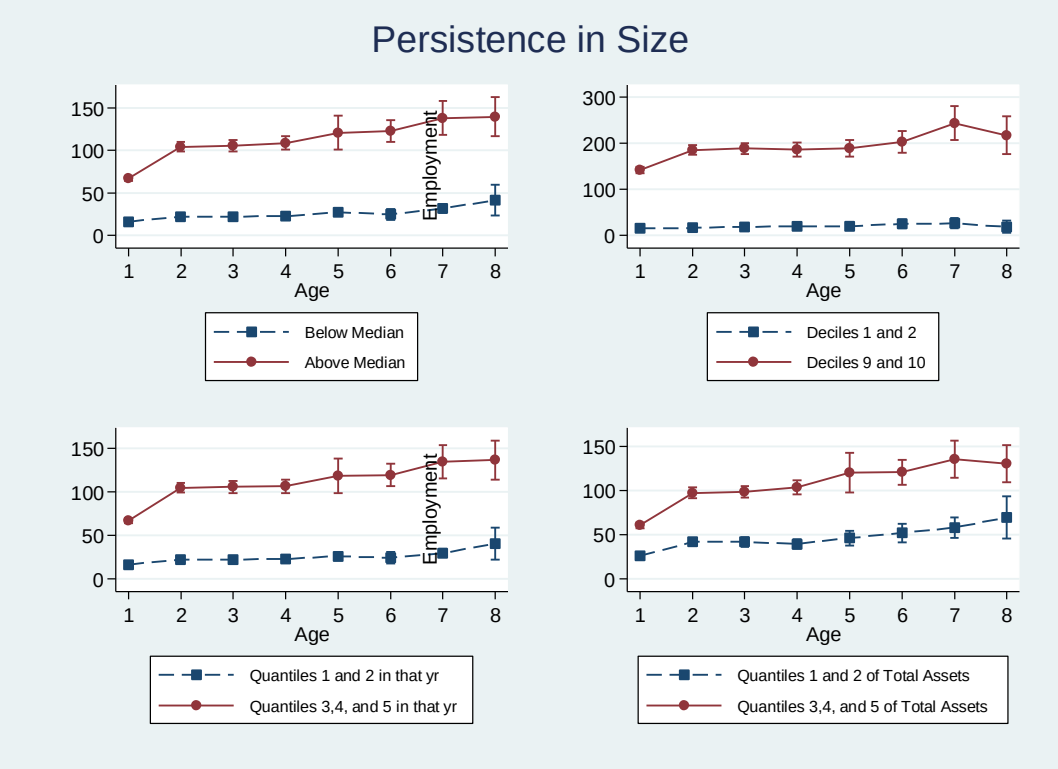


Figure A2: Employment Growth - Alternate Definitions of Large and Small Entrants

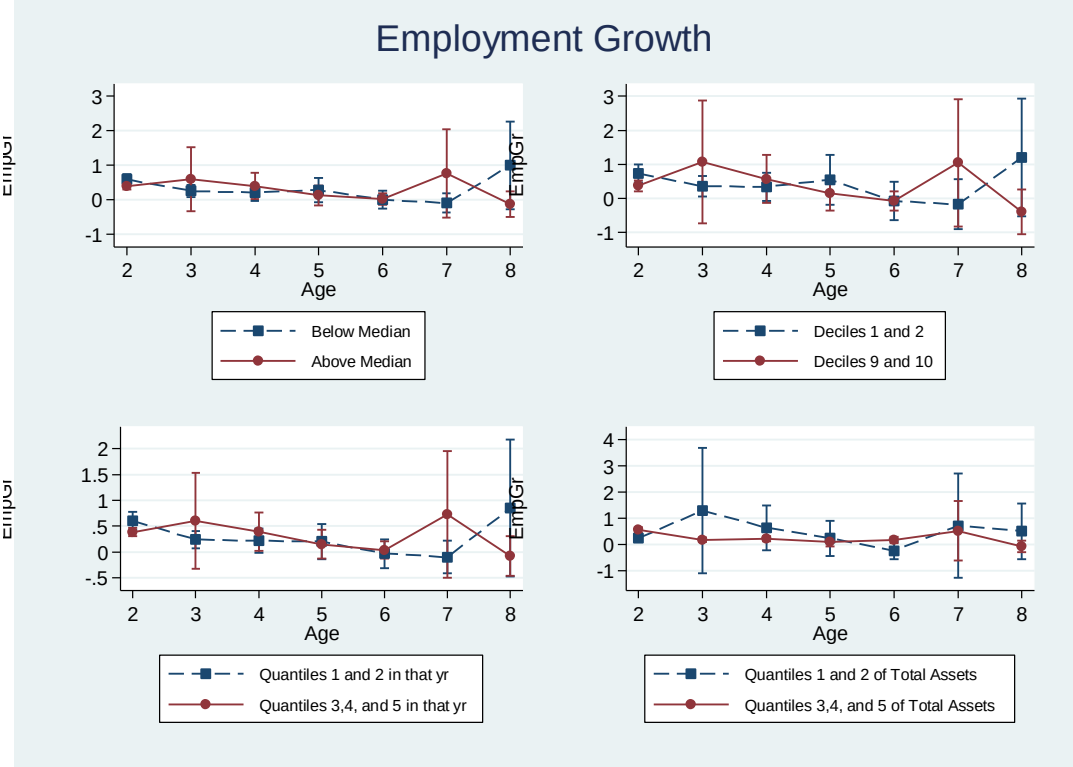


Figure A3: Size and Growth over Early Lifecycle: Large Vs. Small Entrants – Reweighting to take into account panel attrition

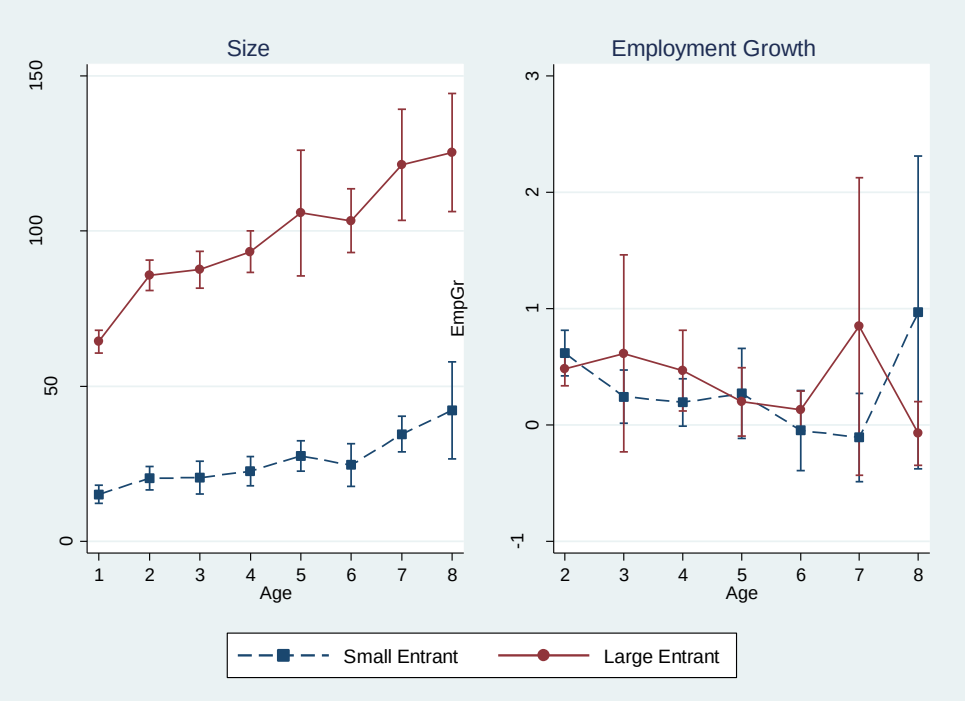
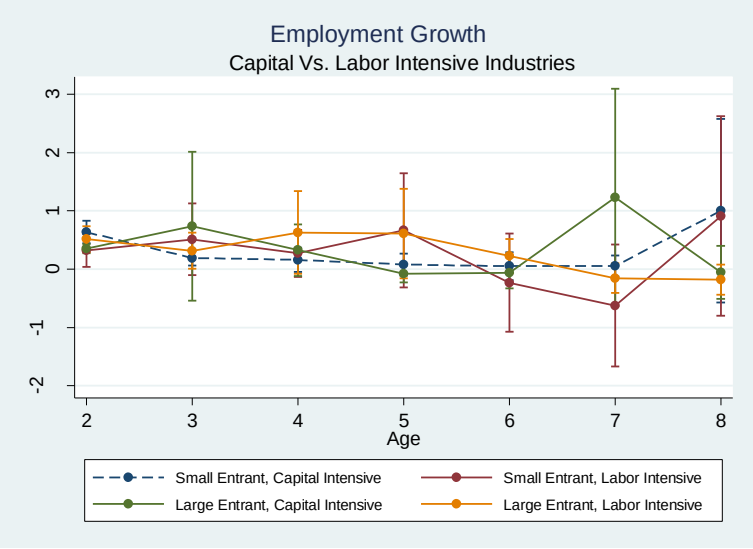
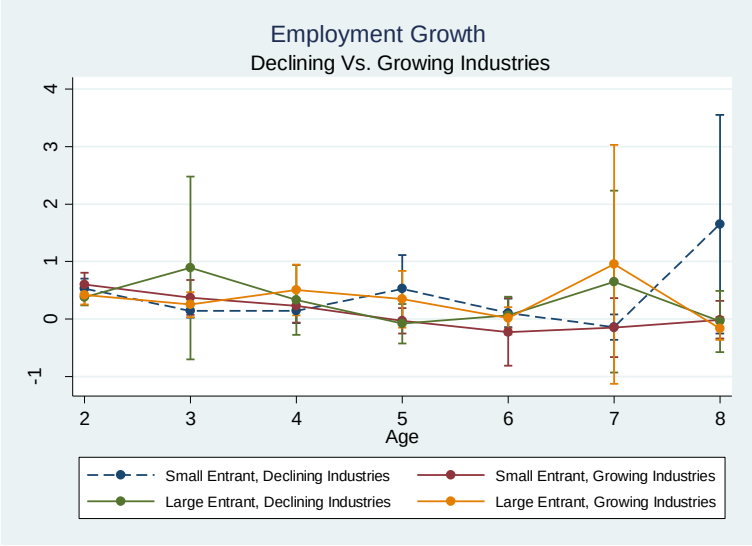


Figure A4: Employment Growth over Early Lifecycle – Large Entrant x Industry Classifications





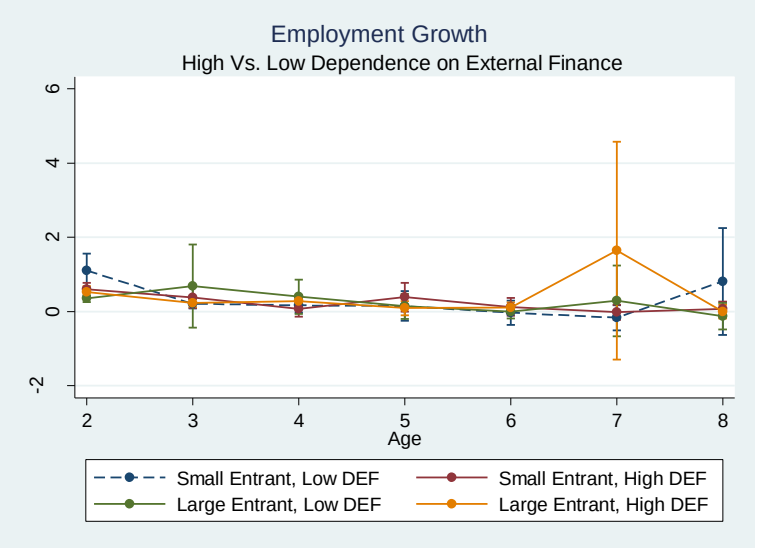


Figure A5: Is there a difference in growth rates across entrants in industries dependent on external finance (High DEF) in states with good financial institutions?