Productivity, Profitability and Growth: an Empirical Analysis of Firm Selection and Dynamics

Giulio Bottazzi∗†, Marco Grazzi†, Angelo Secchi‡, and Federico Tamagni†

†LEM, Scuola Superiore Sant’Anna, Pisa, Italy
‡Università di Pisa, Pisa, Italy

Do not quote or circulate. Comments are welcome

Abstract

JEL codes:

Keywords:

∗Corresponding Author: Giulio Bottazzi, LEM-Scuola Superiore Sant’Anna, Piazza Martiri della Libertà, 33, 56127, Pisa, Italy. Tel: +39-050-883365. E-mail: bottazzi@sssup.it.
1 Introduction

Building on a long and expanding body of micro-level studies, it has become natural to think about aggregate industry evolution as shaped by the working of an underlying selection process which operates on heterogeneous producers, via a reallocation of market shares from the least to the more efficient business units. It is usually argued that empirical evidence in favour of the existence and the power of such a selection mechanism is robust. At the micro (firm or plant) level, the rates of inputs and outputs reallocation are quite large even within 4-Digit industries. Differences in relative efficiency are large and tend to persist. Large entry and exit rates testify of a relevant process of creative destruction, along which one observes that less productive units have lower probabilities to survive and grow.

The theoretical underpinnings of these findings can be found in those models of firm-industry dynamics, which stress the importance of heterogeneous firms’ characteristics as crucial for their survival and growth patterns. Despite the many existing approaches certainly stress specific and different aspects, the core reference framework, shared by most of these models, is that differential efficiencies in production set the stage for differential potential profitabilities which are then realized after market interactions with incumbents, entrants, demand conditions and uncertainty, thereby shaping the balance between growth and exit events. Simplifying, but not too much, the reasoning behind the theories maintains i) that productive efficiency, profitability records and growth of size constitute the three crucial dimensions of firm performance; ii) that these are highly correlated over the selection process; and therefore iii) that they all stand as candidates to explain the outcome of this process.

This paper starts from the premise that there are valid reasons which prevent from considering the existing evidence as conclusive as it might appear. Quite the opposite, studies have started to appear which challenge many of the common wisdom on workings of selection, from various perspectives.

The element of major weakness, we believe, resides in the approach followed by most of the empirical studies, which is almost exclusively focused on the productivity-survival link. This bears a number of consequences, which get in turn reflected into opportunities for further research. First, this approach implicitly takes that productivity is the proper fitness measure, meaning that this is the level where differential survival and growth opportunities originate from. This has become a common practice, despite theoretical models suggest that profitability is the dimension where selection should act upon, because it is at this level, intermediate in between efficiency and growth, that interactions between firms and environment (demand, uncertainty, competitors) occur. It can be argued that the root followed by empirical research

---

1 For more extensive review of this empirical literature see Bartelsman and Doms (2000); Foster et al. (2001); Bartelsman et al. (2005).

2 Such a simplified theoretical view of firm growth and industry dynamics, can in its essence apply to a broad category of models, going back to Alchian (1950) influential paper. Notably, agreement on the basic dynamics is so widespread that it can accommodate such different and competing frameworks as evolutionary-schumpeterian theories, based on boundedly rational firms experiencing “growth of the fitter” along disequilibrium trajectories (see models in Winter, 1971; Nelson and Winter, 1982; Dosi, 2000; and the replicator dynamics in Metcalfe, 1998) as well as more standard theories of firm growth where agents are rational and profit maximizers (the list of these more standard settings is long, going from Lucas, 1978, to the passive and active learning models in Jovanovic, 1982, and Ericson and Pakes, 1995, respectively; other influential formalizations of creative-destruction, variously incorporating the effect of innovation, vintage of installed technology and capital, are Aghion and Howitt, 1992; Hopenhayn, 1992; Caballero and Hammour, 1994; 1996; and Campbell, 1997; recent works by Melitz, 2003, and Melitz and Ottaviano, 2008, model selection of producers in international trade; see also Asplund and Nocke, 2006.)
might have been influenced by some over-simplification of reality suggested by theory itself, since it is indeed true that profits are often modeled as an increasing function of productivity. Still, the results in Foster et al. (2008) suggest that the productivity-profitability link has attracted too limited attention. They indeed analyze a group of sectors characterized by homogeneous products, finding that selection indeed operates on profitability, and that demand conditions, captured as a component of prices at the producer level, represent the most relevant factor in determining survival and exit.

Second, and relatedly, the pursued approach leaves us with limited, if not lacking, evidence about the productivity-growth and profitability-growth links, whose existence is supported by theory, but much less considered in empirical work. Actually, in the few empirical investigations concerned with this links, the usual finding is that selection is much less tightening than one is used to take for granted. On the one hand, very few empirical studies have considered the link between profit and growth. Statistical discrepancy between the properties of the two variables is often cited as a first evidence of lacking relationship (Geroski, 2002). Goddard et al. (2004) find profitability to be important for future growth on a sample of European banks. Conversely, Coad (2007) concludes that exactly the opposite direction of causation might be the true one, performing panel data estimates on French Manufacturing, while no correlation at all is found in Bottazzi et al. (2008). On the other hand, concerning the productivity-growth link, Baily et al. (1996) find that, among U.S. establishments experiencing productivity growth during their period of observation, the contribution of business units experiencing increased size, in terms of employment, is roughly equivalent to the contribution of firms which actually decreased employment. Foster et al. (2001) also find a lacking relationship between productivity (labor and multifactor) and growth on their detailed dataset of US Manufacturing plants. The same happens in two different samples of Italian firms, as shown in Bottazzi et al. (2007) and Bottazzi et al. (2008), while Disney et al. (2003) find that productivity associates with a negative effect on reallocation of market shares between UK establishments.

Third, another relatively unexplored issue concerns the direction of causation between the dimensions considered. As said, the presumption is often that selection goes from productivity to growth via profitability, but it seems natural that complex feedbacks might be in place. Theories that stress learning, for instance, suggest that growth is a mean to gather the needed resources for further efficiency enhancing or innovative investments, which eventually lead to higher profits. The above mention positive effect of on subsequent profits found in Coad (2007) seems indeed consistent with this interpretation.

In this paper, we aim at offering new evidence on the effectiveness of market selection mechanisms through a multidimensional analysis which considers the productivity-profitability-growth links altogether at the same time, on the same dataset. This is inteded to overcome the limitations common to the bulk of the above mentioned empirical studies in the field, which tends to analyse only one specific link at a time. Our approach shares the early intuitions already undertaken in Dosi (2007) and Bottazzi et al. (2008). However, those attempts are far from complete, and we can hopefully extend the analysis in several respects. In particular, we can take advantage of a richer dataset, longer in time and more representative. More remarkably, we move away from contemporaneous links among the variables, and also consider lead-lag relationships, which allow to shed light on the direction of causation in the productivity-profitability-growth sequence. Moreover, we depart from simple unconditional correlations and non-parametric estimates, to comprise panel data regressions, which allow for a more precise assessment of the strength of each (contemporaneous or lead-lag) relation, conditioning out the possibly confounding effects due to control variables and unobserved
characteristics.

To briefly anticipate the results, our findings disconfirms the existence of a supposedly strong relationship among the variables involved in the selection process, thus bringing support to the challenging evidence already described. We indeed find that all the relationships, even when statistically significant, are very weak: unobserved firm specific heterogeneity account for much higher proportion of the explained variance as compared to the other economic regressors. Another finding is that our failure to find a statistically significant productivity-profitability link might be due to an interesting non-linear shape characterizing this relationship. Finally, lead-lag exercises suggest that, even if the effects are weak, productivity and profitability display a negative impact on future growth, whereas present growth positively affects subsequent productivity and profitability. Overall, the evidence is consistent with lacking competition, preventing firms to enjoy higher growth from higher efficiency, but it is also pointing toward a trade-off between ability to make profits and willingness to grow.

We proceed as follows. The next section describes the dataset of Italian firms on which we base our empirical analysis. In Section 3 and Section 4 we present the basic findings of the paper, on contemporaneous and lead-lag relationships, respectively. Both sections move incrementally from descriptive analysis based on graphical results, to non-parametric unconditional correlations, and end up with panel data estimates. Section 5 concludes, sketching some policy implications of the results and suggesting open questions for further research.

2 Data Description

The research we present here draws upon the Micro.3 databank developed by the Italian Statistical Office (ISTAT). The database covers the period 1989-2004 and represents a development of the former Micro.1 dataset, which was based on the census of all the Italian firms with more than 20 employees conducted by ISTAT over the period 1989-1997. (see Dosi, 2007, and the works cited therein for a survey of results obtained on that dataset). Micro.3 embodies Micro.1, but extends the latter forward in time, thorough combining census data with information appearing in firms’ financial statement. Indeed, starting from 1998 onward, the census of firms only concerns companies with more than 100 employees, due a to change in the data collection practises of ISTAT. As far as firms in the range 20-99 employees are concerned, ISTAT directly monitors only a stratified representative sample which is “rotating” every five years. Yet, data on those firms not included in this “rotating sample” are collected by ISTAT resorting to data provided by Centrale dei Bilanci (CEBI). This is the Italian official member of the European Committee of Central Balance-Sheet Data Offices (ECCBSO), having as institutional duty to collect, clean and diffuse financial statement information on corporate enterprise.

As a result of this development process, a twofold source is available for the period 1998-2004. This allows us to accomplish, in a former and preparatory phase of our work, two objectives. On one side, resorting to balance sheet information from CEBI enables us to enhance the completeness of the database for firms and variables when census data are not available. On the other, when data from both sources are available, it is possible to check for the reliability of the balance sheet data, which might arguably be influenced by fiscal policies and accounting practises. The consistency checks, which also had to take into account slight

---

3The data are made available to our team under the mandatory condition of censorship of any individual information.
difference in definition of variables, were largely satisfactory. Indeed, the discrepancy between
the two sources is lower than 2% for about the 90% of the firms, for most of the variables.
Anyhow, when, for a certain variable, both sources are available, we employ the one coming
from the ISTAT census. It is indeed important to understand that, more than a ‘plain’ higher
number of observation, census data collected by ISTAT enables to access information which
is not compulsorily reported by firms in the standard balance sheet. In particular, there
are specific sections to be filled in the census which concern detailed information on firms’
employment characteristics, types of investments and import-export activity. This allows
MICRO.3 to distinguish – for instance – the total number of employees between production
and non-production workers, and their associated wages.

After the linking of census and CEBI data, Micro.3 contains information about a total
number of 134625 Italian firms. This work is focused on Manufacturing only, for whom data
on 60084 firms are available. To have an idea of the precise representativeness of our data,
consider that firms with more than 20 employees account for almost 70% in terms of total
employment in Italy (see Bartelsman et al., 2004). Focusing on Manufacturing, Figure 1
shows that Micro.3 covers more than 50% of nominal GDP generated by all Italian firms
active in this sector, as reported by Eurostat (National Accounts data); and more than 40%
of employment. The increased coverage of our sample observed from 1998 onward is obviously
due to availability of an additional source of data in the second half of the period covered

In the following, we are going to employ those variables which enable to analyse the
relation between productivity, profitability and growth of the firm. We are going to employ
the ratio between value added and number of employees as our proxy of (labour) productivity,
Π. Profitability, \( P \), is measured in terms of return on sales, that is the ratio between gross
operating margin and total sales. Finally, growth is the log difference of total sales in two
consecutive years, \( g_{T S}^{TS} \). To ease the interpretation of results, all the variables are in Euro, even
though, for the early years of the sample, the data were filled in Lira currency.\(^4\)

\(^4\)In the first part of the period under investigation, the currency with legal tender status was the Italian
Lira, whose exchange rate with the Euro was later fixed at 1 € per 1936.27 Lira.
3 Contemporaneous relationships

In this section we look at the productivity-profitability-growth links when all the variables are observed in the same year, gathering evidence about the presence and strength of contemporaneous relationships among the dimensions involved. We move incrementally from graphical analysis and non-parametric estimates of simple correlations, to panel data techniques controlling for unobserved heterogeneity.

3.1 Descriptive results and unconditional correlations

The graphs reported in Figure 2 show the relationships between the relevant variables, as they appear in the raw data. By way of example we take two years, 1996 and 2002, but it is worth noting that the main properties are quite stable in the remaining years also.

The top panel shows the scatter plot of growth (as log-difference of Sales, \( g^{TS} \) in the
graphs) against productivity (in terms of value added per employee, Π). The latter is taken in deviation from the cross-sectional mean of each year. This reduces the support of the variable, and makes it more comparable with the support of the growth rates, allowing an easier visualization of the possible relationship. The same is done in the mid panels, where productivity is compared with profitability (as of return on sales, P in the plots). Bottom plots instead look at the relationship between growth and profitability.

The clouds of points are quite suggestive. Indeed, the presence of very mild or lacking correlation seem to represent the most reasonable approximation for the relationship between growth and the other two variables. Some more structure is instead displayed in the scatter plot relating profitability and productivity, possibly suggesting a tendency toward a positive association between the variables, perhaps non linear. To help the identification of the possible slope characterizing the data, we also divide the plane into the four quadrants. The productivity-profitability relationship, more concentrated in quadrants I and III, confirms that a positive correlation seems to be in place in the data. A clearcut pattern remains instead difficult to identify when looking at the other two relationships, which remain much more dispersed around the x and y axes.

A more quantitative analysis is then performed computing rank correlations between the variables. Table 1 reports the Kendall’s τ coefficients. This represent a non-parametric esti-

<table>
<thead>
<tr>
<th>Year</th>
<th>Π&lt;sub&gt;t&lt;/sub&gt; vs g&lt;sub&gt;t&lt;/sub&gt;&lt;sup&gt;TS&lt;/sup&gt;</th>
<th>Π&lt;sub&gt;t&lt;/sub&gt; vs P&lt;sub&gt;t&lt;/sub&gt;</th>
<th>P&lt;sub&gt;t&lt;/sub&gt; vs g&lt;sub&gt;t&lt;/sub&gt;&lt;sup&gt;TS&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>0.079 0.000</td>
<td>0.321 0.000</td>
<td>0.087 0.000</td>
</tr>
<tr>
<td>1991</td>
<td>0.117 0.000</td>
<td>0.349 0.000</td>
<td>0.120 0.000</td>
</tr>
<tr>
<td>1992</td>
<td>0.136 0.000</td>
<td>0.346 0.000</td>
<td>0.131 0.000</td>
</tr>
<tr>
<td>1993</td>
<td>0.169 0.000</td>
<td>0.343 0.000</td>
<td>0.121 0.000</td>
</tr>
<tr>
<td>1994</td>
<td>0.129 0.000</td>
<td>0.313 0.000</td>
<td>0.105 0.000</td>
</tr>
<tr>
<td>1995</td>
<td>0.186 0.000</td>
<td>0.334 0.000</td>
<td>0.121 0.000</td>
</tr>
<tr>
<td>1996</td>
<td>0.090 0.000</td>
<td>0.385 0.000</td>
<td>0.067 0.000</td>
</tr>
<tr>
<td>1997</td>
<td>0.148 0.000</td>
<td>0.395 0.000</td>
<td>0.125 0.000</td>
</tr>
<tr>
<td>1998</td>
<td>0.117 0.000</td>
<td>0.388 0.000</td>
<td>0.117 0.000</td>
</tr>
<tr>
<td>1999</td>
<td>0.127 0.000</td>
<td>0.395 0.000</td>
<td>0.107 0.000</td>
</tr>
<tr>
<td>2000</td>
<td>0.117 0.000</td>
<td>0.384 0.000</td>
<td>0.084 0.000</td>
</tr>
<tr>
<td>2001</td>
<td>0.111 0.000</td>
<td>0.380 0.000</td>
<td>0.068 0.000</td>
</tr>
<tr>
<td>2002</td>
<td>0.149 0.000</td>
<td>0.395 0.000</td>
<td>0.119 0.000</td>
</tr>
<tr>
<td>2003</td>
<td>0.155 0.000</td>
<td>0.396 0.000</td>
<td>0.137 0.000</td>
</tr>
<tr>
<td>2004</td>
<td>0.154 0.000</td>
<td>0.395 0.000</td>
<td>0.119 0.000</td>
</tr>
</tbody>
</table>

Table 1: Contemporaneous relationships – Kendall’s τ and P-values in different years t. P-value=0.000 means p < 10⁻⁶.
mate of correlation, which is appropriate given the substantial lack of any clearcut structure found in the raw data. The exercise is done comparing pairs of variables in each year of the period 1990-2004. The values of $\tau$ are reported together with the associated $P$-values allowing to assess their statistical significance.\footnote{1989 is excluded simply because growth cannot be computed in the first year. As a further check, estimates of the Spearman’s $\rho$ correlation coefficient were also explored, but the conclusions discussed here remain valid under this different measure.}

The results agree with the picture conveyed by the above graphical analysis. First, the small magnitudes of the estimated $\tau$, all falling in between 0.1 and 0.4, are indeed suggesting that the variables are only mildly correlated. Second, the productivity-profitability relationship turns sensibly stronger than the other two relations. The higher dispersion noted in the graphs exploring the two relationships wherein growth is included, indeed, turns reflected into values of $\tau$ around 0.1, while the values obtained for the productivity-profitability link are about three times bigger. Finally, we do not observe any clearcut trend in the estimates, revealing that the findings are very stable over time.

3.2 Panel data analysis

The foregoing exercises are of course capturing unconditional correlations among the variables. In this section we instead exploit the panel dimension of the data to include the effect of unobserved, firm specific variables that might play a relevant role in the various relationships considered. In particular, we specify a series of bivariate models of the form

$$Y_{i,t} = c + u_i + \alpha X_{i,t} + \epsilon_{i,t} , \quad (1)$$

where $Y$ and $X$ are the two economic dimensions considered in the different specifications, while the term $u_i$ is a firm-specific constant, absorbing unmeasured characteristics. Also note that year dummies are added to control for time effects common to all the firms.

Table 2 reports the Fixed Effects estimates. Bear in mind that the dependent variables are reported in rows, as opposed to independent variables, which are on columns. For instance, the result of the model $g_{i,t}^{TS} = c + u_i + \alpha P_{i,t} + \epsilon_{i,t}$ is displayed in the top right box. The estimates show that inclusion of fixed and time effects is producing interesting differences with respect to the picture emerged from the foregoing descriptive analysis. First notice that the productivity-profitability relationship is not significant, which can be quite surprising given the sensibly higher values of Kendall’s $\tau$ obtained above for this relationship. One possibility is of course that the linear specification of model (1) cannot account properly for the non-linearities found in Figure 2, while the latter are instead captured by the Kendall’s $\tau$, which is non-parametric index of “generic”, rather than linear correlation. Second, moving to the significant relationships, the sign of the estimates reveals the existence of two different patterns. On the one hand, and quite in accordance to what one might expect, productivity and growth turn linked by a positive relationship. On the other hand, there seems instead to be a trade-off between profitability and growth, which are indeed negatively related.

It is also important to understand that a simple comparison of the magnitudes of the estimated coefficients cannot say anything about the relative strenght of relationships. To evaluate this point, we consider a measure which normalizes the regression coefficients for the standard deviation of the variables involved. This is given by

$$S_{Y,X} = |\hat{\alpha}| \frac{\sigma_X}{\sigma_Y} , \quad (2)$$


<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$g_t^{TS}$</th>
<th>Fixed Effects</th>
<th>$g_t^{TS}$</th>
<th>$\Pi_t$</th>
<th>$P_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td></td>
<td>-0.06586 0.000</td>
<td>0.00134 0.000</td>
<td>-0.00090 0.000</td>
<td></td>
</tr>
<tr>
<td>$\Pi_t$</td>
<td></td>
<td>0.15735</td>
<td>0.09210</td>
<td>0.22365</td>
<td>0.22535</td>
</tr>
<tr>
<td>$S_{Y,X}$</td>
<td></td>
<td>0.08915</td>
<td>0.00136</td>
<td>0.56016</td>
<td>0.54857</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>0.11890 0.59794</td>
<td>-0.04899 0.84038</td>
<td>0.00217 0.33994</td>
<td></td>
</tr>
<tr>
<td>$P_t$</td>
<td></td>
<td>0.10236</td>
<td>0.00268</td>
<td>0.14215</td>
<td>0.11095</td>
</tr>
<tr>
<td>$S_{Y,X}$</td>
<td></td>
<td>0.08915</td>
<td>0.00136</td>
<td>0.56016</td>
<td>0.54857</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>0.11890 0.59794</td>
<td>-0.04899 0.84038</td>
<td>0.00217 0.33994</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.10236</td>
<td>0.00268</td>
<td>0.14215</td>
<td>0.11095</td>
</tr>
</tbody>
</table>

Table 2: Contemporaneous relationships – coefficients and associated $P$-values of Fixed Effects estimates of equation (1). A full set of year dummies is also included (coefficients not reported). Significant coefficients are in bold, $P$-value=0.000 means $p < 10^{-6}$. Strength of relationship is measured via the value of $S_{Y,X}$ in expression (2). $R^2$ refers to equation (3).

where $\hat{\alpha}$ is the Fixed Effects estimate of the $\alpha$ coefficient, and $\sigma$ stands for the standard deviation of the two variables included in each regression. Note that $S_{Y,X}$ is equivalent to (the square root of) an $R^2$ measure, but computed ignoring the terms due to annual dummies and fixed effects. Thus, $S_{Y,X}$ gives the amount of variance of the dependent variable which is explained by the “economic” regressor $X$ alone. This is exactly what one is interested into as far as the strength of the different relationships is concerned. The usual $R^2$, obtained as

$$R^2 = 1 - \frac{\sigma^2}{\sigma_Y^2},$$

is also reported, providing a useful benchmark to be compared with $S_{Y,X}$.

Focusing on the statistically significant relationships, the model taking growth as dependent variable, and productivity as explanatory, turns out to be the regression with the higher $S_{Y,X}$ (approximately equal to 0.15). The other relationships do not display big differences, as we obtain quite comparable values of $S_{Y,X}$, close to 0.09 or about to 0.1. The most relevant result is however given by a comparison of $S_{Y,X}$ and $R^2$. In all of the cases considered, $R^2$ is bigger than $S_{Y,X}$, meaning that $u_i$ and time dummies are responsible for a noticeable increase in the explanatory power of the models. This means that the ability of each independent variable $X$ alone to explain each dependent variable is quite low, once other firm specific factors and time are controlled for. Notably, computation of the standard $R^2$ performed via pooled OLS estimates of the models, exactly corresponding to ignore the $u_i$ terms, gives values (not shown in the Table) which are substantially identical to the $S_{Y,X}$ obtained via Fixed Effects.
Figure 3: Scatter plots of Lead-Lag relationships – productivity (Π) at time \( t \), against sales growth \( (g_{TS}^{T}) \) and profitability \( (P) \) at time \( t + 1 \), for \( t = 1996 \) (left) and \( t = 2002 \) (right).

This confirm that the \( u_i \) terms are more responsible than the time dummies for a good amount of the explanatory power. In other words, what we observe is that, in all of the relationships, the regressors \( X \) display only a weak effect, largely residual with respect to the contribution provided by the unobserved components. There are two explanations for this kind of results. A first one is statistical, and it is related with the very nature of panel data modeling, which forces to assume that the firm-specific factors \( u_i \) do not vary with time. Indeed, since the variables involved (like sales, value added, employment) are known to be very sticky over time, it happens almost “by construction” that the terms \( u_i \) are very likely to capture some fraction of the variance of each dependent variable. This effect is surely interacting with a second, more economic, interpretation of the results, which tells that the strength of the relationships largely depends from idiosyncratic characteristics of the firms. Links are weak on average, while it is heterogeneity that matters to understand if and to what extent gains or loss of productivity, for instance, associates with gains or loss in growth.

As already suggested by graphical analysis and kendall’s \( \tau \), the evidence is supporting that selection is only mildly in place. The relationships between the variables at stake in this work are typically conceived to be quite strong, so that fierce selection is operating assigning rewards and punishments. On the contrary, we observe non-productive or non-profitable firms which do not experience significant growth.

4 Lead-Lag relationships

This section address the issue of lead-lag relationships among the variables. Indeed, simultaneous links can be influenced by budget constraint effects which might binding on the decision
4.1 Unconditional analysis

The graphs in Figure 3 show how productivity is related with subsequent profitability and growth performance of the firms. Values of Π measured in deviation from annual cross-sectional mean at time $t$, are plotted against the values of $g^{TS}$ and $P$ in $t+1$. As before, we take $t=1996$ and $t=2002$ by way of example of what we actually observe in all the other years. Concerning the relation with growth, in the top panels, we obtain exactly the same result which was observed in our graphical analysis of contemporaneous linkages. We can indeed confirm that the two variables do not display any clearcut association. The same conclusion can only partially apply to the productivity-profitability link. In this case the scatter plot might indicate a slightly positive slope, but the result is clearly less remarked than in the contemporaneous exercise run in Section 3. Notably, non-linearities are still present, but again less apparent as compared to the foregoing findings.

Second step is to look at whether higher profitability at time $t$ gets translated into higher productivity or higher growth in the subsequent year. Figure 4 presents the scatter plots...
of $P_t$ against $g_{t+1}^{TS}$ and $\Pi_{t+1}$ (this latter in deviation from its annual average), again taking $t=1996$ and $t=2000$ as reference years. Top panels show that the profitability-growth link is a lacking relationship. As already observed for the contemporaneous relationship between these two variables, the scatter plots are indeed very dispersed, so that one cannot identify any precise shape in the way the clouds of points lay in the quadrants. The bottom plots reveal that an increasing relationship is instead in place between profitability and subsequent productivity. This is more apparent than what obtained above comparing $\Pi_t$ with $P_{t+1}$, that is reversing the lead-lag order of the two variables. The result here is more resembling the picture emerged from our foregoing investigation of the contemporaneous relationship. In particular, the shapes of the scatter plots are again suggesting the presence of considerable non-linearities in the profitability-productivity link.

In the graphs of Figure 5 we then consider growth rates of year $t$ against profitability and (de-meaned) productivity in $t+1$. The clearcut finding is that none of the graphs display any precise shape: irrespective of the year considered, growth appears to lack any clearcut relationship with the other two dimensions of firm structure and performance considered.

Overall, graphical analysis of lead-lag relationships is broadly consistent with the analogous analysis conducted on contemporaneous relationships among the variables. The two relationships involving growth seem substantially weak, if not absent at all, no matter the lead-lag order considered between pairs of variables, whereas the productivity-profitability link appears as the only relation present in the row data. A noticeable result is that the relationship seems stronger when we consider how profitability associates with subsequent productivity, as compared to the case where we ask if higher productivity associates with higher profitability.

We now try to quantify the strenght of the various patterns, presenting a non-parametric analysis of lead-lag correlations based on Kendall’s $\tau$ coefficients.
Results are shown in Table 3, for each year of the sample period. Reported $P-$values allow to test if the different values of $\tau$ are statistically different from zero.\textsuperscript{6}

Presentation of findings follows the discussion of graphical analysis. We take each variable at time $t$ and explore its relationships with the other two variables taken in the following year $t+1$. The weakest relationships are those linking present productivity $\Pi_t$ with subsequent profitability $P_{t+1}$, on the one hand, and present profitability $P_t$ with future growth $g_{T_S}^{T_S}$, on the other. In both cases the coefficients are in practice zero, and not significant at all in some years. The two relationships involving $g_{T_S}^{T_S}$ vs future values of the other variables, come second in the ranking. We indeed obtain slightly higher estimates of $\tau$, around 0.08-0.1, which partially refine the impression of lacking relationship emerged from the above graphical investigation. The latter consideration is also valid for the values of $\tau \approx 0.25 - 0.3$, which we observe in the relationship between productivity and future growth. Lastly, similar magnitudes also characterize the correlation between present profitability and subsequent productivity, which is also significant and positive.

The general impression is however that the emerging picture is strongly supporting that all of the relationships considered are very weak. Indeed, the estimates of $\tau$, even when significant, only rarely exceeds 0.3. This pattern is prevailing over the identification of which specific relationship is actually in place.

\textsuperscript{6}Growth at time $t=1989$ is obviously not available.
### Table 4: Lagged relationships – coefficients and associated $P-$values of Fixed Effects estimates of equation (4). A full set of year dummies is also included (coefficients not reported). Significant coefficients are in bold, $P-$value=0.000 means $p < 10^{-6}$. Strength of relationship is measured via the value of $S_{Y,X}$ in expression (5). $R^2$ refers to equation (6).

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$g_{t+1}^{TS}$</th>
<th>$\Pi_t$</th>
<th>$P_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_{t+1}^{TS}$ c</td>
<td>0.07698</td>
<td>0.000</td>
<td>0.00131</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.00149</td>
<td>0.000</td>
<td>-0.00013</td>
</tr>
<tr>
<td>$S_{Y,X}$</td>
<td>0.1578</td>
<td></td>
<td>0.0637</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.20938</td>
<td></td>
<td>0.21897</td>
</tr>
<tr>
<td>$\Pi_{t+1}$ c</td>
<td>52.4991</td>
<td>0.000</td>
<td>51.9588</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>4.72162</td>
<td>0.000</td>
<td>-0.0029</td>
</tr>
<tr>
<td>$S_{Y,X}$</td>
<td>0.0422</td>
<td></td>
<td>0.0013</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.63669</td>
<td></td>
<td>0.55092</td>
</tr>
<tr>
<td>$P_{t+1}$ c</td>
<td>0.20503</td>
<td>0.432</td>
<td>-0.08528</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.84905</td>
<td>0.000</td>
<td>0.00363</td>
</tr>
<tr>
<td>$S_{Y,X}$</td>
<td>0.0235</td>
<td></td>
<td>0.0034</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.14283</td>
<td></td>
<td>0.13402</td>
</tr>
</tbody>
</table>

4.2 Panel estimates

This section further explores the issue of lead-lag relationship through panel data regressions. As already argued, the aim is to move from unconditional correlations to a parametric exercise which allows to estimate strength of relationships conditioning away the possible effect of unobserved variables. We include such factors as a firm specific constant $u_i$ in a simple panel data model

$$Y_{i,t+1} = c + u_i + \alpha X_{i,t} + \epsilon_{i,t}$$, \hspace{1cm} (4)

where we again focus on the relationship between pairs of variables, again taking a 1-year lag between the two. In order to explore all the possible lead-lag relations, each variable is in turn allowed to enter the models as either the dependent variable $Y_{t+1}$ or, alternatively, as the explanatory regressor $X_t$. Further, all the considered specifications are estimated including a full set of year dummies, at least partially controlling for time trends in the variables.

Fixed Effects estimates are displayed in Table 4. To ease comparison of the strength of the relationships across different specifications, we also compute the “standardized” version of the coefficients. In the present exercises, this is defined as

$$S_{Y,X} = |\hat{\alpha}| \frac{\sigma_{X_i}}{\sigma_{Y_{t+1}}},$$ \hspace{1cm} (5)

and can also be interpreted as the (square root of) the fraction of the variance of $Y_{t+1}$ explained
by $X_t$ alone. The overall explanatory power of the models is instead measured by the standard

\[ R^2 = 1 - \frac{\sigma^2}{\sigma^2_{Y_{t+1}}}, \]  

which also considers the contribution of fixed effects and time dummies.

Concerning significance and sign of the estimates, we interpret the results as non-trivial, in the sense that we do not observe the sequence of positive relationships often suggested by theory of firms’ selection and industry dynamics. Indeed, starting from the two regressions where $g_{t+1}$ is the dependent variable, we obtain that both productivity and profitability display a significant and negative effect on future growth. Second, future productivity $\Pi_{t+1}$ is significantly affected only by present growth, with a positive coefficient, whereas the effect of current profitability $\Pi_t$ turns not significant. In turn, the lacking significance of the relationship between these two dimensions is further confirmed also when we look at the estimated effect of current productivity on subsequent profitability $\Pi_{t+1}$. In both cases, we tend to interpret the results as a failure of standard linear panel models to properly account for the already mentioned non-linearities characterizing this specific link. Lastly, we find that present growth exerts a positive effect on future profitability $P_{t+1}$.

The further remarkable finding is that all of the relationships are confirmed to be very weak, even when we get statistically significant effects. This is shown by the small values of $S_{Y,X}$, revealing that the regressors alone never explain more than a rough 15% of the variance of the dependent variable (cfr. the “$g_{t+1}$ vs $\Pi_t$” specification), and actually much less (from 2% to 6%) in the majority of the models considered. Further, values of $S_{Y,X}$ are smaller than corresponding $R^2$, which is suggesting that the contribution of unobserved $u_i$ and time is more important than the effect attributable to each sole economic regressor. However, leaving aside the contribution of $u_i$ only (via $R^2$ of OLS estimates, not reported) is enough to drop the difference between $S_{Y,X}$ and reported $R^2$, suggesting that $u_i$ are much more important than time dummies in explaining such a discrepancy. The result parallels the conclusion drawn from the analysis of contemporaneous relations: independently of the relationship considered, the direct effects of $Y$ and $X$ are all weak, while heterogeneity of firms is what really matters in shaping the way efficiency, profitability and growth affect each other.

5 Conclusion

Summing up our findings, the evidence is in tune with that part of existing empirical research which is challenging the idea that selection is working. To put it differently, selection, if at all in place, seems to be more complex than a simple monotonic process leading from efficiency to growth. In the data, the relations linking together productivity, profitability and growth appear quite weak, or at least weaker than one would have expected to observe, based on theoretical presumption. Particularly interesting is the lacking relationship between productivity and profitability, which seems due to the novel finding that remarkable non-linearities are characterizing this link.

Bearing in mind this overall picture, the sign and direction of causation of the barely significant relationships are nonetheless noticeable. Analysis of contemporaneous effects indeed reveals a positive relationship between growth and productivity, but a negative association between growth and productivity. Further, it appears from lead-lag analysis that both efficiency and profits negatively affect subsequent growth, whereas growth today exerts a positive effect on future efficiency and profitability performance.
The natural implication following from this picture is that the workings of selection is not only weak, but also much less efficiency enhancing than expected. Negative correlations of productivity and profitability with subsequent growth, indeed, represent a worrying signal that reallocation of market shares does not necessarily occur in favor of best performing firm. This can be read as evidence of too little competition, so that there are mechanisms preventing the more efficient firms to grow more. Still, positive effect of growth on future productivity and profitability is supporting that firms enjoy of growth to “learn” efficiency, but then more productive firms might prefer profitability to the alternative to pursue further growth. It is not obvious which of the two effects – barriers to competition vs unwillingness to growth – is prevailing. In both cases, there is certainly room for policy intervention. Measures must however be carefully designed, with a twofold objective. On the one hand, easing competition is essential to allow more efficient firms to grow more, if necessary providing support to alleviate social consequences of displacement of least efficient businesses. On the other hand, however, there might also be a need to reduce the emphasis usually devoted to improving efficiency, and set up incentives for the firms to pass efficiency gains onto investment and growth, rather than onto profitability only. From such a policy perspective, it is also instructive that, as noted in our brief literature review, the failure to find a clearcut positive correlation between productivity and growth does not represent an Italian specificity. It is indeed emerging in countries –such as the US and the UK – where competition forces are arguably stronger, and barriers to entry less stringent. This point further confirms that firms’ ability and willingness to identify further growth as an objective might matter more than fostering competition and efficiency by itself. In order to reach more conclusive results, one should consider an important missing link, namely the role played by investments into the observed productivity-profitability-growth chain of relationships. In particular, we suggest that a direct exploration of the “quality” (innovative vs imitative, intensive or non-intensive in skilled labour, and so on) of the investment pursued by the different firms, would be a very appropriate topic for further research. The aim is to further clarify whether learning is indeed behind the uncovered sequence of causation, from present growth to subsequent productivity and profitability, but not leading to further growth.

There are however other open questions. First, it is not clear a priori whether our results are affected by the fact that we do not observe price and quantities separately. Prices showing up in all the measures employed should induce a statistical co-movement between pairs of variables. Then, observed correlations are overstating “true” correlations, thereby confirming that relationships in reality might be even weaker than observed. However, recent evidence (Foster et al., 2008) lend support that plant-level prices are inversely correlated with physical efficiency, suggesting that the profitability-productivity link might indeed pass through increased demand attracted via lower prices, for entrants in particular. Still, such findings apply to sectors characterized by homogeneous products, and it is unclear if they are relevant in other markets. Of course, one needs price-quantity data, which are unfortunately difficult to find.

Further, another possible explanation of the findings, in turn deserving further research, is that different sectors, more narrowly defined than overall Manufacturing, are characterized by different selection processes. Firms operating in different industries possibly face different growth opportunities and specific competitive environments, due to, for instance, different degrees of maturity and openness to trade, different type of inputs, varying technological levels, or, most obviously, different market structures. This are certainly complex phenomena. Adding sectoral dummies or re-estimating the models within each industry would give at least a first valuable hint to asses whether they can affect the overall picture of weak selection emerged at
the aggregate level. Again, one needs the appropriate data, which must be detailed enough to reach meaningful levels of sectoral aggregation.

Finally, it would be important to link results to macroeconomic conditions. There is evidence (Foster et al., 2001) that workings of selection might differ along different phases of the business cycles. The relatively long length of our sample period might hopefully be enough to capture some of this cyclical components.

The possible impact of these and other factors is left for further analysis. They all cast a reasonable chance to offer some explanation for the present evidence that selection is not as selective as expected, which remains as the main result of our investigation.
References


