Micro Data and Macro Technology*

Ezra Oberfield

Princeton University

edo@princeton.edu

Devesh Raval

Federal Trade Commission

devesh.raval@gmail.com

August 7, 2020

Abstract

We develop a framework to estimate the aggregate capital-labor elasticity of substitution by aggregating the actions of individual plants. The aggregate elasticity reflects substitution within plants and reallocation across plants; the extent of heterogeneity in capital intensities determines their relative importance. We use micro data on the cross-section of plants to build up to the aggregate elasticity at a point in time. Interpreting our econometric estimates through the lens of several different models, we find that the aggregate elasticity for the US manufacturing sector is in the range of 0.5-0.7, and has declined slightly since 1970. We use our estimates to measure the bias of technical change and assess the decline in labor's share of income in the US manufacturing sector. Mechanisms that rely on changes in the relative supply of factors, such as an acceleration of capital accumulation, cannot account for the decline.

KEYWORDS: Elasticity of Substitution, Aggregation, Labor Share, Bias of Technical Change.

*First version: January 2012. We are grateful for the guidance of Daron Acemoglu and four anonymous referees. We also thank Enghin Atalay, Bob Barsky, Ehsan Ebrahimy, Ali Hortacsu, Konstantin Golyaev, Michal Kolesar, Sam Kortum, Lee Lockwood, Eduardo Morales, Harry J. Paarsch, Richard Rogerson, Chad Syverson, Hiroatsu Tanaka, and Nate Wilson for their comments on this paper, and Loukas Karabarbounis for his comments as a discussant. We are grateful to Frank Limehouse for all of his help at the Chicago Census Research Data Center and to Jon Samuels and Dale Jorgenson for providing us with capital tax information and estimates. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau or the Federal Trade Commission and its Commissioners. All results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

Over the last several decades, labor's share of income in the US manufacturing sector has fallen by more than 15 percentage points, and its share of aggregate income has declined by 8 percentage points. A variety of mechanisms have been proposed to explain declining labor shares. Some have proposed that the decline originated in movements in relative supplies of capital and labor.[1] Others have proposed that the change reflects a decline in demand for labor relative to capital, due to changes in the pace of automation or offshoring.

The aggregate capital-labor elasticity of substitution is a crucial input into assessing the relevance of these mechanisms, as Hicks (1932) first pointed out. Armed with this elasticity, one can measure the shift in demand for labor relative to capital and its contribution to the labor share decline, or assess the contribution of any change in relative supply.

Unfortunately, obtaining the elasticity is difficult; Diamond et al. (1978) proved that the elasticity cannot be identified from time series data on output, inputs, and marginal products alone. Instead, identification requires factor price movements that are independent of the bias of technical change.[2]

One approach to the Diamond et al. (1978) critique is to place parametric assumptions on the bias of technical change to identify the elasticity with aggregate time series data. The most common assumptions have been that there has been no bias or a constant bias over time. Leon-Ledesma et al. (2010) demonstrated that, even under such assumptions, it is difficult to obtain the true aggregate elasticity. Not surprisingly, estimates using this approach range widely.[3]

The second approach uses micro production data with more plausibly exogenous variation

[1] Piketty (2014) maintained that declining labor shares resulted from increased capital accumulation, and Karabarbounis and Neiman (2014) argued that they stem from investment-specific technical change which also increased the supply of capital.

[2] This is one incarnation of the familiar point that to estimate an elasticity of demand (in this case, relative factor demand), one needs an instrument for supply.

[3] Although Berndt (1976) found a unitary elasticity of substitution in the US time series assuming neutral technical change, Antras (2004) and Klump et al. (2007) subsequently found estimates from 0.5 to 0.9 allowing for biased technical change. Karabarbounis and Neiman (2014) estimate an aggregate elasticity of 1.25 using cross-country panel variation in capital prices. Piketty (2014) estimates an aggregate elasticity between 1.3 and 1.6. Herrendorf et al. (2015) estimate an elasticity of 0.84 for the US economy as a whole and 0.80 for the manufacturing sector. Alvarez-Cuadrado et al. (2018) estimate an elasticity of 0.78 for the manufacturing sector. See Chirinko (2008) and Raval (2017) for a meta-analysis of the estimates.

in factor prices, and yields the micro capital-labor elasticity of substitution. Houthakker (1955), however, famously showed that the micro and macro elasticities can be very different; an economy of Leontief micro units can have a Cobb–Douglas aggregate production function.[4] Given Houthakker's result, it is unclear whether the micro elasticity can help answer the many questions that hinge on the aggregate elasticity.

In this paper, we show how the aggregate elasticity of substitution can be recovered from the plant-level elasticity. Building on Sato (1975), the aggregate elasticity is a convex combination of the plant-level elasticity of substitution and the elasticity of demand in our baseline model.[5] In response to a wage increase, plants substitute towards capital. In addition, capital-intensive plants gain market share from labor-intensive plants. The degree of heterogeneity in capital intensities determines the relative importance of within-plant substitution and reallocation. For example, when all plants produce with the same capital intensity, there is no reallocation of resources across plants. In addition, the increased wage could cause the entry of capital-intensive plants and exit of labor-intensive plants.

Using this framework, we build the aggregate capital-labor elasticity for the US manufacturing sector from its individual components. We estimate micro production and demand parameters. Since Levhari (1968), it has been well known that Houthakker's result of a unitary elasticity of substitution is sensitive to the distribution of capital intensities. Rather than making distributional assumptions, we directly measure the empirical distribution using cross-sectional micro data.

Thus, given the set of plants that existed at a point in time, we estimate the aggregate elasticity of substitution for the manufacturing sector *at that time*.[6] Our strategy allows both the aggregate elasticity and the bias of technical change to vary freely over time, opening up a new set of questions. Because our identification does not impose strong parametric assumptions on the time path of the bias, our approach is well suited for measuring how

[4]Houthakker (1955) assumed that factor-augmenting productivities follow independent Pareto distributions. The connection between Pareto distributions and a Cobb-Douglas aggregate production function is also emphasized in Jones (2005), Lagos (2006), Luttmer (2012), Mangin (2017), and Boehm and Oberfield (2018).

[5]Sato (1975) showed this for a two-good economy. See also Miyagiwa and Papageorgiou (2007) and Rognlie (2014).

[6]To avoid tedious repetition, we occasionally use the term "aggregate elasticity" to refer to the aggregate elasticity for the manufacturing sector.

the bias has varied over time and how it has contributed to the decline in labor's share of income. We can also examine how the aggregate elasticity has changed over time.

In Section 3 we estimate the aggregate elasticity for the US manufacturing sector using the US Census of Manufactures. As in Raval (2019), We use the estimates of cross-sectional differences in local wages to estimate plant-level elasticities of substitution. To account for the potential endogeneity of the local wage to technological differences, we employ three different instruments for the supply of labor facing manufacturing plants. These instruments include a set of climate-based amenities as well as shift-share variables from Bartik (1991) and Beaudry et al. (2012). Across years and specifications, estimates of the plant-level elasticity of substitution lie roughly between 0.3 and 0.5. Given the heterogeneity in capital shares and our estimates of other parameters, the aggregate elasticity for the manufacturing sector is between 0.5 and 0.7. Reallocation thus accounts for roughly one-third of substitution. Examining changes over time, we find that the aggregate elasticity has fallen slightly from 1972 to 2007.

In Section 4, we incorporate additional channels of adjustment to factor price changes. We show that, in an environment that incorporates entry and exit, our baseline estimate is an upper bound for the aggregate elasticity: our estimates based on cross-sectional variation reflect how the local average of factor shares responds to local factor prices, and thus capture both within-plant substitution and entry and exit. We can also derive a lower bound using dynamic panel estimates which capture only within-plant substitution. Given our estimates, the range between our lower and upper bound are relatively small.

We next allow for shifts in the technological frontier, and show the elasticity that incorporates the change in the frontier is between our baseline estimate and one. We also consider several additional margins of adjustment, including adjustment costs and substitution to intangible capital, and examine the implications of these forces for the aggregate elasticity for the manufacturing sector.

In Section 5, we use our estimate of the aggregate elasticity for the manufacturing sector to decompose the decline in labor's share of income in that sector since 1970. The bias of technical change contributes to a decline in the labor share of 20 percentage points from 1970 to 2010, with an ongoing decline that accelerates starting in 2000. The rising cost of labor relative to capital did not fully offset the bias. In contrast to the acceleration of the bias after 2000, the contribution of factor prices to the labor share exhibited little variation over time, and does not match the accelerating decline of the labor share.

2 Theory

This section characterizes the aggregate elasticity of substitution between capital and labor in terms of production and demand elasticities of individual plants. We begin with a simplified environment in which we describe the basic mechanism and intuition, and then enrich the model by incorporating materials and allowing for heterogeneity across industries. We then incorporate entry and exit. A number of features—shifts in the technological frontier, adjustment costs, wedges, and intangible capital—are omitted from our benchmark model. We postpone a discussion of these until Section 4.

2.1 Simple Example

Consider a large set of plants I whose production functions share a common, constant elasticity of substitution between capital and labor, σ. A plant produces output Y_i from capital K_i and labor L_i using the following CES production function:

$$Y_i = \left[(A_i K_i)^{\frac{\sigma-1}{\sigma}} + (B_i L_i)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{1}$$

Productivity differences among plants are factor augmenting: A_i is i's capital-augmenting productivity and B_i i's labor-augmenting productivity.

Consumers have Dixit–Stiglitz preference across goods, consuming the bundle $Y = \left(\sum_{i\in I} D_i^{\frac{1}{\varepsilon}} Y_i^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}$. Plants are monopolistically competitive, so each plant faces an isoelastic demand curve with a common elasticity of demand $\varepsilon > 1$.

Among these plants, aggregate demand for capital and labor are defined as $K \equiv \sum_{i\in I} K_i$ and $L \equiv \sum_{i\in I} L_i$ respectively. We define the aggregate elasticity of substitution, σ^{agg}, to be the partial equilibrium response of the aggregate capital-labor ratio, K/L, to a change in relative factor prices, w/r:[7]

$$\sigma^{agg} \equiv \frac{\mathrm{d}\ln K/L}{\mathrm{d}\ln w/r} \tag{2}$$

We neither impose nor derive a parametric form for an aggregate production function. Given the allocation of capital and labor, σ^{agg} simply summarizes, to a first order, how a change in factor prices would affect the aggregate capital-labor ratio.

Let $\alpha_i \equiv \frac{rK_i}{rK_i+wL_i}$ and $\alpha \equiv \frac{rK}{rK+wL}$ denote the cost shares of capital for plant i and in aggregate. The plant-level and aggregate elasticities of substitution are closely related to the changes in these capital shares:

$$\sigma - 1 = \frac{\mathrm{d}\ln rK_i/wL_i}{\mathrm{d}\ln w/r} = \frac{\mathrm{d}\ln \alpha_i/(1-\alpha_i)}{\mathrm{d}\ln w/r} = \frac{1}{\alpha_i(1-\alpha_i)}\frac{\mathrm{d}\alpha_i}{\mathrm{d}\ln w/r} \tag{3}$$

$$\sigma^{agg} - 1 = \frac{\mathrm{d}\ln rK/wL}{\mathrm{d}\ln w/r} = \frac{\mathrm{d}\ln \alpha/(1-\alpha)}{\mathrm{d}\ln w/r} = \frac{1}{\alpha(1-\alpha)}\frac{\mathrm{d}\alpha}{\mathrm{d}\ln w/r} \tag{4}$$

The aggregate cost share of capital can be expressed as an average of plant capital shares, weighted by size:

$$\alpha = \sum_{i\in I} \alpha_i \theta_i \tag{5}$$

where $\theta_i \equiv \frac{rK_i+wL_i}{rK+wL}$ denotes plant i's expenditure on capital and labor as a fraction of the aggregate expenditure. To find the aggregate elasticity of substitution, we can simply differentiate (5):

$$\frac{\mathrm{d}\alpha}{\mathrm{d}\ln w/r} = \sum_{i\in I} \frac{\mathrm{d}\alpha_i}{\mathrm{d}\ln w/r}\theta_i + \sum_{i\in I} \alpha_i \frac{\mathrm{d}\theta_i}{\mathrm{d}\ln w/r}$$

Using equations (3) and (4), this can be written as

$$\sigma^{agg} - 1 = \frac{1}{\alpha(1-\alpha)} \sum_{i\in I} \alpha_i(1-\alpha_i)(\sigma-1)\theta_i + \frac{1}{\alpha(1-\alpha)} \sum_{i\in I} \alpha_i\theta_i \frac{\mathrm{d}\ln\theta_i}{\mathrm{d}\ln w/r} \tag{6}$$

The first term on the right hand side is a substitution effect that captures the change in

[7] Since production and demand are homogeneous of degree one, a change in total spending would not alter the aggregate capital-labor ratio. We address non-homothetic environments in Web Appendix G.2.4.

factor intensity holding fixed each plant's size, θ_i. The plant-level substitution elasticity σ measures how much an individual plant changes its mix of capital and labor in response to changes in factor prices. The second term is a reallocation effect that captures how plants' sizes change with relative factor prices. By Shephard's Lemma, a plant's cost share of capital α_i measures how relative factor prices affect its marginal cost. When wages rise, plants that use capital more intensively gain a relative cost advantage. Consumers respond to the subsequent changes in relative prices by shifting consumption toward the capital intensive goods. This reallocation effect is larger when demand is more elastic, because customers respond more to changing relative prices. Formally, the change in i's relative expenditure on capital and labor is

$$\frac{\mathrm{d}\ln\theta_i}{\mathrm{d}\ln w/r} = (\varepsilon-1)(\alpha_i-\alpha) \tag{7}$$

After some manipulation (see Appendix A for details), the industry elasticity of substitution is a convex combination of the micro elasticity of substitution and elasticity of demand:

$$\sigma^{agg} = (1-\chi)\sigma + \chi\varepsilon \tag{8}$$

where $\chi \equiv \sum_{i\in I}\frac{(\alpha_i-\alpha)^2}{\alpha(1-\alpha)}\theta_i$.

The first term, $(1-\chi)\sigma$, measures substitution between capital and labor within plants. The second term, $\chi\varepsilon$, captures reallocation between capital- and labor-intensive plants.

We call χ the heterogeneity index. It is proportional to the cost-weighted variance of capital shares and lies between zero and one.[8] When each plant produces at the same capital intensity, χ is zero and there is no reallocation across plants. Each plant's marginal cost responds to factor price changes in the same way, so relative output prices are unchanged. In contrast, if some plants produce using only capital while all others produce using only labor, all factor substitution is across plants and χ is one. When there is little variation in capital intensities, within-plant substitution is more important than reallocation.

[8]A simple proof: $\sum_{i\in I}(\alpha_i-\alpha)^2\theta_i = \sum_{i\in I}\alpha_i^2\theta_i - \alpha^2 \leq \sum_{i\in I}\alpha_i\theta_i - \alpha^2 = \alpha - \alpha^2 = \alpha(1-\alpha)$. It follows that $\chi = 1$ if and only if each plant uses only capital or only labor (i.e., for each i, $\alpha_i \in \{0,1\}$).

2.2 Baseline Model

This section describes the baseline model we will use in our empirical implementation. The baseline model extends the previous analysis by allowing for heterogeneity across industries and using a production structure in which plants use materials in addition to capital and labor.

Let N be the set of industries and I_n be the set of plants in industry n. We assume that each plant's production function has a nested CES structure.

Assumption 1 *Plant i in industry n produces with the production function*

$$F_{ni}\left(K_{ni}, L_{ni}, M_{ni}\right) = \left(\left[\left(A_{ni}K_{ni}\right)^{\frac{\sigma_n-1}{\sigma_n}} + \left(B_{ni}L_{ni}\right)^{\frac{\sigma_n-1}{\sigma_n}}\right]^{\frac{\sigma_n}{\sigma_n-1}\frac{\zeta_n-1}{\zeta_n}} + \left(C_{ni}M_{ni}\right)^{\frac{\zeta_n-1}{\zeta_n}}\right)^{\frac{\zeta_n}{\zeta_n-1}} \tag{9}$$

so its elasticity of substitution between capital and labor is σ_n. Further, i's elasticity of substitution between materials and its capital-labor bundle is ζ_n.

We also assume that demand has a nested structure with a constant elasticity at each level of aggregation. Such a structure is consistent with a representative consumer whose preferences exhibit constant elasticities of substitution across industries and across varieties within each industry:

$$Y \equiv \left[\sum_{n\in N} D_n^{\frac{1}{\eta}} Y_n^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}, \qquad Y_n \equiv \left(\sum_{i\in I_n} D_{ni}^{\frac{1}{\varepsilon_n}} Y_{ni}^{\frac{\varepsilon_n-1}{\varepsilon_n}}\right)^{\frac{\varepsilon_n}{\varepsilon_n-1}} \tag{10}$$

This demand structure implies that each plant in industry n faces a demand curve with constant elasticity ε_n. Letting q be the price of materials, each plant maximizes profit

$$\max_{P_{ni},Y_{ni},K_{ni},L_{ni},M_{ni}} P_{ni}Y_{ni} - rK_{ni} - wL_{ni} - qM_{ni}$$

subject to the technological constraint $Y_{ni} = F_{ni}\left(K_{ni}, L_{ni}, M_{ni}\right)$ and the demand curve $Y_{ni} = Y_n(P_{ni}/P_n)^{-\varepsilon_n}$, where $P_n \equiv \left(\sum_{i\in I_n} D_{ni}P_{ni}^{1-\varepsilon_n}\right)^{\frac{1}{1-\varepsilon_n}}$ is the price index for industry n.

The industry-level elasticity of substitution between capital and labor for industry n measures the response of the industry's capital-labor ratio to a change in relative factor prices:

$$\sigma_n^N \equiv \frac{\mathrm{d}\ln K_n/L_n}{\mathrm{d}\ln w/r}$$

The derivation of this industry elasticity of substitution follows Section 2.1 up to (6). As before, $\alpha_{ni} = \frac{rK_{ni}}{rK_{ni}+wL_{ni}}$ is plant i's capital share of non-materials cost and $\theta_{ni} = \frac{rK_{ni}+wL_{ni}}{rK_n+wL_n}$ plant i's share of industry n's expenditure on capital and labor. We will show that reallocation depends on plants' expenditures on materials. We denote plant i's materials share of its total cost as $s_{ni}^M \equiv \frac{qM_{ni}}{rK_{ni}+wL_{ni}+qM_{ni}}$. Because producers of intermediate inputs use capital and labor, changes in r and w would affect the price of materials. We define $\alpha^M \equiv \frac{\mathrm{d}\ln q/w}{\mathrm{d}\ln r/w}$ to be the capital content of materials.

Proposition 1 *Under Assumption 1, the industry elasticity of substitution is:*

$$\sigma_n^N = (1-\chi_n)\sigma_n + \chi_n\left[(1-\bar{s}_n^M)\varepsilon_n + \bar{s}_n^M\zeta_n\right] \tag{11}$$

where $\chi_n = \sum_{i\in I_n}\frac{(\alpha_{ni}-\alpha_n)^2}{\alpha_n(1-\alpha_n)}\theta_{ni}$ *and* $\bar{s}_n^M = \frac{\sum_{i\in I_n}(\alpha_{ni}-\alpha_n)(\alpha_{ni}-\alpha^M)\theta_{ni}s_{ni}^M}{\sum_{i\in I_n}(\alpha_{ni}-\alpha_n)(\alpha_{ni}-\alpha^M)\theta_{ni}}$

The proofs of all propositions are contained in Appendix A.

Relative to (8), the demand elasticity is replaced by a convex combination of the elasticity of demand, ε_n, and the elasticity of substitution between materials and the capital-labor bundle, ζ_n. This composite term measures the change in i's share of its industry's expenditure on capital and labor, θ_{ni}. Intuitively, a plant's expenditure on capital and labor can fall because its overall scale declines or because it substitutes towards materials. The cost share of materials determines the relative importance of each. As materials shares approach zero, all shifts in composition are due to changes in scale, and Proposition 1 reduces to (8). In contrast, as a plant's materials share approaches one, changes in its cost of capital and labor have a negligible impact on its marginal cost, and hence a negligible impact on its sales. Rather, the change in its expenditure on capital and labor is determined by substitution between materials and the capital-labor bundle.

If we dispense with the assumption that production functions take the nested CES functional form of Assumption 1, and maintain only that they exhibit constant returns to scale, the resulting expression for the industry elasticity of substitution is identical to (11) except that the parameters σ_n and ζ_n are replaced with weighted averages of the plants' local elasticities, $\bar{\sigma}_n \equiv \sum_{i\in I_n} \frac{\alpha_{ni}(1-\alpha_{ni})\theta_{ni}}{\sum_{i'\in I_n}\alpha_{ni'}(1-\alpha_{ni'})\theta_{ni'}}\sigma_{ni}$ and $\bar{\zeta}_n \equiv \sum_{i\in I_n} \frac{(\alpha_{ni}-\alpha_n)(\alpha_{ni}-\alpha^M)s^M_{ni}\theta_{ni}}{\sum_{i'\in I_n}(\alpha_{ni'}-\alpha_n)(\alpha_{ni'}-\alpha^M)s^M_{ni'}\theta_{ni'}}\zeta_{ni}$.[9]

The aggregate elasticity parallels the industry elasticity; aggregate capital-labor substitution consists of substitution within industries and reallocation across industries. Proposition 2 shows that expression for the aggregate elasticity parallels the expressions for the industry elasticity in Proposition 1 with plant and industry variables replaced by industry and aggregate variables respectively.

Proposition 2 *The aggregate elasticity between capital and labor, $\sigma^{agg} = \frac{d\ln K/L}{d\ln w/r}$, is:*

$$\sigma^{agg} = (1-\chi^{agg})\,\bar{\sigma}^N + \chi^{agg}\left[(1-\bar{s}^M)\eta + \bar{s}^M\bar{\zeta}^N\right] \tag{12}$$

where $\chi^{agg} \equiv \sum_{n\in N}\frac{(\alpha_n-\alpha)^2}{\alpha(1-\alpha)}\theta_n$, $\bar{s}^M \equiv \sum_{n\in N}\frac{(\alpha_n-\alpha)(\alpha_n-\alpha^M)\theta_n}{\sum_{n'\in N}(\alpha_{n'}-\alpha)(\alpha_{n'}-\alpha^M)\theta_{n'}}s^M_n$,
$\bar{\sigma}^N \equiv \sum_{n\in N}\frac{\alpha_n(1-\alpha_n)\theta_n}{\sum_{n'\in N}\alpha_{n'}(1-\alpha_{n'})\theta_n}\sigma^N_n$, *and* $\bar{\zeta}^N \equiv \sum_{n\in N}\frac{(\alpha_n-\alpha)(\alpha_n-\alpha^M)\theta_n s^M_n}{\sum_{n'\in N}(\alpha_{n'}-\alpha)(\alpha_{n'}-\alpha^M)\theta_{n'}s^M_{n'}}\zeta^N_n$.

Substitution within industries depends on $\bar{\sigma}^N$, a weighted average of the industry elasticities of substitution defined in Proposition 1. $\bar{\zeta}^N$ is similarly a weighted average of industry level elasticities of substitution between materials and non-materials (we relegate the expression ζ^N_n to Appendix A). χ^{agg} is the cross-industry heterogeneity index and is proportional to the cost-weighted variance of industry capital shares.

2.3 Entry and Exit

This section incorporates entry and exit by introducing entry and overhead costs. A large continuum of entrepreneurs can each draw a random technology by paying an entry cost of f^E units of final output. The technology τ is a constant-returns-to scale production function drawn from an exogenous distribution with CDF $T(\tau)$. After observing the draw, the entrepreneur can operate a plant with that technology if she is willing to pay an overhead

[9]The local elasticities σ_{ni} and ζ_{ni} are defined to satisfy $\sigma_{ni} - 1 \equiv \frac{d\ln rK_{ni}/wL_{ni}}{d\ln w/r}$ and $\zeta_{ni} - 1 \equiv \frac{1}{\alpha^M-\alpha_{ni}}\frac{d\ln\frac{qM_{ni}}{rK_{ni}+wL_{ni}}}{d\ln w/r}$. The definition of σ_{ni} is straightforward but ζ_{ni} is more subtle; see Appendix A for details. If i's production function takes a nested CES form as in Assumption 1, σ_{ni} and ζ_{ni} would equal σ_n and ζ_n respectively. Here, however, these elasticities are not parameters of a production function. Instead, they are defined locally in terms of derivatives of i's production function evaluated at the cost-minimizing input bundle. Exact expressions for σ_{ni} and ζ_{ni} are given in Web Appendix G.1.

cost of f^O units of final output.[10] Free entry determines the price level. Entry costs are incurred before production, and we assume that they are not recorded in our data.

Let E_τ be an indicator of whether plant τ chooses to operate. Should the plant enter, we denote its capital share by α_τ and its expenditure on capital and labor relative to the average expenditure by θ_τ. Thus the aggregate capital share can be expressed as $\alpha = \int \alpha_\tau \theta_\tau E_\tau \mathrm{d}T(\tau)$. Again, we can derive an expression for the aggregate elasticity of substitution by differentiating each side with respect to relative factor prices. One subtlety that emerges in this context is that aggregate production is no longer homogeneous of degree one in primary inputs because gains from variety can alter the cost of intermediate inputs relative to the cost of capital and labor. We follow Robinson (1933) and define the aggregate elasticity as the derivative holding fixed the level of aggregate output. We show in Web Appendix H that in an economy with a single industry, that aggregate elasticity of substitution is

$$\sigma^{agg} = (1-\chi)\bar{\sigma} + \int \frac{\alpha_\tau - \alpha}{\alpha(1-\alpha)} \frac{\mathrm{d}E_\tau}{\mathrm{d}\ln w/r} \theta_\tau dT(\tau) + \chi \bar{s}^M \bar{\zeta} + \chi(1-\bar{s}^M)\varepsilon \tag{13}$$

where $\chi \equiv \frac{\int (\alpha_\tau-\alpha)^2 \theta_\tau E_\tau \mathrm{d}T(\tau)}{\alpha(1-\alpha)}$, $\bar{\sigma} \equiv \frac{\int \alpha(1-\alpha)\theta_\tau E_\tau \sigma_\tau \mathrm{d}T(\tau)}{\int \alpha(1-\alpha)\theta_\tau E_\tau \mathrm{d}T(\tau)}$, $\bar{s}^M \equiv \frac{\int (\alpha_\tau-\alpha)(\alpha_\tau-\alpha^M)\theta_\tau s_\tau^M \mathrm{d}T(\tau)}{\int (\alpha_\tau-\alpha)(\alpha_\tau-\alpha^M)\theta_\tau \mathrm{d}T(\tau)}$, $\bar{\zeta} \equiv \frac{\int (\alpha_\tau-\alpha)(\alpha_\tau-\alpha^M)\theta_\tau s_\tau^M \zeta_\tau \mathrm{d}T(\tau)}{\int (\alpha_\tau-\alpha)(\alpha_\tau-\alpha^M)\theta_\tau s_\tau^M \mathrm{d}T(\tau)}$, and $\alpha^M \equiv \frac{\mathrm{d}\ln P/r}{\mathrm{d}\ln w/r}$.

The difference from (11) is an extra term that captures the change in aggregate factor shares due to entry and exit; an increase in the wage induces labor-intensive plants to exit and capital-intensive plants to enter.

3 Aggregation in the Baseline Model

The methodology developed in the previous section shows how to recover an aggregate capital-labor elasticity from micro parameters and the distribution of plant expenditures. We now use plant-level data on US manufacturing plants to estimate all of the micro components. We then assemble these components to estimate the aggregate capital-labor elasticity of substitution for the US manufacturing sector under the assumptions of our baseline model.

[10] In this section we assume that entry and overhead costs use final output. While the existing literature provides little guidance on the factor content of entry and overhead costs, we study several alternative assumptions theoretically and quantitatively in Web Appendix H.

3.1 Data

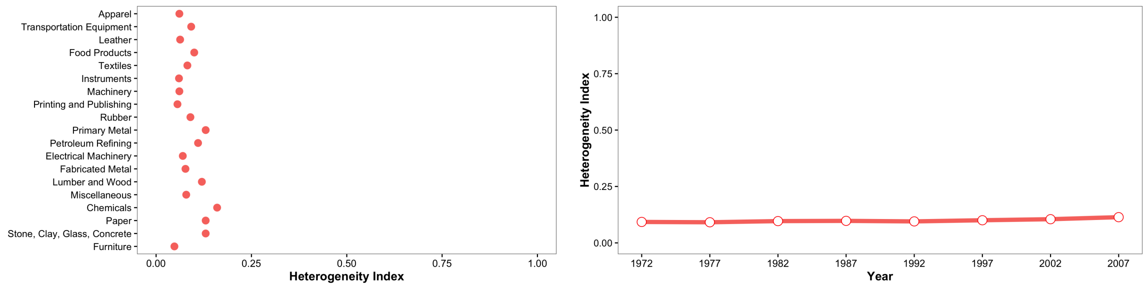
Our main sources of micro data on manufacturing plants are the US Census of Manufactures and Annual Survey of Manufactures (ASM). The Census of Manufactures is a census of all manufacturing plants conducted every five years with more than 180,000 plants per year.[11] The Annual Survey of Manufactures tracks about 50,000 plants over five year panel rotations between Census years, and includes the largest plants with certainty.

We measure capital using perpetual inventory measures created by the Census. Capital costs consist of the total stocks of structures and equipment capital multiplied by the appropriate rental rate, using a Jorgensonian user cost of capital based upon an external real rate of return of 3.5 percent as in Harper et al. (1989). In the ASM subsamples, we include machinery rents as part of capital costs as well. For labor costs, both surveys contain total salaries and wages at the plant level, but the ASM subsample includes data on supplemental labor costs including benefits as well as payroll and other taxes. For details about data construction, see Web Appendix B.

The Census of Manufactures, unlike the ASM subsample, only contains capital data beginning in 1987. Further, industry definitions change from SIC to NAICS in 1997. We thus take the following approach to estimating the aggregate elasticity.

We use the Census of Manufactures from 1987 through 2007 to estimate the micro elasticities and examine their robustness. We then use the ASM in each year for the relevant information on the composition of plants – the heterogeneity indices and materials shares – to extend the analysis from 1972 to 2007. So, for example, to compute the aggregate elasticity of substitution in 1977, we combine estimates of micro parameters from later Censuses with information from the 1977 ASM. We define industries in our aggregation framework at the two digit SIC or three digit NAICS level, and allow plant elasticities and heterogeneity indices to vary by industry.

[11] We exclude small Administrative Record plants with fewer than five employees, for whom the Census only tracks payroll and employment, in line with the rest of the literature using manufacturing Census data. We also omit plants in Alaska and Hawaii because we do not have amenity instruments for Alaska and Hawaii.



(a) Across Industries

(b) Over Time

Figure 1 Heterogeneity Indices

Note: The left figure displays the heterogeneity index, χ_n, in 1987 for each industry. The right figure displays the average heterogeneity index over time.

3.2 Micro Heterogeneity

The heterogeneity index measures the extent of heterogeneity in capital intensities, and so determines the relative importance of within-plant substitution and reallocation. Figure 1a depicts these indices for each industry in 1987. Across industries, the indices average 0.1 and are all less than 0.2. Similarly, Figure 1b shows how the average heterogeneity index evolves over time. While heterogeneity indices are rising, they remain relatively small. Industry capital shares exhibit even less variation; the cross-industry heterogeneity index χ^{agg} is 0.07.[12,13]

Given this level of heterogeneity, the plant-level elasticity of substitution between capital and labor is the primary determinant of the aggregate elasticity. Therefore, we begin with estimates of this elasticity.

[12]It may seem surprising that these heterogeneity indices are so small. Because capital shares are less than one, their variance is smaller than their standard deviation. In Web Appendix D.1, we calibrate a log-normal distribution for the capital cost to labor cost ratio and show that there would need to be much more dispersion in capital shares than we see in the data in order to have values of χ substantially above what we report in Figure 1a. For example, for χ to be 0.2, the 90-10 ratio for the capital cost to labor cost ratio would have to be more than three times larger than observed.

[13]Although classical measurement error would lead us to overstate the heterogeneity indices, measurement error due to imputation (White et al., 2018) would lead us to understate these indices. Note, however, that the numerator of the heterogeneity index is a *cost-weighted* variance of capital shares, and measurement error is likely less important for larger plants.

3.3 Plant-Level Elasticity of Substitution

We obtain the plant-level elasticity of substitution from Raval (2019).[14] We describe the methodology and estimates from that paper in detail in order to explain how we map the theory to the data.

Given cost minimization, the relationship between relative expenditures on capital and labor rK_{ni}/wL_{ni} and relative factor prices w/r identifies the plant-level elasticity. We exploit wage differences across local areas in the US in order to identify the micro elasticity of substitution between capital and labor. Because these wage differences are persistent, with a correlation between 1990 and 2000 of 0.93, they identify plants' long-run response to a permanent change in factor prices. Our measure of local areas is the commuting zone; commuting zones are clusters of US counties designed to have high commuting ties within cluster, so workers in the same commuting zone should face similar wages.

We run the regression:

$$\log \frac{rK_{nic}}{wL_{nic}} = \beta_n \log w_c + \gamma_n X_{nic} + \epsilon_{nic}$$

where w_c is the wage for commuting zone c in which plant i in industry n is located. Under our baseline model, $\beta_n = (\sigma_n - 1)$. The regression only uses plants in a single year; the implicit assumption is that capital is mobile so all plants face the same cost of capital. Since the local wage should reflect the cost of an efficiency unit of labor, local wages are estimated controlling for observable measures of skill. To obtain the local wages, we first use data from the Population Censuses and the American Community Surveys (ACS) to estimate a residual wage for each person after controlling for education, experience, and demographics. We then average this residual within each commuting zone.[15] X_{nic} are additional controls; all regressions control for 4 digit SIC or 6 digit NAICS industry fixed effects, as well as plant age and multi-unit status.

This specification has several attractive properties. First, the dependent variable uses a plant's wage bill rather than a count of employees. If employees supply efficiency units

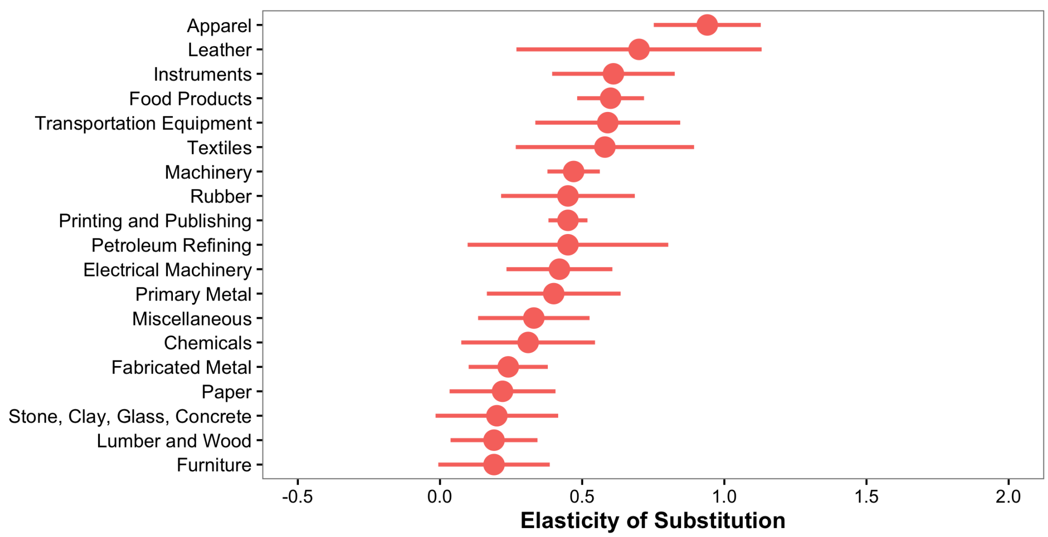
[14] Figure 2, Table I, and Table II were originally published in Raval (2019).

[15] For details about how we construct this wage, see Web Appendix B.3.

of labor, using the wage bill automatically controls for differences in skill across plants. Second, the local wage and plant wage bills are calculated using different data sources, so we avoid division bias from measurement error in the wage in the dependent and independent variables.[16]

Using all manufacturing plants, Raval (2019) estimates a plant level elasticity of substitution ranging from 0.3 to 0.5 across Census years. In addition, we allow the elasticity to vary by industry through separate regressions in each industry. Figure 2 displays the estimates by industry for 1987 along with 95 percent confidence intervals.[17] Most of the estimates range between 0.3 and 0.7.

Figure 2 Plant Elasticity of Substitution by Industry, 1987



Note: For each industry, this graph plots the plant level elasticity of substitution between capital and labor as estimated in Raval (2019), together with the 95 percent confidence interval for each estimate. Estimates based on wages derived from the Population Census, as described in the text. Standard errors are clustered at the commuting zone level.

3.3.1 Instruments

We have used differences in wages across locations to estimate the plant-level elasticity of substitution. Our variation therefore captures how average factor shares in a location vary

[16] An alternative approach would be to instrument for the plant-level wage using the local wage. We show in Web Appendix C.7 that if regions differ in efficiency units of labor per worker in a way that is correlated with the local wage, such an estimate would be biased. In general, we find lower estimates of the elasticity when we instrument for the plant-level wage than in our baseline estimates.

[17] We list these estimates in Web Appendix C.

with the local wage. This group-level design brings up an important endogeneity issue. If local wages are correlated with local non-neutral productivity or unobserved skill differences, then our estimate of the elasticity of substitution would be biased towards one as $\sigma - 1$ would be attenuated.[18]

To address such endogeneity problems, we examine three sets of instruments for the labor supply facing manufacturing plants which we assume are independent of local area confounding variables. First, we use a version of Bartik's (1991) instrument for labor market conditions, which is based on the premise that local areas differ in their industrial composition. When an industry expands nationwide, commuting zones more heavily exposed to that industry experience larger increases in labor demand, raising local workers' outside options. Thus given each area's initial industrial composition, we can construct the change in each area's labor demand due to the change in each industry's nationwide employment. Because we want an instrument for the supply of labor facing manufacturing plants, we construct the instrument using non-manufacturing industries only.[19]

Second, we use two instruments for labor market conditions from Beaudry et al. (2012). Like Bartik's (1991) instrument, these instruments are based on a model in which local industrial composition affects the labor supply facing manufacturing plants through workers' outside options. While the local wage premium, defined as local shares of each industry multiplied by each industry's wage premium, is endogenous, Beaudry et al. (2012) develop two instruments for it: the interaction of predicted changes in industry employment shares with initial national industry wage premia, and the interaction of national changes in industry wage premia with predicted contemporaneous industry employment shares.[20]

[18]Even though our identifying variation uses a group level design, our strategy estimates the *average* response of plant-level factor shares, not the response of a location's *aggregate* factor shares. The latter incorporates changes in the scale of capital-intensive plants relative to labor-intensive plants, which are not captured in these regressions. See Section 4.4 for a discussion of estimates from specifications using aggregated data.

[19]Formally, the instrument is constructed as follows: Let $g_n(t) = \frac{1}{10}\ln(L_n(t)/L_n(t-10))$ be the national growth rate of industry n, and let $\omega_{j,n}(t)$ be the share of area j's employment working in industry n. The instrument is the interaction between initial area employment shares and 10 year national employment growth rates: $Z_j(t) = \sum_{n \in N^S} \omega_{j,n}(t-10) g_n(t)$, where N^S is the set of non-manufacturing four-digit SIC or six-digit NAICS industries.

[20]Formally, let $v_n(t)$ be the national wage premium in industry n in time t, $\omega_{j,n}(t)$ be the share of local area j's employment working in industry n, and $\hat{\omega}_{j,n}(t)$ be the predicted share of local area j's employment working in industry n. The predicted employment share $\hat{\omega}_{j,n}(t)$ is predicted based on national employment changes in the same way as the Bartik instrument. The first instrument is then $\sum_{n \in N^S} v_n(t)(\hat{\omega}_{j,n}(t) - \omega_{j,n}(t-10))$ and

One issue with both the Bartik and BGS instruments is that they do not account for input-output linkages; a shock that leads to growth in non-manufacturing industries could affect local manufacturing plants that sell to or buy from them. We thus include a third set of instruments that are not affected by this issue; measures of local amenities based on climate and geography. Workers would accept a lower wage in locations with better amenities (Rosen, 1979; Roback, 1982). Building on Albouy et al. (2016), we include measures of the slope, elevation, relative humidity, average precipitation, average sunlight, average dew point, the number of hot degree days and cold degree days, and temperature day bins for each commuting zone.[21]

For the instruments for labor market conditions, we measure local wages using data on average payroll to employment across all establishments in a commuting zone from the Longitudinal Business Database (LBD). The Population Censuses are only conducted every 10 years in different years from the Economic Censuses, so the wages from the Population Censuses do not match the year of the Economic Census. For most of our specifications, this mismatch is not a problem because our wage variation is highly persistent. This mismatch becomes a problem if we want to use short-run variation in wages. While the LBD-based wages do not control for differences in individual worker characteristics, the instruments should be orthogonal to the measurement error in wages. We examine whether this is the case using the amenity instruments, and find that the estimates of the elasticity using the LBD-based wages are only slightly higher than estimates using Population Census-based wages.

Table I contains estimates of the elasticity of substitution using these instruments.[22] The first two columns report OLS estimates, based upon the Population Census-based wages in

the second instrument is $\sum_{n \in N^S} \hat{\omega}_{j,n}(t)(v_n(t) - v_n(t-10))$, where N^S is the set of non-manufacturing four-digit SIC or six-digit NAICS industries. National industry wage premia are fixed effects from a regression of establishment wages on industry dummy variables.

[21] The amenities in Albouy et al. (2016) were collected at the PUMA level; we aggregate them to the commuting zone level by taking an average across PUMAs in the same commuting zone, weighting PUMAs by their population in the commuting zone. We exclude amenities based on distance to the coast or lakes as these may also affect import and export possibilities, and thus the productivity of the plant and the competition it faces.

[22] Because industry definitions change over time, we often have to use slightly different years for the instrument, or the 5 year instrument and its lag rather than the 10 year instrument. The exact time periods underlying each instrument are detailed in Appendix B.

the first column and LBD-based wages in the second column. On average, the elasticities using LBD-based wages are about 0.13 higher than those based on Population Census-based wages. Estimates using the Population Census-based wages are likely lower because they control for differences in worker skills across areas, unlike the LBD-based wages.

Columns 3 and 4 estimate the elasticity instrumenting with the Bartik and BGS instruments given LBD-based wages, while Columns 5 and 6 use the amenity instruments given the Population Census-based and LBD-based wages, respectively. The last column uses all three sets of instruments. Reassuringly, the estimates of the elasticity using instruments range from 0.3 to 0.6, similar to the range of the OLS estimates. In addition, estimates of the elasticity using amenity based instruments are similar to those with labor demand based instruments, indicating that input-output linkage issues are not a major concern.

Table I IV Estimates of The Plant Capital-Labor Substitution Elasticity

Year	OLS		Bartik	BGS	Amenities		All
1987	0.44 *(0.04)*	0.54 *(0.03)*	0.52 *(0.04)*	0.45 *(0.09)*	0.45 *(0.07)*	0.48 *(0.06)*	0.51 *(0.04)*
1992	0.47 *(0.03)*	0.52 *(0.03)*	0.45 *(0.04)*	0.48 *(0.04)*	0.57 *(0.06)*	0.55 *(0.05)*	0.50 *(0.03)*
1997	0.29 *(0.05)*	0.48 *(0.04)*	0.41 *(0.11)*	0.36 *(0.08)*	0.28 *(0.09)*	0.40 *(0.07)*	0.41 *(0.05)*
2002	0.31 *(0.06)*	0.48 *(0.05)*	0.31 *(0.10)*	0.37 *(0.06)*	0.33 *(0.13)*	0.42 *(0.11)*	0.42 *(0.06)*
2007	0.45 *(0.04)*	0.58 *(0.03)*	0.51 *(0.05)*	0.56 *(0.05)*	0.49 *(0.09)*	0.53 *(0.07)*	0.54 *(0.04)*
Wage	Pop Census	LBD	LBD	LBD	Pop Census	LBD	LBD

Note: Standard errors are in parentheses and are clustered at the commuting zone level. All regressions include industry dummies, age fixed effects, and a multiunit status indicator. Instruments are as defined in the text. Wages used are the average log wage for the commuting zone. In the first and fifth columns, the wage is computed as wage and salary income over total number of hours worked adjusted for differences in worker characteristics using the Population Censuses; in all other cases, the wage is computed as payroll/number of employees at the establishment level using the LBD.

3.3.2 Other Threats to Identification

Our estimate of the micro elasticity of substitution would be biased if rental rates vary systematically with local wages. Rental rates might vary with wages for three reasons. First, the cost of some kinds of capital, such as structures, may reflect local wages. To examine this, we estimate the elasticity of substitution between labor and equipment capital, which is more plausibly mobile across locations.[23] Second, the cost of capital could vary

[23] Because the Census only collects separate equipment and structures capital data for all manufacturing plants before 1997, we only estimate this specification for 1987 and 1992.

because of differences in lending rates from banks in different locations, or from differences in firm creditworthiness or access to capital markets. To control for these differences, we add firm fixed effects. Third, individual states could have different capital taxes or investment subsidies; we control for these by including state fixed effects.

Table II compares our baseline approach to these alternative specifications. The first two columns contain our baseline least squares estimates. The first column contains the average plant-level elasticity when we estimate separate elasticities for each industry.[24] The second column estimates a common elasticity across all industries in manufacturing. Columns (3)-(5) contain estimates for our robustness checks. The estimates including only equipment capital or including state fixed effects are quite close to those from the baseline specification. The estimates including firm fixed effects are somewhat higher, with elasticities between 0.55 and 0.65 and are about 0.21 higher on average compared to the baseline OLS estimates.[25] One reason that the firm fixed effect estimates might overstate the true elasticity is that firm-wide wage setting procedures might compress wage differences within firms, which would bias the estimate toward 1.[26]

Another concern is that plant level capital stocks in the Census are measured with error. We examine this concern in two ways; first, we use only plants in the ASM, as these plants tend to have longer investment histories so that perpetual inventory measures of capital are better measured. Second, we use a book value measure of capital as our capital stock measure. We report these results in columns (6) and (7) of Table II; we find a slightly higher elasticity on average using only ASM plants (about 0.08 higher compared to baseline), and a slightly lower elasticity on average using book value of capital (0.05 lower compared to

[24]This average, along with other averages across industries, is a weighted average where the weight on industry n is $\frac{\alpha_n(1-\alpha_n)\theta_n}{\sum_{n\in N}\alpha_n(1-\alpha_n)\theta_n}$ as in Proposition 2.

[25]The sample of plants differ between columns (1)-(4) and column (5) because adding firm fixed effects means that our we identify the elasticity using only plants in multi-plant firms. However, if we use the sample of column (5) but omit firm fixed effects, we obtain estimates close to our baseline. It is thus the inclusion of these fixed effects that leads to the difference.

[26]If firms are constrained to pay similar wages across plants in different locations, then our measured local differences in wages would overstate the true differences in wages. This would attenuate our estimate of $1-\sigma$, i.e., it would bias our estimate toward one. As an extreme example, if firms were constrained to pay the same wage in every location, firms would face the same factor prices in every location, and we would observe no response of factor shares to the local wage. We indeed show in Web Appendix C.7 that wage differences across locations among plants within the same multi-plant firm are smaller than wage differences among all plants across those locations. See Silva (forthcoming) for further evidence of fairness norms in multi-plant firms and Hjort et al. (2020) for further evidence of firm-wide wage setting procedures.

Table II Robustness Checks for Plant Capital-Labor Substitution Elasticity

	(1) Separate OLS	(2) Single OLS	(3) Equipment Capital	(4) Firm FE	(5) State FE	(6) ASM Only	(7) Book Value Capital
1987	0.43	0.44 *(0.04)*	0.45 *(0.03)*	0.57 *(0.07)*	0.39 *(0.04)*	0.40 *(0.08)*	0.42 *(0.04)*
1992	0.48	0.47 *(0.03)*	0.47 *(0.03)*	0.65 *(0.06)*	0.31 *(0.03)*	0.67 *(0.07)*	0.39 *(0.03)*
1997	0.34	0.29 *(0.05)*		0.66 *(0.06)*	0.32 *(0.05)*	0.42 *(0.09)*	0.27 *(0.05)*
2002	0.34	0.31 *(0.06)*		0.59 *(0.06)*	0.41 *(0.07)*	0.52 *(0.09)*	0.22 *(0.07)*
2007	0.38	0.45 *(0.04)*		0.55 *(0.07)*	0.48 *(0.05)*	0.37 *(0.07)*	0.39 *(0.04)*

Note: The table contains seven specifications, all of which we estimate using OLS regressions. In (1), we average separately-estimated plant elasticity of substitution for each industry using the cross-industry weights. In (2), we estimate a single common elasticity of substitution for the entire manufacturing sector. In (3), we only use equipment capital. In (4), we include firm fixed effects. In (5), we include state fixed effects. In (6), we only use ASM plants and weight using the ASM weights. In (7), we measure capital using book values.

All regressions include industry fixed effects, age fixed effects, and a multi-unit status indicator. Wages used are the average log wage for the commuting zone, computed using the Population Censuses as wage and salary income over total number of hours worked adjusted for differences in worker characteristics. Standard errors, in parentheses, are clustered at the commuting zone level.

baseline).

In Web Appendix C.2 we relax the assumption that plants in each industry share the same common, constant elasticity of capital-labor substitution. We use two methods to estimate the weighted average of local elasticities of substitution, $\bar{\sigma}$—using quantile regression and using a transformation of the dependent variable—and find estimates in a narrow range around our baseline estimates.

3.4 Aggregation

We now estimate the remaining plant-level production and demand parameters and use our baseline framework to aggregate to the manufacturing-level elasticity of substitution. The reallocation effect depends upon both the plant elasticity of substitution between materials and non-materials inputs, ζ, and the elasticity of demand, ε.

To identify ζ, we use the same cross-area variation in the wage. Across commuting zones, the local wage varies but the prices of capital and materials are fixed. Cost minimization implies that ζ measures the response of relative expenditures of materials and non-materials costs to changes in their respective prices: $1-\zeta = \frac{\mathrm{d}\ln[(rK_{ni}+wL_{ni})/qM_{ni}]}{\mathrm{d}\ln(\lambda_{ni}/q)}$, where λ_{ni} is the cost index of i's capital-labor bundle. Holding fixed the prices of materials and capital, a change in the local wage would increase these relative prices by $\mathrm{d}\ln\lambda_{ni}/q = (1-\alpha_{ni})\mathrm{d}\ln w$. To a first-order approximation, the response of $(rK_{ni} + wL_{ni})/qM_{ni}$ to the local wage would be $(1-\zeta)(1-\alpha_{ni})$. We therefore estimate ζ using the regression:

$$\log \frac{rK_{nic} + wL_{nic}}{qM_{nic}} = (1-\zeta)(1-\alpha_{nic})(\log w_c) + \gamma_n X_{nic} + \epsilon_{nic}.$$

Table III contains these estimates. Because we use the full Census for each estimate, our estimate of ζ is common across industries. The first column contains our baseline estimates. These estimates range between 0.6 and 1 across years.[27]

Table III Plant-Level Elasticities of Substitution between Materials and Non-Materials for the Manufacturing Sector

1987	1.03 *(0.12)*
1992	0.83 *(0.10)*
1997	0.69 *(0.07)*
2002	0.78 *(0.08)*
2007	0.57 *(0.06)*

Note: Standard errors are in parentheses. All regressions include industry fixed effects, age fixed effects and a multi-unit status indicator, and have standard errors clustered at the commuting zone level.

Within industries, the demand elasticity tells us how much consumers substitute across plants when relative prices change. We estimate the elasticity of demand using the implications of profit maximization; optimal price setting behavior implies that the markup over marginal cost is equal to $\frac{\varepsilon}{\varepsilon-1}$. Thus, we invert the average markup across plants in an industry to obtain the elasticity of demand. The assumption of constant returns to scale implies that each plant's markup is equal to its ratio of revenue to cost. Figure 3 displays the elasticities of demand for each manufacturing industry in 1987.[28] Across industries, elasticities

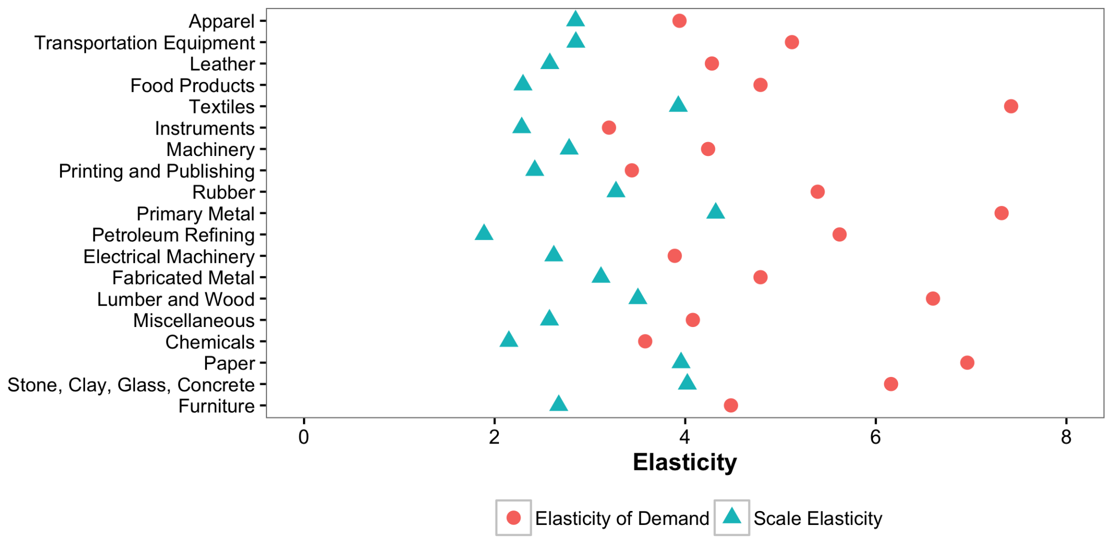
[27] This estimation strategy implicitly assumes a national market for materials. We have run additional specifications for 1987 and 1992 that adjust the regression to account for the fact that some materials are sourced locally; an increase in the local wage would raise the cost of such locally sourced materials. These estimates are only 0.02 to 0.03 than our baseline estimates of the materials-non materials elasticity. See Web Appendix D.4 for details. Atalay (2017) pursued a complementary approach using differences in materials prices across plants in the US Census of Manufactures and finds an estimate of 0.65, within the range of Table III. See Appendix F of his paper. This differs from his main estimate of this elasticity which uses shorter-run industry-level variation and includes non-manufacturing industries, and so may not reflect the long-run, plant-level elasticity for manufacturing required for our model.

[28] We list these estimates in Web Appendix D.2. In addition, in Web Appendix D.3 we examine alternative approaches to estimate the demand elasticity. We also show that our baseline strategy is robust to assuming diminishing returns to scale, and that less than 100% pass through will mute the scale response.

of demand vary between three and seven in 1987.

The overall scale elasticity is $\bar{s}_n^M \zeta + (1 - \bar{s}_n^M)\varepsilon_n$. $\bar{s}_n^M$ is an average of materials shares, which are high in manufacturing; the average across industries in 1987 is 0.52. Figure 3 contains our estimates of the scale elasticities; they average 2.9 across industries in 1987.

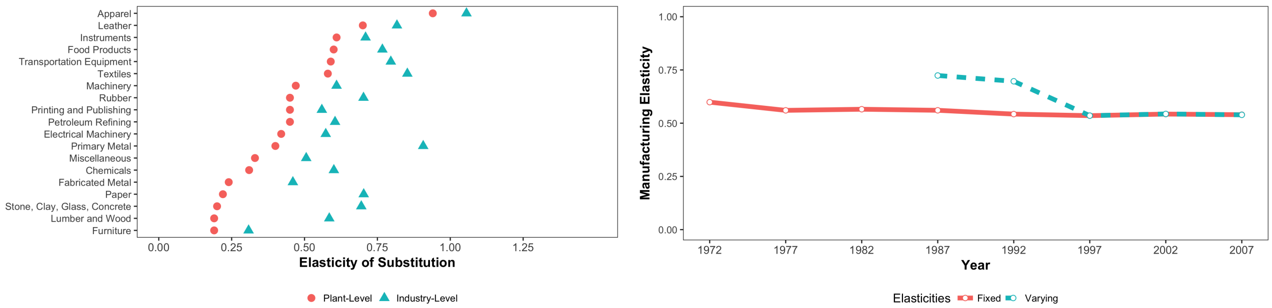
Figure 3 Elasticity of Demand and Scale Elasticity by Industry, 1987



Note: For each industry, this graph plots both the elasticity of demand, estimated from revenue-cost ratios, and the scale elasticity $\bar{s}_n^M \zeta + (1 - \bar{s}_n^M)\varepsilon_n$.

To aggregate across industries, we need one more elasticity, the cross industry elasticity of demand η. We estimate this elasticity using industry-level panel data by regressing quantity on price, using average cost as an instrument for supply. Web Appendix D.5 contains the details of this analysis. Across specifications, we find estimates centered around one. We thus set η to one. As we would expect, the cross-industry demand elasticity η is much lower than the plant-level demand elasticities; varieties in the same industries are better substitutes than varieties in other industries.

We can now combine the substitution and reallocation effects to estimate the industry and manufacturing sector level elasticities of substitution. In Figure 4a, we depict the plant-level and industry-level elasticities of substitution. Because the heterogeneity indices tend to be small, the industry-level elasticities of substitution are only moderately higher than the plant-level elasticities. The average industry elasticity is 0.68 and the overall manufacturing level elasticity of substitution is 0.72 in 1987. Within-plant substitution accounts for 68



(a) Plant and Industry Elasticities

(b) Manufacturing Sector Elasticity Across Time

Figure 4 Elasticities of Substitution and Aggregation

Note: The left figure displays the plant level elasticity and industry level elasticity of substitution for each industry. The right figure displays the manufacturing level elasticity of substitution for each Census year from 1972-2007.

percent of industry substitution and 58 percent of overall substitution for manufacturing.[29]

Our methodology allows us to examine how much changes in the composition of plants caused the aggregate elasticity of substitution to vary over time as the manufacturing sector has evolved.[30] Figure 4b depicts the aggregate elasticity of substitution from 1972 to 2007. We estimate this elasticity in two ways. First, we freeze all elasticities at their values in 1997, the year for which we can estimate elasticities using both SIC and NAICS industry definitions, but let all of the sufficient statistics vary over time. In that case, the red solid line in Figure 4b, the aggregate elasticity has fallen slightly from 0.60 to 0.54 from 1972 to 2007. Alternatively, between 1987 and 2007 we can allow all elasticities to vary by year. The resulting time path of aggregate elasticities, the blue dashed line, reveals a larger fall in the elasticity, from 0.72 to 0.54, over time, although the time path for 1997 and after is virtually identical to the previous case.

This decline primarily reflects two changes. First, the average demand elasticity falls from 5.0 in 1987 to 3.3 in 2007, which explains approximately half of the decline. Second, the average plant elasticity falls substantially between 1992 and 1997, from 0.48 in 1992 to

[29] While our estimate is not directly comparable to estimates of aggregate elasticities for the entire private sector, we speculate about the implications for such an elasticity in Web Appendix E.2. In Web Appendix E.3 and Web Appendix E.4 we discuss the differences between our methodology and one that estimates an aggregate elasticity using aggregate time series as well as the methodology of Karabarbounis and Neiman (2014).

[30] After 1997, industry definitions switch from two digit SIC basis to three digit NAICS basis; we estimate plant elasticities using SIC industries for 1987 and 1992 and NAICS industries for 1997, 2002, and 2007.

0.34 in 1997.

4 Additional Margins of Adjustment

We now enrich our theoretical model to incorporate several additional margins of adjustment. First, a factor price change may induce some plants to exit and other plants to enter. Here, the degree of adjustment will vary across plants: the contribution of a plant on the margin of entering or exiting is different from that of a plant that is inframarginal. Second, adjustment costs may mean that plants' factor usage does not reflect static cost minimization. Third, factor price changes may induce the creation of intermediates that complement particular technologies, shifting the technological frontier. Finally, we examine other considerations such as substitution to intangible capital. For each channel, we discuss how our estimates would change.

4.1 Entry and Exit

The last section provided an estimate of the aggregate elasticity of capital-labor substitution for the manufacturing sector using a model with a fixed set of firms. We now interpret the evidence using a model that incorporates entry and exit that was introduced in Section 2.3. The expression for the aggregate elasticity (13) can be rearranged as

$$\sigma^{agg} = (1-\chi)\left[\sigma + \frac{\int (\alpha_\tau - \alpha)\frac{\mathrm{d}E_\tau}{\mathrm{d}\ln w/r}\theta_\tau dT(\tau)}{\int \alpha_\tau(1-\alpha_\tau)\theta_\tau dT(\tau)}\right] + \chi \bar{s}^M \zeta + \chi(1-\bar{s}^M)\varepsilon \qquad (14)$$

The term in brackets captures the response of plants' capital shares to an increase in the cost of labor relative to capital coming from the intensive and extensive margins: within-plant substitution, and a contribution from the entry of capital intensive plants and exit of labor intensive plants. The second term captures changes in scale coming from substitution between materials and primary inputs, while the third term comes from changes in scale coming from changes in revenue.

In this section, we use this formula to bound the true elasticity. We show that the upper bound is quite close to our baseline estimate, and derive a lower bound using dynamic panel estimates of within-plant substitution. The implied range for the aggregate capital-labor elasticity for the manufacturing sector averages $[0.35, 0.65]$ across years.[31]

Upper Bound

Our baseline estimate of the aggregate elasticity of substitution is an upper bound on the true aggregate elasticity for three reasons. First, our cross-sectional estimates of the micro elasticity of substitution incorporate both within-plant substitution and changes due to entry and exit, i.e. the entire term in brackets in (14). At root, our cross-sectional estimates capture how the average capital share varies with the local wage, and changes in this average reflect both the intensive and extensive margins. In Web Appendix H we provide an explicit characterization and describe how selection causes an upward bias if marginal firms tend to be more labor intensive than average.

Second, the estimated micro elasticity of substitution between intermediate and primary inputs, $\hat{\zeta}$, reported in Table III, is larger than $\bar{\zeta}$. $\bar{\zeta}$ captures only the intensive margin—substitution within plants—while $\hat{\zeta}$ uses cross-sectional variation and incorporates both the intensive and extensive margins.

Third, our baseline strategy overstates how a plant's scale responds to a change in its marginal cost because part of this cost—the overhead cost—is fixed. Formally, we had estimated this response from plants' revenue to cost ratio, which we interpreted in our model as the markup, $\frac{\varepsilon}{\varepsilon-1}$. Here, cost includes both variable cost and fixed overhead cost, so the ratio of revenue to cost for a plant that operates technology τ is $\frac{\hat{\varepsilon}_\tau}{\hat{\varepsilon}_\tau-1} = \frac{\frac{\varepsilon}{\varepsilon-1}\text{Variable Costs}}{\text{Variable Costs}+\text{Overhead Costs}} \leq \frac{\varepsilon}{\varepsilon-1}$, or $\hat{\varepsilon} \geq \varepsilon$.

Thus, our cross-sectional estimates provide an upper bound for the true aggregate elasticity for the US manufacturing sector that averages 0.65 across years.[32]

[31] While we have assumed that entry and overhead costs both use final output, we explore alternative assumptions about their factor content in Web Appendix H. We show that if plants' overhead costs have the same factor content as their variable costs, or if the overhead cost used labor, the upper bound remains valid and the lower bound is slightly lower. We also examine the case in which both entry costs and overhead costs require the entrepreneur's labor, but these costs—the opportunity cost of the entrepreneur's time—do not appear on the plant's wage bill. In such a world, one can differentiate between an aggregate elasticity of substitution that captures how measured factor shares respond to changes in factor prices and one that captures how underlying resource usage (which incorporates unmeasured labor) responds. In practice, the two elasticities are fairly close. We show that the former corresponds to our baseline estimate, while the latter is about 0.1 higher than our baseline estimate.

[32] To derive the upper bound, we compute the aggregate elasticity in each year using our baseline formula but using the estimated cross-sectional elasticity from column (4) of Table C.2. This uses an estimate of the

Lower Bound

We next characterize a lower bound. The changes in plants' capital and labor usage can be divided into two parts, the intensive and extensive margins, i.e., the two terms in brackets in (14). The extensive margin must be non-negative. To quantify the intensive margin, we now exploit the panel structure of our data to examine how individual plants respond to changes in factor prices. Because this adjustment may be slow, the long-run response to a factor price change should be larger than the short-run adjustment. We estimate the following econometric model for plant i and time period t:

$$\log \frac{K_{itc}}{L_{itc}} = \rho_5 \log \frac{K_{it-5c}}{L_{it-5c}} + \rho_{10} \log \frac{K_{it-10c}}{L_{it-10c}} + \beta \log(w_{tc}/r_t) + \eta_i + \delta_t + \gamma_{n(i)} t + \epsilon_{itc} \quad (15)$$

where η_i is a plant fixed effect, ρ_5 and ρ_{10} measure the degree of persistence in the capital-labor ratio through two lags, and β measures the short-run elasticity of substitution. We estimate this relationship in terms of the capital-labor ratio, and not the capital cost - labor cost ratio, so the long run capital-labor elasticity is $\frac{\beta}{1-\rho_5-\rho_{10}}$.[33] Because we examine plants over time, we decompose the bias of plant i's technology into a plant fixed effect, η_i, a time fixed effect, δ_t, an 3-digit industry specific trend, $\gamma_{n(i)} t$, and a residual ϵ_{itc}.

We then use the Blundell-Bond model to estimate this relationship instrumenting for the wage-rental ratio using all of the instruments used earlier.[34] The estimate of the short run elasticity using the unbalanced panel is 0.21 with a standard error of 0.09. The implied long run elasticity is then 0.34, substantially lower than our cross-sectional estimates.[35] Our long-run intensive margin estimates are similar to those found in the literature using alternative methods.[36]

$\bar{\sigma}$ derived from regressing α_i, which we show is the appropriate theoretical object in this context, rather than $\ln \frac{\alpha_i}{1-\alpha_i}$, on the log of the local wage. In practice, the estimates of $\bar{\sigma}$ using each approach are very similar.

[33] We measure the labor input at a plant as the wage bill divided by the local wage.

[34] Local amenities instrument for the local wage level, while the Bartik and BGS shocks instrument for both wage levels and changes. Wages are based on establishment data in order to match the same year as the Economic Census.

[35] In Web Appendix C.5, we examine several additional specifications and find qualitatively similar conclusions.

[36] An alternative strategy to estimate the micro elasticity is to use variation in the user cost of capital over time stemming from changes in tax-laws or the price of capital that differentially affect asset types; see Chirinko (2008) for a survey of this literature. Chirinko et al. (2011) and Barnes et al. (2008) provided estimates that are the conceptually closest to ours, as they used long-run movements in the user cost of capital to identify the long-run micro elasticity for US and UK public firms respectively. Their approach

These, together with the conservative restrictions that $\varepsilon > 1$ and $\zeta > 0$, deliver a lower bound for the aggregate elasticity which averages 0.35 across years.

4.2 Adjustment Frictions and Distortions

In Web Appendix I, we study models in which heterogeneity in cost shares of capital are due to adjustment costs or misallocation frictions. In these models, the only difference with our baseline is the scale response; since factor usage does not reflect static cost minimization, we can no longer use Shephard's Lemma to characterize the response of scale.[37] If all heterogeneity is due to "exogenous wedges" (Hsieh and Klenow, 2009), our baseline elasticity recovers the true elasticity to a first order. With capital adjustment frictions, our baseline elasticity would slightly overestimate the actual elasticity. Finally, with plant-specific input prices, our baseline elasticity recovers the response of factor shares to uniform changes in factor prices.

4.3 Shifts in the Technological Frontier

Shifts in factor prices may induce changes in the technological frontier, as outlined by Acemoglu (2002), Acemoglu (2003), and Acemoglu (2010). Holding the technological frontier fixed, an increase in the wage would change the economy's capital-labor ratio. This would change the size of the market for innovations that complement each factor, and the subsequent adjustment of the technological frontier could amplify or dampen the initial wage increase. We articulate an explicit model of directed technical change in Web Appendix J.1. In that context, we can distinguish between the short-run aggregate elasticity which holds the technological frontier fixed and the long-run elasticity which includes shifts in the frontier. If the innovations that complement each factor are available nationwide, then our baseline estimate corresponds to the short-run elasticity because they are based on cross-sectional differences at a point in time. We show that if the short-run aggregate elasticity is less than

estimates the elasticity using the capital first order condition and allows for biased technical change at the industry level. Each estimated a micro elasticity of substitution of 0.4.

[37] One can identify each plant with a production function and a history of shocks. The micro elasticities of substitution in both our baseline formula and our empirical estimates correspond to how each such plant's factor usage would be different if factor prices were permanently different.

one, then the long-run elasticity is between the short-run value and one; the induced shift in the technological frontier dampens but does not reverse the sign of the initial shift in factor shares.

4.4 Other Considerations

In addition, in Web Appendix E.2, we estimate industry-level elasticities of substitution by aggregating the micro data to the industry level in each location. With 4-digit SIC / 6-digit NAICS industries, the estimates are relatively close to the implied industry-level elasticities that we built up from individual plants. When we use broader industry definitions, the estimates are much less consistent across specifications, likely due to differences in composition across locations. In Web Appendix J.2, we study an economy in which each plant uses intangible capital as well as physical capital in production and find negligible changes to our estimates, as the compensation of intangible capital accrues to the factors that produce it.[38]

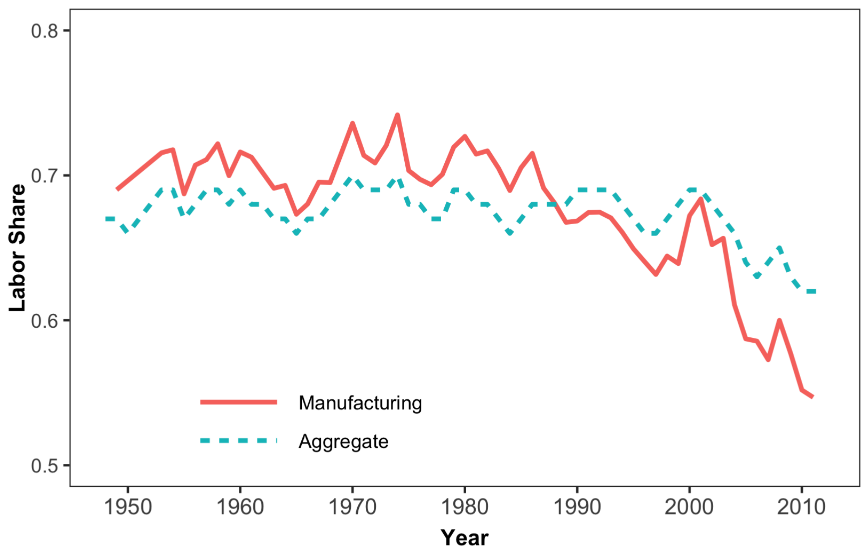
In Web Appendix J.4, we examine several reasons why the response of plants' capital-labor ratios to local factor prices might differ from the response to a national change. Our identification strategy relied on measuring the response to local factor prices, but an aggregate elasticity captures the response to national factor prices. One plausible reason for such differences is endogenous location choice, which would lead to an upward bias in our baseline estimate of the aggregate elasticity. Another is national shifts in the technological frontier, which would imply a downward bias, but that the true elasticity is less than one. While it is difficult to say anything definitive, one way to gauge the importance of these mechanisms is to ask how factor shares in a location respond to the wage in nearby locations. We do not find a large response to the wage in nearby locations after controlling for the wage in the plant's own location.

[38]We show that our strategy is robust to modeling intangible capital as increasing productivity or demand.

5 The Decline of the Labor Share

Figure 5 depicts the labor share of income in the US for the manufacturing sector and for the aggregate economy in the post-war period.[39] The labor share for manufacturing has fallen since 1970, from about 0.73 to 0.55 in 2011. The steepest decline has been since 2000; the labor share fell from roughly 0.65 to 0.55 in one decade. The labor share has fallen for the overall economy as well, though not by as much, falling from 0.70 in 1970 to 0.62 in 2011. In this section, we will examine the decline in the labor share for manufacturing.

Figure 5 Labor Share over Time



Note: The solid line is the labor share for manufacturing based on data from the BLS Multifactor Productivity series. The dashed line is the labor share for the overall economy from Fernald (2012).

Along a balanced growth path in the neoclassical growth model, labor-augmenting technical change and the induced rise in wages exactly offset to keep factor shares constant. The labor share has recently been falling, which means that the race has not been even.

Some mechanisms that have recently been proposed to explain the decline in the labor share—such as an acceleration in investment specific technical change as in Karabarbounis and Neiman (2014) or increased capital accumulation as in Piketty (2014)—would affect

[39] As Gomme and Rupert (2004), Krueger (1999), and Elsby et al. (2013) point out, the major issue with calculating the labor share is whether proprietors' income accrues to labor or capital. Both series assume that the share of labor for proprietors is the same as for corporations. For the manufacturing sector, proprietors' income represented 1.4 percent of income on average since 1970.

relative factor supply but not the bias of technical change. When we use the term "bias of technical change," we mean any change in the economy that would alter relative factor demand. Other mechanisms, such as changes in preferences, markups, demographics, technology, or trade barriers, would also alter the implied bias of technical change.

To examine the role of these two classes of mechanisms, we decompose the change in the labor share into a contribution from factor prices and a contribution from the bias of technical change. Given that we estimate the elasticity of substitution is less than unity and wages have risen relative to rental rates of capital, factor price changes have raised the labor share holding technology fixed. However, this was true both when the labor share was relatively stable in the middle part of the twentieth century and also when the labor share declined more recently. To draw conclusions about which class of mechanisms caused the decline in the labor share, it is important to examine the *changes* in the behavior of each component over time.

One advantage of our strategy for estimating the aggregate capital-labor elasticity of substitution is that, unlike most other work, it does not impose restrictions on the behavior of the bias of technical change over time. For example, most time series estimates of the aggregate elasticity impose that the bias of aggregate technical change follows a linear trend. Because we impose no such restrictions, our approach is well-suited to studying changes in the trend of the bias of technical change.

Formally, let $s^{v,L}$ denote labor's share of value added. Then the change in the labor share can be decomposed into two terms

$$\mathrm{d}s^{v,L} = \underbrace{\frac{\partial s^{v,L}}{\partial \ln w/r}\mathrm{d}\ln w/r}_{\text{contr. from factor prices}} + \underbrace{\left(\mathrm{d}s^{v,L} - \frac{\partial s^{v,L}}{\partial \ln w/r}\mathrm{d}\ln w/r\right)}_{\text{contr. from bias}} \tag{16}$$

The first measures the contribution of changes in factor prices. The second measures a residual that we label the "bias of technical change". In executing this decomposition, we use the long-run manufacturing-level elasticity of substitution, and compute the long-run contribution from factor price changes.

To execute this decomposition, we need measures of factor prices and expenditures. We base our rental prices on NIPA deflators for equipment and structures, and wages on average compensation per hour worked adjusted for changes in worker quality over time. Our measures of value added and input expenditures are based on NIPA data. Our estimates of expenditures on fixed capital combine NIPA data on equipment and structures capital with our rental prices.[40]

The black line in Figure 6 displays the cumulative contribution of the bias of technical change to the decline in the labor share. Using our baseline aggregate elasticity, the biased residual contributes to a decline of 20 percentage points from 1970 to 2010. The contribution of the bias accelerates starting in 2000; about half of the cumulative contribution of the bias is from the 2000-2010 period.[41]

We then demonstrate that the value of the aggregate elasticity is crucial to measure the magnitude of the bias of technical change by doing the same decomposition using alternative values for the aggregate elasticity. First, if we used the value of plant-level elasticity as the aggregate elasticity, the cumulative contribution of the bias would be two percentage points higher than our baseline. Second, with a much higher aggregate elasticity of 2, the bias of technical change does not contribute to the labor share decline prior to 2000, and the overall cumulative contribution is much smaller at 7 percentage points. For all three values of the elasticity, the bias of technical change accelerates after 2000.

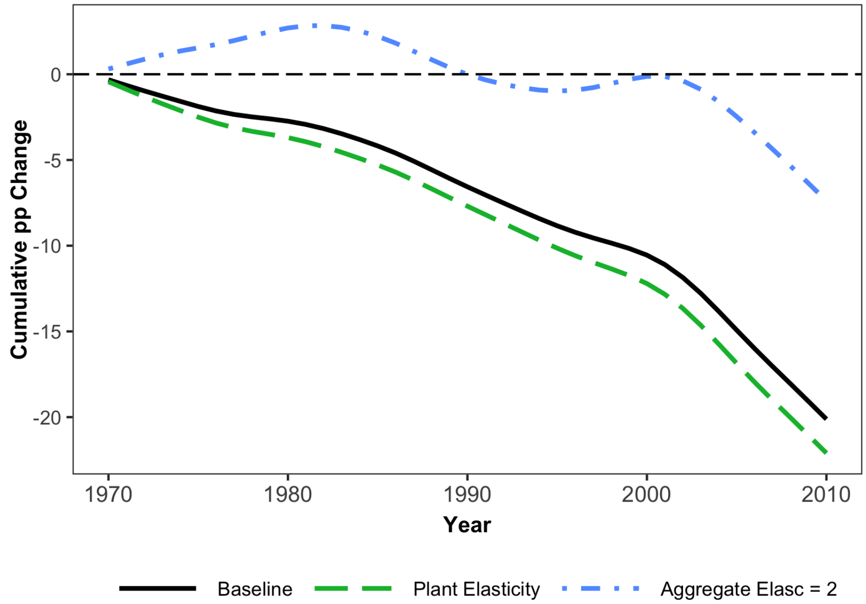
An interesting feature is that the break in trend is much less drastic with our baseline elasticity than with an aggregate elasticity of 2. A more steady contribution of biased technical change is consistent with, among other factors, slow, continuous adoption of automation technologies (Acemoglu and Restrepo, 2018).

We then examine the contribution of changes in factor prices to the labor share. Using our baseline aggregate elasticity, the contribution of changes in factor prices has been rela-

[40]See Web Appendix F for details of construction of the factor price series, the details of this calculation, and decompositions that use alternative rental price series, alternative assumptions about the interest rate faced by firms, and factor shares from production data rather than NIPA.

[41]In this decomposition, we decompose total income into payments to labor, payments to fixed capital, and a residual that we label "profit". The bias incorporates contributions from changes in labor's share of cost as well from changes in the profit share. Both accelerate after 2000, although the change in the annual contribution from the profit share is more stark. In Web Appendix F.2 we pursue an alternative approach put forward by Karabarbounis and Neiman (2019) which assumes that this profit share reflects unmeasured payments to capital. In that case, the cumulative contribution of the bias rises by 3 percentage points.

Figure 6 Cumulative Contribution of the Bias of Technical Change to the Decline in the Labor Share



Note: All changes are in percentage points. Each contribution to the labor share change has been smoothed using a Hodrick-Prescott filter. The contributions are based on different values for the aggregate elasticity of substitution, and are as defined in the text.

tively constant over the past 40 years, raising the labor share by 0.10 percentage points per year in the 1970-1999 period and 0.08 percentage points per year in the 2000-2010 period. This means that explanations that work purely through factor supplies would have trouble matching the timing of the decline in the labor share.[42]

Finally, while real wage growth has indeed slowed after 1970 compared to the immediate post-war period, we argue that this slowdown accounts for only a small share of the decline in the labor share. From 1970-2011, annual growth in real wages in the manufacturing sector was 1.9 percentage points lower than 1954-1969. We use our estimated elasticities to assess how much of the decline in the labor share can be explained by the slowdown in wage growth.[43] We find that if wages grew faster, the manufacturing labor share in 2010 would be 0.59, compared to its actual value of 0.55. Thus, the slowdown in wages can explain about

[42] In an environment with directed technical change, a change in relative factor supply might induce a change in the bias of technology. Nevertheless, as discussed in Section 4.3, the long-run aggregate elasticity incorporating directed technical change is between our estimated elasticity and unity, muting the effect on factor shares. Thus incorporating directed technical change would not alter our conclusion that explanations stemming from changes in relative factor supply would have trouble matching the decline in the labor share.

[43] This counterfactual involves extrapolating our local estimate of the aggregate elasticity of substitution.

20 percent of the total decline.

6 Conclusion

This paper has developed a new approach to estimate the aggregate elasticity of substitution between capital and labor by building up from micro structural parameters and the cross-sectional distribution of plant-level expenditures. Our approach has several advantages. First, when we estimate micro-level parameters, we can use variation in the supply of factors that is plausibly independent of technological differences, which is difficult to do at the aggregate level. We can also estimate changes in the aggregate elasticity and the bias of technology over time. In addition, with cross-sectional micro data, there are many ways to examine the robustness of our model and identification strategy.

We then applied our methodology to data on plants in the US manufacturing sector. While the average plant-level elasticity lies roughly between 0.3 and 0.5, our estimates indicated that the aggregate elasticity of substitution between capital and labor for the manufacturing sector has been between 0.5 and 0.7, with a slight decline from 1972 to 2007. While we have applied our methodology to capital-labor substitution, one could posit a richer production structure and estimate elasticities between varieties of labor and capital. We believe this would be a fruitful way to gain insight into the evolution of skill premia and inequality.

References

Acemoglu, Daron, "Directed Technical Change," *Review of Economic Studies*, 2002, *69* (4), 781–809.

_ , "Labor- And Capital-Augmenting Technical Change," *Journal of the European Economic Association*, 2003, *1* (1), 1–37.

_ , "When Does Labor Scarcity Encourage Innovation?," *Journal of Political Economy*, 2010, *118* (6), 1037–1078.

_ **and Pascual Restrepo**, "The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 2018, *108* (6), 1488–1542.

Albouy, David, Walter Graf, Ryan Kellogg, and Hendrik Wolff, "Climate Amenities, Climate Change, and American Quality of Life," *Journal of the Association of Environmental and Resource Economists*, 2016, *3* (1), 205–246.

Alvarez-Cuadrado, Francisco, Ngo Van Long, and Markus Poschke, "Capital-Labor Substitution, Structural Change and the Labor Income Share," *Journal of Economic Dynamics and Control*, 2018, *87*, 206–231.

Antras, Pol, "Is the US Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution," *Contributions to Macroeconomics*, 2004, *4* (1).

Atalay, Enghin, "How Important are Sectoral Shocks?," *American Economic Journal: Macroeconomics*, 2017, *9* (4), 254–80.

Barnes, Sebastian N, Simon Price, and Maria Barriel, "The Elasticity of Substitution: Evidence from a UK Firm-level Data Set," Technical Report 2008.

Bartik, Timothy J., *Who Benefits from State and Local Economic Development Policies?*, W. E. Upjohn Institute for Employment Research, 1991.

Beaudry, Paul, David A. Green, and Benjamin Sand, "Does Industrial Composition Matter For Wages? A Test Of Search And Bargaining Theory," *Econometrica*, 2012, *80* (3), 1063–1104.

Berndt, Ernst R., "Reconciling Alternative Estimates of the Elasticity of Substitution," *Review of Economics and Statistics*, 1976, *58* (1), 59–68.

Boehm, Johannes and Ezra Oberfield, "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," Technical Report, National Bureau of Economic Research 2018.

Chirinko, Robert S., "Sigma: The Long and Short of It," *Journal of Macroeconomics*, 2008, *30*, 671–686.

— , **Steven M. Fazzari, and Andrew P. Meyer**, "A New Approach to Estimating Production Function Parameters: The Elusive Capital-Labor Substitution Elasticity," *Journal of Business and Economic Statistics*, 2011, *29* (4), 587–594.

Diamond, Peter, Daniel McFadden, and Miguel Rodriguez, "Measurement of the Elasticity of Factor Substitution and Bias of Technical Change," in Melvyn Fuss and Daniel McFadden, eds., *Production Economics: A Dual Approach to Theory and Applications*, 1978, chapter 5.

Elsby, Michael W. L., Bart Hobijn, and Aysegul Sahin, "The Decline of the U.S. Labor Share," *Brookings Papers on Economic Activity*, 2013.

Fernald, John, "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," September 2012. Mimeo.

Gomme, Paul and Peter Rupert, "Measuring Labor's Share of Income," 2004. Policy Discussion Paper - No 7, Federal Reserve Bank of Cleveland.

Harper, Michael J., Ernst R. Berndt, and David O. Wood, "Rates of Return and Capital Aggregation Using Alternative Rental Prices," in D.W. Jorgenson and R. London, eds., *Technology and Capital Formation*, Cambridge, MA: MIT Press, 1989.

Herrendorf, Berthold, Christopher Herrington, and Akos Valentinyi, "Sectoral technology and structural transformation," *American Economic Journal: Macroeconomics*, 2015, *7* (4), 104–33.

Hicks, John R., *The Theory of Wages*, MacMillan and Co, 1932.

Hjort, Jonas, Xuan Li, and Heather Sarsons, "Across-Country Wage Compression in Multinationals," Technical Report, National Bureau of Economic Research 2020.

Houthakker, Hendrik, "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," *The Review of Economic Studies*, 1955, *23* (1), pp. 27–31.

Hsieh, Chang-Tai and Peter Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Jones, Chad, "The Shape of Production Functions and the Direction of Technical Change," *Quarterly Journal of Economics*, 2005, *120* (2), 517–549.

Karabarbounis, Loukas and Brent Neiman, "The Global Decline of the Labor Share," *Quarterly Journal of Economics*, 2014, *129*, 61–103.

— **and** — , "Accounting for Factorless Income," *NBER Macroeconomics Annual*, 2019, *33* (1), 167–228.

Klump, Rainer, Peter McAdam, and Alpo Willman, "Factor Substitution and Factor Augmenting Technical Progress in the US," *Review of Economics and Statistics*, 2007, *89* (1), 183–192.

Krueger, Alan B., "Measuring Labor's Share," *American Economic Review*, 1999, *89* (2), 45–51.

Lagos, Ricardo, "A Model of TFP," *Review of Economic Studies*, October 2006, *73* (4), 983–1007.

Leon-Ledesma, Miguel A., Peter McAdam, and Alpo Willman, "Identifying the Elasticity of Substitution with Biased Technical Change," *The American Economic Review*, 2010, *100* (4), 1330–1357.

Levhari, David, "A Note on Houthakker's Aggregate Production Function in a Multifirm Industry," *Econometrica*, 1968, *36* (1), pp. 151–154.

Luttmer, Erzo G. J., "Growth with a Fixed Factor," 2012. Mimeo.

Mangin, Sephorah, "A Theory of Production, Matching, and Distribution," *Journal of Economic Theory*, 2017, *172*, 376–409.

Miyagiwa, Kaz and Chris Papageorgiou, "Endogenous Aggregate Easticity of Substitution," *Journal of Economic Dynamics and Control*, 2007, *31* (9), 2899–2919.

Piketty, Thomas, *Capital in the Twenty-First Century*, Belknap Press, 2014.

Raval, Devesh, "What's Wrong with Capital in the Twenty-First Century's model," in Heather Boushey, J. Bradford DeLong, and Marshall Steinbaum, eds., *After Piketty: The Agenda for Economics and Inequality*, Harvard University Press, 2017, chapter 4, pp. 75–98.

Raval, Devesh R, "The Micro Elasticity of Substitution and Non-Neutral Technology," *The RAND Journal of Economics*, 2019, *50* (1), 147–167.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Robinson, Joan, *The Economics of Imperfect Competition*, MacMillan and Co, 1933.

Rognlie, Matthew, "A note on Piketty and diminishing returns to capital," Technical Report 2014.

Rosen, Sherwin, "Wages, Rents, and the Quality of Life," in Peter M. Mieszkowski and Mahlon R. Straszheim, eds., *Current Issues in Urban Economics*, Baltimore: Johns Hopkins University Press, 1979, p. 74–104.

Sato, Kazuo, *Production Functions and Aggregation*, Elsevier, 1975.

Silva, Rui, "Internal Labor Markets, Wage Convergence and Investment," *Journal of Financial and Quantitative Analysis*, forthcoming.

White, T Kirk, Jerome P Reiter, and Amil Petrin, "Imputation in US Manufacturing Data and its Implications for Productivity Dispersion," *Review of Economics and Statistics*, 2018, *100* (3), 502–509.

Appendix

A Proofs of Propositions

We first derive an expression for the industry-level elasticity of substitution. In Section 2, we assumed that plants' production functions took a nested CES functional form. Here we relax this functional form assumption. While we maintain that each plant's production function has constant returns to scale, we impose no further structure.

Assumption 1′ *Plant i in industry n uses the constant returns to scale production function $F_{ni}(K_{ni}, L_{ni}, M_{ni})$.*

The purpose of this is twofold. First, it provides a generalization of the formulas in Proposition 1 that we will use later in Section C.2. Second, we will later use the formulas here to aggregate across industries and derive an expression for the aggregate elasticity of substitution between capital and labor for the manufacturing sector as a whole.

For plant i, we define the local elasticities σ_{ni} and ζ_{ni} to satisfy

$$\sigma_{ni} - 1 = \frac{\mathrm{d}\ln rK_{ni}/wL_{ni}}{\mathrm{d}\ln w/r} \tag{A.1}$$

$$\zeta_{ni} - 1 = \frac{1}{\alpha^M - \alpha_{ni}} \frac{\mathrm{d}\ln \frac{qM_{ni}}{rK_{ni}+wL_{ni}}}{\mathrm{d}\ln w/r} \tag{A.2}$$

If i's production function takes a nested CES form as in Assumption 1, σ_{ni} and ζ_{ni} would equal σ_n and ζ_n respectively.[44] Under Assumption 1′, they are defined locally in terms of derivatives of F_{ni} evaluated at i's cost-minimizing input bundle. Exact expressions for σ_{ni} and ζ_{ni} are given in Web Appendix G.1. Proposition 1′ characterizes the industry elasticity of substitution σ_n^N. The resulting expression is identical to Proposition 1 except the plant elasticities of substitution are replaced with weighted averages of the plants' local elasticities, $\bar{\sigma}_n$ and $\bar{\zeta}_n$.

Proposition 1′ *Under Assumption 1′, the industry elasticity of substitution $\sigma_n^N = \frac{\mathrm{d}\ln K_n/L_n}{\mathrm{d}\ln w/r}$ is:*

$$\sigma_n^N = (1-\chi_n)\bar{\sigma}_n + \chi_n\left[(1-\bar{s}_n^M)\varepsilon_n + \bar{s}_n^M\bar{\zeta}_n\right] \tag{A.3}$$

where $\chi_n \equiv \sum_{i\in I_n} \frac{(\alpha_{ni}-\alpha_n)^2\theta_{ni}}{\alpha_n(1-\alpha_n)}$ and $\bar{s}_n^M \equiv \sum_{i\in I_n} \frac{(\alpha_{ni}-\alpha_n)(\alpha_{ni}-\alpha^M)\theta_{ni}}{\sum_{j\in I_n}(\alpha_{nj}-\alpha_n)(\alpha_{nj}-\alpha^M)\theta_{nj}} s_{ni}^M$ as in Proposition 1 and

$$\bar{\sigma}_n \equiv \sum_{i\in I_n} \frac{\alpha_{ni}(1-\alpha_{ni})\theta_{ni}}{\sum_{j\in I_n}\alpha_{nj}(1-\alpha_{nj})\theta_{nj}}\sigma_{ni}$$

$$\bar{\zeta}_n \equiv \sum_{i\in I_n} \frac{(\alpha_{ni}-\alpha_n)(\alpha_{ni}-\alpha^M)\theta_{ni}s_{ni}^M}{\sum_{j\in I_n}(\alpha_{nj}-\alpha_n)(\alpha_{nj}-\alpha^M)\theta_{nj}s_{nj}^M}\zeta_{ni}$$

[44] The definition of σ_{ni} is straightforward but ζ_{ni} requires some elaboration. Suppose that F_{ni} takes the nested CES form of Assumption 1. Let $\lambda_{ni} \equiv \left[(r/A_{ni})^{1-\sigma_n} + (w/B_{ni})^{1-\sigma_n}\right]^{\frac{1}{\sigma_n - 1}}$ be the marginal cost of i's capital-labor bundle. Then cost minimization implies $\frac{qM_{ni}}{rK_{ni}+wL_{ni}} = \left(\frac{q/C_{ni}}{\lambda_{ni}}\right)^{1-\zeta_n}$. Since $\frac{\mathrm{d}\ln q/r}{\mathrm{d}\ln w/r} = 1-\alpha^M$ and, from Shephard's Lemma, $\frac{\mathrm{d}\ln \lambda_{ni}/r}{\mathrm{d}\ln w/r} = 1-\alpha_{ni}$, we have that $\frac{\mathrm{d}\ln\frac{qM_{ni}}{rK_{ni}+wL_{ni}}}{\mathrm{d}\ln w/r} = (\zeta_n - 1)(\alpha^M - \alpha_{ni})$.

Proof. As discussed in the text, the definitions of σ_{ni} and σ_n^N imply

$$\begin{aligned}\sigma_{ni}-1 &= \frac{\mathrm{d}\ln\frac{rK_{ni}}{wL_{ni}}}{\mathrm{d}\ln w/r}=\frac{1}{\alpha_{ni}(1-\alpha_{ni})}\frac{\mathrm{d}\alpha_{ni}}{\mathrm{d}\ln w/r}, \qquad \forall i\in I_n\\ \sigma_n^N-1 &= \frac{\mathrm{d}\ln\frac{rK_n}{wL_n}}{\mathrm{d}\ln w/r}=\frac{1}{\alpha_n(1-\alpha_n)}\frac{\mathrm{d}\alpha_n}{\mathrm{d}\ln w/r}\end{aligned}$$

Since $\alpha_n=\sum_{i\in I_n}\alpha_{ni}\theta_{ni}$, we can differentiate to get

$$\begin{aligned}\sigma_n^N-1 &= \frac{1}{\alpha_n(1-\alpha_n)}\left[\sum_{i\in I_n}\frac{\mathrm{d}\alpha_{ni}}{\mathrm{d}\ln w/r}\theta_{ni}+\sum_{i\in I_n}\alpha_{ni}\frac{\mathrm{d}\theta_{ni}}{\mathrm{d}\ln w/r}\right]\\ &= \frac{\sum_{i\in I_n}\alpha_{ni}(1-\alpha_{ni})\theta_{ni}(\sigma_{ni}-1)}{\alpha_n(1-\alpha_n)}+\frac{\sum_{i\in I_n}\alpha_{ni}\frac{\mathrm{d}\theta_{ni}}{\mathrm{d}\ln w/r}}{\alpha_n(1-\alpha_n)}\end{aligned}$$

Using the definition of $\bar{\sigma}_n$ and $\sum_{i\in I_n}\frac{\mathrm{d}\theta_{ni}}{\mathrm{d}\ln w/r}=0$, we have

$$\sigma_n^N-1=(\bar{\sigma}_n-1)\sum_{i\in I_n}\frac{\alpha_{ni}(1-\alpha_{ni})\theta_{ni}}{\alpha_n(1-\alpha_n)}+\frac{\sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\frac{\mathrm{d}\theta_{ni}}{\mathrm{d}\ln w/r}}{\alpha_n(1-\alpha_n)} \tag{A.4}$$

Since $\chi_n=\sum_{i\in I_n}\frac{(\alpha_{ni}-\alpha_n)^2\theta_{ni}}{\alpha_n(1-\alpha_n)}$, one can verify that $\sum_{i\in I_n}\frac{\alpha_{ni}(1-\alpha_{ni})\theta_{ni}}{\alpha_n(1-\alpha_n)}=1-\chi_n$, which gives

$$\sigma_n^N=(1-\chi_n)\bar{\sigma}_n+\frac{\sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\frac{\mathrm{d}\theta_{ni}}{\mathrm{d}\ln w/r}}{\alpha_n(1-\alpha_n)}+\chi_n \tag{A.5}$$

We now find an expression for $\frac{\mathrm{d}\ln\theta_{ni}}{\mathrm{d}\ln w/r}$. θ_{ni} can be written as

$$\theta_{ni}=\frac{rK_{ni}+wL_{ni}}{\sum_{j\in I_n}rK_{nj}+wL_{nj}}=\frac{(1-s_{ni}^M)z_{ni}}{\sum_{j\in I_n}(1-s_{nj}^M)z_{nj}}=\frac{(1-s_{ni}^M)z_{ni}}{(1-s_n^M)z_n}$$

where $z_{ni}\equiv rK_{ni}+wL_{ni}+qM_{ni}$ and $z_n\equiv rK_n+wL_n+qM_n$. Since $\mathrm{d}\ln\frac{1-s_{ni}^M}{s_{ni}^M}=\frac{1}{s_{ni}^M}\mathrm{d}\ln(1-s_{ni}^M)$, the definition of ζ_n implies

$$\frac{\mathrm{d}\ln(1-s_{ni}^M)}{\mathrm{d}\ln w/r}=s_{ni}^M(\zeta_n-1)(\alpha_{ni}-\alpha^M)$$

The change in i's expenditure on all inputs depends on its expenditure shares:

$$\begin{aligned}\frac{z_{ni}}{z_n} &= \frac{rK_{ni}+wL_{ni}+qM_{ni}}{\sum_{j\in I_n}rK_{nj}+wL_{nj}+qM_{nj}}=\frac{\frac{\varepsilon_n-1}{\varepsilon_n}P_{ni}Y_{ni}}{\sum_{j\in I_n}\frac{\varepsilon_n-1}{\varepsilon_n}P_{nj}Y_{nj}}=\frac{\frac{\varepsilon_n-1}{\varepsilon_n}P_{ni}^{1-\varepsilon_n}Y_nP_n^{\varepsilon_n}}{\sum_{j\in I_n}\frac{\varepsilon_n-1}{\varepsilon_n}P_{nj}^{1-\varepsilon_n}Y_nP_n^{\varepsilon_n}}\\ &= \left(\frac{P_{ni}}{P_n}\right)^{1-\varepsilon_n}\end{aligned}$$

The change in i's price depends on the change in its marginal cost

$$\frac{\mathrm{d}\ln P_{ni}/r}{\mathrm{d}\ln w/r} = \frac{\mathrm{d}\ln\frac{\varepsilon_n}{\varepsilon_n-1}mc_{ni}}{\mathrm{d}\ln w/r}=\frac{\mathrm{d}\ln mc_{ni}}{\mathrm{d}\ln w/r}=(1-s_{ni}^M)(1-\alpha_{ni})+s_{ni}^M(1-\alpha^M) \tag{A.6}$$

Putting these pieces together, we have $\theta_{ni} = \frac{1-s^M_{ni}}{1-s^M_n}\left(\frac{P_{ni}}{P_n}\right)^{1-\varepsilon_n}$, so differentiating and using $\sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\theta_{ni} = 0$ yields

$$\begin{aligned}
\sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\theta_{ni}\frac{\mathrm{d}\ln\theta_{ni}}{\mathrm{d}\ln w/r} &= \sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\theta_{ni}\left[\frac{\mathrm{d}\ln 1-s^M_{ni}}{\mathrm{d}\ln w/r} + (1-\varepsilon_n)\frac{\mathrm{d}\ln P_{ni}/r}{\mathrm{d}\ln w/r}\right] \\
&= \sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\theta_{ni}\left\{\begin{array}{c} s^M_{ni}(\zeta_{ni}-1)(\alpha_{ni}-\alpha^M) \\ +(1-\varepsilon_n)\left[(1-s^M_{ni})(1-\alpha_{ni}) + s^M_{ni}(1-\alpha^M)\right]\end{array}\right\} \\
&= \sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\theta_{ni}\left\{s^M_{ni}(\zeta_{ni}-\varepsilon_n)(\alpha_{ni}-\alpha^M) + (1-\varepsilon_n)(1-\alpha_{ni})\right\}
\end{aligned}$$

Using the definitions of $\bar{\zeta}_n$, $\bar{s}^M_n$, and χ_n, this becomes

$$\begin{aligned}
\sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\theta_{ni}\frac{\mathrm{d}\ln\theta_{ni}}{\mathrm{d}\ln w/r} &= \sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\theta_{ni}\left[s^M_{ni}(\bar{\zeta}_n-\varepsilon_n)(\alpha_{ni}-\alpha^M) + (1-\varepsilon_n)(1-\alpha_{ni})\right] \\
&= \sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\theta_{ni}\left[\bar{s}^M_n(\bar{\zeta}_n-\varepsilon_n)(\alpha_{ni}-\alpha^M) + (1-\varepsilon_n)(1-\alpha_{ni})\right] \\
&= \sum_{i\in I_n}(\alpha_{ni}-\alpha_n)\theta_{ni}\left[\bar{s}^M_n(\bar{\zeta}_n-\varepsilon_n)(\alpha_{ni}-\alpha_n) + (1-\varepsilon_n)(\alpha_n-\alpha_{ni})\right] \\
&= \alpha_n(1-\alpha_n)\chi_n\left[(\bar{\zeta}_n-\varepsilon_n)\bar{s}^M_n - (1-\varepsilon_n)\right]
\end{aligned}$$

Finally, we can plug this back into (A.5) to get

$$\sigma^N_n = (1-\chi_n)\bar{\sigma}_n + \chi_n\left[\bar{s}^M_n\zeta_n + (1-\bar{s}^M_n)\varepsilon_n\right]$$

■

To build up to the aggregate elasticity of substitution between capital and labor, we proceed in exactly the same way as with the industry-level elasticity. Define ζ^N_n to satisfy $(\zeta^N_n-1)(\alpha_n-\alpha^M) = \frac{\mathrm{d}\ln\frac{1-s^M_n}{s^M_n}}{\mathrm{d}\ln w/r}$. The claim below will give an expression for ζ^N_n in terms of plant level elasticities and choices.

Proposition 2′ *Under Assumption 1′, the aggregate elasticity between capital and labor is*

$$\sigma^{agg} = (1-\chi^{agg})\bar{\sigma}^N + \chi^{agg}\left[\bar{s}^M\bar{\zeta}^N + (1-\bar{s}^M)\eta\right]$$

where χ^{agg}, $\bar{\sigma}^N$, $\bar{\zeta}^N$, $\bar{s}^M$, and ζ_n^N are defined as

$$\begin{aligned}
\chi^{agg} &= \sum_{n\in N} \frac{(\alpha_n-\alpha)^2\theta_n}{\alpha(1-\alpha)} \\
\bar{\sigma}^N &= \sum_{n\in N} \frac{\alpha_n(1-\alpha_n)\theta_n}{\sum_{n'\in N}\alpha_{n'}(1-\alpha_{n'})\theta_{n'}}\sigma_n^N \\
\bar{\zeta}^N &\equiv \frac{\sum_{n\in N}(\alpha_n-\alpha)(\alpha_n-\alpha^M)\theta_n s_n^M \zeta_n^N}{\sum_{n\in N}(\alpha_n-\alpha)(\alpha_n-\alpha^M)\theta_n s_n^M} \\
\bar{s}^M &= \frac{\sum_{n\in N}(\alpha_n-\alpha)(\alpha_n-\alpha^M)\theta_n s_n^M}{\sum_{n\in N}(\alpha_n-\alpha)(\alpha_n-\alpha^M)\theta_n} \\
\zeta_n^N &= \sum_{i\in I_n}\theta_{ni}\left[\zeta_{ni}\frac{s_{ni}^M(\alpha_{ni}-\alpha^M)}{s_n^M(\alpha_n-\alpha^M)} + \varepsilon_n\left\{1-\frac{s_{ni}^M(\alpha_{ni}-\alpha^M)}{s_n^M(\alpha_n-\alpha^M)}\right\}\right]
\end{aligned}$$

Proof. Note that since $\frac{z_{ni}}{z_n} = \frac{P_{ni}Y_{ni}}{P_nY_n}$, we have both $\frac{\mathrm{d}\ln Y_n}{\mathrm{d}\ln w/r} = \sum_{i\in I_n}\frac{z_{ni}}{z_n}\frac{\mathrm{d}\ln Y_{ni}}{\mathrm{d}\ln w/r}$ and $\frac{\mathrm{d}\ln P_n/r}{\mathrm{d}\ln w/r} = \sum_{i\in I_n}\frac{z_{ni}}{z_n}\frac{\mathrm{d}\ln P_{ni}/r}{\mathrm{d}\ln w/r}$. With the first, we can derive an expression for the change in cost that parallels the within-industry expression:

$$\begin{aligned}
\frac{\mathrm{d}\ln z_n/r}{\mathrm{d}\ln w/r} &= \sum_{i\in I_n}\frac{z_{ni}}{z_n}\frac{\mathrm{d}\ln z_{ni}/r}{\mathrm{d}\ln w/r} \\
&= \sum_{i\in I_n}\frac{z_{ni}}{z_n}\frac{\mathrm{d}\ln \frac{\varepsilon_n-1}{\varepsilon_n}Y_{ni}P_{ni}/r}{\mathrm{d}\ln w/r} \\
&= \sum_{i\in I_n}\frac{z_{ni}}{z_n}\left[\frac{\mathrm{d}\ln Y_{ni}}{\mathrm{d}\ln w/r} + (1-s_{ni}^M)(1-\alpha_{ni}) + s_{ni}^M(1-\alpha^M)\right] \\
&= \frac{\mathrm{d}\ln Y_n}{\mathrm{d}\ln w/r} + (1-s_n^M)(1-\alpha_n) + s_n^M(1-\alpha^M)
\end{aligned}$$

With the second, we can derive an expression for the change in the price level that parallels (A.6)

$$\begin{aligned}
\frac{\mathrm{d}\ln P_n/r}{\mathrm{d}\ln w/r} &= \sum_{i\in I_n}\frac{z_{ni}}{z_n}\frac{\mathrm{d}\ln P_{ni}/r}{\mathrm{d}\ln w/r} \\
&= \sum_{i\in I_n}\frac{z_{ni}}{z_n}\left[(1-\alpha_{ni})(1-s_{ni}^M) + s_{ni}^M(1-\alpha^M)\right] \\
&= (1-\alpha_n)(1-s_n^M) + s_n^M(1-\alpha^M)
\end{aligned}$$

Thus following the exact logic of Proposition 1′, we have that

$$\sigma^{agg} = (1-\chi^N)\bar{\sigma}^N + \chi_N\left[\bar{s}^M\bar{\zeta}^N + (1-\bar{s}^M)\eta\right]$$

It remains only to derive the expression for ζ_n^N. Begin with $1-s_n^M = \sum_{i\in I_n}(1-s_{ni}^M)\frac{z_{ni}}{z_n}$. Differentiating each side gives:

$$\frac{\mathrm{d}\ln(1-s_n^M)}{\mathrm{d}\ln w/r} = \sum_{i\in I_n}\frac{(1-s_{ni}^M)z_{ni}}{(1-s_n^M)z_n}\left[\frac{\mathrm{d}\ln(1-s_{ni}^M)}{\mathrm{d}\ln w/r} + \frac{\mathrm{d}\ln z_{ni}/z_n}{\mathrm{d}\ln w/r}\right] = \sum_{i\in I_n}\theta_{ni}\left[\frac{\mathrm{d}\ln(1-s_{ni}^M)}{\mathrm{d}\ln w/r} + \frac{\mathrm{d}\ln z_{ni}/z_n}{\mathrm{d}\ln w/r}\right]$$

Using $(\alpha_{ni}-\alpha^M)(\zeta_{ni}-1)=\frac{1}{s_{ni}^M}\frac{\mathrm{d}\ln(1-s_{ni}^M)}{\mathrm{d}\ln w/r}$ and $(\alpha_n-\alpha^M)(\zeta_n^N-1)=\frac{1}{s_n^M}\frac{\mathrm{d}\ln(1-s_n^M)}{\mathrm{d}\ln w/r}$ gives

$$s_n^M(\alpha_n-\alpha^M)(\zeta_n^N-1)=\sum_{i\in I_n}\theta_{ni}\left[s_{ni}^M(\alpha_{ni}-\alpha^M)(\zeta_{ni}-1)+\frac{\mathrm{d}\ln z_{ni}/z_n}{\mathrm{d}\ln w/r}\right]$$

Finally, we have

$$\begin{aligned}
\frac{\mathrm{d}\ln z_{ni}/z_n}{\mathrm{d}\ln w/r} &= \frac{\mathrm{d}\ln Y_{ni}}{\mathrm{d}\ln w/r}+(1-s_{ni}^M)(1-\alpha_{ni})+s_{ni}^M(1-\alpha^M)-\frac{\mathrm{d}\ln Y_n}{\mathrm{d}\ln w/r}-(1-s_n^M)(1-\alpha_n)-s_n^M(1-\alpha^M)\\
&= (-\varepsilon_n)\frac{\mathrm{d}\ln P_{ni}/P_n}{\mathrm{d}\ln w/r}+(1-s_{ni}^M)(1-\alpha_{ni})+s_{ni}^M(1-\alpha^M)-(1-s_n^M)(1-\alpha_n)-s_n^M(1-\alpha^M)\\
&= (1-\varepsilon_n)\left[(1-s_{ni}^M)(1-\alpha_{ni})+s_{ni}^M(1-\alpha^M)-(1-s_n^M)(1-\alpha_n)-s_n^M(1-\alpha^M)\right]\\
&= (\varepsilon_n-1)\left[s_n^M(\alpha_n-\alpha^M)-s_{ni}^M(\alpha_{ni}-\alpha^M)+(\alpha_{ni}-\alpha_n)\right]
\end{aligned}$$

Plugging this in, we have

$$s_n^M(\alpha_n-\alpha^M)(\zeta_n^N-1)=\sum_{i\in I_n}\theta_{ni}\left\{\begin{array}{c} s_{ni}^M(\alpha_{ni}-\alpha^M)(\zeta_{ni}-1)\\ +(\varepsilon_n-1)\left[s_n^M(\alpha_n-\alpha^M)-s_{ni}^M(\alpha_{ni}-\alpha^M)+(\alpha_{ni}-\alpha_n)\right]\end{array}\right\}$$

Using $\sum_{i\in I_n}\theta_{ni}(1-\varepsilon_n)(\alpha_{ni}-\alpha_n)=0$, dividing through by $s_n^M(\alpha_n-\alpha^M)$ and subtracting 1 from each side gives the result. ■