Supplemental Material: Annu. Rev. Econ. 2022. 14:93-124
https://doi.org/10.1146/annurev-economics-080921-103046
The Elusive Explanation for the Declining Labor Share
Grossman & Oberfield

A Estimating the Elasticity of Substitution

The elasticity of substitution between capital and labor figures prominently in the literature on the labor share. Many qualitative results hinge on whether the elasticity exceeds, equals, or falls short of unity. Nearly all quantitative exercises that purport to explain the decline in the labor share rely on an estimate of σ. But this parameter is notoriously difficult to estimate. In this appendix, we provide a brief critical discussion of some of the methodological issues that arise in estimating σ.

The central identification issue is articulated most clearly by Diamond and McFadden (1965), who show that it is impossible to separately identify the elasticity of substitution from the factor bias of technical change in smooth, time-series data. Their observation is analogous to the classic point that one cannot estimate the slope of a demand curve by simply observing changes in prices and quantities, inasmuch as the same data might be generated by shifts in the demand curve or shifts in the supply curve. To estimate the slope of a demand curve, one needs an instrument for supply. Similarly, to estimate the elasticity of substitution, which reveals how relative input demands (K/L) respond to changes in relative prices (w/r) requires an instrument for relative factor supplies.

To make this concrete, consider the CES production function,

$$Y_t = \left[(A_t K_t)^{\frac{\sigma-1}{\sigma}} + (B_t L_t)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}.$$

Minimizing the dual cost function requires $\frac{\theta_t}{1-\theta_t} = \frac{w_t L_t}{r_t K_t} = \left(\frac{w_t/B_t}{r_t/A_t} \right)^{1-\sigma}$, where θ_t is the labor share. Now consider the regression of relative factor shares on relative prices,

$$\ln \frac{\theta_t}{1-\theta_t} = (1-\sigma) \ln \frac{w_t}{r_t} + \varepsilon_t.$$

Here, the error term reflects the relative factor productivities, $\varepsilon_t = (\sigma - 1) \ln (B_t/A_t)$. This term is bound to be correlated with the regressor if productivity shocks that alter relative factor demands for given factor prices are correlated with equilibrium factor prices.[1] To estimate σ from this regression, one would need an instrument for w_t/r_t that is orthogonal to B_t/A_t.

In practice, three methods have been used to estimate σ. The most common approach uses the aggregate time series of input usage and factor prices. Typically, researchers assume that the bias

[1] Along a balanced growth path, factor shares are constant, so that relative factor price movements completely offset changes in relative productivity.

in technical change follows a linear time trend and thus they include time t in the regression to proxy for the omitted variable, $\ln(B_t/A_t)$. Then they compute σ from the regression[2]

$$\ln\frac{K_t}{L_t} = \sigma \ln\frac{w_t}{r_t} + \phi t + \eta_t \tag{1}$$

Researchers that have taken this approach to estimate σ include Antràs (2004), Klump et al. (2007), León-Ledesma (2010), McAdam and Willman (2010), Mallick (2012), Lawrence (2015), Herrendorf et al. (2015), and Alvarez-Cuadrado et al. (2018). They typically find capital and labor to be complements in production, i.e., $\sigma < 1$.[3]

However, replacing B_t/A_t by a function of time does not solve the endogeneity problem, it simply assumes it away. The Diamond-McFadden Impossibility Theorem tells us that assumptions about the factor bias of technical change cannot be tested unless we already know σ, which clearly we do not when we are trying to estimate it.

A further concern about (1) is that the time trend might soak up all the long-run movement in factor demands and factor prices, leaving an estimate of σ that reflects only short-run responses to transitory movements in factor prices. Samuelson (1947) suggested with his Le Chatelier's principle that long-run responsiveness to a permanent change might exceed the short-run response to a transitory change, so estimates of σ from (1) might understate the true, long-run elasticity of substitution.[4]

A second approach, adopted by Karabarbounis and Neiman (2014), uses long-run variation in a cross-section of countries. They use the first-order condition for optimal capital usage, $1-\theta_t = \left(\frac{r_t/A_t}{p_t}\right)^{1-\sigma}$, to derive the regression equation

$$\Delta\ln(1-\theta_t) = (1-\sigma)\,\Delta\ln\frac{r_t}{p_t} + \varepsilon_t, \tag{2}$$

where $\varepsilon_t \equiv (\sigma-1)\,A_t$. Then they estimate σ using long differences in factor shares and in the price of investment relative to output in a sample of countries. To deal with the potential endogeneity of relative prices, Karabarbounis and Neiman assume either that technical change is Harrod neutral, so that A_t is constant, or that it is Hicks neutral, so that A_t can be measured directly by the Solow residual. In either case, they find an elasticity above one.

Two criticisms can be levelled at their approach. First, they too have not meaningfully addressed the endogeneity issue, but rather have swept it away. Many of the proposed reasons for the decline in the labor share do not imply changes in factor demand that are either Hicks-neutral or Harrod-neutral in form. Second, as Glover and Short (2020) argue, even if one accepts that productivity growth is Harrod-neutral, the imputed rental rate for capital, r_t in (2), should incorporate changes in

[2] There are more sophisticated versions of this approach that allow for more flexible (but still imposed) evolution of relative factor productivity over time. León-Ledesma et. al (2010), suggest using all of the first-order conditions for the optimal inputs of the various factors, not just the ratio of the first-order conditions for labor and capital.

[3] See Raval (2018) and Knoblach et al. (2021) for recent meta analyses of estimates.

[4] Chirinko and Mallick (2008) use only low-frequency variation in relative factor prices in their analysis, and so they are exempt from this critique. They find an industry-level elasticity of substitution of approximately 0.4.

the real interest rate. Moreover, they show that the Karabarbounis-Neiman estimates are sensitive to the criteria for data selection. In their estimation using similar techniques, but allowing for alternative data selection or adjustments for the real interest rate, Glover and Short estimate a value of σ that is slightly smaller than one.

A third approach attempts to compute an economy-wide elasticity of substitution by aggregating estimates derived using industry or firm-level data. This approach recognizes that macro-level instruments are difficult to come by, whereas researchers often can find variation in factor prices that arguably is orthogonal to individual plants' production technologies. For example, Raval (2019) uses shifts in labor supply facing manufacturing plants in a location that reflect changes in local amenities, as well as a shift-share variable that captures local impacts of national changes in non-manufacturing industries as instruments in his regressions. He finds long-run plant-level elasticities of substitution in the range from 0.3 to 0.5.

However, Houthakker (1957-58) pointed out long ago that a macro-level elasticity of substitution can differ substantially from the underlying micro elasticities. Oberfield and Raval (2021) show how to combine an estimate of the micro elasticity of capital-labor substitution with the distribution of capital shares across plants to compute a macro elasticity. In the simplest version, the aggregate elasticity is a weighted average of the micro elasticity of substitution and the elasticity of demand facing individual plants, with the relative importance of each depending on a statistic that is proportional to a cost-weighted variance of capital shares across plants. That is, in response to an increase in the wage, individual plants will substitute toward capital and capital-intensive plants will grow relative to labor-intensive plants, and the extent of heterogeneity in capital intensities determines the relative importance of each adjustment margin. One potential criticism of their approach is that it relies more heavily on the structure of the model than other approaches, and is therefore more susceptible to model uncertainty. However, it turns out that when capital shares do not vary too much across plants, as is the case in the U.S. manufacturing sector, estimates of the sector-level elasticity using different models fall within a fairly narrow range, mostly between 0.5 and 0.7.

As difficult as it may be to identify the elasticity of substitution between a homogeneous labor input and a homogeneous capital input, the challenges multiply when labor and capital are heterogeneous inputs or when there are variable returns to scale. First, there is a question of definition: With more than two inputs, there are many ways to define an elasticity of substitution between any two of them, depending on what is being held constant (e.g., prices of other inputs, quantities of other inputs, quantities of output, which of the two input prices changes).[5] There are some special cases where the elasticity of substitution between capital and labor remains a well-defined concept. For example, the set of capital inputs might be separable in the production function from the set of labor inputs, i.e., we might be able to write $F(K_1, ..., K_J, L_1, ..., L_I)$ as $\tilde{F}(G(K_1, ..., K_J), H(L_1, ..., L_I))$. But such a specification of technology is fairly restrictive. More generally, two natural ways to define the capital-labor elasticity might be in terms of the change

[5] See Morishima (1967) or Blackorby and Russell (1981).

in total compensation of capital relative to total compensation of labor in response to proportional change in quantities of all types of capital relative to quantities of all types of labor, or its dual, the change in relative compensation when the prices of all types of capital change equiproportionately and so do the prices of all types of labor.[6] Such a definition can be useful, but it limits the set of questions that might be addressed, for example it would not help with understanding the effects of a fall in the price of robots relative to that of structures.

In circumstances where labor or capital is heterogeneous and relative prices within a grouping change, it becomes difficult to infer a capital-labor substitution elasticity from changes in the quantity of one particular input. Two examples from the literature illustrate this point.

Grossman et al. (2021) posit a production function in which labor hours L and labor skill s enter the production function differently, so that the technology can be represented as $F(K, L; s)$. For example, s could be the fraction of individuals that have a college degree or the educational attainment of the representative worker. The authors focus on a setting in which capital and skill are complementary, i.e., $\varphi \equiv \frac{d \ln F_s/F_L}{d \ln K} > 0$, which means that there is no separability between capital and labor in the aggregate production function. In such circumstances, an exogenous decline in the cost of capital would alter the capital share not only directly, but also indirectly by inducing a change in s, such that

$$\frac{d \log(1-\theta)}{d \ln R} = (1-\sigma) + \sigma \frac{\varphi}{\theta} \frac{F_s}{F} \frac{ds}{d \ln R},$$

where σ is defined as the elasticity of substitution between capital and hours, holding s constant. As a result, an estimate of σ from a regression of the capital share $(1-\theta)$ on R would be upwardly biased if it does not take into account the induced change in skills. Karabarbounis and Neiman (2014) and Glover and Short (2020) do not control for workers' human capital in their regressions, which may have contributed to their finding of a larger elasticity than what Oberfield and Raval (2021) estimate when they do control for changes in the skill premium.

Meanwhile, Hubmer (2020) emphasizes the fact that equipment prices have fallen faster than the price of structures. Industries that use equipment capital more intensively faced a larger decline in their weighted average cost of capital. He shows that those that used equipment more intensively relative to structures saw larger increases in their capital shares, leading him to conclude that, on average, the industry-level elasticity of capital-labor substitution is greater than one.

[6] With a constant returns to scale production with two inputs, $F(K, L)$ with unit cost function $c(r, w)$, the elasticity of substitution satisfies $1 - \frac{1}{\sigma} = \frac{d \ln \frac{F_K(K,L)K}{F_L(K,L)L}}{d \ln K/L}$ and $1 - \sigma = \frac{d \ln \frac{r c_r}{w c_w}}{d \ln r/w}$. With multiple inputs, with production function $F(K_1, ..., K_J, L_1, ..., L_I)$ and its unit cost function $c(r_1, ..., r_J, w_1, ..., w_I)$, one could similarly define

$$1 - \frac{1}{\sigma} = \left. \frac{d \ln \frac{\sum_j F_{Kj}(tK_1, ..., tK_J, L_1, ..., L_I) tK_j}{\sum_i F_{Li}(tK_1, ..., tK_J, L_1, ..., L_I) L_i}}{d \ln t} \right|_{t=1}$$

or its dual

$$1 - \sigma = \left. \frac{d \ln \frac{\sum_j tr_j c_{r_j}(tr_1, ..., tr_J, w_1, ..., w_I)}{\sum_i w_i c_{w_i}(tr_1, ..., tr_J, w_1, ..., w_I)}}{d \ln t} \right|_{t=1}.$$

These are not generically the same.

Such a conclusion would be warranted if capital inputs are separable from labor inputs, i.e., if industry production functions take the form $Y_i = F_i\left(G_i\left(K^e, K^s\right), L\right)$. Outside of this case, such a regression would not reveal whether capital and labor are complements or substitutes (if we define complementarity to mean that an increase in the wage rate raises the labor share in revenues holding the rental price of each type of capital constant). The implications of such a regression for whether capital and labor are complements or substitutes is even less clear if, for example, equipment is complementary to skilled labor but a substitute for unskilled labor.

References

Alvarez-Cuadrado F, Van Long N, Poschke M. 2018. Capital-labor substitution, structural change and the labor income share. *J. Econ. Dyn. Control* 87:206-31

Antràs P, 2004. Is the US aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution. *Contributions in Macroecon.* 4(1)

Blackorby C., Russell, R. (1981). The Morishima elasticity of substitution; symmetry, constancy, separability, and its relationship to the Hicks and Allen elasticities. The Review of Economic Studies, 48(1):147-58

Chirinko RS, Mallick D. 2008. *The marginal product of capital: A persistent international puzzle.* Unpublished manuscript, Univ. Ill., Chi

Diamond PA, McFadden D. 1965. *Identification of the elasticity of substitution and the bias of technical change: An impossibility theorem.* Unpublished manuscript, Univ. Calif., Berkeley

Glover A, Short J. 2020. Can capital deepening explain the global decline in labor's share? *Rev. of Econ. Dyn.* 35:35–53

Grossman GM, Helpman E, Oberfield E, Sampson T. 2021. Endogenous education and long-run factor shares. *Am. Econ. Rev. Insights* 3(2):215–232

Herrendorf B, Herrington C, Valentinyi A. 2015. Sectoral technology and structural transformation. *Am. Econ. J. Macroecon.* 7(4):104–133

Houthakker HS. 1955. The Pareto distribution and the Cobb-Douglas production function in activity analysis. *Rev. Econ. Stud.* 23(1):27-31

Hubmer J. 2020. *The race between preferences and technology.* Unpublished manuscript, Univ. Pa.

Karabarbounis L, Neiman B. 2014. The global decline of the labor share. *Q. J. Econ.* 129(1):61–103

Klump R, McAdam P, Willman A. 2007. Factor substitution and factor-augmenting technical progress in the United States: A normalized supply-side system approach. *Rev. Econ. Stat.* 89(1):183-92

Knoblach, M., Roessler, M., & Zwerschke, P. (2020). The elasticity of substitution between capital and labour in the US economy: A meta-regression analysis. Oxford Bull. of Econ. and Stat. 82(1), 62-82.

Lawrence RZ. 2015. *Recent declines in labor's share in US income: a preliminary neoclassical account.* NBER Work. Pap. 21296

León-Ledesma MA, McAdam P, Willman A. 2010. Identifying the elasticity of substitution with biased technical change. *Am. Econ. Rev.* 100(4):1330-57

Mallick D. 2012. The role of the elasticity of substitution in economic growth: A cross-country investigation. *Labour Econ.* 19(5):682-94

Morishima, M. (1967). A few suggestions on the theory of elasticity. *Keizai Hyoron (Economic Revie*w*)*, 16, 144-150.

Oberfield E, Raval D. 2021. Micro data and macro technology. *Econometrica* 89(2):703–732

Raval, D. (2018). What's wrong with capital in the twenty-first century's model?. In *After Piketty* (pp. 75-98). Harvard University Press

Raval D. 2019. The micro elasticity of substitution and non-neutral technology. *RAND J. Econ.* 50(1):147–167

Samuelson P. 1947. *Foundations of economic analysis*. Cambridge, MA: Harvard Univ. Press