Increasing Differences Between Firms: Market Power and the Macroeconomy

John Van Reenen

I. Introduction

Increasingly, researchers have been using microeconomic data on firm differences to better understand macroeconomic outcomes. It is now widely accepted that there is enormous cross-sectional heterogeneity in establishment performance even within narrowly defined industries. But what is less recognized is that these inequalities between firms have been increasing over time, not just in terms of productivity, but also in terms of sales—leading to a substantial increase in sales concentration across a wide range of markets. This polarization has also been evident in wages, with almost all of the increase in earnings inequality happening between firms with rather little change in inequality within firms.

The aim of this paper is to document some of the emerging facts on firm heterogeneity especially in terms of recent changes. It is designed to “set the scene” for discussions in terms of macro implications, causes and policy responses. I do not aim to give definitive answers to what these changes imply in terms of specific policies, but I do speculate on some potential policy implications toward the end of the paper.
A theme of this paper is whether the increasing gap between large and small firms reflects an increase in market power due to a reduction in competition arising from (for example) weakened antitrust enforcement. The fall in the labor share of gross domestic product (GDP) and the rise in estimated aggregate price-cost markups are consistent with a rise in market power. A generalized rise in market power can have several important (and often harmful) macroeconomic consequences. For example, there is much evidence that weakened competition undermines productive efficiency and it is clear that productivity growth has been very disappointing over the last decade across Organisation for Economic Co-operation and Development (OECD) countries. Even without cost inefficiencies, greater monopoly power is still a negative supply shock to the economy. In the long run real wages will need to fall to restore an inflation rate consistent with the central bank’s target. This will mean a lower participation rate (as work is less attractive) and/or higher unemployment.

Although such arguments over rising market power from lax antitrust enforcement should be taken seriously, I suggest a more nuanced view. There are other explanations of the increasing differences that do not rest on a generalized fall in product market competition. Indeed, an equally strong case could be made that the forces of globalization and new technologies have changed the nature of competition without necessarily diminishing it across the board. For example, if more markets are becoming “winner take all” as with digital platform competition, this will generate the dominance of “superstar firms” such as Amazon, Apple, Facebook, Google and Microsoft. The success of such firms may be as much due to intensified competition “for the market” rather than anti-competitive mergers or collusion “in the market.” Furthermore, even in lower-tech markets like retail and wholesale, rapid falls in quality-adjusted information and communication technologies (ICT) prices may give larger firms—who can invest heavily in developing proprietary software—major advantages in logistics and inventory control management.

We will discuss several pieces of evidence that are suggestive of some role for the superstar firm hypothesis. First, using firm-level data to decompose the changes in aggregate markups and labor shares, the vast
majority of the changes are due to reallocation between firms toward larger, more productive and profitable firms. Most American firms have seen either no increase or a fall in their markups and labor shares. Second, the industries growing most concentrated appear to have rising productivity and innovation which is consistent with reallocation to more efficient and innovative firms. Third, the qualitative trends of concentration and markups seem similar across countries, which suggests global changes, rather than country specific institutional changes such as the relative weakening of U.S. competition policy compared to Europe. None of these are dispositive, so we also look at other explanations—such as an increasing role for intangible capital.

I focus on long-run secular changes rather than how price-cost markups change over the business cycle. Short-run fluctuations in markups are a very important research area, especially for New Keynesian models, which focus on sticky prices implying countercyclical markups. The empirical evidence, however, casts doubt on this as markups appear acyclical or even pro-cyclical (Nekarda and Ramey 2013). Some of the methodological issues in estimating markups will also appear in the discussion below.

The structure of this paper is as follows. Section II lays out some facts on the level of firm heterogeneity and Section III focuses on changes over time. In Section IV, I discuss alternative explanations of the increase in concentration and markups based on (i) declining competition and (ii) superstar firms. Section V gives some tentative policy implications and Section VI concludes.

II. Productivity Variation

II.i. TFP Differences at a Point of Time

Chart 1 shows GDP per worker and total factor productivity (TFP) for a large number of countries where the values are normalized to be 1 in the United States (so a number like ½ on the vertical axis implies that a country has 50 percent of the TFP of the United States). Two things stand out. First, it is clear that those countries with high TFP are also the countries with high GDP per worker, implying that capital accumulation cannot explain all of the differential wealth of nations. Second, it is striking that there is such a wide dispersion in TFP.
Taken literally this suggests that it takes a Liberian worker a month to produce what an American worker can produce in a day with the same inputs.

Many scholars have looked to firm heterogeneity to understand these stark developmental differences (e.g. see the review in Hopenhayn 2014). Firm heterogeneity has a long history in social science. The first systematic empirical analysis focused on the firm size distribution measured by employment, sales or assets. Gibrat (1931) characterized the size distribution as approximately log normal and sought to explain this with reference to simple statistical models of growth. In fact, the firm size distribution is closer to a power law is now well documented in every country in the world where data are available.3

Today we are lucky to live in a world of large-scale (frequently near population) data on firms and establishments.4 These are usually from national statistical agencies that collect microdata primarily to build aggregate information either at the industry or macro level. Increasingly, researchers have been granted confidential access to such government data (e.g. the Longitudinal Business Database of U.S. establishments). A second major source is from the private sector. For example, Bureau van Dijk’s Orbis

---

Notes: 2010 data; Total Factor Productivity (Labor augmenting with common share $\alpha = 1/3$).
Sources: Penn World Tables 8.0; Jones (2015).
dataset has accounting panel data from about 200 million firms from almost every country in the world (e.g., Bajgar et al. 2018). Since this includes private companies, researchers can move beyond a focus on simply publicly listed firms such as Compustat. Liberalization of access to administrative data and rapid increases in computer power have enormously enhanced our capacity to store and interrogate these new microdata sources.

One of the robust facts emerging from the analysis of large-scale firm-level databases is the very high degree of productivity dispersion. For example, Syverson (2004, 2011) analyzed labor productivity (value-added per worker) in U.S. manufacturing establishments in the 1997 U.S. Economic Census and shows that on average, an establishment at the 90th percentile of the distribution had four times higher labor productivity than one at the 10th percentile in the same four-digit sector. Even after controlling for other inputs the TFPR ratio is still 2 to 1. Bartelsman, Haltiwanger and Scarpetta (2013) show that even larger differences exist for other OECD countries.

The aggregate TFP differences documented in Chart 1 are influenced by how different economies allocate output to firms of different productivity levels. For example, India, China and Mexico appear to allow less efficient firms to survive for much longer than in the U.S. (Hsieh and Klenow 2009, 2014). A large number of possible explanations present themselves that we will later examine, such as competitive intensity in the product market, labor, capital or housing market frictions, size-related regulations and other distortions due to corruption and arbitrary tax.

**II.ii. Aggregate Changes in Productivity Over Time**

Solow (1957) found that the vast majority of the growth of output per worker in the United States was due to growth in TFP rather than capital accumulation. The finding that TFP is at least as important as observable factors of production in such growth accounting exercises has been replicated for numerous countries. The traditional view is to analyze the productivity growth of a representative firm. This *within-firm* growth could be from innovation expanding the
technological frontier outward or from the adoption of existing ideas by follower firms.

By contrast, the Schumpeterian tradition has long emphasized the between-firm component. Much of aggregate productivity growth is from the reallocation of output away from less productive firms toward more productive firms. This reallocation can take place on the extensive margin as less productive firms exit and more productive firms enter. This is the traditional notion of creative destruction, which is a Darwinian force of natural selection. But reallocation can also take place on the intensive margin as market shares get reallocated among incumbents away from the least efficient and toward the more efficient firms. In either case these are between-firm effects that are distinct from the traditional within-firm effects.

Bailey, Hulten and Campbell (1992) found that over a five-year period about half of a typical U.S. industry’s TFP growth was due to the reallocation of output between establishments rather than ongoing incumbent within plant productivity growth. There are multiple ways in which to perform such statistical decompositions of productivity growth into within-firm and between-firm components. But whatever way this is performed, subsequent work has confirmed that there is always a substantial between-firm/reallocation component.

II.iii. Where Do Firm Productivity Differences Come From?

What could explain these differences in productivity, and how can they persist in a competitive industry? One explanation is that the differences are just measurement error. Although there is certainly some measurement error, substantial productivity differences show up clearly even for quite homogeneous well-measured goods. It is also the case that higher measured TFP is positively related to firm size, growth and survival probabilities, which suggests that there is some signal. Further, there is substantial persistence in productivity, which would not be the case if measured TFP was just transitory errors.

What lies behind these firm-level TFP differences? There are two levels to addressing this issue. One level is the proximate causes of the differences and the second is more fundamental causes. This is like peeling the layers of an onion. If we discovered that all labor
productivity differences were due to fixed capital like plant and machinery (i.e., no TFP differences) we would then have to address the question of why these differed. But at least observable capital would give us a proximate explanation. Consider one of the proximate causes of TFP differences—new technologies. The generation of technological innovation (as proxied by measures of R&D or citation-weighted patents) or the adoption of technologies (e.g., robotics, AI, other ICTs, hybrid corn, new drugs, etc.) would therefore be the things to focus on. There is a huge literature on such observable measures of innovation and diffusion and their relationship to firm performance.

Differences in the generation and use of new technologies, however, are not able to fully account for firm productivity spreads. First, even after controlling for a host of observable technology measures there remains a very large TFP residual. Second, the impact of observable technologies seems to vary systematically with the management and organization of the firm. This has most clearly been seen in studies of the effect of ICT on productivity (e.g., Bresnahan, Bjornjolfsson and Hitt 2002; Bloom, Sadun and Van Reenen 2012). There is a very wide range of effects of ICT on productivity and the impact depends very much on the organizational and management practice in the firm.

There is now a considerable body of evidence suggesting that management practices are a major reason for TFP differences. To measure management practices, we developed a methodology (World Management Survey, WMS) first described in Bloom and Van Reenen (2007). In summary, we use an interview-based evaluation tool that defines and scores from 1 (“worst practice”) to 5 (“best practice”) across 18 key management practices relating to data collection/monitoring, targets and incentives. The WMS now covers 34 countries and multiple sectors (e.g., manufacturing, retail, health care and education). More recently, we have developed the questions in a form that can be sent out with the standard Census Bureau surveys. These Management and Organizational Practice Surveys (MOPS) are being run in nine countries, with the largest samples in the United States
These surveys show large heterogeneity in firm-level management quality within every country. For example, Chart 2 shows this firm-level dispersion for the different countries in the WMS, which broadly mimics the variation observed in productivity. As with the average productivity levels in Chart 1, the United States has a very high management score, but there is large variation within the United States and indeed every country. The American advantage over India is not because every U.S. firm has managerial superiority over every Indian firm.

These management differences are strongly correlated with measures of firm performance such as productivity, size and survival. And a number of randomized control trials and natural experiments suggests that this correlation is causal (e.g., Bloom et al. 2013, Bruhn et al. 2017 and Giorcelli 2017). Across countries, around a third of the aggregate TFP differences in Chart 1 are accounted for by the management scores, and around half of the U.S.-EU TFP differences in the decade after 1995 (see Bloom, Sadun and Van Reenen 2017).

If these management and technological capabilities are so beneficial to firms, why are they not quickly imitated? There is certainly diffusion of capabilities over time, but there are many barriers to diffusion. This is the central topic of organizational economics which emphasizes issues of information, complementarities, incentives and collective action problems within the firm (Gibbons and Roberts 2013, Rivkin 2000).

III. Increasing Differences Between Firms Over Time

Having established the fact of a large degree of firm heterogeneity in the cross section, I now move to a less well-known set of facts—that there has been an increase in many dimensions of firm inequality in the last few decades.

III.i. Concentration Trends

As discussed in the previous section, the most basic measure of heterogeneity is firm size. Gabaix and Landier (2010) were among
Chart 2
Management Varies Heavily Within Countries

Note: Firm level average management scores, 1 (worst practice) to 5 (best practice).
the first to argue for an increase in average firm size in their explanation of increasing CEO pay. However, their measure of firm size was stock market value, which can fluctuate for many reasons unrelated to fundamentals. But more critically, their sample was only over U.S. publicly listed firms from Compustat. This has the problem that (i) it covers only around 30 percent of U.S. employment (the very largest firms) and (ii) the Compustat sample changes in a very nonrandom way over time due to declining IPOs and the addition of many smaller high tech start-ups (see Davis et al. 2007).

In Autor et al. (2017, 2018) we calculate measures of concentration for a consistent set of four-digit (SIC) industries using the U.S. Economic Census between 1982 and 2012. We do this for the six main Economic Census sectors that cover over 80 percent of private sector jobs. Chart 3 presents the data for each sector showing weighted average changes in four-firm (CR4) and 20-firm (CR20) concentration measures. The data show sharp increases in concentration across the whole U.S. economy in the last 30 years, with the growth generally stronger in the second half of the sample. A similar picture of generally rising concentration emerges from alternative measures such as the Herfindahl Index or CR1.

This finding is consistent with other papers examining concentration trends. Council of Economic Advisers (2016) discuss various industry studies such as airlines and telecoms which have shown significant consolidation. Grullon et al. (2016) also conclude that more than three-quarters of U.S. industries have had an increase in concentration since 1997.

Chart 3 shows that concentration has risen across industries within all six of the Economic Census sectors in a broadly secular way. Notice that the upper lines are for sales and the lower ones are for employment. Both measures of concentration have tended to rise, but sales concentration is greater than employment concentration at any point of time. This illustrates that many firms have “scale without mass”—e.g., many high-revenue firms like Google and Facebook have relatively few employees (see Brynjolfsson et al. 2008). Furthermore, unlike other sectors, in manufacturing the growth in employment concentration is pretty flat compared to sales concentration. Some
Chart 3
Rising Concentration in the United States

A. Manufacturing Sector

B. Retail Trade

Notes: Weighted average of four-digit industries within each large sector. Manufacturing: 388 inds; Retail: 58. Source: Autor, Dorn, Katz, Patterson and Van Reenen (2017).
C. Services

D. Wholesale Trade

Notes: Weighted average of four-digit industries within each large sector. Wholesale: 56. Services: 95.
Source: Autor, Dorn, Katz, Patterson and Van Reenen (2017).
E. Utilities Plus Transportation Sector

<table>
<thead>
<tr>
<th>Year</th>
<th>Top Four Concentration</th>
<th>Top 20 Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F. Finance Sector

<table>
<thead>
<tr>
<th>Year</th>
<th>Top Four Concentration</th>
<th>Top 20 Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Weighted average of four-digit industries within each large sector. Utilities & Transport: 48; Finance 31
Source: Autor, Dorn, Katz, Patterson and Van Reenen (2017).
commentators have underestimated the growth of firm concentration by focusing on employment instead of sales (or value added) and on just the manufacturing sector instead of the whole economy.

Another fact about concentration is that the trends are still there, but are more muted when we aggregate over the whole economy (or even look at the two- or three-digit level industry level). This illustrates that it is not simply growth of conglomerates that dominate multiple industries causing these trends, but rather that firms are focusing in particular sectors where their core competency is stronger.

Finally, note that one might be less concerned if this increased concentration was accompanied by an increase in firm mobility. In other words, it could be very different firms who were in the top four in an industry in one year compared to being in the top four five years earlier. There is certainly turnover at the top. However, there is no evidence in our data that this turnover has increased over time. If anything, firms in the top group in one Census year are increasingly likely to remain there five years later (so-called “persistent dominance”).

What about other countries? The most comprehensive analysis of this issue is in a series of OECD reports using MULTIPROD, which is a firm-level database they have constructed in cooperation with the national statistical agencies in many countries and so are broadly similar to the U.S. Economic Census in Chart 3. Chart 4 is taken from Criscuolo (2018) who uses these data to show that, on average, within the nine EU countries where comprehensive data are available, sales concentration has risen since 2000. This remains true when adding other non-EU OECD countries such as Australia, Japan and Switzerland. Some of the countries are small relative to the United States, so one might be concerned that the relevant market is geographically much wider. So an alternative is to look at concentration in each industry across Europe as a whole. Using this alternative measure the OECD also find increases in concentration since 2000.

III.ii. Productivity Dispersion Trends

The increase in productivity dispersion in the United States has been pointed out in several papers (e.g., Decker, Haltiwanger, Jarmin
Increasing Differences Between Firms: Market Power and the Macroeconomy

Chart 4
Like U.S., Sales Concentration has Increased in the EU

Note: Year effects from regressions with country-industry dummies and year dummies (BEL, DEU, DNK, FIN, FRA, HUN, NOR, PRT, SWE).

and Miranda 2018)—see Chart 5 for example. Decker et al. (2018) see this as one of many indicators suggesting a fall in U.S. business dynamism (particularly since 2000). These indicators include a fall in the fraction of workers employed in young firms; slowing job mobility across firms and geographical areas; slower employment responsiveness of firms to TFP shocks and a falling number of start-ups. Bils, Klenow and Ruane (2017) have argued that observed increases in productivity dispersion are misleading because they could be generated by increased measurement error. This seems unlikely, however, as the LBD series in Chart 5 is based on administrative rather than survey data and there is little direct evidence that classical measurement error has increased over time. Furthermore, as White et al. (2018) show, although there are errors in the raw Census data, the extensive cleaning and imputations performed by the U.S. Census actually tend to underestimate the true level of productivity dispersion rather than overestimate it.

Again, the OECD has examined this in 14 other developed countries. Chart 6 (Andrews et al. 2017) is also taken from MULTI-PROD administrative data and documents an increase in both labor
Chart 5
Rising U.S. Productivity Dispersion (manufacturing)

Notes: Standard Deviation of log(real sales/employment) normalized in a NAICS six digit industry-year. HP filtered series in dashed lines. LBD is population whereas ASM is corrected for sample selection. Weights are employment weights. Source: Decker, Haltiwanger, Jarmin and Miranda (2018, figure A6).

Chart 6
Change in Firm-Level Productivity Dispersion, 2001-12

Increasing Differences Between Firms: Market Power and the Macroeconomy

Productivity and TFP dispersion, qualitatively similar to the U.S. trends in the previous graphs (they also find these patterns in BVD Orbis company accounts).

III.iii. Trends in Firm-Level Pay Dispersion

Labor economists have long pointed to large wage differences across individuals and, more recently, to the big increase in individual earnings inequality that began in the United States in the late 1970s. Most other countries followed, some more quickly (like the U.K.) than others (e.g., Card et al. 2013, show that the increase in German inequality only really started in the mid-1990s). A leading factor behind the wage inequality increase is the increasing return to skill, mainly driven by skill-biased technical change. Institutional changes such as declining union power and falls in the real value of the minimum wage also played a role.

Many authors long suspected that there was a large between-firm component to wage inequality (e.g., Davis and Haltiwanger 1991, Faggio et al. 2010), but this was difficult to convincingly demonstrate as matched comprehensive employer-employee data were unavailable in those countries where inequality had grown the most. This meant that one could not track within firm wage inequality across individuals. Further, the average wage at the plant level in Census type data is payroll divided by workers (or hours) and this can change due to changing worker composition (e.g., by skill group, age, etc.—variables which most Census data do not track).

Recently, Song et al. (2017) have used U.S. Social Security Administration (SSA) data which are near population data following the same workers over time and across firms. This allows them to decompose the overall increase in individual earnings inequality into between firm and within firm components. Chart 7 shows that just about all of the increase in earnings inequality has happened between firms rather than within firms (except maybe for the top percentile, dominated by the CEO). In other words the oft-cited differences within companies between high- and low-paid workers explain very little of the increase in overall U.S. earnings inequality.
There could be many reasons for this—Song et al. (2017) emphasize sorting of workers of different skill types between firms, with highly skilled workers increasingly only working in firms employing other highly skilled workers (e.g., McKinsey vs. McDonald’s). But another reason could be that employees of high-performing firms share in more of the quasi-rents generated by these firms. For example, if a firm creates a new technology and thereby has a temporary market advantage, it will earn some above-market return and workers may share in this. There is strong evidence that such rent-sharing is important, especially in innovative firms (see Van Reenen 1996, Kline et al. 2017 and Card and Kline 2018 for a recent survey).

Indeed, in their study of Germany using similar social security data, Card, Henning and Kline (2013) also find that a big part of the increase in individual earnings inequality is also due to between-establishment differences. But they argue that a large chunk of this is due to workers sharing rents with increasingly different firms rather than just worker sorting.
III.iv. Summary on Increasing Differences

Not only are there enormous differences between firms at a point of time in terms of performance (as shown in Section II), these differences appear to be increasing over time in the United States and other developed countries. I have focused on the evidence for dispersion in size, productivity and wages—three key indicators—but there is also evidence of increased dispersion of other firm measures such as rates of return on investment in public firms (Council of Economic Advisers 2016).

IV. The Weakening Competition Explanation

One popular explanation of the patterns documented in the previous section is that they reflect a generalized increase in product market power. Competition authorities often use the degree of sales concentration as a signal of market power—for example mergers between large firms are subject to additional scrutiny if they are in very concentrated markets. The fact that the productivity gap has widened could reflect anti-competitive practices to prevent the less productive firms from catching up. And rent-sharing would mean a privileged elite of workers were able to share in the bounty from monopolistic practices (an example might be the high pay of workers in the financial services sector).

In this section we evaluate the weakening competition argument by examining the evidence on markups (IV.i); look at possible causes of declining competition (IV.ii) and briefly mention welfare implications (IV.iii), which we link more explicitly with policy in Section V. In subsections IV.iv and IV.v we detail an alternative perspective—superstar firm models and look at the empirical evidence that could distinguish this model from the weakening competition story (IV.vi). Finally we examine some alternative explanations in subsection IV.vii.

IV.i. Evidence—Markups and the Labor Share

Concentration is a crude measure of market power partly because of the difficulty of defining the relevant market. A more direct approach is to try to measure producer price-cost markups. The
well-documented decline in the labor share of GDP—see Chart 8—is broadly consistent with a rise in markups. To see this, note that in a wide class of imperfect competition models (e.g., Hall 1988)\textsuperscript{12}, one can write the markup $\mu_i$ of firm $i$’s ratio of its price ($p_i$) to its marginal cost ($c_i$) as:

$$\mu_i = \frac{\alpha_i v}{s_i v}$$

(1)

where $\alpha_i v$ is the output elasticity with respect to a variable factor $v$ and $s_i v$ is the factor share, the cost of factor $v$ in total revenue (i.e., $s_i v = \frac{w_i x_i v}{p_i q_i}$ where $w$ is the factor price, $x$ the factor quantity and $q$ the firm’s output). This implies that the markup can be estimated from just a production function parameter and a variable factor share. For example, in a representative firm model with a time invariant Cobb-Douglas production function we can write the markup as a function of the labor share at time $t$ as:

$$\mu_t = \frac{\alpha L}{s_t L}$$

(2)

Hence, the fall in the labor share is the inverse of the rise in the markup. Using Chart 8 we see the labor share fell from about 64 percent in 1982 to 58 percent in 2016. If $\alpha L = \frac{2}{3}$, for example, this implies that the markup rose from 4 percent in 1982 to 15 percent by 2016 (1.04 to 1.15).

There are many issues with accurately measuring the labor share of GDP such as the treatment of the property sector; the income of the self-employed and business owner-managers; differences between net and gross GDP. Although such corrections affect the magnitude of the fall of the labor share, the fact that there has been a fall in the United States and other OECD countries seems robust to different ways of dealing with these issues (e.g., Karabarounis and Neiman 2014 and Autor et al. 2018 focus on the corporate sector).

An alternative to relying on equation (2) is to take a more direct approach of measuring capital. Barkai (2017) uses macroeconomic data and calculates the share of profits ($\Pi$) in GDP ($PY$) since the early 1980s using the accounting formula:

$$\left( \frac{\Pi}{PY} \right)_t = 1 - \left( \frac{w^L L}{PY} \right)_t - \left( \frac{rK}{PY} \right)_t$$

(3)
He finds that the profit share has risen because both the labor share and the capital share have declined, which is consistent with rising market power. The capital share is more challenging to calculate than the labor share as one needs to robustly calculate the opportunity cost of capital ($r$). Karabarbounis and Neiman (2018) revisit Barkai’s calculations over a longer period (from 1960) and broadly replicate the results, but label the residual profit share as “factorless income” highlighting the uncertainty over the measurement of the risk-adjusted user cost of capital and the capital stock. They also emphasize that profit shares appear high in the 1960s and 1970s, before falling in the early 1980s, a pattern driven mainly by sharp swings in the interest rate.

Given the difficulty with macroeconomic estimates, De Loecker and Eeckhout (2017) use U.S. Compustat firms to estimate a version of equation (1). They econometrically estimate two-digit industry specific production functions to obtain the output elasticity with respect to the sum of variable costs and then divide these estimates by the factor revenue shares. They reach the dramatic conclusion that aggregate markups have risen from 1.1 to 1.6 between 1980 and 2015. One challenge is separating variable costs from fixed costs in
the company accounts. They use “cost of goods sold” (COGS) as their measure of variable costs. This is potentially problematic because this may include some fixed cost components and it may also miss out other variable costs. Traina (2018) shows that using all operating expenses results in markups that do not grow very dramatically as the COGS share of revenue has declined a lot. However, an even more important concern with their paper (as noted in the previous section) is that generalizing from publicly listed can be dangerous because of the serious sample selection problems both in a point of time and over time, making it very hard to describe robust macroeconomic trends.

Others have estimated markups using administrative data which are much more representative of the economy as a whole. Hall (2018) uses industry level data and also finds an increase in aggregate markups over time. Autor et al. (2018) use Census microdata to estimate markups using both accounting approaches (as in Antras et al. 2017) and also econometric estimation (as in De Loecker and Warzynski 2012). Like De Loecker and Eeckhout (2017) and Hall (2018) they also find increases in the aggregate markup. However, they find little change in the unweighted average or median firm markup (either from direct production function estimation or accounting exercises). This mirrors their finding on the changes in the labor share—the fall in the labor share is due to reallocation towards large, high margin firms rather than a general increase in the markup across all firms.

Looking more generally across 26 OECD countries between 2001 and 2014, Calligaris et al. (2018) use equation (2) to estimate markups and also find that although the size-weighted average (i.e., aggregate) markup has risen, most firms have not seen increases in their individual markups. This is also the conclusion of Baqaee and Fahri (2017) and is again consistent with the reallocation effects stressed by Autor et al. (2018). We discuss the interpretation of this more below.

**IV.ii. Implications of Weakening Competition**

A generalized increase in market power has many worrying implications (see De Loecker and Eeckhout 2017). Because of lower
allocative efficiency it directly leads to higher prices, an inflationary force that cannot be effectively counteracted by monetary policy (see Section V). Secondly, there is loss of efficiency as market power tends to inhibit productivity which will be a further upward pressure on costs and prices. Third, these forces will all feed into lower real wages. Fourth, lower average wages makes working less attractive and so could lead to lower participation rates, something observed in the United States, especially for men.

**IV.iii. Possible Causes of Rising Market Power**

What might have caused an increase in market power? Proponents have a difficult time pinning down plausible explanations, but antitrust and regulation are the two most cited culprits.

*Antitrust Policy*

One view is that U.S. antitrust enforcement has weakened over time. Grullon et al. (2017) claim that there was a significant decline in antitrust enforcement during the administrations of George W. Bush and Barack Obama (e.g., Harty, Shelanski and Solomon 2012, Crane 2012). Use of Section 2 of the Sherman Act, which allows antitrust agencies to prevent increase in market power of existing dominant firms, has declined from an average of 15.7 cases per year over 1970-1999 to fewer than three over 2000-14. Global activity in mergers and acquisitions surpassed $5 trillion in 2015, about $2.5 trillion of which was in the United States, the highest amount in a year on record. Between 1998 and 2008 the Federal Trade Commission essentially stopped enforcing mergers when there were at least five firms remaining in the market (Kwoka 2017). Wollman (2018) points to the amendment to the Hart-Scott-Rodino Act in late 2000, which allowed many more exemptions for merger notification and shows evidence that it lead to more anti-competitive mergers.

Other measures of antitrust enforcement, however, exhibit opposite trends. For instance, Council of Economic Advisers (2016) provide evidence of increased cartel enforcement in the form of fines/penalties and prison sentences (e.g., from $0.36 billion in 2004 to $1.3 billion in 2014). In the 1990s under Bill Clinton there was a
spike in antitrust enforcement, perhaps most famously in the moves against Microsoft in the browser wars. And the competition authorities in the EU have actually been strengthened in the last 20 years leading to many high profile cases, such as the recent $5.1 billion fine on Google.\textsuperscript{18}

A related issue is common ownership of many stocks by financial companies such as Blackrock. These institutional owners have become larger and more concentrated themselves. Azar et al. (2015, 2016) have argued that increasing common ownership of corporations by banks and other financial institutions has softened price competition between firms. Yet, there is little direct evidence that these large shareholders are at all active in pushing for such coordination.

\textit{Regulation}

Regulation could reduce competition in at least three ways: (i) regulations often have a big fixed cost component which benefits larger firms; (ii) regulation may introduce barriers to entry; and (iii) increased rent seeking may allow larger firms to affect regulation through lobbying, thereby strengthening their position as leaders.

Some have argued that the United States has become substantially more regulated in recent decades. For example, financial service regulation has increased enormously since the credit crunch (e.g., Dodd-Frank), as has health-care regulation (e.g., American Care Act, ACA) and environmental regulation. Gutierrez and Philippon (2017) show a secular increase in the Mercatus Regulation index since 1970.\textsuperscript{19} Kleiner and Krueger (2013) show a large increase in occupational licensing—this creates a barrier to entry making it harder to start a new business in some services. For example, the share of workers in occupations requiring some sort of state license grew fivefold over the last half of the 20th century.

Yet not all indicators show moves to greater regulation. The more standard OECD index of product market regulations shows broad stability for the United States from the late 1990s to today. And this index also shows substantial deregulation in the EU over this time period. Furthermore, as with antitrust, one could also point to many
ways in which there have been significant deregulatory activities over time, especially under Ronald Reagan in the 1980s when concentration was also rising. In many ways the regulation under Dodd-Frank and the ACA were in response to crises arising, in part, from the heavy concentration of the banking sector (“too big to fail”) and health-care markets. So the new regulations may be more effects than causes of greater concentration.

**IV.iv. An Alternative Perspective on Market Power: Superstar Firms**

The rising market power argument has many attractions as it seems to explain disparate and surprising macro trends. But on one level it is still quite puzzling. As discussed above, it is not clear that institutional factors such as antitrust and regulation have moved in a strong anti-competitive direction. And many other major changes would lead observers to believe that markets have become more competitive over time. Trade costs have declined over the last 40 years as more countries have signed up to the rules-based trading system under the World Trade Organization (WTO). In particular, China’s re-integration into the global trading system and in particular its accession to the WTO in December 2001 has been a major pro-competitive shock to the OECD markets (e.g., Draca et al. 2016, Autor et al. 2013). Nontariff trade barriers have also fallen with “deep” regional integration such as the development of the EU’s Single Market.

Technological change has helped reduce the frictions to product markets within and between countries. Shipping is less expensive and with digital goods essentially costless and communication costs have fallen dramatically. The internet has enabled firms to enter each other’s markets more aggressively (think of Amazon) and for consumers to more easily compare prices (and characteristics) of goods and services online.

It may seem more natural to think that these new technologies and globalization trends would make markets more competitive, rather than less competitive.
An important insight from classic debates in Industrial Organization between Bain (1951) and Demsetz (1973) over the Structure-Conduct-Performance paradigm (see Schmalensee 1987) has been lost in many recent discussions. Increases in market toughness can go hand in hand with many of the trends we have documented. For example, in Autor et al. (2018) we show this in the context of a simple model with heterogeneous firms and imperfect competition, but the intuition is straightforward. As discussed in Section II if firms differ in their productivity and markets are not perfectly competitive more productive firms will have bigger market shares. Furthermore, these large “superstar” firms will tend to have higher profit margins and lower labor shares of value added. If market competition rises (e.g., consumers become more price sensitive) then more output is allocated to the larger, most productive firms—i.e., concentration rises. This can be through the extensive margin (less productive badly managed firms exit) and the intensive margin (among the survivors, high productivity firms get even larger market shares). Hence an increase in competition could easily lead to rising concentration.

What about the aggregate labor share and price cost margin? For any individual firm a rise in competition will mean its price-cost margin tends to fall (and labor share of value added rises) when market toughness increases. But offsetting this “within firm” effect is the “between firm reallocation” effect that moves more market share toward the high margin, larger, more productive firms when competition rises. If the underlying skewness of productivity is great enough, this reallocation effect dominates and aggregate markups rise and the labor share falls when market toughness increases. In this case we will see an increase in concentration and an increase in the aggregate/industry markup arising from an increase in competition.

The lesson from this simple analysis is that we cannot conclude from aggregate industry wide changes in concentration and markups, etc. that competition has fallen. One needs to look into more detail at the microdata. As noted above, Autor et al. (2018) find that the labor share and markup has hardly changed in 30 years for the typical U.S. firm. What explains the fall in the aggregate labor share and rise in the markup is a reallocation toward the larger and more productive firms.
A more specific version of this superstar firm hypothesis is to consider “winner takes most” markets. Higher competition in general will give firms with a cost or quality advantage a large share of the market. But the growth of platform competition in digital markets has led to dominance by a small number of firms such as internet search (Google), ride sharing (Uber), social media (Facebook, Twitter), operating systems for cellphones (Apple, Android), home sharing (AirBnB), etc. Network effects mean that small quality differences can tip a market to one or two players who earn very high profits.

The growth of such industries does not mean that competition has disappeared, rather its nature has changed. There is more competition “for the market” rather than “in the market.”

**IV.v. The Role of Information and Communications Technology (ICT)**

We have discussed the superstar firm explanation of the trend toward higher concentration in the context of (i) high tech digital markets engaging in platform competition and (ii) a general increase in market toughness due to globalization and/or the Internet. But a third reason why larger firms may expand more relates to the ICT revolution. Quality adjusted ICT prices have been falling dramatically for some time, meaning that there were huge opportunities for firms who could exploit this opportunity effectively. For example, retailers like Walmart were able to develop deeply integrated supply chain networks based on proprietary logistical software. More efficient logistics, higher turnover of inventory, and greater product variety at lower cost gave these big box retailers major advantages over smaller chains and independent mom-and-pop retailers (e.g., Foster, Haltiwanger and Kirzon 2006). This proprietary own-account software is also a feature of many other industries, such as banking and online shopping (Amazon), giving large players competitive advantage.

Bessen (2017) finds that U.S. industries since 2002 with greater software developer employment (a proxy for proprietary software) had faster growth in concentration ratios, profit margins and plant sizes). Bauer and Lashkari (2018) analyze French firm-level data with rich measures of IT software and hardware. They find that larger
firms use IT much more intensively than smaller firms and show that there has been a large reallocation of output (and rising IT intensity) in these firms. This helps explain the fall in labor shares in France.

This is suggestive evidence that part of the growth of superstar firms is related to the fall in price of ICT (as suggested by Karabarbounis and Neiman 2013). But the mechanism is not through a general substitution toward capital, but rather a reallocation of output toward highly profitable and more IT intensive companies.

**IV.vi. Distinguishing Between Declining Competition vs. Superstar Firms Explanations**

Is the increase in aggregate markups and concentration due to a general fall in competition or rather a change in the economic environment reallocating more activity towards superstar firms?

If the increase in the aggregate markup reflects weakened antitrust enforcement this will generally lead to worse allocative efficiency, higher prices and lower productivity as discussed above. On the other hand, if it is due to tougher markets reallocating more output to the more efficient firms this should lead to higher productivity.

We can shed light on these alternative explanations by examining the changing characteristics of the industries that are becoming more concentrated. Autor et al. (2018) find that the U.S. four-digit industries with the greatest increases in concentration have also had higher TFP growth and a faster growth in innovation (as measured by such indicators as cite-weighted patents). Ganapati (2018) and Besen (2017) also find a positive relationship between concentration growth and productivity changes. Looking across all the Economic Census data 1977-2012 Ganapati (2018) also finds little correlation between increases in prices and changes in concentration, as would be expected if concentration was being driven by weaker competition (Peltzman 2018 reaches the same conclusion).

These pieces of evidence do not seem consistent with the simple story of falling competition lying behind increasing size differences, as the concentrating industries appear to be the more dynamic sectors.
One implication of this discussion is that to understand the aggregate productivity slowdown we have to look elsewhere for explanations than simply declining competition. Fortunately, there is no shortage of alternative hypotheses such as mismeasurement, the rapid buildup of intangible capital investments in new technologies such as AI and automation, declining R&D productivity, secular stagnation, continued financial frictions, demographic changes or even a return to the 1974-1994 trend. Unfortunately, there is no consensus (yet) on which ones—if any—matter the most (see Syverson 2018 for an overview).

Finally, it is useful to reflect on the finding that the main trends of increasing differences appear to be broadly common across the OECD. Institutional changes across countries—for example, antitrust, regulation, union power, the minimum wage—have evolved in very different ways in the United States and EU. Although these institutions clearly may play some role, the global similarity of the changes suggest that something more fundamental is at play.

**IV.vii. Other Explanations**

*Overhead Fixed Costs and Intangible Capital*

The evidence suggests that aggregate markups over marginal costs have increased. However, it may be that there has also been a rise in fixed overhead costs. If this is the case, higher margins are necessary for firms to stay in business over the longer run. Indeed, in the long run almost all fixed costs become variable. Perhaps the most obvious candidates for these costs are those related to “intangible capital.” Many types of capital are hard to measure (e.g., firm-specific human capital from training and goodwill capital from advertising). Corrado, Hulten and Sichel (2009) describe the technological and managerial competencies discussed in Section II as scientific capital and organizational capital respectively. Since these investments are generally treated as current expenses in company and national accounts, this mismeasurement could help explain some of the patterns we observe. In terms of equation (3), the argument would be that true $K$ has risen much faster than observed $K$, keeping the profit share constant. Certainly, the evidence in Corrado et al. (2018)
suggests that intangible capital is becoming an ever-larger share of total capital.

In one sense, intangible capital is simply a form of measurement error and it is always possible to rationalize any empirical pattern in this way. As is well known, the biases could just as easily lead to overestimate of TFP as an underestimate.\textsuperscript{27} Karabarbounis and Neiman (2018) consider how an intangible capital model could generate the measured rise in the aggregate markup discussed above. They argue that it is hard to rationalize the trends in this way (intangible capital would have had to be particularly high in 1960-1980 for example).

Do the larger and more productive firms simply have more intangible capital than their smaller counterparts? Of course, one would need to discuss why larger firms were accumulating more intangible capital in this way. The Corrado et al. (2009) framework is a neoclassical representative firm approach and so is silent on this. As noted above, Bauer and Lashkari (2018) find that large firms are more IT intensive and De Loecker and Eeckhout (2017) argue that they are also more selling, general and administrative expense (SG&A) intensive (one measure of overhead or intangible capital). However, the latter find that the increase of SG&A as a fraction of sales is too low to rationalize the increase in markups over marginal costs.

One difficult issue is that there may be increases in expenditure on the sunk costs of entry. It is unclear how these are accounted for. If entry costs are reflected in R&D then in principle these should be picked up, but it is likely that not all R&D is measured especially for new firms and, in any case, entry costs go beyond formal R&D expenditure.

Ultimately, the only way to adjudicate the role of intangible capital is to try to improve measurement and see how much of the increased variance in size, productivity and wages can be explained with these factors. There has been progress here. R&D and software are now capitalized in the national accounts of the United States. As noted in Section I, we have been able to better account for more of the cross-firm and country-sectional differences in measured TFP with better indicators of technology and management.
Outsourcing/Offshoring

Firms are increasingly outsourcing and offshoring activities into global value chains. This may be related to some of the trends we observe. For example, Song et al. (2016) speculate that increased between-firm earnings dispersion might be due to high wage/skill firms increasingly outsourcing their low wage jobs such as cleaners to specialized service firms. As with intangible capital, however, to account for increasing differences the story would have to be not only that these sourcing activities have become easier, but also that larger firms have a particular advantage in them. While this is likely for offshoring, it is not so clear for domestic outsourcing. Smaller firms are increasingly able to outsource online many activities such as payroll, accounting, hiring and marketing that would formerly have to be performed in-house.

In summary, although offshoring may account for some of the rise in markups, the fact that we also observe rises in markups for purely domestic firms and industries suggests that this mechanism cannot fully explain the patterns we observe.

Slowing Technological Spillovers

Andrews et al. (2015) argue that the increasing TFP gap between leading and laggard firms is due to a slower rate of diffusion. We would expect that over time, small/low productivity firms catch up with large/high productivity leaders through imitation and adoption of the know-how of the leading firms. If this has become increasingly hard, inequality will increase between firms. So rather than a change in market environment this is more like a change in the ability to imitate. The problem with this story is that one would have thought that better ICT should lead to faster rather than slower technological diffusion. It is possible that there are other barriers (if erected by leading firms then we are closer to the diminishing competition hypothesis) such as stronger IP, lower interfirm mobility of skilled labor, etc. But there is no clear evidence so far that I know of that would corroborate this.
V. Some Policy Implications

I have focused more on positive than normative issues in this paper. But I speculate a little on policy issues in this section.29

VI. Antitrust Policy

If the increase in concentration and markups reflects weakening antitrust enforcement then an obvious policy solution is to strengthen it again. Similarly if overly onerous regulations or lobbying are the root causes then the policy implications are obvious.

Even if it is the case that the world is closer to the superstar firm model, however, this does not mean that antitrust should be weakened. If superstar firms attain their dominant positions on the merits, it does not mean that they will always use their market power for the good of consumers. They have incentives to entrench their position through lobbying, erecting entry barriers and buying up future rivals. In the technology sector, the main exit strategy of venture capital-backed start-ups is now to be eventually acquired by a tech titan rather than aim for an IPO.30 The current merger guidelines focus on horizontal competition, but the risk may be that future competition is weakened by such acquisitions. The acquisitions by Facebook of Instagram in 2012 and WhatsApp in 2014, for example, could be seen as problematic because these social media platforms may have become a major future rival to Facebook’s platform in the absence of the merger. This implies that antitrust policy needs to be re-thought in the era of superstar firms. There are many actions of dominant incumbents to maintain their market power not just through mergers and acquisitions but also through other strategies which can be used to strengthen incumbent advantage (such as the control of consumer data).

This leaves an important role for the modernization of competition rules to reflect the changing nature of product market competition.31 But what are the implications for monetary policy makers? We consider these next.
V.ii. Monetary Policy in the Long Run and the Short Run

If the increase in concentration and markups reflects institutional changes such as weakening antitrust enforcement or poor regulations then, as discussed above, this will lead to inefficiently higher prices and a lower equilibrium real wage. This is a structural supply side problem that monetary policy authorities can do little to directly offset in the long run. In the short run, things depend on how nominal wage setting changes. With some nominal wage rigidity, we would expect a central bank to have to tighten monetary policy in the face of a negative supply shock in order to induce a negative output gap and so bring inflation back to target. The degree to which it does so will depend on the slope of the Phillips curve and the central bank’s relative weight on output stabilization vs. departures from its inflation target. Eventually, the new constant inflation equilibrium will be characterized by lower real wages and lower employment. On the other hand, if nominal wages respond flexibly to the negative supply shock of increased market power, the real wage consistent with stable inflation will be reached more quickly, and the central bank may not need to tighten policy.

On the other hand, if the trends of increased concentration and markup levels reflect technological and globalization changes favoring superstar firms, the implications are not so malign. Rather than lower real wages, we may expect to eventually observe higher productivity, lower prices and higher real wages.

Since the global financial crisis, inflation has been quite dormant despite the continuation of the trend toward rising markups and concentration rates. At a crude level, this would suggest that the negative supply shock story is missing something. One reason could be that superstar firm models better explain the trends. Of course, there may be many alternatives such as offsetting shocks to worker bargaining power and secular weaknesses in investment.

The basic point, however, is that it matters for monetary policy whether one perceives the product market trends as a negative supply shock or something potentially more benign.
Viii. Monetary Policy Effectiveness

Finally, consider a central bank seeking to stimulate investment in a downturn by reducing interest rates. To the extent that the mechanism works through (at least temporarily) lowering the cost of capital, the size of the markup could matter. In standard macro models, markups are constant across firms and time so the level of the markup will not matter, because a 1 percent decrease in marginal cost is passed through to a 1 percent decrease in prices. But if markups are variable (as is suggested by the trade evidence on incomplete markups and the firm heterogeneity literature) then a firm will generally take some of a cost decrease in the form of higher profits rather than just increasing inputs. Thus, a firm with a high markup will increase investment less than a more competitive firm when interest rates fall. This will tend to make any given decrease in interest rates less effective in a high markup economy. Higher markups and incomplete pass-through also slow reallocation and so further reduce productivity growth (Decker et al. 2018).

VI. Conclusions

Analyzing the macroeconomy often requires getting “under the hood” and understanding the vast diversity of firm experience. We have shown that there are enormous differences in the productivity of firms within narrow sectors and that this is linked to their technological and managerial capabilities. This fundamental heterogeneity helps explain differences in the wealth of nations across countries and also aggregate productivity growth over time.

In recent years, firm heterogeneity appears to have increased rather than narrowed. These increasing differences are most obvious in terms of size: sales concentration has mushroomed across most U.S. industries, but it is also discernible in terms of wages and productivity.

Increased concentration brings with it the concern of market power and indeed, some have argued that many of the economic ills we face today in terms of sluggish productivity and real wage growth are due to rising monopoly power. My view is that this conclusion is premature. Rising aggregate markups and concentration may also reflect changes in the nature of competition where superstar firms are
rewarded with greater market share in “winner take most” markets. I have offered some evidence more in line with the nuanced superstar firm model than a general fall in competition due to antitrust and regulation. But this is for sure not the final paper in this area, however, and there are substantial uncertainties.

A final word of warning. Even if it was the case that the world is closer to the superstar firm model, this does not mean that antitrust policy should be relaxed. Even if superstar firms attain their currently dominant positions on their merits of out-competing rivals, it does not mean that they will always use their power for the good of consumers. They may well try to entrench their position through lobbying, erecting entry barriers and buying up future rivals. As larger parts of the modern economy become winner take most/all, it is important that competition authorities develop better tools for understanding harm to innovation and future competition, rather than the traditional emphasis on the pricing decisions of current rivals.

Author’s Note: This paper was prepared for the 2018 Jackson Hole Symposium. It has benefitted from comments by my discussant, Valerie Ramey. The paper draws extensively on work with many co-authors, especially David Autor, Nick Bloom, David Dorn, Larry Katz, Christina Patterson and Raffaella Sadun. Particular thanks to Wendy Carlin, Chiara Criscuolo and Iwan Werning for discussions on earlier drafts.
Endnotes

1See the survey in Van Reenen (2011a) for example.

2Development accounting (e.g., Caselli 2005) focuses on how to account for these large cross-sectional differences across countries. It is the cross-sectional analog of the Solow growth accounting approach. Gennaioli et al. (2013) perform development accounting using cross-sectional data from the regions within a large number of countries. They argue that an expanded view of human capital (which includes managerial/entrepreneurial skills) can account for most of the TFP differences.

3For example Hernández-Pérez et al. (2006), Axtell (2011) and Garicano et al. (2016).

4For simplicity I will use “firms” and “establishments” more or less interchangeably for convenience. This abstracts from a growing literature looking at reallocation across plants within the same firms (e.g., Bloom et al. 2017).

5The OECD team of Berlingieri et al. (2017) describe similar large cross-firm differences using 15 OECD countries in their MULTIPROD database, using moments collected from censuses run by the National Statistical Agencies of many countries.

6Analysis of entrants has found that their measured productivity is surprisingly low, usually no better than incumbents. However, this appears to be due to an overestimation of their output price, because firm specific prices are usually unobserved and researchers use industrywide price deflator instead. Foster et al. (2008) show in industries where plant-specific prices are observed entrants typically price below the average incumbent, so revenues deflated by industry prices will lead to an underestimate of entrant output and therefore also their productivity.

7See, for example, Olley and Pakes (1996) and Melitz and Polanec (2015).

8Foster, Haltiwanger and Syverson (2008) study 11 seven-digit homogeneous goods (including block ice, white pan bread, cardboard boxes and carbon black) where they have access to plant specific output prices. They find that conventionally measured revenue based TFP (“TFPR”) numbers actually understate the degree of “true” quantity-based productivity dispersion (“TFPQ”) especially for newer firms as the more productive firms typically have lower prices and are relatively larger. Bartelsman et al. (2009) show that measured TFPR will generally be correlated with true TFPQ but also with the firm specific price shocks. In the Hsieh and Klenow (2009) model heterogeneous TFPQ produces no difference in TFPR because the more productive firms grow larger and have lower prices, thus equalizing TFPR. But this is a knife-edge case.

9Bartelsman and Dhrymes (1998, Table A.7) show that over a five-year period around one-third of plants stay in their productivity quintile. White et al. (2018) emphasize that the imputations the Census makes to deal with nonreporting of some items in the Economic Census causes TFP dispersion to be underestimated rather than overestimated.
Gutierrez and Philippon (2017, 2018) have argued the opposite—that concentration has been falling in the EU since 2000. The reason for the difference is that Gutierrez and Philippon use BVD Orbis data to calculate concentration rather than the near-population data used by the OECD. Orbis does a reasonable job at tracking sales in the largest firms, but (especially in the late 1990s and early 2000s) has very incomplete coverage of small and medium-sized firms in many countries (Bajgar et al. 2018). Hence Orbis overestimates overall industry sales growth as it includes the increase in industry sales arising through expanding sample coverage. The OECD can reproduce Gutierrez and Philippon’s falling EU concentration when they use Orbis for both the numerator and denominator of concentration. But when they use the true industry size from population data in the denominator they reverse this result and find rising concentration as illustrated in Chart 4.


Formally, equation (1) requires only cost minimization with respect to a variable factor (no adjustment costs) and a production function that is continuous and twice differentiable in its factor inputs.

We discuss the measurement of intangible capital below. Karabarbounis and Neiman (2018) prefer the explanation of an increase in the user cost as the most plausible explanation. However, it is very unclear why the risk premium should have risen so much since 1980.

This is the De Loecker and Warzynski (2012) method, which is an application of Hall (1988) to microdata.

The other big operating expense apart from COGS is Sales and General and Administrative Expenses, SGA. Traina shows that SGA has increased substantially for firms over time.

Unfortunately, his instrumental variables (categories of government defense spending) have no variation across industries and would be collinear with time dummies. Not including time dummies raises the concern of omitted variables violating the exclusion restriction. The change in markup is imprecisely estimated.

Calligaris et al. (2018) use BVD data (firm accounts). De Loecker and Eeckhout (2018) also use firm accounts (Worldscope data) from developing countries as well as OECD countries and claim that markups have risen in every continent. None of these sources is based on administrative data, so the sample selection problems are an issue again, especially for developing countries.

Gutierrez and Philippon (2018) correctly point to various ways in which EU competition policy has been toughened and contrast this with the United States where they argue it has deteriorated. They then argue that there has been a fall...
in concentration in the EU compared to a rise in the United States. As discussed above, however, OECD analysis of near population administrative datasets (as opposed to their selected BVD subsamples) actually shows a rise in concentration in the EU as well as the United States. This casts doubt on a purely institutional explanation of the trends.

19 In 2002-12 they show a (weak) positive relationship between increases in this measure of regulation and concentration in very broad industries (NAICS 3).

20 The model is a generalization of Melitz and Ottaviano (2008) to allow for more general productivity distributions (rather than just Pareto) and any demand structure that satisfies “Marshall’s Second Law of Demand” that the absolute elasticity is falling in consumption (rather than just linear demand).

21 The exception to this is if markups are the same across all firms because the elasticity of product demand is constant as in the Dixit-Stiglitz CES preferences case. Although a workhorse macro model, this is actually a knife-edge case. For any utility function consistent with “Marshall’s Second Law of Demand” (demand is more inelastic at lower prices), more productive firms will have higher markups because they face a less elastic part of the demand curve. Autor et al. (2018) show that markups are higher (and labor shares lower) for establishments and firms within four-digit industries.

22 The same conclusion on the importance of reallocation is in Kehrig and Vincent (2017), Hartman-Glaser et al. (2016) and De Loecker and Eeckhout (2017).

23 Formally, they argue that the production function is nonhomothetic across different factor inputs. In particular, a proportionate increase in firm size has a larger proportional effect on the demand for IT capital than non-IT capital and labor. They test for this by estimating factor demand equations with size on the right-hand side instrumented by market size shocks (e.g., from foreign markets).

24 This is consistent with the empirical evidence discussed earlier that product market power tends to generate lower productivity. The most compelling evidence from this body of work looks at natural experiments which exogenously shift competition rather than simple correlations with concentration in product markets (which as noted could come from decreases or increases in competition).

25 A caveat to this conclusion is that the evidence of faster productivity growth in the concentrating sectors does not imply that the simple “tougher competition” is correct. We would expect to see stronger falls in prices in these sectors under this model. The fact that prices have not fallen more robustly could be interpreted to mean that leading firms are increasing productivity and market power together (although it may also mean that prices are being overestimated by failing to properly account for quality changes). This would push toward one of the other versions of the superstar firm model based on platform competition and/or ICT.
From 2005 through 2015, U.S. labor productivity growth averaged 1.3 percent per year, down from a trajectory of 2.8 percent average annual growth sustained over 1995-2004. It was 2.7 percent per year from 1947-1973, then fell to 1.5 percent per year over 1974-1994.

Assume intangible capital is treated as an expense, such as an intermediate input. During a period when current intangible investment growth exceeds intangible capital stock growth this will cause measured TFP growth to be too low compared with real TFP growth (this is one explanation of the current productivity slowdown). However, when the stock growth is faster than the current flows, the opposite is true—we overestimate actual TFP growth (see Brynjolfsson and Syverson 2017, for example).

Domestic outsourcing cannot of course explain the fall in the aggregate labor share as these outsourced workers will still show up in other firms (unless these workers were earning some wage rents before—see Godschmidt et al. 2017).

For a more formal treatment of the welfare effects of rising markups see Edmond, Midrigan and Xu (2018).

Cunningham et al. (2018) show theoretical and empirical evidence that many acquirers kill off promising innovations in the acquired as these threaten their monopoly rents from incumbent technologies.

For some recent thoughts on how to do this see Tirole (2017) and the recent Chicago Stigler Center conference https://research.chicagobooth.edu/stigler/events/single-events/antitrust-competition-conference-digital-platforms-concentration. Tirole emphasizes that the burden of proof in U.S. cases must be shifted, as the authorities face a near impossible standard in proving that the dominant firm’s actions will very likely lead to the foreclosure of future markets.

For simple expositions of this idea, see Carlin and Soskice (2015) or Carney (2017).

Markups are the same across all firms in Dixit-Stiglitz models of preferences which is common in macro models with monopolistic competition models. But there is now much evidence that markups are different across firms (as with the evidence from Autor et al. 2018, discussed above). In particular they are greater for larger firms as is consistent with trade models such as Melitz and Ottaviano (2008). The extensive pass-through literature also suggests that costs are incompletely passed on to consumers, which is consistent with variable markups (Arkolakis et al. 2018, De Loecker et al. 2016).

Gutierrez and Philippon (2017) argue that the fall in investment as a share of GDP in the United States. despite increases in Tobin’s Q is a sign of weaker product market competition.
References


Increasing Differences Between Firms: Market Power and the Macroeconomy


Peltzman, Sam. 2018. “Productivity and Prices in Manufacturing During an Era of Rising Concentration,” University of Chicago mimeo.


