Haltiwanger, John

Article
Firm dynamics and productivity growth

EIB Papers

Provided in Cooperation with:
European Investment Bank (EIB), Luxembourg

Suggested Citation: Haltiwanger, John (2011) : Firm dynamics and productivity growth, EIB Papers, ISSN 0257-7755, European Investment Bank (EIB), Luxembourg, Vol. 16, Iss. 1, pp. 116-136

This Version is available at:
http://hdl.handle.net/10419/54671

Standard-Nutzungsbedingungen:
Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen. Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Countries differ substantially in the extent to which more productive firms are large and/or are becoming larger and less productive firms are small and/or becoming smaller. A challenge for both emerging and advanced economies is that achieving such static and dynamic allocative efficiency requires an ongoing process of restructuring and reallocation. Such restructuring and reallocation is by its very nature costly. Market structure and institutions that promote well-functioning business dynamism are, accordingly, critical for economic performance. In the 1980s and 1990s, the US exhibited a robust pace of business dynamism that contributed substantially to US productivity and job growth. There are, however, some disturbing trends in the nature of US business dynamism – for example, the pace of business start-ups has declined secularly especially over the last decade. The decline in the pace of business dynamism may be contributing to the anaemic US recovery from the recent recession.

John Haltiwanger (haltiwan@econ.umd.edu) is professor of economics at the University of Maryland, a research associate at the National Bureau of Economic Research (NBER) and a senior research fellow at the Center for Economic Studies at the Bureau of the Census. The author would like to thank Alan Finkelstein Shapiro for excellent research assistance and Hubert Strauss for excellent comments. This paper, without implication, draws heavily on collaborative work with Eric Bartelsman, Steven Davis, Haiyan Deng, Jason Faberman, Lucia Foster, Ron Jarmin, C.J. Kritzan, Javier Miranda, Stefano Scarpetta and Chad Syverson. The views and opinions expressed here are of the author alone.
Firm dynamics and productivity growth

1. Introduction

A ubiquitous feature of market economies is that there are large differences in productivity across businesses even within narrowly defined sectors. These differences in productivity exhibit considerable persistence. An obvious question is how low-productivity firms persist in a well-functioning, market economy. To help answer this question, it is instructive to note two key features of well-functioning, market economies. First, in these economies the lowest-productivity businesses are more likely to exit. Second, among surviving businesses, the most productive businesses are the largest businesses. These features imply that while low-productivity businesses do exist in equilibrium due to a variety of frictions, market forces in a well-functioning, market economy allocate more outputs and inputs to the more productive businesses. That is, well-functioning market economies exhibit a high degree of allocative efficiency.

There is increasing evidence that the success of an economy depends critically on the extent to which the market structure, business climate and institutions promote such allocative efficiency. Allocative efficiency means that resources are allocated to their highest valued use. Achieving high allocative efficiency is not just a static problem but a dynamic one. The reason is that the economic environment is constantly changing, requiring an ongoing process of restructuring and reallocation. One manifestation of such change is that while the differences in productivity across businesses are persistent, there is a process of continuous change in the distribution of productivity. As such, in well-functioning market economies there is a high pace of ongoing reallocation of outputs and inputs across businesses wherein resources are shifted away from less productive to more productive producers. The empirical evidence shows that in well-functioning economies the ongoing pace of reallocation is productivity enhancing. One needs to be careful about making causal inferences here – it is not reallocation per se that yields productivity growth but rather the process of productivity growth requires ongoing productivity-enhancing reallocation. The reason is that there is need for experimentation and trial and error in both developing new products and processes and in adapting to changes in the economic environment.

Allocative efficiency thus involves both static and dynamic dimensions. Static allocative efficiency is associated with more productive businesses being larger. Dynamic allocative efficiency is associated with businesses that have become more (less) productive expanding (contracting). In addition, achieving allocative efficiency also inherently involves keeping the costs of such business dynamism low.

By its very nature the reallocation of outputs and inputs across firms is costly – it is costly to businesses in terms of adjustment frictions and it is costly to households as workers are caught up in this reallocation and also because households own the businesses incurring costs. Workers impacted by reallocation often spend time in unemployment and if unemployment is prolonged, it is often accompanied by substantial and persistent earnings losses. Substantial costs are born by businesses in terms of the time and resources associated with changing activity, whether via firm entry and exit or contraction and expansion. Some of these time and resource costs are an inherent component of the process of reallocation but market structure and the regulatory and institutional framework play a critical role in determining the extent to which the reallocation is productivity enhancing.

---

1. The evidence suggests this partly reflects idiosyncratic choices of product quality and mix, location of the business, organizational practices and the like. It likely also reflects differences in entrepreneurial and managerial ability. In addition, it likely reflects a form of luck – being in the right place at the right time with a product and process that is of high value and can be produced in a cost-effective manner. In what follows, as a short-hand we mostly refer to all these factors as differences in productivity (broadly defined) across businesses.
A key theme of the paper is that the extent to which a country exhibits static and dynamic allocative efficiency without incurring high costs depends critically on market structure and the institutions that govern economic activity. While distortions to allocative efficiency are present in all economies, countries with strongly distorted product, capital and labour markets and poorly functioning institutions exhibit worse outcomes in terms of allocative efficiency. In turn, highly distorted economies have lower output per capita.

In this paper, I summarize the theoretical and empirical literature underlying the challenges of promoting allocative efficiency on the one hand and minimizing the disruption costs of ongoing reallocation. Section 2 provides an overview of the basic facts on firm dynamics. Section 3 presents conceptual underpinnings. Section 4 discusses policy challenges. Section 5 provides some concluding remarks.

2. The relationship between productivity and reallocation

2.1 Basic facts

It is useful to start with basic facts about the distribution of productivity and size across businesses. There is much evidence that even within narrowly defined sectors there is substantial dispersion in both productivity and size of businesses.

Figure 1. The distribution of productivity across businesses in the same industry

Interquartile range within narrow industries is over 30 log points

Productivity of businesses

Note: The above is a hypothetical depiction of the shape of the productivity distribution reflecting the empirical finding that productivity is approximately log normally distributed. The reported interquartile range is based on the distribution of US manufacturing establishments (see Syverson 2004).

Figure 1 illustrates a hypothetical productivity distribution within industries that reflects the patterns that have been observed in the data. For example, Syverson (2004) shows that the inter-quartile range of measures of within-industry establishment-level total factor productivity (TFP) is about 30 log points. This implies that if the firm at the 75th percentile has productivity equal to 100 then the firm at the 25th percentile has productivity equal to 74. Foster et al. (2008) show that the dispersion of establishment-level total factor productivity within detailed product classes that abstracts from variation in plant-level prices is at least as large. Similarly, there is substantial dispersion in business size. For example, Bartelsman

2 In what follows, some of the evidence is about establishments and some is about firms. By establishments, we mean specific physical locations of production activity. By firms, we mean all activity under common operational control. As an example, an individual Wal-Mart store is an establishment while the firm is the activity of all Wal-Mart stores as well as other establishments owned and controlled by Wal-Mart (e.g. distribution facilities). Both establishment and firm-level evidence is relevant. For job reallocation, the establishment level is likely preferred since the frictions in the labour market are very much about moving workers away from one location to another. In addition, most establishment-level job reallocation is between-firm reallocation. For other purposes, analyzing activity at the firm level is preferable. For example, when discussing financial-market frictions, the relevant level of activity is the firm not the establishment. The discussion in this paper specifies whether results refer to the establishment level or the firm level. Note that theoretical models often do not make this distinction – that is they don’t formally model multi-establishment firms.

3 Foster et al. (2008) examine 11 detailed product classes for the US where direct measurement of physical output and prices
et al. (2009, 2011) show that within US industries, firms in the top quartile of the size distribution are on average 80 times larger than firms in the first quartile of the within-industry size distribution.

The large dispersion in productivity and size provide ample scope for there to be differences across countries, time periods within countries and industries within countries in “static” allocative efficiency. By the latter, we mean the extent to which in the cross section resources are allocated to their highest-valued use which in this case implies that the most productive firms should be the largest firms. Figure 2a based on Bartelsman et al. (2009, 2011) shows there are large differences in the within-industry covariance of size and productivity across countries. For example, the covariance in firm size and firm productivity in the US is high and positive while it is lower in Western Europe and still lower in Eastern Europe. Interestingly, while the covariance between size and productivity is low in Eastern Europe, Figure 2b shows that the covariance has been increasing substantially over the last couple of decades. Bartelsman et al. (2009, 2011) also show that these differences in the size/productivity covariances are potentially quite important in accounting for differences in output per capita across countries.

**Figure 2.** The relationship between size and productivity

<table>
<thead>
<tr>
<th>2a. Average within-industry covariance between size and productivity, 1992-2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
</tr>
<tr>
<td>0.6</td>
</tr>
<tr>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2b. Change in within-industry covariance between size and productivity, from 1992-1996 to 1997-2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
</tr>
<tr>
<td>0.6</td>
</tr>
<tr>
<td>0.1</td>
</tr>
</tbody>
</table>

Source: Bartelsman et al. (2011), see Tables 1 and 2

is feasible. They find that the dispersion of physical productivity is slightly larger than the dispersion of revenue productivity (essentially price times physical productivity). Interestingly, the reason is that physical productivity and price are inversely correlated at the establishment level. This latter pattern is consistent with models of product differentiation such as those in Melitz (2003) and Melitz and Ottaviano (2008).
The covariance measures depicted in Figures 2a and 2b are a component of a productivity decomposition developed by Olley and Pakes (1996). The Olley-Pakes decomposition decomposes an index of industry level productivity into an unweighted mean of productivity at the firm level and the covariance of size and productivity. Figures 3 and 4 show Olley-Pakes decompositions of within-industry productivity for Colombia (using TFP as the measure of productivity) based on a study by Eslava et al. (2004) and for China (using labour productivity) based on a study by Deng and Haltiwanger (2008). Both countries underwent substantial market reform in the sample periods for these analyses. It is striking that in both countries the covariance between size and productivity rose substantially. Moreover, it is especially striking that in 1998 the covariance between size and productivity in China was negative. The interpretation is that at that point the largest firms were relatively low productivity firms. Figure 4 suggests that an important part of China’s rapidly growing productivity is a movement of the covariance from negative to slightly above zero. A covariance at or around zero is still quite low relative to say the US, leaving considerable opportunities in China to improve allocative efficiency.

China and Colombia underwent market reform and have seen the covariance between size and productivity of their firms rise substantially.

Figure 3. Olley-Pakes decompositions of productivity for Colombian manufacturing, 1982-1998

Figure 4. Olley-Pakes decomposition of labour productivity for China, 1998-2005
The within-industry cross-sectional patterns of productivity and size across countries are of critical interest and importance but offer an incomplete picture. That is, on the basis of the cross-sectional evidence alone one might conclude that there is relatively stable within-industry size and productivity distribution in the sense that high-productivity firms remain high-productivity firms and large firms remain large firms and so on. While there is persistence in both firm size and firm productivity, there also is considerable reallocation and movements within the distributions. Estimates of the persistence of idiosyncratic or productivity shocks suggest first order yearly autocorrelation of about 0.8 (see e.g. Foster et al. 2008). Along with estimates of dispersion, this estimate of persistence implies estimates of the standard deviation of innovations to productivity shocks of about 0.20 (in terms of log total factor productivity). 4

Complementing the high variance of innovations to productivity shocks is a high pace of reallocation of outputs and inputs. Figure 5 based on Haltiwanger et al. (2010) shows an annual establishment-level gross job creation rate of about 17 percent (as a percentage of employment) and an annual establishment-level gross job destruction rate of 15 percent in the US. This implies in any given year a gross job reallocation rate of about 32 percent – that is about 32 percent of jobs are reallocated each year in the US. Figure 5 also shows that entry and exit of firms as well as entry and exit of establishments of existing firms play an important role in this reallocation. Bartelsman et al. (2009, 2011) show that such patterns are present in a range of advanced and emerging economies. In addition, Davis and Haltiwanger (1999) and Haltiwanger et al. (2010) show that much of this reallocation is within industries (about 90 percent of job reallocation in the US is within 6-digit NAICS or 4-digit SIC industries). Thus, reallocation largely reflects the contribution of business entry, exit, expansion and contraction within industries.

Figure 5. Annual job creation and destruction in the US private sector, 1980-2009 (percent of employment)

<table>
<thead>
<tr>
<th>Job Creation</th>
<th>Job Destruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuing Establishments</td>
<td>10.3</td>
</tr>
<tr>
<td>New Establishments (Existing Firms)</td>
<td>3.2</td>
</tr>
<tr>
<td>New Firms</td>
<td>3.1</td>
</tr>
<tr>
<td>Exiting Establishments (Continuing Firms)</td>
<td>2.4</td>
</tr>
<tr>
<td>Exiting Firms</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Source: Tabulations from the BDS combined with tabulations from Haltiwanger et al. (2010)

Just as there is a relationship in the cross-sectional distribution of size and productivity, there is a relationship between the pace of reallocation and productivity shocks. In well-functioning economies, outputs and inputs are being reallocated away from the lower-productivity to higher-productivity businesses. Figure 6 shows that about a third of the productivity growth within a manufacturing industry over a five-year period of time is accounted for by such reallocation in the US (this is captured by summing the middle and right bars in Figure 6). Foster et al. (2001) show that over longer horizons (ten years) the contribution is even larger (about 50 percent).

4 This statistic is consistent with the evidence in Foster et al. (2008).
Young firms exhibit an up-or-out dynamic—they either grow fast on average or they exit.

An important component of this reallocation is entry and exit of establishments and firms. Given the importance of entry, it is instructive to characterize the post-entry dynamics of young firms. Figure 7 based on Haltiwanger et al. (2010) shows how job destruction and net employment growth at the firm level vary with age of the parent firm. Among surviving firms, young firms grow very fast in absolute terms and relative to their more mature counterparts. However, the job destruction rate from firm exit is also much higher for young firms. Taken together, the implication is that young firms exhibit an “up or out” dynamic—they either grow fast on average or they exit.

How do these “up or out” dynamics relate to productivity? Figure 8 (based on Foster et al. 2006) shows the relationship between productivity and continuing and exiting for all and single-unit establishment firms in retail trade. Exiting young establishments and firms have very low productivity while surviving

---

**Figure 6.** Components of TFP growth over five-year horizons in selected US manufacturing industries, 1977-1997

<table>
<thead>
<tr>
<th>Component</th>
<th>Change in Log TFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>5.13</td>
</tr>
<tr>
<td>Within</td>
<td>3.44</td>
</tr>
<tr>
<td>Reallocation among existing establishments</td>
<td>0.35</td>
</tr>
<tr>
<td>Net entry</td>
<td>1.35</td>
</tr>
</tbody>
</table>

Source: Foster et al. (2008)

**Figure 7.** Up-or-out dynamics of young US firms

Source: Haltiwanger et al. (2010)

Note: Firm age is defined as the age of the oldest establishment. For example, firm age category “1” is based on firms where the oldest establishments entered in the prior year (start-ups are categorized as firm age equal to “0”).

---

5 These patterns show dynamics at the firm not establishment level.

6 A related key message of Haltiwanger et al. (2010) is that firm age rather than firm size is the more theoretically and empirically relevant characteristic of businesses for job creation. That is, the conventional wisdom that small businesses create most jobs is better understood in terms of the job creation prowess of business start-ups and the rapidly growing survivors in Figure 7. Business start-ups and young firms are small so the conventional wisdom is picking up the role of start-ups and young firms. On the flip side, small, mature firms are not disproportionate creators of jobs.
young firms and establishments have above average productivity. As such, the “up or out” dynamic in Figure 7 contributes to productivity growth through moving resources towards the more productive and away from the less productive young businesses.

In short, in well-functioning economies there is evidence of not only static allocative efficiency (more productive businesses are larger) but dynamic allocative efficiency (resources are being moved from less to more productive businesses). A key theme in the remainder of the paper is that the extent to which a country exhibits patterns of both static and dynamic efficiency will depend on market structure and institutions.

Another theme emphasized in this paper is that accommodating the micro volatility as evidenced by the ongoing need to reallocate workers to more efficient producers becomes disrupted in economic slumps. The nature of this disruption will be elaborated on below. In addition, micro volatility can change the nature of macro volatility. For example, periods of intense restructuring in the economy can dampen aggregate activity as resources are being used for restructuring and reallocation rather than current production. In a related fashion, periods of intense restructuring are often associated with periods of heightened uncertainty, which can slow down the adjustment dynamics from both aggregate and micro shocks. These relationships are also discussed below.

2.2 The impact on workers

As noted in the introduction, the ongoing reallocation is not costless, with workers and businesses bearing substantial time and resource costs in accommodating the reallocation, even if it is productivity enhancing. Both types of resource costs need to be taken into account in evaluating the extent to which a country is achieving static and dynamic allocative efficiency.

In good times in well-functioning economies, the impact on workers is not too adverse in terms of employment and earnings outcomes. For this purpose, we focus on the evidence in the US.\(^7\) Figures 9 and 10 (based on Davis et al. 2011) help highlight several key patterns. These figures show that in good economic times, much of the job destruction in the US takes the form of worker quits instead of layoffs.

\(^7\) See Davis et al. (2010) and references therein.
This pattern is consistent with related evidence that shows that in good economic times, many separations of workers are associated with either no spell of joblessness or a short spell of joblessness and often result in an increase in earnings relative to the prior job. The latter is consistent with the perspective that the workers are reallocating away from lower-productivity firms or low-quality matches to higher-productivity firms or matches. In good times, the typical worker switching jobs experiences an increase in earnings (see e.g. Fallick et al. 2011).

All the potential problems with the dislocation of workers are significantly exacerbated in economic downturns even in otherwise well-functioning economies. Not surprisingly, as seen in Figure 9, job destruction increases and job creation decreases in an economic downturn. Job destruction in downturns is accommodated mostly through layoffs, yielding spells of unemployment that are often protracted. The current economic downturn in the US offers ample evidence of these challenges. Figure 11 shows unemployment inflow rates and escape rates from unemployment. In normal times, the average duration of unemployment in the US is about two months (this is roughly 1 over the escape rate). In the current economic downturn, it is closer to ten months. Empirical evidence also shows that the persistent earnings losses that displaced workers experience are worse in recessions. 

---

In distorted economies, reallocation is stifled in good and bad times, with adverse effects on allocative efficiency.

Figure 11. Unemployment inflow and escape rates, US, 1967-2011 (percent)

The above discussion pertains to well-functioning economies. In highly distorted economies, reallocation is generally not well accommodated, regardless of whether we focus on expansions or contractions. One often observes an effort to stifle reallocation in these economies. This is rationalized with concerns about long-term unemployment and the impact of displacement on earnings. However, as we discuss below, stifling such reallocation has adverse effects on static and dynamic allocative efficiency.

3. Guidance from economic theory on allocative efficiency

3.1 Core models of firm dynamics

We begin with considering why firms of different sizes and productivity co-exist within an industry. One reason is that even though more productive firms have an incentive to become larger, there may be some form of decreasing returns due to economies of scope and control (e.g. Lucas 1978). Another reason is that firms produce and provide somewhat different products even in the same industry. Models of product differentiation such as those in Melitz (2003) and many antecedents have this feature. Such product variation need not be differences in physical products but can also include differences in the bundling and the way of providing the goods and services in question (including the location of delivery). For example, it may be that firms differ in their reliability and timeliness of delivery.

With such models as a backdrop, there is a rich set of models that help us understand the observed industry and firm dynamics. Jovanovic (1982) posits that, at entry, firms don’t fully know their productivity (or other aspects of profitability). Thus, an important part of firm dynamics, especially for growing industries, is the selection and learning dynamics of young firms. For example, those firms that learn they have a good location, good product or process, survive and grow. Those that learn they are not profitable contract and exit. Since the evidence on firm dynamics shows that reallocation and restructuring is not confined to young firms, additional theories need to be used to understand such dynamics. Ericson and Pakes (1995) and a variety of other papers (see Syverson 2011 for a recent survey) develop models that help account for the ongoing reallocation and productivity dynamics. Ericson and Pakes (1995) postulate that every time a firm makes a major change in its way of doing business...
(either by adopting a new technology or in responding to some major change in economic conditions like higher energy costs), the firm begins the learning and selection dynamics anew. That is, they need to learn about their profitability with the new product or process.

The more general notion as illustrated in models such as Hopenhayn (1992) and Hopenhayn and Rogerson (1993) is that the productivity shocks firms face are persistent but that firms are constantly subject to new productivity and profitability shocks. Viewed from this richer perspective, firms are constantly forced to adjust and adapt to changing economic circumstances and while their past successes can help in forecasting their ability to adjust and adapt, firms are constantly required to reinvent themselves. Those that reinvent themselves well survive and grow. Those that adapt and adjust poorly contract and exit.

3.2 Scope for misallocation

Much of the above discussion paints a picture of the potentially important role of productivity-enhancing reallocation for economic growth. More recent work has emphasized many factors that can go wrong as countries try to achieve both static and dynamic allocative efficiency. Banerjee and Duflo (2005), Restuccia and Rogerson (2008), Hsieh and Klenow (2009) and Bartelsman et al. (2009, 2011) all emphasize that there are a host of distortions to static and dynamic allocative efficiency. Such distortions include barriers to entry and exit, regulations that deter job destruction, poorly functioning product, capital and labour markets, weak rule of law, poor public infrastructure for communication and transportation, as well as problems with graft and corruption or otherwise arbitrary and capricious behaviour of governments. The consequences of such distortions can be severe. As discussed above, in a well-functioning economy the most productive firms are the largest firms. In a distorted economy with poor institutions, the largest firms may not be the most productive ones but rather the best connected or perhaps the best at navigating the distortions within a country.9

This recent literature has shown that the misallocation that results from the type of distortions discussed above can account for a substantial fraction of the observed differences in proxies for allocative efficiency (such as the size/productivity covariance discussed in Section 2) as well as differences in aggregate output and consumption per capita.

Such misallocation distortions have adverse consequences in their own right, but can also potentially yield a variety of second-best problems for economic reforms. For example, consider trade reform. While the Melitz (2003) and related models make a case why trade liberalization can yield productivity-enhancing reallocation, in the presence of these distortions the impact of piecemeal economic reforms is less clear. If it is difficult to start a business, difficult to expand, difficult to avoid having rents extracted from any profits unless one stays sufficiently small, and difficult to contract or exit, the productivity enhancing reallocation highlighted by Melitz (2003) and others can be derailed.

In like fashion, not only might the reallocation be derailed but it may be especially costly. As emphasized by Caballero and Hammour (2000), distortions can be such that creation and destruction get decoupled in time – that is, market reform (including trade reform) might induce downsizing and exit by less productive businesses but the accompanying creation and expansion by the more productive businesses may be delayed or derailed. When there is such decoupling, the cost to workers can be especially high, since in an economy with lots of destruction but not much creation (at least for a period of time) there is by construction an economic downturn with many dislocated workers.

---

9 Bartelsman et al. (2009, 2011) provide evidence on differences across countries for a wide range of distortions.
One caveat heard regarding the arguments expressed above is that the role of reallocation for productivity growth may be more of an issue for advanced market economies than emerging economies. One argument is that it is economies at the frontier of technology that are inherently engaged in the experimentation and creative destruction process. Following this reasoning, the argument for emerging economies is that, if technology could simply be brought up to levels from the past in advanced economies where methods and business practices are well understood, this would be still be a substantial improvement in emerging economies.

There are several reasons why this line of argument is not persuasive. For one, the evidence shows that in all economies (advanced and emerging) we observe large within-sector differences in productivity across businesses (see e.g. Bartelsman et al. (2009, 2011), and Hsieh and Klenow 2009). If anything, within-sector dispersion in productivity is larger in emerging economies reflecting, as Hsieh and Klenow (2009) emphasize, the effects of misallocation. The point is there is much scope for productivity-enhancing reallocation in emerging economies. Furthermore, while the sources of within-industry differences in productivity across businesses are still under investigation, it is clear that they don’t simply stem from access to different “blueprints” for how to produce specific goods and services. Rather, differences in productivity reflect differences in managerial ability, organizational capital, management practices and other intangible factors (see e.g. Corrado et al. 2005) along with potentially random factors about choosing the right combination of location, products, and processes. The implication is that productivity differences across businesses reflect idiosyncratic factors that are not simply a matter of blueprints – and that such differences are pervasive not only in high-tech and low-tech sectors but also in advanced and emerging economies.

While this discussion highlights the considerable progress made in our understanding of these issues theoretically and empirically, there remain many open questions on these issues. Identifying the potential benefits from economic reforms in terms of improved allocative efficiency and their costs in terms of transition costs and worker dislocation is an active area of research.

3.3 Different dimensions of volatility

Much of the discussion about volatility has focused on two dimensions of volatility: First, the large dispersion of productivity/profitability across businesses; and second, the ongoing reallocation of outputs and inputs across businesses. In terms of the latter, it is useful to note that such reallocation reflects an important form of dispersion across businesses – specifically, dispersion in output and input growth rates across businesses. That is, reallocation reflects resources flowing from contracting businesses (those with negative growth rates in outputs and inputs) to expanding businesses (those with positive growth rates in outputs and inputs). Entry and exit rates represent the extremes of the output and input growth rate distributions and obviously contribute substantially to volatility.

It is natural to focus on dispersion in profitability/productivity on the one hand and dispersion in output and input growth rates on the other hand. The core models discussed in Section 3.1 largely treat the dispersion in productivity/profitability as exogenous while treating the dispersion of output and input growth rates as endogenous. As highlighted in the discussion of Sections 3.1 and 3.2, a critical factor impacting aggregate outcomes is how well an economy accommodates the idiosyncratic productivity/profitability shocks: Are those businesses with favourable shocks growing and those with less favourable shocks shrinking and is such reallocation accomplished without too much disruption?

There are other closely related dimensions of volatility. An obvious candidate is dispersion in earnings across workers. It is well known that in advanced economies there has been an increase in the dispersion of the level of earnings across workers – and the evidence suggests this is associated with changing
Much of the increase in personal-income inequality in the US reflects increased dispersion in productivity and earnings between establishments.

technology favouring more skilled workers (i.e. skill-biased technological change) as well as corresponding changes in trade patterns (the off-shoring of lower-skilled jobs). This rise in earnings inequality is closely related to the discussion on firm dynamics in prior sections. For example, a number of studies (e.g. Davis and Haltiwanger, 1991, Dunne et al. (2004) and Barth et al. (2010)) have found that much of the increase in earnings inequality in the US is associated with an increase in the between-establishment dispersion in earnings. Moreover, these studies show that the establishments with higher earnings are more productive, more highly skilled and more likely to have adopted advanced technology.

What do we know about changes in volatility over time as well as differences in volatility across countries? They may reflect many factors. Variations in volatility may reflect changes in the driving forces of profitability as well as changes in the adjustment dynamics. For the latter, an important issue in the current context is whether the differences in volatility reflect the relative flexibility of an economy. Greater flexibility might take many different forms. It might be that workers in a more flexible economy are more geographically mobile so that there is even more reallocation of labour in response to a given set of shocks. Alternatively, it might be that wages become more flexible (e.g. with greater reliance on flexible-pay mechanisms) so that a given set of shocks is reflected more in wages than in the reallocation of employment. These examples highlight the fact that appropriate caution is needed in assessing differences in measures of volatility across time and countries.

The evidence on changes in volatility is primarily for the US which has extensive longitudinal panels of businesses and workers covering many decades. For the US, there is evidence that the volatility of output and employment growth rates of publicly traded firms has increased over the last few decades (see e.g. Comin and Phillippon 2006). However, interestingly when the entire economy is considered (in the US, publicly traded firms account for about 30 percent of employment and 40 percent of output), there is actually a pronounced decline in the volatility of employment growth rates (see Davis et al. 2007, Davis et al. 2010, Davis et al. 2011, and Haltiwanger et al. 2011). The evidence in Figure 12, drawn from Haltiwanger et al. (2011), shows that a decline in the pace of business start-ups accompanies this decline in volatility. On average, start-ups accounted for 3.5 percent of employment annually in the 1980s, 3.0 in the 1990s and declined to 2.6 percent post-2000. This reflects a decline in new jobs from start-ups of more than one million jobs per year. Given that the average net increase in jobs is only a little over two million jobs this is a substantial decline.

Figure 12. Declining business dynamism in the US: Trends in gross flows and net job creation, 1980-2009

Source: Haltiwanger et al. (2011)
Does the lower volatility imply the US has become less flexible over time? This is an open research question. There is some evidence that it may reflect in part a different form of flexibility. Lemieux et al. (2009) show there has been a pronounced increase in the use of flexible-pay mechanisms (bonus pay, stock options, etc.) in the US so this may reflect increased earnings flexibility. However, the evidence in Davis et al. (2007) suggests this is unlikely to be the whole story. For example, they find that the lower volatility in part reflects the increasing shift in sectors like retail trade to large, national firms (e.g. Wal-Mart) who are much less volatile than Mom-and-Pop retailers. There is evidence that the shift to large, national chains reflects the type of technological change and reallocation discussed in prior sections as large, national chains have been able to take greater advantage of advances in information technology for distribution networks and inventory control. However, it may also be that large, national chains are less nimble in adjusting to changing economic conditions. The more general point is that a decline in volatility in the US may reflect a less dynamic US economy (and thus an economy less able to respond to changing economic conditions).

The findings of a decline in the pace of business start-ups raise related concerns. If start-ups and young businesses are an important source of experimentation and innovation, the fall in volatility associated with the decline in start-ups may bode ill for future US growth. It is an open question why there has been a secular decline in the pace of start-ups. It may be related to an increased fraction of activity being accounted for by large, multi-national firms as discussed above.

There is also evidence that volatility increased dramatically in the 1990s in the transition economies (see e.g. Faggio and Konings 1999, Jurajda and Terrell 2002 and Haltiwanger and Vodopivec 2003). When looking at the performance of these economies, it was clear that this was disruptive with adverse aggregate consequences as most transition economies first experienced a downturn in aggregate economic activity. Moreover, the evidence suggests that there was a non-trivial lag between the burst of job destruction and job separations early in the reforms and the subsequent recovery of job creation and hires. The patterns exhibited in these economies were consistent with the discussion and concerns about decoupling of job creation and destruction in Section 3.2. Still, the evidence is that, for the most part, the transition economies weathered this storm and recovered with robust growth. It likely helped that the world economy exhibited robust growth in the second half of the 1990s.

Another important issue in terms of changes in volatility over time within countries is that periods of more intense restructuring are often associated with periods of heightened uncertainty. Bloom (2009) has stressed that recessions differ in the extent of uncertainty, which impacts how fast the economy recovers from the downturn. Bloom et al. (2010) have emphasized that the Great Recession of 2007-09 is a period of especially heightened uncertainty due to the collapse of financial markets and the accompanying intense period of restructuring associated with the downturn (e.g. shifts away from construction activity and restructuring of financial markets). Such heightened uncertainty contributes to particularly slow recoveries, since even businesses with potential profit opportunities are reluctant to invest and hire new workers under these circumstances.

Comparing the level of volatility across countries has proven to have substantial measurement and conceptual challenges. The working conjecture is that the US, being a very flexible economy, would have a higher dispersion of growth rates of outputs and inputs than other countries. However, the evidence on this is mixed. Part of the reason for this is measurement difficulties (see Bartelsman et al. (2009, 2011)). However, another reason might be flexibility manifesting itself in different dimensions. As Bertola and Rogerson (1997) emphasize, countries with rigid labour regulations also often have centralized wage bargaining. The former should dampen employment volatility while the latter should increase employment volatility, so the final outcome is ambiguous.
This discussion of different dimensions of volatility highlights the difficulties of simply comparing measures of volatility across countries or across time. As discussed earlier, one approach that overcomes the measurement and conceptual challenges of comparing measures of volatility is to focus on whether the volatility (reallocation) is productivity enhancing. Differences across time and across countries on whether reallocation is productivity enhancing are of unambiguous importance. Another approach to making cross-country comparisons is to focus on the impact of the business climate on volatility. For example, Haltiwanger, Scarpetta and Schweiger (2010) use differences in volatility across industries and size classes within countries to show that countries with more rigid labour markets have less employment reallocation. One can identify this effect not with the cross-country variation but with the within-country variation between industries and size classes.

4. Policy lessons and challenges

The policy lessons in broad terms are clear but the actual implementation imposes many challenges. The broad policy lesson is that, in order to function well, the economy needs to be sufficiently flexible to permit productivity-enhancing reallocation while minimizing the disruption costs from such reallocation in a manner that does not stifle the reallocation. Few countries achieve the economic environment that is consistent with this broad lesson. One could argue that the US has the market structure and economic institutions that closely approximate this objective in normal economic times. But the recent Great Recession has reminded us that even in the US, there is fragility in the system and disruptions in key markets (like financial markets) affect the nature and consequences of the natural economic volatility that is part of any ongoing process of technological progress. Thus, one of the policy challenges is how to maintain the market structure and economic institutions that operate in normal economic times while permitting intervention when markets get disrupted. This challenge of countercyclical policy is not the primary focus of this paper but we discuss some issues along these lines below.

For emerging economies, the challenges are potentially enormous. As discussed in Pagés (2010) and Pagés et al. (2009), one important challenge evident in many emerging economies is the role of informality, which can often also go hand-in-hand with what they call the “missing middle”. In highly distorted economies where the burden of poor institutions and market distortions weigh down on businesses, there tend to be very small businesses, very large businesses but not as many medium-sized businesses as in well-functioning market economies. The authors argue the reason for this phenomenon is that only the very large businesses have the resources to deal with the highly distorted economic environment (or worse, are large simply because of the existing distortions, which are often associated with crony capitalism and government-private sector connections). They argue that small businesses – even those with great potential in terms of productivity – stay small to fly below the radar. That is, businesses stay small and informal to circumvent the regulatory burden and corruption practices that often characterize countries with weak institutions.\textsuperscript{10}

Reallocation has little chance of being productivity-enhancing in highly-distorted economic environments. The challenges then are that many components need to be simultaneously in place for economies to successfully grow. This is particularly the case when undertaking market reforms. These challenges are present in both emerging and advanced market economies. The list of components for “success” for any country is long:

\textsuperscript{10} There may be an ameliorating effect on the duration of joblessness in economies with large informal sectors to the extent that workers dislocated by restructuring and reallocation can quickly find jobs in the informal sector. It is not clear that this is indeed beneficial to the extent it reflects workers and firms in the informal sector being underemployed for the reasons discussed in the text.
• Labour markets need to be sufficiently flexible to permit reallocating workers from less productive to more productive establishments without long spells of unemployment.

• As part of this flexibility, safety nets need to be in place so that workers adversely impacted by reallocation can be assisted in finding new employment without distorting the process of reallocation.

• The infrastructure needs to be of sufficiently high quality to insure that existing and starting-up businesses that seek to grow are not thwarted by factors such as poor transportation and communication infrastructure.

• Product markets need to be sufficiently competitive so that firms are not large for reasons related to market power or due to favourable treatment by the government.

• Financial markets need to be sufficiently developed to provide funding to starting-up and expanding businesses and to be able to deal with the inevitable failure of young and small businesses.

• Regulation has to provide appropriate oversight without imposing onerous time and resource costs on individuals who start or close down a business.

• The legal system has to work sufficiently well so that property rights are well established and bankruptcy and business failure can be accommodated.

• The rule of law and the role of the government need to be such that graft, corruption and other forms of criminal activity don’t thwart private sector businesses from starting and growing.

These are just examples of the many components that need to be in place for an economy to operate efficiently. With all of these components in place, opening up to markets and competing in world markets is much more likely to be productivity-enhancing without the costs of reallocation being too high for businesses and workers.

Getting all these pieces in place simultaneously is obviously a challenge on many dimensions. Given such challenges, governments often try to intervene to facilitate growth and/or to protect workers and businesses from some of the adverse effects of volatility. The message of this paper is that policies and institutions that stifle reallocation can yield very poor outcomes. Another related message of the paper is that well-intended industrial policies that try to aid the private sector must confront the facts associated with the large dispersion of productivity across businesses (and the associated productivity-enhancing reallocation that works in well-functioning market economies). Recall that dispersion of productivity in narrowly defined sectors in advanced economies like the US is very large and even larger in less developed economies. Industrial policies that (perhaps inadvertently) support the low-productivity businesses in a sector will lower aggregate productivity and make it difficult for a country to increase its productivity over time if, for example, it is difficult for governments to let go companies they have supported in the past. The government is in no better position than the market to pick winners and, given the evidence on dispersion, the risks of picking and supporting low-productivity businesses is non-trivial. As an alternative to industrial policies, policies that seek to address the distortions and market failures in the country may be more promising.

Another challenge is how to handle crises. Crises tend to distort the dynamics of reallocation and restructuring, regardless of whether we look at advanced or emerging economies. In crises there is lots of job destruction but not much job creation, with accompanying high unemployment. In financial crises, financial markets are not facilitating the reallocation of resources away from less productive to more productive businesses.11 Such productivity-enhancing reallocation requires financial markets

11 A recent paper that explores these issues is Eslava et al. (2010). They find that exits are less related to productivity in times of financial crises.
providing funding to start-ups and to young, small businesses which have the potential to grow fast. Such credit channels break down in recessions that are associated with financial crises.

5. Concluding remarks

The evidence strongly supports the view that static and dynamic allocative efficiency is critical for the aggregate economic performance of a country. In the cross-section, we observe a very dispersed and skewed size distribution of activity in advanced market economies that is accompanied by a very dispersed and skewed distribution of productivity. In a well-functioning economy, these two distributions should be strongly positively correlated – that is, the most productive businesses should be the largest businesses. In addition, in a well-functioning economy, resources tend to be reallocated from less productive businesses to more productive businesses. The evidence shows there is considerable variation across countries in the extent to which size and productivity are correlated and reallocation is productivity enhancing.

The evidence shows that countries that undergo market reform improve their static and dynamic allocative efficiency and, in turn, achieve higher productivity. The covariance between size and productivity rises in response to market reform and market selection improves with market reform. By market selection, we mean that less productive businesses are more likely to exit and more productive businesses are more likely to survive. This improved market selection contributes positively and substantially to productivity growth.

Many things can go wrong that either mitigate or potentially limit the gains from market reform. In a highly distorted economy, there are second-best problems so that piecemeal market reform will not be as effective. Distortions may arise in the legal system and the rule of law as well as in regulation and in product, labour and financial markets. A poorly-functioning labour market makes the response to reallocation very costly. Reallocation yields inherent costs on businesses and workers as it induces workers to relocate across businesses. In a poorly-functioning labour market, this can be very costly.

Even in advanced market economies that are normally well-functioning, the reallocation dynamics of workers can become distorted in severe economic downturns. Addressing the difficulties of managing reallocation dynamics during economic downturns without distorting the potential for productivity enhancing reallocation in the long run is a continuing challenge. The recent crisis has highlighted the importance of well-functioning financial markets. In times of financial crises, financial markets are less able to facilitate the selection and growth dynamics of businesses – for large, mature as well as young and small businesses alike. Perhaps ironically, the globalization of financial markets has made the problem more challenging during economic crises given the flight to quality becoming increasingly global during crises.

Financial regulation that helps monitor the health of the financial services industry and provides safeguards against financial collapses is undoubtedly needed. Some caution about how to design such safeguards is provided by the underlying message of this paper. Successful new, young firms need equity investors. The development of venture capital, angel financing, and other markets that target start-ups and young and small businesses has facilitated productivity-enhancing reallocation. Financial-sector reform should avoid increasing the barriers to the financial sector in finding new instruments and creative ways of providing funding to high growth businesses and, more generally, to productivity-enhancing reallocation.

The recent economic crisis has also highlighted the potential importance of heightened uncertainty during economic crises in dampening the pace of economic recovery. The key insight from economic
theory that also enjoys empirical support, especially in the recent crisis, is that heightened uncertainty will slow down recoveries due to caution and waiting effects. That is, even businesses with profit opportunities will delay and/or reduce the amount of investment and hiring due to heightened uncertainty.

Is the discussion in this paper on productivity and firm dynamics related to the slow recovery of the US economy from the recent economic crisis? It may be. There is indeed a potentially disturbing trend in business dynamism in the US: Whilst having achieved rapid productivity and job growth with a high pace of reallocation, the US has been experiencing a secular decline in the pace of business dynamism highlighted by a secular decline in start-ups over the last three decades. The decline has become especially pronounced in the post-2000 period. While understanding this decline is an open research question, it bodes ill for future US growth given that flexibility, business dynamism and a high pace of reallocation contributed so much to US growth over the past several decades. Moreover, the decline in flexibility would also imply the US will struggle to restructure and reinvent itself in the way which it successfully did in previous crises. This may be a contributing factor to the anaemic US recovery from the Great Recession.
References


