

Panel on Changing Market Structure and Implications for Monetary Policy

Chad Syverson

I would like to make four points.

Before I do, I want to be clear about how I am defining terms in the panel's topic. My research background is different from many of those in the room, so I just want to make sure you understand my assumptions. It does sound, though, from the context of the conversations so far, that my definitions are largely in line with others'.

I read "changing market structure" as referring to changes in the distribution of economic activity across firms within markets, industries, or sectors. An example getting a lot of attention lately is concentration, the distribution of revenues across companies. In addition, my sense based on both the research literature and the policy debate that market structure also has a connotation regarding market power, so I will hold that in mind too.

"Implications for monetary policy" means to me implications about multiple related things: the price level, inflation, pass-through of changes in capital prices, and the Philips curve.

Now to my four points of emphasis.

I. A Basic Paradox

I start by pointing out a seeming paradox I have noticed underlying the empirical patterns and economic research related to the theme of this symposium.

I begin with what is essentially an accounting identity: price, P , equals markup, m , times cost, C :

$$P = m \cdot C$$

The theory of profit maximization says for a firm, the cost C ought to equal marginal cost, and the markup m should be a function of customers' price sensitivity (less sensitivity, higher markup; technically speaking, the markup is a function of the slope of the inverse residual demand curve). 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 m as whatever multiplicative factor makes the relationship hold (m could even be less than 1 if price is less than cost for some reason). It is in this sense that the relationship is an identity and as such holds by definition.

We can express the relationship in terms of growth rates (percentage changes), which also holds by definition:

$$\text{Growth in } P \approx \text{Growth in } m + \text{Growth in } C$$

The growth rate relationship is approximate, but in situations with relatively modest growth rates that we are interested in, it is close to exact.

Consider the behavior of each of these growth rates over the past 10-15 years, as based on the best available economic measurement and research.

The left-hand side, the growth rate of prices, is inflation. What do we know about measured inflation over the past decade and a half? It has been low—in absolute terms, and perhaps more to the point, low relative to what past relationships between inflation and forcing variables would have implied.

The first term on the right-hand side, the growth rate of the markup, is one of the reasons for the theme of this symposium. Profit-maximization theory implies the markup reflects the degree of market power. The fact that broadly measured markups have been on an upward trend in recent years (departing from steady or even falling markups before that) has raised concerns about the aggregate implications of rising market power.

Finally, there is the growth rate of costs. One of the most important determinants of costs is productivity, and productivity growth has been in a slump since the mid-2000s. Productivity, as a measure of the amount of output obtained per unit of input use, is inversely related to costs; when output per unit input rises, companies have to buy fewer inputs to make a unit of output, lowering their costs. Thus when productivity grows more slowly than usual, as has lately been the case, cost growth is unusually high (or equivalently though perhaps more precisely stated, costs fall more slowly than usual).

Thus the data appear to say, separately, that a) price growth has been unusually low, b) markup growth has been unusually high (by being positive at all), and c) cost growth—at least its productivity component—has been unusually high. Each of these three growth rate patterns has been documented as an empirical phenomenon in multiple economies around the world. Thus, the paradox: How can two growth rates that have been unusually high (those of markups and costs) sum to a growth rate that has been unusually low (that of prices)? Something doesn't add up, quite literally.

One issue that needs to be resolved is that price and costs are nominal measures while productivity is a real measure. (Markups, as a ratio of nominal prices to nominal costs, are unitless.) We need to think about how real productivity affects nominal costs. This connection arises through the other component of costs besides productivity: input prices. Input prices can be expressed in nominal terms. Input price dynamics, along with productivity growth, affect the growth rate of nominal costs. We can investigate this by looking at unit labor costs, which conveniently combine both productivity and input price (wage) effects on costs, at least for labor inputs. Unit labor costs are the ratio of total compensation per hour worked to labor

productivity (that is, the output quantity per hour worked). They are therefore the nominal labor compensation required to build one unit of output.

In the U.S. aggregate data, while productivity growth slowed starting in the mid-2000s, unit labor cost growth stayed relatively steady. It in fact fell slightly, from 1.4 percent per year growth over 1995-2004 to 1.2 percent per year after 2005. Presuming (unmeasured) nominal capital costs did not stray too far from this pattern, this suggests that nominal cost growth over the past couple decades might have been pretty steady—a combination of slow productivity growth and equally slow input price growth.

Steady cost growth does not resolve the paradox. We still have unusually low measured price growth in the face of unusually high markup growth:

$$\text{Growth in } P \approx \text{Growth in } m + \text{Growth in } C$$

$$[\text{unusually low}] = [\text{unusually high}] + [\text{steady}]$$

I do not have an obvious resolution to the paradox. I have suggestions for partial resolutions, but I do not think they are sufficient.

A potential (partial) resolution can be elucidated by distinguishing among the types of cost in C . Productivity and unit labor cost probably most closely track average cost. But suppose prices are typically set in a profit-maximizing fashion so they depend on marginal cost. If average costs were falling slower than usual while marginal costs were falling more quickly, it would then be possible for unit labor cost growth to 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. The ratio of average cost to marginal cost is the scale elasticity of the cost function. If it is greater (less) than one, there are (dis)economies of scale because average costs fall (rise) as quantity rises. Under cost minimization, a relationship must hold between the markup and the scale elasticity;

namely, 1 minus the markup-elasticity ratio must equal pure profits' share of revenue (pure profits are revenues not paid to suppliers of labor, capital, and intermediates). Given some of the most commonly cited values for the increase in measured markups, the size of the increase in scale economies necessary to explain the observed changes in profits' revenue share is very large. In other words, the differential in the trends of average and marginal costs necessary to quantitatively resolve the differential trends of prices, markups, and costs is quite large, perhaps implausibly so.

Mismeasurement is another possible source of explanations for the paradox. Most everything I have seen about inflation measurement seems to indicate that if it is mismeasured, the true value is probably lower than the measured number, not higher. That would only deepen the paradox. The accuracy of markup measures from recent research is still being hotly debated. I and others have looked closely at whether true productivity growth is faster than the laggard rate measured over the past 15 years due to measurement problems. Our conclusion, arrived at separately and through diverse methods, is that it was not.

The upshot is that we appear to still be left with the paradox. Hopefully future research will help clear this up.

II. Heterogeneity Matters

Both papers today raise an important issue about heterogeneity and understanding changes in aggregates. Averages can obscure. One needs to recognize the roles of heterogeneity and shifts in composition within industries to grasp what shapes important aggregate values.

A ubiquitous empirical pattern is the incredible amount of differences among producers, within even narrowly defined industries or markets. This heterogeneity exists along multiple dimensions. Total factor productivity levels typically vary by a factor of two or more within industries. Labor productivity variation is even greater. Differences in size (measured by, say, output or employment) are larger still. Average worker earnings range extensively, again within narrowly defined industries.

All of these differences suggest, and empirical evidence bears out, that producers in an industry differ markedly in their behavior, including in their responses to even common external influences. This means that aggregate (market, industry, or economywide) changes do not generally reflect a common change across all producers within the market, industry, or economy. Rather, they reflect the summation of what are typically very different responses. For example, if industry output grows 2 percent year over year, this is not because every producer in that industry saw its output grow 2 percent, even approximately. Instead, some producers likely grew very fast, some more modestly, some not at all, and some shrank, perhaps considerably. Moreover, new firms would have come into the industry, and some existing producers would have exited. Similar patterns would hold for aggregate productivity growth, price changes, employment shifts, and so on. Shifts in activity among producers within an industry can and do have a role in shaping aggregates that is as large, or even larger, than changes within producers.

The Van Reenen and Crouzet-Eberly papers vividly demonstrate this principle. The Van Reenen paper shows that changes in industry concentration, average productivity, wage inequality, labor's share of income are not the result of common changes across all producers. Instead, they are all driven by within-industry reallocation of activity across heterogeneous producers. Crouzet and Eberly's paper demonstrates that the coincident changes in concentration and average intangible intensity of industries result from within-industry reallocations of activity across producers with differing amounts of intangible investment.

To understand why these changes are happening and the potential consequences of such shifts, one must understand the fundamental fact that industry producers are different from one another and will evolve differentially going forward. The experience of the median producer (or even the average producer, if producers are equally weighted) is uninformative 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.

III. Pass-Through and Monetary Policy

For my third point, I took the panel's title to heart and tried to write down, as someone who studies market power, the simplest framework I could imagine for thinking about how market power affects the efficacy of monetary policy.

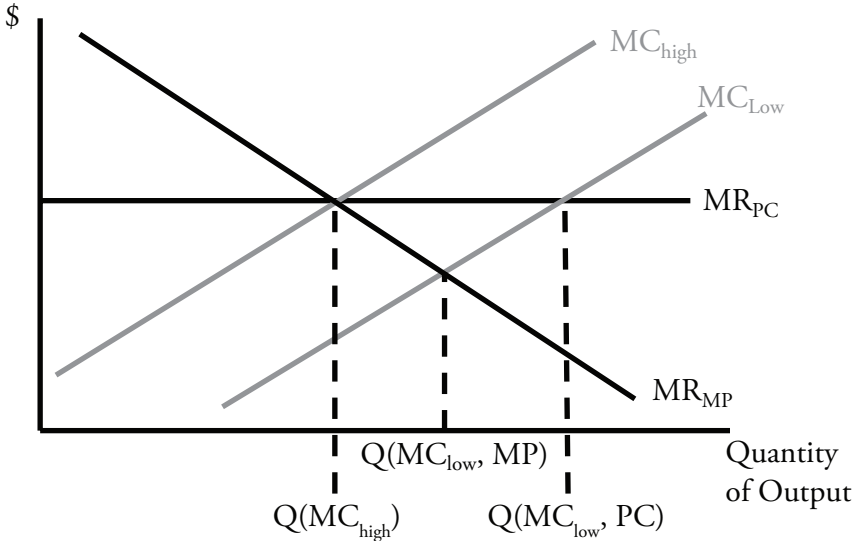
A key mechanism of action for monetary policy is that the changes in capital costs induced by policy should cause companies to adjust their real activity levels. Reductions in capital costs due to expansionary policy should lower companies' cost levels, leading them to invest more and expand employment and output levels. Contractionary policies raise firms' costs and have the opposite effect.

Of course, the magnitude of companies' responses to policy-induced cost changes depends on the ties between companies' costs and their desired activity levels. Standard economic theory offers guidance on the nature of those ties and, of particular interest to this symposium, their connection to market power.

Profit maximization implies firms should operate at output levels that equate their marginal revenues from that output with their marginal costs. Monetary policy, by affecting firms' capital costs and as such their marginal costs, will shift firms' profit-maximizing output levels accordingly. Note that marginal costs in this analysis include capital costs, so the analysis spans a suitably long horizon that producers can adjust their levels of capital inputs in response to monetary policy. This is of course implicitly the horizon over which standard analyses of monetary policy's effects on investment are made, so I am not departing substantially from typically studied response periods.

Chart 1 shows how the absence or presence of market power can affect the size of the response to monetary policy. The chart shows two marginal cost curves: one when capital costs (and therefore marginal costs) are relatively high, MC_{high} , and one where costs are relatively low, MC_{low} . The former might be thought of as conditions under which monetary policy is relatively tight, the latter when it is relatively loose.

Chart 1
Firms' Responsiveness to Monetary Policy With
and Without Market Power



The profit-maximizing output level depends not just on marginal costs but also on marginal revenue.

For a firm operating in a perfectly competitive market, marginal revenue is the same at all output levels: the firm can simply sell every unit it would like at the going market price. (Because of this, a perfectly competitive firm's marginal revenue curve is the same as its residual demand curve—flat, at the level of the market price.) The marginal revenue of a company in a perfectly competitive market is MR_{PC} in Chart 1. When capital costs are such that the perfectly competitive company's marginal cost curve is MC_{high} , its profit-maximizing output quantity is $Q(MC_{high})$. If instead marginal costs were MC_{low} , its profit-maximizing output quantity would be $Q(MC_{low}, PC)$. Thus the expansionary effect of monetary policy that reduced capital costs enough to shift marginal costs from MC_{high} to MC_{low} would be to increase the firm's output from $Q(MC_{high})$ to $Q(MC_{low}, PC)$. The sum of similar such changes across other firms would give the aggregate output change due to the expansionary monetary policy.¹

We can contrast this effect in a perfectly competitive market with one where companies have market power. In this case, companies' marginal revenues are no longer invariant to output. Their marginal revenue curves instead slope down; the more output they make, the lower is the marginal revenue from making an additional unit of output. Because firms with market power face downward sloping (residual) demand curves, when they expand output they also reduce the market price for their product, driving down marginal revenue.

Chart 1 also shows a marginal revenue curve for a firm with market power, MR_{MP} . For the purposes of comparison, this has been drawn to intersect with MR_{PC} at MC_{high} . Generally, a company with market power will produce less output than it would in a perfectly competitive market. However, because I am interested in comparing output *changes* here—specifically, in response to changes in monetary policy—there is no loss of generality in norming their outputs to be equal at some reference cost level.

Consider what happens when monetary policy is loosened and the marginal cost curve of a company with market power falls from MC_{high} to MC_{low} . Its profit-maximizing output quantity rises from $Q(MC_{high})$ to $Q(MC_{low}, MP)$. This increase in quantity is smaller than that obtained under perfect competition, where the same change in costs expanded output from $Q(MC_{high})$ to $Q(MC_{low}, PC)$. This is a key result of the analysis: companies with market power expand less in response to a given loosening in monetary policy than do those in a perfectly competitive market.²

The same logic holds for a contraction in monetary policy: companies with market power cut output by less than would similar firms in a perfectly competitive market.

While not shown in the chart for the sake of avoiding clutter, one can easily see that if the marginal revenue curve were even steeper than MR_{MP} the response to the same change in monetary policy and marginal costs would be smaller than the $Q(MC_{high})$ to $Q(MC_{low}, MP)$ quantity difference in the chart. Thus the change in real activity induced by a given shift in monetary policy will decline as the marginal revenue curve steepens further. Because of this, one might think that

as companies gain still more market power, their responsiveness to monetary policy would shrink even more. This need not be so, however. While increases in market power do imply steeper residual demand curves, they do not necessarily imply steeper marginal revenue curves. The steepness of the marginal revenue curve does depend on the steepness of the demand curve, but it also depends on the *change in* the steepness of the demand curve (i.e., whether it is flattening or steepening as output changes) as well as the size of the change in the firm's profit-maximizing quantity as market power changes. Depending on the shape of the demand curve, these two additional factors can cut the other way, sometimes causing marginal revenue curves to flatten even as market power increases.

Therefore, while moving from a situation with perfect competition to one with market power does steepen marginal revenues (from flat to downward sloping) and thus reduce the real expansionary effects of loosening market power, it is ambiguous whether further increases in market power further limit that expansionary effect. To answer that question definitively, one needs to know what companies' demand curves look like. (For those familiar with the price-cost pass-through literature, this is the same math; we are simply exploring the quantity manifestation of it.) I regret that this analysis cannot offer a more definitive comparative static, but that is just the way it is.

IV. Measuring Competition with Concentration

I would like to close with a discussion of the measurement of competition.

Concentration is a market outcome, not a market primitive. In other words, concentration is not an immutable core determinant of how competitive an industry or market is. One cannot look at concentration and infer directly from it the extent of competition and the degree of companies' market power. Competition and concentration are related because the nature and intensity of competition in an industry combines with many other supply and demand factors to determine concentration. However, competition drives concentration, not the other way around.

This does not just mean that concentration is a noisy barometer of market power. Rather, we cannot even be generally clear about the direction of the barometer's orientation. Concentration can be associated with less competition, but it can also be associated with more competition.

An entire class of commonly used models in economics, where firms with differing costs/productivities sell differentiated goods (i.e., models that embody the heterogeneity discussed in my second point above), predicts a positive correlation between competition and concentration. These models all imply that an increase the elasticity of substitution among the outputs of industry producers—things that make their products more substitutable to buyers or makes it easier for buyers to discover, like reduced trade barriers, search costs, or transport costs—will shift output from higher-cost/lower-quality to lower-cost/higher-quality firms. This output shift generally raises concentration. Often, also, markups decline because the increased willingness or ability of buyers to substitute to other sellers reduces industry producers' market power. The market simultaneously becomes both more competitive and more concentrated.

One often hears a complaint that it is possible to write a model do anything. I am not enough of a theorist to prove *that* proposition as a theorem, but I conjecture it is correct. Nevertheless, this connection between greater competition and more concentration is not just a theoretical curiosity. There have been many empirical studies in varied settings that show exogenous increases in substitutability/competitiveness (e.g., reductions in trade, transport, or search costs) lead to concentration-increasing shifts in activity away from smaller, higher-cost/lower-quality producers and toward larger, lower-cost/higher-quality producers. I have participated in that literature, but I am hardly alone. I do not think it is an exaggeration to say that there are scores, perhaps hundreds, of such studies.

I work a lot in the industrial organization research literature. Industrial organization is the field of economics that focuses on studying market power. The field used to do a lot of comparing market outcomes like prices, margins, and profit rates to concentration levels. This was the so-called structure-conduct-performance

research literature. However, researchers in the field essentially stopped doing such comparisons about 35 years ago. There was a recognition that one needed to be *very* careful in inferring causal effects of differences in concentration on market power and related outcomes, especially when making comparisons across markets or industries, with their differing demand and technology fundamentals. Differences in fundamentals drive both the extent of competition and the degree of concentration in the industry, and it can lead to either a positive or negative correlation between the two depending on the circumstances.

Actually, I am being diplomatic with my description. Many in industrial organization view comparing prices, markups, or profit rates to across-market differences in concentration—and thinking that comparison reveals the effects of market power—as the field’s original sin. A sin for which we should ritually self-flagellate to remind ourselves not to err as our forebears did.

While I am on the less strident about this line of thought than many other industrial organization researchers, I believe the caution about reading concentration as a proxy for market power is well warranted. There were good reasons for the field to swear it off (particularly, again, for across-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.

Endnotes

¹If the marginal cost reductions are marketwide, as we might expect from monetary policy expansions, this would shift the entire market's marginal cost curve and reduce the equilibrium price. This would shift down firms' marginal revenue curves, adding to the size of their quantity expansions.

²Again, the aggregate change would sum firm-level responses across all producers in the market, and a marketwide reduction in marginal cost would reduce the equilibrium price and drive additional expansion. For a given market demand curve, this expansion would generally be no larger than that in a perfectly competitive market, preserving the "market power leads to a smaller response than perfect competition" result of the firm-by-firm analysis here.

