HUGO A. HOPENHAYN
University of California at Los Angeles

Firm Size and Development

ABSTRACT Firm size increases with GDP per capita. The paper develops a simple framework to explore three alternative sources of variation that may explain this correlation: (1) excessive entry; (2) differences in the distribution of firm productivities; and (3) differences in returns to scale. The results show that all these sources of variation lead to substantial differences in firm size. GDP per capita is also significantly affected, but by an order of magnitude less.

JEL classifications: O11, E13

Keywords: aggregate productivity; firm size distribution; firm dynamics; misallocation

One of the greatest challenges in economics is explaining the disparity in income per capita across countries. There is a large literature documenting the development gap. For example, Caselli reports a twenty-fold gap (1.00 to 0.05) between the per capita gross domestic product (GDP) of the top and bottom 10 percent.[1] After controlling for differences in resource endowments, he finds a corresponding gap in total factor productivity (TFP) on the order of 1.00 to 0.30. Using data for a subset of Latin American countries, I find a 3.5 ratio between the top quarter (Uruguay, Mexico, and Panama) and the bottom quarter (Honduras, Nicaragua, and Bolivia).[2]

Aggregate output and TFP result from the allocation of resources to a very heterogeneous set of productive units or firms. This, in turn, determines the size distribution of firms. A recent paper that uses a comprehensive data set for over a hundred countries finds an elasticity of 0.3 between GDP per capita and average firm size.[3] When the sample is restricted to countries in Latin America, the elasticity is on the order of 0.5 These results are consistent with a widespread view that firm size distribution differs considerably according to the stage of development. This paper examines the link between firm size

1. Caselli (2005).
2. This includes Argentina, Bolivia, Brazil, Colombia, Ecuador, Honduras, Mexico, Nicaragua, Panama, Peru, Uruguay, and Venezuela. The source of these data is Bento and Restuccia (2015).
3. Bento and Restuccia (2015). A similar result is found by Poschke (2010).

distribution and aggregate TFP, through the lens of a model that has become standard in this literature.[4]

I develop a stylized version of a firm heterogeneity model that has as a special case Lucas's span-of-control model and a simple static version of a model I developed earlier.[5] The model has one period and three stages. The first stage involves the entry decision by firms. In the second stage, after observing their productivity draws, firms decide whether to stay or exit. All firms with productivity above a threshold stay, and this determines the degree of selection for the remaining firms that produce in the third stage. Labor is allocated to the entry cost of firms and to production and overhead for those firms that decided to stay in the second stage. Output per capita in this economy is thus determined by three margins: the number of producing firms per capita, the distribution of firms' productivities, and the allocation of labor across these firms. It follows that for a given distribution of firm productivity draws, entry and exit decisions determine both GDP per capita and average firm size.

I explore three alternative sources of variation to explain the correlation between GDP per capita and average firm size. First, I consider distortions that lead to excessive entry (or insufficient exit) of firms over and beyond the equilibrium values, which in this model are the ones that maximize GDP per capita. As the number of firms increases beyond this optimal level, GDP per capita falls monotonically, and average firm size decreases. The second source of variation is changes in the distribution of firm productivities. This is consistent with the findings of Hsieh and Klenow showing that differences in firm productivities can account for half of the TFP gap between the United States and India.[6] More specifically, I consider the special case of a Pareto distribution and examine the effects of an increase in the shape parameter that implies a lower mean and lower tail of the distribution. This has the effect of decreasing both GDP per capita and average firm size. The final source of variation is differences in economies of scale. This is considered by Banerjee and Duflo, who argue that borrowing constraints in India prevent firms from adopting production technologies with higher returns to scale.[7] As I show, in the model an increase in returns to scale decreases GDP per capita and, for the case of a Pareto distribution of productivities, also decreases firm size.

4. Lucas (1978); Jovanovic (1982); Hopenhayn (1992a, 1992b); Melitz (2003).
5. Lucas (1978); Hopenhayn (1992a).
6. Hsieh and Klenow (2009).
7. Banerjee and Duflo (2005).

To perform the quantitative exploration, I consider a baseline economy with a Pareto distribution of shocks with parameter β and returns to scale α. I set α to match the values used by the development literature, which vary between 0.65 and 0.85. Given the cost of entry and overhead, the only remaining parameter is β, which is chosen to match average firm size in manufacturing in the United States, the benchmark economy. The overhead cost (in units of labor) is set to one, which is consistent with the span-of-control model. As I show, the cost of entry has no effect on average firm size (remarkably, it is offset exactly by changes in the exit threshold), so I choose an arbitrary value.

The results show that all the sources of variation considered lead to substantial differences in firm size. GDP per capita is also significantly affected, but by an order of magnitude less. For example, an increase in the number of firms that leads to a 50 percent fall in average size in the United States (consistent with several of the Latin American economies considered) leads to a meager fall in GDP per capita of less than 2 percent. A further increase that reduces average firm size by 25 percent in the United States (consistent with the bottom quartile of the Latin American sample) leads to a decrease in GDP between 6 percent and 9 percent. For the second source of variation, as I increase the parameter β so that average firm size falls 50 percent in the United States, GDP per capita falls by 22 to 33 percent. This is ten times more than the effect of excessive entry, but still short of the observed differences in the data; for those economies in the sample of countries with employment at half the level of the United States, GDP per capita is between 15 and 27 percent of that in the United States. For the third source of variation, a drop in returns to scale α to match half the U.S. average firm size decreases GDP per capita by 12 to 21 percent. The flip side of these results is that the implied elasticities of average firm size to GDP per capita are huge: over 6.5 for the first exercise and over 1.5 for the remaining two. This is five times larger than the elasticities reported by Bento and Restuccia and three times larger than for the Latin American countries in the sample.[8] So while the results suggest that the driving forces considered cannot account for the lion's share of the GDP gap, they are very effective in explaining differences in average firm size.

Few papers address the development fact discussed above. The most notable exception is Bento and Restuccia, who consider the impact of correlated distortions (that is, policies that favor small firms versus larger ones) on productivity decisions.[9] These distortions imply an implicit tax on productivity,

8. Bento and Restuccia (2015).
9. Bento and Restuccia (2015).

making firms invest less in productivity enhancement. In this way, the authors provide a foundation for the differences in the distribution of firm productivities discussed above, as well as a good quantitative explanation for the cross-country elasticity between average firm size and GDP per capita. Hsieh and Klenow, as well as Bello, explore a similar idea, and a number of papers study a similar mechanism.[10]

The paper is organized as follows. The next section discusses the evidence. The paper then develops the model and proves many of the theoretical results that are used in the numerical experiments. The final section concludes.

Evidence

Early evidence on firm size and development seemed to be mixed, but recent papers suggest a strong connection between the two variables. The major problem is the lack of good data that are comparable across countries. According to Alfaro, Charlton, and Kanczuk, who use Dun and Bradstreet WorldBase data, the correlation is significantly negative; Bollard, Klenow, and Li find a mildly negative connection using data from the United Nations Industrial Development Organization (UNIDO); and Poschke finds a significantly positive relation based on the Global Entrepreneurship Monitor (GEM) survey and the Amadeus database (with an "eyeballed" elasticity of 0.3).[11] The most compelling evidence is from a recent paper that constructs a standardized database on establishment and firm sizes using hundreds of separate sources for a total of 134 countries.[12] Consistent with Poschke, the authors find an income elasticity of establishment size of 0.29.

Figure 1 plots a subset of these data corresponding to countries in Latin America.[13] The estimated elasticity as given by the fitted line is 0.53, somewhat higher than that observed for the rest of the world.[14] According to the data, the average Latin American country in the sample has roughly half the average firm size as the United States, and eight of the twelve countries considered are below this threshold.

10. Hsieh and Klenow (2014); Blyde, Restuccia, and Bello (2011); Tavares, Restuccia, and Da-Rocha (2014); Gabler and Poschke (2013); Ranasinghe (2014).

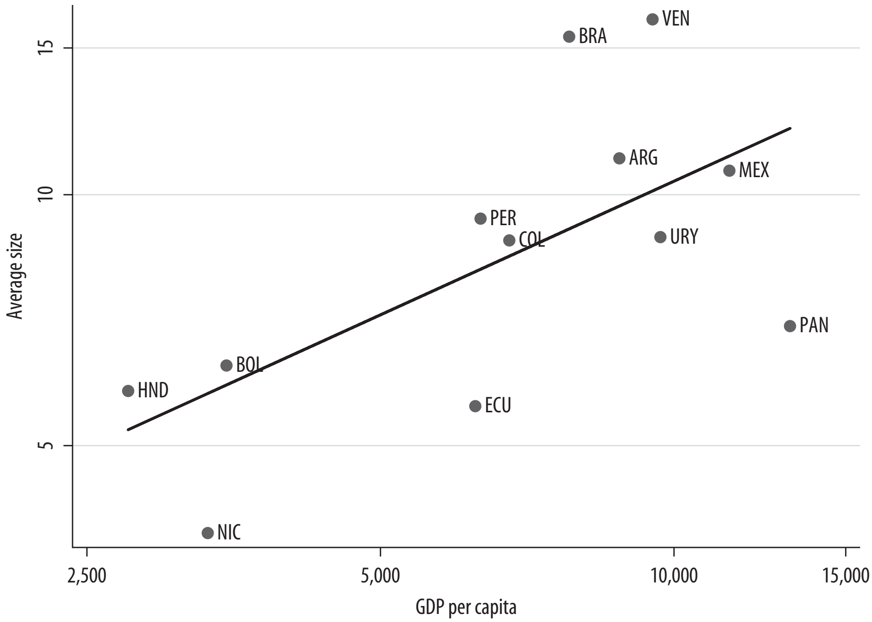
11. Alfaro, Charlton, and Kanczuk (2008); Bollard, Klenow, and Li (2014); Poschke (2010).

12. Bento and Restuccia (2015).

13. I am very grateful to Bento and Restuccia for providing me with these data.

14. Interestingly, the data point for the United States—omitted for scaling reasons—falls almost exactly in the fitted line.

FIGURE 1. Firm Size and GDP per Capita



Source: Bento and Restuccia (2015).

These differences are also reflected when considering the size distribution of firms. A recent study by the Inter-American Development Bank (IDB) provides information on the size distribution of manufacturing firms for a subset of Latin American economies and the United States.[15] Table 1 gives the population-weighted average size distribution of firms for a subset of Latin American countries and for the United States. The main difference is accounted for by the larger share of small establishments (10–19 employees) in Latin America. The disparity in the share of small firms is even larger when considering those establishments under 10 employees, as shown in table 2. While the share of firms in this category is roughly 50 percent in the United States, it exceeds 80 percent for all Latin American countries in this group. Measured by employment, the share is over five times larger than the corresponding value for the United States.

In the following sections, I examine different alternative theories that might account jointly for the differences in firm size and GDP per capita.

15. Pagés (2010, tables 4.1 and 4.2).

TABLE 1. Size Distribution of Firms

Size bracket	*Latin America*[a]	*United States*
10–19 employees	38.5	31.9
20–49 employees	28.3	32.4
50–99 employees	14.3	16.2
100–249 employees	11.3	12.8
250+ employees	7.6	6.8

Source: Pagés (2010, table 4-1).
a. The Latin American average is weighted by population and includes Argentina, Bolivia, Chile, Colombia, Ecuador, El Salvador, Mexico, Uruguay, and Venezuela.

TABLE 2. Share of Firms with Fewer than Ten Employees

Country	*Share of firms*	*Employment share*
Argentina	84.0	22.0
Bolivia	91.7	43.6
El Salvador	82.0	17.7
Mexico	90.5	22.7
United States	54.5	4.2

Source: Pagés (2010, table 4-21).

Model

This section presents a stylized model that nests models by Lucas and Hopenhayn.[16] The model features firm heterogeneity and a fixed endowment of labor to be allocated across firms. The size of the population is N. Firms produce a homogeneous good according to production function $y = zn^{\alpha}$, maximizing profits for a given wage w.[17] In addition, there is an overhead/fixed cost $f \geq 0$ in terms of labor.

There is only one period.[18] The timing of decisions is as follows. At the beginning of the period, entry decisions are made. The firms that enter pay an entry cost c and make a draw for their productivity z from distribution G. After observing this realization, they decide whether to exit or to continue; in the latter case, they choose n to maximize profits.

In the span-of-control model (or the Lucas model), agents in the population are endowed with entrepreneurial/managerial abilities z that are distributed

16. Lucas (1978); Hopenhayn (1992a).
17. This model is equivalent for all the analysis below to the alternative standard model of firm heterogeneity based on monopolistic competition, as in Melitz (2003).
18. The model can easily be extended to multiple periods, as in Hopenhayn (1992b).

according to a cumulative distribution function (CDF), G.[19] While agents differ in managerial abilities, they all have the same ability as workers. An agent with ability z can produce with n workers according to the production function given above, forgoing the opportunity to earn a wage as a worker. My model specializes to the Lucas model when $c = 0$ and $f = 1$. The latter represents the overhead given by the entrepreneur's input. As will become clear, $c = 0$ since there is no real technology for creating firms, so the set of active firms is determined by selection of agents into entrepreneurship.

The second model I focus on is an entry and exit model from my earlier work (or the Hopenhayn model).[20] There is a technology for creating firms at cost c. The productivity of a firm is revealed after paying this cost—after entry—and drawn independently from a common distribution G. In the earlier paper, this shock evolves over time, and there is entry and exit in a long-run steady state equilibrium. Here I simplify the structure to one period. The model still allows for determining the number of firms and the selection of firms, through the exit decision. As shown below, the latter plays a very similar role to selection into entrepreneurship in the Lucas model. The main difference between the two models for the purpose of this analysis is the added margin of entry in the Hopenhayn model.

Equilibrium

Let $n(z, w)$ denote the optimal employment decision of a firm with productivity z when the wage rate is w. Let $\pi(z, w)$ denote its gross profits, prior to paying the fixed costs. It is easy to verify that profits are increasing in z, so there is a threshold z^* such that only firms with $z \geq z^*$ will be active, where $\pi(z^*, w) = wf$.

An equilibrium is given by a wage w (the only relative price in this economy), optimal employment decisions $n(z, w)$, a threshold for exit z^*, and entrants M_0, such that

$$\pi(z^*, w) - wf = 0;$$

$$\int_{z^*}\left[\pi(z, w) - wf\right]dG(z) = wc, \text{ if } M_0 < N \text{ or, if greater than } c, \text{ then } M_0 = N;$$

$$M_0\int_{z^*}\left[n(z, w) + f\right]dG(z) + cM_0 = N.$$

19. Lucas (1978).
20. Hopenhayn (1992a).

The first condition was described above and corresponds to the optimal exit rule. The last condition is market clearing, where the sources of labor demand are productive and overhead workers, and cM labor is allocated to entry. Finally, the second condition is a zero-profit condition for entering firms when $M_o < N$. It states that expected profits must equal the cost of entry, wc.[21] The last part of this condition is included to accommodate the Lucas model, where I set $c = 0$ and $M = N$. In reality, this condition plays no role in the Lucas model since there is no creation of firms, so the only relevant decision determining the set of active firms is the threshold z^*. Also in the Lucas model, since $M_o = N$. the last condition reduces to

$$\int_{z^*}\left[n(z, w) + f\right]dG(z) = 1.$$

Moreover, using that fact that $f = 1$ in the Lucas model (that is, the only overhead labor is the manager/entrepreneur), $[1 - G(z^*)]$ is the share of entrepreneurs/firms in the population. The above equation is then equivalent to

$$\int_{z^*} n(z, w)\, dG(z) = G(z^*),$$

which is the share of productive workers. For the Hopenhayn model, while M_o firms enter, all those with $z < z^*$ are inactive (that is, they exit), so the total number of producing firms is actually $M_o[1 - G(z^*)]$.

I now turn to the analysis of equilibrium. I start by considering the second stage, once entry and exit decisions have been made. Entry and exit are considered later in the paper.

Aggregation

Consider an economy where there are M firms with productivity distribution F and an endowment of labor L to be allocated among the firms, not considering overhead labor. The distribution F and number of firms M may represent the net effect of entry and selection due to exit in the stages prior to production.[22]

Let $a = 1/(1 - \alpha)$. For a wage rate w, it follows easily that profit maximization gives employment choices that are proportional to z^a, so that $n(z) = bz^a$ for a value b to be determined and

21. All fixed and entry costs are denoted in units of labor and thus multiplied by the wage rate w in the above expressions.

22. Given entry and exit decision rules (M_o, z^*), it follows that $M = M_o[1 - G(z^*)]$, F is the distribution of productivities G conditional on $z \geq z^*$ and $L = N - M_oc - M_o[1 - G(z^*)]f$.

$$y(z) = z(bz^a)^\alpha = b^\alpha z^{1+aa} = b^\alpha z^a \tag{1}$$

is also proportional to z^a. Furthermore, the labor market clearing condition

$$M\int n(z)\,dF(z) = Mb\int z^a\,dF(z) = L,$$

implies that

$$b = \frac{L}{M\int z^a\,dF(z)}.$$

Total output $y = M\int y(z, w)dF(z)$, which after substituting equation 1 and the above equation for b gives

$$y = \bar{z}M^{1-\alpha}L^\alpha, \tag{2}$$

where $\bar{z} = (Ez^{1/(1-\alpha)})^{1-\alpha}$ is the geometric mean of productivities over all active firms, M is the number (mass) of producing firms, L is the number of production workers (not including fixed or entry cost), and F is the distribution of firm productivities. The aggregate production function inherits the structure of the firm-level production function with returns to scale α, together with the geometric average productivity across firms $\bar{z}$. Production is Cobb-Douglas in the stock of firms M and labor L, the latter with the same coefficient α as in the firms' production function. It also follows that labor share $wL = \alpha y$, so the equilibrium wage w consistent with these allocations can be determined from the above equation. Equation 2 makes very clear the role of firms as another input into production.[23]

The average productivity $\bar{z}$ is solely determined by the productivity distribution F.[24] However, this distribution will, in general, not be independent of M. In particular, in the span-of-control model where all agents above some threshold z^* become entrepreneurs, $M = [1 - G(z^*)]N$, so increases in M are

23. Here, α represents total returns to scale. For example, if the production function is extended at the firm level to a Cobb-Douglas function with capital and labor $n^{\beta 1}\,k^{\beta 2}$, then $\alpha = \beta_1+\beta_2$.

24. The definition of TFP depends on the treatment of M. If it is considered an input, then TFP equals $\bar{z}$. In contrast, if M were not accounted as an input, TFP would equal $(\bar{z}\ M)^{1-\alpha}$. Conceptually, I think the first alternative makes more economic sense, but from the measurement perspective, the answer depends on the extent to which firm capital (sometimes denoted as intangibles) is accounted for in aggregate capital stock.

associated with a fall in z^* and thus in the geometric mean $\bar{z}$. Similarly, in the Hopenhayn model, $M = M_o[1 - G(z^*)]$ and (M_o, z^*) are jointly determined. These effects are taken into consideration below in the analysis of entry and exit/selection.

THE SIZE DISTRIBUTION OF FIRMS. The equilibrium exit condition, $\pi(z^*, w) = wf$, has implications for the size distribution of firms. It follows from the homogeneous production function that $wn(z^*) = \alpha y^*$ and $\pi(z^*, w) = (1 - \alpha)y^*$, so

$$\frac{wf^*}{wn(z^*)} = \frac{\pi(z^*,w)}{wn(z^*)} = \frac{1-\alpha}{\alpha}.$$

This implies that the size of the smallest firm $n(z^*) = [\alpha/(1 - \alpha)]f$. In particular, it is independent of the cost of entry and also independent of the distribution of firms' productivity! I turn now to the size distribution of firms. As mentioned above, $n(z)$ is a linear function of z^a, so

$$n(z) = \left(\frac{z}{z^*}\right)^a n(z^*) = \left(\frac{z}{z^*}\right)^a \left(\frac{\alpha f}{1-\alpha}\right).$$

The size distribution of firms is thus fully determined from z^* and the distribution G. In particular, average firm size is given by

$$\bar{n} = \left(\frac{\alpha f}{1-\alpha}\right)\left\{\frac{\int_{z^*} z^{1/(1-\alpha)}\, dG(z)}{\left[1-G(z^*)\right](z^*)^{1/(1-\alpha)}}\right\}. \tag{3}$$

Entry and Exit

I now consider the determination of entry and exit in the first stage. Rather than solving for the equilibrium, I find the allocations through a planner's problem. The competitive equilibrium maximizes total output—or equally output per capita—subject to a resource constraint, which I now consider. The allocations to be chosen are the entry rate M_o and the exit/selection threshold z^*. Given a pair (M_o, z^*), the number of firms $M = M_o[1 - G(z^*)]$ and

$$\bar{z} = \left[\frac{\int_{z^*} z^{1/(1-\alpha)}\, dG(z)}{1-G(z^*)}\right]^{1-\alpha}.$$

Substituting in equation 2,

$$(4) \qquad y = \left[\int_{z^*} z^{1/(1-\alpha)} \, dG(z) \right]^{1-\alpha} M_0^{1-\alpha} L^{\alpha},$$

and the resource constraint is

$$(5) \qquad M_0 \left\{ c + \left[1 - G(z^*) \right] f \right\} + L \leq N.$$

Equilibrium (and optimal) entry and selection z^* can be obtained by maximizing equation 4 with respect to $M_0 \leq N$, z^*, and L subject to equation 5. Letting λ denote the multiplier of this constraint, the first-order conditions are

$$(6) \qquad (1-\alpha)\frac{y}{M_0} \geq \lambda \left\{ c + f\left[1 - G(z^*) \right] \right\} \text{ and } \lambda (N - M_0) = 0;$$

$$(7) \qquad \alpha \frac{y}{L} = \lambda;$$

$$(8) \qquad (1-\alpha)\frac{y}{\int_{z^*} z^a \, dG(z)} (z^*)^a = \lambda M_0 f;$$

where $a = 1/(1-\alpha)$.

I now use the above conditions to derive specific implications for each of the two models. In the Lucas model, $M = N$, $L = G(z^*)N$, and $c = 0$. Using the last two conditions it follows that

$$(9) \qquad \frac{\alpha f}{1-\alpha} \int_{z^*} \left(\frac{z}{z^*} \right)^a dG(z) = G(z^*).$$

This equation determines uniquely z^*, as the left-hand-side term is strictly decreasing and the right-hand-side term increasing in z^*. Average firm size equals $G(z^*)/[1 - G(z^*)]$ without taking into account overhead and $[1 - G(z^*)]^{-1}$ otherwise. It is obviously increasing in z^*. The following proposition summarizes the main implications.

Proposition 1. *In the Lucas model, z^* and average firm size are increasing in f and α.*[25]

25. In the Lucas model, $f = 1$, so this is a slight generalization allowing for additional overhead workers besides the manager.

Consider now the Hopenhayn model, where all conditions are satisfied with equality. Using the first two and the resource constraint, it follows that

$$L = \alpha N \tag{10}$$

and

$$M_0\left\{c + f\left[1 - G\left(z^*\right)\right]\right\} = \left(1 - \alpha\right)N. \tag{11}$$

This result is standard for a Cobb-Douglas production function, where α and $(1 - \alpha)$ are the share of resources (in this case of N) spent in each of the two inputs (M_0, L), where $\{c + f\,[1 - G(z^*)]\}$ represents the price of M in units of labor. Using the first and last conditions,

$$\int_{z^*}\left(\frac{z}{z^*}\right)^{a} dG\left(z\right) = \frac{c}{f} + \left[1 - G\left(z^*\right)\right]. \tag{12}$$

This equation uniquely determines z^*. Average firm size (not including overhead) is

$$\bar{n} = \frac{L}{M_0\left[1 - G\left(z^*\right)\right]}.$$

Using equations 10 and 11, it becomes

$$\begin{aligned} \bar{n} &= \left(\frac{\alpha}{1-\alpha}\right)\left\{\frac{c + f\left[1 - G\left(z^*\right)\right]}{1 - G\left(z^*\right)}\right\} \\ &= \left(\frac{\alpha}{1-\alpha}\right)\left[\frac{c}{1 - G\left(z^*\right)} + f\right], \end{aligned} \tag{13}$$

and including overhead it is just $\bar{n} + f$. The following proposition summarizes the main implications for the Hopenhayn model.

Proposition 2. *In the Hopenhayn model, z^* is decreasing in c and increasing in f. It is also increasing in α. Average firm size is increasing in f. The effect of c on average size is ambiguous: it increases if $E\,[(z/z^*)^a \mid z \geq z^*]$ increases as a result.*

Distortions on Entry and Exit

This section discusses distortions on entry and exit as a potential explanation for the correlation between GDP per capita and average firm size. A complete analysis would require taking into account the specific policies that lead to these distortions, as they might imply other wedges between equilibrium and optimal allocations. Here I take a shortcut by taking into account all first-order conditions except those that pin down the variable under consideration. I consider, first, the effect of *excessive entry* (high M_0) and, second, the effect of *too little exit* (low z^*). In the Lucas model, the distinction really does not exist, since the number of firms equals $[1 - G(z^*)]N$ and thus is determined only by the threshold z^*. A decrease in z^* leads to an excessive number of firms, smaller average size, and lower GDP per capita, making it a potential candidate to explain the main fact.

The effect of an increase in entry in the Hopenhayn model can be analyzed combining the first-order conditions for L and z^* for fixed M_0, which gives

$$\left(\frac{\alpha f}{1-\alpha}\right)\frac{\int_{z^*} z^a \, dG(z)}{\left[1-G(z^*)\right](z^*)^a} = \frac{L}{M_0\left[1-G(z^*)\right]}$$

or

$$\left(\frac{\alpha f}{1-\alpha}\right)E\left[\left(\frac{z}{z^*}\right)^a \middle| z \geq z^*\right] = \frac{L}{M_0\left[1-G(z^*)\right]}.$$

An increase in M_0 decreases L and increases z^*. The overall effect on average size (the term on the right-hand side) depends on how the expectation on the left changes with increases in z^*. For example, when z follows a Pareto distribution, $E[z^a/(z^*)^a]$ is independent of z^* so average size does not depend on M_0. Consider now a decrease in z^* (too little exit) in the Hopenhayn model. Equation 13 still holds since it does not rely on the first-order condition for determining z^*. An increase in z^* increases the unit cost of firms $\{c + [1 - G(z^*)]f\}$ and by equation 11 decreases M_0. The net effect is still an increase in the number of firms $M_0\,[1 - G(z^*)]$ and a decrease in average size (L does not change). The suboptimal degree of selection will also decrease GDP. As in the Lucas model, this is a possible channel to explain the observed positive correlation between GDP and average firm size.

Proposition 3. *An increase in the number of firms (entrepreneurs) decreases average size in the Lucas model and increases average size in the Hopenhayn model if* $E\,[(z/z^*)^a \mid z \geq z^*]$ *increases with* z^* *as a result. A decrease in the exit rate in the Hopenhayn model decreases average size and increases the total number of operating firms.*

Pareto Distribution

I specialize results to the case where $1 - G(z) = z^{-\beta}$ so G is a Pareto distribution with parameter β. This specification is very tractable and is the one used for my computations in the next section. Substituting for the distribution G in equation 9 yields

$$\frac{\beta\alpha(z^*)^{-\beta} f}{(1-\alpha)(\beta-\alpha)} = 1 - (z^*)^{-\beta}.$$

Average firm size is thus

$$\frac{1-(z^*)^{-\beta}}{(z^*)^{-\beta}} = \frac{\beta\alpha f}{(1-\alpha)(\beta-1)}.$$

In the pure Lucas model $f = 1$, and a simple expression for the fraction of firms is

$$(z^*)^{-\beta} = \frac{\beta(1-\alpha)-1}{\beta-1},$$

which is obviously increasing in β and decreasing in α. This, in turn, implies that average firm size decreases in β. It follows immediately that GDP also decreases in β (since higher β represents a worse distribution of productivities), thus providing another source of variation consistent with the observed positive correlation between GDP and average firm size.

The expression for average firm size is exactly the same in the Hopenhayn economy. It is independent of the cost of entry, increasing in α and f and decreasing in β. More specifically, the allocations in the Hopenhayn economy are

$$L = \alpha N;$$

$$M_0 = \left(\frac{1}{\beta}\right)\left(\frac{N}{c}\right);$$

$$(z^*)^{-\beta} = \left(\frac{c}{f}\right)\left[\beta(1-\alpha)-1\right].$$

Using these equations, the total number of productive firms is

$$M_0\left[1-G(z^*)\right] = \left(\frac{1}{f}\right)N\left(1-\alpha-\frac{1}{\beta}\right),$$

so changes in the cost of entry c have no final effect on active firms; the fall in M_0 is exactly compensated by the fall in z^*. The following proposition summarizes the results derived in this section.

Proposition 4. *An increase in* β *decreases GDP and the average size of firms in both models. An increase in* α *increases average size of firms in both models.*

Explaining GDP/Average Firm Size Correlation

I now use the results in the previous section to examine several alternative mechanisms that might contribute to an explanation of the facts presented earlier. I consider three mechanisms: (1) an increase in the number of firms over and above the undistorted equilibrium/optimal level for less developed economies; (2) an improvement in the distribution of firm productivity (that is, a decrease in β for the Pareto distribution) with development; (3) an increase in α (that is, higher returns to scale) for more developed economies. The results in this section are mostly quantitative and are given to provide orders of magnitude for these different channels. Throughout, the analysis considers a Pareto distribution for productivities with parameter β. Together with the returns to scale α, these are the only two parameters in the Lucas model, whereas the Hopenhayn model also includes fixed cost f and entry cost c. I choose the fixed cost $f = 1$ for consistency with the Lucas model (one entrepreneur per firm), so that both give the same average firm size. For simplicity, $c = 1$. (Larger values of c result in lower output, but as discussed in the previous section, they have no impact on average size with the Pareto distribution.)

In the baseline scenario, I choose two alternative values for α: $\alpha = 0.65$ and $\alpha = 0.85$. The latter is the value most frequently used in the development/macroeconomic literature. The former is consistent with the elasticity of substitution used by Hsieh and Klenow, which is a standard reference in the

TABLE 3. GDP, Average Size, and Entry

	GDP	
Average size	$\alpha = 0.65$	$\alpha = 0.85$
20	1.00	1.00
10	0.98	0.99
5	0.91	0.94
2.5	0.74	0.80
2	0.64	0.73

development literature.[26] The value of β is chosen so that the average size is approximately twenty, matching the value for U.S. manufacturing. This gives values of $\beta = 3.14$ for the economy with $\alpha = 0.65$ and $\beta = 9.15$ for $\alpha = 0.85$.

Excessive Entry

Earlier, the paper described the theoretical effects of excessive entry, which reduces average size and at the same time lowers GDP per capita. In the baseline scenario, average size is approximately twenty. In the Lucas model, an increase in the number of firms means a decrease in the threshold z^*, which immediately reduces the average size of firms, $1/[1 - G(z^*)]$. This latter expression includes the entrepreneur in the employment account. This has three effects on output: it increases the number of operating firms (positive); it reduces the number of production workers (negative); and it decreases the average quality of firms (negative). The net effect is negative: as z^* decreases from the value in the optimal allocation, GDP decreases monotonically.

Results are given in table 3. Decreasing z^* from the baseline to one-tenth of the baseline results in a decline of GDP per capita of 27 percent and 36 percent, depending on the value of α. This is a sizable effect, but it is still very small compared to the standard development gap. For instance, an increase in the number of firms that leads to a fall in average size to half the level in the United States (consistent with several of the Latin American economies considered) leads to a meager fall in GDP per capita of less than 2 percent. A further increase that results in average firm size that is a quarter of the U.S. level (consistent with the bottom quartile of the Latin American sample) leads to a decrease in GDP of between 6 percent and 9 percent. This small effect can be explained by the fact that deviations from optimal entry

26. Hsieh and Klenow (2009).

are second order (at the optimum) for GDP per capita but first order for average firm size. Only when deviations are very large do they have a sizable effect on GDP per capita. For example, when entry rates increase so that average size drops from 5.0 to 2.5, the corresponding drop in GDP per capita is on the order of 15 percent.

The values generated through this experiment imply an elasticity of average size with respect to GDP per capita of between 6.5 and 8.5, which is huge compared with the 0.3 value reported by Bento and Restuccia or the 0.5 found in the sample of Latin American countries.[27] This result is explained by the fact that very large excessive entry has a relatively moderate, albeit significant, impact on GDP per capita. I could not perform this experiment for the Hopenhayn economy: as shown above, increases in the entry rate M_o have no effect on the average size of firms because they are compensated exactly with increases in the exit threshold z^*. In contrast, the direct effect of a decrease in z^* does increase the total number of firms, though the overall effect is less than in the Lucas model, as it is mitigated by a fall in the rate of entry M_o.

The Impact of the Distribution of Firm Productivity

To assess the impact of changes in the distribution of firm productivities for the Pareto distribution, the parameter β is increased from the baseline value. This shifts the weight of the distribution to lower values, thus lowering average firm productivity and reducing the tail of the distribution of firm sizes. As stated in proposition 4, this results in reductions in average size and GDP in both models. While this exercise takes the changes in β as exogenous, there are many sources of variation across countries that could explain the different distributions of firm productivity. For instance, they might reflect differences in the quality of human capital, the existence of institutions that complement and enhance entrepreneurial ability, differences in the degree of transfer of knowledge between entrepreneurs, incentives to invest to enhance a firm's productivity, and so forth.[28]

As in the previous section, this exercise considers the two values for α and the corresponding values for β indicated above. The values of β are then increased to find the impact on average size and GDP per capita. Table 4 gives results for this quantitative exercise based on the Lucas model.[29] Note that

27. Bento and Restuccia (2015). The flip side is that the inverse, the elasticity of GDP with respect to average size, is only between 0.12 and 0.15, while in the data it exceeds 3.0.

28. On investment incentives, see Bento and Restuccia (2015).

29. The results for the Hopenhayn model are very similar and thus are omitted.

TABLE 4. GDP and Average Size with Changes in β

	GDP	
Average size	$\alpha = 0.65$	$\alpha = 0.85$
20	1.00	1.00
15	0.83	0.91
10	0.67	0.78
6.7	0.53	0.65**
2.9	0.25*	

* Achieved for $\beta = \infty$ when $\alpha = 0.65$.
** Achieved for $\beta = \infty$ when $\alpha = 0.85$.

there is a minimum value for average size in the limit as $\beta \to \infty$. The impact on GDP relative to the change in average size is larger than in the previous exercise. The elasticity of GDP to average size is now 0.37 (for $\alpha = 0.85$) and 0.68 (for $\alpha = 0.65$), which is far from the observed values in the data and about four times larger than the values obtained before. The catch is that the values for β that generate these smaller average firm sizes are extremely large, so they would imply a very small dispersion in average sizes for firms in less developed countries. This is contrary to the data for the sample of Latin American countries used here. As mentioned above, the smallest firm in the Lucas model has a size equal to $\alpha/(1 - \alpha)$, a value of 5.7 for $\alpha = 0.85$, and a value of 2.14 for $\alpha = 0.65$.

When the values of average size in table 4 are compared with this minimum, it is apparent that as β increases, the gap between the average and minimum size narrows considerably. This reflects the fact that the ratio between the mean and the lowest value in the Pareto distribution is $\beta/(\beta - 1)$, which is clearly decreasing in β and converges to one as $\beta \to \infty$. Employment also follows a Pareto distribution with parameter $\gamma = \beta - [1/(1-\alpha)]$, and for large values of β this distribution has very little variance. Moreover, the empirical results reported by Hsieh and Klenow show that firm-level productivity is more dispersed in India than the United States: the standard deviation of the log of TFP is over 35 percent higher and is similarly large for the ratio of TFP between the ninetieth and tenth percentiles.[30] A similar picture emerges from the evidence for Latin American countries reported in the IDB study.[31] In contrast, in this setting the log of firm productivity follows an exponential distribution with parameter β, which has a standard deviation of $1/\beta$ and is

30. Hsieh and Klenow (2009).
31. Pagés (2010).

thus decreasing in β. There are two caveats to this comparison. One is measurement error, which might be much larger in India and the Latin American economies than in the United States. Second, the results are obviously dependent on the Pareto distribution, and a further exploration would require backing out the empirical distribution of productivities.[32] What seems to be true in the data is that the distribution of firm-level productivity has a larger left tail in India than in the United States, and this might be driving the results. As explained by Hsieh and Klenow, the differences in the distribution of firm productivities between the United States and India still account for about 50 percent of the overall difference in aggregate TFP.[33]

The Impact of Returns to Scale

With regard to the impact of changes in the returns to scale parameter α, proposition 4 states that increases in α are associated with higher average size. As explained below, they are also likely to increase GDP and may thus provide a possible channel for explaining the observed correlation. Why would countries differ in their returns to scale? One possible explanation is that entrepreneurs face a menu of alternative technologies that differ in economies of scale. For example, Banerjee and Duflo consider the choice of technology between a low-fixed-cost/low-α technology and a high-fixed-cost/high-α technology.[34] In their model, the latter is more cost effective, but requires significant initial investments that entrepreneurs facing borrowing constraints cannot afford. As the authors argue, such a model can simultaneously explain smaller firm size and lower aggregate productivity.

Yet another reason why returns to scale might vary across economies relies on a deeper foundation. Returns to scale are usually considered a characteristic of technology. Decreasing returns to scale are thus associated with the existence of some fixed factor that cannot be replicated. It is hard to find such fixed factors, especially if they are part of the human capital stock. Experts can breed new experts, and firms and individuals make considerable investment in learning and thus replicating existing knowledge. But if the extent to which factors are fixed depends on their replicability, this ceases to be a property of technology and becomes an economic decision of whether or not to bear the costs of replication. Incentives and costs for doing so will imply varying degrees of "fixedness" of inputs and corresponding differences in

32. See Hsieh and Klenow (2009).
33. Hsieh and Klenow (2009).
34. Banerjee and Duflo (2005).

TABLE 5. GDP and Average Size with Changes in α

	GDP			
	Lucas model		Hopenhayn model	
Average size	$\beta = 3.137$	$\beta = 9.150$	$\beta = 3.137$	$\beta = 9.150$
20	1.00	1.00	1.00	1.00
15	0.91	0.95	0.91	0.95
10	0.79	0.87	0.80	0.88
5	0.62	0.73	0.65	0.75
2.5	0.50	0.61	0.54	0.65

returns to scale. Likewise, when decreasing returns to scale on the revenue function come from a downward-sloping demand, one can think of differences across countries related to variations in the ease by which demand can be expanded, that is, consumers replicated.

To explore the impact of an increase in α on aggregate GDP, I start from the aggregate production function:

$$y = \left[\int_{z^*} n(z)^a dG(z)\right]^{1/a} M_0^{1-\alpha} L^{\alpha}.$$

By the envelope condition, it is sufficient to consider the direct effect of an increase in α without changing the decision variables—$n(z)$, z^*, M_o, L. The first term is increasing in α as a direct implication of Lyapunov's inequality. Since $a = 1/(1 - \alpha)$, the first term is also increasing in α. The remaining term $M_o^{1-\alpha} L^{\alpha}$ increases in α if and only if its log value, $(1 - \alpha)\ln M_o + \alpha \ln L$, does too, and this occurs if $L > M$ or equivalently, if average firm size (not including overhead) is greater than one. This is obviously the relevant empirical case.

The results of the quantitative exercise are reported in table 5. For each of the two models, the table reports the corresponding values of α for the baseline that are $\alpha = 0.65$ and $\alpha = 0.85$, respectively, and decreases α moving down in the table. The minimum values for α corresponding to average size 2.5 are 0.41 (0.48 in Hopenhayn) and 0.54 (0.64 in Hopenhayn), respectively. The large range of changes in average size are matched again with moderate changes in GDP per capita. The implied elasticities of GDP to average size range between 0.24 and 0.30 depending on the case. So again, the impact of a reduction in returns to scale α is very effective in reducing average size and has an important effect on GDP, but the latter impact is moderate compared with the data. The results in both sets of models are very similar, though the required range of returns to scale to generate the range of variation in average firm size is narrower in the Hopenhayn model.

There is one negative side effect of differences in α that could be acting as a driving force. In measuring firm-level productivities, both Hsieh and Klenow and the IDB study assume that all economies have the same returns to scale.[35] This measure is misspecified if returns to scale are different for different countries. In particular, if the true value is α_o and the value used in the calculation is $\alpha > \alpha_o$, then the ratio of measured to true productivity is $n(\alpha)^{\alpha_0 - \alpha}$. Therefore,

$$\ln(\hat{a}) = \ln(a) + (\alpha_0 - \alpha)\ln[n(a)],$$

where a is the true productivity and $\hat{a}$ measured productivity. It follows that the variance of measured productivity is

$$\begin{aligned} \operatorname{var}[\ln(\hat{a})] = {} & \operatorname{var}[\ln(a)] + (\alpha_0 - \alpha)^2 \operatorname{var}\{\ln[n(a)]\} \\ & + (\alpha_0 - \alpha)\operatorname{cov}[\ln(a), \ln(n)]. \end{aligned}$$

Recall that $n(a) =/ ba^{1/(1-\alpha_0)}$, so $\ln(a) = \ln(b) + [1/(1 - \alpha_o)\ln(a)$. This implies that

$$\operatorname{var}[\ln(\hat{a})] = \left[1 + \left(\frac{\alpha_0 - \alpha}{1 - \alpha_0}\right)^2 + \left(\frac{\alpha_0 - \alpha}{1 - \alpha_0}\right)\right] \operatorname{var}[\ln(a)],$$

and if $\alpha_o < \alpha$, then $\operatorname{var}[\ln(\hat{a})] < \operatorname{var}[\ln(a)]$. If the only source of variation between two countries is that one has the true value of α while the other has α_o, and if α is the value used to back out firm productivities, then the country with lower returns to scale would have lower measured variation of log productivity. This is again in contrast to the results found in the data, as reported above.

Final Remarks

Average firm size increases with the level of development, as established by a series of recent papers. This paper explores the correlation through the lens of a very simple model of entry and exit/selection. It considers the role

35. Hsieh and Klenow (2009); Pagés (2010).

of excessive entry, variation in the distribution of firm productivities, and variation in returns to scale to account for this fact. The results suggest that average firm size is very sensitive to these sources of variation, while GDP per capita is so to a much lesser degree. The analysis, while parsimonious, is limited in several respects. First, I rely on the class of Pareto distributions for capturing differences in the distribution of firm productivities across countries, and this is certainly not very flexible in capturing specific variations, such as lower left tails in the distribution. A more satisfying approach is to back out the distribution of firm productivity from the data.[36] Second, the model has only one period. A multi-period version would potentially give more importance to entry decisions and the role of option value. Third, these experiments represent reduced forms for policies or institutional factors that can affect the distribution of firm productivity or economies of scale. While useful as a benchmark, it leaves open the question of what specific policies might underlie these structural differences. The most promising work in this direction comes from Bento and Restuccia and from Hsieh and Klenow, among others cited in the introduction.[37]

36. As in Hsieh and Klenow (2009) and Pagés (2010).
37. Bento and Restuccia (2015); Hsieh and Klenow (2014).

References

Alfaro, Laura, Andrew Charlton, and Fabio Kanczuk. 2008. "Plant-Size Distribution and Cross-Country Income Differences." Working Paper 14060. Cambridge, Mass.: National Bureau of Economic Research.

Banerjee, Abhijit V., and Esther Duflo. 2005. "Growth Theory through the Lens of Development Economics." *Handbook of Development Economics,* vol. 1, edited by Philippe Aghion and Steven Durlauf, pp. 473–552. Amsterdam: Elsevier.

Bento, Pedro, and Diego Restuccia. 2015. "Misallocation, Establishment Size, and Productivity." Working Paper 537. University of Toronto, Department of Economics.

Blyde, Juan S., Diego Restuccia, and Omar D. Bello. 2011. "Venezuela's Growth Experience." *Latin American Journal of Economics* 48(2): 199–226.

Bollard, Albert, Peter J. Klenow, and Huiyu Li. 2014. "Entry Costs Rise with Development." Working Paper 518. Stanford University, Center for International Development.

Caselli, Francesco. 2005. "Accounting for Cross-Country Income Differences." *Handbook of Economic Growth,* vol. 1, edited by Aghion and Durlauf, pp. 679–741.

Gabler, Alain, and Markus Poschke. 2013. "Experimentation by Firms, Distortions, and Aggregate Productivity." *Review of Economic Dynamics* 16(1): 26–38.

Hopenhayn, Hugo A. 1992a. "Entry, Exit, and Firm Dynamics in Long-Run Equilibrium." *Econometrica* 60(5): 1127–50.

———. 1992b. "Exit, Selection, and the Value of Firms." *Journal of Economic Dynamics and Control* 16(3–4): 621–53.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124(4): 1403–48.

———. 2014. "The Life Cycle of Plants in India and Mexico." *Quarterly Journal of Economics* 129(3): 1035–84.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica* 50(3): 649–70.

Lucas, Robert E. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9(2): 508–23.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71(6): 1695–725.

Pagés, Carmen, ed. 2010. *La era de la productividad: Cómo transformar las economías desde sus cimientos.* Washington: Inter-American Development Bank.

Poschke, Markus. 2010. "The Regulation of Entry and Aggregate Productivity." *Economic Journal* 120(549): 1175–200.

Ranasinghe, Ashantha. 2014. "Impact of Policy Distortions on Firm-Level Innovation, Productivity Dynamics and TFP." *Journal of Economic Dynamics and Control* 46(C):114–29.

Tavares, Marina Mendes, Diego Restuccia, and José-Maria Da-Rocha. 2014. "Policy Distortions and Aggregate Productivity with Endogenous Establishment-Level Productivity." Working Paper 523. University of Toronto, Department of Economics.