# Finance and Development: A Tale of Two Sectors

*By* Francisco J. Buera, Joseph P. Kaboski, and Yongseok Shin*

*We develop a quantitative framework to explain the relationship between aggregate/sector-level total factor productivity (TFP) and financial development across countries. Financial frictions distort the allocation of capital and entrepreneurial talent across production units, adversely affecting measured productivity. In our model, sectors with larger scales of operation (e.g., manufacturing) have more financing needs, and are hence disproportionately vulnerable to financial frictions. Our quantitative analysis shows that financial frictions account for a substantial part of the observed cross-country differences in output per worker, aggregate TFP, sector-level relative productivity, and capital to output ratios. (JEL O11, O16, O41)*

Income per capita differences across countries are primarily accounted for by low total factor productivity (TFP) in poor countries (Peter J. Klenow and Andrés Rodríguez-Clare 1997, Robert E. Hall and Charles I. Jones 1999). More disaggregate data on relative prices suggest that the TFP gap between rich and poor countries varies systematically across industrial sectors of the economy. For instance, less developed countries are particularly unproductive in producing manufactured goods, including equipment investment (Chang-Tai Hsieh and Peter Klenow 2007).

Given the stark differences in financial development across rich and poor countries, a large body of empirical and theoretical work has stressed the role of financial markets in economic development. Our goal in this paper is to capture and quantify this role.

Our main contribution is to present a rich quantitative framework and analyze the role of financial frictions in explaining a set of empirical regularities in economic development: poor countries' low per-capita income, low aggregate TFP, and large differences across industrial sectors in relative prices and implied sector-level productivity. Relative to the existing quantitatively-oriented work in the literature, our framework emphasizes two features: sector-specific non-convexities (fixed costs) in micro-level production technologies that cause differences in the scale of production across sectors; and the ability to overcome financial constraints with internal funds or self-financing through forward-looking savings behavior. These two elements not only enable a clean mapping between our model and data that disciplines our quantitative analysis, but also turn out to play critical quantitative roles.

We discover that financial frictions explain a substantial part of the above development regularities. Essentially, financial frictions distort the allocation of capital across heterogeneous production units and also their entry/exit decisions, lowering aggregate and sector-level TFP. While self-financing can alleviate the resulting misallocation, it is

* Buera: University of California–Los Angeles, Department of Economics, 8283 Bunche Hall, Los Angeles, CA 90095 (e-mail: fjbuera@econ.ucla.edu); Kaboski: University of Notre Dame, Department of Economics, 712 Flanner Hall, Notre Dame, IN 46556 (e-mail: jkaboski@nd.edu); Shin: Washington University in St. Louis, Department of Economics, Saint Louis, MO 63130, and Federal Reserve Bank of St. Louis (e-mail: yshin@wustl.edu). We thank three anonymous referees for helpful suggestions. We would also like to thank Philippe Aghion, Chang-Tai Hsieh, David Lagakos, Omar Licandro, Juan Pablo Nicolini, Roberto Rigobon, Richard Rogerson, and Hye Mi You for their comments on earlier versions of this paper. Buera and Shin gratefully acknowledge the financial support of the National Science Foundation (SES-0820318 and SES-0946647).

inherently more difficult to do so in sectors with larger scale and larger financing needs. Thus, sectors with larger scale (e.g., manufacturing) are affected disproportionately more by financial frictions.

We build a model with two sectors that differ in the per-period fixed costs of operating an establishment. This difference in fixed costs leads to a difference in the scale of establishments in the two sectors. Scale is the defining characteristic of a sector in our theory. To map the large-scale vs. small-scale sectors in our model into data, we use the manufacturing-services dichotomy. In the model, individuals choose in each period whether to operate an establishment in either sector (entrepreneurship) or to supply labor for a wage. They have different levels of entrepreneurial productivity and wealth. The former evolves stochastically and generates the need to reallocate capital and labor from previously-productive entrepreneurs to currently-productive ones. Financial frictions—which we model in the form of endogenous collateral constraints founded on imperfect enforceability of contracts—hinder this reallocation process.

We discipline our quantitative analysis by requiring that a benchmark model with well-functioning financial markets matches the US data on the establishment size distribution across and within sectors (e.g., averages and thick right tails) and the dynamics of establishments, among many others. We then quantify the relationship between financial development (measured by the ratio of external finance to GDP) and economic development.

We find that financial frictions have sizable effects on output per worker, aggregate and sector-level TFP, and capital to output ratios.

The variation in financial development can explain a factor-of-two difference in output per worker across economies, or almost 80 per cent of the difference in output per worker between Mexico and the US. One thing to note is that the agricultural sector is not modeled or analyzed here. This factor-of-two difference goes a long way in explaining the factor-of-five difference in non-agricultural output per worker between the richest and the poorest fifth percentiles of countries. Consistent with the consensus view in the literature, most of the output per worker differences in our model are accounted for by the low TFP in economies with underdeveloped financial markets. Our model predicts that the aggregate TFP of the country with the least financial development will be almost 40 per cent below the US level.

The impact of financial frictions is particularly large in the large-scale, manufacturing sector. While the sector-level TFP declines by less than 30 per cent in services, it declines by more than 50 per cent in manufacturing, a result broadly in line with the available sector-level productivity data from 18 OECD countries. The differential impacts of financial frictions on sector-level productivity are reflected on the higher relative prices of manufactured goods to services in financially underdeveloped economies. The model accounts for a quarter of the relationship between relative prices and financial development observed over a much larger set of countries.

Financial frictions also have a significant impact on the investment rate, when measured at common fixed prices across economies with different degrees of financial development. In our model simulations, the capital to output ratio (measured at common fixed prices) declines by 15 per cent with financial frictions. Consistent with the data, this decline is almost entirely driven by the higher relative prices of manufactured investment goods in financially underdeveloped economies (Stephen L. Parente and Edward C. Prescott 2000, Hsieh and Klenow 2007): The investment rates measured at respective equilibrium prices are roughly constant across economies with varying degrees of financial development.

Our quantitative analysis provides a clear decomposition of the main margins distorted by financial frictions. First, for a given set of heterogeneous production units in operation, financial frictions distort the allocation of capital among them (misallocation of capital).

Second, for a given number of production units in operation, financial frictions distort the selection into entrepreneurship, with productive-but-poor individuals delaying their entry and incompetent-but-rich entrepreneurs remaining in business (misallocation of talent). Third, financial frictions distort the number of production units for a given distribution of entrepreneurial talent in an economy. In our simulations, whereas the misallocation of capital is responsible for 90 per cent of the effect of financial frictions on the service sector TFP, it is the misallocation of talent that accounts for more than 50 per cent of the effect on the manufacturing sector TFP.

The differential impacts of financial frictions across sectors in our model produce an interesting testable implication on the establishment size distribution of each sector. Financial frictions, together with the resulting higher relative price of manufactured goods and lower capital rental rates and wages in the equilibrium, lead to too few entrepreneurs and too large establishments in manufacturing, and too many entrepreneurs and too small establishments in services. To evaluate this implication, we perform a detailed, disaggregate-level case study of Mexico and the US, and find empirical support for the model prediction.

Finally, we show that the two main elements of our model—sector-specific non-convexity and self-financing, which the conventional specifications in the literature lack—play important quantitative roles. Using a comparable one-sector model, we find that the effect of the non-convexities is itself convex: The aggregate impact of large fixed costs borne by one sector of the economy (manufacturing) is larger—by as much as one-third—than that of the small fixed costs spread over the whole economy. We then work out a comparable two-period overlapping-generation model, in which individuals only have very limited opportunities to overcome financial frictions over time through self-financing—because they live for only two periods. We find that constricting the self-financing channel overstates the aggregate impact of financial frictions by as much as 50 per cent.

## Empirical Underpinnings

Our theory is built on two premises: cross-country differences in financial development and cross-sector differences in the scale of establishments. Both of these underlying premises have strong empirical support.

The first premise, cross-country differences in financial development—underdevelopment in poor countries in particular, has been well established in the literature. Robert G King and Ross Levine (1993) and Thorsten Beck, Asli Demirgüç-Kunt and Ross Levine (2000) show that aggregate measures of credit and financial development are closely correlated with output per capita across countries, while Rafael La Porta, Florencio Lopez de Silanes, Andrei Shleifer and Robert W. Vishny (1998) document that these macro indicators are strongly related to underlying institutional differences such as contract enforcement and creditor protection. Abhijit V. Banerjee and Esther Duflo (2005) review the literature documenting micro-level evidence for credit constraints in poor countries and the resulting misallocation of capital. In his detailed analysis of Thailand, Robert M. Townsend (2010) links observed misallocation to micro-level credit constraints and shows how their relaxation through financial development leads to faster economic growth.

One empirical contribution of our paper is to establish the second premise: cross-sector differences in scale, defined as workers per establishment. Using detailed sector-level data from the OECD countries, we document that the average size of establishments varies substantially across broadly-defined sectors. For example, the average establishment in manufacturing is more than three times as large as that in services.

In addition, we carry out a detailed case study of Mexico and the US to evaluate our model prediction on the impact of financial frictions on the establishment size distribution. We use data from the economic censuses of the two countries (based on the common North

American Industrial Classification System) and a survey of small businesses in Mexico. We examine the data at a more disaggregate industry level than the manufacturing-services dichotomy. The average establishments in Mexico are substantially smaller, especially in the service industries. However, industries with large-scale establishments in the US (mostly in manufacturing) tend to have even larger establishments in Mexico.

One could define and explore sectoral heterogeneity along other characteristics than scale, but few are as easily-measured or robust as scale. We present two sets of empirical findings that further support our decision to focus on scale. Firstly, sectoral scale seems to be closely related to sector-level productivity: Using sector-level TFP data from a subset of OECD countries and price data from a broader cross-section of countries, we show that, even at a more disaggregate level of industry classification, less developed countries are particularly unproductive and have higher relative prices in industries with larger scales. Secondly, in our model, financial frictions have differential impacts on sectors with different scales because our notion of scale (establishment size) translates directly into financing needs. The most widely-used empirical metric of financing needs is "external dependence" of Raghuram G. Rajan and Luigi Zingales (1998). We compute the external dependence for broadly-defined sectors in the US, and find that sectors with larger scales indeed have larger external dependence.

## Contribution to the Literature

Our paper contributes to a vast theoretical literature relating financial frictions, entrepreneurship, and economic development. See Kiminori Matsuyama (2007) for an excellent, extensive survey.

There have been relatively fewer quantitatively-oriented studies. Xavier Giné and Robert M. Townsend (2004) and Hyeok Jeong and Robert M. Townsend (2007) have pioneered the quantitative analysis of the link between financial frictions and development. They estimate models of the theoretical literature using data on Thailand's growth experience.

Pedro S. Amaral and Erwan Quintin (2010) and Jeremy Greenwood, Juan M. Sanchez and Cheng Wang (2010) are recent contributions that quantify the long-run impact of financial frictions.[1] They study models that allow for sharp analytical characterizations of the equilibrium. In obtaining such tractability and lucidity, they downplay the role of entrepreneurs' self-financing. In our model, entrepreneurs can partly overcome financial constraints over time with self-financing, and the presence of such an avenue is shown to materially affect the quantitative results. More important, we complement their work by building a multi-sector model with non-convex technologies at the establishment level. These features not only generate richer disaggregate-level predictions that can be evaluated empirically, but also help better capture the aggregate impact of financial frictions.

This paper is closely related and complementary to three others in the literature that emphasize the differential effects of financial frictions on different industries. Rajan and Zingales (1998) create an index of dependence on external sources of financing for various manufacturing industries, and test whether industries that are particularly dependent on financing grow relatively faster in countries with more developed financial markets. We reconstruct their measure of industry-specific financial dependence for our analysis, and show that our measure of sectoral scale (workers per establishment) is closely re-

[1] Andrés Erosa (2001) is an earlier contribution that abstracts from microeconomic heterogeneity but focuses on the wedge between deposit and loan rates. Francesco Caselli and Nicola Gennaioli (2005) focus on low frequency dynastic shocks requiring the reallocation of capital from incompetent heirs to talented-but-poor entrepreneurs.

lated to external dependence both in our model and in the data.[2] Andrés Erosa and Ana Hidalgo Cabrillana (2008) show theoretically that financial frictions can have differential effects on the productivity of manufacturing industries with different fixed cost requirements. We study a broader set of sectors, and quantify the impact of financial frictions by introducing scale as an empirical measure related to fixed costs and financing needs. Rui Castro, Gian Luca Clementi and Glenn MacDonald (2009) start from the premise that different sectors (consumption vs. investment goods) are characterized by different volatility of underlying idiosyncratic productivity shocks. They show that in economies with less risk-sharing, sectors with more volatile shocks (investment goods) are particularly unproductive. Methodologically, they use a tractable model to obtain insightful analytical characterizations. We view our approach as complementary to theirs. We focus on sectoral scale differences and cross-country differences in external finance—readily-measured and robust features in the data. Our model, with its rich dynamics and heterogeneity, eludes analytical tractability, but facilitates a clear mapping into data and a direct interpretation of the quantitative results.

Finally, we add to the broader literature on the effect of micro-level distortions on aggregate productivity (Hugo A. Hopenhayn and Richard Rogerson 1993, Nezih Guner, Gustavo Ventura and Yi Xu 2008, Diego Restuccia and Richard Rogerson 2008). Our results complement the empirical findings of Chang-Tai Hsieh and Peter Klenow (2009) in particular, who find a factor-of-two difference in manufacturing TFP due to the misallocation of capital and labor. We explicitly model one source of such misallocation (financial frictions), and include the service sector in our analysis. Furthermore, we explore how distortions affect the entry and exit decisions of establishments.

## I. Facts

This section documents the key empirical facts motivating our study. First, we show that the price of manufactured goods is high relative to services in less developed countries, and that such relative prices are closely linked to sector-level relative productivity. Second, we point out that there are large differences in scale (establishment size) across sectors, with the manufacturing sector having much larger establishments on average than the service sector. Next, we emphasize this sectoral pattern in relative scale and relative productivity by showing that this relationship holds at a more disaggregate level as well: Goods and services that are produced with large-scale technologies tend to be relatively more expensive in less developed economies, implying their low relative productivity in these industries.

### *A. Relative Prices, Relative Productivity, and Economic Development*

In poor countries, the final prices of manufactured goods are high relative to services.[3] The left panel of Figure 1 shows this fact by plotting the relative price of manufactured goods to services from the 1996 ICP against PPP output per worker from Penn World Table 6.1.[4] The relative price of manufactured goods to services has a strong negative

[2]Thorsten Beck, Asli Demirgüç-Kunt, Luc Laeven and Ross Levine (2008) examine the independent effects of scale and financing needs.

[3]A closely related fact, first documented by Bela Balassa (1964) and Paul A. Samuelson (1964), is that the relative price of tradable goods is higher in poor countries.

[4]ICP stands for the International Comparison Programme of the United Nations. There are 115 ICP benchmark countries in 1996. To maintain a consistent sample, we show results using the 102 countries for which Beck, Demirgüç-Kunt and Levine (2000) constructed financial development data. Relative prices are compiled by creating Geary-Khamis aggregated prices for manufactured goods and services out of 27

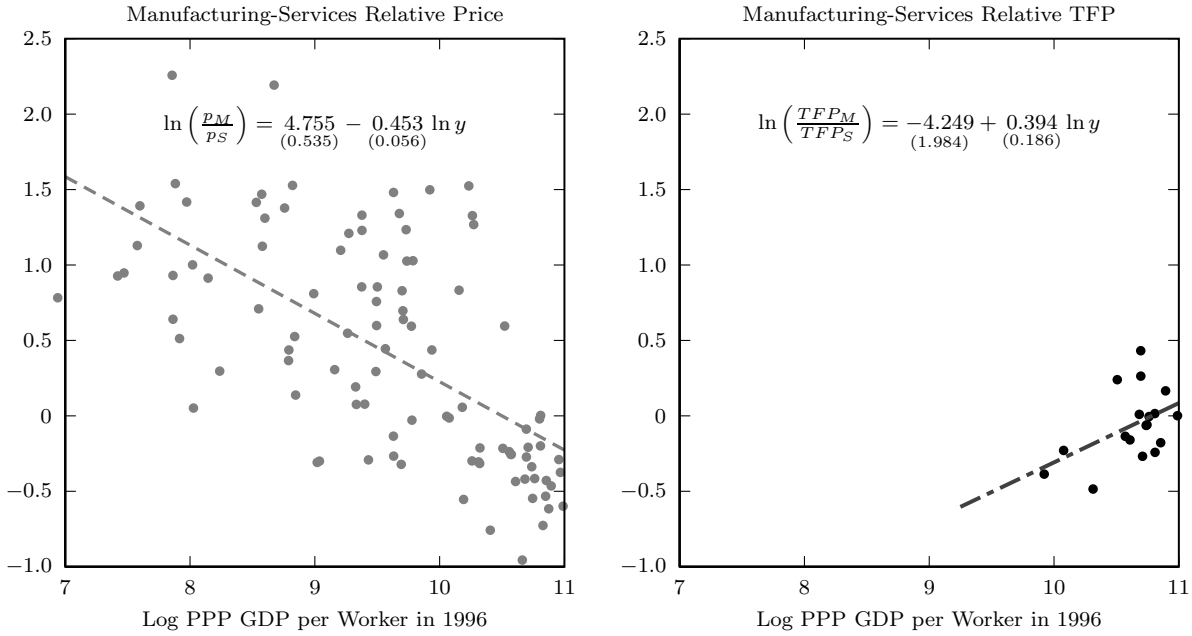


FIGURE 1. RELATIVE PRICES AND RELATIVE PRODUCTIVITY AGAINST OUTPUT PER WORKER

relationship with the output per worker across countries. The slope coefficient of -0.45 is highly significant, and the $R^2$ is 0.40.

Within many models, this relationship can be interpreted as lower relative TFP of manufacturing to services in less developed countries.[5] Indeed, in models with constant-returns-to-scale production functions and equal factor shares across sectors, relative prices equal the inverse of relative TFP (Hsieh and Klenow 2007). Differences in factor shares and the relative supply of factors could break this relationship, but empirically factor shares do not vary much across sectors (Ákos Valentinyi and Berthold Herrendorf 2008).

Furthermore, the available data on sector-level productivity across countries support the relative productivity interpretation of relative prices. Using the Productivity Level Database (Robert Inklaar and Marcel P. Timmer 2008) of the Groningen Growth and Development Centre (GGDC), we construct sector-level TFP measures for manufacturing and services value-added for 18 OECD countries.[6] The log relative sector-level TFP of manufacturing to services is plotted against log GDP per worker in the right panel of Figure 1. The regression coefficient is 0.39. Given the smaller sample size, the standard

disaggregate product categories. Goods categories are clothing, nine food/beverage categories, footwear, fuel, furniture/floor coverings, household appliances, household textiles, other household goods, machinery/equipment, tobacco, transportation equipment. Services categories are communication, education, medical/health, recreation/culture, rent, water, restaurants/hotels, transportation services.

[5] Alternative explanations include taxes, tariffs, and transportation costs that differ across sectors and countries. See Hsieh and Klenow (2007) for more discussions on these issues.

[6] The database contains data for 19 of the 30 OECD countries (Australia, Austria, Belgium, Czech Republic, Denmark, Spain, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Luxembourg, the Netherlands, Portugal, Sweden, the UK and the US) plus Slovenia. A double deflation method is used to construct real value-added by sector. While domestic intermediates are deflated by their source sector, imported intermediates are not. Imported intermediates are important for two small countries with very high trade to GDP ratios, Luxembourg (112 per cent) and Slovenia (107 per cent). These two are also outliers in terms of their sectoral TFP (in opposite directions), and are excluded from our analysis.

error is larger (0.19), but the estimate is still significant at the five-percent level, with an $R^2$ of 0.22. Thus, we find that the relative TFP of manufacturing to services is positively correlated with output per worker, just as relative prices are negatively correlated with output per worker, and that the magnitudes are similar (0.39 and -0.45, respectively). This is comforting in two ways. Firstly, it confirms the predicted relationship between relative prices and relative productivity, albeit for a limited set of fairly developed countries. Secondly, it gives evidence for value-added rather than final goods prices: We explore the patterns in relative productivity, relative prices, and output per worker with a model that applies more transparently to value-added.

Nevertheless, we acknowledge that no direct evidence on relative sectoral productivity is available for the vast majority of poor and less developed countries. The absence of such evidence is made conspicuous by the large tracts of emptiness in the right panel of Figure 1.

In this context, we note that Martin N. Baily and Robert M. Solow (2001) discuss evidence from a handful of case studies on selected manufacturing and service industries in Brazil, France, Germany, Japan, Korea, the Netherlands, the UK, and the US.[7] They find that the productivity of the service industries studied (airlines, modern retail, retail banking, and telecom) is substantially lower in the two developing countries (Brazil and Korea), while their low productivity in these services is commensurate with their low productivity in manufacturing.

However, it is important to understand that the case studies are focusing on modern service industries that look more like manufacturing in term of the average size of establishments or firms.[8] In our theory, establishment size, or scale, is the defining characteristic of a sector. Though we map our model into data using the manufacturing-services dichotomy—based on the observation that average manufacturing establishments (firms) are larger than those in services, with 47 (57) vs. 14 (17) employees—we also acknowledge substantial heterogeneity in scale across industries within these sectors.

The McKinsey Global Institute, the coordinator of the case studies, did not consider more conventional, small-scale service industries, except for "mom-and-pop" retail stores. The study on the productivity of the retail sectors in Korea and the US finds that, in the early 1990s, Korea was most unproductive in modern formats of large-scale retailing like specialty stores (e.g., Benetton, Gap) and department stores (e.g., Macy's, Saks), while in mom-and-pop retailing, the US stores barely had an edge (Baily and Solow 2001, p. 163).

In summary, the existing evidence congruently points to the linkage among sectoral scale, relative productivity, and economic development, which we discuss subsequently in detail.

### *B. Relative Prices and Financial Development*

A common measure of a country's level of financial development is its ratio of external finance to GDP, where external finance is defined as the sum of private credit, private bond market capitalization, and stock market capitalization.[9] As is well documented in the literature, this measure is very closely correlated with output per worker: The

[7]Of the eight, only Brazil is poorer than the least rich country in the GGDC database, Hungary.

[8]The number of employees per establishment (firm) is 100 (165) for air transportation, 38 (59) for supermarkets, 20 (226) for commercial banking, and 32 (279) for wired telecommunication carriers, according to the 2002 US Census. These industries are almost outliers in terms of scale among service industries.

[9]We use the data from La Porta et al. (1998), Rajan and Zingales (1998), and Beck, Demirgüç-Kunt and Levine (2000). The market value of equity overstates the book value which is conceptually closer to the financed capital in our model. We multiply the reported stock market capitalization by the average

regression of log GDP per worker on external finance to GDP yields a slope coefficient of 1.08, with a standard error of 0.13 and an $R^2$ of 0.48. Therefore it is unsurprising that the relationship between log relative prices and external finance to GDP ratios is similar to the one between log relative prices and log GDP per worker in the left panel of Figure 1. The estimated slope coefficient is -0.67, with a standard error of 0.10 and an $R^2$ of 0.38. For the 18 OECD countries with the sector-level TFP data, the regression coefficient of the log manufacturing-services relative TFP on external finance to GDP ratios is 0.08. While it has the same sign as the slope in the right panel of Figure 1, it is not significant with a standard error of 0.08.

The strength of the relationship suggests that financial development is potentially closely related to the patterns in relative prices, relative productivity, and output per worker across countries. In the model we develop, it is financial development that is the causal force behind these cross-country differences.

### *C. Scale Differences across Sectors*

Another pivotal empirical fact for our study is the clear differences in the scale of production units across broadly-defined sectors. One empirical contribution of our paper is to extend the work of Francisco J. Buera and Joseph P. Kaboski (2008) in establishing cross-sector differences in scale for a broad cross-section of countries. Our interpretation is that the observed sectoral scale differences reflect differences in micro-level production technologies across sectors. We will argue that these technological differences interact with financial development, so that financial development affects large-scale (e.g., manufacturing) and small-scale sectors (e.g., services) differently.

Here we use two measures of scale: workers per establishment and workers per enterprise. Establishments are locations of business, so that a single enterprise (i.e., firm) may have multiple establishments. Table 1 presents measures of average scale across broadly-defined final goods sectors in the US and other OECD countries, along with other sector-level characteristics like financial dependence and factor intensities.[10]

Table 1—Manufacturing vs. Services

| | Workers per Establishment | Workers per Enterprise | External Dependence | Capital Share |
|---|---|---|---|---|
| | US | OECD | US | US |
| Manufacturing | 47 | 28 | 0.21 | 0.31 |
| Services | 14 | 8 | 0.09 | 0.27 |

The first column is based on data from the 2002 US Economic Census, which uses an establishment basis and the NAICS classification. The second column is based on the OECD Structural Statistics for Industry and Services (SSIS) data for 2002. These data follow the common ISIC 3.2 four-digit classification, enabling comparison across countries. Enterprise-level data permit comparison over the largest set of countries. The OECD data cover nine countries: Czech Republic, France, Germany, Hungary, Poland, Portugal, Slovakia, the UK, and the US.

book-to-market ratio in the data (0.33). Our cross-country regressions are robust to this correction, because stock markets are important for only a few countries (notably the US) in the data. We also note that some private credit is used for consumer credit (e.g., credit cards). In the US flow of funds data, this amounts to 9 per cent of private credit during the 1990s. As we do not have data on consumer credit for most other countries—and this number is likely to be even smaller for them, we do not adjust for it.

[10] The sectors are constructed to be consistent with the final goods categories in the ICP data.

Whether establishments or enterprises are the unit of measurement, the average scale varies considerably across sectors. In the US census data, establishments in manufacturing are on average 3.4 times as large as those in services: 47 vs. 14. The relative scale of manufacturing to services is even larger with enterprises as the unit of observation. For the OECD average, the ratio is 3.5: 28 vs. 8. Establishments are our preferred unit of analysis because we think they embody production technologies, although we acknowledge that some technologies (e.g., the distribution system of Walmart) may be at the firm level. Data availability dictates which measure we use in certain cases.

In our model, financial frictions have differential impacts on sectors with different scales because our notion of scale (workers per establishment) closely correlates with financing needs. Support for this interpretation can be obtained by comparing more direct measures of financing needs across sectors. The third column reports the measures of external dependence (Rajan and Zingales 1998) that we construct using the US Compustat data for 1993–2003.[11] We find that firms producing final manufacturing goods are substantially more financially-dependent than those producing services, with a median of 0.21 vs. 0.09.[12] While external dependence may be a more direct measure of financing needs or investment requirements, we decide to focus on scale (employment), a measure that is available for all firms in the economy, as opposed to just the publicly-traded firms in Compustat. See Steven J. Davis, John C. Haltiwanger, Ron Jarmin and Javier Miranda (2006) for an example where conclusions drawn from the universe of firms and from the set of publicly-traded firms are very different.

Finally, the fourth column reports the sectoral differences in factor intensities. These numbers are from Valentinyi and Herrendorf (2008), and correspond to the capital share of gross output in each of these broadly-defined sectors, calculated using input-output data. Unlike differences in scale and external dependence, the variation in factor intensity is relatively small. We follow the recent literature (V. V. Chari, Patrick J. Kehoe and Ellen R. McGrattan 1997, Hsieh and Klenow 2007) and build a model that abstracts from this difference. Our focus is instead on the large observed differences in scale, and the corresponding financing needs, across sectors.[13]

### *D. Relative Prices, Productivity, and Scale*

We have presented evidence that the large-scale, manufacturing sector has higher relative prices and lower relative productivity in less developed economies. A natural question to ask is whether relative prices, relative productivity, and the scale of production technologies are related at a more disaggregate level as well.

We examine this issue using disaggregate ICP price data from its 1996 benchmark. The scale of an industry is constructed by averaging across nine countries for which

[11] Rajan and Zingales measure the ratio of the difference between capital expenditures and cash flow to capital expenditures. To negate the influence of outliers in noisy firm-level data, they take the total capital expenditures and total cash flow over the sample period to compute firm-specific numbers, and then pick the median of an industry as the industry-specific value. Note that they only study the manufacturing sector.

[12] Alternative measures of financing needs give a similar picture. We devised a measure of setup costs. For each firm in Compustat, we located the first period with positive excess investment, and computed the average excess investment over following consecutive periods with positive excess investment. In manufacturing, the ratio of this measure to annual sales has a median of 0.65. In services, the median is 0.12.

[13] The distinctions in Table 1 also hold at a more disaggregate level. Broadly-defined sub-sectors within manufacturing tend to have larger scales and higher dependence on external finance than services and construction, but all have comparable factor shares. See the earlier version of this paper (Francisco J. Buera, Joseph P. Kaboski and Yongseok Shin 2009).

comparable data (OECD SSIS) are available.[14] We then map ICP categories into closely related groups of industries and calculate the average scale for these industry groups. Finally, we run a cross-country regression of 2,794 disaggregated ICP price data from 112 countries on log output per worker, log industry scale, and their interaction. The estimation result is (with standard errors in parentheses):

$$\ln\left(\frac{p_{i,j}}{PPP_i}\right) = \underset{(0.56)}{-7.48} + \underset{(0.05)}{0.29}\ln y_i + \underset{(0.15)}{1.01}\ln\ \bar{l}_j - \underset{(0.02)}{0.10}\ln y_i \ln \bar{l}_j, \quad R^2 = 0.22,$$

where $p_{i,j}$ is the 1996 PPP price of industry $j$ output in country $i$, $PPP_i$ is the average (Geary-Khamis) 1996 PPP price level in country $i$, $y_i$ is the output per worker of country $i$ in 1996 international prices, and $\bar{l}_j$ is the average number of workers engaged per enterprise for industry $j$. The negative coefficient on the interaction term indicates that prices of the output of industries with larger scales are relatively higher in low income countries. Given the log difference between manufacturing and service scales in Table 1—i.e., $\ln(28/8) = 1.25$, the coefficient of -0.10 implies a relative price elasticity with respect to output per worker of 0.13, about one-third of the full relationship in Figure 1.

We work out a parallel exercise using the disaggregate industry-level TFP data (29 industries) for the 18 countries in the GGDC database. We obtain a consistent result:

$$\ln\left(\frac{TFP_{i,j}}{TFP_i}\right) = \underset{(0.28)}{12.18} - \underset{(0.32)}{1.10}\ln y_i - \underset{(1.00)}{3.52}\ln\ \bar{l}_j + \underset{(0.09)}{0.32}\ln y_i \ln \bar{l}_j, \quad R^2 = 0.08.$$

The magnitude of the interaction term is substantially larger. With the difference between manufacturing and service scales in Table 1, the coefficient of 0.32 implies a relative TFP elasticity with respect to output per worker of 0.40, almost the full estimated elasticity in the right panel of Figure 1.[15]

The general magnitude and significance of the above regression results are quite robust.[16] We conclude that, even at a more disaggregate level, less developed countries have relatively high prices and low productivity in industries with larger scales. This is further evidence supporting our emphasis on sectoral scale differences.

## II. Model

We model an economy with two sectors, $S$ (small scale, services) and $M$ (large scale, manufacturing). The output of the service sector is used for consumption only. The manufactured goods are used for consumption and investment, and are the numeraire.

[14] We use number of persons engaged per enterprise as our scale measure. For any given disaggregate-level industry, there is a strong correlation in scale across countries. We average across countries to smooth out idiosyncracies that may arise from local market structures, regulations, and so on.

[15] Note that the result from Baily and Solow's (2001) large-scale vs. small-scale retailing case study is consistent with this regression. A related conclusion emerges from Martin N. Baily and Eric Zitzewitz (2001), who study the productivity of service industries across the eight countries covered in Baily and Solow (2001). Additionally, Argentina, Colombia, Mexico, and Venezuela are in the retail banking study, and Argentina is in the telecom study. They find that the productivity in the developing countries are substantially lower in large-scale service industries (airlines, modern retailing, retail banking, and telecom).

[16] The significance of the results at the five-percent level is robust to clustering standard errors by country or by ICP category. Also, we can use $p_{i,j}$ and $TFP_{i,j}$ without scaling by $PPP_i$ and $TFP_i$ respectively. More generally, alternative specifications that use country-specific fixed effects instead of controlling for $\ln y_i$ or industry-specific fixed effects instead of controlling for $\ln \bar{l}_j$ yield very similar results. Finally, substituting countries' ratios of external finance to GDP for $\ln y_i$ produces again very similar coefficients.

There are measure $N$ of infinitely-lived individuals, who are heterogeneous in their wealth and the quality of their entrepreneurial ideas or talent, $\mathbf{z} = (z_S, z_M)$. Individuals' wealth is determined endogenously by forward-looking savings behavior. The vector of entrepreneurial ideas is drawn from a distribution $\mu(\mathbf{z})$. Entrepreneurial ideas "die" with a constant hazard rate of $1 - \gamma$, in which case a new vector of ideas is independently drawn from $\mu(\mathbf{z})$; that is, $\gamma$ controls the persistence of the entrepreneurial idea or talent process. The $\gamma$ shock can be interpreted as changes in market conditions that affect the profitability of individual skills.

In each period, individuals choose their occupation: whether to work for a wage or to operate a business in sector $S$ or $M$ (entrepreneurship). Their occupation choices are based on their comparative advantage as an entrepreneur ($\mathbf{z}$) and their access to capital. Access to capital is limited by their wealth through an endogenous collateral constraint, because capital rental contracts may not be perfectly enforceable in our model.

One entrepreneur can operate only one production unit (establishment) in a given period. Entrepreneurial ideas are inalienable, and there is no market for managers or entrepreneurial talent. The way we model an establishment draws upon the span of control of Robert E. Lucas, Jr. (1978) and per-period fixed costs as in Esteban Rossi-Hansberg and Mark L. J. Wright (2007).

### A. *Preferences*

Individual preferences are described by the following expected utility function over sequences of consumption $\mathbf{c}_t = (c_{S,t}, c_{M,t})$:

$$(1) \qquad U(\mathbf{c}) = \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(\mathbf{c}_t)\right], \quad u(\mathbf{c}_t) = \frac{1}{1-\sigma}\left(\psi c_{S,t}^{1-1/\varepsilon} + (1-\psi)c_{M,t}^{1-1/\varepsilon}\right)^{\frac{1-\sigma}{1-1/\varepsilon}},$$

where $\beta$ is the discount factor, $\sigma$ is the coefficient of relative risk aversion (and the reciprocal of the elasticity of intertemporal substitution), $\varepsilon$ is the intratemporal elasticity of substitution between services and manufactured goods, and $\psi$ controls the share of services in overall consumption expenditure. The expectation is over the realizations of entrepreneurial ideas ($\mathbf{z}$), which depend on the stochastic death of ideas ($1 - \gamma$) and on draws from $\mu(\mathbf{z})$.

### B. *Technology*

At the beginning of each period, an individual with vector of entrepreneurial ideas $\mathbf{z}$ and wealth $a$ chooses whether to work for a wage $w$ or operate a business in either sector $j = S, M$. To operate a business in a sector, individuals must pay a sector-specific per-period fixed cost of $\kappa_j$, in units of the sector's output. The crucial assumption is that the fixed cost to run an establishment in the manufacturing sector is higher than that in the service sector, $\kappa_M > \kappa_S$. This will generate the scale difference between the two sectors that we observe in the data (Table 1). Note that $\kappa_j$ is fixed costs that need to be paid in every period of operation. We discuss in Section III.C how the results will change if one were to introduce one-time fixed setup costs of starting a business.

After paying the fixed cost, an entrepreneur with talent $z_j$ produces using capital ($k$) and labor ($l$) according to:

$$z_j f(k, l) = z_j k^{\alpha} l^{\theta},$$

where $\alpha$ and $\theta$ are the elasticities of output with respect to capital and labor, and $\alpha+\theta < 1$, implying diminishing returns to scale in variable factors at the establishment level. Note that the factor elasticities are assumed to be the same in both sectors, consistent with the empirical findings in the literature (Chari, Kehoe and McGrattan 1997, Valentinyi and Herrendorf 2008).

Given factor prices $w$ and $R$ (rental rate of capital), the profit of an entrepreneur is:

$$\pi_j\left(k,l;R,w,\mathbf{p}\right)=p_j z_j k^{\alpha}l^{\theta}-Rk-wl-(1+r)p_j\kappa_j,$$

where $r$ is the interest rate and $p_j$ is the price of sector $j$ output. We normalize $p_M$ to one. For later use, we define the optimal level of capital and labor inputs when production is not subject to financial constraints:

$$\left(k_j^u(z_j),l_j^u(z_j)\right)=\arg\max_{k,l}\left\{p_j z_j k^{\alpha}l^{\theta}-Rk-wl\right\}.$$

The key feature of this technology is that the fixed costs introduce non-convexity. For any strictly positive fixed cost $\kappa_j$, the technology is feasible only if operated above the minimum scale; that is, $z_j k^{\alpha}l^{\theta}\geq(1+r)\kappa_j$.[17]

### C. *Credit and Rental Markets*

Individuals have access to competitive financial intermediaries, who receive deposits, rent capital $k$ at rate $R$ to entrepreneurs, and lend entrepreneurs the fixed cost $p_j\kappa_j$. We restrict the analysis to the case where both borrowing and capital rental are within a period—that is, individuals' financial wealth is non-negative ($a\geq 0$). The zero-profit condition of the intermediaries implies $R=r+\delta$, where $r$ is the deposit and lending rate and $\delta$ is the depreciation rate.

Borrowing and capital rental by entrepreneurs are limited by imperfect enforceability of contracts. In particular, we assume that, after production has taken place, entrepreneurs may renege on the contracts. In such cases, the entrepreneurs can keep fraction $1-\phi$ of the undepreciated capital and the revenue net of labor payments: $(1-\phi)\left[p_j z_j f\left(k,l\right)-wl+(1-\delta)\,k\right]$, $0\leq\phi\leq 1$. The only punishment is the garnishment of their financial assets deposited with the financial intermediary, $a$. In the following period, the entrepreneurs in default regain access to financial markets, and are not treated any differently despite the history of default.

Note that $\phi$ indexes the strength of an economy's legal institutions enforcing contractual obligations. This one-dimensional parameter captures the extent of frictions in the financial market owing to imperfect enforcement of credit and rental contracts. This parsimonious specification allows for a flexible modeling of limited commitment that spans economies with no credit ($\phi=0$) and those with perfect credit markets ($\phi=1$).

We consider equilibria where the borrowing and capital rental contracts are incentive-compatible and are hence fulfilled. In particular, we study equilibria where the rental of capital is quantity-restricted by an upper bound $\bar{k}^j\left(a,z_j;\phi\right)$, which is a sector-specific function of the individual state $(a,\mathbf{z})$. We choose the rental limits $\bar{k}^j\left(a,z_j;\phi\right)$ to be the

[17] The model can be extended to allow for a choice of technologies within each sector: a technology with a small fixed cost and low productivity, $A_{1,j}z_j k^{\alpha}l^{\theta}-\kappa_1$, and a technology with a large fixed cost and high productivity, $A_{2,j}z_j k^{\alpha}l^{\theta}-\kappa_2$, with $A_{2,j}>A_{1,j}$ for $j=S,M$ and $\kappa_2>\kappa_1$. In this extension, fixed costs are technology-specific but not sector-specific. One can think of our current setup as a case where the productivity gains from the $\kappa_2$-technology are significantly larger for manufacturing; that is, $A_{2,M}\gg A_{2,S}$.

largest limits that are consistent with entrepreneurs choosing to abide by their credit contracts. Without loss of generality, we assume $\bar{k}^j(a, z_j; \phi) \leq k_j^u(z_j)$, where $k_j^u$ is the profit-maximizing capital inputs in the unconstrained static problem in sector $j$.

The following proposition provides a simple characterization of the set of enforceable contracts and the rental limits $\bar{k}^j(a, z_j; \phi)$ for $j = S, M$.

PROPOSITION 1: *Capital rental $k$ in sector $j$ by an entrepreneur with wealth $a$ and talent $z_j$ is enforceable if and only if*

$$\max_l \{p_j z_j f(k,l) - wl\} - Rk - (1+r) p_j \kappa_j + (1+r) a \geq (1-\phi) \left[\max_l \{p_j z_j f(k,l) - wl\} + (1-\delta) k\right]. \tag{2}$$

*The upper bound on capital rental that is consistent with entrepreneurs choosing to abide by their contracts can be represented by a function $\bar{k}^j(a, z_j; \phi)$, which is increasing in $a$, $z_j$, $\phi$.*

Condition (2) states that an entrepreneur must end up with (weakly) more economic resources when he fulfills his credit and rental obligations (left-hand side) than when he defaults (right-hand side). This static condition is sufficient to characterize enforceable allocations because we assume that defaulting entrepreneurs regain full access to financial markets in the following period.

This proposition also provides a convenient way to operationalize the enforceability constraint into a simple rental limit $\bar{k}^j(a, z_j; \phi)$.[18] As long as the unconstrained level of capital rental is not enforceable, the rental limit $\bar{k}^j(a, z_j; \phi)$ is implicitly defined as the larger root of the equation given by the equality in condition (2). Rental limits increase with the wealth of entrepreneurs, because the punishment for defaulting (loss of collateral) is larger. Similarly, rental limits increase with the talent of an entrepreneur because defaulting entrepreneurs keep only a fraction $1 - \phi$ of the output. In the rest of the paper, we restrict individuals' capital inputs to be less than or equal to the rental limit $\bar{k}^j(a, z_j; \phi)$.

While the enforceability of contracts as measured by $\phi$ is not sector-specific, the equilibrium enforceable rental contracts, as captured by the rental limits $\bar{k}^j(a, z_j; \phi)$, do vary across sectors because of the differences in technology and output prices.

### *D. Recursive Representation of Individuals' Problem*

Individuals maximize (1) by choosing sequences of consumption, financial wealth, occupations, and capital/labor inputs if they choose to be entrepreneurs, subject to a sequence of period budget constraints and rental limits.

At the beginning of a period, an individual's state is summarized by his wealth $a$ and vector of talent $\mathbf{z}$. He then chooses whether to be a worker or to be an entrepreneur in sector $S$ or $M$ for the period. The value for him at this stage, $v(a, \mathbf{z})$, is the maximum over the value of being a worker, $v^W(a, \mathbf{z})$, and the value of being an entrepreneur in sector $j$, $v^j(a, \mathbf{z})$, for $j = S, M$:

$$v(a, \mathbf{z}) = \max\left\{v^W(a, \mathbf{z}), v^S(a, \mathbf{z}), v^M(a, \mathbf{z})\right\}. \tag{3}$$

[18] The set of enforceable capital rental dictated by (2) may not coincide with $k \leq \bar{k}^j(a, z_j; \phi)$, if, for example, $p_j \kappa_j > a$. Nevertheless, the solution to the individual problem subject to (2) coincides with the solution to the individual problem subject to the simpler limit $k \leq \bar{k}^j(a, z_j; \phi)$. See the proof in the appendix.

Note that the value of being a worker, $v^W(a, \mathbf{z})$, depends on his assets $a$ and on his entrepreneurial ideas $\mathbf{z}$, which may be implemented at a later date. Similarly, the value of being an entrepreneur in sector $j$, $v^j(a, \mathbf{z})$, depends on the entire vector of entrepreneurial ideas, as he may switch sectors at a later date. We denote the optimal occupation choice by $o(a, \mathbf{z}) \in \{W, S, M\}$.

As a worker, an individual chooses a consumption bundle $\mathbf{c} = (c_S, c_M)$ and the next period's assets $a'$ to maximize his continuation value subject to the period budget constraint:

$$v^W(a, \mathbf{z}) = \max_{\mathbf{c}, a' \geq 0} u(\mathbf{c}) + \beta \{\gamma v(a', \mathbf{z}) + (1-\gamma) \mathbb{E}_{\mathbf{z}'}[v(a', \mathbf{z}')]\} \tag{4}$$
$$\text{s.t. } \mathbf{p} \cdot \mathbf{c} + a' \leq w + (1+r)a,$$

where $w$ is his labor income, and $\mathbf{p}$ denotes the vector of goods prices. The continuation value is a function of the end-of-period state $(a', \mathbf{z}')$, where $\mathbf{z}' = \mathbf{z}$ with probability $\gamma$ and $\mathbf{z}' \sim \mu(\mathbf{z}')$ with probability $1-\gamma$. In the next period, he will face an occupational choice again, and the function $v$ appears in the continuation value.

Alternatively, individuals can choose to become an entrepreneur in sector $j$. The value function of being an entrepreneur in sector $j$ is as follows.

$$v^j(a, \mathbf{z}) = \max_{\mathbf{c}, a', k, l \geq 0} u(\mathbf{c}) + \beta \{\gamma v(a', \mathbf{z}) + (1-\gamma) \mathbb{E}_{\mathbf{z}'}[v(a', \mathbf{z}')]\} \tag{5}$$
$$\text{s.t. } \mathbf{p} \cdot \mathbf{c} + a' \leq p_j z_j f(k, l) - Rk - wl - (1+r) p_j \kappa_j + (1+r)a$$
$$k \leq \bar{k}^j(a, z_j; \phi)$$

Note that an entrepreneur's income is given by period profit $p_j z_j f(k, l) - Rk - wl$ net of fixed costs $(1+r)p_j\kappa_j$ plus the return to his initial wealth, and that his choices of capital inputs are constrained by the rental limit $\bar{k}^j(a, z_j; \phi)$.

### *E. Stationary Competitive Equilibrium*

A stationary competitive equilibrium is composed of: an invariant distribution of wealth and entrepreneurial ideas $G(a, \mathbf{z})$, with the marginal distribution of $\mathbf{z}$ denoted with $\mu(\mathbf{z})$; policy functions $c_S(a, \mathbf{z})$, $c_M(a, \mathbf{z})$, $a'(a, \mathbf{z})$, $o(a, \mathbf{z})$, $l(a, \mathbf{z})$, $k(a, \mathbf{z})$; rental limits $\bar{k}^j(a, z_j; \phi)$, $j = S, M$; and prices $w$, $R$, $r$, $\mathbf{p}$ such that:

1) Given $\bar{k}^j(a, z_j; \phi)$, $w$, $R$, $r$ and $\mathbf{p}$, the individual policy functions $c_S(a, \mathbf{z})$, $c_M(a, \mathbf{z})$, $a'(a, \mathbf{z})$, $o(a, \mathbf{z})$, $l(a, \mathbf{z})$, $k(a, \mathbf{z})$ solve (3), (4) and (5);

2) Financial intermediaries make zero profit: $R = r + \delta$;

3) Rental limits $\bar{k}^j(a, z_j; \phi)$ are the most generous limits satisfying condition (2), with $\bar{k}^j(a, z_j; \phi) \leq k_j^u(z_j)$;

4) Capital rental, labor, services, and manufactured goods markets clear:

$$\text{(Capital rental)} \qquad \frac{K}{N} \equiv \int k(a,\mathbf{z})\, G(da, d\mathbf{z}) = \int a G(da, d\mathbf{z})$$

$$\text{(Labor)} \qquad \int l(a,\mathbf{z})\, G(da, d\mathbf{z}) = \int\limits_{\{o(a,\mathbf{z})=W\}} G(da, d\mathbf{z})$$

$$\text{(Services)} \quad \int c_S(a,\mathbf{z})\, G(da, d\mathbf{z}) = \int\limits_{\{o(a,\mathbf{z})=S\}} \left[ z_S k(a,\mathbf{z})^{\alpha} l(a,\mathbf{z})^{\theta} - \kappa_S \right] G(da, d\mathbf{z})$$

$$\text{(Manufactured goods)} \quad \int c_M(a,\mathbf{z})\, G(da, d\mathbf{z}) + \delta \frac{K}{N} = \int\limits_{\{o(a,\mathbf{z})=M\}} \left[ z_M k(a,\mathbf{z})^{\alpha} l(a,\mathbf{z})^{\theta} - \kappa_M \right] G(da, d\mathbf{z})$$

5) The joint distribution of wealth and entrepreneurial ideas is a fixed point of the equilibrium mapping:

$$G(a,\mathbf{z}) = \gamma \int\limits_{\{(\tilde{a},\tilde{\mathbf{z}})|\tilde{\mathbf{z}}\leq\mathbf{z}, a'(\tilde{a},\tilde{\mathbf{z}})\leq a\}} G(d\tilde{a}, d\tilde{\mathbf{z}}) + (1-\gamma)\,\mu(\mathbf{z}) \int\limits_{\{(\tilde{a},\tilde{\mathbf{z}})|a'(\tilde{a},\tilde{\mathbf{z}})\leq a\}} G(d\tilde{a}, d\tilde{\mathbf{z}}).$$

### *F. Perfect-Credit Benchmark*

To clarify the basic mechanics of the model, we analyze the perfect credit benchmark, $\phi = 1$. This is an economy with unconstrained within-period borrowing and capital rental for production—that is, $\tilde{k}^j(a, z_j, \phi) = k_j^u(z_j)$ for all $a$—but without between-periods borrowing or consumption insurance. We present two results characterizing the production side of the perfect-credit economy under the assumption that entrepreneurial talents for the two sectors follow mutually independent Pareto distributions with the same tail parameter $\eta$, $(z_S, z_M) \sim \eta^2 (z_S z_M)^{-(\eta+1)}$ for $z_j \geq 1$, $j = S, M$. This assumption permits approximate closed-form expressions for net sectoral production functions (sectoral output net of fixed costs), factor shares, and the establishment size distribution. In addition, it implies that the establishment size distribution within each sector exhibits a thick right tail, a salient feature of the data.[19] These characterizations will help us pin down the technological parameters of the model using the US data on establishment size distributions across and within sectors.

The first result is that the net output of a sector is given by a Cobb-Douglas, constant-returns-to-scale function of population size ($N$), and sectoral capital ($K_j$) and labor inputs ($L_j$).

PROPOSITION 2: *Assume that entrepreneurial talents for the two sectors follow mutually-independent Pareto distributions with the same tail parameter $\eta$, $(z_S, z_M) \sim \eta^2 (z_S z_M)^{-(\eta+1)}$ for $z_j \geq 1$, $j = S, M$, and that active entrepreneurs are a small fraction of the population.*

[19] We cannot deviate too strongly from the assumption of mutual independence, or the tail in the service sector will become unrealistically thinner than the tail in the manufacturing sector. In the extreme, if entrepreneurial ideas were perfectly correlated across sectors, with $\phi = 1$, even the smallest manufacturing establishment will have more employees than the largest service establishment.

*Then the output of a sector, net of fixed costs, equals:*

$$Y_j\left(K_j, L_j; N\right) = A_j N^{\frac{1/\eta}{\alpha+\theta+1/\eta}} K_j^{\frac{\alpha}{\alpha+\theta+1/\eta}} L_j^{\frac{\theta}{\alpha+\theta+1/\eta}},$$

$$A_j = \frac{w + p_j\kappa_j\left(\alpha+\theta+\eta^{-1}\right)}{w+p_j\kappa_j}\left[\frac{\eta\left(1-\alpha-\theta\right)}{\eta\left(1-\alpha-\theta\right)-1}\right]^{\frac{1}{1+\eta(\alpha+\theta)}}\left[\frac{w+p_j\kappa_j}{p_j\left(1-\alpha-\theta\right)}\right]^{\frac{1-\eta(1-\alpha-\theta)}{1+\eta(\alpha+\theta)}}.$$

In our calibration of the perfect-credit benchmark, active entrepreneurs are indeed a small fraction of the population (about five per cent). It follows that, as in the standard neoclassical sectoral growth model, the elasticities of output with respect to capital and labor are constant, $\frac{\alpha}{\alpha+\theta+1/\eta}$ and $\frac{\theta}{\alpha+\theta+1/\eta}$, respectively. Unlike in the standard model, however, the elasticities are not equal to the factor shares, because entrepreneurs earn rents. In particular, payments to capital as a share of income equals:

$$s_{K,j} = \frac{RK_j}{Y_j(K_j, L_j; N)} = \frac{\alpha(w+p_j\kappa_j)}{w+p_j\kappa_j\left(\alpha+\theta+\eta^{-1}\right)}.$$

For realistic parameterizations of the model, $\alpha+\theta+1/\eta$ is close to one, and hence factor shares in the two sectors are approximately equal.[20]

Our second result pertains to the establishment size distribution in the perfect-credit benchmark. In particular, we show that the establishment size in each sector follows a Pareto distribution with tail coefficient $\eta\left(1-\alpha-\theta\right)$, and that the overall establishment size distribution in the economy is a mixture of Pareto distributions. We also show that there is a direct mapping between the ratio of fixed costs to wage ($p_j\kappa_j/w$) and the ratio between the average establishment sizes ($\bar{l}_j$) of the two sectors.

PROPOSITION 3: *Assume that entrepreneurial talents for the two sectors follow mutually-independent Pareto distributions with the same tail parameter $\eta$, $(z_S, z_M) \sim \eta^2\left(z_S z_M\right)^{-(\eta+1)}$ for $z_j \geq 1$, $j = S, M$, and that active entrepreneurs are a small fraction of the population. Then the establishment size distribution in each sector follows the power law:*

$$\Pr\left[\tilde{l}_j > l\right] = \left(\frac{l\left(\hat{z}_j\right)}{l}\right)^{\eta(1-\alpha-\theta)}, \quad l \geq l\left(\hat{z}_j\right),$$

*where $l\left(\hat{z}_j\right)$ is the employment in the marginal establishment of sector $j$. Furthermore, the establishment size distribution in the aggregate economy is given by a mixture of Pareto distributions:*

$$\Pr\left[\tilde{l} > l\right] = n_S\left(\frac{l\left(\hat{z}_S\right)}{l}\right)^{\eta(1-\alpha-\theta)} + n_M\left(\frac{l\left(\hat{z}_M\right)}{\max\left\{l, l\left(\hat{z}_M\right)\right\}}\right)^{\eta(1-\alpha-\theta)}, \quad l \geq l\left(\hat{z}_M\right),$$

*where $n_S$ and $n_M$ are respectively the fraction of service and manufacturing establishments in the economy, with $n_S + n_M = 1$. Also, the ratio of the average establishment sizes of the two sectors is:*

$$\frac{\bar{l}_j}{\bar{l}_{j'}} = \frac{p_j\kappa_j + w}{p_{j'}\kappa_{j'} + w}.$$

[20] One technical condition is $1 > \alpha+\theta+1/\eta$. See the proof in the appendix.

This last result suggests a simple way of identifying the relative magnitude of sector-specific fixed costs from their relative scale. In our model, the large scale of a sector arises from the large fixed costs at the establishment level.[21] In addition, the tail of the establishment size distribution identifies the parameter governing the distribution of entrepreneurial talents. These observations will enable us to calibrate the model in a transparent manner.

## III. Quantitative Analysis

In this section, we first calibrate the perfect-credit benchmark of our model economy to the US economy. We then conduct experiments to assess the effect of financial frictions. In particular, we vary $\phi$, the parameter governing the degree of financial frictions, to generate variations in external finance to GDP ratios that are comparable to the range observed in a cross-section of countries. We evaluate our model predictions for aggregate/sector-level TFP, output per worker, and capital to output ratios.

In our quantitative analysis we hold fixed all technological parameters across countries, and vary only the parameter governing financial frictions ($\phi$). In particular, we assume that countries are endowed with the same entrepreneurial talent distribution. We maintain this assumption because our goal is to isolate and quantify the direct impact of financial frictions. One of our main results is that, starting with the same potential pool of entrepreneurs, financial frictions distort the selection into entrepreneurship. The productivity distribution of entrepreneurs in operation therefore differs across countries, with financial frictions lowering the mean and raising the dispersion of this distribution. The effect on the mean conforms to the conventional wisdom of aggregate TFP differences across countries. The increase in dispersion is consistent with the empirical findings of Hsieh and Klenow (2009), who show that less developed countries' establishment-level productivity (TFPQ in their terminology) dispersion is larger than that of the US. It would be straightforward to incorporate cross-country differences in the average productivity of potential entrepreneurs and workers by considering human capital and exogenous TFP differences. It is less obvious how one would discipline exogenous cross-country differences in higher moments of the talent distribution.

### *A. Calibration*

We calibrate preference and technology parameters so that the perfect-credit economy matches key aspects of the US, a relatively undistorted economy. Our target moments pertain to standard macroeconomic aggregates, the establishment size distribution within and across sectors, and establishment dynamics, among others.

We need to specify values for eleven parameters: four technological parameters, $\alpha$, $\theta$, $\kappa_S$, $\kappa_M$, and the depreciation rate $\delta$; two parameters describing the process for entrepreneurial talent, $\gamma$ and $\eta$; the subjective discount factor $\beta$, the coefficient of relative risk aversion $\sigma$, the intratemporal elasticity of substitution $\varepsilon$, and the service share in consumption $\psi$.

Two preference parameters, $\sigma$ and $\varepsilon$, and two technological parameters, $\alpha/(1/\eta+\alpha+\theta)$ and $\delta$, can be set to standard values in the literature. We let $\sigma = 1.5$ and $\varepsilon = 1.0$.[22] The one-year depreciation rate is set at $\delta = 0.06$, and we choose $\alpha/(1/\eta + \alpha + \theta)$ to match the aggregate capital income share of 0.30.[23]

We are thus left with the seven parameters that are more specific to our study. We calibrate them to match seven relevant moments in the US data as shown in Table 2: the

[21] It is straightforward to show that the model analog of Rajan and Zingales's (1998) financial dependence measure increases with the fixed cost $\kappa_j$ in the perfect credit benchmark. We average investment

Table 2—Calibration

| Target Moments | US Data | Model | Parameter |
|---|---|---|---|
| Top 10-percentile employment share | 0.69 | 0.69 | $\eta = 4.84$ |
| Top 5-percentile earnings share | 0.30 | 0.30 | $\alpha + \theta = 0.79$ |
| Average scale in services | 14 | 14 | $\kappa_S = 0.00$ |
| Average scale in manufacturing | 47 | 47 | $\kappa_M = 4.68$ |
| Establishment exit rate | 0.10 | 0.10 | $\gamma = 0.89$ |
| Manufacturing share of GDP | 0.25 | 0.25 | $\psi = 0.91$ |
| Interest rate | 0.04 | 0.04 | $\beta = 0.92$ |

average size of establishments in services and in manufacturing; the employment share of the top decile of establishments; the share of earnings generated by the top five per cent of earners; the annual exit rate of establishments; the share of manufacturing value-added in the absorbed GDP; and the annual real interest rate.

The identification of these seven parameters follows the basic logic given in our discussion of the perfect-credit benchmark. We calibrate the sector-specific fixed costs, $\kappa_S = 0.0$ and $\kappa_M = 4.68$, to match the average establishment size in services and manufacturing (14 and 47, respectively). The per-period fixed cost in manufacturing sector, $\kappa_M = 4.68$, is tantamount to about three times the equilibrium wage in the perfect-credit benchmark. Given the returns to scale, $\alpha + \theta$, we choose the tail parameter of the entrepreneurial talent distribution, $\eta = 4.84$, to match the employment share of the largest ten percent of establishments, 0.69. We can then infer $\alpha + \theta = 0.79$ from the earnings share of the top five percent of earners. Top earners are mostly entrepreneurs (both in the US data and in the model), and $\alpha + \theta$ controls the fraction of output going to the entrepreneurial input. The parameter $\gamma = 0.89$ leads to an annual establishment exit rate of ten per cent in the model. This is consistent with the exit rate of establishments reported in the US Census Business Dynamics Statistics.[24] We set $\psi = 0.91$ to match the share of manufacturing value-added in absorbed GDP. Note that all investment goods are manufactured goods in our model. Finally, the model requires a discount factor of $\beta = 0.92$ to match the annual interest rate of four per cent.

As we start out by assuming that the US is the economy with the best financial market possible in the model ($\phi = 1$), the external finance to GDP ratio is not our target moment. This ratio is 2.3 in our calibrated perfect-credit equilibrium, while it is 2.5 in the US data.

Figure 2 shows the establishment size distribution from the calibrated perfect-credit benchmark, and compares it with the US data. The horizontal axis is the establishment size (number of employees, $l$) in log. For each $l$, we compute the fraction of establishments whose size is larger than or equal to $l$. With our independent Pareto distribution for talent in each sector, the perfect-credit benchmark gives straight lines for services (dashed line) and manufacturing (solid line). We construct a line using all establishments in our perfect-credit benchmark (dotted line). We do the same calculation using the 2002 US Economic Census: asterisks for manufacturing and triangles for services. The model is able to fit the tails of the empirical distribution, the distance between the two within-sector distributions, and the initial concavity in the overall (inclusive of both sectors)

inclusive of fixed costs across entrepreneurs to construct our analog.

[22] See Section III.C for more discussions on the choice of $\varepsilon$.

[23] We are being conservative in choosing a relatively low capital share: The larger the share of capital, the bigger the role of capital misallocation. We are also accommodating the fact that some of the payments to capital in the data are actually payments to entrepreneurial input.

[24] Note that $1-\gamma$ is larger than 0.1, because a fraction of those hit by the idea shock chooses to remain in business. Entrepreneurs exit only if their new idea is below the equilibrium cutoff level in either sector.

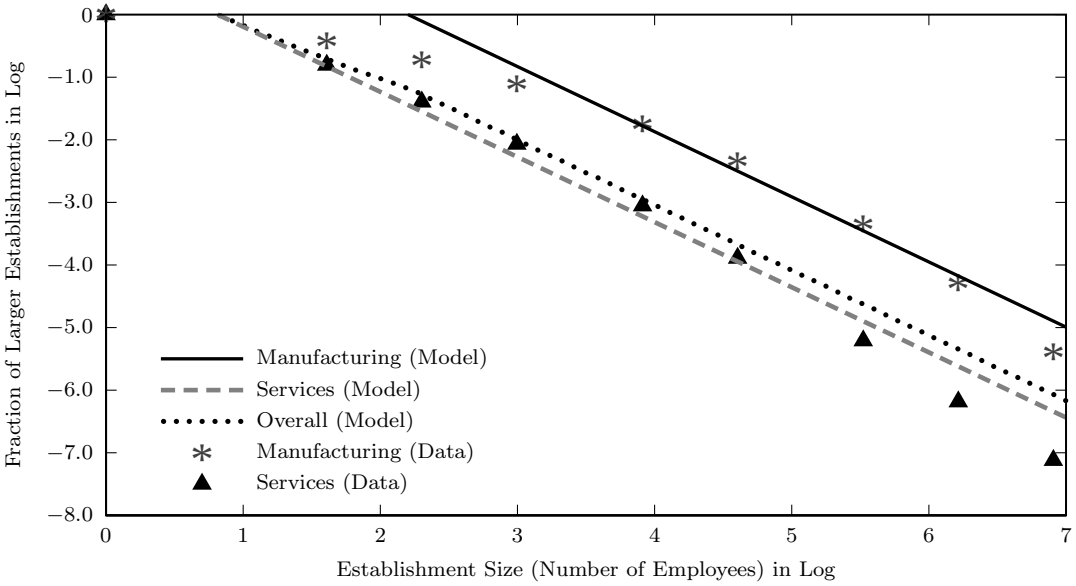


FIGURE 2. ESTABLISHMENT SIZE DISTRIBUTION IN THE MODEL AND DATA

distribution of establishment size. The assumption that the entrepreneurial talents for the two sectors are drawn from the same Pareto distribution generates the identical slope for the right tails. The model cannot capture the initial concavity in the distribution of establishment size within a sector, presumably because we abstract from within-sector heterogeneity in fixed costs.

### *B. Results*

We now quantify the effect of financial frictions on economic development. We first show that financial frictions have a substantial adverse impact on output per worker. In our exercises, the low per-capita income in economies with financial frictions is primarily explained by their low aggregate TFP, with particularly low productivity in manufacturing, consistent with the empirical findings in Section I.

We vary $\phi$—the parameter governing the enforcement of contracts in condition (2)—to span a range of external finance to GDP ratios observed in the data. With quintiles of countries constructed in terms of GDP per worker at PPP, external finance to GDP ratios average 0.1 for the bottom quintile and 2.1 for the top quintile. We use 13 values of $\phi$ ranging from 0 (financial autarky) to 1 (perfect credit), which span external finance to GDP ratios from 0 to 2.3. The parameter $\phi$ itself has no immediate empirical counterpart. Hence we plot our model simulation results against the (endogenous) ratio of external finance to GDP implied by a given $\phi$. The equilibrium external finance to GDP ratio is monotonically increasing in $\phi$, with a lower $\phi$ corresponding to more financial frictions.

The model quantities (e.g., output and TFP) are computed using a common fixed price for output, unless noted otherwise. In particular, we apply the output prices in the perfect-credit benchmark, with manufactured goods as numeraire. Similarly, the quantities in the data are measured at PPP.

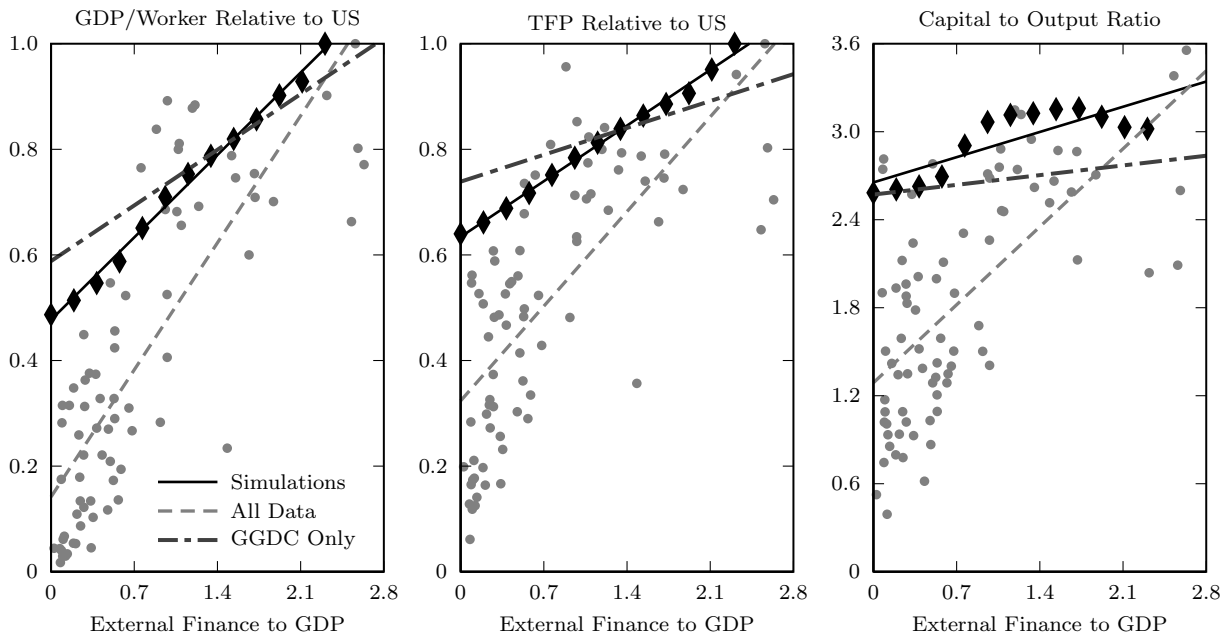


Figure 3. Aggregate Impact of Financial Frictions

## Aggregate Impact of Financial Frictions

Figure 3 plots the effect of financial frictions on output per worker at PPP, aggregate TFP, and capital to output ratios at PPP. The diamonds are from model simulations, and the gray dots represent country-level data.[25] The solid lines are regression lines for the model simulations, and the dashed lines are for the country data. For GDP and TFP, the model quantities are normalized by the perfect-credit level, and the data by the US level.

In our model, the variation in financial frictions can bring down output per worker to less than half of the perfect-credit benchmark level. This is tantamount to the output per worker difference between, say, Malaysia and the US, or about 80 per cent of the US-Mexico difference. While this does not come close to the difference between the US and the poorest countries in sub-Saharan Africa, the magnitude is nevertheless sizable, considering that we are varying one single factor—financial markets—across countries. The regression coefficient of output per worker on external finance to GDP ratios is 0.22 for model simulations (solid line, left panel), and 0.34 for the data (dashed line).[26] Comparing the two coefficients, one may conclude that our model explains as causal two thirds of the cross-country relationship between output per worker and financial development.

As in the data, the per-capita income differences in our model are primarily accounted for by differences in TFP (center panel). Financial frictions can reduce aggregate TFP by 36 per cent in our model. The regression coefficient of aggregate TFP on external finance

[25]The data are from the PWT 6.1. We use a perpetual inventory method and a depreciation rate of six per cent to construct the capital stock, using only the 79 benchmark countries with investment series starting in 1980 or earlier. TFP in 1996 is $YK^{-1/3}L^{-2/3}$, where $Y$ is PPP GDP and $L$ is the number of workers.

[26]The univariate regression of output per worker on external finance in the data has an $R^2$ of 0.64.

to GDP ratios is 0.15 for model simulations (solid line, center panel), and 0.26 for the data (dashed line).[27] Comparing the two coefficients, one may conclude that our model explains about 60 per cent of the cross-country relationship between aggregate TFP and financial development.

Another effect of financial frictions is the impediment of capital accumulation. The right panel of Figure 3 shows the relationship between capital to output ratio and external finance to GDP. In our model, capital to output ratios fall by 15 per cent as we move from the perfect-credit benchmark to financial autarky, when measured at common fixed prices across economies. The regression coefficient of capital to output ratios on external finance to GDP is 0.25 for model simulations (solid line) and 0.76 for the data measured at international prices (dashed line). As we discuss below, financial frictions affect the manufacturing sector and the service sector differentially, with the result that manufactured investment goods become relatively more expensive than services. Consistent with the data (Hsieh and Klenow 2007), this higher relative price of investment explains almost all of the fall in capital to output ratios: The saving and investment rates measured at respective equilibrium prices are roughly constant across our model economies with varying degrees of financial development. This is because of two opposing forces at work. Firstly, financial frictions lead to lower equilibrium interest rates and hence returns to saving, exerting downward pressure on aggregate saving. On the other hand, the resulting lower wages and capital rental rates increase the returns to self-financing for constrained entrepreneurs. These forces seem to offset, leaving the saving and investment rates constant when measured at the respective equilibrium prices.

We also note that the factor-of-two difference in output per worker generated by our model goes a long way if one were to focus on the factor-of-five difference in non-agricultural output per worker between the richest fifth percentile and the poorest fifth percentile of countries.[28] As we do not model the agricultural sector, ideally we would compare our model implications to data on non-agricultural output and productivity. The difficulty here is the dearth of data on sector-level capital stock that are comparable across countries.

We address this issue in two ways. First, we use the data set of Restuccia, Yang and Zhu (2008) covering 82 countries to run a univariate regression of non-agricultural output per worker on external finance to GDP ratios. The regression coefficient is 0.24, and the $R^2$ is 0.45. This coefficient is about 30 per cent smaller than the one using the total output data (dashed line, left panel), and is very close to the slope, 0.22, generated by our model simulations (solid line, left panel): Our model almost fully accounts for the empirical relationship between non-agricultural output per worker and financial development across countries.

Second, because Restuccia, Yang and Zhu do not have data on capital input in the agricultural sector, we again turn to the GGDC Productivity Level Database and analyze the relationship between financial development and the performance of non-agricultural sectors. In each of the three panels, we use a dash-dot line to show the regression line for these 18 OECD countries.[29] These 18 data points are not shown. For the GGDC sample, the regression coefficients of non-agricultural output per worker, productivity, and capital to output ratios on financial development are smaller than those for the overall sample of

[27] The univariate regression of aggregate TFP on external finance in the data has an $R^2$ of 0.51.

[28] See Diego Restuccia, Dennis Tao Yang and Xiaodong Zhu (2008) for a decomposition of cross-country per-capita income differences into agricultural and non-agricultural components.

[29] Among the 18 countries in the sample we use, Hungary is the poorest country, and also has the lowest external finance to GDP ratio (0.32). In the overall sample of the 79 countries analyzed, Hungary ranks in the sixty-sixth percentile of output per worker, and in the thirty-sixth percentile of external finance to GDP.

countries. In fact, for this subset of countries, our model regression estimates are quite comparable to the empirical relationships between the dependent variables and external finance to GDP. The regression coefficients on external finance to GDP are 0.15, 0.07, and 0.09, respectively for output per worker, TFP, and capital to output ratios (dash-dot lines). Recall that the corresponding regression coefficients are 0.22, 0.15, and 0.25 with our model simulations (solid lines).

In summary, financial frictions in our model can explain a large part of the cross-country differences in economic development, measured by output per worker, aggregate TFP, and capital to output ratios. If one were to focus on the non-agricultural sector, our model would account for an even larger part of the cross-country differences.

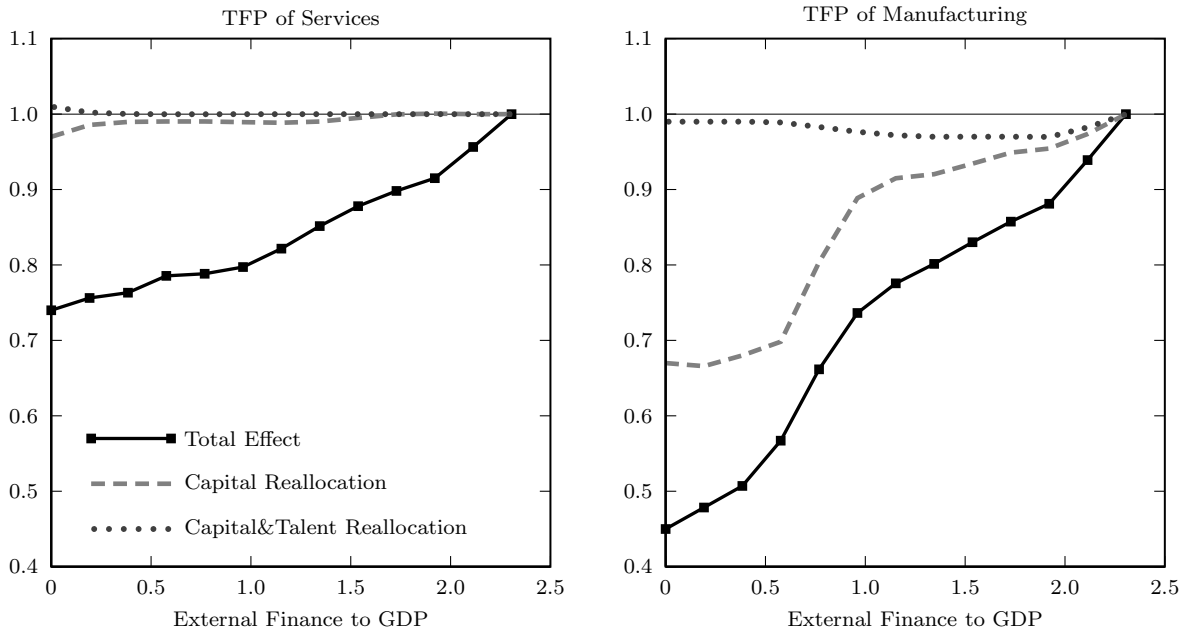


Figure 4. Impact of Financial Frictions on Sector-Level TFP

## Impact on Sector-Level Productivity

Financial frictions have differential impacts on the two sectors. The solid lines in Figure 4 trace the effect of financial frictions on the measured TFP of the service sector (left panel) and the manufacturing sector (right panel). Sector-level TFPs are normalized by their respective levels in the perfect-credit benchmark. While TFP declines by 26 per cent in services, the manufacturing sector TFP declines by 55 per cent with financial frictions. This result is consistent with the empirical observations that productivity differences across countries are sharpest for the large-scale, manufacturing sector.

Next, we examine the driving forces behind these effects on sector-level productivity, which in turn determines the aggregate TFP. Intuitively, financial frictions distort the allocation of productive capital among entrepreneurs in operation. Those with binding collateral constraints will have a marginal product of capital higher than the rental rate. For instance, in the financial autarky ($\phi = 0$), the standard deviation of log marginal product of capital equals 1.07 in services and 1.23 in manufacturing. Financial frictions also

distort the entry and exit decisions of entrepreneurs: Productive-but-poor entrepreneurs delay entry until they can overcome financing constraints, and incompetent-but-wealthy ones remain in business.

In Figure 4, we decompose the effects of financial frictions on sector-level TFP into distortions on the allocation of capital across active entrepreneurs (intensive margin, or misallocation of capital), and into distortions on the allocation of entrepreneurial talent (extensive margin). The extensive margin is further decomposed into the number of active entrepreneurs in each sector, and into the distribution of talent among active entrepreneurs (misallocation of talent).

To quantify this decomposition, we perform three experiments on our model economies. First, we reallocate capital among active entrepreneurs within each sector to equalize the marginal product of capital across them. For a given simulated economy, we hold fixed the number and the talent distribution of existing active entrepreneurs, as well as the total capital and labor employed in each sector. The sector-level TFP after this reallocation is the dashed lines in both panels. For the service sector, almost all of the low TFP is explained by the misallocation of capital among active entrepreneurs. For manufacturing, this intensive-margin distortion explains less than half of the low TFP.

In the second experiment, while holding fixed the number of active entrepreneurs in each sector, we select the most talented individuals into entrepreneurship. We also allocate capital efficiently across the new set of active entrepreneurs, while holding constant the total capital and labor employed in each sector. The resulting sectoral TFP from this reallocation of talent and capital is the dotted lines. The misallocation of talent into entrepreneurship explains more than half of the low TFP in the large-scale, manufacturing sector, and less than one tenth of the low TFP in the small-scale, service sector.

Finally, in addition to the efficient reallocation of talent and capital above, we allow the number of entrepreneurs to adjust in each sector at the perfect-credit equilibrium prices. This additional adjustment affects the TFP only slightly: The dotted lines from the second experiment are already close to the horizontal lines going through one, which represent the respective sector TFP levels in the perfect-credit benchmark. This last experiment suggests that restrictions to entry per se may not have significant quantitative effects unless the distribution of entrants is distorted.

The following conclusions can be drawn from these sector-level exercises. Because of their larger scale and financing needs, establishments in manufacturing are more vulnerable to financial frictions. There is more misallocation of capital and entrepreneurial talent in manufacturing than in services. In particular, the distortions on the entry and exit decisions of entrepreneurs matter vastly more for manufacturing. This result suggests that modeling endogenous entry and exit is pivotal for capturing the full impact of financial frictions.

## Relative Productivity and Relative Prices

The pattern of relative productivity between the two sectors leads to the price of manufactured goods relative to services being higher in countries with underdeveloped financial markets. The left panel of Figure 5 plots the relative price of manufactured goods to services in log from the 1996 ICP against external finance to GDP ratios. The regression coefficients are -0.67 for the data (gray dots, dashed line) and -0.16 for model simulations (diamonds, solid line): Our model accounts for a quarter of the empirical cross-country relationship between the manufacturing-services relative price and financial development, leaving room for other explanations of relative prices that are correlated with financial frictions. As we have seen in Figure 3, this effect on relative prices helps explain the lower capital to output ratios observed in countries with less developed financial markets.

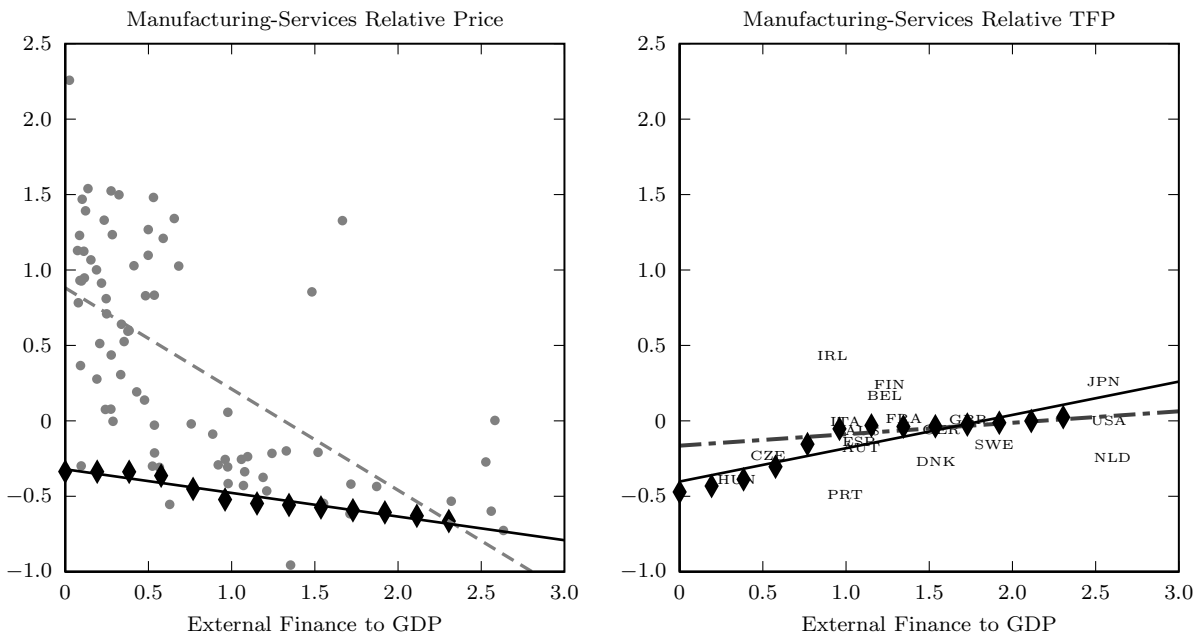


Figure 5. Relative Prices and Relative Productivity against External Finance

In the right panel of Figure 5, we plot the relative productivity of manufacturing to services in log against external finance to GDP ratios. The diamonds are the model simulations, with the corresponding regression line (solid line) that has a slope coefficient of 0.22. We again turn to the GGDC database for the sector-level relative productivity in the 18 OECD countries (three-letter codes). The regression line (dash-dot line) has a slope coefficient of 0.08: Our model generates a relationship between relative sector-level productivity and financial development that is starker than the empirical one found in the available data.

### Impact on Establishment-Level Productivity and Size Distribution

Financial frictions have impacts on the distribution of establishment-level productivity and size as well. The top left panel of Figure 6 plots—against external finance to GDP ratios—the average talent or productivity ($z_j$) of active entrepreneurs in manufacturing (solid line) and services (dashed line), normalized by the average manufacturing entrepreneurial talent in the perfect-credit benchmark. With financial frictions, not only an individual's entrepreneurial talent but also his wealth determine whether he will be an entrepreneur in any given period. As a result, incompetent-but-wealthy entrepreneurs remain in business, and talented-but-poor individuals do not run businesses until they can self-finance the capital needed for a profitable scale. With more financial frictions—and hence less external finance in equilibrium, an individual's wealth has a greater influence upon his decision of entry into and exit from entrepreneurship, and individuals with more diverse entrepreneurial talents will be operating business in equilibrium. There are two consequences. First, the average talent of active entrepreneurs falls. In Figure 4, we have shown that such misallocation of talent is more rampant in the manufacturing sector, which is confirmed here: The average entrepreneurial talent in manufacturing drops by 40 per cent with financial frictions, while it goes down by only 20 per cent in

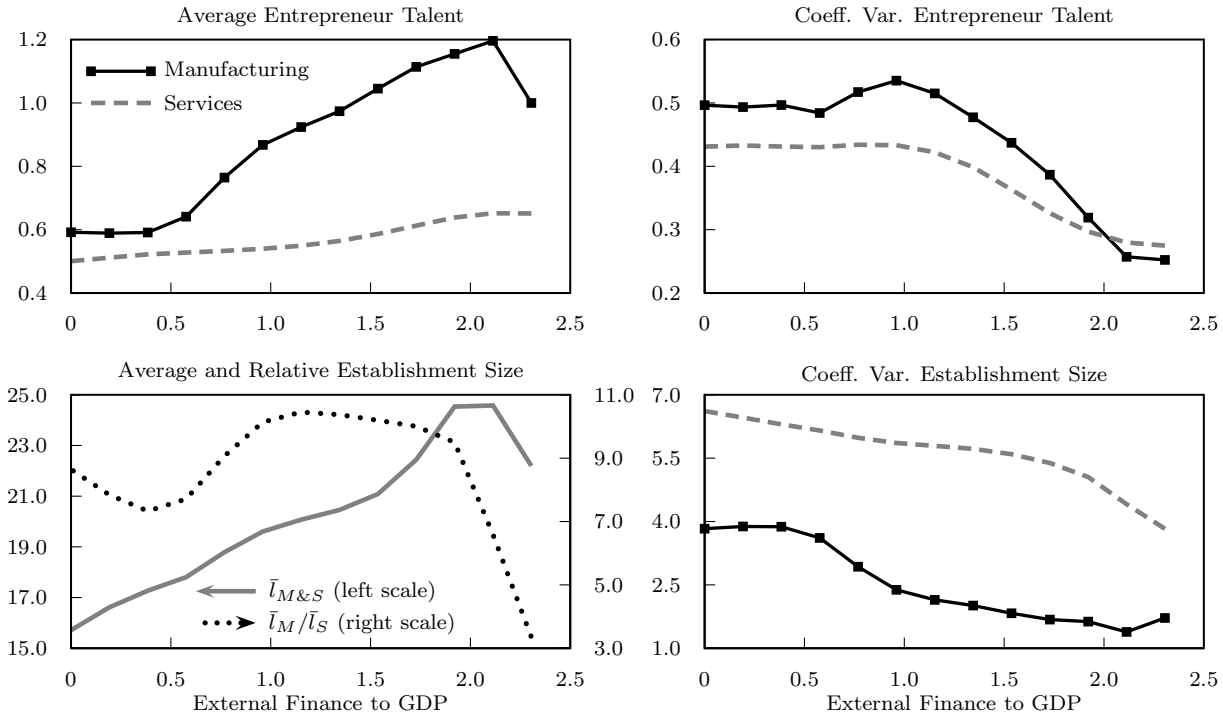


Figure 6. Establishment-Level Impact of Financial Frictions

services.[30] Second, the within-sector distribution of active entrepreneurs' talent becomes more disperse, and its coefficient of variation increases (top right panel).

Next, we explore how financial frictions affect the establishment size distribution. In the bottom left panel, the solid line traces the average establishment size (i.e., number of workers), inclusive of both sectors (left-hand side vertical axis). Financial frictions reduce the average establishment size by up to 30 per cent.[31] This is because lower equilibrium wages (and hence lower opportunity cost of entrepreneurship) owing to financial frictions attract more individuals, including those who are not particularly talented, into entrepreneurship. In this sense, financial frictions generate too many establishments that are too small. This model prediction is consistent with available empirical evidence, although its magnitude is not as large (James R. Tybout 2000).

The marginal entrepreneurs in economies with financial frictions overwhelmingly choose to start businesses in the small-scale, service sector, where the establishment-level production technology commands less financing needs. The larger scale and financing needs

[30]The average entrepreneurial talent in manufacturing is non-monotonic with respect to external finance. With moderate degrees of financial frictions, there seems to be a "positive" selection into entrepreneurship in manufacturing. This is because high-talent entrepreneurs can overcome collateral constraints more easily than low-talent ones. First, for given prices, they make more profits, and can save more. Second, the credit limit $\bar{k}^j$ is increasing in $z_j$. These forces seem to prevail when $\phi$ is close to one. Consistent with this, in the right panel of Figure 4, talent misallocation explains little of the low manufacturing TFP over this region.

[31]As financial frictions intensify, the average establishment size first increases from 22.1 to 24.5, before it goes down to 15.4. The initial increase over the region where external finance to GDP is between 2.3 and 1.9 is the flip side of the fall in the number of establishments, especially in manufacturing.

in manufacturing make it harder for entrepreneurs to start and grow there, reducing the number of manufacturing establishments in equilibrium. At the same time, the relative price of manufactured goods to services increases with financial frictions. As a result, the operating scale of manufacturing establishments relative to service establishments becomes even larger. The dotted line is the ratio of the average establishment size in manufacturing to that in services ($\bar{l}_M/\bar{l}_S$, right-hand side vertical axis). We pursue this prediction next.[32]

For given output and factor prices, the size of an establishment is determined by the entrepreneurial talent and the collateral constraint: Entrepreneurs with the same talent may operate at different scales, as they may have different levels of wealth or collateral. We have seen that financial frictions increase the dispersion of the talent distribution of active entrepreneurs (top right panel), which is now convoluted with the increased dispersion of the establishment size distribution for a given entrepreneurial talent. As a result, the within-sector (bottom right panel) and overall establishment size distributions become more disperse with financial frictions, consistent with the "missing middle" in less developed economies (Tybout 2000).

## Relative Scale of Sectors: A Case Study

One interesting model implication is that financial frictions lead to greater disparity in the average establishment size or scale between manufacturing and services. As discussed above, this result stems not only from the direct effect of financial frictions on entrepreneurial entry and exit decisions, but also from the general equilibrium effects on input and output prices.

We evaluate this prediction using some new evidence from Mexico and the US. We compare establishment size data from the 2002 US Economic Census and the 2003 Mexican Economic Census, both of which follow the NAICS classification and are hence directly comparable.[33] The two censuses have some substantial differences in their coverage. We use Mexico's 1998 National Survey of Micro-Enterprises (ENAMIN) to impute corrections that make the US and the Mexican data fully comparable.[34]

Figure 7 plots the average establishment size in Mexico (in log, vertical axis) against the average establishment size in the US (in log, horizontal axis) for 86 four-digit manufacturing industries and 12 two-digit service industries. The overall average establishment size is substantially smaller in Mexico than in the US, almost by a factor of three. However, many industries (those lying above the 45-degree dashed line) have an average establishment that is larger in Mexico than in the US. Indeed, the data has a slope (solid line) that is significantly steeper than the 45-degree line: The regression coefficient is 1.22 with a standard error of 0.11. That is, the industries that are large scale in the US have an even larger scale in Mexico, while those that are small scale in the US have an even smaller scale in Mexico. With the exception of administration/management services, those above the 45-degree line are manufacturing industries. This finding, hitherto undocumented in the literature, is consistent with our simulation results that the relative

[32] The model of Castro, Clementi and MacDonald (2009) has a similar prediction on the relative scale of sectors, with scale being measured in terms of capital per establishment.

[33] For manufacturing, the classification schemes for Mexico and the US are identical, allowing comparability at the four-digit industry level. For services, the schemes differ and allow comparison at the two-digit level.

[34] These corrections include adjustments to remove non-employers—included in the Mexican Census but not for the US, and adjustments to add small-scale entrepreneurs without a fixed location—included in the US Census, though presumably unimportant, but not for Mexico, where they play an important role.

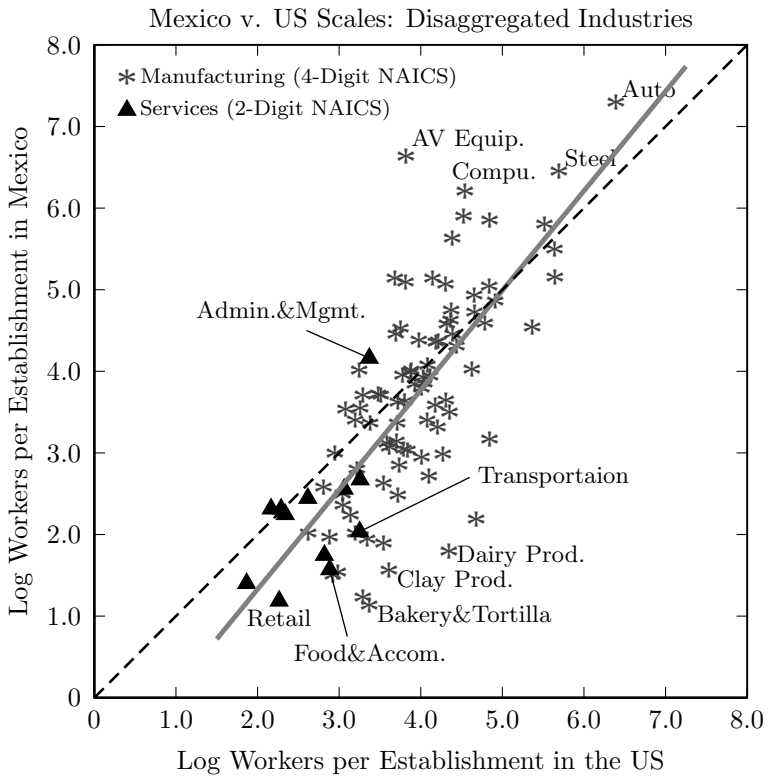


Figure 7. Average Establishment Size of Industries

scale of manufacturing to services is larger in financially less developed economies.[35]

In addition, the fact that, at a more disaggregate industry level, Mexico has wider variations in scale than the US is consistent with our model prediction of more disperse establishment size distribution in financially less developed economies.[36] A formal testing of this prediction, however, requires more detailed establishment-level data that are comparable across countries.

We also conjecture that our mechanism, in which scale differences lead to differential impacts of financial frictions across sectors, will also work at a more disaggregate level, not just at the level of the manufacturing-services dichotomy.

### *C. Discussions on Modeling Choices*

We first consider two other ways of generating sectoral scale differences. We then show that the two main elements of our model—sector-specific non-convexities and self-financing, which the conventional specifications in the literature lack—play important quantitative roles. Finally, we discuss the manufacturing-to-services structural change in

[35]There is other evidence consistent with this result. Using the available OECD SSIS data, we find that financially less developed countries (Czech Republic, Hungary, Poland, Portugal, Slovakia) have a larger relative scale of manufacturing to services (ranging from 3.4 to 4.1) than the US (3.1), where the sector-level scale is defined to be the number of workers per enterprise.

[36]Erosa and Hidalgo Cabrillana (2008) report a related empirical fact: The variance in output per worker across industries within the manufacturing sector is higher in poor countries.

our model context.

### Setup Costs

We have constructed numerical examples for a version of our model where the sectoral scale difference stems from one-time setup costs, rather than per-period fixed costs. We find that financial frictions have an even larger impact on aggregate and sector-level productivity, reflecting the stronger non-convexity that front-loaded setup costs impose. We choose our per-period fixed cost specification to avoid exaggerating the impact of financial frictions, especially when there are no reliable data on setup costs.[37]

### Span of Control Differences

We have considered the possibility that the sectoral scale difference hinges on the establishment-level span of control. In particular, we set $\alpha_j$ and $\theta_j$ such that $\alpha_S + \theta_S < \alpha_M + \theta_M$, with $\alpha_S/\theta_S = \alpha_M/\theta_M$. The latter assumption reflects the empirical facts on factor shares discussed in Section I. In calibrated exercises with span-of-control differences and no fixed cost in either sector, we find that the effect of financial frictions is broadly consistent with our findings in Section III.B, although the magnitude is not as large. We note that, with span-of-control differences, the relative scale of manufacturing to services decreases with financial frictions.

### Comparison with One-Sector Models

One obvious advantage of our two-sector model over one-sector models is its richer implications on relative prices, productivity, and scale between the two sectors. We also find that our two-sector model generates a larger impact of financial frictions on aggregate output and TFP than does a one-sector model, when both models follow the same calibration strategy of Section III.A. For comparability, we pick the fixed cost $\kappa$ in the one-sector model so that the ratios of total fixed costs to GDP are the same in the perfect-credit benchmarks of the two models. Recall that financial frictions can bring down aggregate output by 52 per cent, aggregate TFP by 36 per cent, and the capital to output ratio by 15 per cent (Figure 3). In the calibrated one-sector model, the effect is 39 per cent on output, 30 per cent on TFP, and virtually none on the capital to output ratio.[38]

We conclude that the effect of the non-convexities (i.e., fixed costs) is itself convex: The aggregate impact of the large fixed costs borne by one part of the economy (manufacturing) is larger than that of the small fixed costs spread over the whole economy, holding constant the ratio of total fixed costs to GDP.

### Comparison with Two-Period Models: The Issue of Time Aggregation

Earlier attempts at quantifying the effect of financial frictions often relied on two-period overlapping-generation models. Given their tractability, two-period models are

[37] See, for example, Anna L. Paulson and Robert M. Townsend (2004) and David J. McKenzie and Christopher M. Woodruff (2006) for discussions on setup and fixed costs for small establishments in developing countries. John Sutton (1991) explains the difficulties of estimating setup costs using data from oligopolistic industries in developed countries.

[38] In an alternative one-sector exercise, we assume that there is no fixed cost, while still following the same calibration strategy. The effect of financial frictions are even smaller without fixed costs: 35 per cent on output, 26 per cent on TFP, and again none on the capital to output ratio.

useful for qualitative characterizations of the economic mechanisms at work. However, it is difficult to map these models into data and use them for quantitative purposes. In particular, two-period models nearly preclude one important way of coping with financial frictions: self-financing. With the realistic time horizon in our model, a talented-but-poor individual saves up so that he can overcome the financial constraint over time and operate at the maximal-profit scale. In a two-period model, such a would-be entrepreneur is by construction condemned to the binding collateral constraint for half his life. As a result, the effect of financial frictions in a two-period model is made artificially large. In our own version of a two-period overlapping-generation model, the effect of financial frictions on output and TFP is about 50 per cent larger than in our baseline model with more realistic time horizons.[39] Also, the lack of meaningful self-financing implies virtually no growth in establishment size, and the average establishment size in equilibrium is unrealistically small with financial frictions.

## Structural Change

Empirically, we observe that the service sector becomes more important as economies grow richer (Simon Kuznets 1973). In our framework, one can think of the process of economic development as being driven by exogenous improvement in contract enforcement (an increase in $\phi$) that raises an economy's output and TFP. In our quantitative exercises, the manufacturing share of GDP increases with financial frictions. With perfect credit markets ($\phi = 1$), our calibrated model matches the manufacturing share of the value-added GDP in the US, 25 per cent. As we go to the financial autarky ($\phi = 0$), the manufacturing share of GDP measured at equilibrium prices increases by two percentage points, because investment falls by slightly less than does consumption in relative terms, and investment goods are produced by the manufacturing sector: Our model is at least qualitatively consistent with the observation that manufacturing accounts for a larger share of non-agricultural GDP in less developed economies.[40]

In our analysis, we assume that consumers' elasticity of substitution between manufactured goods and services ($\varepsilon$) is one. If $\varepsilon$ is less than one, consumers will not substitute away as much when the relative price of manufactured goods rises with financial frictions. Therefore, the GDP share of manufacturing will be larger with $\varepsilon < 1$ than with $\varepsilon = 1$ as we intensify financial frictions by lowering $\phi$. To quantify this effect, we have worked out a case with $\varepsilon = 0.75$, while raising the consumption share parameter $\psi$ to 0.96 so that the manufacturing share of GDP with perfect credit is still 25 per cent. We find that, going from the perfect credit case to the financial autarky, the manufacturing share of GDP increases by about three percentage points. In addition, because manufacturing—the sector that is disproportionately affected by financial frictions—now accounts for an even larger part of the economy, the aggregate effect of financial frictions is accordingly larger with $\varepsilon = 0.75$. In fact, in the literature on structural change, $\varepsilon$ is estimated to be close to zero. Francisco J. Buera and Joseph P. Kaboski (2009) and Berthold Herrendorf, Richard Rogerson and Ákos Valentinyi (2009) find that the preference over manufactured goods and services value-added is close to Leontief ($\varepsilon = 0$). A lower $\varepsilon$ would certainly

[39] Our two-period overlapping-generation model follows as much as possible the calibration strategy in Section III.A. The production side parameters are not affected by the two-period assumption. Individuals are born with no wealth, live and work—with occupation choice—for two periods, and a period is thought of as 20 years. The probability of an individual retaining his talent from the first period to the second is $\gamma^{20}$. We compute the perfect-credit case ($\phi = 1$) and the financial autarky ($\phi = 0$) to compute the effect of financial frictions.

[40] According to the World Development Indicators, industry accounts for 30 per cent of non-agricultural GDP in the richest decile of countries in 2000, and 35 per cent in the poorest decile.

improve the performance of our model in terms of the structural change from manufacturing to services along the process of economic development, and would also magnify the impact of financial frictions on output and productivity. We nevertheless choose to be conservative and set $\varepsilon = 1$ for our quantitative analysis.[41]

In a related vein, we wish to remind readers that our calibration is based on the US as a perfect credit benchmark. In particular, the parameter governing manufacturing's share of GDP is chosen to match the recent US level, 25 per cent, which is on the lower end of the spectrum across various countries in different stages of economic development. Had we targeted a higher manufacturing share—e.g., a cross-country average—or introduced an exogenous force leading to a higher manufacturing share in less developed economies, we would have found a much larger effect of financial frictions because manufacturing is the sector that is disproportionately affected by financial frictions. Again, to be conservative and robust, we target the manufacturing share in the US and quantify the effect of financial frictions that is not convoluted with structural change.

Finally, we note that the rise of the service sector is primarily driven by the emergence of a new kind of service industries that look rather like manufacturing in terms of scale. Consistent with this trend, the average establishment size inclusive of all manufacturing and service industries has remained roughly constant over the past 30 years in the US, as large-scale services waxed and manufacturing waned.[42] We emphasize that the key sectoral distinction in our model is scale or establishment size, although we use the manufacturing-services dichotomy to map our model into data. A richer model with scale heterogeneity across industries within the two sectors will be more suitable for the analysis of structural change. In such a model, manufacturing-to-services structural change would be neutral with respect to the relative importance of large-scale vs. small-scale sectors in an economy.

## IV. Concluding Remarks

We have developed a quantitative theory linking financial development to output per worker, aggregate TFP, sector-level relative productivity, and relative prices. Financial frictions distort the allocation of capital and entrepreneurial talent, and have sizable adverse effects on a country's output per worker and aggregate productivity. Establishments in sectors characterized by larger fixed costs operate most efficiently at larger scales, and hence have bigger financing needs. For this reason, they are more vulnerable to financial frictions than those in small-scale sectors. We have shown that this mechanism almost fully explains the relationship between financial development and relative productivity of sectors in the available data. We have also shown that our mechanism is consistent with the larger disparity in average establishment size across sectors observed in financially less developed countries.

Our theory and its quantitative implementation have revolved around the technological difference between two broadly-defined sectors: manufacturing vs. services. We have abstracted from the rich heterogeneity across industries within each sector, but our mechanism appears to play a role at a more disaggregate level as well. Such heterogeneity implies even larger differences in fixed costs or non-convexities across industries, and hence a more disaggregate model may lead to even larger effects of financial frictions on

[41] Herrendorf, Rogerson and Valentinyi (2009) also use data on final consumption, and estimate $\varepsilon$ to be close to one. However, for our value-added specification of technologies, values of $\varepsilon$ close to zero are more relevant.

[42] In 1973, the average US service establishment had 11 employees. In 2007, it employed 15. From 1946 to 1973 a similar trend existed for the average size of firms, the reporting unit during that period: The average size of service firms went from 8 to 12 (County Business Patterns and Economic Census).

economic development.

Our analysis shows how micro-level (firm or establishment) technology differences across sectors interact with financial frictions and help us better understand macroeconomic issues. In this context, we view the study of other micro-level distortions—e.g., size-dependent policies of Guner, Ventura and Xu (2008) or entry barriers of Simeon Djankov, Rafael La Prota, Florencio Lopez-de Silanes and Andrei Shleifer (2002)—and their interaction with financial frictions as promising avenues for future research.

## REFERENCES

**Amaral, Pedro S., and Erwan Quintin.** 2010. "Limited Enforcement, Financial Intermediation, and Economic Development: A Quantitative Assessment." *International Economic Review*, 51(3): 785–811.

**Baily, Martin N., and Eric Zitzewitz.** 2001. "Service Sector Productivity Comparisons: Lessons for Measurement." In *New Developments in Productivity Analysis.* , ed. Charles R. Hulten, Edwin R. Dean and Michael J. Harper, 419–455. Chicago:University of Chicago Press.

**Baily, Martin N., and Robert M. Solow.** 2001. "International Productivity Comparisons Built from the Firm Level." *Journal of Economic Perspectives*, 15(3): 151–172.

**Balassa, Bela.** 1964. "The Purchasing Power Parity Doctrine: A Reappraisal." *Journal of Political Economy*, 72(6): 584–596.

**Banerjee, Abhijit V., and Esther Duflo.** 2005. "Growth Theory through the Lens of Development Economics." In *Handbook of Economic Growth.* Vol. 1A, , ed. Philippe Aghion and Steven N. Durlauf, 473–552. Amsterdam:Elsevier.

**Beck, Thorsten, Asli Demirgüç-Kunt, and Ross Levine.** 2000. "A New Database on the Structure and Development of the Financial Sector." *World Bank Economic Review*, 14(3): 597–605.

**Beck, Thorsten, Asli Demirgüç-Kunt, Luc Laeven, and Ross Levine.** 2008. "Finance, Firm Size, and Growth." *Journal of Money, Credit and Banking*, 40(7): 1327–1554.

**Buera, Francisco J., and Joseph P. Kaboski.** 2008. "Scale and the Origins of Structural Change." UCLA Manuscript.

**Buera, Francisco J., and Joseph P. Kaboski.** 2009. "Can Traditional Theories of Structural Change Fit The Data?" *Journal of the European Economic Association*, 7(2-3): 469–477.

**Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin.** 2009. "Finance and Development: A Tale of Two Sectors." National Bureau of Economic Research Working Paper 14914.

**Caselli, Francesco, and Nicola Gennaioli.** 2005. "Dynastic Management." London School of Economics Manuscript.

**Castro, Rui, Gian Luca Clementi, and Glenn MacDonald.** 2009. "Legal Institutions, Sectoral Heterogeneity, and Economic Development." *Review of Economic Studies*, 76(2): 529–561.

**Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan.** 1997. "The Poverty of Nations: A Quantitative Investigation." Federal Reserve Bank of Minneapolis Staff Report 204.

**Davis, Steven J., John C. Haltiwanger, Ron Jarmin, and Javier Miranda.** 2006. "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately

Held Firms." In *NBER Macroeconomics Annual 2006.* , ed. Mark Gertler and Kenneth Rogoff, 107–156. Cambridge, MA:MIT Press.

**Djankov, Simeon, Rafael La Prota, Florencio Lopez-de Silanes, and Andrei Shleifer.** 2002. "The Regulation of Entry." *Quarterly Journal of Economics*, 117(1): 1–37.

**Erosa, Andrés.** 2001. "Financial Intermediation and Occupational Choice in Development." *Review of Economic Dynamics*, 4(2): 303–334.

**Erosa, Andrés, and Ana Hidalgo Cabrillana.** 2008. "On Finance as a Theory of TFP, Cross-Industry Productivity Differences, and Economic Rents." *International Economic Review*, 49(2): 437–473.

**Giné, Xavier, and Robert M. Townsend.** 2004. "Evaluation of Financial Liberalization: A General Equilibrium Model with Constrained Occupation Choice." *Journal of Development Economics*, 74(2): 269–307.

**Greenwood, Jeremy, Juan M. Sanchez, and Cheng Wang.** 2010. "Quantifying the Impact of Financial Development on Economic Development." Federal Reserve Bank of Richmond Working Paper 10-05.

**Guner, Nezih, Gustavo Ventura, and Yi Xu.** 2008. "Macroeconomic Implications of Size-Dependent Policies." *Review of Economic Dynamics*, 11(4): 721–744.

**Hall, Robert E., and Charles I. Jones.** 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics*, 114(1): 83–116.

**Herrendorf, Berthold, Richard Rogerson, and Ákos Valentinyi.** 2009. "Two Perspectives on Preferences and Structural Transformation." Arizona State University Manuscript.

**Hopenhayn, Hugo A., and Richard Rogerson.** 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy*, 101(5): 915–938.

**Hsieh, Chang-Tai, and Peter Klenow.** 2007. "Relative Prices and Relative Prosperity." *American Economic Review*, 97(3): 562–585.

**Hsieh, Chang-Tai, and Peter Klenow.** 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–1448.

**Inklaar, Robert, and Marcel P. Timmer.** 2008. "GGDC Productivity Level Database: International Comparisons of Output, Inputs and Productivity at the Industry Level." Groningen Growth and Development Centre Working Paper GD-104.

**Jeong, Hyeok, and Robert M. Townsend.** 2007. "Sources of TFP Growth: Occupational Choice and Financial Deepening." *Economic Theory*, 32(1): 197–221.

**King, Robert G, and Ross Levine.** 1993. "Finance and Growth: Schumpeter Might Be Right." *Quarterly Journal of Economics*, 108(3): 717–737.

**Klenow, Peter J., and Andrés Rodríguez-Clare.** 1997. "The Neoclassical Revival in Growth Economics: Has it Gone Too Far?" In *Macroeconomics Annual 1997.* , ed. Ben S. Bernanke and Julio J. Rotemberg, 73–102. Cambridge:MIT Press.

**Kuznets, Simon.** 1973. "Modern Economic Growth: Findings and Reflections." *American Economic Review*, 63(3): 247–258.

**La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert W. Vishny.** 1998. "Law and Finance." *Journal of Political Economy*, 106(6): 1113–1155.

**Lucas, Jr., Robert E.** 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics*, 9(2): 508–523.

**Matsuyama, Kiminori.** 2007. "Aggregate Implications of Credit Market Imperfec-

tions." In *NBER Macroeconomics Annual*. Vol. 22, , ed. Daron Acemoglu, Kenneth Rogoff and Michael Woodford. Cambridge, MA:MIT Press.

**McKenzie, David J., and Christopher M. Woodruff.** 2006. "Do Entry Costs Provide an Empirical Basis for Poverty Traps? Evidence from Mexican Microenterprises." *Development and Cultural Change*, 55(1): 3–42.

**Parente, Stephen L., and Edward C. Prescott.** 2000. *Barriers to Riches.* Cambridge:MIT Press.

**Paulson, Anna L., and Robert M. Townsend.** 2004. "Entrepreneurship and Financial Constraints in Thailand." *Journal of Corporate Finance*, 10(2): 229–262.

**Rajan, Raghuram G., and Luigi Zingales.** 1998. "Financial Dependence and Growth." *American Economic Review*, 88(3): 559–586.

**Restuccia, Diego, and Richard Rogerson.** 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments." *Review of Economic Dynamics*, 11(4): 707–720.

**Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu.** 2008. "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis." *Journal of Monetary Economics*, 55(2): 234–250.

**Rossi-Hansberg, Esteban, and Mark L. J. Wright.** 2007. "Establishment Size Dynamics in the Aggregate Economy." *American Economic Review*, 97(5): 1639–1666.

**Samuelson, Paul A.** 1964. "Theoretical Notes on Trade Problems." *Review of Economics and Statistics*, 46(2): 145–154.

**Sutton, John.** 1991. *Sunk Costs and Market Structure.* Cambridge, MA:MIT Press.

**Townsend, Robert M.** 2010. *Financial Systems in Developing Economies: Growth, Inequality and Policy Evaluation in Thailand.* Oxford:Oxford University Press.

**Tybout, James R.** 2000. "Manufacturing Frims in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature*, 38(1): 11–44.

**Valentinyi, Ákos, and Berthold Herrendorf.** 2008. "Measuring Factor Income Shares at the Sector Level." *Review of Economic Dynamics*, 11(4): 820–835.

## Proofs

PROOF OF PROPOSITION 1:

The rental of capital $k$ in sector $j$ is enforceable iff:

$$\tilde{v}^j\left(k;a,\mathbf{z}\right)\geq v^d\left(a^d,\mathbf{z}\right),$$

where $\tilde{v}^j\left(k;a,\mathbf{z}\right)$ is the value of a non-defaulting entrepreneur with wealth $a$ and ability $\mathbf{z}$ that operates in sector $j$ with rented capital $k$:

$$\tilde{v}^j\left(k;a,\mathbf{z}\right)=\max_{\mathbf{c},a'}\left\{u\left(\mathbf{c}\right)+\beta\left[\gamma v\left(a',\mathbf{z}\right)+\left(1-\gamma\right)\mathbb{E}_{\mathbf{z}'}v\left(a',\mathbf{z}'\right)\right]\right\}$$
$$\text{s.t. }\mathbf{p}\cdot\mathbf{c}+a'\leq\max_{l}\left\{p_jz_jf\left(k,l\right)-wl\right\}-Rk-\left(1+r\right)p_j\kappa_j+\left(1+r\right)a.$$

We define $v^d\left(a^d,\mathbf{z}\right)$ to be the value of a defaulting entrepreneur with ability $\mathbf{z}$ who gets

to keep $a^d = (1-\phi)\left[\max_l \{p_j z_j f(k,l) - wl\} + (1-\delta)k\right]$:

$$v^d\left(a^d, \mathbf{z}\right) = \max_{\mathbf{c}, a'} \{u(\mathbf{c}) + \beta\left[\gamma v(a', z) + (1-\gamma)\mathbb{E}_{z'} v(a', \mathbf{z}')\right]\}$$
$$\text{s.t. } \mathbf{p}\cdot\mathbf{c} + a' \leq (1-\phi)\left[\max_l \{p_j z_j f(k,l) - wl\} + (1-\delta)k\right].$$

It is straightforward to see that $\tilde{v}^j(k; a, \mathbf{z}) \geq v^d\left(a^d, \mathbf{z}\right)$ iff:

$$\max_l \{p_j z_j f(k,l) - wl\} - Rk - (1+r)p_j\kappa_j + (1+r)a$$
$$\geq (1-\phi)\left[\max_l \{p_j z_j f(k,l) - wl\} + (1-\delta)k\right],$$

which is equivalent to:

$$(1+r)(a - p_j\kappa_j) \geq (1-\phi+r+\delta\phi)k - \phi\max_l \{p_j z_j f(k,l) - wl\}$$
$$= -\phi\left[\max_l \{p_j z_j f(k,l) - wl\} - \frac{1-\phi+r+\delta\phi}{\phi}k\right].$$

Note that as long as $f_{kk}(k,l) < 0$, $f_{ll}(k,l) < 0$, $\lim_{k\to 0} f_k(k,l) = \infty$ and $\lim_{k\to\infty} f_k(k,l) = 0$, the right hand side is minimized for some $\hat{k}(z_j;\phi)$ such that $0 < \hat{k}(z_j;\phi) < k_j^u(z_j)$, as it equals $-\phi$ times the profits of an entrepreneur facing a rental price $(1-\phi+r+\delta\phi)/\phi > r+\delta$. The set of enforceable levels of capital rental can be easily characterized by a simple set of rental limits. There are two cases to consider.

If $(1+r)(a - p_j\kappa_j) > -\phi\left[\max_l\left\{p_j z_j f\left(\hat{k}(z_j;\phi), l\right) - wl\right\} - \frac{1-\phi+r+\delta\phi}{\phi}\hat{k}(z_j;\phi)\right]$, there exists a unique function $\bar{k}^j(a, z_j;\phi) \geq \hat{k}(z_j;\phi)$ given by the largest root of the equation:

$$(1+r)(a - p_j\kappa_j) = (1-\phi+r+\phi\delta)\bar{k}^j(a, z_j;\phi) - \phi\max_l\left\{p_j z_j f\left(\bar{k}^j(a, z_j;\phi), l\right) - wl\right\}.$$

Note that if $a - p_j\kappa_j < 0$, there are two positive roots of the equation, with the smaller root satisfying $\underline{k}^j(a, z_j;\phi) \leq \hat{k}(z_j;\phi)$. In this case, the set of enforceable levels of capital rental is $\left[\underline{k}^j(a, z_j;\phi), \bar{k}^j(a, z_j;\phi)\right]$. If $a - p_j\kappa_j \geq 0$, the set of enforceable levels of capital rental is simply $\left[0, \bar{k}^j(a, z_j;\phi)\right]$. It is straightforward to see that $\bar{k}^j(a, z_j;\phi)$ is strictly increasing in $a$, $z_j$, and $\phi$.

If $(1+r)(a - p_j\kappa_j) \geq -\phi\left[\max_l\left\{p_j z_j f\left(\hat{k}(z_j;\phi), l\right) - wl\right\} - \frac{1-\phi+r+\delta\phi}{\phi}\hat{k}(z_j;\phi)\right]$, we set $\bar{k}^j(a, z_j;\phi) = 0$.

PROOF OF PROPOSITION 2:

In an economy with perfect credit markets, selection of individuals into entrepreneurship and sectors is determined by their entrepreneurial talents and relative prices. In particular, there exist two threshold ideas $\hat{z}_j$, $j = S, M$, and a function $\hat{z}_j(z_{-j})$, $(j, -j) = (S, M), (M, S)$, dividing the space of entrepreneurial ideas $(z_S, z_M)$ into workers and entrepreneurs in the $S$ and $M$ sectors. These thresholds are defined by the following three indifference conditions:

$$(6) \qquad (p_j \hat{z}_j)^{\frac{1}{1-\alpha-\theta}} \left(\frac{\alpha}{R}\right)^{\frac{\alpha}{1-\alpha-\theta}} \left(\frac{\theta}{w}\right)^{\frac{\theta}{1-\alpha-\theta}} (1-\alpha-\theta) = w + p_j \kappa_j (1+r), \quad s = S, M,$$

$$(p_j \hat{z}_j (z_{-j}))^{\frac{1}{1-\alpha-\theta}} \left(\frac{\alpha}{R}\right)^{\frac{\alpha}{1-\alpha-\theta}} \left(\frac{\theta}{w}\right)^{\frac{\theta}{1-\alpha-\theta}} (1-\alpha-\theta) - p_j \kappa_j (1+r)$$

$$(7) \qquad = (p_{-j} z_{-j})^{\frac{1}{1-\alpha-\theta}} \left(\frac{\alpha}{R}\right)^{\frac{\alpha}{1-\alpha-\theta}} \left(\frac{\theta}{w}\right)^{\frac{\theta}{1-\alpha-\theta}} (1-\alpha-\theta) - p_{-j} \kappa_{-j} (1+r).$$

Integrating over individual output of entrepreneurs in sector $j$ net of fixed costs, we obtain an expression for the net output of sector $j$,

$$(8) \qquad Y_j = N \int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-j}(z_j)} z_j k(z_j)^{\alpha} l(z_j)^{\theta} \mu(d\mathbf{z}) - \kappa_j (1+r) N \int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-j}(z_j)} \mu(d\mathbf{z}).$$

Using $k(z_j) = \frac{z_j^{\frac{1}{1-\alpha-\theta}}}{Z_j^{\frac{1}{1-\alpha-\theta}}} \frac{K_j}{N}$ and $l(z_j) = \frac{z_j^{\frac{1}{1-\alpha-\theta}}}{Z_j^{\frac{1}{1-\alpha-\theta}}} \frac{L_j}{N}$, which follow from the first order conditions of the entrepreneurs' problem, we can rewrite (8) as:

$$\begin{aligned} Y_j &= N \int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-j}(z_j)} z_j \left[\frac{z_j^{\frac{1}{1-\alpha-\theta}}}{Z_j^{\frac{1}{1-\alpha-\theta}}} \frac{K_j}{N_j}\right]^{\alpha} \left[\frac{z_j^{\frac{1}{1-\alpha-\theta}}}{Z_j^{\frac{1}{1-\alpha-\theta}}} \frac{L_j}{N_j}\right]^{\theta} \mu(d\mathbf{z}) - \kappa_j (1+r) N \int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-j}(z_j)} \mu(d\mathbf{z}) \\ &= \frac{N^{1-\alpha-\theta} K_j^{\alpha} L_j^{\theta}}{Z_j^{\frac{\alpha+\theta}{1-\alpha-\theta}}} \int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-j}(z_j)} z_j^{\frac{1}{1-\alpha-\theta}} \mu(d\mathbf{z}) - \kappa_j (1+r) N \int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-j}(z_j)} \mu(d\mathbf{z}). \end{aligned}$$

Then,

$$(9) \qquad Y_j = N^{1-\alpha-\theta} K_j^{\alpha} L_j^{\theta} Z_j - \kappa_j (1+r) N \int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-j}(z_j)} \mu(d\mathbf{z}),$$

where $Z_j = \left[\int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-j}(z_j)} z_j^{\frac{1}{1-\alpha-\theta}} \mu(d\mathbf{z})\right]^{1-\alpha-\theta}$.

Assuming $\mu(d\mathbf{z}) = \eta^2 (z_S z_L)^{-(\eta+1)}$ for $z_j \geq 1$ and that entrepreneurs are a small fractions of the population—i.e. $\hat{z}_j$ is large for $j = S, M$, we obtain

$$Z_j = \left[\int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-s}(z_j)} z_j^{\frac{1}{1-\alpha-\theta}} \mu(d\mathbf{z})\right]^{1-\alpha-\theta} \approx \left[\int_{\hat{z}_j}^{\infty} z_j^{\frac{1}{1-\alpha-\theta}} \eta z_j^{-(\eta+1)} dz_j\right]^{1-\alpha-\theta},$$

$$\int_{\hat{z}_j}^{\infty} \int_1^{\hat{z}_{-j}(z_j)} \mu(d\mathbf{z}) \approx \hat{z}_j^{-\eta}.$$

Here, we further assume $\frac{1}{\eta} < (1-\alpha-\theta)$ to guarantee that the integral is finite.

Using (6), as well as $\frac{\alpha}{R} = \frac{K_j^{1-\alpha}}{L_j^{\theta} N^{1-\alpha-\theta}} \frac{1}{p_j Z_j}$ and $\frac{\theta}{w} = \frac{L_j^{1-\theta}}{K_j^{\alpha} N^{1-\alpha-\theta}} \frac{1}{p_j Z_j}$, we obtain

$$\hat{z}_j = \left\{ \frac{\left(\kappa_j(1+r) + \frac{w}{p_j}\right) \left[\frac{(1-\alpha-\theta)}{\eta(1-\alpha-\theta)-1}\eta\right]^{\alpha+\theta}}{(1-\alpha-\theta) N^{-\alpha-\theta} K_j^{\alpha} L_j^{\theta}} \right\}^{\frac{1}{1+\eta(\alpha+\theta)}}. \tag{10}$$

Substituting into (9),

$$Y_j = A_j N^{\frac{1}{1+\eta(\alpha+\theta)}} K_j^{\frac{\alpha\eta}{1+\eta(\alpha+\theta)}} L_j^{\frac{\theta\eta}{1+\eta(\alpha+\theta)}},$$

where

$$A_j = \frac{\left[\frac{\eta(1-\alpha-\theta)}{\eta(1-\alpha-\theta)-1}\right]^{\frac{1}{1+\eta(\alpha+\theta)}} \left[1 - \frac{p_j\kappa_j}{p_j\kappa_j + w}\left(1-\alpha-\theta-\frac{1}{\eta}\right)\right]}{\left[\frac{p_j\kappa_j + w}{p_j(1-\alpha-\theta)}\right]^{\frac{\eta(1-\alpha-\theta)-1}{1+\eta(\alpha+\theta)}}}. \tag{11}$$

PROOF OF PROPOSITION 3:

From the first order condition of an entrepreneur of productivity $z_j$ and that of the marginal entrepreneur ($\hat{z}_j$), we obtain

$$l(z_j) = \left(\frac{z_j}{\hat{z}_j}\right)^{\frac{1}{1-\alpha-\theta}} l(\hat{z}_j).$$

Thus,

$$\Pr\left[\tilde{l}_j > l\right] = \Pr\left[z_j > \left(\frac{l}{l(\hat{z}_j)}\right)^{1-\alpha-\theta} \hat{z}_j \middle| z_j \geq \hat{z}_j\right] = \left(\frac{l(\hat{z}_j)}{l}\right)^{\eta(1-\alpha-\theta)}.$$

The aggregate establishment size distribution in the economy is then given by a mixture of Pareto distributions:

$$\Pr\left[\tilde{l}_j > l\right] = n_S \left(\frac{l(\hat{z}_S)}{l}\right)^{\eta(1-\alpha-\theta)} + n_M \left(\frac{l(\hat{z}_M)}{\max\{l, l(\hat{z}_M)\}}\right)^{\eta(1-\alpha-\theta)}, \quad l \geq l(\hat{z}_S).$$

Finally, by integrating $l(z_j) = \frac{z_j^{\frac{1}{1-\alpha-\theta}}}{Z_j^{\frac{1}{1-\alpha-\theta}}} \frac{L_j}{N}$ over $z_j$, we calculate the average establishment size in sector $j$:

$$\bar{l}_j = \frac{L_j}{N(1-\mu(\hat{z}_j))}. \tag{12}$$

The optimal allocation of labor $L_j$ and entrepreneurs $N\left(1-\mu\left(\hat{z}_j\right)\right)$ to sector $j$ implies:

$$\theta p_j N^{1-\alpha-\theta} Z_j K_j^{\alpha} L_j^{\theta-1} = w, \tag{13}$$

$$(1-\alpha-\theta)\, p_j N^{-\alpha-\theta} Z_j K_j^{\alpha} L_j^{\theta} = p_j \kappa_j (1+r)\left(1-\mu\left(\hat{z}_j\right)\right) + w\mu\left(\hat{z}_j\right). \tag{14}$$

Taking the ratio of these two conditions, we obtain:

$$\frac{1-\alpha-\theta}{\theta}\frac{L_j}{N\left(1-\mu\left(\hat{z}_j\right)\right)} = \frac{p_j\kappa_j(1+r)}{w} + \frac{\mu\left(\hat{z}_j\right)}{1-\mu\left(\hat{z}_j\right)}. \tag{15}$$

Substituting (14) into (10), we obtain:

$$\hat{z}_j = \left\{ \frac{\left(\kappa_j(1+r)+\frac{w}{p_j}\right)\left[\frac{(1-\alpha-\theta)}{\eta(1-\alpha-\theta)-1}\eta\right]^{\alpha+\theta}}{\frac{p_j\kappa_j(1+r)(1-\mu(\hat{z}_j))+\frac{w}{p_j}\mu(\hat{z}_j)}{Z_j}} \right\}^{\frac{1}{1+\eta(\alpha+\theta)}}, \tag{16}$$

$$\frac{p_j\kappa_j(1+r)}{w} + \frac{\mu\left(\hat{z}_j\right)}{1-\mu\left(\hat{z}_j\right)} = \left[\frac{p_j\kappa_j(1+r)}{w}+1\right]\frac{\eta\left(1-\alpha-\theta\right)}{\eta\left(1-\alpha-\theta\right)-1}. \tag{17}$$

Combining (12), (15) and (17), we obtain the desired expression:

$$\frac{\bar{l}_j}{\bar{l}_{j'}} = \frac{p_j\kappa_j + w}{p_{j'}\kappa_{j'} + w}. \tag{18}$$