# Markups and Markdowns

*Chad Syverson*[1]

*University of Chicago Booth School of Business and NBER*

*August 2024*

**Abstract**

Interest in market power has recently surged among economists in many fields, well beyond its traditional home in industrial organization. This has focused empirical attention on markups, the ratios of price to marginal cost in product markets, and markdowns, the ratios of inputs' marginal products to their paid wage in factor markets. In this review, I offer a conceptual overview of both metrics and survey recent research examining them. I pay particular attention to the distinct interests that microeconomists and macroeconomists have had regarding these metrics, as well as topics that have bridged and are bridging these often distinct literatures.

Keywords: market power, product markets, factor markets, firms, cyclicality, inflation

---

[1] chad.syverson@chicagobooth.edu; 5807 S. Woodlawn; Chicago, IL 60637. I am grateful to Nathan Miller and James Traina for detailed comments. Disclosure statement: I am serving as an expert witness in a legal case involving market power issues.

A confluence of recent trends has heighted interest in market power among economists in many fields. Long a focus of industrial organization (IO), a growing body of theoretical and empirical research has pointed to market power's potential for explaining economic phenomena in many settings.

A natural focus of market power research is the markup, the ratio of price to marginal cost.[2] Markups are the most direct measure of the textbook definition of market power in product markets: a firm being able to influence the price at which it sells its product (e.g., Pindyck & Rubinfeld 2018, Goolsbee et al. 2024). Profit-maximizing firms use this ability to set price above marginal cost, that is, to set the markup above one. Besides indicating its presence, the how high the profit maximizing markup is above one measures the magnitude of market power.

Markdowns are the buy-side analog of markups, present primarily in factor markets. Markdowns reflect a firm's ability to hold the price it pays for an input below the input's marginal product. While factor market power analysis has traditionally taken a back seat to its product market sibling (modulo some famous exceptions, e.g., Robinson 1932), the past decade has seen burgeoning interest in monopsonistic behavior and its consequences. Analysis of monopsony-driven markdowns is a natural complement to monopoly-driven markups. As I discuss below, they share many theoretical implications and empirical patterns.

In this review, I organize and summarize economic research on markups and markdowns. It is worth prominently noting some disclaimers first. I mean the review to be an introductory overview of the related work. It is not—it cannot be, due to practical space constraints—a discussion of every result in all relevant studies. I choose the results and studies discussed here in the hopes of structuring an instructive and useful narrative which interested readers can use as a guide to dig deeper. Exclusion of particular findings, papers, or articles is not a sign that I view those works as unimportant or wrong. Further, the topical organization of the review is one I find useful, but it is far from the only one. Many potential common threads

[2] Closely related are the difference between price and marginal cost (often referred to as the *margin*) and the Lerner Index, the margin as a share of the price. The conversions among the three metrics are clear enough, but one must be careful when comparing them across different studies, as terminology is not completely standardized and sometimes even the labeling within a study slips.

tie together markup and markdown research. I weave with those I have found most fruitful, but readers may find others. For additional perspectives, Berry et al. (2019), Miller (2024), and Shapiro and Yurukoglu (2024) offer other reviews of the markups literature, primarily from an IO perspective. I also include occasional commentary notes that reflect my views but not necessarily those of other researchers.

The blossoming of market power research and its empirical manifestations in markups and markdowns has greatly enriched understanding across many fields. I am personally excited about this. I hope this review is not the end of readers' travels into this area but rather just a step (perhaps the first) along the way.

## 1. Some Useful Relationships with Markups and Markdowns

Reviewing the conceptual relationships between markups, markdowns, and various other economic objects (primitives and outcomes both) sheds light on the nature of market power and can help explain the variations in markups and markdowns we observe in the data. I discuss several such relationships in this section.

### 1.1. Markups, Costs, and Inflation

One markup relationship is simply an accounting identity yet nevertheless demonstrates how widespread changes in markups can influence other macroeconomic variables. Specifically, one can express a price, $P$, as the product of a markup $\mu$ and cost $C$:

$$P = \mu \cdot C$$

Under profit maximization, the cost $C$ ought to equal marginal cost, and the markup $\mu$ should be a function of consumers' price sensitivity. However, the relationship is still quite useful and applicable even if prices are not set to maximize profits. One can define the markup $\mu$ as whatever multiplicative factor makes the relationship hold between any consistently measured price and cost ($\mu$ could be less than 1 if price is less than cost for some reason).

In growth rates, the relationship is:

$$Growth\ in\ P \approx Growth\ in\ \mu + Growth\ in\ C$$

That is, inflation equals the sum of markup growth and cost growth. I return to this when discussing some of the macroeconomic implications of markups below. This relationship is approximate but close to exact when growth rates are relatively modest.

### 1.2. Markups, Profit Shares, and Scale Elasticities

Another useful decomposition of the markup is virtually assumption free and therefore applies under general conditions. At the same time, it has realistically measurable implications in many types of data.

Its derivation begins by multiplying and dividing the markup, $\mu$, by average costs:

$$\mu \equiv \frac{P}{MC} = \frac{P}{AC}\frac{AC}{MC}$$

Next, $P/AC$ is multiplied and divided by the output quantity to rewrite it as the ratio of revenues to total costs. Further, by definition, the $AC/MC$ ratio is the scale elasticity of the cost function (equivalently, the inverse of the elasticity of costs with respect to quantity).[3] When marginal costs are less than average costs, average costs are falling in quantity, and the scale elasticity is greater than one. If MC > AC, there are diseconomies of scale, and the scale elasticity is less than one.

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

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

Defining pure profit's revenue share as $s_\pi \equiv (R - TC)/R$, we can rewrite the markup as:

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

The markup equals the inverse of one minus profits' share of revenue times the scale elasticity. As mentioned, the relationship applies under very general conditions; all that is required is differentiability of the function that relates output to costs. This function does not even need to be the standard cost function of production theory; i.e., the total cost expression

[3] For any differentiable cost function $C(Q)$, the elasticity of costs with respect to quantity is $C'(Q)(Q/C) = MC(1/AC)$. For homothetic production functions, the scale elasticity equals the returns to scale of the production function.

evaluated at the cost-minimizing factor demands. Whatever cost-quantity ties are implied by the observed production behavior apply.

This relationship should hold at the producer level—that is, the level of the entity whose choices define the cost function. It can serve as a way to measure markups or as an empirical check on markup estimates obtained using other approaches.

The right hand side of this expression is measurable in several settings. Researchers can construct reasonable approximations to profit shares from standard accounting data. Obtaining scale elasticities can be more involved, as it requires estimating the producer's production or cost function. A substantial literature serves as a guide to doing this.

If applying the relationship at more aggregate levels, one needs to recognize that the expression holds at the individual producer level, and Jensen's inequality implies the average of the producer-level markup-to-scale-elasticity ratios will not equal the analogous ratio computed in aggregate data.

Empirical implementation specifics aside, the relationship implies common ties among variations in markups and other objects of interest. For example, if producers are known to have seen substantial markup growth, either pure profit shares of income increased or the producers' scale elasticities must have risen. Likewise, one would expect observed changes in scale elasticities (or differences in the cross section) to show up in markups if profit shares were known to be constant.

The economic intuition behind what is essentially an accounting identity is as follows. Markups cause the producer's revenues to exceed its costs, at least for marginal units. This extra revenue must either lead to higher profits $s_\pi$ or be used to pay off costs incurred in producing inframarginal units where average cost exceeded marginal cost. Equivalently, if the production process involves scale economies at the profit-maximizing quantity, the producer must somehow pay off the excess AC over MC (say because of fixed costs). This would not be possible if all units were sold at P = MC. A markup must be present to yield the extra revenue required to pay off those costs.

### 1.3. Markups and the Elasticity of Demand

One of the best-known relationships involving markups is their tie to the elasticity of demand. This arises from the price-setting problem of a profit-maximizing producer. Facing residual demand curve $q(p)$, the producer's price should conform to the following first order condition

$$q + p\frac{\partial q(p)}{\partial p} = \frac{\partial C(q)}{\partial q(p)}\frac{\partial q(p)}{\partial p}$$

where $C(q)$ is the cost function. Rearranging and recognizing that the derivative of the cost function with respect to quantity is marginal cost yields

$$\mu \equiv \frac{p}{C'(q)} = \frac{|\eta|}{|\eta| - 1}$$

where $|\eta|$ is the absolute value of the elasticity of the residual demand curve. Note that $\eta < 0$, and further, profit maximization requires the firm operate an a portion of its residual demand curve where $\eta < -1$.

This equation expresses the familiar intuition that the profit-maximizing markup falls as the residual demand curve become more elastic. Indeed, in the polar case of perfect competition where the residual demand curve is infinitely elastic, $\mu = 1$ and P = MC.

It is worth noting some points regarding this expression's generality. It assumes profit maximization, so it is not as general as the associations above. Nevertheless, given profit maximization, it applies under quite general competitive conditions. The demand curve $q(p)$ is a residual demand curve, so it embodies the demand faced by the producer conditional on the quantity choices of other firms operating in the market. It therefore applies under many potential market structures, from single-firm monopoly to monopolistic competition, to various oligopolies, to the limit case of perfect competition. It does assume the producer has only one product whose demand is affected by the chosen price, however. That is, the producer makes no substitutes or complements in the demand system. If it did, then the producer would account not just for the effect of a product's price in demand for that product, but also on demand for the producer's other products. Substitutes within the producer's portfolio would weaken the incentive to reduce price to increase quantity demanded, as some of the extra demand would be pulled from the producer's other product(s). This yields a higher markup, all else equal. Complements have the opposite effect, tending to reduce the markup on the reference good.

As discussed further below, using this relationship to measure markups requires only knowledge of the demand side of the market. In some situations, estimating producers' demand elasticities can be easier than estimating their economies of scale, as required to measure markups using the expression in the prior section. Of course, with adequate data availability, the two conditions can be used as checks on one another or combined into a single estimator.

### 1.4. Markups, Factor Cost Shares, and Factor Output Elasticities

Another markup relationship used recently to great empirical effect emerges from cost minimization.

For any factor $X$ that is *flexible*—i.e., the firm can completely adjust usage to its desired level within a period, unlike fixed or quasi-fixed factors—cost minimization implies the following first order condition:

$$p_X = \lambda f_X(\cdot)$$

Here, $p_X$ is the unit price of $X$, $f(\cdot)$ the production function, and $f_X(\cdot)$ the marginal product of $X$. The value $\lambda$ is the Lagrange multiplier in the cost minimization problem, and as such is marginal cost (the derivative of the objective function with respect to the constraint). Multiplying and dividing either side by various convenient variables, recognizing that $f(\cdot)$ is output, and rearranging yields

$$\mu \equiv \frac{p}{\lambda} = \frac{\theta_X}{s_X}$$

where $\theta_X$ is the elasticity of output with respect to $X$, and $s_X$ is expenditures on $X$ expressed as a share of revenue.

One way to think about the intuition behind this relationship is that $\theta_X$ captures what output $X$ allows the producer to make (and sell), while $s_X$ captures what part of that revenue the firm pays to $X$. When revenue from sales exceeds payments to the input that yields that marginal revenue, the markup is greater than one.

Note that the above relationship is tied to the output elasticity of input $X$—how much output quantity increases when additional $X$ is hired. In empirical practice, however, researchers often do not directly observe producers' output quantities, but rather revenues instead. Researchers therefore commonly replaced $\theta_X$ with the *revenue* elasticity of input $X$, the responsiveness of revenues to increases in $X$. However, Bond et al. (2021) show that this is not an innocuous proxy for the output elasticity. Indeed, profit maximization implies a pathological outcome: the ratio of $X$'s revenue elasticity to its revenue share is always 1, regardless of the true value of the output elasticity $\theta_X$.

To see why, denote the revenue elasticity of $X$ as $\rho_X$. Then:

$$\rho_X = \frac{d}{dX}[p(q)q]\frac{X}{pq} = \frac{d}{dX}\left[p\big(f(\cdot)\big)f(\cdot)\right]\frac{X}{pf(\cdot)} = \left(\frac{|\eta|-1}{|\eta|}\right)\theta_X$$

where $|\eta|$ is the absolute value of the price elasticity of the producer's residual demand, as above. This implies the ratio of this elasticity to the revenue share of X is:

$$\frac{\rho_X}{s_X} = \left(\frac{|\eta|-1}{|\eta|}\right)\frac{\theta_X}{s_X} = \left(\frac{|\eta|-1}{|\eta|}\right)\mu$$

But by the profit-maximizing pricing condition derived above, the term in parentheses is $\mu^{-1}$. Substituting this in, we see that $\rho_X/s_X = 1$ regardless of the output elasticity of $X$ or for that matter the price elasticity of demand. Rather than proxying for the output-elasticity-to-cost-share ratio, the revenue-elasticity-based ratio is instead uninformative about the markup.

### 1.5. Markdowns

The derivations above focused on product market power and markups. Analogous relationships exist for markdowns tied to factor market power.

One of the more commonly employed is the following. Suppose a producer has monopsony power in the market for input $X$, reflected in the fact that it faces an upward sloping residual supply curve of $X$. Call this inverse residual input supply curve $w(X)$, where $w$ is the price the producer pays for a unit of $X$. The producer's total expenditures on the input are $w(X)X$.

Let the producer's revenue function be $R(q)$. If the producer hires $X$ to produce output using the production function $q = f(X)$, the first order condition for profit maximization is

$$R'(q)f'(X) = w'(X)X + w(X)$$

Denoting $\varepsilon_X$ as the elasticity of the residual factor supply curve, some manipulation shows:

$$\frac{R'(q)f'(X)}{w(X)} = \varepsilon_X^{-1} + 1$$

The ratio of the input's marginal revenue product to its price equals one plus the inverse elasticity of residual factor supply. When the producer has monopsony power, $\varepsilon_X^{-1} > 0$, so the ratio is above one. That is, the input is paid less than its marginal product. The more factor market power the producer has, the larger the wedge between the input's marginal revenue product and its price.

Note that this wedge exists even if the producer is a price taker in the product market, in which case $R'(q)$ is constant and equals the output price. In the more general case where the producer has both product and factor market power, further manipulation shows:

$$\frac{\theta_X}{s_X}\left(1 - \frac{1}{|\eta|}\right) = \varepsilon_X^{-1} + 1$$

where the output elasticity and revenue shares of $X$ are defined as above, as is the elasticity of demand. Equivalently,

$$(\varepsilon_X^{-1} + 1)\mu = \frac{\theta_X}{s_X}$$

This clarifies that the markup expression derived in Section 1.4 is actually for the special case when the firm has only product market power. When it has market power in both the product and factor markets, the combination relates to the ratio of the firm's output elasticity with respect to $X$ and the share of revenue paid to it. The intuition is as before, but with a new role for factor market power. Some combination of a product market markup and a factor market markdown drives a wedge between what an input is paid and its contribution to production and revenue.

I use these conceptual relationships to augment my discussion of the recent literature on markups and markdowns. I organize the discussion into two broad, familiar swaths—microeconomics and macroeconomics (though as noted below there are plenty of overlaps