Macroeconomics and Market Power: Context, Implications, and Open Questions

Chad Syverson

Prior research on market power had largely been the domain of microeconomists, who focused their analytical microscopes on individual industries or markets. Decades of microeconomic study have built a knowledge base, formed modeling conventions, and standardized empirical practices. However, a robust debate has erupted about whether the influence of monopoly power has grown beyond its traditionally studied microeconomic realm of the single industry or market and into the economy overall. Empirical investigations have found broad growth in measured profit rates, price-cost margins, and market concentration since at least as far back as 2000, if not earlier. Those upward shifts have been accompanied by drops in measured investment rates, firm entry rates, and labor’s share of income. If average levels of market power have indeed grown across the board, this is likely to degrade key metrics of economy-wide well-being, including investment, innovation, total output, and the distribution of income. A related debate, though one in which economists have played a lesser role, involves the consequences of potential broad-based concentration of not just economic but also political and cultural power (for example, Khan 2017).

This article assesses the macroeconomic market power research. I write as someone who has primarily studied market power in microeconomic frameworks.
but who has also done some macro-oriented research on topics such as aggregate productivity trends, albeit not dealing with market power per se from a macro perspective. For various reasons, the recent macro-oriented work has often departed somewhat from the established practices in micro analysis of market power. Part of this shift is surely tied to the obvious difference in the scope of analysis; things that can be done relatively straightforwardly for an individual market are not so easy to do at the economy-wide level. But there are other differences, too.

I will look at the issue of market power from a number of different perspectives. I begin with a theoretical comparison of defining market power in formal terms as a markup of price over marginal cost on the one hand and the often-used approach of using concentration to measure market power on the other. I then discuss how a prominent strand of macro market power research has used accounting data to estimate markups. I look at how markups are necessarily related to prices, costs, scale elasticities, and profits and point out seeming inconsistencies among the empirical estimates of these values in the literature. I then look at some of the research that has linked a rise in market power to lower levels of investment and a lower labor share of income. Throughout this discussion, I characterize the congruencies and incongruencies between macro evidence and micro views of market power and, when they do not perfectly overlap, explain the open questions that need to be answered to make the connection complete. I hope in this article to pull the two bodies of work somewhat closer together.

To preview my conclusion, I believe the macro market literature has established and collected an important and provocative set of facts, some developed by this literature and some built closely upon previous work. The literature has done a service by drawing plausible connections among these facts and showing how they might be tied to increases in the average level of market power. However, I believe the case for large and general increases in market power is not yet dispositive. There are empirical holes to be filled and plausible alternative stories (some with evidence of their own in their favor) that would first need to be rejected. To be clear, this is not to say that I believe the case for market power has failed or ought to be rejected. It remains a leading candidate explanation for several trends in the data. Rather, to my mind there remains considerable empirical uncertainty around the existence and magnitude of any across-the-board increase in market power in the economy. Thus, I finish the discussion by addressing what holes in the existing literature I would like to see filled.

---

1 I focus in this article on research that involves broad-scope empirical examinations of market power and its effects. There continues to be a lot of work examining market power in specific industries or markets. Some have argued for broader implications to be drawn from these market-specific studies. See, for example, Shapiro (2018 and in this issue) or the reaction of Scott Morton and Hovenkamp (2018) to studies like Azar, Schmalz, and Tecu (2018). There does not yet seem to be a consensus about whether market-specific studies themselves, alone or in their collective weight, have macroeconomic implications, but they can exposit the potential mechanisms at work.
Market Power, Markups, and Concentration

The literal textbook definition of market power is a firm having the ability to influence the price at which it sells its product(s) (for example, Pindyck and Rubinfeld 2012; Goolsbee, Levitt, and Syverson 2016). In other words, if a firm does not face a perfectly elastic residual demand curve, it has market power. A connotation of this definition, sometimes left implicit, is that the firm uses this ability to hold the price above marginal cost.\(^2\)

Using this definition, the magnitude of market power is tied to the size of the gap between price and marginal cost at the firm’s profit-maximizing level of output. The size of this gap—typically called the “markup” when expressed multiplicatively and the “margin” when expressed as a difference, though there is some variation in usage—depends on the shape of the firm’s residual demand curve. Steeper inverse demand raises the profit-maximizing margin and implies more market power.

Markups are difficult to measure directly. They require information not just on prices but on hard-to-observe marginal costs. As a result, there are many informal definitions of “market power” and associated metrics used in popular economic writing, and sometimes in economic research as well. Examples include the number of competitors (actual or potential), profit rates, and costs of market entry. Each of these alternatives has its merits, but each is also one step removed from actual pricing power, which in turn can lend itself to shortcomings and ambiguities in practice.

In the recent macroeconomic market power literature, the most frequently used measure of market power is concentration. Measures of market concentration summarize the share of market or industry activity accounted for by large firms. The two most common are the Herfindahl–Hirschman index, which is the sum of firms’ squared market shares, and \(C_n\), which is the combined market share of the largest \(n\) firms.

An advantage of concentration as an empirical tool for studying market power is that it requires data only on revenues and thus is often relatively easy to compute. The corresponding disadvantage is that concentration is about relative revenue and thus includes no information about costs or profits. For example, a monopolistically competitive market can be very unconcentrated and display near-zero levels of economic profit—indeed, monopolistic competition is defined by the atomistic nature of firms combined with possibilities for entry and exit—but firms in such a market can still have very inelastic residual demand curves and hence a lot of market power.

Concentration also necessarily requires a market definition, which is often a point of contention. The macro literature has in some cases measured the extent of concentration within broad industry groupings, which raises the possibility that

\(^2\) I focus here on market power in output markets. Of course, firms can also exercise market power in the markets for their inputs, including labor; for discussions, see Manning (2003), Azar, Marinescu, and Steinbaum (2017), Benmelech, Bergman, and Kim (2018), and Krueger (2018).
increases in concentration in narrower and more relevant markets may be invisible in the broader measures. Moreover, national concentration measures can be particularly misleading for geographically localized markets. For example, a chain restaurant building stores in a number of local markets would tend to increase measures of concentration computed at the national level, even if it reduced concentration in the economically relevant local markets. Rossi-Hansberg, Sarte, and Trachter (2018) find evidence suggesting that a “national concentration, local de-concentration” pattern is occurring in a number of industries.

Perhaps the deepest conceptual problem with concentration as a measure of market power is that it is an outcome, not an immutable core determinant of how competitive an industry or market is. The nature and intensity of industry competition combine with other supply and demand primitives to determine equilibrium concentration. However, the conditions of competition drive concentration, not vice versa.

As a result, concentration is worse than just a noisy barometer of market power. Instead, we cannot even generally know which way the barometer is oriented. Even if researchers agree on a definition of the market, concentration can be associated with either less or more competition.

Research that uses concentration as a measure of market power is implicitly relying on the mechanics of the standard Cournot oligopoly model, in which a number of firms with possibly different marginal costs choose what quantity to make of a homogeneous product whose price is determined by the intersection of the product demand curve and the joint production of the firms. This model implies a positive relationship between market concentration and average market power. More concentration implies less competition. In effect, with fewer firms, each firm has less competition to take into account and more ability to raise price above marginal cost. Furthermore, in this model, welfare is lower in more concentrated markets because of the deadweight loss associated with price-cost margins.3

On the other hand, a large class of commonly used industry models predict a positive relationship between competition and concentration. All involve heterogeneous-cost firms selling differentiated goods. The models build in this differentiation in various ways, ranging from a direct preference parameter to trade, transport, or search costs. More substitutability implies firms’ residual demand curves are more elastic. Reflecting this heightened competition, price-cost margins are lower when substitutability is high. Examples of models in this class include Melitz (2003), Asplund and Nocke (2006), Melitz and Ottaviano (2008), and Foster, Haltiwanger, and Syverson (2008).4

3 Specifically, under this framework, one can show that the share-weighted sum of firms’ Lerner indexes—their price-minus-marginal-cost margin as a share of the price—equals the Herfindahl–Hirschman index divided by the price elasticity of demand. In this issue, Berry, Gaynor, and Scott Morton offer additional discussion of interpreting the connections between competition and the Herfindahl–Hirschman index in the Cournot model.

4 Versions of these types of models where the differentiation is instead in product quality are often isomorphic to the heterogeneous-cost version.
In such models, actual market entrants come from a pool of potential entrants who decide whether to pay a sunk entry cost to draw a cost level from a known distribution. Entrants who choose to receive a draw determine after observing the draw whether to begin production and thus to earn the corresponding operating profits. This setup creates a threshold cost level where only potential entrants receiving sufficiently low-cost draws enter in equilibrium. This basic structure implies a comparative static result where increases in substitutability (consumers are more willing or able to shift to different producers) both reduce margins and make it harder for higher-cost firms to operate. Additionally, because consumers are more responsive to any given cost difference, the responsiveness of market shares to cost differences is larger when substitutability rises. Thus, an increase in substitutability both reduces price-cost margins and increases concentration. In contrast to the Cournot case, the model predicts a negative correlation between market power and concentration.

Two other predictions of these models are relevant to this discussion. First, welfare rises along with substitutability; heightened competition reduces margins and the associated deadweight loss. Second, as substitutability/competition increases, profits of the firms operating in the market actually increase. In the theoretical model, more intense competition reduces the range of operating cost draws that are profitable, reducing successful entry rates. As a result, profits conditional on operating must rise to counterbalance the higher risk of failure. Interestingly, then, higher profits among firms in the market are a sign not of less competition, but more. Profits rise, despite the lower margins, because quantities sold increase markedly as substitutability/competition rises. Models in this class emphasize the earlier point that concentration is an outcome of underlying forces of supply and demand that can play out in various ways.

A negative relationship between market power and concentration is not just a theoretical curiosity. Many empirical studies in varied settings have found that greater substitutability/competition—resulting from, say, reductions in trade, transport, or search costs—shifts activity away from smaller, higher-cost producers and toward larger, lower-cost producers. As an example, in Syverson (2004a, b), I show that increases in the ease with which consumers can substitute among producers—spatial differentiation is limited, or products are more physically similar—force out the least efficient producers and increase skewness in the size distribution. In Goldmanis et al. (2010), we demonstrate that search cost reductions reallocate market share toward lower-cost and larger sellers, increasing market concentration even as margins fall. It is not an exaggeration to say that there are scores, perhaps hundreds, of such studies. Some focus on specific industries; others are broader.\footnote{Changes in production technologies that increase scale economies can also raise concentration. Unlike increases in product substitutability, which by their nature tend to flatten residual demand curves and therefore reduce market power, scale economies have no direct influence on demand. Thus, their equilibrium effect on market power is more ambiguous. However, prices could very well still fall even if markups do not, because the scale economies have reduced marginal costs. Arguably, this mechanism in part accounts for the transformation of the US retail sector over the past several decades, first through...} Perhaps most...
relevant to the current discussion are Autor et al. (2017) and Crouzet and Eberly (2019). These studies find patterns of simultaneous concentration and productivity growth in settings that speak very directly to the recent macro market power literature as well. I will discuss each in more detail below.

In thinking about concentration as a measure of market power, there is a sharp split between the macro and micro market power literatures. From the 1950s through the 1970s, industrial organization often tried to link measures of market concentration to the behavior of firms and to resulting profit, in what was known as the structure-conduct-performance literature. But by the 1980s, given the very real concerns that concentration was likely to be misleading as a measure of market power, the field of industrial organization essentially stopped comparing market outcomes such as prices, margins, and profit rates to concentration levels—especially when making comparisons across markets or industries that differ in demand and technology fundamentals. While I would not call for a blanket ban on the practice of using concentration to measure market power, caution about the practice is well warranted. There were good reasons for industrial organization to choose to forgo it (particularly, again, for cross-industry comparisons). Simply put, the relationship between concentration and markups, prices, or profits is a relationship between market outcomes. These can be uninformative or, worse, misleading about the causal effect of competition.

Below I will speak further to what the microeconomic literature typically does to measure market power, whether it is practical for macro-oriented work, and what other alternatives might be available.

**Direct Measurement of Markups with Accounting Data**

To estimate markups directly, one needs data on prices and marginal cost. Data on prices is relatively straightforward to obtain, but data on costs across a wide range of firms—and especially data on marginal costs—is harder to find. One approach here is to use what researchers refer to as “accounting data”—essentially, data from firms’ financial statements—and then work with this data to develop estimates of marginal costs. In two recent papers, De Loecker and Eeckhout (2018) and De Loecker, Eeckhout, and Unger (2018) take this approach to estimating price–marginal cost markups in the United States and around the world. In the US-centric study, they use the Compustat database, comprising the harmonized financial reports of publicly listed companies for the past several decades (De Loecker, Eeckhout, and Unger 2018). The world study uses Thomson Reuters Worldscope, which spans over 100 countries and contains income statements mostly for publicly traded companies, though it does also include some private firms (De Loecker and Eeckhout 2018).
The simplest method with which one might use accounting data to measure markups is to look at the ratio of revenues to total variable costs, which—when both of these are divided by quantity produced—would be equal to the ratio of price to average variable cost. Average variable cost does not of course generally equal marginal cost, but marginal cost is very hard to measure directly. Only when marginal cost is constant at all quantity levels is it equal to average variable cost. Moreover, accounting cost categorizations do not make it easy to separate variable from fixed costs on a consistent basis.

Thus, the studies by De Loecker and Eeckhout (2018) and De Loecker, Eeckhout, and Unger (2018) move beyond the simple proxy approach to obtain a more sophisticated estimate of markups. They use a firm-level variant of a method Hall (1988, 2018) developed and applied to industry-level data. Hall (1988) shows that under cost minimization, for any variable input (an input that is freely adjustable by firms within any given period, as opposed to an input that is quasi-fixed, as many forms of capital are often thought to be), the firm’s markup will equal the ratio of two values: the elasticity of output to that variable input, and the share of revenues the input is paid. That is,

\[ \mu = \frac{\beta_v}{s_v}, \]

where \( \mu \) is the (multiplicative) markup, \( \beta_v \) is the elasticity of output with respect to the variable input \( v \) (from the firm’s production function), and \( s_v \) is the share of revenues paid to the variable input supplier. Basu (in this issue) overviews this and related “production-function-based” approaches for measuring markups.

The accounting data include a measure called “cost of goods sold” (COGS). De Loecker, Eeckhout, and Unger (2018) use this as a measure of variable inputs. They estimate a production function by regressing revenues on this measure of COGS and on the book value of capital for all firms in an industry. This yields an estimate of the elasticity of output with respect to COGS. The other piece of information necessary to estimate the markup, the share of revenue paid to this category of COGS, is observed directly in the data. They take the quotient of these two elements to obtain an estimate of the markup for every firm-year in their data. (The elasticity \( \beta_v \) is restricted to be the same across all firms in an industry or industry-year depending on the specification. The revenue share \( s_v \) is firm-year specific.)

---

6 Hall (2018) uses industry data and finds mixed support for increasing markups. He estimates an average trend in measured markups between 1988 and 2015 that is positive but statistically insignificant (0.6 percent annual growth, with a standard error of 0.5 percent). Multiple measures of returns to capital rise. There is a low correlation between the levels and growth rates of three measures of market power he constructs, but there is a modest positive correlation between concentration and measured markups in his sample.

7 Production function estimation is itself the subject of a large methodological literature and raises additional measurement issues beyond the scope of our discussion here.
An attention-getting headline number from De Loecker, Eeckhout, and Unger (2018) is that the revenue-weighted average markup in the United States climbed from about 1.2 in 1980 to 1.6 in 2014. They also find increasing skewness in the across-firm distribution of markups over that period, with average markup growth coming from a spreading of the right tail and a shift in revenue shares toward higher-markup firms. Indeed, the median firm-level markup remained essentially constant throughout the time period.

In their study with international data, De Loecker and Eeckhout (2018) find a similarly sized increase in the size-weighted markup, from 1.1 in 1980 to 1.6 in 2016. Some systematic variations in this trend exist across continents, however. While Europe, North America, Asia, and Oceania saw rather steady increases over 1980–2015, average markups in South America had little discernible trend. Markups in Africa jumped up between 2000 and 2005 but were level before and after.

One of the most compelling elements of these studies is that they are using a direct measure of price-cost margins to gauge market power. In terms of vulnerabilities, accounting data are not constructed for the sake of measuring economic categories like variable costs. Accounting data include two primary categories of costs: (1) cost of goods sold and (2) selling, general, and administrative (SG&A) expenses. COGS includes direct costs associated with purchasing and transforming inputs into the product a company sells and as such is thought to be composed primarily of variable costs. The SG&A category includes most other costs and as such captures many fixed costs. That said, some SG&A expenses might plausibly scale with the size of operations, while some costs in COGS might arguably be fixed. Indeed, accounting standards actually allow classification of expenses by COGS and SG&A to vary by sector. In the end, the variable/fixed demarcation is not as clean as one would like it to be for measuring markups.

How one measures variable costs matters empirically. Traina (2018) shows that if the sum of COGS and SG&A is used as the variable input measure, both the estimated levels and, more to the point, the changes in US markups fall. Instead of rising from 1.2 to 1.6 over 1980–2015, Traina’s alternate markups rise from only around 1 to 1.15. Of course, this estimate of markup growth could itself be flawed because of the imperfect mapping between accounting and economic cost categories, as De Loecker, Eeckhout, and Unger (2018) point out. In the end, researchers using this approach are left to make choices among imperfect options.

A separate measurement (and conceptual) issue is that while this ratio of marginal product to revenue share equals the output markup under the assumptions of imperfect competition in the product market and a perfectly competitive market for the variable input, it also equals the monopsony markdown in the wage of the variable factor if instead the product market is perfectly competitive and producers have market power in the factor market. If firms have market power in both the product and factor markets, then the ratio reflects the combination of these two effects. Therefore, reading the ratio as reflecting solely product market monopoly (or only factor market monopsony, for that matter) could misattribute one for the other. Moreover, even recognizing that the measured ratio may reflect
both market power effects, separately quantifying each component requires additional variation and empirical metrics beyond simply constructing the ratio.

**Posing a Paradox: Markups and Their Relations**

A widespread change in markups across an economy will necessarily have implications for other macroeconomic variables, including price inflation, cost growth, and profits. In Syverson (2018), I raise a potential inconsistency among measures of inflation, markups, and cost growth. I can summarize the paradox using the relationship that price, \( P \), equals a markup rate, \( \mu \), times cost, \( C \):

\[
p = \mu \cdot C.
\]

According to firms’ profit-maximization theory, the relevant cost \( C \) ought to equal marginal cost, and the markup \( \mu \) should be a function of consumers’ price sensitivity. However, even if prices are not set to maximize profits, the relationship is still quite general and useful. For any consistently measured price and cost, one can define the markup \( \mu \) as whatever multiplicative factor makes the relationship hold (\( \mu \) could even be less than 1, if price is less than cost for some reason). In this sense, the relationship is essentially an identity.

The same relationship applies to growth rates. That is,

\[
\text{Growth in } P = \text{Growth in } \mu + \text{Growth in } C.
\]

Expressing the relationship in growth rates is an approximation, but it will be close to exact in the situations in which we are interested, where growth rates are relatively modest.

Now consider the empirical patterns observed in each of these growth rates over the past few decades. The left-hand side, the growth rate of prices, is inflation. Measured inflation has been low over the past few decades, especially relative to what many consider as its traditional driving forces. The first term on the right-hand side, the growth rate of markups, has been estimated to be quite high in some studies, which is the object of focus here. But if price growth is relatively low and markups are growing quickly, costs must be falling quickly. It is not clear in the data that this is the case.

Two factors affect the growth rate of costs: productivity and factor prices. Productivity growth has been in a slump since the mid-2000s. Productivity is inversely related to costs, so when productivity grows more slowly than usual, cost growth will tend to be higher than usual. As for factor price trends over the past couple of decades, wage growth has been slow, if anything (more so for the middle and lower end of the distribution than for the high end), and interest rates have fallen to historic lows. In isolation, those factor price patterns would tend to slow the growth rate of cost, but they are countervailed by slowing productivity growth.
We can investigate the net effect of these two influences by looking at “unit costs,” which conveniently combine both productivity and factor price effects on costs. Unit labor costs are the ratio of total compensation per hour worked to labor productivity—the nominal labor compensation required to build one unit of output. According to Bureau of Labor Statistics data, the growth in US aggregate unit labor cost has been somewhat slower than inflation (this is also reflected in the labor compensation’s falling share of income). This opens the possibility that low labor cost growth has “made room” for higher markups. However, the timings of the two trajectories do not line up well. Much of the decline of labor’s share has occurred since 2000, while the period with the fastest increases in markups was 1980–2000. Furthermore, nominal unit cost growth for other factors may have accelerated over the period. The average growth rate of the “unit nonlabor payments” series from the Bureau of Labor Statistics, which includes capital payments, taxes on production, and profits, has slowly risen during the past two decades. This could in part reflect profits from increased markups showing up in addition to actual costs. Unfortunately, the Bureau of Labor Statistics does not break profits out of unit nonlabor payments for the entire nonfarm business sector. They do for nonfinancial corporations, however. These indicate increasing unit capital costs for such firms, with unit nonlabor costs having grown faster than inflation for the past 20 years.

In short, measures of costs do not seem to behave in the way implied by the measured trends in inflation and markups. One potential resolution of the paradox comes from parsing types of cost. Productivity and unit cost measures probably most closely reflect average cost. If marginal costs were rising at a slower rate than average costs, it is possible that unit cost growth could be steady even as inflation remained unusually low. The former would reflect steady changes in average cost; the latter would reflect faster reductions in marginal cost.

This story has the right qualitative features to resolve the paradox. However, it is unclear that it can quantitatively account for the differential patterns in prices, markups, and costs. A decomposition of the price-cost markup, first made by Susanto Basu in an earlier discussion of De Loecker and Eeckhout’s work, is instructive about this.

Rewrite the markup expression by multiplying and dividing it by average costs:

\[ \mu \equiv \frac{P}{MC} = \frac{P}{AC} \frac{AC}{MC} \]

Multiplying and dividing \( P/AC \) by the output quantity makes it clear that the markup is equivalent to the ratio of revenues to total costs. The \( AC/MC \) ratio is, by definition, the scale elasticity of the function that relates a firm’s costs to its output (that is, the inverse of the elasticity of costs with respect to quantity). When marginal costs are

---

8 I am grateful to Susanto Basu for conversations regarding this decomposition.

9 This is straightforward to verify. The elasticity of costs with respect to quantity for any differentiable cost function \( C(Q) \) is \( C'(Q) (Q/C) = MC(1/AC) \). For homothetic production functions, the scale elasticity equals the returns to scale of the production function. Note that \( C(Q) \) is the function that relates the
less than average costs, average costs are falling in quantity and the scale elasticity is greater than one. Conversely, if $MC > AC$, there are diseconomies of scale, and the scale elasticity is less than one.

We therefore have, using $\nu$ to denote the scale elasticity,

$$\mu = \frac{R}{TC} \nu.$$ 

Define pure profit’s share of revenues as

$$s_\pi \equiv \frac{R - TC}{R}.$$ 

We can rewrite the markup as

$$\mu = \frac{1}{1 - s_\pi} \nu.$$ 

Thus, the markup must equal the inverse of one minus profit’s share of revenue times the scale elasticity. Note that the only assumption required to derive this expression is that the cost function $C(Q)$ is differentiable; the other manipulations were just algebra or identities.

This expression reveals an empirical discipline on measures of markups at the firm level. Namely, markup levels must also imply something about profit shares, scale elasticities, or both. If a firm sees a substantial increase in markups over time, there must also be an increase in pure profit’s share of its income or in its scale elasticity.

It is often difficult to obtain firm-level estimates of scale elasticities, as common technologies are typically imposed across firms by researchers in order to estimate an elasticity. Thus, investigating the markup profit-share scale-elasticity relationship firm by firm can be hard. Exploring its aggregate version can still be informative, though this should be accompanied by the caveat that, as they are ratios, the nonlinearity of the markup and the scale elasticities implies that weighted averages of firm-level values will not generally exactly add up to their aggregate analogs. I make this aggregate comparison here, noting this proviso.

As noted above, DeLoecker, Eckhout, and Unger (2018) report that US average markups grew from 1.21 to 1.61 between 1980 and 2016.\(^{10}\) Suppose that the production technology remained stable enough over the period so that the scale firm’s production cost to its output in practice; it need not necessarily be the cost function (that is, the one that assumes the firm has chosen the minimum-cost bundle of inputs required to make any given $Q$). Thus, the expression still applies even if firms aren’t cost-minimizing.

\(^{10}\) De Loecker, Eckhout, and Unger (2018) report several sets of markup estimates. I am using their benchmark “PF1” specification. The alternative estimates exhibit quantitative behaviors similar to those described here.
elasticity didn’t change. Then pure profit’s share of revenues in 2016 must be the following function of its 1980 share:

\[
\frac{1.61}{1.21} = \frac{1}{1 - s_{\pi, 2016}} \frac{1}{1 - s_{\pi, 1980}},
\]

\[s_{\pi, 2016} = 0.25 + 0.75 s_{\pi, 1980}.
\]

Even if profit’s revenue share was zero in 1980, the observed change in markups—in the absence of any increase in scale economies—would imply the profit share in 2016 would be 25 percent. Given that this is a share of revenue and that total aggregate revenues (that is, sales) are roughly double aggregate value added, this implies that profits were roughly half of all value added in 2016. This would be unrealistically large. By any measure, labor’s share of value added is greater than this. Even if we were to consider all capital income as pure economic profit (that is, capital’s competitive return was zero), the increase in measured markups does not make empirical sense in the absence of substantial changes in scale economies.

What if scale economies did increase over the period? Fixed costs may have grown, or the output product mix may have shifted in composition toward products with lower marginal costs (like software and pharmaceuticals). An empirical test is feasible here. Pure profit rates can be estimated, although doing so requires assumptions about how to measure capital’s competitive return. One can estimate production functions to obtain scale elasticities. The recent literature contains some estimates along these lines. Barkai (2017) constructs a measure of pure profit’s share of value added, finding that from 1908 to 2014 it grew from 3 to 16 percent, and its revenue share thus grew from about 1.5 to 8 percent. (I will discuss this study in detail below.) De Loecker, Eeckhout, and Unger (2018) estimate changes in the average scale elasticity for firms in their sample, finding that it rose from 1.03 to 1.08 during 1980–2016. Plugging these values into the relationship above and taking their ratio yields:

\[
\frac{\mu_{2016}}{\mu_{1980}} = \left( \frac{1 - s_{\pi, 1980}}{1 - s_{\pi, 2016}} \right) \frac{\nu_{2016}}{\nu_{1980}} = \frac{1.61}{1.21} \frac{1.08}{1.03} = \frac{1.33}{1.05} = 1.33 = 1.14.
\]

While the relationship is still some distance from implying consistency, it is closer to equality, suggesting that growth in scale economies is part of the story. In addition, there is the caveat that I am mixing aggregates and firm-level averages when the relationship should hold firm by firm.
The relationship between markup, profit share, and scale elasticity is a tool that can be applied more generally, whether among firms in the cross section or over time. While there are practical hurdles, estimates of the necessary components are generally feasible to obtain in the data. The relationship imposes a useful consistency check on empirical estimates in this area.

**Market Power and Low Investment Rates**

Corporate profit rates and Tobin’s \( q \) (the ratio of a firm’s market value to the book value of its assets) have both been relatively high since 2000. However, during the same period, the investment rate has been low relative to its historical connections to profits and Tobin’s \( q \). In a pair of papers, Gutiérrez and Philippon (2017a, b) marshal evidence suggesting market power may be behind the low investment rates.

Gutiérrez and Philippon (2017b) run a horse race between alternative hypotheses for low investment: a rise in financial frictions, changes in the nature of investment (like intangibles replacing measured capital investment or globalization shifting investment abroad), increased short-termism in management, and decreased competition. Each class of explanation has multiple specific measures. They find that, at least in terms of ability to explain statistically the unusually low observed investment rate, the rising importance of intangibles accounts for about one-third of the drop, while corporate ownership structure (what fraction of company stock is held by likely long-term investors) and increased industry concentration explain the rest. In their framework, measures of financial frictions have no explanatory power.

To address the question of causality more directly, Gutiérrez and Philippon (2017a) use natural experiments and instrumental variables techniques to link changes in competition to investment. The natural experiments involve two measures of increased competition from Chinese imports. The instrumental variable is a measure of “excess entry” in an industry in the 1990s. The logic of the instrument is that the go-go US startup environment of the latter part of that decade in particular led to a large amount of essentially random volatility in entry rates across markets. They show that the amount of 1990s entry relative to fundamentals (both current and in expectation) is correlated with industry concentration a decade later, but uncorrelated with observable shocks that occurred in the interim. Instrumenting for industry concentration using excess entry, they find that concentration is negatively correlated with investment rates.

These papers are persuasive in their case that measured investment is low relative to standard explanatory variables. Moreover, they demonstrate the potential for market power not only to create inefficiencies and reduce output today but also, through its investment effects, to reduce future growth rates. But I believe the case is not yet proven. A few critiques present themselves here.

If intangible capital has become more important over the past couple of decades, and the composition of investment has shifted toward it as a result, the
quantitative response of measured investment (which of course does not include intangibles) to traditional variables like corporate profits and Tobin’s \(q\) would decline. But this would be a measurement change, not necessarily an economic one. Gutiérrez and Philippon (2017b) do consider this possibility, and they seek to address it using proxies for intangible capital (including “tangible intangibles”—the capitalized R&D, software, and artistic originals series constructed by the Bureau of Economic Analysis). Their proxy for intangibles does explain some of the drop in the measured investment rate, but given the uncertainties involved in measuring intangible investment, a different proxy might explain still more.

However, a deeper issue is that intangible investment need not just be associated with (or caused by) concentration; in addition, it can causally affect industry concentration. Thus, intangibles aren’t just another factor in addition to concentration that might explain low measured investment. They might be affecting concentration directly.

Crouzet and Eberly (2019) point out that an intangibles-concentration connection can occur through two mechanisms, with very different economic implications. One is that increased concentration, in this case reflecting less competition, reduces the incentives of firms to invest, and this might be coincidentally (or perhaps even causally) correlated with growth in intangible intensity. The other mechanism reverses the potential causality between intangibles and concentration and has diametrically opposed implications for welfare. If a company invests in intangibles that allow it to deliver a higher-quality product at a lower price (by reconfiguring its organizational structure and internal processes, for example), market share will naturally shift toward that company, creating coincident growth of intangible intensity and industry concentration. However, this rise in concentration would be efficiency enhancing, as the total resources required to deliver a given amount of product quality (and consumer welfare) would have fallen.

What suggestive evidence might be brought to bear on these two possibilities? Intangible investment intensity is highest and grows fastest for the largest and fastest-growing firms in an industry, according to the estimates from Crouzet and Eberly (2019). In addition, they compare sector-level trends in labor productivity levels and Hall’s (2018) industry-level markup estimates. They find that within the manufacturing and consumer sectors (the latter combining wholesale and retail trade as well as agriculture), estimated markups were flat over 1990–2015, while labor productivity rose. They therefore attribute the coincident increases in concentration and intangible intensity observed in those sectors to efficiency-enhancing mechanisms. On the other hand, they find that both markups and labor productivity grew steadily in the health-care and high-tech sectors, indicating elements of both market power and efficiency gains at work.

These results suggest that the connections between lower measured investment and concentration do not reflect an across-the-board influence of rising market power but instead are an amalgamation of differing mechanisms with quite different economic interpretations. Sector-specific mechanisms would be consistent with, for example, the notions that globalization has increased competitive pressures in
manufacturing (Feenstra and Weinstein 2017) while merger waves have reduced competition in healthcare (Gaynor 2018). Some recognition of the heterogeneity underlying aggregate patterns seems clearly warranted, both for understanding the phenomenon and for drawing welfare implications. This insight also vividly evokes the aforementioned issues involved with assuming that concentration is a useful measure of market power.

**Market Power and the Labor Share of Income**

It seems fair to say there is a consensus in the profession that labor’s share has been trending down, while corporate profits have risen. While some have raised concerns about specifics of measurement, the trends have been documented in multiple ways. The macro market power literature has raised the possibility that these changes may be related to a rise in market power, markups, and pure profit.

In an example of research along these lines, Barkai (2017) decomposes aggregate factor income into three elements: labor’s share, capital’s share, and pure profit. Labor income is taken directly from national income accounts and is therefore measured in standard ways. To compute capital income, Barkai multiplies the observed aggregate capital stock by a user cost of capital. The user cost equals a real interest rate (constructed as average blue-chip bond yields in a period minus a measure of expected inflation) plus a measure of the depreciation rates. This user cost is supposed to reflect the competitive return earned by capital inputs. Any remaining income is considered pure profit. The results of this approach indicate that the drop in the share of income paid to labor was accompanied by a slight drop in capital’s share. Meanwhile, the pure profit residual increased substantially, from 3 percent of national income in 1985 to 16 percent by 2014.

The study ties this shift in factor income shares to market power, using regressions conducted at the six-digit North American Industry Classification System (NAICS) level. Industries that saw larger increases in concentration saw bigger drops in labor’s share of income. Barkai (2017) interprets this as evidence that declining competition has been responsible at least in part for the secular decline in labor’s share.

Eggertsson, Robbins, and Getz Wold (2018) also emphasize the pure profit approach by augmenting a standard neoclassical model with increasing market power/markups. They show that, suitably parameterized and in the presence of a decreasing natural rate of interest, the model can qualitatively and quantitatively explain the falling labor’s share of income as well as several other phenomena: the increase in the pure profit rate, growth in the financial wealth-to-output ratio, an increase in Tobin’s $q$ without associated investment, and a divergence between the marginal and the average return on capital. The mechanism, briefly stated, is that growing market power (what the paper terms the “emergence of a non-zero-rent economy”) leads directly to the increase in pure profit through higher markups. Financial wealth and Tobin’s $q$ both reflect future claims on profits, so these rise as well. The increase in pure profit’s share decreases both labor’s and capital’s shares.
Because higher profits increase the return on capital, however, there must be a countervailing influence in order to generate the roughly constant returns that have been observed in the data. This is where the falling natural rate of interest comes in.

Other papers start with the same patterns but emphasize different explanations. The model of Farhi and Gourio (forthcoming) shares several basic structural elements with Eggertsson, Robbins, and Getz Wold (2018), but the analysis allows for and emphasizes the role of a changing risk premium in explaining several of the observed patterns. Relatedly, Karabarbounis and Neiman (2018) argue that a shrinking labor share and rising pure profit (what they call “factorless income”) can best be explained by a rising rental rate for capital. They point out that explaining a lower labor share of income via higher pure profit and/or returns to intangible capital, while consistent with many empirical patterns in recent decades, implies implausible empirical patterns when applied to the 1960s and 1970s. The mismeasured rental rate explanation avoids these counterfactual predictions and implies some other empirical patterns more consistent with observed data. At the same time, the study lacks a sharp explanation for what would cause the true rental rate to vary as it would need to in order to produce the macroeconomic outcomes that have occurred over the prior 60 to 70 years.

Other papers raise the general point that concentration can be related to macroeconomic outcomes through mechanisms other than market power; in particular, in Autor et al. (2017) and Bessen (2017), higher concentration and lower labor income may be part of an efficiency-enhancing shift. Both papers argue that concentration has grown because changes in market factors have created an environment that increases skewness—in revenues (as measured by rising concentration) and productivity, certainly, and perhaps in other dimensions. Something has flattened firms’ residual demand curves or marginal cost curves, be it increased scale economies, network effects, or improved abilities of consumers to find low-cost or high-quality firms. These changes lead to increased concentration (“superstar firms” in the parlance of Autor et al. 2017) but do not necessarily imply growth in market power. Increased scale economies may come from reductions in marginal cost that reduce the amount of inputs necessary to produce output—an efficiency enhancement. On the other hand, scale economies also require enough market power in equilibrium for firms to pay fixed costs and production costs of their inframarginal units. Network effects also have implications for both efficiency and market power. Consumers can obtain a utility benefit from network effects, but network effects can also cause lock-in, which gives firms pricing power. Improving consumers’ abilities to choose from whom they buy—which may come from changes in search, transport, or trade costs, for example—is likely to be efficiency enhancing.

Both Autor et al. (2017) and Bessen (2017) present evidence bolstering the case for an efficiency-enhancing mode of concentration being the primary actor in their data. Autor et al. find that industries that saw greater increases in concentration also saw on average faster growth in patent rates, capital intensity, and productivity. Bessen ties use of information technology systems to concentration as well as to
more skewed operating margins and productivity levels in an industry. To the extent that gains in concentration have been accompanied by efficiency gains, caution is again warranted when using concentration as a metric to infer market power.

Filling in the Macro Market Power Literature

In my discussion of various aspects of the macro market power literature, I have described some current vulnerabilities that in my opinion keep the literature’s conclusions from being dispositive. What might be done to fill holes and round out the evidence in a way that would allow more definitive conclusions?

One logical place to look for new threads that the literature could pick up is in the best practices of the well-developed microeconomic literature on market power. This literature typically starts with a recognition that the optimal price-cost markup depends on the slope of the inverse residual demand curve facing the firm. If that slope can be estimated, the implied profit-maximizing price–marginal cost margin can be backed out from that. Most of the microeconomic literature follows this logic and estimates the demand system for the products in the market (if the products are differentiated, this is typically accommodated by using a discrete choice demand system where the product attributes are included as demand shifters). While using demand-side data to infer marginal costs may seem surprising, in many settings the richness of the demand system offers the ability to estimate demand with some precision, and therefore implied margins, in ways that cost data alone could not. Moreover, one can typically jointly estimate both the demand and supply sides by parameterizing costs (again as a function of attributes if products are differentiated) and by using the restriction that the observed product price must equal the estimated marginal cost times the profit-maximizing markup implied by estimated residual demand.

However, the demand-system-estimation approach may not be feasible in the macro market power literature, with its broad combinations of industries and market settings. Specifying a realistic demand system typically takes a fair amount of knowledge about the nature of the product and the institutional details of the market. It is not practical to do this in studies that look across hundreds of very different markets. Taking an analogous approach with aggregate data would not work either. Pricing power depends on the slope of firms’ residual demand curves, not the slope of the market/industry demand curve. Backing out an implied markup and market power from a market/industry demand curve would be conceptually and empirically incorrect.

If microdata are available, one might imagine a more parametric approach to measuring marginal cost whereby a cost function is specified and estimated using observed variation in costs. Marginal costs would then be derived from this estimated function. This approach is limited by several factors, however. Many producer-level datasets report only revenues, which combine quantity and price, and quantity data of some type are required to estimate a production function.
For highly differentiated products, there may not be enough data to characterize
the cost function fully, given the multiple attributes that could shift costs. Also, to
estimate a cost curve, instruments that exogenously shift quantities are needed, and
these are not easy to find in many settings.

What is one to do, then? When it comes to estimating markups or measures of
market power for broad swaths of the economy, there may be no silver bullet. One
is left with a menu of imperfect choices.

Sometimes one can obtain direct measures of plausibly exogenous differences
in competition. In that case, concentration might be instrumented using those
measures, or alternatively, those measures could be used directly as explanatory
variables themselves.

If there does not seem to be an alternative to concentration as a measure of
market power, researchers should strive to demonstrate using ancillary evidence that
increases in concentration do in fact correspond to more market power rather than
efficiency in the market(s) they are studying. As an example of how this distinction
might be made, Autor et al. (2017) show that concentration is associated with inno-
vation, capital deepening, and productivity, which bolsters the case for efficiency
mechanisms. Alternate findings from this methodology would have supported a
market power interpretation.

Another area for ongoing research in the macro market power literature is
to characterize heterogeneity more fully, both across and within markets. As
mentioned above, the results from Crouzet and Eberly (2019) suggest that market
power can act broadly within some sectors but not others. They find the health-care
sector, for example, seems to have seen the influence of market power, while the
manufacturing and consumer sectors are not showing the signs of market power.
In turn, broad analysis across sectors can be compared to market-specific studies in
the micro literature; for example, Cabral, Geruso, and Mahoney (2018) and work
described in Gaynor (2018) support a finding of rising market power in the health-
care sector. Characterizing such sectoral differences and explaining where they
come from is important for understanding the mechanisms behind, and the effects
of, market power in macro settings.

Within industries, the skewness results shown by De Loecker, Eeckhout, and
Unger (2018)—that the increase in average measured markups is driven exclusively
by increases in the right tail of the distribution—are an example of the necessity of
understanding within-industry heterogeneity. Averages can obscure. Producers in an
industry differ markedly in their behavior, including in their responses to common
external influences. Market-, industry-, or economy-wide changes do not always, nor
likely even usually, reflect a common change experienced by all producers. Rather,
they reflect the summation of what are typically very different responses, which
includes reallocations of activity across heterogeneous producers. The experience
of the median producer (or even the average producer, if producers are equally
weighted) may not be informative about changes at the industry level. One cannot
simply rely on producer-level variation “canceling out” when looking at aggregate
changes. That variation is what creates the aggregate changes.
Conclusion

The macro market power literature has offered an immense service by documenting and emphasizing the potential connections between several trends: labor’s declining share of income, increasing corporate profits, increasing margins, increasing concentration, slower productivity growth, decreasing firm entry and dynamism, and reduced investment rates. While none of these is a perfect metric for market power, many (but not all) have been replicated in multiple venues with multiple techniques and as such can be considered reasonably robust. The fact that these changes are so noticeable and have been trending for so long (each for over a decade at a minimum, some approaching four decades now)—often in contrast to very different patterns before—creates an inherent interest and importance.

The market power story is very much a viable candidate explanation for the documented trends, especially in specific industries or sectors. However, I believe more evidence is yet required to make a broad-based increase in average market power the undisputed leading candidate explanation. Empirical gaps still need to be closed. There are plausible alternative stories, some accompanied by controveting empirical evidence to the market power hypothesis, that need to be rejected. Ultimately, indeed, it may be that the sources of the patterns are multicausal—some combination of greater intangible intensity, changing product-market substitutability, greater scale economies, and higher entry costs, all with potential implications for market power (though in possibly different directions). Moreover, the relative contribution of each could vary across sectors. Regardless, stronger conclusions will be warranted if researchers can make further progress in both qualifying and quantifying the roles of market power and alternatives.

I thank Susanto Basu, Steven Berry, Judy Chevalier, Dennis Carlton, Jan De Loecker, Jan Eeckhout, Gauti Eggertsson, Marty Gaynor, Germán Gutiérrez, Bob Hall, Adam Looney, Thomas Philippon, Fiona Scott Morton, Carl Shapiro, and the editorial staff of this journal for comments.

References


Barkai, Simcha. 2017. “Declining Labor and


Shapiro, Carl. 2018. “Antitrust in a Time of


