

# The Missing Middle, Revisited

James Tybout  
Pennsylvania State University and NBER

October 20, 2014

## 1 Overview

The "missing middle" (MM) story has been present in the development literature for decades, though it has varied substantially from telling to telling.<sup>1</sup> Here is one common version: When policies are imperfectly enforced, producers tend to stay small in order to fly under the government radar, avoiding taxes and costly regulations. Some high-productivity firms do grow large, anticipating that the costs of doing so will be more than offset by the extra operating profits associated with a large customer base, access to the legal system, and modern factory technologies. But mid-sized firms are rare ("missing") because they are large enough to be regulated and pay higher taxes, yet too small to generate large operating profits.

Hsieh and Olken (2014) look for evidence of the MM phenomenon by asking whether manufacturing plant size distributions in India, Indonesia, and Mexico are bimodal. They show that they are not and infer that the MM is a misconception. Of course, one might question whether findings from these three countries provide a sufficient basis for generalization, particularly since much of the MM literature concerns Africa. But this issue aside, I find it problematic to equate the MM with bimodality. The features of the business environment that can induce a MM are varied, and many are not rigidly linked to firm employment level. Size-related policies also evolve through time and are enforced with varying degrees of rigor. Accordingly, it is unlikely that countries that fit the MM story will exhibit clear breaks in their plant size distributions.

## 2 What Constitutes a Missing Middle?

Rather than looking for a bimodal plant size distribution, I think a better test of the MM perspective is to ask whether the share of mid-sized firms, as opposed to small or large firms, is smaller than the share one would observe in an undistorted economy. Figures 1a and 1b depict several hypothetical plant size

---

<sup>1</sup>For a few examples, refer to Little, et al. (1987), Rauch (1992), Tybout (2000), and Hsieh and Olken (2014). Given India's distinctive policy bias toward small-scale producers, it is important to distinguish missing middle analyses of this country from others.

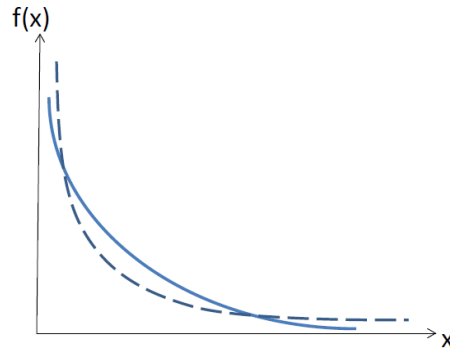


Figure 1a: A Missing Middle

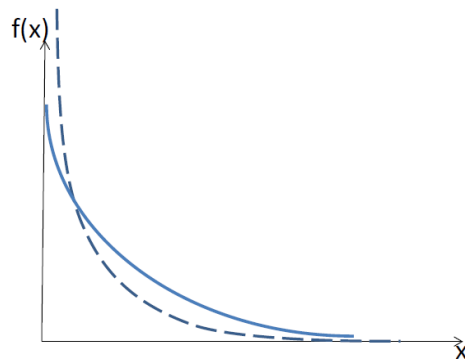


Figure 1b: A Concentration of Small Plants

distributions that illustrate this notion of a MM. The former shows an undistorted distribution (solid line) and a distribution that exhibits a missing middle (dotted line). Note that the distorted distribution is unimodal, and would not be considered to exhibit a MM by Hsieh and Olken (2014). The latter depicts a country that has a preponderance of small plants (dotted line), but does not fit the MM story because both mid-size *and* large plants are scarce relative to the undistorted distribution (solid line).

### 3 An Alternative Approach to the Data

How might one look for the type of MM shown in figure 1a? It is difficult to say what an undistorted size distribution should look like, since low-income countries tend to have small firms for a variety of reasons that have little to do with distortions, including the mix of products they consume and the relatively low degree of urbanization. But the literature on firm size distributions suggests it is reasonable to approximate the undistorted cumulative size distribution as

Pareto (e.g., Axtell, 2001; Luttmer, 2007). I will do so and write this object as:

$$F(x) = 1 - x^{-k}$$

where  $x$  is plant size measured as number of workers, and I assume the smallest feasible plant has a single worker. Then, the general tendency toward small scale production in poor countries can be crudely controlled for by allowing the shape parameter  $k$  to vary across countries, with a larger value for  $k$  implying a smaller size distribution.

With information on the number of plants of each size in a particular country, one could readily investigate whether deviations from the Pareto shape imply a missing middle by my definition. But it is easier to obtain data on the share of the work force in different plant size categories, so I will instead examine deviations of these shares from the shares implied by a Pareto shape. Specifically, if plant sizes are distributed Pareto, the share of the work force employed at plants in the size range  $\ell_i \leq x < \ell_j$  should be:<sup>2</sup>

$$\hat{s}(\ell_i, \ell_j|k) = \ell_j^{1-k} - \ell_i^{1-k} \tag{1}$$

Using the share data reported in Hsieh and Olken (2014), as well as data from the U.S. (as a reference point) and India (as a robustness check), I calibrate equation (1) country by country, choosing  $k$  to minimize the Euclidean distance between the log of the vector of actual shares,  $s$ , and the log of  $\hat{s}$ .<sup>3</sup> The results appear in Table 1. Clearly, for India and Indonesia, the middle category is less populated than predicted by the Pareto distribution, while the smallest and largest size categories are not. In contrast, the Mexican distribution is nearly Pareto, and the U.S. distribution is "fatter" in the middle than predicted by Pareto.<sup>4</sup> Note also that  $k$  is monotonically and negatively related to per capita GDP, reflecting the well-known fact that small firms are relatively common in poor countries.

---

<sup>2</sup>With measure  $M$  producers, total employment at plants in any size interval  $\ell_1 \leq x \leq \ell_2$  is:

$$L(\ell_1, \ell_2|k, M) = M \int_{\ell_1}^{\ell_2} x \cdot f(x) dx = \frac{Mk}{k-1} (\ell_1^{1-k} - \ell_2^{1-k})$$

where  $f(x) = F'(x)$ . The share of employment within the interval is thus:

$$\hat{s}(\ell_1, \ell_2|k) = \frac{L(\ell_1, \ell_2|k, M)}{L(1, \infty|k, M)} = \ell_1^{1-k} - \ell_2^{1-k}$$

<sup>3</sup>An alternative approach to calibration chooses  $k$  so that the actual and predicted shares match exactly in the smallest size category. This approach forces any deviation from the Pareto shape in the middle size class to be offset by an equivalent deviation in the opposite direction for the largest size class. Another alternative approach is to measure shares in levels rather than logs. These alternatives do not yield qualitatively different results, so I do not report them here.

<sup>4</sup>The fact that mid-sized firms are over-represented in the United States might seem to contradict Axtell's (2001) famous finding that the U.S. plant size distribution is exactly Pareto. One explanation may be that Axtell studies all tax-paying enterprises, regardless of sector, while the results I present here are specific to manufacturing.

**Table 1: Predicted versus Actual Employment Shares in the Hsieh-Olken Countries and the United States**  
(middle category: 10-49 workers)

	GDP/pop*	$k$	$s_{<10}$ $-\hat{s}_{<10}$	$s_{10-49}$ $-\hat{s}_{10-49}$	$s_{>50}$ $-\hat{s}_{>50}$
India 2006 (Kothari, 2013)	797	1.47	0.081	-0.089	0.006
India 2011 (Hsieh/Olken)	1,086	1.39	0.073	-0.084	0.011
Indonesia 2006 (Hsieh/Olken)	1,324	1.30	0.061	-0.085	0.024
Mexico 2006 (Hsieh/Olken)	8,113	1.11	0.005	-0.030	0.026
U.S. 2006 (Kothari, 2013)	45,863	1.02	-0.001	0.134	-0.131

\*All GDP per capita figures are taken from the World Bank's (2014) World Development Indicators and expressed in 2005 U.S. dollars.

**Table 2: Predicted versus Actual Employment Shares in Selected Asian Countries and the United States**  
(middle category: 50-200 workers)

	GDP/pop*	$k$	$s_{<50}$ $-\hat{s}_{<50}$	$s_{50-200}$ $-\hat{s}_{50-200}$	$s_{>200}$ $-\hat{s}_{>200}$
India, 2005 (ADB, 2009)	797	1.43	0.029	-0.029	0.000
Philippines, 2005 (ADB, 2009)	1,403	1.28	0.028	-0.032	0.004
Indonesia, 2006 (ADB, 2009)	1,324	1.24	0.040	-0.048	0.008
S. Korea, 2004 (ADB, 2009)	22,236	1.21	-0.094	0.128	-0.034
Thailand, 2007 (ADB, 2009)	2,813	1.16	-0.012	0.020	-0.008
Taiwan, 2006 (ADB, 2009)	16,450	1.15	-0.057	0.108	-0.051
Malaysia, 2005 (ADB, 2009)	5,553	1.09	-0.025	0.114	-0.089
China, 2004 (ADB, 2009)	1,564	1.08	-0.029	0.154	-0.126
U.S., 2006 (Kothari, 2013)	45,863	1.08	-0.046	0.268	-0.220

\*All GDP per capita figures except Taiwan taken from the World Bank (2014) and are expressed in 2005 U.S. dollars. The Taiwanese GDP per capita figure is taken from the IMF's World Economic Outlook and is expressed in 2006 U.S. dollars.

Given the simplicity of this exercise, I find it irresistible to repeat it with alternative data. Table 2 does so using employment shares for the various Asian countries that are discussed in Asian Development Bank (2009). As in Hsieh and Olken (2014), these shares are based on data sets that cover micro enterprises. Importantly, however, they are constructed using a different set of plant size cut-offs. Specifically, what was considered the "middle" in table 1 is now absorbed into the lowest size category, and the new middle captures plants with between 50 and 200 workers.

Several patterns emerge from Table 2. First, shifting the cut-offs weakens but does not reverse the MM result for India. Second, the MM phenomenon remains concentrated in low-income countries, while high-income countries continue to exhibit a middle that is fatter than predicted by the Pareto distribution. Third, however, there are some cases of low income countries that do not exhibit a MM. The most striking example is China in 2004. This may reflect the fact that circa 2004, China was extraordinarily industrialized for a country with its per capita income, and much of its production was destined for higher-income countries.

Of course, my assumption that undistorted distributions are Pareto is restrictive. But it seems reasonable to conjecture that one would find similar cross-country contrasts in labor shares for the 10-50 worker category, or even the 50-200 worker category, under a variety of other distributional assumptions. Accordingly, I am less inclined than Hsieh and Olken (2014) to dismiss the MM phenomenon as a misperception.

## 4 A link to Policy Distortions?

Can we draw policy lessons from these patterns? The business environment in low-income and emerging market economies is distinctive in many ways, and its features vary substantially from country to country. So without detailed country-specific analyses, I do not think it is possible to draw strong conclusions about the sources of the MM phenomenon (as I define it), or the extent to which it reflects distortionary policies. But a number of studies do delve into these details in selected countries, and many conclude that the MM is at least partly policy-induced.<sup>5</sup> The findings of Hsieh and Olken (2014) notwithstanding, this line of research continues to strike me as useful.

## References

- [1] Asian Development Bank (2009) *Key Indicators for Asia and the Pacific*: 2009.
- [2] Axtell, Robert (2001). "Zipf Distribution of U.S. Firm Sizes," *Science* 293(5536): 1818-1820.
- [3] Hasan, Rana and Karl Robert Jandoc (2010) "The Distribution of Firm Size in India: What Can Survey Data Tell Us?" Asian Development Bank, Economic Working Paper Series No. 213.
- [4] Goyette, Jonathan and Giovanni Galipoli (2013) "Corruption and the Size Distribution of Firms," Working Paper, Department of Economics, University of British Columbia.

---

<sup>5</sup>Some of this literature is summarized in Tybout (2000), and Hsieh and Olken (2014) provide additional references. Recent examples not cited in either paper include Hasan and Jandoc (2010), Goyette and Galipoli (2013), and Martin et al. (2014).

- [5] Hsieh, Chang-Tai and Benjamin Olken (2014) "The Missing 'Missing Middle.'" *Journal of Economic Perspectives* 28(3): 89-98.
- [6] Little, Ian, Dipak Mazumdar, and John Page (1987). *Small Manufacturing Enterprises: A Comparative Analysis of India and Other Economies*, New York: Oxford University Press.
- [7] Kothari, Siddharh (2013) "The Size Distribution of Manufacturing Plants and Development," Stanford University, Dept. of Economics.
- [8] Luttmer, Enzo (2007) "Selection, Growth, and the Size Distribution of Firms." *Quarterly Journal of Economics* 1103-1144.
- [9] Martin, Leslie, Shanthi Nataraj, and Ann Harrison (2014) "In With the Big, Out With the Small: Removing Small Scale Reservations in India." NBER Working Paper 19942.
- [10] Rauch, James (1991) "Modeling the Informal Sector Formally," *Journal of Development Economics* 35: 33-47.
- [11] Tybout, James (2000) "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature* 38(1): 11-44.
- [12] World Bank (2014) *World Development Indicators, 2013*.