NBER WORKING PAPER SERIES

THE MISSING "MISSING MIDDLE"

Chang-Tai Hsieh
Benjamin A. Olken

Working Paper 19966
http://www.nber.org/papers/w19966

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2014

Prepared for the Journal of Economic Perspectives. We thank Arianna Ornaghi, Nick Tsivanidis, Pedro-José Martínez-Alanis and especially Donghee Jo for outstanding research assistance and Abhijit Banerjee, Esther Duflo, Santiago Levy, Pete Klenow, and John Van Reenan for helpful comments. The results from the Mexican Census have been screened by Mexico's INEGI to ensure no confidential data is released. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

The Missing "Missing Middle"
Chang-Tai Hsieh and Benjamin A. Olken
NBER Working Paper No. 19966
March 2014
JEL No. E23,H25,O11,O47

ABSTRACT

Although a large literature seeks to explain the "missing middle" of mid-sized firms in developing countries, there is surprisingly little empirical backing for existence of the missing middle. Using microdata on the full distribution of both formal and informal sector manufacturing firms in India, Indonesia, and Mexico, we document three facts. First, while there are a very large number of small firms, there is no "missing middle" in the sense of a bimodal distribution: mid-sized firms are missing, but large firms are missing too, and the fraction of firms of a given size is smoothly declining in firm size. Second, we show that the distribution of average products of capital and labor is unimodal, and that large firms, not small firms, have higher average products. This is inconsistent with many models in which small firms with high returns are constrained from expanding. Third, we examine regulatory and tax notches in India, Indonesia, and Mexico of the sort often thought to discourage firm growth, and find no economically meaningful bunching of firms near the notch points. We show that existing beliefs about the missing middle are largely due to arbitrary transformations that were made to the data in previous studies.

Chang-Tai Hsieh
Booth School of Business
University of Chicago
5807 S Woodlawn Ave
Chicago, IL 60637
and NBER
chsieh@chicagoBooth.edu

Benjamin A. Olken
Department of Economics, E17-212
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
bolken@mit.edu

I. Introduction

The notion that the distribution of firm size in poor countries is characterized by a bimodal distribution with a "missing middle" is a widely accepted stylized fact in development. The missing middle "fact" is cited as evidence for two broad stories of development. Perhaps the most prominent one is that most firms in poor countries do not have the resources necessary to grow. For example, advocates of microfinance argue that small firms have difficulty accessing capital and thus do not grow to become medium sized firms.

A second story for the missing middle is the dual economy view. There are two versions of this story. One version is that large formal firms are subject to onerous regulations and higher input costs. For example, the classic work by Harris and Todaro (1970) posits that formal firms are forced to pay wages that are above the market clearing wage, or are subject to a variety of regulations that small firms are exempt from. A second version is that the fixed cost of becoming a large firm is high in poor countries. For example, Banerjee and Duflo (2005, 2011) argue that the fixed costs of adopting modern technologies are prohibitively high in poor countries.

What is missing in this literature is evidence on the distribution of firm size. An empirical challenge is that available microdata on firm surveys rarely include small informal firms, and it has previously been difficult to access microdata of complete economic censuses that capture all firms. For example, the widely used Chilean Manufacturing data surveys firms with more than ten employees, and the World Bank's Economic Surveys that are widely available for many countries only surveys formal firms. As a result, the little evidence we have on the distribution of all firms tends to be based on a few aggregate tabulations of the data, which we show can lead to misleading answers.

In this paper we present evidence on the firm size distribution from India, Indonesia, and Mexico where we have representative firm micro data on all manufacturing firms. Importantly, the datasets we use includes small informal firms that account for a large share of employment in these countries. We use the microdata on these firms, so can look at the complete distribution of firm size, rather than arbitrary bins of the data.

We present three facts. First, there is no evidence of a missing middle in India, Indonesia, or Mexico, regardless of how we slice the data. To be sure, there 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 ones, there is no missing middle in the sense of a bimodal distribution. The absence of a missing middle suggests that the two broad stories of development that cite the missing middle "fact" are not correct, at least without modifications.

Second, the average product of labor and capital is significantly lower in small firms when compared to larger firms. This suggests that, to the extent that the average and marginal products of capital move together, the view that small firms do not grow because they face high marginal costs of capital is not likely to be correct. However, this fact supports the dual economy view that large firms face constraints that raise the effective cost of their inputs (relative to the costs faced by small firms.) Put differently, it is not that small firms cannot grow because they faced large fixed costs or do not have access to credit. Instead, it is the large firms that are more stifled and cannot grow.

Third, there is no evidence of meaningful discontinuities anywhere in the firm size distribution. In particular, we focus on areas where we believe there to be thresholds in

enforcement – a size threshold of 100 employees in India where various labor regulations kick in, and a revenue threshold in Indonesia above which firms are required to pay VAT, and a revenue threshold in Mexico above which firms face higher tax rates – and find no economically meaningful bunching of firms around these thresholds.

The absence of bunching around these thresholds 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. For example, if firms above a certain size are legally obligated to pay high minimum wages or pay higher taxes, implementation of the minimum wage or enforcement of the tax law is imperfect or the burden they impose is not large enough to significantly constrain firm growth.

Where does this misconception about the missing middle come from? We suggest it comes from the combination of two transformations that had previously been made to the data. While there is much discussion of the missing middle, the main citation is a table on the distribution of employment shares in a number of countries in Tybout's (2000) well-known survey article, much but not all of which is drawn from Leidholm and Mead (1987), which in turn summarized a variety of other studies. Due to data limitations, these tabulations are for most countries 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 (e.g. firms over a size are differentially taxed, firms have trouble getting credit 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. That is, if we bin the firm size distribution into these three broad bins, it still 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 paper proceeds as follows. We first present the theories of development for which the missing middle is a central fact, and discuss the auxiliary implications of these theories. We then present facts from our data from India, Mexico, and Indonesia. We then discuss the source of the perception of the existence of the missing middle. The final section discusses the implications of our new facts for theories of economic development.

II. The Missing Middle and Development

There are at least two models of development where the missing middle is an important stylized fact. Perhaps the most prominent is the view that the institutional environment in poor countries discriminates against small firms and favors big firms. There are several versions of such models. The most common is based on credit constraints, where small firms are credit constrained and large firms are not. Closely related mechanisms are based on the idea that property rights are protected for formal firms but not for informal firms (De Soto, 1989) and that large firms have better access to intermediate inputs and output markets. Other models are based on the idea that government interventions benefit large firms. This can be because large firms

are state owned firms, or because industrial policy targets large firms, or because large firms are the beneficiaries of protectionism and entry barriers.

We note that in addition to the missing middle, 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. This is because there are many small firms that are constrained so they have high marginal products. For example, a number of papers show that the return to capital is very high in small firms in developing countries (de Mel, McKenzie, and Woodruff (2008), Anagol and Udry (2006), Kremer et. al (2013)). In addition, while there is significant evidence that there is more dispersion in the return to capital in developing countries, the question however is whether the return to capital for small firms is higher than the return to capital for large firms.

A second model of development that generates the missing middle is what we call the "dual economy" view. For example, McKinsey (2001) argues that the most productive firms in poor countries are as productive as the firms in rich countries but the vast majority of firms in poor countries are low productivity ones. Bloom and Van Reenan (2007) provide similar evidence, focusing on the distribution of the quality of management. This view harkens back to Arthur Lewis' (1954) 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."[1] Banerjee and Duflo (2005 and 2011) make this idea explicit. In their view, the marginal return from increasing scale is low for firms using "primitive" technologies and high in firms using "modern" technologies. However, because the fixed cost of modern technologies is prohibitively high in poor countries, only a small number of firms adopt such technologies. These firms are Arthur Lewis' "islands of capitalist employment" while the "vast sea of subsistence workers" are employed in firms utilizing low fixed cost primitive technologies.

Another version of the dual economy view is that large firms are subject to constraints and regulations that small firms are able to evade. The classic paper by Harris and Todaro (1970) was the first to make this idea explicit by positing 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 (2009), McKinsey (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 are able to evade.

We make following points about the empirical predictions of the dual economy model. First, in the dual technology version of the model, the missing middle is generated by the simultaneous presence of modern firms with high fixed costs and "primitive" firms with low fixed costs. Second, in models where the average product of capital and labor are proportional to their marginal products (such as Cobb-Douglas) and where the factor intensities are the same in the types of firms, the average product of capital and labor will also be the same in large vs. small firms. However, if high fixed cost technologies are also more capital intensive, then the average product of labor would be higher and the average product of capital would be ***lower*** in large firms (compared to small labor intensive firms). In addition, if some of capital measured in

[1] Lewis (1954) page 147.

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. 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.

In the "large firms are constrained" version of the model, the missing middle is due to the fixed threshold above which firms face higher taxes or subject to onerous regulations. However, if 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. This is true even if taxes and regulations are imperfectly enforced (and thus there is no missing middle).

III. Results

Data

Here we present evidence from the firm size distribution in India, and Indonesia, and Mexico. We use these three countries since we could obtain complete, representative microdata on the entire manufacturing sector for these countries, including both formal and informal enterprises. These countries differ substantially in terms of GDP per capita, with Mexican real GDP per capita about 4 times higher than India and Indonesia in the year our data was collected.[2]

We use micro-data 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). 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% 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. For each country, we present the most recent wave of data available. The key variable we use is the number of workers, including unpaid family workers.

The Firm Size Distribution

Figure 1 presents the distribution of firm size in bins of 10 workers. The first column presents the size distribution of ***all*** firms. The first row presents the distribution for India (2011), the second row for Indonesia (2006), and the third row for Mexico (2008). The figure shows that

[2] Real per capita GDP in PPP terms in the year of our data was $3,700 in India (2011), $3,600 in Indonesia (2006), and $14,200 (2008) (World Bank World Development Indicators 2013).

the vast majority of firms in all three countries are small, with no evidence of bimodality in the firm size distribution. The next columns focus on different samples of the data so that the patterns are more easily visible. Specifically, we restrict the sample to firms with 10 to 200 workers (column 2), 20 to 200 workers (column 3), 50 to 200 workers (column 4), and 200 to 3000 workers (column 5). 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 visually indistinguishable from 100 percent. GDP per capita in Mexico is about four times higher than in India and Indonesia. This suggests that development is associated with a decline in the right skew of size distribution, but not with a less bimodal firm size distribution.

It is worth noting that if one compares these countries to the US, there is a marked difference: the US distribution of manufacturing firms has as its mode mid-sized firms with about 45 employees (see Figure 14 in Hsieh and Klenow 2014), whereas the mode in each of these countries are firms with 1 worker. There are fewer mid-sized firms in India, Indonesia, and Mexico than in the US. 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 Products

Instead of looking at the firm size distribution, we can also look for evidence for bimodality in the forces that lie behind potential 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. For example, one might expect small unconstrained firms and large unconstrained firms to have low returns to capital, but firms that are hitting the constraint – the firms that would grow to be mid-sized firms but are "missing" – 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 property rights of formal firms are better protected, all imply that the return to all the resources used by the firm is bimodal, with one set of unconstrained firms with low returns and one 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. If revenue is generated by a 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. We note however that these assumptions do necessarily hold. In the dual technology model, the average product of capital will be 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. 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 ratio of value-added to capital (column 1), value-added to labor (column 2), and the ratio of gross output to the value of intermediate inputs (column 3).[3] We truncate the top and bottom percentile to make the histograms more easily viewable. The distribution of the average product of capital and labor is not bimodal as suggested by theories of capital constraints or labor costs. The distribution of the average product of intermediate inputs is also not bimodal, but is roughly right skewed. This 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. The first column presents the non-parametric relationship (from a Fan (1992) regression) between the average product of capital with firm employment. The dashed lines in each figure represent 95% 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, say because these firms have difficulty accessing capital. 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 high-fixed cost technologies. 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 are Leontief 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 are theoretically impossible, but they are not necessarily what one would have expected from the most standard versions of these theories.

The second column in Figure 3 plots the non-parametric 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 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 (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's (2008) also find that average labor productivity increases with firm size (Table X and XI in La Porta and Shleifer, 2008) 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.

As discussed earlier, 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

[3] 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 the appendix, we show qualitatively similar patterns when we separately examine firms with 20 or more employees and firms with less than 20 employees.

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 vs. large firms, Figure 3 indicates that markups are no higher in large firms. In turn, this suggests that that the higher average product of capital and labor for large firms do not reflect higher markups but 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 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. 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 that there is a tax or regulatory notch. If firms above a certain size are subject to regulations, that would create a bunching of firms at the notch point, and a missing distribution of firms above the kink point. There are many examples of this: 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.

Although a casual examination of the histograms in Figure 1 suggests no discontinuities, it is possible that if we zoom in on places where we know ex-ante there are notches in the regulatory environment we will detect something. In our setting, there are three such notches we can examine. In India, the Industrial Disputes Act requires firms employing more than 100 workmen (i.e. 100 workers other than managers) to obtain government permission before laying off workers. This suggests a discrete notch in labor regulation at 100 non-managerial employees, which some have suggested is an important reason for the small size of firms in India.[4]

In Indonesia, firms below a given revenue threshold are exempt from paying the 10% VAT. This again creates a discrete notch, where we would 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 IDR 600 million (about USD 65,000), and was not increased 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 USD 125,000 in 2008) pay a flat tax of about 2 percent of their sales and are exempt from payroll taxes, income

[4] For example, *The Economist* recently noted that "Manufacturing firms need to obtain government permission to lay off workers from factories with more than 100 staff. This partly explains why most firms are so small." *Economist, 2007*. http://www.economist.com/node/9955756. Also see Krueger (2009) for the same view, Besley and Burgess (2004) for an empirical assessment of the importance of the employment law, and Guner, Ventura, and Xi (2007) for a quantitative model.

taxes, and VAT taxes. Firms above the 2 million peso threshold are subject a 15% VAT, a 38% income tax, and a 35% payroll tax.[5]

Figure 4 shows the distribution of non-managerial employment in India in 2011.[6] We zoom in on the range from 60 to 140 non-managerial 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.

Visually inspecting the leftmost panel of Figure 4, there is perhaps a slight bit of bunching at 100 employees, but it is economically very small. In the bin just below the cutoff (97-100 workers), there are a total of 1370 firms. In the next bin (101-104 firms), there are 1013 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 2 tenths of one 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 about 2.5 tenths of one-percent of all informal firms – or 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 IDR 600 million. One would expect more heaping in revenue than in employment, since presumably revenue is easier to adjust in order to stay under the threshold (e.g. firms can choose what year to realize revenue from a given sale). The top 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 IDR 600 million, the figures in the bottom panel are easier to read. In each panel, the left panel shows the distribution for all firms with less than IDR 40 billion in revenue (about USD 4.3 million); to zoom in closer to the discontinuity, the right panel zooms in on firms with less than IDR 1.8 billion in revenue (about USD 200,000).

The figure (particularly the bottom-right), which zooms in on the relevant part of the cutoff shows no bunching at the discontinuity in VAT eligibility. Since 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 see if there are any changes in firm size associated with the different cutoffs that were in place over the years. In the online appendix, we show this figure, with the relevant cutoff line shown each year. Again, we never find any substantial bunching at the discontinuities.

Next, we focus on the potential discontinuity above 2 million pesos in Mexico due to the simplified tax regime for small firms. 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 the discontinuity of 2

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

[6] For non-formal firms, we do not have employment separately by managerial and non-managerial, so we report total employment for these firms.

million pesos (the vertical line) after which firms legally switch from a flat 2% sales tax regime to the combination of the VAT, 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 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.

It is worth noting that there are a small number of other papers that have found some bunching of firms around similar notches, but in most of these cases the quantitative magnitude of the bunching is small. For example, Onji (2009) examines the introduction of VAT 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 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. Similarly, Garicano, LeLarge and Van Reenen (2013) estimate the impact of lifting French regulations that apply to firms with 50 or more 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, though the assumption of fully inflexible wages yields substantially larger estimates. The evidence we present from India, Indonesia, and Mexico is consistent with the generally small magnitudes of bunching observed in other contexts, with the possible exception of France.

IV. Where did the misconception come from?

Given the facts presented in this paper, a natural question is where the misconception about the missing middle – in the sense of the bimodality of the distribution – comes from. We suggest it comes from the combination of two transformations that had previously been made to the 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 5 or 6 bins of firms instead. The data 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 (i.e. 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 and Liedholm and Mead tables and the facts we present here. First, the existing tables refer to the employment share, i.e. what fraction of total manufacturing employment comes from firms of a given size, rather than the distribution of firm size. That is, 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.

The employment share 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 both countries.

The second transformation is that instead of showing the data as nonparametric histograms, due to data limitations, the tabulations reported in these papers are arbitrarily 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 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 3 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 our data, 54 percent of total employment is in firms with 1-9 employees, 12 percent is in firms with 10-49 employees, and 34 percent in firms with 50 or more employees – the same missing middle phenomenon reported by Tybout (2000). Thus, the existing facts about the missing middle seem to come from the combination of these two transformations to the data – the transformation from the distribution of firms to the aggregate employment share, and the arbitrary binning of the employment share distribution rather than examining the distribution non-parametrically.

V. Implications for Theories of Development

Ultimately, the main reason we care about whether the size distribution in developing countries is bimodal or not is what the size distribution tells us about the relevance of alternative theories of development. The absence of bimodality in the size distribution suggests that neither the "small firms are constrained" nor dual economy theories of development are correct 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? It is not enough to simply posit alternative, more capital-intensive production technologies for large vs. small firms, as it is likely that large firms use more capital intensive technologies that, all else equal, would lower the average product of capital in large firms. To make such a dual-economy model fit the facts, 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 in the high 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, say 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% 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 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 there is a confluence of factors that make enforcement 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 correct, the fact that firm size distribution in poor countries is dominated by small firms is because these firms ***choose*** not to exert the effort necessary to grow because their marginal cost would rise if they did grow.

Put differently, the problem of development is how to relieve the differential constraints faced by large firms and not how to relax the constraints faced by small firms. In turn, this view of the world suggests that programs such as microcredit or simplified tax regimes that benefit only small firms will worsen the development problem by further increasing the incentive to stay small.

References

Anagol, Santosh, and Christopher Udry, "The Return to Capital in Ghana," *American Economic Review Papers and Proceedings*, February 2006.

Banerjee, A., & E. Duflo (2005): "Growth Theory through the Lens of Development Economics," Handbook of Economic Growth, in: Philippe Aghion & Steven Durlauf (ed.), Handbook of Economic Growth, edition 1, volume 1, chapter 7, 473-552 Elsevier.

Banerjee, A., and E. 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* (2004) 119 (1): 91-134.

Bloom, Nicholas and John Van Reenan, "Measuring and Explaining Differences in Management Practices across Countries," *Quarterly Journal of Economics*. November 2007, Vol 122(4), pg. 1351-1408.

De Loecker, Jan and F. Warzynski (2012), "Markups and Firm-level Export Status", *American Economic Review*, Vol. 102, No. 6. (October), 2437-2471.

De Mel, Suresh, McKenzie, David, and Christopher Woodruff, "Returns to Capital in Microenterprises: Evidence from a Field Experiment", *Quarterly Journal of Economics, November 2008, Vol 123(4), pp. 1329-1372.*

De Soto, Hernando (1989), The Other Path: the Invisible Revolution in the Third World, New York, NY: Harper and Row.

Fan, Jianqing (1992), "Design-adaptive Non-Parametric Regression," *Journal of the American Statistical Association*, Vol. 87, No. 420. (Dec., 1992), pp. 998-1004.

Garicano, L., LeLarge, C., & J. Van Reenen (2013), "Firm Size Distortions and the Productivity Distribution: Evidence from France," NBER Working Paper No. 18841.

Guner, N., Ventura, G., & X. Yi (2008): "Macroeconomic Implications of Size-Dependent Policies," *Review of Economic Dynamics*, 11(4), 721-744.

Harris, J. & M. Todaro (1970): "Migration, Unemployment and Development: A Two Sector Analysis," *American Economic Review*, 40, 126-142.

Hsieh, Chang-Tai and Peter Klenow (2014), "The Lifecycle of Manufacturing Plants in India and Mexico," University of Chicago mimeo.

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, 2013, pp. 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: Papers and Proceedings* 103 (3).

Krueger, Anne (2007), "The Missing Middle," ICRIER Working Paper 230.

La Porta, Rafael and Andrei Shleifer (2008), "The Unofficial Economy and Economic Development," Brookings Papers in Economic Activity, 275-352.

Leidholm, C., & D. Mead. (1987) *Small- Scale Industries in Developing Countries: Empirical Evidence and Policy Implications*, International Development Paper 9, Agricultural Economics Department, Michigan State University.

Levy, Santiago (2008), Good Intentions, Bad Outcomes, Washington DC: Brookings Institution.

Lewis, Arthur (1954): "Economic Development with Unlimited Supplies of Labour," *Manchester School*, 22: 139-191.

McKinsey and 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), 766-775.

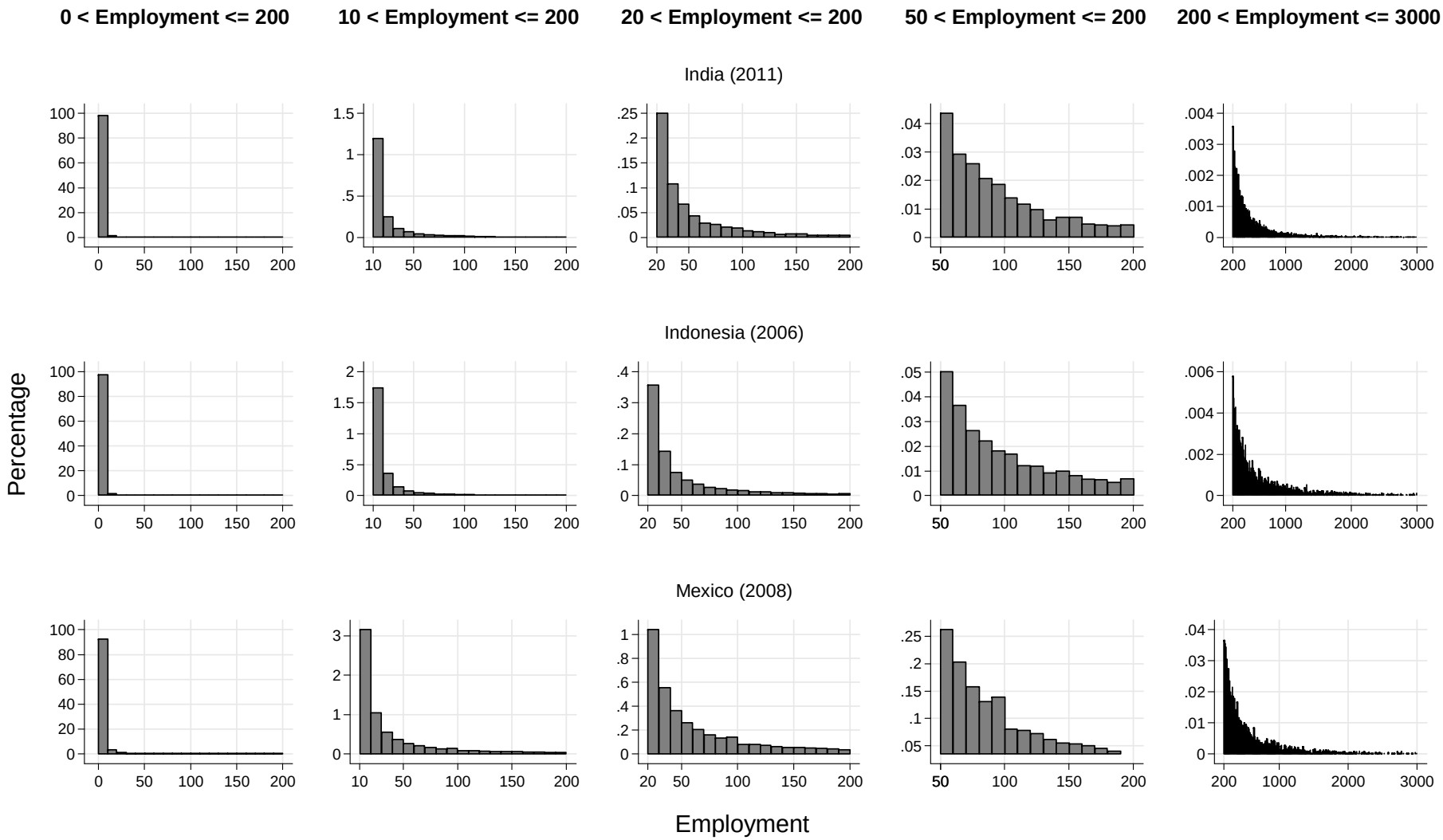
Rauch, James (1991): "Modeling the informal sector formally," *Journal of Development Economics*, 35(1), 33–47.

Sanchez-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." IADB mimeo.

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

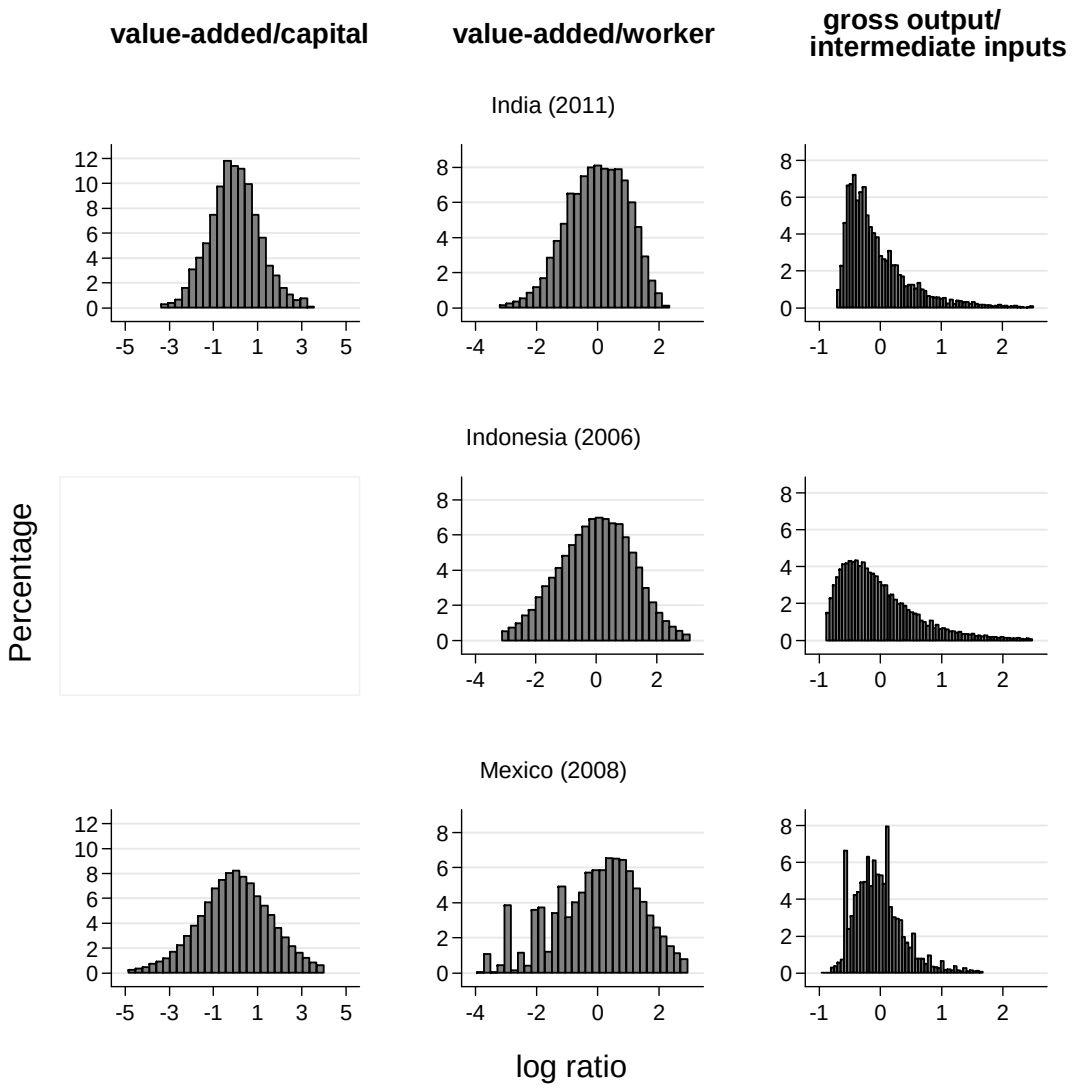
Tybout, James (2000): "Manufacturing firms in developing countries: How well do they do, and why?" *Journal of Economic Literature*, 38(1), 11–44.

Figure 1: Distribution of Firm Employment



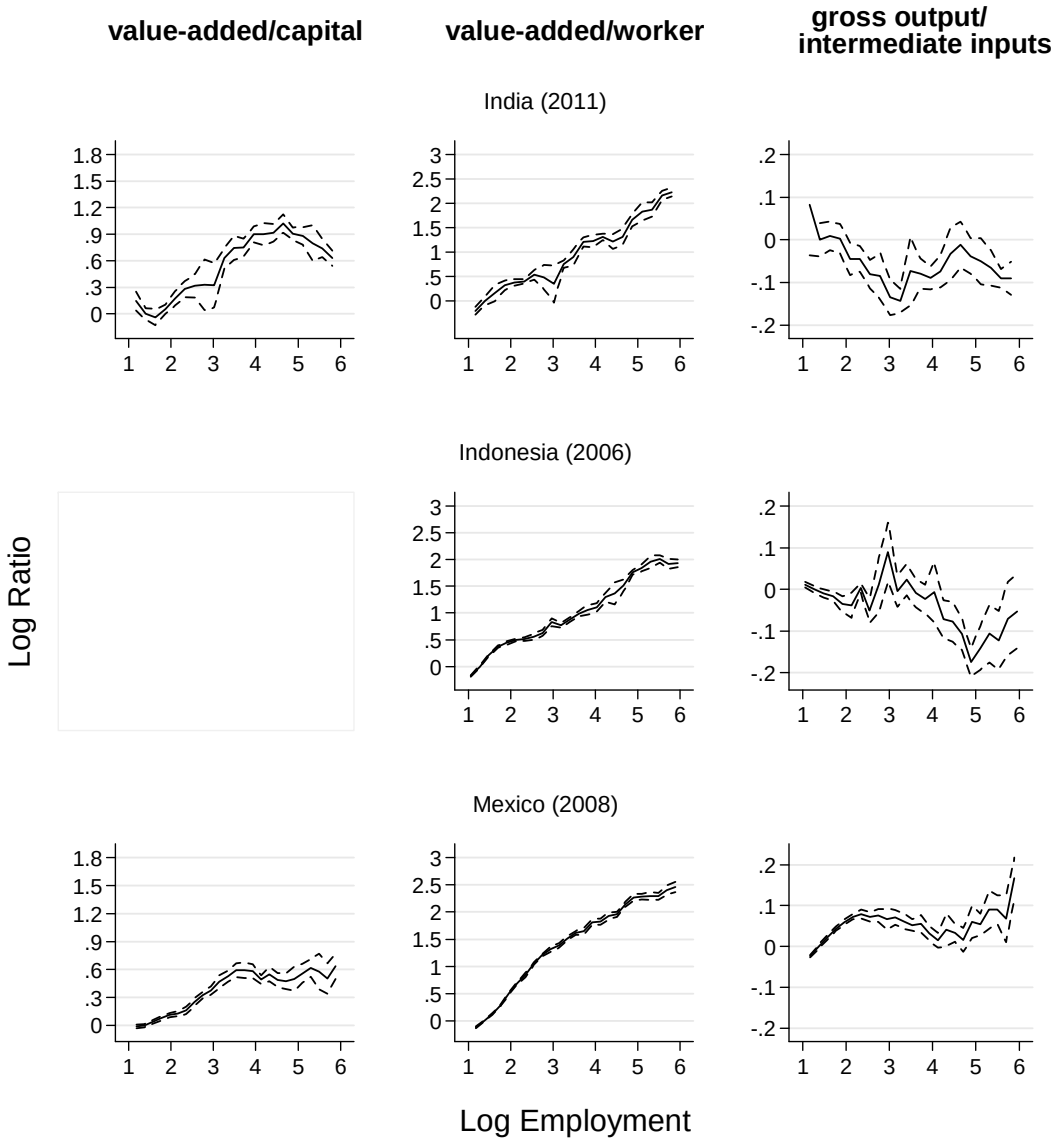
Note: Figure shows distribution of firm size measured by the number of workers. Bin size is 10 workers and each bin contains the upper bound and not the lower bound.

Figure 2: Distribution of Average Products



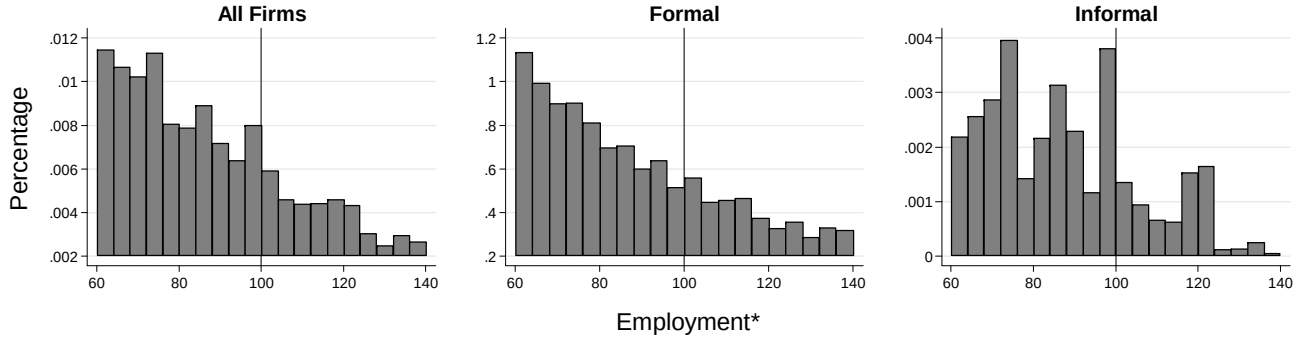
Note: Figure presents distributions of the demeaned average product of capital (column 1), average product of labor (column 2), and ratio of revenues to intermediate inputs (column 3). The bin size is the same in each column and chosen such that the histograms for Mexico have 50 bins. We drop the bottom and top 1% in each sample.

Figure 3: Average Product and Firm Size



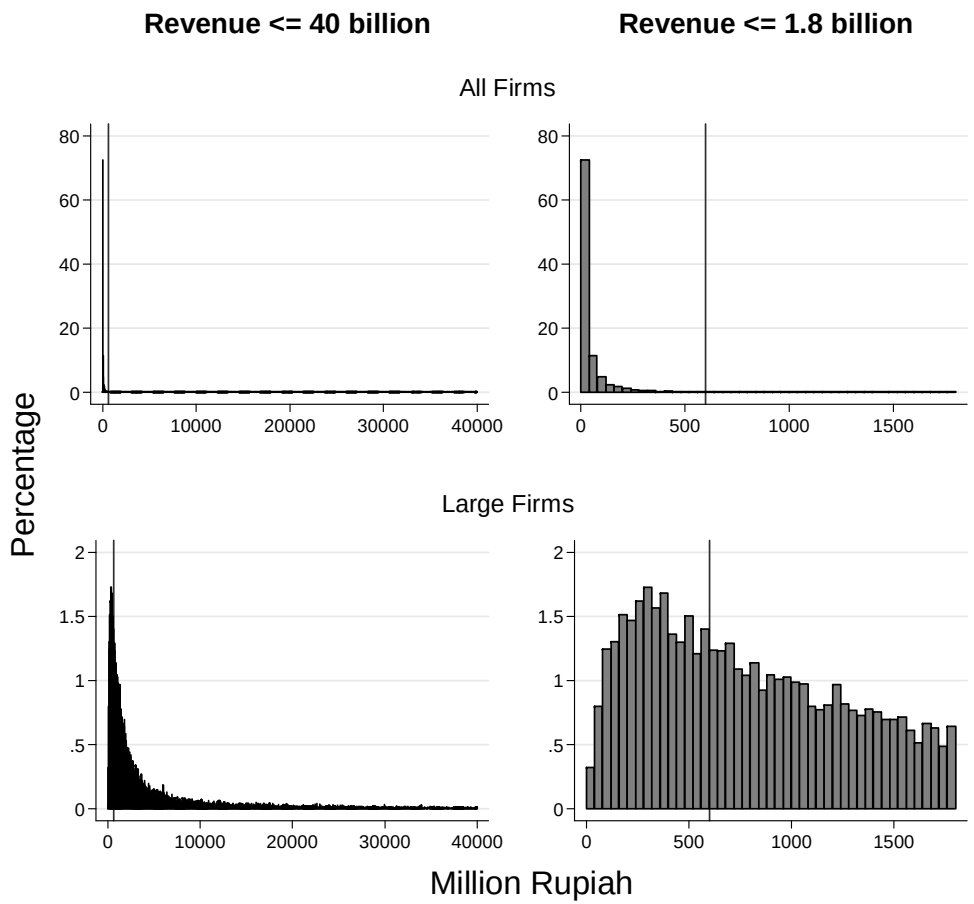
Note: Figure shows local linear regressions of log average product on log employment. We subtract the mean of the fitted value at log(employment)=4. Dashed lines represent 95 percent confidence bounds.

Figure 4: Distribution of Indian Firm Size and Labor Regulations



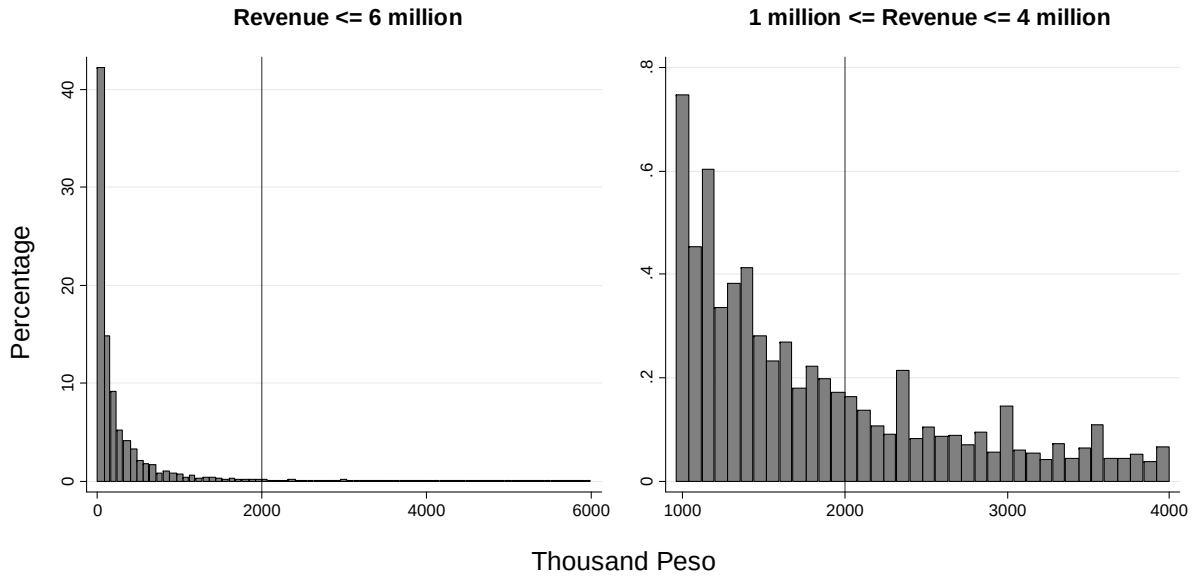
Note: Figure shows size distribution of Indian firms around firms with 100 workers. We exclude managerial workers in the sample of formal firms (ASI). The bin size is 4 workers and each bin contains the upper and not the lower bound.

Figure 5: Distribution of Indonesian Firm Size and the VAT threshold



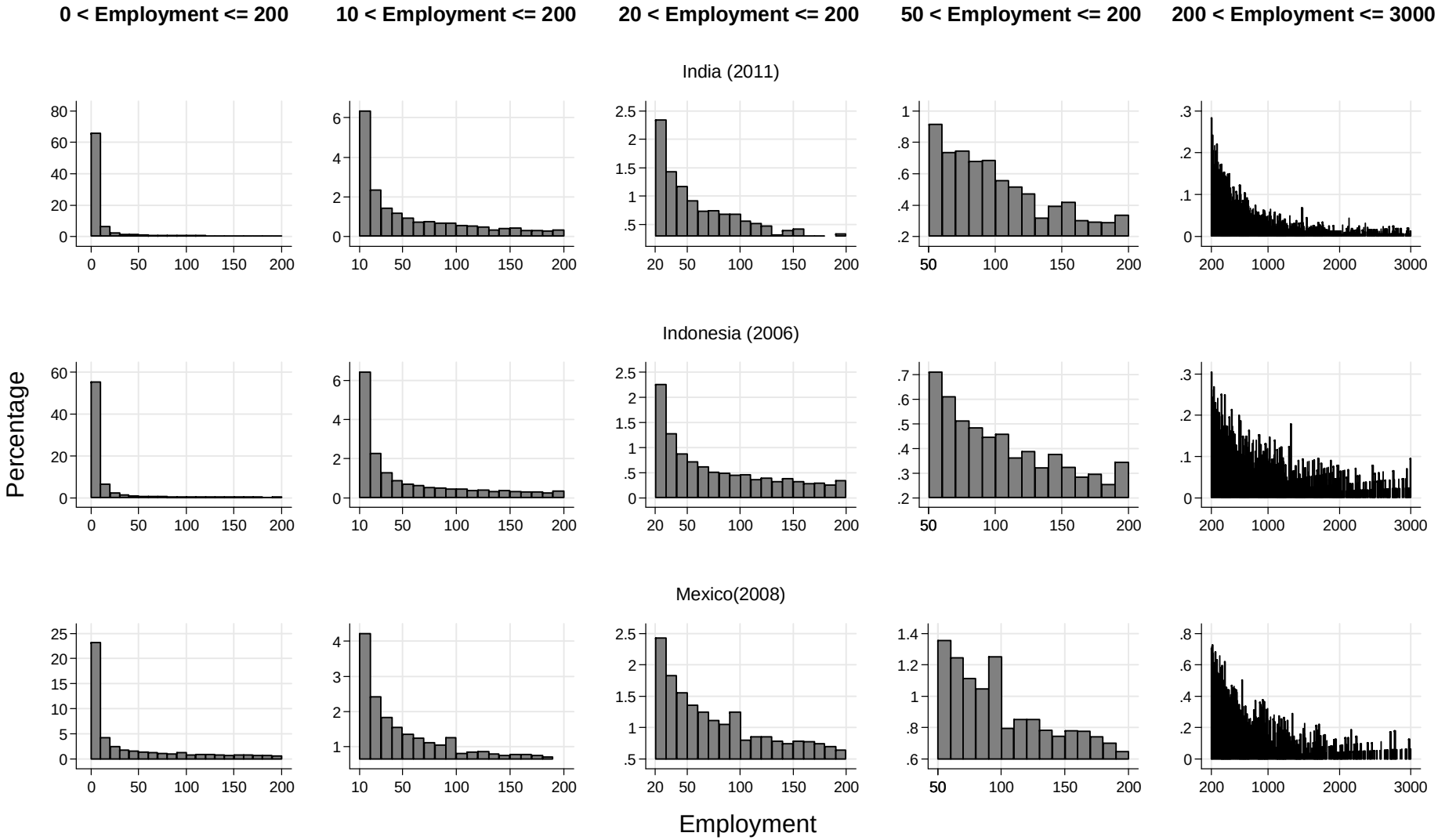
Note: Figure shows distribution of the revenue of Indonesian firms. Vertical line (600 million Rupiah) denotes VAT threshold. The bin size is forty million Rupiah, and each bin contains the upper bound but not the lower bound.

Figure 6: Distribution of Mexican Firm Size and the Simplified Tax Regime Threshold



Note: Figure shows the distribution of revenues of Mexican firms. Vertical line (2 million pesos) denotes threshold for simplified tax regime. The bin size is eighty thousand pesos, and each bin contains the upper bound and not the lower bound.

Figure 7: Distribution of Employment by Firm Size



Note: Figure shows the employment share of the firms in each bin. Bin size is 10 throughout, and each bin includes firms at its upper bound but not those at the lower bound.

Table 1: Distribution of Firm 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 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