

Revisiting Marshall's Third Law: Why Does Labor's Share Interact with the Elasticity of Substitution to Decrease the Elasticity of Labor Demand?

Saul D. Hoffman

Abstract: The third Marshall–Hicks–Allen rule of elasticity of derived demand purports to show that labor demand is less elastic when labor is a smaller share of total costs. As Hicks, Allen, and then Bronfenbrenner showed, this rule is not quite correct, and actually is complicated by an unexpected negative relationship involving labor's share of total costs and the elasticity of substitution. The standard intuitive explanation for the exception to the rule presented by Stigler and referenced in many textbooks describes a situation rather different than the one described in the rule. The author presents an example that illustrates the peculiar negative impact of labor's share operating via the elasticity of substitution and then explains why the unexpected relationship between labor's share of total cost, the elasticity of substitution, and the elasticity of labor demand holds.

Keywords: labor demand, Hicks–Marshall rules
JEL codes: J010, J200

The third Marshall–Hicks–Allen rule of the elasticity of derived demand has bedeviled labor economists for many decades. This rule purports to show that the elasticity of labor demand is less when labor is a smaller share of total costs. It is often referred to by the aphorism that “it is important to be unimportant,” meaning that when labor's share is small, the resulting demand curve will be more inelastic, giving labor (or labor unions) more power to increase wages with less of a reduction in employment. A common example contrasts the demand situation facing an industrial union (large share) and a craft union (small share), with the prediction that, *ceteris paribus*, the industrial union would face a more elastic demand curve.

The problem is that this rule is not quite correct, a point that is not universally recognized even among labor economists. It holds, as Hicks (1932) and then Allen (1938) pointed out, only when the elasticity of final product demand (η) is greater than the elasticity of substitution (σ). The explanation by Stigler (1966) that is included in two well-known labor economics textbooks (Ehrenberg and Smith

Saul D. Hoffman is professor of economics and chair of the Department of Economics at the University of Delaware (e-mail: hoffmans@lerner.udel.edu). Copyright © 2009 Heldref Publications

2009 and Borjas 2008), is incorrect. It describes a situation rather different than the one described in Marshall's controversial third rule. As far as I can tell, no straightforward and reasonably intuitive economic explanation exists.¹

In an interesting article more than four decades ago, Bronfenbrenner (1961) presented a clear historical account of the elasticity rules that focused on the derivation of the rules from Marshall (1920) to Hicks's *Theory of Wages* (1932, Appendices 3 and 4) to Allen's *Mathematical Analysis for Economics* (1938), with special emphasis on the pesky third rule. But Bronfenbrenner deftly stepped aside from explaining the rule, preferring to emphasize the related important influence of labor's share in determining the overall elasticity. He wrote that the "variation of the elasticity of derived demand for a productive service with the relative weight of that service in total cost is only a side issue. The important role . . . is in determining the relative weights of η and σ " (Bronfenbrenner 1961, 259).

I pick up where Bronfenbrenner left off, first explaining the problem of the third rule and also its common, but incorrect, explanation. I then document the exception to Marshall's third law with a simple numerical example using a Cobb–Douglas production function. In particular, I show that the output-constant response to a wage change is negatively related to labor's share for a given elasticity of substitution. Finally, I provide an economic explanation of why labor's share affects the impact of the elasticity of substitution on the elasticity of labor demand. The explanation turns out to be relatively simple, turning on the relationship between the share of labor in total costs and a firm's cost-minimizing ratio of labor to capital or, equivalently, on the common marginal rate of technical substitution (MRTS) of firms facing the same input prices, but with different initial labor–capital ratios.

BACKGROUND

Marshall, and later Hicks and Allen, established four famous rules of the determinants of the elasticity of derived factor demand, rules that have been taught to generations of economics students.² The first two rules, which are by far the most famous, relate to the elasticity of final demand and the ease of substitution in production, both of which increase the elasticity of derived demand. The fourth relates to the elasticity of supply of the other factor or factors of production. This rule is often treated as a minor issue—Allen ignored its role entirely. The third and most controversial rule relates labor's share in total cost to the elasticity of derived demand. Marshall argued that labor demand was more inelastic when labor's share of total costs was smaller. The underlying intuition is that any given increase in the wage will have a bigger impact on average cost and thus price when labor is a more important share of total cost. For any given elasticity of final demand, the impact on quantity demanded and eventually on labor demanded will then be greater. Thus, it is "important to be unimportant," where "important" means to face a more inelastic demand curve and "unimportant" means to be a small share of total cost. That much seems reasonable.

Unfortunately, as Hicks and then Allen showed, this third law is not strictly true. It holds only if the elasticity of final demand is greater than the elasticity of substitution. The issue is easiest to see in Allen's version. The Allen equation for

the elasticity of labor demand is $\eta_L = (1-S)\sigma + S\eta_Q$, in which η_L is the absolute value of the elasticity of labor demand, S is the share of labor in total cost, σ is the elasticity of substitution, and η_Q is the elasticity of final demand, also treated as a positive number.³ Allen's equation clearly shows that the elasticity of labor demand is a weighted average of substitution and scale effects, represented by σ and η_Q , respectively. Taking the derivative of η_L with respect to s reveals the complication first discovered by Hicks using his far more complex formulation: $\partial\eta_L / \partial S = \eta_Q - \sigma$. This derivative is not unambiguously positive or negative, but rather depends on the relative size of the elasticity of final product demand and the elasticity of substitution. If $\eta_Q < \sigma$, it is apparently important to be important! Note also, that $\partial\eta_L / \partial\sigma = (1-S)$ and $\partial\eta_L / \partial\eta_Q = S$, so that these effects depend negatively and positively, respectively, on labor's share.

It is no surprise that the elasticity of labor demand depends positively on the elasticity of the demand for the final product, as Allen's equation shows. The positive impact of S through η_Q is also straightforward. A larger S means that any increase in wages has a bigger effect on average cost and thus price, and eventually, via η_Q on the amount of labor demanded.

It is also no surprise that the elasticity of labor demand depends on the elasticity of substitution, because that has already been established by the second rule. But the Allen equation suggests two puzzles about the role of labor's share in the elasticity of labor demand. First, why should labor's share of total costs have anything at all to do with the effect of the elasticity of substitution on the elasticity of labor demand or, equivalently, why does the elasticity of substitution have anything to do with the effect of labor's share on demand elasticity? And, second, even more unexpectedly, why is the effect negative? Why, for example, isn't $\eta_L = (\sigma + S\eta_Q)$ or even $S(\sigma + \eta_Q)$? Those relationships seem far more sensible. Indeed, Bronfenbrenner (1961, 258) wrote that "Common sense appears to suggest that a high elasticity of demand for the product and a high elasticity of substitution between services should *reinforce* rather than offset each other in increasing the 'importance of being unimportant'" (emphasis added).

I have never seen an adequate explanation for why this peculiar negative relationship holds, and I suspect that very few labor economists actually understand it.⁴ Most labor economics textbooks do not include the qualification, which may, frankly, be subtler than is necessary for many undergraduate audiences (see, without assignment of any blame whatsoever, Hyclak, Johnes, and Thornton 2005, 51; Hamermesh and Rees 1993, 146, and Reynolds, Master, and Moser 1998, 91).

Even when the exception is noted and addressed, it is not explained correctly.⁵ The most common explanation comes from Stigler in his *Theory of Price* (1966). His explanation is reported virtually identically with attribution in long footnotes in two well-regarded U.S. labor economics textbooks (Ehrenberg and Smith 2009, 100; Borjas 2008, 131). This account involves carpenters of identical skill who are classified by their ancestry, e.g., "African-, Asian-, German-, Hispanic-, Irish-, and Italian-American" (Ehrenberg and Smith, 2009, 100). If each such group was treated as a separate factor of production, its share of total cost would be small, but because so many perfect substitutes exist, its elasticity of substitution would be very high (presumably ∞). Thus, if any one carpenter group attempted to increase

its own wage, demand would be highly elastic. In contrast, if all the groups were treated as one factor of production, they would account for a larger share of total cost, but because they face fewer substitutes in production, their demand curve would be less elastic. Stigler's (1966, 244) key point is, in fact, that the elasticity of substitution typically varies along with labor's share, with a small share implying greater substitutability.

This explanation is perfectly correct as an explanation for why a factor for which perfect production substitutes exist would be likely to face an elastic demand curve, courtesy of the elasticity of substitution. But it is not an explanation for Marshall's law in the form presented by Allen or Hicks. In Allen's equation, the elasticity of substitution is held constant as labor's share varies: $\eta_L = (1-s)\sigma + S\eta_Q$ and $\partial\eta_L/\partial S = \eta_Q - \sigma$. Stigler's verbal example represents a different formula for the elasticity of labor demand, one in which σ is replaced by $\sigma(s)$, with $\sigma' < 0$. The corresponding Stiglerian demand elasticity-labor share derivative is $\partial\eta_L/\partial s = \eta_Q - \sigma + (1-S)\sigma'$. The third term is negative and thus makes the relationship more likely to be negative; in this example, it is important to be important. In fact, one might imagine from the example and formula that the relationship between labor's share and demand elasticity was more often negative than positive. Is it important to be unimportant or to be important? And why?

ANALYSIS

In this section, I provide a simple numerical example using a Cobb–Douglas production function to show that the peculiar negative impact of labor's share on the elasticity of labor demand via the elasticity of substitution result is, in fact, true. There is no doubt that it is true, because that is what Allen's equation (and Hicks's more complicated equation) reveals. It is instructive and reassuring, nevertheless, to see an example that confirms it with concrete and easily understood computations. I then present two complementary explanations that explain why it is true. The Cobb–Douglas case is easy to work with, but it does not restrict the applicability of the explanation.

A Numerical Example

Start with a Cobb–Douglas production function $Q = L^a K^{(1-a)}$. The corresponding cost-minimizing choices of L and K must always satisfy

$$[a/(1-a)] \times (K^*/L^*) = w/r \tag{1}$$

in which K^* and L^* represent the best choices of L and K and the left hand side of this equation is the MRTS. Rewriting the cost-minimization condition to emphasize the labor-capital ratio,

$$L^*/K^* = [a/(1-a)] \times (r/w) \tag{2}$$

For simplicity and without any loss of generality, let $r/w = 1.0$ and consider labor demand for two production functions, one with $a = 0.8$ and the other with $a = 0.2$. For $a = 0.8$, it follows from Equation (2) that $L^* = 4K^*$; conversely, for $a = 0.2$, $L^* = K^*/4$. In the first case, labor's share of total cost is 80 percent and in the second case, it is 20 percent. In a firm whose production function has a larger value of a , labor's share will always be greater, as will the labor–capital ratio. In this example, labor's share of total cost and labor's share of total input are equivalent because the factor prices are equal. If $w \neq r$, labor's share of total cost would not be equal to labor's share of total input, but it would be proportional to it.

Continuing with the case of $a = 0.8$ and $w/r = 1.0$, let $K^* = 10$. Using the cost-minimization condition of Equation (2), it follows that $L^* = 40$. This is shown in the first row of Table 1 as the Baseline case. The corresponding output is 30.314 and, of course, $L^*/K^* = 4$. Now let w increase by 10 percent, so that r/w falls by 10 percent. Because $\sigma = 1.0$ for a Cobb–Douglas production function, L^*/K^* will fall by 10 percent, in this case from $L^*/K^* = 4$ to $L^*/K^* = 3.6$. Using the production function and the necessary relationship between L^* and K^* for cost minimization yields the corresponding new choice point along the original isoquant, $L^* = 39.166$ and $K^* = 10.879$. It can readily be verified that this input bundle yields the original output and satisfies the new necessary labor–capital ratio ($L^*/K^* = 3.6$). The resulting percentage change in L^* , shown in the table, is -2.1 percent. If instead, w decreased by 10 percent, the new cost-minimizing labor–capital ratio would be 4.4, and the corresponding input choices would be $L^* = 40.770$ and $K^* = 9.266$. Now the percentage change in L^* is 1.9 percent. These entries are also shown in the table.

Now look at the bottom portion of the table for the case where labor's share is small ($a = 0.2$). In the Baseline situation, $L^* = 10$ and $K^* = 40$, exactly the reverse of the original case. Output is exactly the same as above. If w increased by 10 percent, r/w would fall by 10 percent and L^*/K^* would fall by 10 percent from .25 to .225. Now, as seen in the table, L^* will decrease from 10 to 9.192 and K^* will increase to 40.852. This is a -8.1 percent change in L^* . Similarly, if w decreased by 10 percent, r/w would increase by 10 percent and the new labor–capital ratio would be .275. The corresponding input choices are 10.792 and 39.245. L^* increases by 7.9 percent.

Interestingly and appropriately, the exact same pattern holds in reverse for the other input. Its proportionate change is large when its share, proxied by $(1-a)$, is small as in Case 1 and smaller when its share is larger, as in Case 2. These impacts are also shown in Table 1.

The arithmetic in the table indeed establishes that S does interact with σ in an unexpected way in determining the elasticity of derived labor demand. As a decreased from 0.8 to 0.2, decreasing labor's share of total cost proportionately, the output-constant and elasticity of substitution-constant employment effect of a 10 percent increase in the wage rate increased from approximately 2 percent to 8 percent. Thus, when S is larger, the cost-minimizing, output-constant response to a given change in the wage rate is smaller in percentage terms. As far as this

TABLE 1. The Impact of Labor's Share on Labor Demand Elasticity via the Elasticity of Substitution

Variable	S	L^*/K^*	L^*	K^*	Q
Case 1: labor's share (large) ^a					
Baseline	0.8	4	40	10	30.314
10% increase in w	0.8	3.6	39.166	10.879	30.314
Percent change after increase		10.0	-2.1	8.8	
10% decrease in w	0.8	4.4	40.770	9.266	30.314
Percent change after decrease		-10.0	1.9	-7.3	
Case 2: labor's share (small) ^b					
Baseline	0.2	0.25	10	40	30.314
10% increase in w	0.2	0.225	9.192	40.852	30.314
Percent change after increase		10.0	-8.1	2.1	
10% decrease in w	0.2	0.275	10.792	39.245	30.314
Percent change after decrease		-10.0	7.9	-1.9	

Note. $L^*/K^* = (a/(1-a)) \times (r/w)$; $r/w = 1$; $\sigma = 1$.

^aProduction function for Case 1 is $Q = L^8K^2$.

^bProduction function for Case 2 is $Q = L^2K^8$.

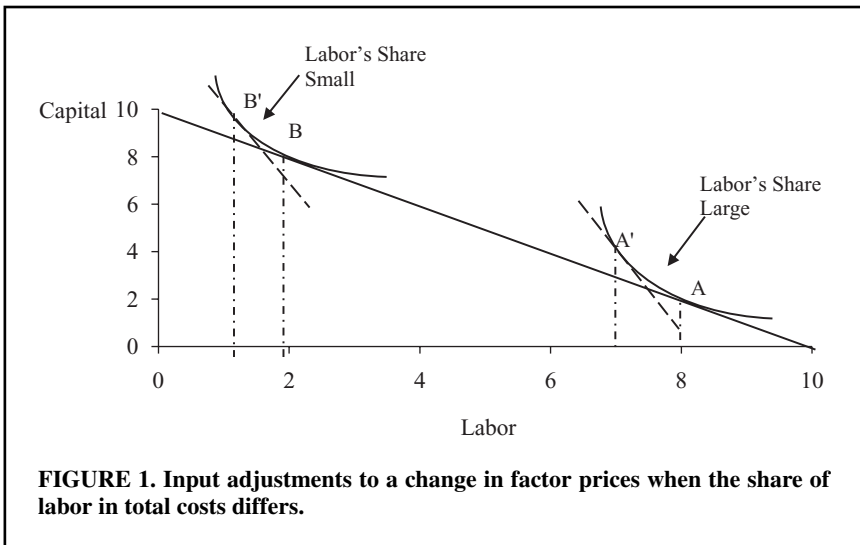
portion of the elasticity formula is concerned, it is definitely important to be important.

Explanation

What is happening to create this unexpected effect? The explanation is something relatively mundane, namely the arithmetic of percentage changes. For a given w and r , labor's share and the labor-capital ratio are monotonically positively related, so that when labor's share is high, so, too, is the labor-capital ratio. This, in turn, necessarily means that the amount of capital is relatively small. Thus, when L and K are adjusted following a wage change so as to make the labor-capital ratio equal the new factor price ratio, a small absolute change in K makes a big impact on the labor-capital ratio. As a result, only a small change in L is required for cost-minimization. And because the amount of labor is large to begin with, this is a small percentage change.

But when labor's share is small, K^* is necessarily large. Now a change in the labor-capital ratio cannot be as easily achieved by changes in K , but require relatively larger changes in L as well. Because L^* was originally small, the percentage change is larger.

Another and perhaps easier way to understand what is happening is to think in terms of the MRTS. Let the two cases correspond to two different firms facing the same factor prices and with the same elasticity of substitution. Because their production functions differ, they choose different input combinations, exactly as



in the example above. Figure 1 illustrates this. In the current example, the MRTS equals one, because $w = r$. Firm 1 chooses Point A, and Firm 2 chooses Point B.

But both firms necessarily have the same MRTS at their optimal choice, because they face the same input prices. Whatever the change in w is, the responses of the two firms will therefore be identical in terms of absolute changes. For example, for a very small change in w in the vicinity of the current choices, $\Delta K^* = -\Delta L^*$. This is approximately true in Table 1. In Case 1, when w increases by 10 percent, L^* falls by 0.834 and K^* increases by 0.879; in Case 2, the corresponding changes are $L^* = -0.818$ and $K^* = 0.852$. Whatever the ΔL^* is, it will always be a larger percentage change when L^* is smaller, which corresponds to the case where labor's share is smaller, and a smaller percentage change when L^* is larger, which is the case where labor's share is large. Thus, Firm 1 moves to A' and Firm 2 moves to B'. The percentage change in L^* is much smaller for Firm 1 (large initial labor share) than for Firm 2 (smaller initial labor share).

Other Considerations

The arithmetic was simplified in these examples by setting $w = r$. But the result would hold for any relationship between w and r . No matter what the factor price ratio is, when labor's share of total cost is greater, so is the labor-capital ratio, holding w/r constant. That is the critical factor. The share of labor in total cost is monotonically related to the labor-capital ratio.

The Cobb-Douglas production function used in the example has constant returns to scale and an elasticity of substitution equal to one. Neither of these properties affects the results here. With other than constant returns to scale, the cost-minimization condition in Equation (1) would have a second parameter b rather than $(1-a)$, but all the arithmetic would go through. Similarly, any other

value for σ would change the absolute size of the factor adjustments, but not the relative percentage changes.

CONCLUSION

The examples I use show that the impact of labor's share of total cost on the elasticity of derived demand does depend negatively on the elasticity of substitution. When labor's share is greater, the cost-minimizing output-constant and the elasticity of substitution-constant responses to a given change in the wage rate are smaller in percentage terms, an effect that reduces the elasticity of labor demand. As far as this portion of the elasticity formula is concerned, it is definitely important to be important. The standard textbook explanation of this odd relationship by Stigler is not correct, because in his example, σ varies with S , a relationship that is certainly plausible, but not in the spirit of the Hicks-Marshall rule.

The unexpected relationship occurs because labor's share is equivalent to or proportional to the labor-capital ratio and because all firms, facing the same factor prices, have the same MRTS at their cost-minimizing choices, even though the choices themselves differ. When the labor-capital ratio is large, the resulting absolute change in the amount of labor necessary to reach cost-minimization following a change in factor prices is a small percentage of the original. When the labor-capital ratio is smaller, the identical absolute change in the amount of labor is a larger percentage of the original. When labor is "important" (labor's share is large), the impact of the elasticity of substitution is attenuated by the "unimportance" of the other input. But when labor is "unimportant," the percentage impact of the elasticity of substitution is greater.

As far as the total impact of labor's share of total cost on the elasticity of labor demand, that depends, exactly as Hicks, Allen, and Bronfenbrenner noted, on whether or not the elasticity of final demand is greater than the elasticity of substitution. If that is true, then an increase in labor's share makes the labor demand curve more elastic. If it does not hold, then an increase in labor's share makes the labor demand curve less elastic. A priori, it is not at all clear whether it is "important to be unimportant" or "important to be important."

NOTES

1. This article was inspired when my graduate students in labor economics asked me for an explanation. Despite having taught labor economics for nearly three decades, I had no explanation whatsoever to offer, and I could not find one in any standard labor economics sources or on a Web search. This provides additional evidence that teaching does give rise to research, as asserted by Becker and Kennedy (2006).
2. Bronfenbrenner (1961) showed that the derivations and explanations of the various authors are not identical, although the resulting rules are. Marshall's rule relating to ease of substitution predated the development of the elasticity of substitution, and thus was not originally stated in those terms.
3. Bronfenbrenner (1961) noted that Allen's equation is a special case of the far more complicated equation from Hicks, corresponding to a situation in which the supply of the other factor of production is perfectly elastic. Analytically, this case might correspond to the demand response of a firm that takes the price of the other input as given or of a competitive industry for which the other factor is not a specialized input. In either case, the supply of the other input is perfectly elastic,

so that changes in input choices will not have effects on the price of the other factor that must be considered.

4. A Google search of the Hicks–Marshall rules uncovers some very nice lecture notes, but no convincing explanations. The account by Hicks (1961) is not very helpful. This particular relationship is possibly the only element of labor demand not explained in Hamermesh (1993, 24–25, n. 2, which refers back to Stigler).
5. Bronfenbrenner (1961, 258) showed that neither Hicks nor Robertson, both of whom offered explanations, got it correct.

REFERENCES

- Allen, R. G. D. 1938. *Mathematical analysis for economists*. London: Macmillan.
- Becker, W. E., and P. Kennedy, 2006. The influence of teaching on research in economics. *Southern Economic Journal* 73 (1): 747–59.
- Borjas, G. 2008. *Labor economics*. Boston: McGraw-Hill Irwin.
- Bronfenbrenner, M. 1961. Notes on the elasticity of derived demand. *Oxford Economic Papers* 13(3): 254–61.
- Ehrenberg, R. G., and R. S. Smith. 2009. *Modern labor economics: Theory and public policy*. Boston: Pearson.
- Hamermesh, D. S. 1993. *Labor demand*. Princeton, NJ: Princeton University Press.
- Hamermesh, D. S., and A. Rees. 1993. *The economics of work and pay*, 5th ed. New York: Harper-Collins.
- Hicks, J. R. 1932. *The theory of wages*. London: Macmillan.
- . 1961. Marshall’s third rule: A further comment. *Oxford Economic Papers* 13(3): 262–65.
- Hyclak, T., G. Johnes, and R. Thornton. 2005. *Fundamentals of labor economics*. Boston: Houghton Mifflin.
- Marshall, A. 1920. *Principles of economics*, 8th ed. London: Macmillan.
- Reynolds, L., S. Masters, and C. Moser. 1998. *Labor economics and labor relations*. Upper Saddle River, NJ: Prentice Hall.
- Stigler, G. J. 1966. *The theory of price*. New York: Macmillan.

Copyright of Journal of Economic Education is the property of Taylor & Francis Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.