

**SMALL FIRMS IN PORTUGAL: A SELECTIVE SURVEY OF STYLIZED
FACTS, ECONOMIC ANALYSIS, AND POLICY IMPLICATIONS**

Luis M. B. Cabral

New York University and CEPR

December 2005

Abstract

I survey a number of stylized facts pertaining to the dynamics of firm entry, growth and exit in competitive industries. I focus particularly on data for Portugal, though I also consider, for comparison purposes, data from other countries. I then present a series of theoretical models that attempt to explain the stylized facts and evaluate the welfare impact of market distortions. Finally, I derive a number of policy implications, all centered around the notion of economic mobility.

1. *Introduction*

Most of the study of industrial organization over the past few decades has focused on concentrated industries and firms with market power.¹ To some extent, the influx of game theory into industrial organization, for all its benefits, has created a bias away from industries where market power is of secondary importance.

A second reason why competitive industries have largely been ignored is that, from our study of the perfect competition model, there would seem to be very little more to say about industry dynamics and the properties of the long-run equilibrium beyond what we learn in an economics principles course. In fact, firm behavior under the perfect competition model is almost trivial: remain active if price is higher than the minimum of average cost; and, conditional on remaining active, choose output so that price equals marginal cost. What else is there to study?

Recent empirical studies of competitive industries reveals a wealth of empirical regularities hitherto ignored. In parallel, a series of theoretical models have attempted to explain these stylized facts. Altogether, these developments have turned the study of small-firm dynamics into an exciting area of research within industrial organization — and in fact a natural bridge between industrial organization and macroeconomics.

In this paper, I present a selective survey of this emergent literature, with an emphasis on Portugal, both from a data point of view and in terms of policy implications. I start (Section 2) with a few stylized facts regarding small firm dynamics in a variety of countries. In fact, one of the striking empirical observations is how regular the main stylized facts are across countries. I then consider (Section 3) a series of models that attempt to explain the main stylized facts. One important component of this analysis is the welfare evaluation of market distortions. This naturally leads into the next section (Section 4), where I consider policy implications of the empirical and theoretical results. The bottom line of this section is the importance of economic mobility, a broad concept that denotes the absence of distortions to the activity of small firms.

¹ See, for example, Tirole (1988).

2. *Industry dynamics: stylized facts*

What does the model of perfect competition have to say about entry, exit, and firm size?² The long-run equilibrium under perfect competition is a limit point that industries converge to by means of successive entry and exit. If active firms make positive profits, then new firms are attracted to the industry. If, on the contrary, active firms make losses, then some of those firms exit. Finally, in the long-run equilibrium, price equals the minimum of the long-run average cost. Since technology (i.e., the cost function) is the same for all firms (because of the equal access assumption), each firm receives zero supra-normal profits and there is neither entry nor exit.

Concerning the size distribution of firms, the perfect competition model is either rather extreme or extremely scant: Assuming plant-level cost functions are U-shaped, all plants must be of the same size in the long run (i.e., there is only one output level that minimizes average cost). If the managerial costs of owning more than one plant are positive, then each firm owns one plant only and all firms are of the same size. If, on the contrary, managerial costs are zero, then there exists a virtually uncountable number of possible industry configurations, all of which are consistent with the model.³

The empirical evidence from various industries with "many" small firms is widely at odds with the above view of industry dynamics: in any given period and industry, entry and exit take place simultaneously; and the size distribution of firms displays a number of regularities and is not concentrated on a single size. In the following subsections, I present data on these and related stylized facts. Before proceeding, I address an important preliminary question: how good is the available data on industry dynamics? Most of these data are based on official census or registries. This implies that they miss some of activity in the informal economy. Since much of my analysis compares Portugal to other countries, the presence of the informal economy is a concern to the extent that it varies greatly across countries.

² The following paragraphs are adapted from Cabral (2000).

³ In fact, there may be economies of scale in multi-plant firms—e.g., large-scale purchasing discounts—that counteract increased managerial costs. But this would take back to a U-shaped cost curve and the prediction that all firms have the same number of plants. See Lucas (1978).

Figure 1 plots the values of the share of the informal economy and the level of Gross National Income (GNI) for all countries with a per-capita GNI greater than US\$5,000. The figure suggests that Portugal is not particularly different from other countries.

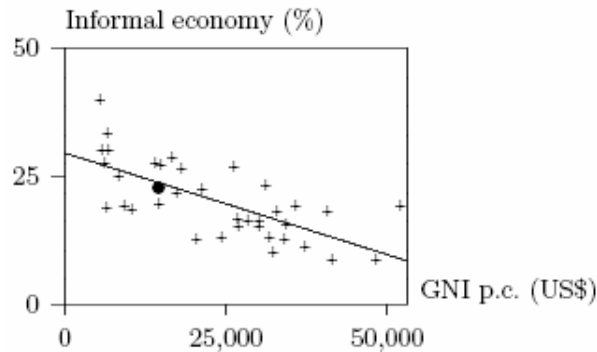


Figure 1: Informal economy and per-capita GNI (countries with per-capita GNI greater than \$5,000). The data point for Portugal is represented by a bullet point.

Source: World Bank.

In fact, correcting for GNI p.c. level, Portugal seems to be right on average.⁴ Insofar as we compare Portugal to developing countries (as I will do for the most part in this paper) then we should be aware of the potential bias introduced by the informal economy, about 23% in Portugal versus 18% for countries with per-capita GNI greater than US\$10,000.

2.1. *Entry, growth and exit*

One of the most robust stylized facts of competitive industry dynamics is that, in any given year, entry and exit occur simultaneously. To put it differently, net entry rates are a small fraction of gross entry rates. Moreover, one finds that entry and exit rates are highly correlated across industries; that is, industries with higher than average entry rates also exhibit higher than average exit rates.

⁴ The line corresponds to the regression of the share of the informal economy with respect to GNI per capita. The regression has an R^2 of 0.49; the GNI per capita coefficient has a p value of 1.4E-06.

Cable and Schwalbach (1991) developed one of the earliest surveys of studies on entry and exit in different countries. Table 1 includes some of the results in their Tables 14.1 and 14.2. As can be seen, entry and exit rates are much higher than net entry rates (the difference between gross entry and gross exit rates). One possible interpretation for this phenomenon might be that it is an artifact of aggregation: some industries might have high entry rates while other industries have high exit rates. However, the data on cross-industry correlation between entry and exit rates suggests that this is not the case. Notice moreover that the results persists at the 3-, 4- and 5-digit industry classification level. Portugal, interestingly, exhibits some of the higher values of entry and exit rates.

Table 1: Annual gross entry and exit rates (in %) and correlation.

Country	Entry	Exit	Corr.	Period	Data*
Portugal	12.3	9.5	.030	83-86	234/5/E/E
Belgium/Man	5.8	6.3	.660	80-84	130/3/E/E
Belgium/Serv	13.0	12.2		80-84	79/3/E/E
Canada	4.0	4.8	.039	71-79	167/4/E/S
FRG	3.8	4.6	.342	83-85	183/4/F/S
Korea	3.3	5.7	**	76-81	62/4,5/F/S
Norway	8.2	8.7	.488	80-85	80/4/F/S
U.K.	6.5	5.1	.318	74-79	114/4/F/S
U.S.	7.7	7.0	.270	63-82	387/4/F/S

(*) Number of industries/aggregation level (no. digit industries)/ Firm or Establishment level/Employment or Sales data.

(**) -.409 in 1976-1978, +.350 in 1979-1981.

Source: Cable and Schwalbach (1991).

Table 2 presents data from a more recent source. The second column, turnover rate, corresponds to the sum of entry and exit rates in term of employment. The entry rate, for example, is given by the total number of new jobs created by entrants divided by the total number of workers in the industry. The correlation rates between entry and exit are also weighted by employment level.

Notice that turnover rates are quite different in Tables 1 and 2. For example, Table 1 implies a turnover rate (in terms of number of firms) of 21.8% for Portugal, whereas Table 2 shows a much lower value, 9.3%. The main reason for the discrepancy is that Table 2 presents employment-weighted rates, whereas Table 1 refers only to the number of firms.

Table 2: Turnover rate (employment-weighted) and correlation between entry and exit rates.

Country	Turnover	Correlation
Portugal	9.3	.64
United States	7.0	.86
Western Germany	3.9	.87
France	7.0	.73
Italy	8.6	.53
Denmark	10.2	.75
Finland	11.9	.75

Sources: Bartelsman et al (2003); Bartelsman et al (2004).

Table 3: Entrants and exiters relative size (as a percentage of incumbent firms)

Country	Entrants	Exiters
Portugal	8.0	11.8
Belgium/Manufacturing	28.5	21.3
Belgium/Services	32.8	32.2
Canada	9.6	7.8
FRG	22.1	18.8
Korea	12.1	
Norway	12.6	11.3
U.K.	44.9	61.2
U.S.	6.7	6.9

Source: Cable and Schwalbach (1991).

The lower rates in Table 2, compared to Table 1, suggest that entrants and exiters are of smaller size than incumbent firms. In fact, a second important stylized fact is that market penetration rates are a small fraction of entry rates. Table 3 includes data on entrants and exiters relative size in different countries. In all countries, the entrants initial size is less than one-half the industry average size. In 6 of 8 countries, it is less than one

quarter. Except for the U.S., Portugal has the lowest ratio of entrant and exiter size with respect to industry average size.

Turning to the dynamics of each individual firm, a third stylized fact is that survival rates tend to be increasing in firm size and in firm age. For example, Mata and Portugal (1994) estimate baseline hazard rates (that is, conditional probabilities of exit) after t years of .19, .14, .12 and .11 for the first four years. Based on the same dataset, Mata et al. (1995) also show that the probability of survival is increasing in current size. As with the previous stylized facts, similar results have been obtained for other countries. For the U.S., Evans (1987) estimates that a 1 percent change in firm size and a 1 percent change in firm age lead, respectively, to a 7 and a 13 percent change in the probability of survival over a 5-year period.

To conclude this subsection, I present a stylized fact regarding firm growth: growth rates are typically decreasing in size, especially for small size levels; and decreasing with age. Evans (1987) estimates that, for the U.S. and over a 10-year period, a 1 percent increase in initial size leads to a 0.68 percent increase in ending-period size, that is, growth is considerably less than proportionate. He also finds that, over the same period, a 1 percent increase in initial age implies a 1.42% decrease in final size. Hall (1987) derives qualitatively similar results based on a different sample of U.S. firms. Finally, Dunne and Hughes (1994) estimate that, over a 5-year period, a 1 percent increase in initial size leads to a 0.93 percent increase in ending-period size (this would correspond to $0.93^2 \approx 0.86$ over a period of 10 years; compare with Evans' results).

To summarize, this subsection depicts a typical industry as having many entrants and exiters each period. A typical entrant is smaller than the industry average and grows faster than the industry average. Entrants are more likely to exit than older entrants, especially when they remain small in size. Finally, these facts are fairly robust both across industries and across countries. In particular, Portugal seems fairly typical.

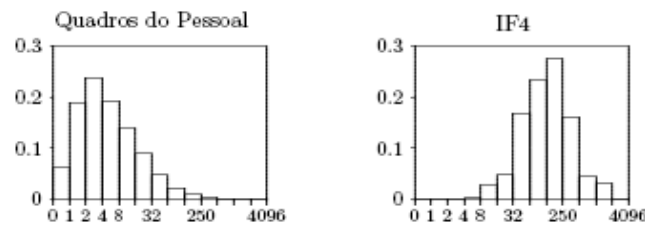


Figure 2: Firm size distribution in Portugal based on two different datasets.

Source: Cabral and Mata (2003).

2.2. Firm size

The patterns of the firm size distribution depend critically on the type of data source one considers. Cabral and Mata (2003) show that, when considering the universe of Portuguese manufacturing firms, the size distribution is skewed to the right. This contrasts with the distribution of firms from commonly used databases, such as Compustat or Amadeus, for which the lognormal is a better approximation. Figure 2 illustrates this contrast. On the left we have the size distribution based on a comprehensive dataset (from Quadros do Pessoal). On the right, the distribution from a dataset with similar characteristics to Compustat and Amadeus. One cannot reject that the right-hand distribution is lognormal, but the one on the left is certainly more right-skewed than a lognormal.

The lognormal distribution has traditionally played an important role in the study of the firm size distribution. Empirically, the distribution seems to fit well data from Compustat and similar databases. Theoretically, the lognormal is the limiting distribution of a firm growth process such that growth rates are independent of firm size.⁵ When one looks at more comprehensive datasets, including small and micro firms, then lognormality is no longer the rule; nor are growth rates independent of size (see the previous section).

⁵ See Gibrat (1937), Sutton (1997).

The pattern of right-skewness, with proportionately more small firms than large firms (with respect to the lognormal distribution), is not unique to Portugal. Figure 3 depicts the distribution of firm size for manufacturing countries in Portugal and the U.S. Despite the significant differences in size and level of economic development, the distributions are relatively similar.

In recent research, Bartelsman et al (2003) present some systematic evidence on firm size and firm size distribution across a series of countries. Table 4 presents the relative weight of firms with fewer than 20 employees. As can be seen, the number of firms in that category is virtually identical in Portugal and the U.S. (consistently with Figure 3).

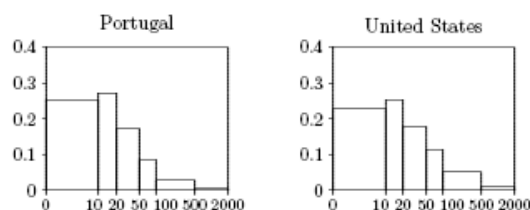


Figure 3: Firm size distribution in Portugal and in the U.S. Source: Cabral and Mata (2003).

Table 4: Relative importance of firms with fewer than 20 employees: fraction of total economy (%).

	Number of firms	Number of employees
Portugal	86.3	27.7
United States	86.7	16.6
Western Germany	87.9	23.6
France	78.6	34.4
Italy	93.1	38.1
Denmark	90.0	30.2
Finland	92.6	25.8
Netherlands	95.8	31.2

Source: Bartelsman et al (2003).

Table 5: Average firm size (number of employees).

	Economy	Manufacturing	Services
Portugal	16.8	31.0	11.4
United States	26.4	80.3	21.4
Western Germany	17.0	39.1	11.5
France	33.5	32.1	35.7
Italy	10.5	15.3	6.8
Canada	12.7	40.5	12.0
Denmark	13.3	30.4	12.7
Finland	13.0	27.8	9.9
Netherlands	6.5	18.3	5.3

Source: Bartelsman *et al* (2003).

However, when we consider the fraction of employees accounted for by firms with fewer than 20 employees, then the number is much lower for the U.S. This reflects the fact that the right-tail of the U.S. distribution of firm size is much thicker.

Table 5 presents data on average firm size in various countries. Again, Portugal seems broadly in line with other European countries. Table 5 also presents data on the subcategories Manufacturing and Services. The cross-country patterns of firm size seem broadly robust to the type of firms considered. In fact, Bartelsman *et al* (2003) show that industry composition accounts for a small fraction of the cross-country differences in firm size.

The results from Table 5 may seem puzzling. After all, Portugal's largest firms pale in size when compared to their European counterparts; and accordingly one might expect average firm size to be smaller in Portugal (and the employment share of small firms to be larger). The solution to this apparent puzzle, at least to some extent, is to be found in what we might call the extreme-statistic fallacy. Consider two random variables, X and Y . Let $E(X, m, n)$ be the average value of the m highest values of x out of a sample of n values. Then, even if the distribution of X and Y were the same, we would inevitably obtain $E(X, m, n_X) > E(Y, m, n_Y)$ so long as $n_X > n_Y$. So the fact that the five largest German firms are greater than the five largest Portuguese firms does not necessarily imply that German firms are on average greater than Portuguese firms; just like the fact the five tallest German men are taller than the five tallest Portuguese men does not imply Germans are on average taller.

In summary, Portugal does not seem to be an outlier in terms of firm size and the firm size distribution. It is often said that there are too many small firms in Portugal, and that average firm size is too small. However, when looking at SME it is hard to find a significant difference between Portugal and the average OECD country.

2.3. Productivity

There is extensive evidence, though not extensive systematic evidence, that productivity in Portugal is substantially lower than the average of all European countries. A report by McKinsey benchmarks Portugal with reference to a group of four other European countries and estimates a productivity gap of about 30%.

I am not aware of any systematic estimate of the distribution of productivity levels by industry in Portugal. Estimates for the U.S. indicate high indices of variability. Considering how similar different countries are on other dimensions, it seems reasonable to expect Portuguese industries to exhibit similar levels of dispersion of productivity levels.

An estimate based on a survey of Portuguese innovators indicates a coefficient of variation greater than 1.⁶ Such high value of the coefficient of variation has important implications. In particular, it makes little sense to use the difference in means and generalize that Portuguese firms are not productive: the fallacy of the average.

A simple calculation will help make the point. Assume that the distribution of productivity levels is lognormal (which makes sense based on the distributions in other countries). Assume moreover that the standard deviation is equal to the mean. Then even if Portuguese firms have an average productivity 50% lower than the European average, 20% of Portuguese firms have a productivity level higher than the European median; or 11% above European average.

These calculations are illustrated in Figure 4. In this figure, I normalize units so that the average productivity level of European firms is 1. I assume that both distributions are lognormal with a coefficient of variation of 1. As can be seen, about 10% of the mass of Portuguese firms lies to the right of 1 (the average). Since the median of the lognormal is

⁶ Pedro Conceição (private communication).

lower than the average, the percentage of Portuguese firms above the median is even higher.

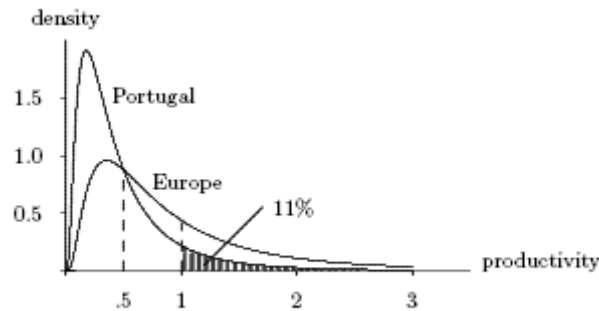


Figure 4: Distribution of productivity levels: an example.

One of the benefits of competition is that it leads to an efficient allocation of resources across firms within an industry. In particular, firms with higher productivity levels will tend to increase in size, whereas firms with lower productivity level will tend to decrease in size or exit. In this regard, it is useful to measure the gap between weighted and un-weighted productivity. If this gap is zero, we conclude that there is little correlation between productivity and size. If this gap is positive, however, we conclude that firms with higher productivity levels are larger, reflecting the benefits of competitive selection.

Table 6 displays the values of this index. Specifically, I follow Olley and Pakes (1996) in defining the gap as

$$G = \left(\sum_i s_i P_i \right) / \left(\frac{1}{N} \sum_i P_i \right) - 1,$$

where s_i is firm i 's share (in number of employees), P_i is firm i 's productivity level, and N is the number of firms.

Considering the values for other OECD countries, the value for Portugal is fairly high, though smaller than in the U.S. Two other noticeable differences are the very high values for southeast Asian countries and the very small (even negative) values for eastern European transition economies. This suggests that, particularly in these countries, there is much to gain from reallocating resources across firms within each industry.

At this point, it is worth noting the caveat with which I started my presentation of empirical evidence: the numbers I present are based on official

Table 6: Gap between weighted and un-weighted labor productivity (1990s) and share of informal economy.

Country	Productivity gap	Informal economy
Portugal	.45	22.6
United States	.60	8.8
Western Germany	.34	16.3
France	.30	15.3
U.K.	.20	12.6
Finland	.27	18.3
Netherlands	.28	13.0
Argentina	.21	25.4
Chile	.46	19.8
Colombia	.47	39.1
Taiwan	.58	19.6
Korea	.60	27.5
Indonesia	.62	19.4
Slovenia	-.05	27.1
Latvia	.02	39.3
Romania	.05	34.4
Hungary	.15	25.1
Estonia	.33	N/A

Source: Bartelsman et al (2005) and World Bank, respectively.

Table 7: Barriers to economic activity.

Index	Unit	Best	Portugal	Worst	Rank
Proc. enforce contract	number	0	22	62	72/142
Time to enforce contract	days	7	365	1460	104/141
Time to resolve insolvency	years	0	2.6	11.3	54/126
Time to start a business	days	2	95	215	126/146
Per capita GDP	1995 \$	54,652	16,039	483	45/186

Source: World Bank Indicators (2003).

statistics, and thus miss much of the informal economy. Insofar as "informal" firms have lower than average productivity, the real gap is lower than the one reported in Table 6. Since Portugal exhibits a higher fraction of informal economy activity, the difference

between the productivity gap for Portugal and other European countries may not be so high—in fact, it may well be negative.

2.4. *Distortions to economic activity*

It is common wisdom that the Portuguese economy is subject to myriad economic distortions, and that Portuguese firms must overcome numerous barriers to succeed. In Table 7, I summarize some of the information from the World Bank's World Economic Indicators on indices that refer to the economic environment in which firms operate.

One must take these numbers with a grain of salt. For example, the number of days to enforce a contract in Portugal, 365, suggests that the number was not calculated with great precision; another suspicious example is that the lowest number of procedures to enforce a contract is Indonesia at zero (Australia follows with 11); and so forth.

I also include in Table 7, pro memoria, the level of GDP per capita. It is noticeable that Portugal is an outlier in the sense that the ranking in terms of barrier indices is significantly lower than that in terms of per capita GDP.

An alternative source of data on entry barriers is the World Bank's database on doing business. Table 8 presents a few variables of interest, both for Portugal and for the remaining countries in the sample.⁷

⁷ See Djankov et al's (2002) for various notes on this dataset. Note that there are some discrepancies between Tables 7 and 8 regarding the data for Portugal. First, the values of GDP per capita are different; however, one must consider that the value is measured for different years and in different units. Second, the time to start a business is lower in Table 8; but here the measure is in business days, not calendar days, so the difference is not that great

Cost of starting a business (% of GNI per capita)

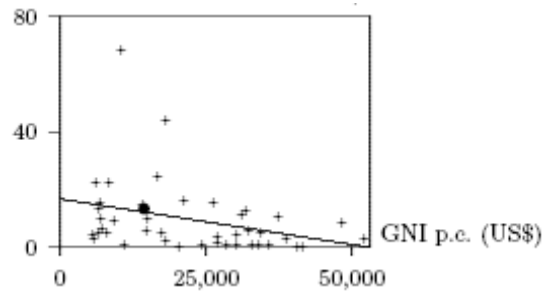


Figure 5: Cost of starting a business and per-capita GNI (countries with per-capita GNI greater than \$5,000). The data point for Portugal is represented by a bullet point. Source: World Bank.

Time to start a business (days)

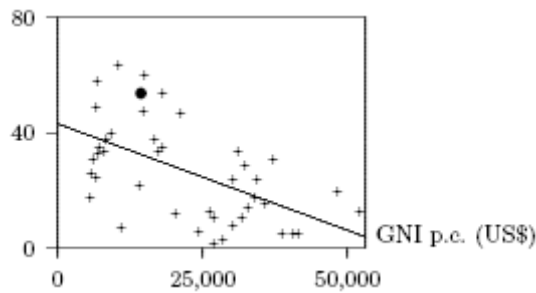


Figure 6: Time to start a business and per-capita GNI (countries with per-capita GNI greater than \$5,000). The data point Portugal is represented by a bullet point. Source: World Bank.

Table 8: Entry costs. Source: Djankov et al (2002)

Description	Portugal	Sample Average
Number of procedures to obtain legal status, by category:		
Safety and health	12	10.48
Environment	0	.34
Taxes	2	2.04
Labor	2	1.94
Screening	8	6.04
Time to obtain legal status (days)	76	47.7
Cost of obtaining legal status	.1844	.4708
Total cost	.4884	.6598
GDP per capita	10,600	8,226

In order to investigate the possibility that Portugal is an outlier in terms of distortions to economic activity, I consider two graphs, Figures 5 and 6. In these graphs, I select countries with a Gross National Income (GNI) higher than \$5,000 and plot the cost of starting a new business (Figure 5) and time to start a business (Figure 6), as well as the level of each country's GNI. I also plot the regression line between the two variables.

Both figures suggest that barriers to entry tend to be smaller in more developed countries. Regarding the cost of entry, Portugal seems to be right on average.⁸ However, regarding time to start a business, Portugal is very much on the outer edge of the distribution. In summary, bureaucratic barriers to entry are very large in Portugal, even when controlling for the level of economic development.

Similar calculations can be performed on the sample of 85 countries considered by Djankov et al (2002). Table presents some summary statistics. Broadly speaking, the numbers for Portugal don't seem too far off the sample average. What is particularly striking, however, is the fact the total cost in Portugal is 2.64 times higher than the monetary cost of entry, whereas the ratio of the averages is only 1.40. In other words, the

⁸ The lines in each figure are the estimated value from regressing the vertical-axis variable on the horizontal-axis variable. The values of R² are 30 and 10%, respectively; the coefficients relating the two variables in each graph have p values of .003 and .000, respectively.

entry cost in Portugal seems disproportionately due to the cost of time in getting the necessary approvals for entry.

Table 9: Indexes of Product Market Regulation and Barriers to Entrepreneur-ship.

Concept	Country	1998	2003
PMR index	Portugal	2.1	1.6
	OECD average	2.1	1.5
BTE	Portugal	1.8	1.3
	OECD average	1.9	1.5

Source: Conway et al (2003).

A separate and very different source of data is the OECD data set on product market regulation.⁹ Table 9 presents values for the overall index (Product Market Regulation) as well as a component index, "Barriers to Entrepreneur-ship," which has a weight of approximately 30% in the overall index. Two facts stand out from the table: first, the recent trend towards a lower degree of regulation.¹⁰ Second, the relatively small difference between Portugal and the OECD average.¹¹

3. *Industry dynamics: economic analysis*

Good policy analysis must be based on good knowledge of data and stylized facts; but it must also be grounded on a coherent model of reality. In this section, I summarize some of the recent developments in the economic analysis of industry dynamics, especially as it addresses the stylized facts presented in the previous section: industry turnover, firm size, productivity. I will conclude with the analysis of how market distortions impact the dynamics of firm entry and exit.

⁹ See Conway et al (2003).

¹⁰ This tendency is marginally more pronounced in the component "state control."

¹¹ In 2003, the standard deviation of the PCR and BTE indices was 0.43 and 0.42, respectively, so the differences 0.1 and 0.2 are economically and statistically small.

3.1. *Explaining industry turnover*

In order to explain the stylized facts described in the previous section, we need to relax some of the assumptions of the model of perfect competition. I will maintain that (i) firms are price takers; (ii) the product is homogeneous; (iii) information about prices is perfect. However, in contrast with the perfect competition model, suppose now that: (iv) firms must pay a sunk cost in order to enter; (v) not all firms have access to the same technology.

Specifically, suppose that *different firms have different degrees of efficiency*, which in turn correspond to different cost functions: more efficient firms have a lower marginal cost schedule. These differences may result from a variety of factors. For example, some managers are more efficient in organizing resources than others (more on this below).

Suppose moreover that *each firm is uncertain about its own efficiency*. When a firm first enters an industry, it has only a vague idea of what its efficiency is. As time goes by, and based on each period's experience, the firm gradually forms a more precise estimate of its true efficiency. In each period, the firm chooses optimal output based on its current expectation of efficiency; roughly speaking, the output level such that price is equal to expected marginal cost.

Given the above elements, we conclude that firms which get a series of bad signals (high production costs) gradually become "pessimistic" about their efficiency level, gradually decrease their output, and, eventually, may decide to exit the industry (as variable profit is insufficient to compensate for the fixed cost). By contrast, firms that receive a series of good signals (low production costs) remain active and gradually increase their output.

This model of *competitive selection* is consistent with several of the stylized facts described in the previous section. In particular, the model is consistent with the stylized fact of *simultaneous entry and exit in the same industry*. Firms which accumulate a series of very unfavorable productivity signals hold a very unfavorable estimate of their own efficiency. As a result, their expected value from remaining active is negative, which in turn leads them to exit. New entrants have no information regarding their efficiency. Their expected efficiency is therefore much better than the exiting firms': no news is better than

bad news. This justifies that their expected value from being active is positive, in fact, greater than the entry cost. In summary, it is possible for a firm with no information about its efficiency to enter while a firm with unfavorable information about efficiency exits.

Efficient firms are firms with a low marginal cost function. Since firms equate price to (expected) marginal cost, it follows that more efficient firms sell a higher output. Together with the previous results, this implies that exiters (the active firms with lowest expected efficiency) are also the firms with lower output. By selection, the firms that remain active have an efficiency higher than average. In particular, higher than the average entrant's. It follows that entrants' output is lower than the surviving firms' average output. In this way, the model is also consistent with the stylized fact that *firms that enter and firms that exit are smaller than average*.

Finally, the competitive selection model is also consistent with the empirical observation that the firm size distribution is neither single-valued nor indeterminate, as the perfect competition model would imply. In fact, a given population distribution of efficiency levels implies a particular distribution of firm sizes.¹²

3.2. *Explaining the evolution of new firms*

As mentioned in Section 2.2, new firms are characterized by a very right-skewed distribution. As time moves on, the size distribution of that cohort shifts to the right and becomes closer to a log-normal. One first natural interpretation of this pattern is that the smallest of the small firms are exiting in greater numbers, leaving a set of survivors that has a more symmetric distribution (in particular with fewer very small firms). However, using Portuguese data Cabral and Mata (2003) reject this interpretation. Figure 7 plots three curves: the size distribution of 1984 entrants; the 1991 distribution of 1984 entrants who survived until 1991; and the 1984 size distribution of the 1984 entrants that survived until 1991. As can be seen the first and the third densities are fairly similar, which contradicts the hypothesis that selection is doing most of the work.

¹² It may be worth to point out that the competitive selection model does not depend on firms being asymmetric with respect to costs. We could alternatively assume that some firms' products are better than others'.

Cabral and Mata (2003) propose an alternative hypothesis, namely that financing constraints play an important role. Among the set of entrants, some firms start off very small because they don't have the resources to start off at their efficient level. As they gradually become less financially constrained, the distribution moves to the right and becomes more symmetric (as found in the data).

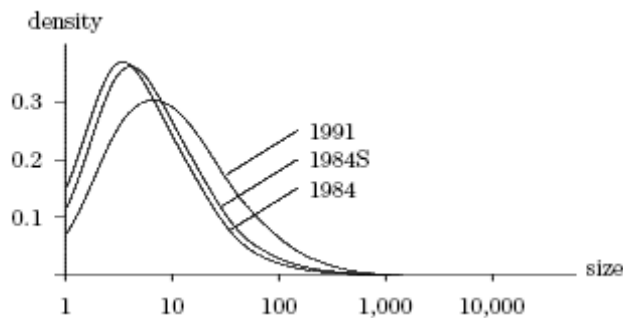


Figure 7: Size distribution of the 1984 cohort of entrants: densities based on 1984 and 1991 data as well 1984 data for the firms that survived through 1991.

Barrios et al (2005) corroborate this explanation for data from Ireland. They show that the above evolution of firm size skewness is found in Irish firms but not in multinational firms. This seems consistent with the hypothesis that multinational firms are less financially constrained and therefore not subject to the above dynamics. However, Fagiolo and Luzzi (2004), using Italian data, reject the hypothesis that financing constraints have a significant effect. Based on firm level survey data where respondents indicate if they are financially constrained, they construct two distributions of new firm size distribution: that of financially constrained firms and that of not financially constrained firms. The difference between the two distributions is minimal.

An alternative explanation for the evolution of firm size focuses on the role of sunk costs. Cabral (1995) provides an theoretical explanation for the negative relation between firm size and firm growth among new entrants. The idea is that capacity and technology choices involve some degree of sunkness (that is, investments for which value is foregone upon exit). Since small entrants are more likely to exit than are large entrants, it is optimal for small entrants to invest more gradually, and thus experience higher expected growth rates upon entry than do large entrants.

3.3. *Explaining variability in productivity*

The characterization of the firm size distribution provided by Jovanovic's model is, to a great extent, tautological: the distribution of efficiency levels is assumed rather than derived; a more satisfactory model would also explain the distribution of efficiency levels. One possibility is to assume that firms invest in R&D and that efficiency levels result from these R&D investments. This is what the model by Ericson and Pakes (1995) does.

Ericson and Pakes (1995) consider a model where each firm's productivity results from rationally chosen R&D investments. Specifically, in each period firms decide whether to remain active and, if so, how much to invest in R&D. Investment leads to a stochastic improvement in the firm's type, which can be interpreted in a way similar to Jovanovic (1982).¹³

Ericson and Pakes (1995) show that there exists a unique rational expectations Markov equilibrium; moreover, this equilibrium is ergodic. In the long-run, there will be a distribution of firm types and firm sizes. In other words, instead of assuming a distribution of types, as in Jovanovic (1982), we now derive such distribution as an equilibrium result. The key to heterogeneity of firm types is therefore the randomness of the investment process.

It seems reasonable to assume differences in efficiency and productivity result from luck of the draw in the R&D process. Still one may wonder how such large differences in productivity as those reported in the previous section can be sustained in the long run. After all, one of the main effects of competition is precisely to weed out the under-performers. One possible explanation, advanced by Syverson (2004a), is that product differentiation allows firms to survive with below-par productivity levels. Consistently with this explanation, Syverson (2004a) shows that industries with greater degree of product differentiation are also industries with greater variability in productivity.

¹³ Ericson and Pakes (1995) also consider the possibility of non-competitive behavior by firms.

An alternative explanation is that not all industries are that competitive. If the more productive firms price above cost then less efficient firms may be able to survive (if with lower margins). Syverson (2004b) tests this hypothesis in the concrete industry. He shows that in geographical areas where there is less competition there is also greater variability in productivity levels.

A related test is that of Asplund and Nocke (2005). They show that in geographical markets where demand density is greater, and thus market competition more intense, the average life span of an entrant is lower. While they don't directly present theoretical or empirical results on productivity level, this result is consistent with Syverson's (2004a,b) on the disciplining effects of competition.

In summary, firm heterogeneity (namely in terms of productivity level) is a combination of exogenous firm attributes (managerial ability, company culture, etc) and luck in the investment process.¹⁴ Although typical industries exhibit a significant degree of variation in productivity levels, such dispersion is smaller in more competitive industries (e.g., industries with lower degree of product differentiation).

3.4. Market distortions and welfare

The fundamental theorem of welfare economics states that, under perfect competition, the market solution is efficient. But, as we've seen in Section 2, the perfect competition model does not stand up to the facts very well. What then can be said of the market efficiency result? To my knowledge, Hopenhayn (1992) provides the most general extension of the fundamental theorem. Based on an extension of Jovanovic's (1982) model, he shows that the market equilibrium is efficient if firms are price takers — even if efficiency varies across firms and across periods. This is a strong result. In particular, if there are no entry barriers then the equilibrium level of firm turnover (possibly with high firm turnover compared to the net entry rate) is socially optimal.

¹⁴ Pakes and Ericson (1998) test the relative importance of these two sources of heterogeneity. They show that firm type is ergodic in manufacturing but not in services. This is consistent with the interpretation that Jovanovic's (1982) story does a better story at explaining the dynamics of firms in the services sector, whereas Ericson and Pakes (1995) is a better model of firms in manufacturing.

In other words, for all its differences with respect to perfect competition, competitive selection maintains one important property: efficiency. So long as firms act as price takers, the equilibrium solution, absent any artificial barriers, is socially efficient. Each firm's output decision in each period is efficient: price equal to expected marginal cost is the most efficient output decision, that is, the one that maximizes total surplus. Moreover, it can be shown that the firm's entry and exit decisions are also optimal from a social point of view. The basic idea is the same as in the model of perfect competition: a very small firm has a negligible impact on other firms and on price. It follows that it internalizes all of the costs and benefits from entering or exiting the industry: what is good for the firm is good for society.

It might seem inefficient to have firms entering and exiting a given industry simultaneously. But we must remember that firms are uncertain about their efficiency. The only way to determine a firm's efficiency is to actually enter the industry. A central planner who attempted to maximize total surplus would not be able to do better than the market. In summary, it can be shown that the equilibrium under competitive selection is efficient.

It follows that distortions to the natural workings of the market lead to a lower level of social welfare. A particularly careful illustration of this point is provided by Hopenhayn and Rogerson (1993). Based on the Hopenhayn's (1992) model of competitive selection, Hopenhayn and Rogerson (1993) estimate the impact of a distortion to the process of firm creation and destruction. Specifically, they consider the impact of firing costs. The results are staggering: for example, they estimate that a tax on dismissals equivalent to 1 year's wages reduces steady-state utility by over 2% measured in terms of consumption. The welfare loss comes about from a 8% reduction in firm turnover. This implies that less efficient firms are active whereas other, more active firms, remain inactive. Also, the labor adjustment cost implies that some firms are smaller or greater (in terms of number of employees) than it would be efficient. Hopenhayn and Rogerson (1993) estimate that, for more than 90% of the firms, the gap between the marginal productivity of labor and wage would be greater than 5%.

I suspect that Hopenhayn and Rogerson's (1993) analysis yields a lower bound of the welfare loss from barriers to entry and mobility. The reason is that they assume all firms are subject to the same barrier. But casual observation suggests that, just as there is

significant variability in firm productivity, there is also significant variability in the barriers to entry and mobility that each firm faces.

In order to get an idea of the magnitude of this effect, consider the following simple model. Suppose there are 1,000 price taking firms each with capacity 1.¹⁵ I assume that marginal cost is constant and normally distributed with mean 100. Consistently with the empirical evidence, I assume a coefficient of variation of 1, so standard deviation is also 100. Market inverse demand is given by $p = 200 - .01Q$, where Q is total output (number of active firms).

I first compute the equilibrium in this economy. This amounts to ordering firms by marginal cost, thus obtaining market supply; and then finding the supply-demand equilibrium.

Suppose now that this economy is subject to a series of distortions. Specifically, firm i 's marginal cost is changed by t_i , where t_i is normally distributed with mean μ and standard deviation σ . I assume the value $t_i q_i$ is a transfer, so the only social cost implied by σ is the distortion it creates.

The social cost of distortion t_i can be divided into two terms. First, assuming no change in costs, t_i leads to a gap between price and marginal cost and the corresponding Harberger excess burden triangle (allocative inefficiency). Second, for a given output level, t_i also implies an increase in production cost compared to the minimum total production cost (productive inefficiency).

¹⁵ The model can easily be extended to firms with different capacities; simply assume that some firms have multiple establishments, each consisting of one of the "firms" that I consider.

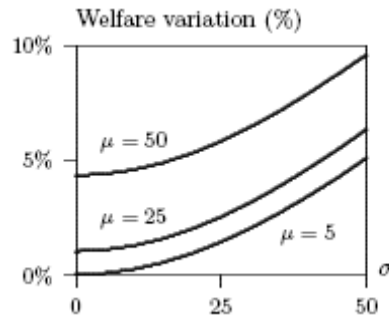


Figure 8: Welfare loss from cost distortions. Each firm's cost is increased by a stochastic variable with distribution $N(\mu, \sigma)$.

Figure 8 shows the welfare effects of distortions for various values of μ and σ . Consider first the case when $\sigma = 0$. As we increase the value of μ , the welfare loss increases. Notice that the relation between μ and welfare loss is convex. If σ is low, then the welfare loss is limited to allocative inefficiency, which is proportional to $t_i^2 = \mu^2$. It follows that a small value of μ implies a very small welfare loss. This is, in essence, the underlying argument in Harberger's (1954) claim that the welfare loss from market distortions in the U.S. is very small.

Figure 8 suggests that the welfare loss is also convex with respect to σ . For example, if $\sigma = 50$, then welfare is as large as when $\mu = 50$ and $\sigma = 0$. In fact, even if μ is very small, a high value of σ implies a high welfare loss.

In summary, the welfare cost of distortions depends not only on the size of distortions (μ) but also on how these vary across firms (σ). There are various sources of cross-firm variation, including tax evasion and more generally the informal economy. I will return to these in the next section.

4. Policy implications

The theoretical ideal of perfect competition includes, among others, the presumption that firms are price takers and that there are no barriers to mobility. While the assumption of price-taking behavior is a reasonable approximation to many industries, the ideal of a level playing field with no barriers to entry and mobility is far from reality.

Competition policy has made great inroads in the treatment of the classical "problematic" cases: mergers and acquisitions in concentrated industries, public utilities,

and so forth. But a lot still needs to be done in a variety of so-called "competitive" industries. Here is where, in my opinion, the main focus of microeconomic policy should lie.

In this section, I further develop this point. First, I focus on the idea of industry turnover as a characteristic of the normal behavior of the economy. Next I argue that turnover is actually a very important channel for the process of industry productivity growth. Finally, I present the ideal of economic mobility as one of the main goals of competition policy.

4.1. Turnover and welfare

One of the most common misperceptions regarding micro policy is that it should protect firms from failing and exiting. It is true that the immediate effect of saving a firm from exiting is to save a number of jobs equal to that firm's employment. But such policy would imply a significant welfare cost in terms of resource misallocation. A firm that's artificially kept alive implies a firm that will not be created, knowing that the latter would probably be more efficient than the former. In this regard, the welfare cost estimates by Hopenhayn and Rogerson (1993) are particularly telling.

One variation of the above misperception is that public policy should protect small firms especially, since their turnover rate is typically higher than average (as documented in Section 2). But again, a high turnover rate may well be part of the natural process of experimentation inherent to a healthy competitive industry. In fact, the cost of experimentation is smaller for small firms, and so it is only natural that in equilibrium we observe higher entry and exit rates for small firms. So, as a matter of principle one cannot say that because their turnover rate is higher small firms should be particularly protected.

Having said that, it is true that some market imperfections hit small firms in a particular way. For example, imperfect credit market conditions are likely to bias the market against small firms. In general, any imperfections that increase a firm's fixed cost slant the field against small firms. The solution is then to correct as much as possible for those credit market imperfections.

4.2. Productivity growth

One of the most important results from the analysis of time series productivity data is the importance of industry turnover in the process of productivity growth. There are essentially two ways in which to increase average productivity in a given industry. One is to increase the productivity level of each firm; the other one is to increase the relative weight of higher productivity firms. Public policy frequently heralds the former and places less emphasis on the latter. Governments programs to improve the quality of human and physical capital inputs have greater political impact; whereas fostering a more fluid process of firm entry and exit, if anything, carries a political cost. This is unfortunate, for the empirical evidence is that resource reallocation is the primary source of productivity growth.

Olley and Pakes (1996) offer an interesting case study of the role of turnover in productivity growth. They look at the 1980s deregulation process in the U.S. telecommunications equipment industry. The common wisdom is that deregulation improves efficiency by forcing incumbent firms to become more efficient. Although telecommunications equipment firms indeed became more efficient following deregulation, the greatest source of industry improvement was the process of capital reallocation amongst incumbent firms.

The exercise of productivity growth accounting leads to different results in different economies and industries; it also depends on the particular definition of productivity.¹⁶ However, the common message of all of these exercises is the importance of the process of industry turnover, both in terms of entry and exit and in terms of resource reallocation among incumbents.

This naturally leads to the question, What can public policy do about it? I next argue that one of the primary areas of competition policy ought to be guaranteeing the basic conditions of economic mobility: a level playing field where competition may work effectively towards the selection of the more efficient firms.

¹⁶ See Ahn (2001) for a survey.

4.3. *Economic mobility*

The Fundamental Theorem of Welfare Economics is remarkably robust. Although it is usually formulated in the context of perfect competition, which is based on a lot of assumptions, one only really needs to assume (a) price taking behavior and (b) free entry, exit and mobility. In particular, as shown by Hopenhayn (1992), market efficiency is consistent with firm heterogeneity and imperfect information.

Most firms in the Portuguese or any other economy, especially small and medium-sized firms, face a fairly flat demand curve. It follows that price taking behavior is a fairly good first-order approximation. We are thus left with condition (b). Unlike price-taking behavior, the ideal of economic mobility is very far from the reality of the Portuguese and most other economies.

Contrary to the increasingly discredited advocates of a strong industrial policy (e.g., government investment in national champions), a growing majority of economics scholars advocates that the best microeconomic policy is a good competition policy. And within micro policy I would advocate that promoting economic mobility plays a central role. Competition policy typically focuses on concentrated industries, industries where firms have market power. Competitive industries (that is, industries where firms are price takers) are frequently given less importance. But the term "competitive" can be deceiving: price taking behavior is an important step towards efficient competition, but it is not the only one. One must make sure that "competitive" industries are truly competitive. This is where economic mobility comes in.

I denote by *economic mobility* the set of conditions ensuring that, in competitive industries (industries where firms are price takers), competition leads to an efficient equilibrium. Barriers to entry, such as the bureaucratic costs of creating a new firm, are an obvious instance of a distortion that drives the market equilibrium away from the efficient outcome idealized by the Fundamental Theorem. If all firms were equal, barriers to entry would not be a very big problem: fewer firms might enter than with no barriers, but the loss

would not be that great if the size of the barrier were not that great.¹⁷ But all firms are not equal; and so, more than higher prices, the implication of a barrier to entry is a lower rate of turnover and thus a less than perfect replacement of less efficient incumbents.

But economic mobility is not just about barriers to entry. The empirical evidence suggests that much of the reallocation of productive resources takes place among active firms. Any artificial barrier that encumbers this process has an effect on turnover similar to a barrier to entry. One example is given by severance payments and, more generally, legal and economic restrictions to layoffs; but there are more such examples.

Another point suggested by the analysis in the previous section is that a crucial aspect of economic mobility is not so much the size of distortions but how they differ across firms. In fact, the loss of productive efficiency is more likely to come from the variation in the size of distortions than its size. It's bad enough if potential entrants must pay an extra cost to become active; but it's much worse if some potential entrants must pay a higher cost than others.

There are several sources of cross-firm variation in barriers to entry and mobility. For example, when entry entails complicated bureaucratic steps, the worst thing that can happen is that some firms may be able to evade those bureaucratic costs. Other important sources include fiscal evasion and more generally avoidance of government imposed regulations (labor, environmental, and so forth). The point is that fiscal evasion, for example, is not simply a problem of fairness; it's also an efficiency problem.

In sum, economic mobility is a set of conditions that create a level playing fields, one where firms can easily enter and exit, grow and contract, according to their relative efficiency; an environment where market selection leads to efficient selection.

¹⁷ This is in essence the point of Harberger's (1954) estimate of the social cost from monopoly: if the distortion is small then the welfare loss is of second order.

5. *Final remarks*

Most of microeconomics research in the past few decades has been devoted to showing when and why markets do not work. The game theory revolution — promptly taken up by industrial economists — stresses the importance of market power. Other scholars, from Akerlof to Spence to Stiglitz, focus on asymmetric information and market failures. In the 1980s, work by Arthur and others highlights the importance of increasing returns and non-ergodic market outcomes. It would seem that classical and neo-classical economics are dead.

In fact, the Fundamental Theorem of Welfare Economics is alive and well. True, there are industries with special problems of market power, asymmetric information, natural-monopoly structures, and so forth. But for the most part the assumption of price taking behavior is a fairly good first-order approximation. And the Fundamental Theorem is then the best guide to public policy: create a level playing field and let the market do the rest.

References

- AHN, SANGHOON (2001), "Firm Dynamics and Productivity Growth: A Review of Micro Evidence From OECD Countries," OECD Economics Department Working Paper 297.
- ASPLUND, MARCUS, AND VOLKER NOCKE (2005), "Firm Turnover in Imperfectly Competitive Markets," forthcoming in *Review of Economic Studies*.
- BARRIOS, SALVADOR, HOLGER GORG, AND ERIC STROBL (2005), "The Evolution of the Firm Size Distribution and Nationality of Ownership," *Economics Bulletin* **12**, 1-11.
- BARTELSMAN, ERIC, JOHN HALTIWANGER, AND STEFANO SCARPETTA (2004), "Microeconomic Evidence of Creative Destruction in Industrial and Developing Countries," IZA Discussion Paper 1374.
- BARTELSMAN, ERIC, JOHN HALTIWANGER, AND STEFANO SCARPETTA (2005), "Measuring and Analyzing Cross-country Differences in Firm Dynamics," Free University Amsterdam, University of Maryland, World Bank.
- BARTELSMAN, ERIC, STEFANO SCARPETTA, AND FABIANO SCHIVARDI (2003), "Comparative Analysis of Firm Demographics and Survival: Micro-Level Evidence for the OECD Countries," OECD Economics Department Working Paper 348.
- CABLE, JOHN, AND JOACHIM SCHWALBACH (1991), *International Comparisons of Entry and Exit*, in Geroski and Schwalbach (1991).
- CABRAL, LUÍS (1995), "Sunk Costs, Firm Size and Firm Growth," *Journal of Industrial Economics* **43**, 161-172.
- CABRAL, LUÍS (2000), *Introduction to Industrial Organization*, Cambridge, MA: MIT Press.
- CABRAL, LUÍS (2005), "Economic Mobility," New York University.
- CABRAL, LUÍS, AND JOSÉ MATA (1996), "On the Evolution of the Firm Size Distribution: Facts and Theories," *American Economic Review* **93**, 1075-1090.
- CONWAY, PAUL, VERONIQUE JANOD, AND GIUSEPPE NICOLETTI (2003), "Product Market Regulation in OECD Countries: 1998 to 2003," OECD Economics Department Working Paper 419.

- DJANKOV, SIMEON, RAFAEL LA PORTA, FLORENCIO LOPEZ-DE-SILANES, AND ANDREW SHLEIFER (2002), "The Regulation of Entry," *Quarterly Journal of Economics* **117**, 1-38..
- DUNNE, TIMOTHY AND A. HUGHES (1994), "Age, Size, Growth and Survival: UK Companies in the 1980s," *Journal of Industrial Economics* **17**, 115-140.
- DUNNE, TIMOTHY, MARK ROBERTS, AND LARRY SAMUELSON (1989), "The Growth and Failure of U.S. Manufacturing Plants," *Quarterly Journal of Economics* **104**, 671-688.
- ERICSON, RICHARD AND ARIEL PAKES (1995), "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economics Studies* **62**, 53-82.
- EVANS, DAVID (1987), "The Relation Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries," *Journal of Industrial Economics* **35**, 567-581.
- FAGIOLO, GIORGIO, AND ALESSANDRA LUZZI (2004), "Do Liquidity Constraints Matter in Explaining Firm Size and Growth? Some Evidence from the Italian Manufacturing Industry," *LEM Working Paper* 2004/08, Sant'Anna School of Advanced Studies, April.
- GEROSKI, PAUL (1991), *Market Dynamics and Entry*, Oxford: Blackwell.
- GEROSKI, PAUL (1996), "What Do We Know About Entry?," *International Journal of Industrial Organization* **13**, 421-440.
- GEROSKI, PAUL, AND J. SCHWALBACH (Eds) (1991), *Entry and Market Contestability*, Oxford: Basil Blackwell.
- GIBRAT, R. (1931), *Les Inegalites Economiques. Applications: Aux Inegalites des Richesses, a la Concentration des Entreprises, Aux Populations des Villes, Aux Statistiques des Familles, etc., d'une Loi nouvelle: La Loi de L'ffect Proportionnel*, Paris: Sirey.
- HALL, BRONWYN (1987), "The Relationship Between Firm Size and Firm Growth in the U.S. Manufacturing Sector," *Journal of Industrial Economics* **35**, 583-606.
- HARBERGER, ARNOLD C. (1954), "Monopoly and Resource Allocation," *American Economic Review* **44**, 77-87.
- HOPENHAYN, HUGO (1992), "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica* **60**, 1127-1150.

- HOPENHAYN, HUGO, AND RICHARD ROGERSON (1993), "Job Turnover and Policy Evaluation: A General Equilibrium Analysis," *Journal of Political Economy* **101**, 915-938.
- JOVANOVIĆ, BOYAN (1982), "Selection and Evolution of Industry," *Econometrica* **50**, 649-670.
- LUCAS, ROBERT (1978), "On the Size Distribution of Business Firms," *Bell Journal of Economics* **9**, 508-523.
- MATA, JOSÉ AND PEDRO PORTUGAL (1994), "Life Duration of New Firms," *Journal of Industrial Economics* **42**, 227-246.
- MATA, JOSÉ, PEDRO PORTUGAL, AND PAULO GUIMARAES (1994), "The Survival of New Plants: Start-Up Conditions and Post-Entry Evolution," *International Journal of Industrial Organization* **13**, 459-481.
- OLLEY, STEVEN G., AND ARIEL PAKES (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica* **64**, 1263-1297.
- PAKES, ARIEL, AND RICHARD ERICSON (1998), "Empirical Implications of Alternative Models of Firm Dynamics," *Journal of Economic Theory* **79**, 1-45.
- SUTTON, JOHN (1997), "Gibrat's Legacy," *Journal of Economic Literature* **35**, 40-59.
- SYVERSON, CHAD (2004a), "Product Substitutability and Productivity Dispersion," *Review of Economics and Statistics* **86**, 534-50.
- SYVERSON, CHAD (2004b), "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy* **112**, 1181-1222.
- TIROLE, JEAN (1988), *The Theory of Industrial Organization*, Cambridge, MA: MIT Press.