The Missing "Missing Middle"[†]

Chang-Tai Hsieh and Benjamin A. Olken

The notion that the distribution of firm size in poor countries is character-ized by a bimodal distribution with a "missing middle" is a widely accepted fact in development economics (for example, see Krueger 2013). The idea of the missing middle is that there are a large number of small firms, some large firms, but very few medium-sized firms.

The purported fact about the missing middle is cited as evidence for two broad stories of why many countries are poor. Perhaps surprisingly, these models look for the cause of the missing middle in two fundamentally different places. One approach suggests that *small firms* are disfavored in low-income countries—for example, by a lack of access to financial capital—and thus face difficulties in growing to become middle-sized firms. The other approach posits a "dual economy" of large high-productivity firms and small low-productivity firms, and then suggests that *larger firms* are disfavored in low-income countries—for example, by having to bear large fixed costs of regulation—which make it difficult for middle-sized firms to become established. We begin by reviewing these theories of development based on the purported fact of a "missing middle." We then explore the evidence on distri-bution of firm sizes in more detail and challenge the presumption that a "missing middle" occurs at all.

We present three main facts. First, there is in fact no evidence of a missing middle in detailed and comprehensive data on the size of manufacturing firms in India, Indonesia, or Mexico, regardless of how we slice the data. To be sure, there

■ *Chang-Tai Hsieh is the Phyllis and Irwin Winkelried Professor of Economics, University of Chicago, Chicago, Illinois. Benjamin A. Olken is Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. Their email addresses are chsieh @chicagobooth.edu and bolken@mit.edu.*

are many more small firms in developing countries—but while medium-sized firms are missing in the data, large firms are missing as well. Put differently, while there are fewer middle-sized firms in developing countries than developed countries, there is no missing middle in the sense of a bimodal distribution.

Second, the average product of labor and capital is significantly lower in small firms when compared to larger firms. This is important because some theories say that small firms do not grow because they face high marginal costs of capital; if so, the marginal product of the capital that they do have should be higher. While we do not directly observe the marginal product of capital, it appears that the average product of labor and capital is significantly *lower* in small firms when compared to larger firms. To the extent that marginal and average costs move together, this fact suggests that large firms rather than small firms are the ones suffering the large fixed costs or shortage of capital that could stifle their growth.

Third, we consider the possibility that regulatory obstacles generate a missing middle, but find no evidence of meaningful discontinuities in the firm size distribution. We focus on regulations that kick in at a certain size threshold and test whether there are an unusually large number of firms right under the threshold and an unusually small number of firms right above the threshold: specifically, we focus on a size threshold of 100 employees in India where various labor regulations kick in; a revenue threshold in Indonesia above which firms are required to pay value-added tax; and a revenue threshold in Mexico above which firms face higher tax rates. However, we find no economically meaningful bunching of firms around these thresholds, which suggests that stories based on thresholds due to formality or regulations are unlikely to be causing major distortions in the economy. This evidence does not rule out the possibility that such forces are present, but it suggests that if fixed costs or thresholds are important, they must vary substantially across firms.

Given these facts, a natural question is how this misconception about the missing middle arose in the first place. We suggest that the misconception about the missing middle comes from the two transformations that have been made to data in generating the main evidence cited for the missing middle. Specifically, the main citation for the existence of the missing middle is a table on the distribution of employment shares in a number of countries in Tybout (2000). Due to data limitations, these tabulations are binned into three groups: firms with less than 10 employees, 10–49 employees, and 50 or more employees. The "missing middle" refers to the fact that there is less employment in the middle category (10–49 employees) than in the other two bins. In addition, these tabulations present the distribution of *employment share* by firm size and not the distribution of the number of *firms* by size. However, the relevant theories for which the missing middle is a key fact are about the firm itself (for example, theories that firms over a size are differentially taxed, or firms have trouble getting credit and so can't grow above a certain size) and not about the employment share, which is instead more relevant for understanding where the typical worker in the economy works.

We show that the widely cited facts about the missing middle come from the product of these two transformations of the data (using three broad bins and the employment share); neither one alone will produce the effect. When we bin the

firm size distribution into these three broad bins, it appears unimodal. Similarly, when plotted flexibly as a histogram, the employment share distribution appears unimodal. Only when one groups the employment share distribution into these three bins does the "missing middle" pattern emerge.

The absence of a missing middle suggests that the theories of development that cite the missing middle "fact" are not correct, at least without substantial modifications. Our evidence suggests that a major problem of economic development in low-income countries may not be how to relax the constraints faced by small firms, but instead how to relieve the differential constraints faced by large firms. In turn, this view of the world suggests that programs such as microcredit or tax regimes that seek to benefit small firms can worsen the development problem by further increasing the incentive for firms to remain small.

The Missing Middle in Theories of Development

The missing middle is an important presumed fact behind two models of development, one that emphasizes that small firms are disfavored relative to large ones and another that emphasizes that medium and large firms are disfavored relative to smaller ones.

The first model is the view that the institutional environment in poor countries discriminates against small firms and favors big firms. Such models come in several versions. The most common version, often put forward by supporters of micro-lending, is based on a claim that small firms are credit constrained—that is, they would like to borrow a larger quantity of funds at the prevailing interest rate but are unable to find a lender willing to lend—and large firms are not credit constrained. Closely related mechanisms are based on the idea that property rights are protected for formal firms but not for informal firms (De Soto 1989), or that large firms have better access to output markets. Other related models are based on the idea that government interventions benefit large firms, perhaps because the large firms are state-owned firms, or because industrial policy targets large firms, or because large firms are the main beneficiaries of protectionism and entry barriers.

A central prediction of many of these models is that the marginal return to resources should be higher in small firms compared to large firms. If small firms are constrained in their ability to obtain capital, they will also have high marginal products for the capital that they do have. Indeed, a number of papers estimate a very high return to capital in small firms in developing countries (de Mel, McKenzie, and Woodruff 2008; Udry and Anagol 2006; Kremer, Lee, Robinson, and Rostapshova 2013).

A second model of development that generates a missing middle harkens back to the "dual economy" view, as expressed in Arthur Lewis's (1954, p. 147) famous characterization of poor countries as islands of capitalist employment, "surrounded by a vast sea of subsistence workers . . . a few industries highly capitalized, such as mining or electric power, side by side with the most primitive techniques; a few high class shops, surrounded by masses of old style traders; a few highly capitalized

plantations, surrounded by a sea of peasants." In more recent manifestations of this view, McKinsey Global Institute (2001) argues that the most productive firms in low-income countries are as productive as the firms in high-income countries, but the vast majority of firms in low-income countries are low-productivity ones. Bloom and Van Reenan (2007) provide similar evidence, focusing on the distribution of the quality of management.

The theories behind this view often emphasize that in many low-income countries, medium and large firms face fixed costs or constraints that small firms do not face. While some large firms are able to spread such fixed costs over a large enough volume of sales, or perhaps to offset regulatory costs with counterbalancing government subsidies, medium-sized firms lack such abilities and face differential barriers. For example, Banerjee and Duflo (2005, 2011) argue that the marginal return from increasing scale is low for firms using "primitive" technologies but is high in firms using "modern" technologies. In their model, this situation arises because the fixed cost of modern technologies is prohibitively high in poor countries. Therefore only a small number of firms adopt such technologies.

The classic paper by Harris and Todaro (1970) was the first to model the dual economy view that large firms are subject to constraints and regulations that small firms are able to evade. Their model posits a "modern" sector that pays above-market wages and a "traditional" sector that pays market wages. Rauch (1991) formally shows how this mechanism can generate a "missing middle" by assuming a fixed threshold due to minimum wage laws or labor unions above which firms have to pay above-market wages. Krueger (2013), McKinsey & Co. (2005), and Levy (2008) are recent versions of the same idea, where large firms pay taxes and are subject to regulations (in India, Brazil, and Mexico, respectively) that smaller firms can evade.

The dual economy model generates specific empirical predictions, too. First, if some firms are capital-constrained while others are not, then when one plots the distribution of returns to capital, there should be a "barbell" shape with one group of firms showing high returns to capital and the other showing lower returns. Second, the question of whether one should expect the returns to inputs (capital or labor) to be higher in larger or in smaller firms depends on the type of production function assumed. But if it is the large firms that are constrained in a dual-economy-type model, then under reasonable assumptions (spelled out below), one would expect to find a higher rate of return to capital in those firms. Third, if the missing middle is due to the fixed threshold above which firms face higher taxes or are subject to onerous regulations but these taxes or regulations are imperfectly enforced (say because of an inefficient bureaucracy), the outcome will be a right-skewed firm size distribution instead of a bimodal distribution. The "constrained large firm" model also predicts that the marginal return to resources will be *lower* in small firms compared to large firms.

With this theoretical framework in mind, the next three sections discuss facts about the distribution of firm size, the patterns of returns to capital and labor inputs, and whether one observes a "kink" in the distribution of firms at points where one might expect to find such a kink based on prominent regulations affecting large firms.

The Size Distribution of Firms: India, Indonesia, Mexico

Much of the work on the size distribution of firms in low-income countries looks at partial datasets—in particular, data that lacks much or any coverage of the informal sector. But we were able to obtain complete, representative microdata on the entire manufacturing sector for India, Indonesia, and Mexico, including both formal and informal enterprises. We use microdata from the manufacturing sector in the Mexican Economic Census, the Indonesian Economic Census, and India's Annual Survey of Industries and National Sample Survey (Schedule 2). For each country, we present the most recent wave of data available.[1] The key variable we use is the number of workers, including unpaid family workers. These countries do differ substantially in terms of GDP per capita in the year our data was collected: specifically, real per capita GDP in purchasing power parity terms was $3,700 in India (2011), $3,600 in Indonesia (2006), and $14,200 (2008) in Mexico (World Bank World Development Indicators 2013).

Figure 1 presents the distribution of firm size in bins of ten workers. The first row presents the distribution for India (2011), the second row for Indonesia (2006), and the third row for Mexico (2008). The first column presents the size distribution of *all* firms (since the tails have so few firms, we truncate the range of the graph at size 200 to make it visible). The next columns focus on different ranges of the data so that the patterns are more easily visible. Specifically, we restrict the range to firms with 10 to 200 workers (column 2), 20 to 200 workers (column 3), 50 to 200 workers (column 4), and 200 to 3,000 workers (column 5). The figure shows that the vast majority of firms in all three countries are small, with no evidence of bimodality in the firm size distribution. In all cases, the distribution of firm size is right skewed and generally smoothly declining in firm size, with no evidence of bimodality or discontinuity. This is the first key fact: there is no evidence of a "missing middle" of firms when one examines the raw distributions of firm size in any of these three countries.

Comparing the three countries, the fraction of small firms is lower in Mexico than in India and Indonesia. About 90 percent of firms in Mexico employ less than 10 workers. In India and Indonesia, the fraction of firms with less than 10 workers is almost visually indistinguishable from 100 percent. Given that the GDP per capita in Mexico is about four times higher than in India and Indonesia, the pattern suggests that development is associated with a decline in the skew of the firm size distribution.

For comparison, the US distribution of manufacturing firms has as its mode mid-sized firms with about 45 employees (Hsieh and Klenow 2014, Figure 14), whereas the mode in each of these countries are firms with one worker. There are

---

[1] The Mexican Economic Census is a complete enumeration of fixed establishments. The Indonesian Economics Census is a complete enumeration of all establishments with 20 or more employees (medium and large firms) and a random 5 percent sample of establishments with 20 or fewer employees (small firms). We combine these two samples to get a complete picture of the entire Indonesian manufacturing sector, including both formal and informal enterprises. The Indian Annual Survey of Industries is a census of formal establishments with more than 100 employees and a random survey of formal establishments with less than 100 employees. The National Sample Survey is a survey of informal establishments. We combine the data from the two surveys when we present evidence from India.

Figure 1
Distribution of Firm Size as Measured by Number of Workers
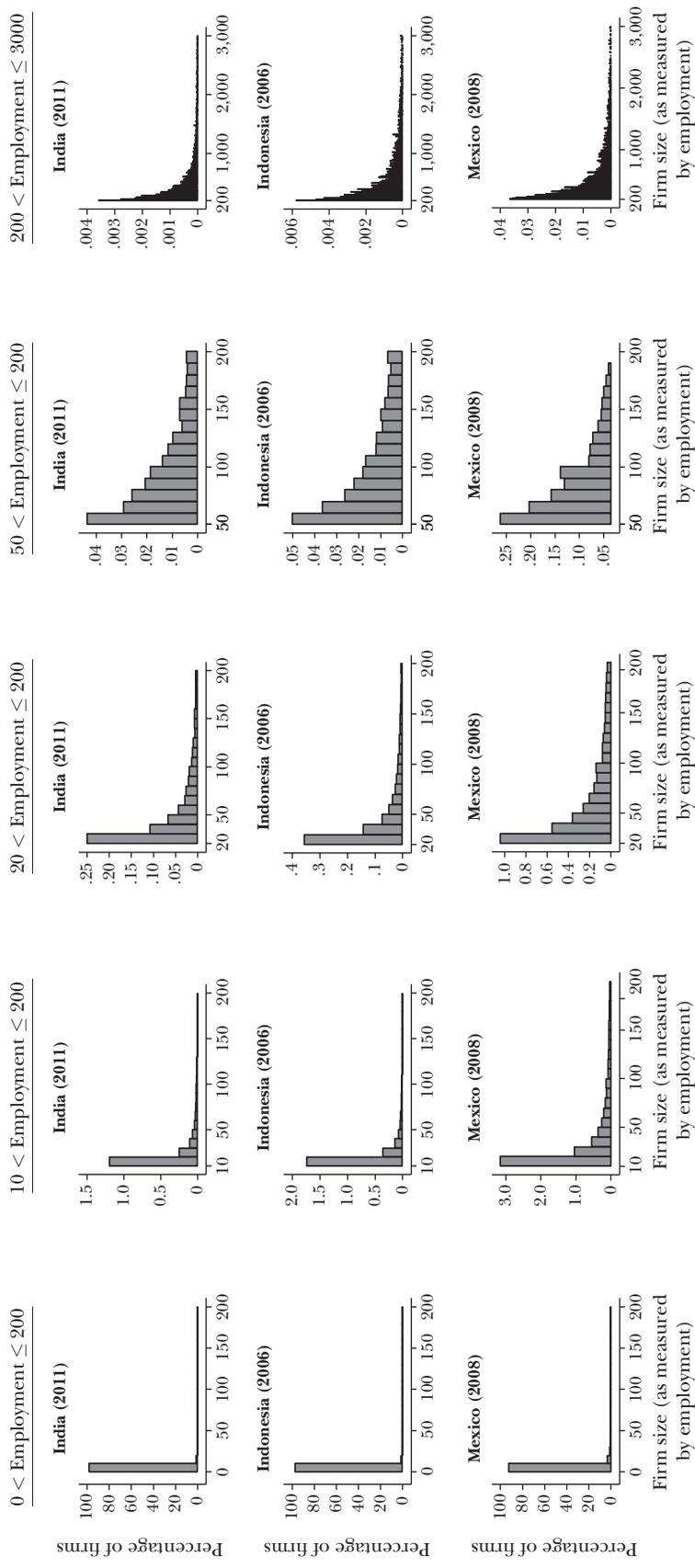
Source: We use microdata from the manufacturing sector in the Mexican Economic Census, the Indonesian Economic Census, and India's Annual Survey of Industries and National Sample Survey (Schedule 2). See footnote 1.

Notes: The figure shows distribution of firm size measured by the number of workers. The bin size is 10 workers, and each bin contains the upper bound and not the lower bound. The different columns truncate the x-axis in different ways to focus on different parts of the distribution. For all graphs, the y-axis indicates the share of all firms in the specified size.

fewer mid-sized firms in India, Indonesia, and Mexico than in the United States. But the overwhelming fact is that most firms are small in our three developing countries—large firms are also missing, and there is no missing middle in the sense of a bimodal distribution.

The Distribution of Average Return to Inputs

Even though the evidence shows no evidence of bimodality in the firm size distribution, another approach is to look also for evidence for the supposed forces that would lie behind the purported bimodality in the firm size distribution. For example, models where capital constraints generate a bimodal size distribution also imply that the return to capital is bimodal: that is, small unconstrained firms and large unconstrained firms would have low returns to capital, but firms that are hitting the constraint—the firms that would have grown to be the allegedly "missing" mid-sized firms—would have much higher returns. Other theories, such as those based on the idea that large firms face higher labor costs, those based on the notion that large firms have better access to intermediate inputs, and those based on De Soto's (1989) hypothesis that the property rights of formal firms are better protected, similarly imply that the return to all the resources used by the firm is bimodal, with one set of unconstrained firms with low returns and another set of constrained firms with high returns.

We do not directly measure the marginal return to inputs, but we can measure the average return to capital, labor, and intermediate inputs. However, interpreting the findings of such an exercise requires some caution, because different production functions have different implications for the relationship between average and marginal products. Here are several possibilities.

First, if revenue is generated by a single Cobb–Douglas function of the factor inputs, factor-intensities and markups are constant, and fixed costs are zero, the marginal return of each input is proportional to its average product.

Second, in a dual technology model with high-capital intensity and low-capital intensity technologies and a Cobb–Douglas production function, the average product of capital will be generally lower, and the average product of labor higher, in firms that utilize more capital-intensive technologies. In addition, the average product of the sum of variable and fixed capital will be lower in firms with high fixed-cost technologies. In addition, if some of capital measured in the data includes the fixed cost of the modern technology, this would further lower the average product of capital (the sum of fixed and variable capital) in large firms.

Although a dual technology model with Cobb–Douglas production technologies predicts that the average product of capital is lower for large modern firms, this prediction does not generalize to arbitrary production functions. For example, imagine that the production function for the two technologies is Leontief (so the marginal product of capital and labor for a given technology is zero), and the average product of capital and labor with the modern technology is higher than in firms using traditional technologies. Here, although the average product of capital

and labor in the traditional firm is low, the marginal return from switching to the modern technology is presumably high.

Third, if markups vary across firms—say, because more-productive firms produce higher-quality products that are more price inelastic—then the average product of capital and labor will be higher in larger firms that produce high-quality products.

Figure 2 looks for evidence of bimodality in the average product of factor inputs. Specifically, it plots the distribution of the log ratios of value-added to capital (column 1), value-added to labor (column 2), and gross output to the value of intermediate inputs (column 3). We truncate the top and bottom percentile to make the histograms more easily viewable. Note that we do not have a comparable capital series for large and small Indonesian firms, so we omit Indonesia from the first column. The distributions of the average product of capital and labor are not bimodal, as suggested by theories of capital constraints or labor costs. The distributions of the average product of intermediate inputs are also not bimodal, but are roughly right skewed. This pattern is consistent with theories where a large number of firms use few intermediate inputs.

Figure 3 looks directly at the correlation between the average product of inputs and firm size (measured as log employment). The first column presents the nonparametric relationship (from a regression using the approach of Fan 1992) with the average product of capital as the independent variable and firm employment as the explanatory variable. The dashed lines in each figure represent 95 percent confidence intervals.

As can be seen, the average product of capital is increasing with firm employment. If the average product of capital is proportional to the marginal product of capital, this suggests that the marginal cost of capital is *higher* in large firms relative to small firms. This fact is inconsistent with a widely held view that the return to capital is high in small firms in poor countries. Put differently, if the return to capital is high in small firms, the evidence in Figure 3 suggests that the return to capital in large firms is even higher.

This fact would be surprising if one believed in the dual technology view that large firms operate capital-intensive technologies with high fixed costs. For the large firms to have higher average products of capital in this story, it would either need to be that the modern firms have high average products of capital but low marginal products (so they use an L-shaped Leontief production function or close to it), or that modern firms also face higher marginal capital costs and the net effect of the higher marginal cost of capital outweighs the effect of capital-intensive technologies and the higher fixed cost on the average product of capital. Neither of these stories is theoretically impossible, but they are not necessarily what one would have expected from most standard versions of these theories.

The second column in Figure 3 plots the nonparametric relationship between the average product of labor with firm employment. The relationship is positive, as if the marginal cost of labor inputs is increasing with firm employment. This is a prediction of the Banerjee–Duflo (2005, 2011) dual technology model if modern technologies are more capital-intensive, although we note that this model is not supported by the evidence that the average product of capital is also higher in
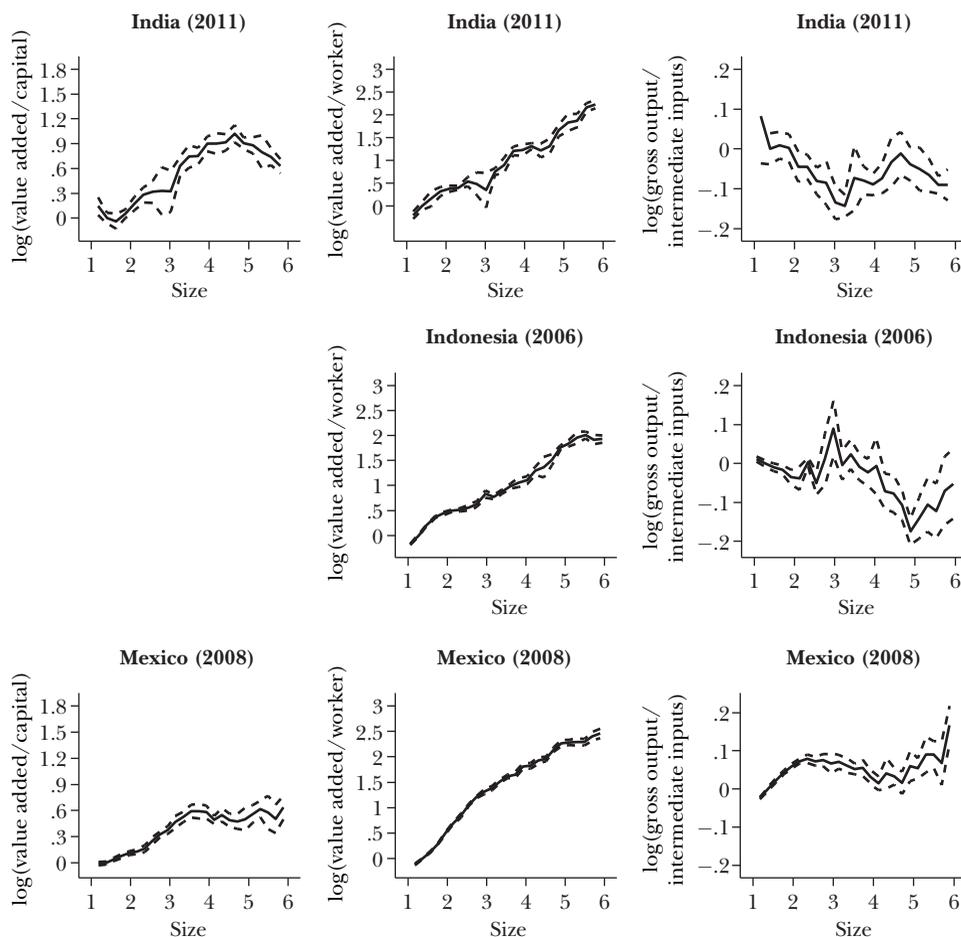
Figure 2
Distribution of Average Products
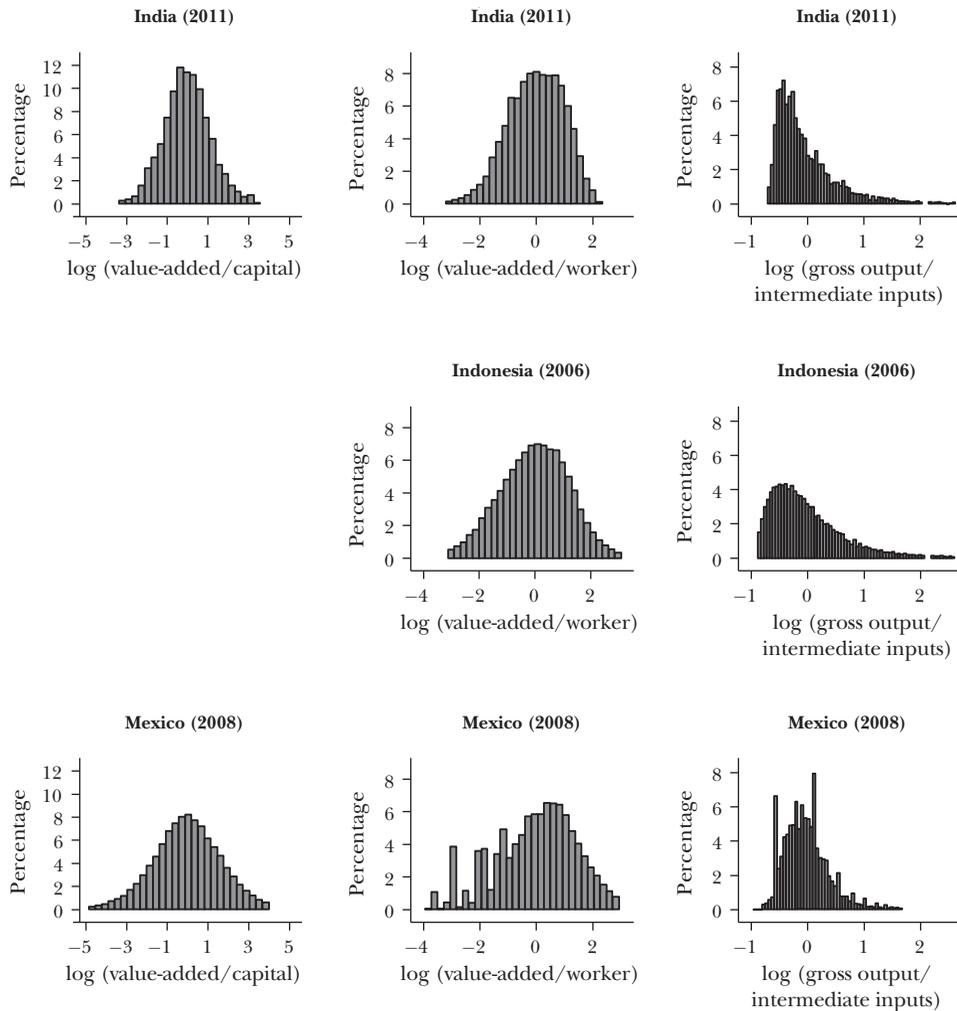
Source: See Figure 1 for sources.
Notes: The figure presents distributions of the demeaned log average product of capital (column 1), log average product of labor (column 2), and the log ratio of revenues to intermediate inputs (column 3). The bin size is the same in each column and chosen such that the histograms for Mexico have 50 bins. We drop the bottom and top 1 percent in each sample. In Indonesia, the questionnaire administered to firms with fewer than 20 employees asked about capital differently than the questionnaire administered to firms with 20 or more employees so we cannot construct a consistent measure of the capital stock across these two samples. In an online Appendix available with this paper at http://e-jep.org, we show qualitatively similar patterns when we separately examine firms with 20 or more employees and firms with less than 20 employees.

Average Product and Firm Size
(size measured as log(employment))

Source: See Figure 1 for sources.
Notes: Figure shows local linear regressions of log average product on log employment. We normalize the y-axis by taking the value of the function at log(employment) = 1.4 to be zero. Dashed lines represent 95 percent confidence bounds. Size is measured as log (employment).

larger firms. The fact that the average product of labor is higher in larger firms also supports the story by Harris and Todaro (1970), McKinsey & Co. (2005), and Levy (2008) that large firms pay above-market labor costs, except that there is no clear discontinuity in this relationship. We note that La Porta and Shleifer (2008, tables X, XI) also find that average labor productivity increases with firm size in the World Bank Enterprise Surveys, except that we interpret the positive relationship as indicating that large firms behave as if they face higher marginal labor costs.

An alternative explanation for why the average product of labor and capital might be higher in large firms is that larger firms charge higher markups. De Loecker

and Warzynski (2012) show that in this case, the markup will be proportional to the ratio of gross output to spending on intermediate inputs. The third column in Figure 3 shows the relationship between revenue per intermediate input and firm size. Here, there is no evidence that the average product of intermediate inputs is higher or lower in large firms relative to small firms. If the marginal cost of intermediate inputs is the same for small versus large firms, Figure 3 indicates that markups are no higher in large firms. In turn, this suggests that the higher average product of capital and labor for large firms do not reflect higher markups, but rather higher marginal costs. Of course, it is possible that large firms charge higher markups but the effect of the higher markup on the average product of intermediate inputs is exactly offset by a lower marginal cost of material inputs.

Together, Figure 2 and Figure 3 produce our second set of stylized facts: the average product of labor and capital is lower in small firms than in large firms, and there is no obvious bimodality in any of these distributions. In this sense, these developing countries look much like the United States, where it also the case that larger firms appear to be more productive (Brown and Medoff 1989; Idson and Oi 1999; Hsieh and Klenow 2014). But the fact that the larger firms are more productive is at odds with the frequent view in the development space that the key constraints are with small firms. If we believed there was a "missing middle" of constrained firms with high returns that could not grow, many models would have predicted either small firms to have higher average products, or potentially an inverted U-shape, with a mass of high average product and constrained firms in the middle of the distribution. However, the data does not support this view.

## Discontinuities in Firm Size from Tax and Regulatory Notches

A frequently cited reason for the existence of the purported missing middle is the existence of a tax or regulatory notch that affects firms above a certain size. In this situation, one might expect to find a bunching of firms at the size determined by the regulation or tax, and a missing distribution of firms just above the kink point. There are many possible examples: firms with few employees are frequently exempt from labor regulations (such as benefits and hiring and firing costs), and there is often preferential tax treatment for firms below a certain size threshold.

In our setting, we examine three such notches. In India, the Industrial Disputes Act requires firms employing more than 100 workmen (that is, 100 workers other than managers) to obtain government permission before laying off workers. This suggests a discrete notch in labor regulation at 100 nonmanagerial employees, which some have suggested is an important reason for the small size of firms in India (for example, *The Economist* 2007; Krueger 2013).[2]

In Indonesia, firms below a given revenue threshold are exempt from paying the 10 percent value-added tax. This again creates a discrete notch where we would

---

[2] Besley and Burgess (2004) offer an empirical assessment of the importance of Indian employment law more generally, and Guner, Ventura, and Xu (2007) present a quantitative model.

Figure 4
Distribution of Indian Firm Size and Labor Regulations
(size as measured by employment)
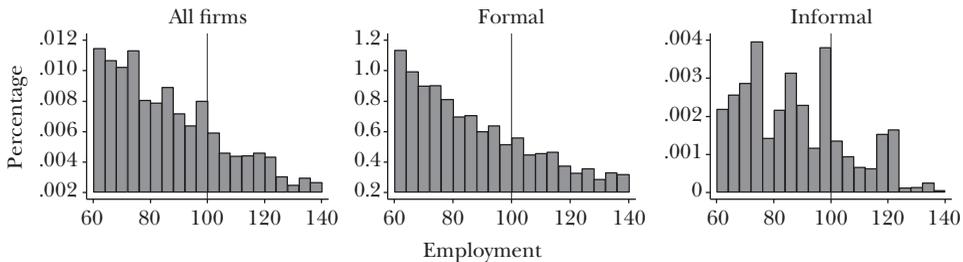
Source: The data comes from India's Annual Survey of Industries and National Sample Survey.
Notes: Figure shows size distribution of Indian firms around firms with 100 workers. We exclude managerial workers in the sample of formal firms from the Annual Survey of Industries. The bin size is four workers, and each bin contains the upper and not the lower bound.

expect bunching of firms below this cutoff (Kleven and Waseem 2013). The cutoff is not indexed for inflation; instead, it is adjusted discretely by the government periodically. Adjustments were made in 1992 (50 percent nominal increase); 1995 (100 percent nominal increase); 2001 (50 percent nominal increase); and 2004 (66 percent nominal increase). In 2006, the year of our census, the threshold was still where it was in 2004, at 600 million Indonesian rupiah (about $65,000 in US dollars); it was not raised again until 2013.

In Mexico, we focus on the revenue threshold due to the simplified tax regime for small firms. From 1998 until 2013, firms with sales below 2 million pesos (about $125,000 in 2008) pay a flat tax of about 2 percent of their sales and are exempt from payroll taxes, income taxes, and value-added taxes. Firms above the 2 million peso threshold are subject to a 15 percent value-added tax, a 38 percent income tax, and a 35 percent payroll tax.[3]

Although a casual examination of the histograms in Figure 1 does not suggest any discontinuities, it is possible that if we zoom in on these kinds of notches in the regulator environment we will see something. Figure 4 shows the distribution of nonmanagerial employment in India in 2011. (For nonformal firms, we do not have employment separately by managerial and nonmanagerial categories, so we report total employment for these firms.) We zoom in on the range from 60 to 140 nonmanagerial employees so we can focus on the 100 worker cutoff (shown by the vertical line). We focus on the distribution of all firms (left panel) but also show the distribution of formal firms (center panel) and informal firms (right panel). Since the regulation applies only to formal firms, it is possible that even if the regulation doesn't affect the total firm size distribution, it affects the decision to switch from formal to informal.

---

[3] The tax rate under the simplified tax regime (Repecos) varies across states but averages 2 percent. The simplified tax regime is administered at the state level. See Sánchez-Vela and Valero-Gill (2011).

Visually inspecting the leftmost panel of Figure 4, there is perhaps a slight bit of bunching at 100 employees, but it is small. In the bin just below the cutoff (97–100 workers), there are a total of 1,370 firms. In the next bin (101–104 firms), there are 1,013 firms. Even abstracting from the fact that the overall distribution is downward sloping, so one would expect fewer firms with 101–104 workers than 97–100 workers, the difference amounts to at most two-tenths of 1 percent of all Indian firms—a few hundred firms in all of India out of the 17,177,148 total firms.

Inspecting the central panel, there is no discontinuity whatsoever in formal firms; if anything, there is a slight spike of firms with more than 100 workers. There *is* bunching of informal firms just below 100, but again the economic magnitude is small: the difference between the number of firms with 97–100 workers and 101–104 workers is a little more than two-tenths of 1 percent of all informal firms—at most about 418 firms in total for all of India. Thus, while there may be a small amount of bunching induced by the regulation, the amount we can detect in the data does not suggest that it is an important driver of small firm size in India.

For Indonesia, we focus on the discontinuity in revenue at 600 million Indonesian rupiah. One might expect more heaping in revenue than in employment, since presumably revenue is easier to adjust in order to stay under the threshold (for example, firms may have some flexibility in deciding what year to realize revenue from a given sale). The top left panel of Figure 5 shows the distribution of revenue for all Indonesian firms; the bottom panel zooms in on firms with more than 20 employees. Since virtually no firms with fewer than 20 employees have revenue close to 600 million rupiah, the figures in the bottom panel are easier to read. The two left panels show the distribution for all firms with less than 40 billion Indonesian rupiah in revenue (about $4.3 million in US dollars). To zoom in closer to the discontinuity, the right panels consider only firms with less than 1.8 billion in Indonesian rupiah in revenue (about $200,000 in US dollars).
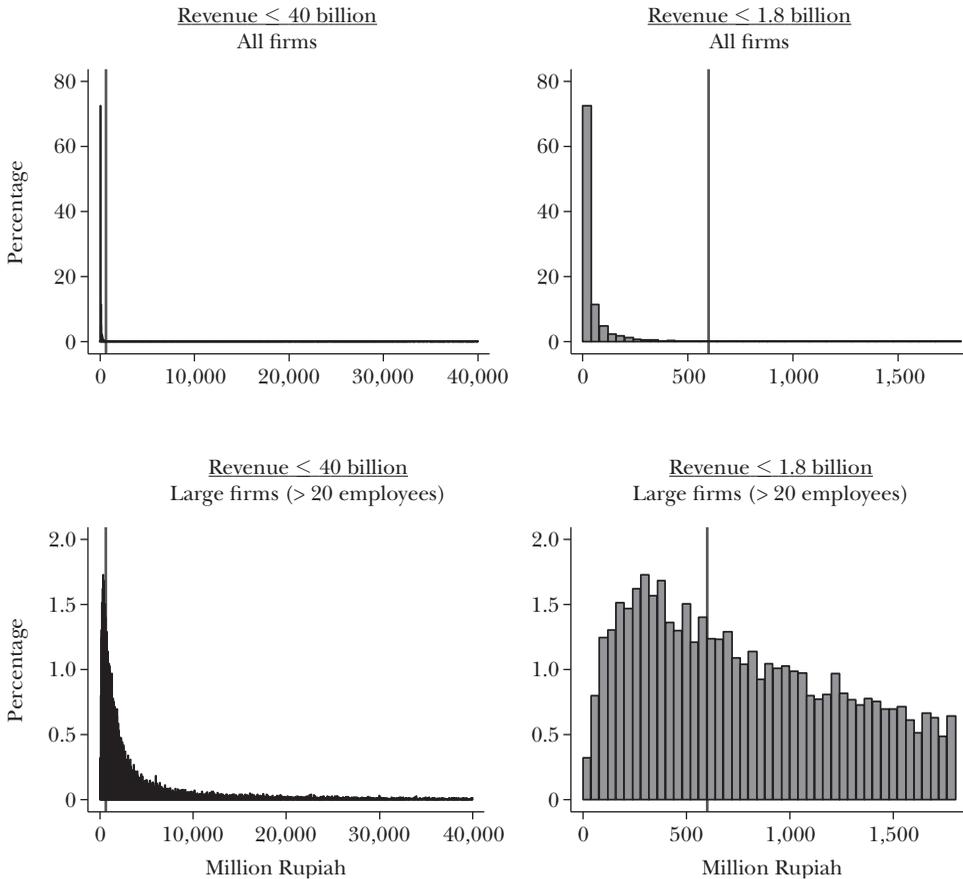
The figure (particularly the bottom-right), which zooms in on the relevant part of the cutoff shows no bunching at the discontinuity in eligibility to pay the value-added tax. Because virtually all firms in the relevant part of the revenue distribution are from the large firm survey, which is conducted annually, we can re-generate the zoomed-in graph for large firms for each year back to 1990 and check for any changes in firm size associated with the different cutoffs that were in place over the years. In an online Appendix available with this paper at http://e-jep.org, we show this figure, with the relevant cutoff line shown each year. We never find any substantial bunching at the discontinuities.

Next, we focus on a potential discontinuity in Mexico due to a simplified tax regime for firms with less revenue than 2 million pesos. The left panel in Figure 6 shows the distribution of sales for all Mexican firms with less than 6 million pesos in sales; the right panel zooms in on firms with sales between 1 and 4 million. As can be seen, there is no bunching at 2 million pesos (the vertical line) after which firms legally switch from a flat 2 percent sales tax regime to the combination of the value-added tax, income tax, and payroll tax regime.

Combined, the evidence from India, Mexico, and Indonesia suggest a third important fact: at least as we can measure it in our data, we do not see important

Figure 5
**Distribution of Indonesian Firm Size and the VAT Threshold**

Revenue ≤ 40 billion
All firms

Revenue ≤ 1.8 billion
All firms

Revenue ≤ 40 billion
Large firms (> 20 employees)

Revenue ≤ 1.8 billion
Large firms (> 20 employees)

Million Rupiah

Million Rupiah

Source: Indonesian Economic Census. See footnote 1.
Notes: Figure shows distribution of the revenue of Indonesian firms. The vertical line (600 million rupiah) denotes the VAT threshold. The bin size is 40 million rupiah, and each bin contains the upper bound but not the lower bound.

discontinuities in firm size, either in general when looking at the distributions or when we zoom in around the places where one would expect them a priori based on regulatory and tax notches.

Of the other papers in the literature that have looked at similar notches, a small number of papers have found some bunching, but in most of these cases the quantitative magnitude of the bunching is small. For example, Onji (2009) examines the introduction of a value-added tax threshold in Japan and looks for bunching around the threshold, much as we do in Indonesia. Although he does find evidence of bunching, the magnitude appears very small: the share of firms below the threshold falls by less than 0.5 percent. Similarly, Schivardi and Torrini (2008) examine a discontinuity in Italian employment regulations that applies to

Figure 6
Distribution of Mexican Firm Size and the Simplified Tax Regime Threshold
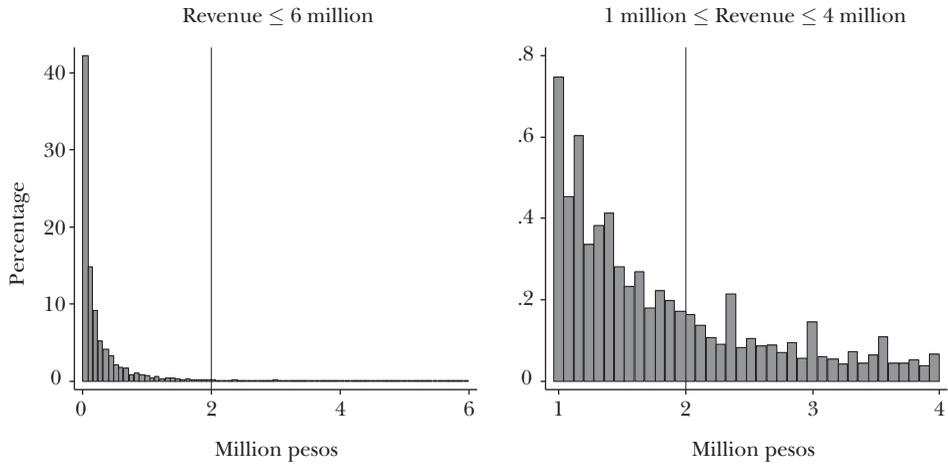
Source: We use microdata from the manufacturing sector in the Mexican Economic Census. See footnote 1.
Notes: Figure shows the distribution of revenues of Mexican firms. The vertical line (2 million pesos) denotes the threshold of a simplified tax regime for small firms. The bin size is 80,000 pesos, and each bin contains the upper bound and not the lower bound.

firms greater than 15, much as we do in India. They estimate that after removing the threshold, average firm size would increase by less than 1 percent. Garicano, LeLarge, and Van Reenen (2013) estimate the impact of lifting French regulations that apply to firms with more than 50 workers. Their model implies that about 3 percent of workers are reallocated from firms of size 50 or more to firms of size 49 and below. Under the assumption of flexible wages, their model estimates an output loss of 0.16 percent of GDP associated with this change, although the assumption of fully inflexible wages yields substantially larger estimates. Thus, the evidence we present from India, Indonesia, and Mexico is consistent with the small magnitudes of bunching observed in other contexts.

How Did the "Missing Middle" Misconception Arise?

Given the facts presented in this paper, a natural question is: Where does the misconception about the missing middle—in the sense of the bimodality of the distribution—come from? We suggest it comes from the combination of two transformations that had previously been made to the available data.

In the economics literature, the main evidence typically cited for the missing middle is table 1 of Tybout (2000). In that table, Tybout shows the distribution of employment shares across plant sizes for manufacturing firms for 19 countries. For most countries in the table, he shows the number of workers in firms of size 1–9,

10–49, and 50+; for a few countries, he includes five or six bins of firms instead. The data in the table is, in turn, drawn from other calculations done by a variety of other authors, most notably Leidholm and Mead (1987), who compile similar tabulations from other studies. The "missing middle" refers to the fact that in most developing countries, there is substantially lower employment share in the mid-sized category (that is, firms of 10–49 employees) than in either the small category (fewer than 10 employees) or the large category (50 or more employees). For example, in Indonesia in 1977, the table shows 77 percent of total manufacturing employment is in firms of size 1–9, 7 percent is in firms of size 10–49, and 16 percent is in firms of size 50 or more.

There are two important differences between the facts reported in the Tybout (2000) and Liedholm and Mead (1987) tables and the facts we present here. First, these earlier tables refer to the employment share—that is, what fraction of total manufacturing employment comes from firms of a given size—rather than the distribution of firm size. The employment share distribution reveals in what size firm a typical worker in the economy works, whereas the firm size distribution reveals the distribution of firms. To compute the employment share statistic, one multiplies the number of firms in each bin with the average employment size of firms in the bin. While the employment share statistic is interesting for understanding the aggregate distribution of employment, most theories about the existence of the missing middle discussed above are about firm size itself. For example, theories about tax and regulatory notches and credits constraints are all about whether firms should grow above a certain size, not about the employment share in aggregate.

We start with our data from Figure 1 about the distribution of firm size, and transform it into one on the distribution of employment share by firm size, as shown in Figure 7. This transformation, in itself, does not create a missing middle. Figure 7 shows the distribution of employment share, analogous to what is shown in Figure 1 for the distribution of firm size. Although it is shifted to the right (mechanically) from the firm size distribution, it still appears unimodal in all three countries.

Second, in the earlier literature, because of data limitations, the tabulations about employment share are typically binned into a small number of groups: for most countries, the authors report the totals for three bins, firms with less than 10 employees, 10–49 employees, and 50 or more employees.

To see what difference this choice of bins makes, Table 1 reports the distribution of firm size (Panel A) and the distribution of employment shares (Panel B) from our data, grouped into these same three categories. Panel A shows that the firm size distribution, even when binned, shows the same pattern as the histograms—the density of firms is monotonically declining in firm size. But Panel B shows that when we apply the arbitrary binning transformation to the *employment share* distribution, the pattern from Tybout (2000) re-emerges. For example, in Indonesia in 2006, 54 percent of total employment is in firms with 1–9 employees, 12 percent is in firms with 10–49 employees, and 34 percent is in firms with 50 or more employees—that is, the missing middle phenomenon now appears. Thus, the existing facts about the missing middle seem to come from the combination of these two transformations to

Figure 7
Distribution of Employment Share by Firm Size
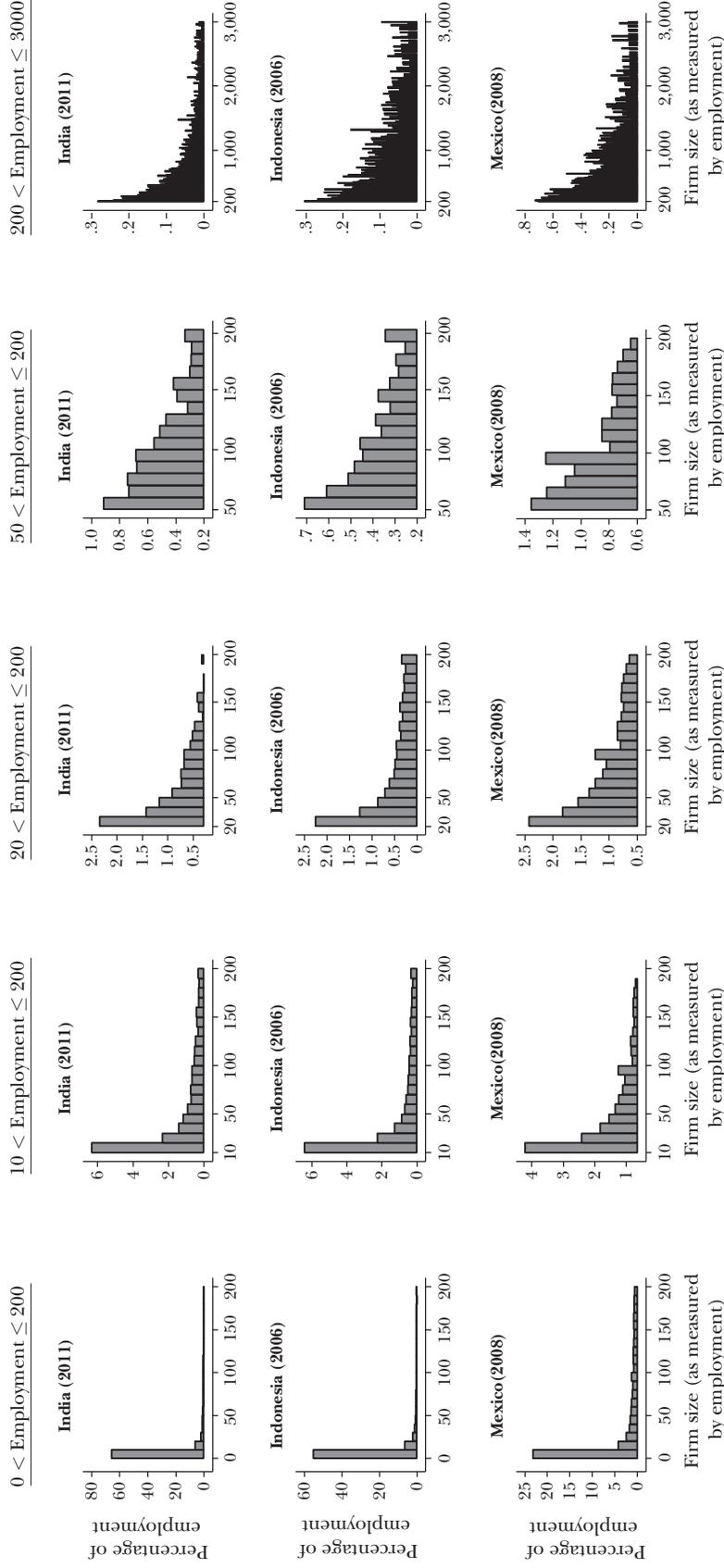
Source: See Figure 1 for sources.

Notes: The figure shows the distribution of employment share across firms of different size as measured by number of employees. The bin size is 10 workers, and each bin contains the upper bound and not the lower bound. The different columns truncate the x-axis in different ways to focus on different parts of the distribution.

Table 1
Distribution of Firms and Employment Shares in Bins

Firm Size (Employment)	India 2011	Indonesia 2006	Mexico 2008
Panel A: Distribution of Firm Size			
1–9	97.88	96.78	91.74
10–49	1.85	2.83	5.85
50+	0.28	0.39	2.41
Panel B: Distribution of Employment Share by Firm Size			
1–9	64.77	53.95	22.45
10–49	12.10	12.04	10.55
50+	23.13	34.01	66.99

Source: See Figure 1 for sources.

the data: the transformation from the distribution of firms to the aggregate employment share, and the arbitrary binning of the employment share distribution.

Implications for Theories of Development

Ultimately, the main reason that economists and policymakers care about the size distribution of firms in developing countries is what it may reveal about alternative theories of firm development and in turn what that implies for policy. The lack of a "missing middle"—that is, the lack of bimodality in the size distribution of firms—suggests that neither the "small firms are constrained" nor the dual economy theories of development are correct, at least not in their simplest form. In addition, the fact that the average returns to capital and labor are lower in small firms suggests that the view that small firms are constrained—say because they have difficulty accessing capital and thus have a high return to capital—is inconsistent with the simple versions of these models.

What would it take to reconcile the models to the facts? One tempting alternative is to explore the implications of more capital-intensive production technologies for larger firms: after all, it is likely that large firms use more capital-intensive technologies that, all else equal, would tend to lower the average product of capital in large firms. To make a dual-economy model fit the facts without also asserting that large firms are constrained, one would need the high-productivity firms to have high average products of capital but low marginal products of capital, and vice-versa. Moreover, one would need substantial heterogeneity across firms in the employment size of such high capital-intensity firms in order to avoid generating bimodality in the firm size distribution. It is theoretically possible to write down such models, but the facts presented here substantially constrain the types of models one can write down.

An alternative theory that fits all our facts is the view that large firms are constrained, perhaps by taxes or regulations, but that implementation of these

barriers is imperfect. Levy (2008), for example, documents that the vast majority of small and mid-size firms in Mexico evade the 35 percent payroll tax. This view is consistent with the evidence that there is little meaningful discontinuity in the size distribution, even at thresholds at which one would expect a discontinuity if taxes or regulations were perfectly enforced. This view also implies that the problem is unlikely to be the (relatively easy to fix) notch in the tax or regulatory code; rather, it suggests that it a confluence of factors make enforcement of such rules easier in larger firms so that costs from regulation are rising smoothly in firm size. Another key prediction of the "large firms are constrained" view is that the marginal return to resources would be higher in large firms, which is supported by the fact that the average product of capital and labor is consistently higher in large firms when compared to small firms. If so, the fact that the firm size distribution in poor countries is dominated by small firms is explained by firms *choosing* not to exert the effort necessary to grow because their marginal cost would rise if they did grow.

In sum, the evidence we present in this paper suggests that the problem of economic development in low-income and middle-income countries is how to relieve the differential constraints faced by large firms, not how to relax the constraints faced by small firms. Indeed, this view suggests that programs such as microcredit or simplified tax regimes that benefit only small firms may worsen the development problem by further increasing the incentive to stay small.

References

Banerjee, Abhijit V., and Esther Duflo. 2005. "Growth Theory through the Lens of Development Economics." Chap. 7 in *Handbook of Economic Growth*, volume 1, edited by Philippe Aghion and Steven Durlauf. Elsevier.

Banerjee, Abhijit V., and Esther Duflo. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. Public Affairs.

Besley, Timothy, and Robin Burgess. 2004. "Can Labor Regulations Hinder Economic Performance? Evidence from India." *Quarterly Journal of Economics* 119(1): 91–134.

Bloom, Nicholas, and John Van Reenan. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122(4): 1351–1408.

Brown, Charles, and James L. Medoff. 1989. "The Employer Size-Wage Effect." NBER Working Paper 2870.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-level Export Status." *American Economic Review* 102(6): 2437–71.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." *Quarterly Journal of Economics* 123(4): 1329–72.

De Soto, Hernando. 1989. *The Other Path: The*

Invisible Revolution in the Third World. New York, NY: Harper and Row.

Economist, The. 2007. "India's Economy: A Himalayan Challenge." October 11. http://www.economist.com/node/9955756.

Fan, Jianqing. 1992. "Design-adaptive Nonparametric Regression." *Journal of the American Statistical Association* 87(420): 998–1004.

Garicano, Luis, Claire LeLarge, and John Van Reenen. 2013. "Firm Size Distortions and the Productivity Distribution: Evidence from France." NBER Working Paper 18841.

Guner, Nezih, Gustavo Ventura, Xu Yi. 2008. "Macroeconomic Implications of Size-Dependent Policies." *Review of Economic Dynamics* 11(4): 721–44.

Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis." *American Economic Review* 60(1): 126–42.

Hsieh, Chang-Tai, and Peter Klenow. 2014. "The Lifecycle of Manufacturing Plants in India and Mexico." http://faculty.chicagobooth.edu/chang-tai.hsieh/research/HsiehKlenow_LifeCycle.pdf. (Forthcoming in *The Quarterly Journal of Economics*).

Idson, Todd L., and Walter Y. Oi. 1999. "Workers Are More Productive in Large Firms." *American Economic Review* 89(2): 104–108.

Kleven, Henrik J., and Mazhar Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics* 128(2): 669–723.

Kremer, Michael, Jean Lee, Jonathan Robinson, and Olga Rostapshova. 2013. "Behavioral Biases and Firm Behavior: Evidence from Kenyan Retail Shops." *American Economic Review* 103(3): 362–68.

Krueger, Anne O. 2013. "The Missing Middle." Chap. 9 in *Economic Reform in India: Challenges, Prospects, and Lessons* edited by Nicholas C. Hope, Anjini Kochar, Roger Noll, and T. N. Srinivasan, Cambridge University Press.

La Porta, Rafael, and Andrei Shleifer. 2008. "The Unofficial Economy and Economic Development." *Brookings Papers on Economic Activity*, no. 2 (Fall), pp. 275–352.

Leidholm, Carl, and Donald C. Mead. 1987. "Small Scale Industries in Developing Countries: Empirical Evidence and Policy Implications." Food Security International Development Papers 54062, Michigan State University.

Levy, Santiago. 2008. *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico*. Washington DC: Brookings Institution.

Lewis, Arthur. 1954. "Economic Development with Unlimited Supplies of Labour." *Manchester School* 22(2): 139–91.

McKinsey & Co. 2005. "Eliminando as barreiras ao crescimento econômico e à economia informal no Brasil."

McKinsey Global Institute. 2001. *India: The Growth Imperative*.

Onji, Kazuki. 2009. "The Response of Firms to Eligibility Thresholds: Evidence from the Japanese Value-Added Tax." *Journal of Public Economics* 93(5–6): 766–75.

Rauch, James E. 1991. "Modeling the Informal Sector Formally." *Journal of Development Economics* 35(1): 33–47.

Sánchez-Vela, Claudia, and Jorge Valero-Gil. 2011. "The Effect of Firm-Size Dependent Policies on the Economy: The Case of the Repecos Law in Mexico." IDB Working Paper 216.

Schivardi, Fabiano, and Roberto Torrini. 2008. "Identifying the Effects of Firing Restrictions through Size-Contingent Differences in Regulation." *Labour Economics* 15(3): 482–511.

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

Udry Christopher, and Santosh Anagol. 2006. "The Return to Capital in Ghana." *American Economic Review* 96(2): 388–93.

This article has been cited by:

1. Luis Garicano, Claire Lelarge, John Van Reenen. 2016. Firm Size Distortions and the Productivity Distribution: Evidence from France. *American Economic Review* **106**:11, 3439-3479. [Abstract] [View PDF article] [PDF with links]
2. Hunt Allcott, Allan Collard-Wexler, Stephen D. O'Connell. 2016. How Do Electricity Shortages Affect Industry? Evidence from India. *American Economic Review* **106**:3, 587-624. [Abstract] [View PDF article] [PDF with links]