WHY IS PRODUCTIVITY SO DISPERSED?

RACHEL GRIFFITH
IFS, UCL, and AIM

JONATHAN HASKEL
Queen Mary, University of London, AIM, CEPR, and IZA

ANDY NEELY
Cranfield School of Management and AIM

Many papers have documented wide variations in productivity even in narrowly defined industries. Some have argued that this primarily reflects measurement problems due, for example, to comparing across different products. Others argue that this reflects persistent differences in performance due, for example, to management. This paper looks at productivity differences not within an industry but within a firm. We use data on productivity of different branches within lines of business of a major UK-based wholesaler. Using these productivity data for comparisons is, we argue, more likely to compare like with like than comparing between firms. We document sustained differences in productivity even between branches within the same line of business. We also discuss the extent to which they are correlated with differences in management and find that such differences ‘account’ for around 40 per cent of the difference in productivity.

I. INTRODUCTION

A long-standing empirical puzzle in economics is why we see so much persistent variation in productivity, even within very narrowly defined industries. Economic theory suggests that in well-functioning markets poorly performing firms should be unviable, and therefore exit.

When comparing productivity at a more aggregate level—for example, between countries or broadly defined industries—the measured gap (say between

---

1 E-mail addresses: r.griffith@ifs.org.uk; j.e.haskel@qmul.ac.uk; a.neely@cranfield.ac.uk

Financial support for this research comes from the ESRC/EPSRC Advanced Institute of Management Research, grant number RES-331-25-0030. This work contains statistical data from the Office for National Statistics (ONS) which is Crown copyright and reproduced with the permission of the controller of HMSO and Queen’s Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.
the UK and the USA) may be due to different circumstances in each country: the skills base, the regulatory environment, and the industry mix. Productivity variation between industries may be explained by, for example, different technology and different levels of competition. But the wide and persistent differences that exist across establishments within very narrowly defined industries remain a puzzle.²

For example, Syverson (2004) computes the ratio of productivity at the 90th percentile (very well performing firm) to the 10th percentile (very poorly performing firm) within very narrowly defined manufacturing industries in the USA in 1997.³ The average of these ratios is over 4: in other words, the top firms are over four times as productive as the bottom firms. Criscuolo et al. (2003) perform a similar calculation for UK manufacturing in 2000 and find that the top firms are over five times as productive as the bottom firms.

Does persistence in such variation matter? One view is that it is illusory: simply an indicator that we do not measure productivity very well. There are a whole host of difficulties in measuring productivity—government industrial classifications may not accurately capture firms that are undertaking similar activities or competing in the same markets and we may be mis-measuring inputs or outputs. Alternatively, productivity variation may reflect real differences in productivity. If it does, we need to understand why variation persists and, crucially, whether it is the same firms that consistently underperform, or whether firms move in and out of bad performance—for example, do firms start as poor performers but then move up as they gain more experience? This latter situation would be one where there is productivity dispersion, but where it is quite consistent with a vigorous market.

In this paper we aim to shed light on these issues by looking at productivity differences within a single firm. We focus on the retail/wholesale industry. Recent attention has focused on this sector as it accounts for close to 20 per cent of the UK’s aggregate productivity gap with the USA.⁴ There is a wide dispersion in productivity in this sector, with the firm at the 90th decile having just under five times higher productivity than the firm at the 10th percentile. The firm that we look at is a national firm with many hundreds of establishments located throughout the United Kingdom operating in the wholesaling of building and plumbing equipment. Even within this narrowly defined (four-digit) industry, when we look across firms in the UK we see that the firm at the 90th percentile is around 2.7 times more productive than the firm at the 10th percentile.

These calculations are for firms within the same industry. In this paper we compare branches within the same firm. By doing so, we can be more certain that we are comparing ‘like for like’—branches sell basically the same thing—and we can control for many other forms of measurement error (the data are all collected in a similar way within the firm). When we do this we still find substantial variation in sales per worker and profit before income tax in these establishments, with the best-performing establishment having about five times larger sales per employee than the worst performing.

This variation appears to be persistent. While some poorly performing branches improve, most branches remain in their relative position over years. This leaves us with a new puzzle—why do these differences in performance persist? There is a large number of candidate reasons (which we enumerate below). One topical theory is that differences in productivity might at least partly be explained by differences in management. We investigate the role that local branch managers have on performance, using data on management from the firm itself. These data are scores that branches have achieved, collected as part of the firm’s balanced scorecard performance-monitoring programme (they score, for example, communication with staff, employee satisfaction, customer service, etc.). We find a

² See, for example, Dunne et al. (1989), Davis and Haltiwanger (1991), Baily et al. (1992), Bernard and Jensen (1995), Davis et al. (1996), Bartelsman and Doms (2000), Foster et al. (2002), and Disney et al. (2003).
³ Industries are classified into one- two-, three-, etc. digit divisions, with the divisions becoming finer and finer. This study uses 443 four-digit US manufacturing industries. Some examples of four-digit US manufacturing industries are manufactured ice, dog and cat food, and animal foods excluding those for dogs and cats.
⁴ See, inter alia, Inklaar et al. (2003) and Griffith et al. (2003).
strong correlation between this score and productivity. Of course, this correlation does not imply causation. Indeed, it might reflect the sorting of managers between branches—good branch managers are asked to manage the better branches which achieve high levels of performance both on the balanced scorecard metrics and the productivity metrics.

What are the quantitative effects that we find? A movement of a firm from the lower- to the upper-quartile management scores ‘accounts’ for about 40 per cent of the inter-quartile differences in productivity. In their study, using data on externally assessed management scores between firms, Bloom and Van Reenen (2006) find that management ‘accounts’ for around 33 per cent of inter-quartile differences in their productivity measure. We comment more on these differences below.

The rest of this paper proceeds as follows. The next section discusses the observed productivity spreads in the wholesaling industry and in our firm. Section III investigates how much of this variation can be explained by differences in management. Section IV concludes.

II. PRODUCTIVITY TRENDS

We start by briefly reviewing what the existing literature shows in terms of productivity dispersion. Table 1 summarizes a number of papers.

The first row shows the Syverson (2004) result referred to above—labour productivity in US manufacturing industries. The second row shows the spread of total factor productivity (TFP). The third row shows that the labour productivity variation also shows up in UK manufacturing. One reason for this might be that labour productivity varies owing to different employment of other inputs, most notably capital and other materials. We would expect, for example, that workers producing pencils are working with very different capital to workers producing aircraft. To control for this, the second and fourth rows in Table 1 shows dispersion in TFP. The dispersion still remains. The final row shows dispersion in both manufacturing and services, which is, in fact larger, than that in manufacturing.

Why does such variation exist? One view is that one is not comparing like with like. Firms might be selling such different products that comparing pencils and

---

Table 1
Productivity Dispersion

<table>
<thead>
<tr>
<th></th>
<th>Standard deviation</th>
<th>90th/10th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log TFP, 1997</td>
<td>0.34</td>
</tr>
<tr>
<td>Criscuolo et al. (2003)</td>
<td>UK, manufacturing, log (gross output/employment), 2000</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>TFP, 2000</td>
<td>0.18</td>
</tr>
<tr>
<td>Oulton (1998)</td>
<td>UK, whole economy, log (sales/employment)</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>UK, wholesale</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Criscuolo et al. and Syverson data are averages of deviations within four-digit industries.*

---

5 TFP generalizes the single factor (labour) productivity measure and so is a multi-factor productivity measure. Rather than dividing by just labour input, it allows productivity to be affected by labour, capital, and material inputs, with each input weighted by its share in total costs. Such weighting is appropriate from an economic or index number point of view.
aircraft, even adjusting for other inputs, is not appropriate. Since comparisons are often made within four-digit industries, this view has less force, but it is clear that there are still some significant differences between firms even within the same four-digit industries. One reason for these differences is the use of different technologies. If capital is well measured this could be controlled for, but measurement problems in capital are legion, and controlling for capital does not control for other factors. Additional complexity is added when one considers that there are other—often called intangible—assets, that are poorly measured. These intangible assets include organizational, reputational, and managerial capital that may account for productivity differences.

(i) Productivity in Wholesaling

How do we measure productivity in this paper? As in much of the literature, our basic measure is output over labour inputs. We have two output measures: sales and sales minus the costs of goods sold (this differs from value-added in that several other intermediate goods are not deducted). We measure labour inputs using the wage bill, since we do not have accurate measures of numbers employed in all branches in all periods. This is not quite labour productivity as conventionally measured, but the use of wages has the advantage of controlling for the quality of labour, which might plausibly vary between stores. We use nominal sales. This means that a branch that sells more of a good with a higher mark-up will appear more productive than one selling a product that is priced nearer to cost. We do not have information on the branch-specific composition of sales, or prices charged. However, the branches within each division of the firm carry very similar products, and over a year it is likely that the mix of goods sold will even out. In any case, it will be much more comparable than when looking across different firms selling quite different goods.

While we focus on the productivity spread within the firm, at the outset it is useful to compare the firm with other firms in the same industry. We focus on the two largest divisions of the wholesaler which sell building and plumbing materials. We compare these with data from the Annual Respondents Database (ARD; see Data Appendix) for the industries SIC 5153 and 5154, which are, respectively, ‘wholesale of wood, construction materials, and sanitary equipment’, and ‘wholesale of hardware, plumbing and heating equipment, and supplies’.

Figure 1 compares the distribution of productivity within branches of each division of the firm with the distribution of productivity in establishments within the relevant industry. The top graph in the top panel, headed ‘Plumb: Company’ is a histogram of productivity for the plumbing division of the company and the lower graph in the top panel headed ‘Plumb: Industry’ is for the four-digit industry. The lower panel has the analogous graphs for the building materials industry. A number of points stand out. There is a wide spread of productivity within both the firm and within these narrowly defined industries. The 90/10 ratio is about 3 to 1 within each industry. There is a similar spread even within the company and even within the different divisions of the company. Note that the company spread is smaller, about 2 to 1, and that the company average is above that of the industry.

Table 2 sets out some further details about the distribution of productivity within the industry and firm. This shows the dispersion in productivity as measured by the ratio of productivity in the 90th to the 10th percentile. As the sectors become more disaggregated, the dispersion falls. The branches within the company are less dispersed than for the sector as a whole, but even at this quite disaggregated level productivity is still very dispersed. In 2003, the branch at the 90th percentile was more than twice as productive as that at the 10th percentile in plumbing and 1.5 times as productive in building.

(ii) How Do Establishments Move in the Productivity Distribution?

As we emphasized above, whether or not persistent dispersion in productivity is a cause for concern depends on whether it is the same branches or firms that are persistently poor performers, or whether individuals move around the distribution. Consider the analogy to the personal income distribution—if some people spend their whole life poor this may be a cause for concern, whereas if everyone is poor for a few years (say when they are students) but then moves up the distribution, this is of much less concern. How persistent is a branch’s position in the cross-section productivity distribution over time?
Figure 1
Comparing Productivity in the Firm with the Industry

Table 2
90/10 Ratio

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>6.44</td>
<td>5.46</td>
<td>9.71</td>
<td>9.55</td>
</tr>
<tr>
<td>51</td>
<td>6.66</td>
<td>5.42</td>
<td>5.14</td>
<td>4.76</td>
</tr>
<tr>
<td>515</td>
<td>5.09</td>
<td>4.39</td>
<td>4.38</td>
<td>3.65</td>
</tr>
<tr>
<td>5153_4</td>
<td>3.01</td>
<td>2.76</td>
<td>2.98</td>
<td>2.73</td>
</tr>
<tr>
<td>5153</td>
<td>3.01</td>
<td>2.74</td>
<td>3.01</td>
<td>2.65</td>
</tr>
<tr>
<td>5154</td>
<td>3.14</td>
<td>2.62</td>
<td>3.08</td>
<td>2.87</td>
</tr>
<tr>
<td>Company</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>2.25</td>
<td>2.10</td>
<td>2.16</td>
<td>2.09</td>
</tr>
<tr>
<td>Building</td>
<td>1.69</td>
<td>1.65</td>
<td>1.63</td>
<td>1.57</td>
</tr>
<tr>
<td>Plumbing</td>
<td>2.11</td>
<td>1.95</td>
<td>2.07</td>
<td>1.94</td>
</tr>
</tbody>
</table>
To examine changes over time in the distribution we first look at how the average level of productivity evolves for branches at a certain point in the distribution. This is set out in Figure 2.

Consider the upper panel for Building. We take all branches in the building division that exist in 2000 and rank them by their position in the productivity distribution in 2000. We then split the plants into ten equal size groups (deciles) and calculate the average productivity of each decile. The graph then shows the average productivity, in each year, of the same group of plants according to their membership in the initial year deciles. The graph shows that the ranking of the plants is quite stable. The group of plants in the top decile, those who start with the
Figure 3
Transition Matrices, Productivity by Brand

Note: These graphs compare where each branch is in the productivity distribution in 2000 (on the horizontal axis) and in 2004 (on the vertical axis). We calculate margins per labour cost (sales less cost of bought-in goods divided by the wage bill) for every branch, and divide branches into 10 equal sized groups (deciles), with the first group being those branches with the lowest value and the tenth group those with the highest values. We do this in 2000 and 2004. Zero in 2000 indicates branches that opened between 2000 and 2004, zero in 2004 indicates branches that closed between 2000 and 2004. The size of the dot indicates the number of branches in that cell.

highest average, ends up with the highest average productivity, although slightly lower than where they started. The group in the lowest decile ends up still at the lowest point, but their average productivity has risen. The middle groups tend to stay as they are, suggesting that the position in the productivity distribution is quite persistent, but that neither the highest nor the lowest productivity levels are sustained; the highest levels cannot be maintained and the lowest levels improve towards the middle. The lower panel shows the same picture for Plumbing division and shows a similar pattern, although there is more convergence.

A second way to assess persistence is by looking at transition matrices. Figure 3 sets these out in diagrammatic form.

To construct these we start by allocating all branches in the Building division to their productivity ranking deciles in 2000. We then calculate which decile these firms end up in in 2004 and allocate them to the
appropriate point in the figure. For example, a firm in the top decile in 2000 which is also in the top decile in 2004 would be in the very top right-hand cell. In the figure, the diameter of the dot corresponds to the number of firms in that cell. A zero in 2000 indicates that the branch was opened after 2000, and a zero in 2004 indicates that the branch shut down before 2004. As the figure shows, there is a concentration of branches (i.e. the widest dots) in the 10/10 and 1/1 cells, indicating that firms in the 10th and 1st decile tend to stay there. Indeed, there is somewhat of a concentration of firms along the diagonal and near diagonal, indicating persistence in performance. Entry and exit tend to be into and from the lowest deciles.

It is slightly difficult to compare these results with others, since other studies have mostly been for manufacturing and have used labour productivity data. Haskel and Martin (2002) look at gross output per person averaged over 3-year intervals for the 1980 and 1990s in UK manufacturing, and find that nearly half of firms tend to stay where they are in the productivity distribution. For example, 48 per cent of firms who were in the bottom quintile of the productivity distribution were still there 3 years later; likewise 50 per cent of firms in the top quintile were still there 3 years later.

Finally, we can partition the variance of productivity into that part that is due to branches having permanently different levels of productivity, and that part that is due to productivity in individual branches fluctuating over time (around their own mean). These two effects are commonly termed the ‘between variance’ and ‘within variance’, respectively. In both divisions the between variation is about twice the within variation. Thus, the bulk of the variation in productivity is due to the fact that some branches are permanently more productive than other branches.

III. THE ROLE OF MANAGEMENT

So far we have shown that there is persistent variation in productivity even when we compare very similar branches where issues of measure error are likely to be minimal. As we stated earlier, there is potentially a large number of reasons why this might be true. One reason that we are interested in investigating here is what role the management skills of the branch manager might play.

Because we are looking within a firm we cannot investigate the role of management overall. A few papers consider this. Bloom and Van Reenen (2006) use a survey of company management practice where they ask a series of questions on practice, each requiring an answer on a scale of 1 to 5. They then normalize the practice to mean zero and standard deviation one and take the unweighted average across all normalized scores as the measure of overall managerial practice. Womack et al. (1990) and Oliver et al. (1996) compare the quality and productivity performance of firms in the automotive industry, arguing that performance variations can be traced back to management practices and the adoption of lean-thinking principles.

In our study, an important question is how to measure the quality of management. A commonly used framework distinguishes between input, process, and outcome measures. Input measures include classic measures of human capital, such as skill levels, educational qualifications, attainment, and experience. Process measures, such as those used by Bloom and Van Reenen (2006), focus on the management processes used in the organization—e.g. the extent to which targets are set, incentives are available, lean manufacturing methods are adopted, etc. Outcome measures focus on the outcomes of management—e.g. customer satisfaction, employee satisfaction, operational performance, etc. Clearly, performance in terms of outcome measures cannot be attributed solely to management, but for the purpose of this paper we assume that management has significant influence over performance outcomes. Hence we use a series of performance-outcome measures as a proxy for management, assuming that better performance in terms of outcomes is correlated with better management.

To measure the performance of its branches the firm has adopted a balanced scorecard (Kaplan and Norton, 1992). The balanced scorecard is a widely used measurement framework, which consists of four perspectives—financial, customer, internal, and innovation and learning. In essence the balanced scorecard is designed to provide a ‘balanced’ view of an organization’s performance by looking at it.
from a variety of perspectives. Individual firms using the balanced scorecard are encouraged to select those performance measures that are most appropriate for their context. In this case the firm involved had 17 performance measures on its balanced scorecard. Together these 17 measures reflect all of the important dimensions of performance for the business. Six of the measures directly contain sales and/or labour cost and so we exclude these for the purpose of our calculations. The remaining 11 measures cover issues such as customer satisfaction, stock availability, and operational standards (see Table 3). For each of these measures the firm has set targets. A traffic-light reporting system is used (green for excellent, amber for acceptable, and red for unacceptable). To measure management performance we calculate how well each individual branch performs against the targets defined by the business. Hence, the best-performing managers will operate in branches that are in the green zone for all of their measures, while the worst-performing managers will operate in branches that are in the red zone for all of their measures.

Figure 4 shows a plot of the branch’s average performance on the 11 measures over the calendar year 2003 against the branches’ productivity performance over the same period. We see a clear positive correlation between the two. This graph shows a correlation between two variables and not causation. It might be that managers are assigned to stores depending on store performance. Suppose, for example, the best managers are assigned to the worst-performing branches. Then we would expect a negative relation (at first) between management scores and performance, followed by a positive relation, but only for those initially poorly performing branches. Thus in Figure 5 we split branches into those that had higher than median productivity in the two preceding years (2000–1) and those that were below median. In both cases, we see a positive relationship here as well.

How do these data compare with other findings? Bloom and Van Reenen (2006) control for other factors besides management (capital, materials, etc.) and find that a movement from the lower to the upper quartile of management scores between firms (0.971 points) is associated with an increase in TFP of around 5 per cent. In their actual data, the difference in TFP between the lower quartile and upper quartile of the firms is 31.9 per cent. Hence management ‘accounts’ for around 33 per cent of inter-quartile differences in their productivity measure. In our data, a movement of a firm from the lower to the upper quartile management scores (0.15 points) is associated with an increase in our productivity measure of 0.32 points. The difference between the upper and lower quartile of productivity

Note: Each dot represents a branch. Data are for the period 2002–4.
<table>
<thead>
<tr>
<th>Management Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Customer measures</strong></td>
</tr>
<tr>
<td>Customer satisfaction</td>
</tr>
<tr>
<td>Customer retention</td>
</tr>
<tr>
<td>Sales mix</td>
</tr>
<tr>
<td>Availability of stock range</td>
</tr>
<tr>
<td><strong>Internal measures</strong></td>
</tr>
<tr>
<td>Operational efficiency</td>
</tr>
<tr>
<td>Operational standards</td>
</tr>
<tr>
<td>Inter-company cooperation</td>
</tr>
<tr>
<td><strong>People measures</strong></td>
</tr>
<tr>
<td>Staff retention</td>
</tr>
<tr>
<td>Employee satisfaction</td>
</tr>
<tr>
<td>Communication</td>
</tr>
<tr>
<td><strong>Supplier measures</strong></td>
</tr>
<tr>
<td>Spend with approved suppliers</td>
</tr>
</tbody>
</table>
is 0.79 points. Thus our management differences ‘account’ for about 40 per cent of the inter-quartile differences in productivity. This is remarkably close to the Bloom and van Reenen number. A few points are worth making. First, our measure of productivity does not correct for materials and capital in the way that the Bloom study does. To the extent that better-managed firms have more capital, then we might overstate the fraction of (total factor) productivity accounted for by management. Second, our data are a much more like-for-like comparison of output than the Bloom and Van Reenen (2006) sample. To the extent that their cross-section comparison is obscured by comparisons of firms with different products, they might overstate the contribution of management to TFP.6

IV. SUMMARY AND CONCLUSIONS
This paper has used a new data set to compare productivity not between firms in an industry, but

6 If competition is not perfect and better managers are able to achieve higher mark-ups, then the measured productivity in the better firms might reflect, in part, higher prices rather than more output.
within a firm. We argue this comparison is more of a like-for-like comparison than other work since different firms likely produce different goods. We document a persistent productivity spread, not unlike that observed in the between-firm work. We then relate this spread to management measures and find that the differences in management account for around 40 per cent of the observed productivity spread. This correlation cannot, of course, inform us about the causal effect of management on productivity, but it does suggest the relation is worth further investigation.

V. DATA APPENDIX

(i) ARD Data

The ARD data are a sample of the business register, the Interdepartmental Business Register (IDBR) (the register is a file of the addresses of all UK businesses, compiled using a combination of tax records, information lodged at Companies House, Dun & Bradstreet data, and information built up from other surveys). The IDBR holds information on the structure of the enterprise/enterprise group, i.e. the addresses of the relevant enterprise and shops and their industries. It holds some employment and some output records. The Annual Business Inquiry (ABI) is the annual survey on inputs and outputs and the ARD consists of the panel micro-level information obtained from the ABI. To reduce compliance costs, however, the ABI is not a census of all shops. This is in two regards. First, an enterprise with many shops may decide to report information for a number of shops combined (a ‘reporting unit’). In practice, most reporting units are single-shop firms, but most employment is in firms with many shops. These multi-shop firms usually report on the whole chain in one reporting unit. Second, all reporting units above a certain employment (currently 250) are sent an ABI form. Smaller reporting units are sampled by size–region–industry bands.7 To match the ARD data with the company data we use sales as defined in the ARD and sales less cost of bought-in goods. Note that the latter is not value added as measured on the ARD, since that also subtracts other intermediate inputs, such as heating and lighting, etc. Labour costs are measured in the ARD as wage bill plus employer taxes. Employment usually refers to December and sales and costs data are typically for the calendar year. We use the closest four-digit industry for comparison with our chain-level data. However, it is important to note that the ARD data are by firm, not by store. Thus productivity differences between firms likely mask differences between stores.

REFERENCES


7 The employment size bands are 1–9, 10–19, 20–49, 50–99, 100–249, the regions are England and Wales combined, Scotland, and Northern Ireland (NI). Within England and Wales, industries are stratified at the four-digit level, NI is at two-digit level, and Scotland is at a hybrid two-/three-/four-digit level (oversampling in Scotland and NI is by arrangement with local executives). See Partington (2001).


