ECON 3399: AI and Robots at Work

Sergio Ocampo Díaz

University of Western Ontario

Fall 2026

Draft lecture notes. Please report any errors or unclear passages to socampod@uwo.ca.

This notes build and reproduce results in various academic papers, review articles, and handbook chapters. The exposition is my own but I try to follow established results as closely as possible.

Contents

Part I

Aggregate Production Functions

The first part of the course introduces the key economic concepts of how labor enters into production. We start with the canonical aggregate production function in which labor and capital enter as aggregate quantities. The key question there is how the elasticity of substitution between labor and capital affects the distribution of income between labor and capital. We then specialize this question to the most common functional forms, Cobb–Douglas and CES, and discuss their empirical relevance. We then open up the production function by considering the role of different types of labor for production. Again, the substitutability between those inputs plays a key role.

1. Labor and Capital as Aggregate Inputs: Substitutability

We start from the most standard way in which we model how labor and capital are used in production. This involves abstracting from the particularities of the workplace and what happens in it. Instead, we consider the aggregate labor amount, L, and the aggregate amount of capital, K, used in production. In practice this requires a way for us to measure the work hours of different types of people in a single unit. Similarly, we need to account for the presence of machines, structures, vehicles, software and other forms of capital in a single aggregate. These indexes suppress heterogeneity between people and between forms of capital in order to isolate substitution between broad factor classes.

Part of the objective of the task-based production function we develop later in the course is to separate the different types of labor and capital by distinguishing *what they do* and not just *who they are*. For now, we start with the aggregate quantities and a description of the technology that combines them to create output.

1.1. The neoclassical production function

We start from the most general description of the production technology, embodied in the production function,

$$Y = zF(K, L), \tag{1.1}$$

where Y is output, K is aggregate capital, L is aggregate labor, and $z > 0$ is total factor productivity. The function F describes how physical inputs become output; z measures how effective the entire production process is. Later on we will introduce *factor-augmenting productivity* that affects a given input directly.

Formally, let $F : \mathbb{R}^2_+ \to \mathbb{R}_+$ be twice continuously differentiable and satisfy

(a) **Productive inputs:** Holding the other input fixed, more capital or labor raises output.

$$F_K(K, L) > 0, \qquad F_L(K, L) > 0.$$

(b) **Marginal products diminish:** Adding more machines or workers increases output, but ever less so.

$$F_{KK}(K, L) < 0, \qquad F_{LL}(K, L) < 0.$$

(c) **Constant returns to scale:** Doubling all inputs doubles output.

$$F(\lambda K, \lambda L) = \lambda F(K, L), \qquad \lambda > 0.$$

(d) **The production function is jointly concave.** We need this assumption to ensure the firm's problem is well posed mathematically.

Constant returns to scale and Euler's theorem. The fact that the production function has constant returns to scale has a very useful consequence that follows from Euler's Theorem for Homogeneous Functions: The production function can be expressed in terms of the marginal product of capital and labor.

$$\text{Euler's Theorem:} \quad F(K, L) \;=\; F_K(K, L)K \;+\; F_L(K, L)L. \tag{1.2}$$

The reason this identity is so useful is that when markets are competitive we can express output in terms of the cost of inputs. In a competitive market, factors receive the value of their marginal products, so $r = pzF_K$ and $w = pzF_L$. So, if output sells at price p we can write

$$rK + wL \;=\; pzF_K(K, L)K + pzF_L(K, L)L \;=\; pz\left(F_K(K, L)K + F_L(K, L)L\right) \;=\; pY \tag{1.3}$$

Thus constant returns imply that all revenue is paid to factors according to their marginal products in competitive markets. In other words, profits are zero when the production function has constant returns to scale. So, the real question is not about profits, but about how labor and capital are remunerated. Going forward we will focus on how the firm chooses inputs, how this choice is affected by input prices, and how inputs are remunerated.

1.2. The firm's input choice

Even if the firm makes zero profits we can establish how it will optimally choose its inputs. We do this by solving the *cost minimization* problem. For this, let r be the rental price of one unit of capital and w the wage paid for one unit of labor. Suppose that the firm wants to produce $\overline{Y}$ at the lowest possible cost. Its problem is

$$\underbrace{C(\overline{Y}, r, w)}_{\text{Cost Function}} = \min_{K,L} rK + wL \qquad \text{subject to} \qquad zF(K, L) \geq \overline{Y}. \tag{1.4}$$

We can analyze this problem graphically by looking at the properties of the cost and the output constraint. Our graph deals with the choice of capital and labor and so those will be the axes. The cost function is linear in the inputs and so we can graph different levels of cost with what we call an *isocost* curve:

Definition 1.1 (Isocost curve). *It is the combinations of capital and labor (K, L) that deliver the same cost level $\overline{c}$:*

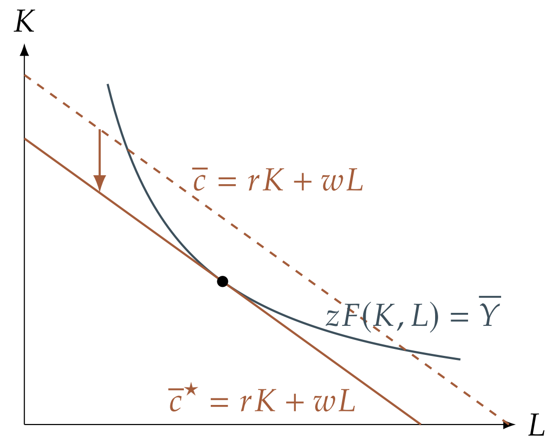
$$\overline{c} = rK + wL \qquad \longleftrightarrow \qquad K = \frac{\overline{c}}{r} - \frac{w}{r}L\,. \tag{1.5}$$

The isocost is a downward-sloping line because increasing either input raises the cost, and so the other input must go down to keep the cost constant. The slope of the isocost tells the rate at which the two inputs must be exchanged to keep the cost constant. That is, the rate at which we can exchange one input for another at market prices. The slope is therefore the relative price, in our case w/r.

The firm's objective is to choose the lowest isocost curve that satisfies the production constraint. That constraint is characterized by an *isoquant* curve:

Definition 1.2 (Isoquant curve). *It is the combinations of capital and labor (K, L) that deliver the*

Figure 1. Finding the lowest feasible isocost



Notes: The blue isoquant holds output fixed. The two red isocosts have the same slope because factor prices are unchanged. Moving to the lower isocost reduces expenditure until the isocost is tangent to the isoquant.

same output level $\overline{Y}$:

$$\overline{Y} = zF(K, L) \ . \tag{1.6}$$

The isoquant is a downward-sloping curve because increasing either input raises output (reflecting positive marginal products), and so the other input must go down to keep output constant. The shape of the isoquant depends on the properties of the production technology.

The graphical representation of the problem is in Figure 1. The firm will take as given the isoquant and move to lower isocost curves as much as possible.

1.3. Isoquants and the marginal rate of transformation (MRT)

How do we know when we have moved to the right combination of capital and labor? The answer comes from examining the cost minimization problem more closely. We will get to the graph in a moment, but first we stick with the mathematical derivations we have to characterize the defining property of the isoquant embodied in the *marginal rate of* *transformation (MRT)*. Taking a total differential of the constraint gives

$$F(K, L) = \overline{Y}/z \quad \longrightarrow \quad F_K \, dK + F_L \, dL = 0. \tag{1.7}$$

Therefore the slope of the isoquant is,

$$-\left.\frac{dK}{dL}\right|_Y = \frac{F_L}{F_K}. \tag{1.8}$$

This slope is a special concept for all the topics that follow because it directly measures how much you have to exchange capital and labor to keep output constant. The concept is so important that we give it a name.

Definition 1.3 (Marginal rate of transformation (MRT)). *The marginal rate of transformation of labor for capital is*

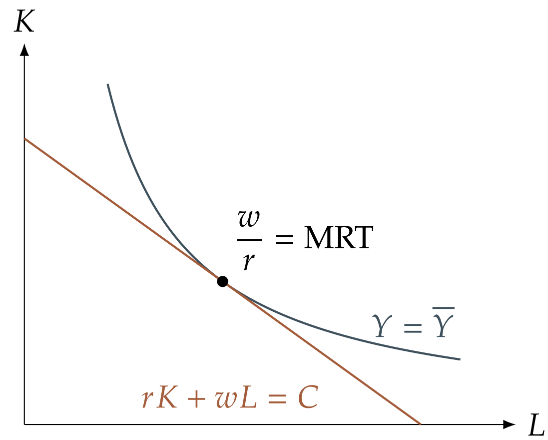
$$\text{MRT}_{L,K} \equiv \frac{F_L}{F_K}. \tag{1.9}$$

It is the amount of capital that can be removed when labor rises by one small unit while keeping output unchanged.

The MRT is a local concept. It answers a question about a small movement from the firm's current input mix. If $\text{MRT}_{L,K} = 2$, then near that point one additional unit of labor allows the firm to use approximately two fewer units of capital without changing output. In consumer theory the analogous object is called the marginal rate of substitution (the ratio of the marginal utilities).

Equation (1.23) now has a visual interpretation. An isocost line has slope $-w/r$ when K is on the vertical axis. At the cost-minimizing point, the isoquant has the same slope. If labor becomes more expensive relative to capital, the isocost becomes steeper and the firm moves toward a more capital-intensive technique.

Figure 2. Cost minimization



Notes: The firm chooses the tangency between the output isoquant and the lowest feasible isocost. At this point the marginal rate of transformation equals the relative factor price w/r.

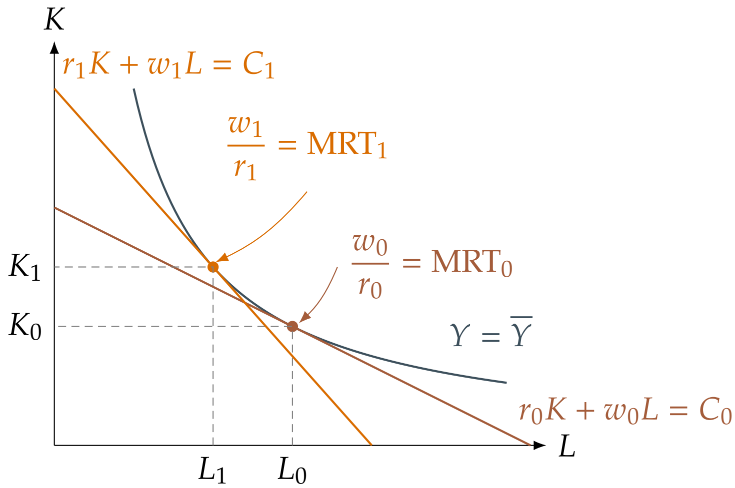
1.4. Substitution between capital and labor

We now get to the key question in this section (and really in all of the course): How substitutable are capital and labor? This turns out to be an incredibly deep question because it really depends on how we model the way in which labor is used for production. However, we can make progress now by asking a more specialized question:

If labor becomes 1 percent more expensive relative to capital,
by what percentage will a firm change its capital–labor ratio?

That might sound like a weird way to pose the question, but it really does speak to the way in which capital and labor are determined. To see it, let's return to Figure 2 and see what happens when labor becomes more expensive relative to capital. This can be because wages go up, or because the cost of capital goes down. In any case, the ratio $\frac{w}{r}$ goes up. This ratio is exactly the slope of the isocost curve, which is now steeper. So the new optimal combination of inputs involves less labor and more capital as in Figure 3. But, by how much?

Figure 3. A change in relative input prices



Notes: The blue isoquant holds output fixed. A higher w/r steepens the isocost and moves the optimum from (L_0, K_0) to (L_1, K_1). The firm uses less labor and more capital; the callouts show the initial and final tangency conditions.

The difficulty in giving a direct answer is that, while we intuitively understand that the firm will substitute labor for capital, we only know that the firm will do so while staying on the same isoquant. It does not tell us how far the firm will move along the isoquant when relative factor prices change. Two technologies can have the same MRT at the current input mix and the current input prices and very different responses to a higher wage or a cheaper machine. To distinguish them we need to measure *how the MRT changes as the input mix changes*.

The key to providing an answer is in the optimality conditions of the firm. Capital and labor will be always chosen so that the MRT equals the ratio of input prices $\left(MRT = \frac{w}{r}\right)$ as we see in Figure 3. We get this directly from (1.23). This gives you a direct link between changes in market prices and the properties of the firm's production technology. A change in the relative price must be matched by a change in the marginal rate of transformation. So, we can change our question to one about the production function:

If the MRT increases by 1 percent, by what percentage will the capital–labor ratio change?

This question has a direct mathematical formulation, and its answer is given in terms of the *elasticity of substitution*:

Definition 1.4 (Elasticity of substitution). *The Hicks elasticity of substitution between capital and labor is*

$$\sigma \equiv \frac{\partial \log(K/L)}{\partial \log \mathrm{MRT}_{L,K}} = \left.\frac{d \log(K/L)}{d \log(w/r)}\right|_{\overline{Y},z}. \tag{1.10}$$

Technology and output are held fixed; the second equality uses the firm's optimality condition $\mathrm{MRT} = w/r$.

A small change in the MRT rotates the tangent to the isoquant; the associated change in capital to output ratio, $k \equiv K/L$, tells us how much the technique changes. This "associated change" is the elasticity of substitution. The elasticity is a percentage response. If $\sigma = 2$, a one-percent increase in w/r leads the firm to raise K/L by approximately two percent. Labor has become more expensive relative to capital, and the firm responds strongly by using a more capital-intensive technique. If $\sigma = 0.2$, the same relative price change produces only a 0.2 percent change in K/L.

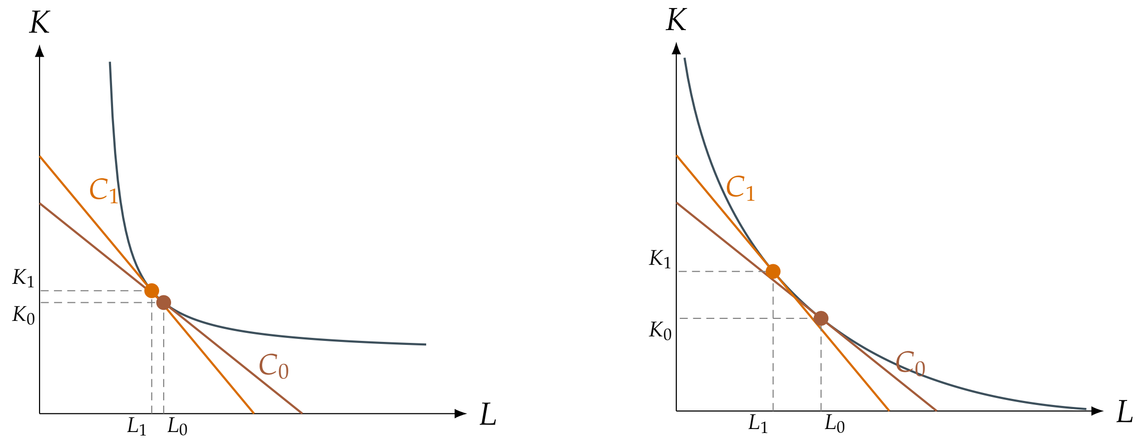
The MRT describes the technological tradeoff between capital and labor at a point.
The elasticity of substitution describes how the chosen input ratio changes when that tradeoff—and therefore the relative factor price—changes.

Three cases will be useful:

- If $\sigma > 1$, the input ratio responds more than proportionally. We will call capital and labor relatively easy to substitute.

Figure 4. Substitution under low and high elasticities

(a) Low elasticity: $\sigma = \frac{1}{2} < 1$ **(b) High elasticity:** $\sigma = 2 > 1$



Notes: The red initial isocost C_0 and orange final isocost C_1 have the same slopes in both panels. The low-elasticity technology produces the smaller change in the capital–labor ratio; the high-elasticity technology produces the larger change.

- If $\sigma = 1$, the input ratio responds proportionally. This is the value generated everywhere by the Cobb–Douglas technology.
- If $0 < \sigma < 1$, the input ratio responds less than proportionally. The inputs are harder to substitute than in the Cobb–Douglas benchmark.

Figure 4 contrasts two technologies: one with low elasticity, $\sigma = \frac{1}{2}$, and one with high elasticity, $\sigma = 2$. The panels use the same change in relative prices, so their isocosts have the same slopes. The outcome is that the same relative-price change produces different substitution patterns between capital and labor.

1.5. Payment to factors: How does inequality change between capital and labor?

Say capital becomes cheaper, for example because of advances in computation. Our previous results show this makes firms substitute away from labor and into capital, pursuing cost reductions. As this substitution happens, there is also a change in the payments to the factors of production. It might be that the substitution increases the payments to capital relative to those of labor. After all, capital demand does go up. But, does it go up enough? Remember that the firm is demanding more capital because it got cheaper. That reduction in the price of capital tends to reduce payments to capital.

What happens to the relative payments to capital and labor when capital becomes cheaper?

To answer this question we first introduce the concept of the *input share*.

Definition 1.5 (Input cost shares). *The share of cost going to each input:*

$$\alpha_K \equiv \frac{rK}{rK + wL} \qquad \alpha_L \equiv \frac{wL}{rK + wL}. \tag{1.11}$$

The input shares sum to one: $\alpha_K + \alpha_L = 1$.

When the technology has constant returns to scale total cost equals revenue and we also get input shares in total revenue or income. Recall from (1.3) that $pY = rK + wL$.

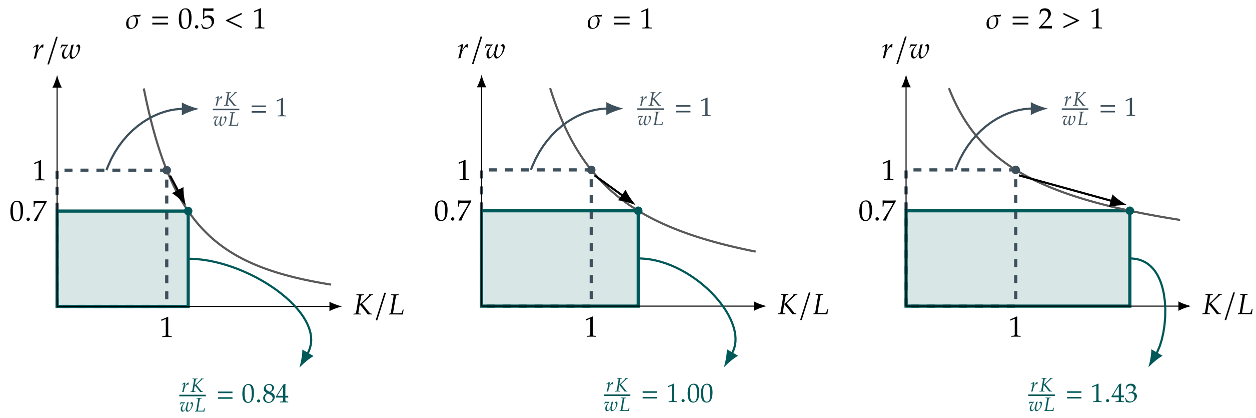
Our question asks how the relative income share of capital and labor changes with the relative price of labor, w/r. The key for this is noticing that the relative payments to capital and labor satisfy the accounting identity

$$\frac{\alpha_K}{\alpha_L} = \frac{rK}{wL} = \frac{K/L}{w/r}\,. \tag{1.12}$$

We can now answer questions about how income is divided between capital and labor. If the relative income share α_K/α_L goes up we know that α_K itself went up and α_L went down because they have to sum to 1.

Seeing relative payments as an area. Figure 5 gives the accounting identity a geometric interpretation. With K/L on the horizontal axis and r/w on the vertical axis, the rectangle from the origin to a chosen bundle has area $(r/w)(K/L) = rK/(wL)$. When capital becomes

Figure 5. Relative factor payments as rectangles: A drop in r/w and its effect on α_K/α_L



Notes: Illustrative constant-elasticity relative-demand curves $K/L = (w/r)^{\sigma}$ start at $K/L = r/w = 1$; the rental price falls to $r/w = 0.7$, holding output and technology fixed. The dashed blue rectangle shows the initial payment ratio; the shaded teal rectangle shows the ratio after capital becomes cheaper. Each rectangle measures $\frac{rK}{wL}$, so a larger area also means a higher capital share. From left to right, the rectangle area and capital share fall, remain unchanged, and rise.

cheaper, the rectangle gets shorter because r/w falls, but wider because the firm chooses more capital relative to labor. The question is whether its area increases or decreases. If substitution is weak, the price decline dominates and capital's relative payment falls; if substitution is strong, the quantity response dominates and its relative payment rises. At unit elasticity the two changes exactly offset. These are relative payments, not the level of total spending on either input.

A price change: does substitution offset the higher price?. Suppose w/r rises. This raises demand for capital relative to labor, but does it also raise expenditure on capital relative to labor? We can use the accounting identity above, take logarithms and differentiate to obtain the answer:

$$d\log\frac{\alpha_K}{\alpha_L} \;=\; d\log\frac{K}{L} - d\log\frac{w}{r} \;=\; (\sigma-1)\,d\log\frac{w}{r}. \tag{1.13}$$

Whether the relative expenditure in capital increases or not depends on the elasticity of substitution. When w/r rises there is a direct price effect in favor of labor's share, but a substitution effect toward capital. If $\sigma < 1$, quantities adjust less than prices and labor's cost share rises. If $\sigma > 1$, the quantity response dominates and labor's cost share falls. At $\sigma = 1$, the two effects exactly offset and input shares remain constant.

An endowment change: does more capital raise capital's share?. Now consider a different experiment: an economy with a fixed, constant-returns technology, competitive factor markets, and fully employed factor supplies. Increase K/L and allow w/r to adjust to marginal products. Under constant returns to scale the ratio of capital to labor determines the MRT and the elasticity of substitution. We can therefore use the same results we have derived so far but reason in the opposite direction. We obtain

$$d\log(w/r) \;=\; \frac{1}{\sigma}\, d\log(K/L), \qquad d\log\frac{rK}{wL} \;=\; \left(1-\frac{1}{\sigma}\right) d\log(K/L). \tag{1.14}$$

More capital per worker raises capital's income share if $\sigma > 1$, leaves it locally unchanged if $\sigma = 1$, and lowers it if $\sigma < 1$. The growing factor earns less per unit relative to the other factor; the elasticity tells us whether this relative-price decline outweighs its greater quantity. Here cost shares are also shares of output revenue because constant returns and competitive marginal-product payments imply $rK + wL = pY$.

1.6. Input shares and input responses

The results we just derived give us a way to see how the combination of capital and labor responds optimally to changes in market prices. But we also want to know how capital and labor respond individually. The elasticity of substitution and the income shares determine how strong the response is of the *levels* of each input.

To see this we start from the output constraint that defines the isoquant in (1.7). We take advantage of the fact that $\log x = \frac{dx}{x}$ to write

$$KF_K\, d\log K \;+\; LF_L\, d\log L \;=\; 0 \longrightarrow \quad rK\, d\log K \;+\; wL\, d\log L \;=\; 0 \tag{1.15}$$

$$\begin{aligned}
&\longrightarrow \quad \frac{rK}{rK + wL}\, d\log K \;+\; \frac{wL}{rK + wL}\, d\log L \;=\; 0 \\
&\longrightarrow \quad \alpha_K\, d\log K \;+\; \alpha_L\, d\log L \;=\; 0
\end{aligned}$$

The only issue outstanding is that we need a way to connect K and L. The elasticity of substitution does that for us. From its definition and the optimality condition of the firm we write

$$d\log K - \; d\log L \;=\; \sigma\, d\log \mathrm{MRT} \;=\; \sigma\, d\log \frac{w}{r}. \tag{1.16}$$

This equation gives a link between the percentage change in capital and labor (approximated by the change in their logs). The final step is using it to get isolated expressions for the response of each input. For labor we have

$$\begin{aligned}
\alpha_K\, d\log K \;+\; \alpha_L\, d\log L \;=\; 0 \longrightarrow \quad & \alpha_K\, d\log K \;+\; (1-\alpha_K)\; d\log L \;=\; 0 \\
\longrightarrow \quad & \alpha_K \left(\sigma\, d\log \frac{w}{r} \;+\; d\log L\right) \;+\; (1-\alpha_K)\; d\log L \;=\; 0 \\
\longrightarrow \quad & \alpha_K \sigma\, d\log \frac{w}{r} \;+\; d\log L \;=\; 0 \\
\longrightarrow \quad & d\log L \;=\; -\alpha_K \sigma\, d\log \frac{w}{r}
\end{aligned} \tag{1.17}$$

Similar steps deliver

$$d\log K \;=\; \alpha_L \sigma\, d\log \frac{w}{r} \tag{1.18}$$

These are local responses at fixed output and technology. For example, if capital's cost share is 0.4 and $\sigma = 0.5$, a one-percent rise in the wage with the rental price fixed reduces conditional labor demand by approximately 0.2 percent and raises conditional capital demand by approximately 0.3 percent. The ratio rises by 0.5 percent.

The responses depend on the input share of the other input, or in other words, on the complement input shares. The higher the input share of the input, the less scope the firm has to substitute into it. Similarly, the larger the input share of the other inputs the more substitutability there can be.

1.7. Solving the firm's problem: Lagrangian approach

We solved the problem graphically, but we can also tackle the problem directly using constrained optimization. The Lagrangian of the cost minimization problem is

$$\mathcal{L} = rK + wL + \lambda \left(\overline{Y} - zF(K, L) \right). \tag{1.19}$$

The first-order conditions are

$$r = \lambda z F_K(K, L), \tag{1.20}$$

$$w = \lambda z F_L(K, L), \tag{1.21}$$

$$\overline{Y} = zF(K, L). \tag{1.22}$$

The multiplier λ is marginal cost: by the envelope theorem, $\lambda = \frac{\partial C(\overline{Y}, r, w)}{\partial \overline{Y}}$.

The solution is characterized by the ratio of the first two conditions, which we obtain by dividing (1.21) by (1.20):

$$\frac{F_L(K, L)}{F_K(K, L)} = \frac{w}{r}. \tag{1.23}$$

Under constant returns, the first-order conditions determine factor proportions but not a unique scale.

1.8. Substitutability cheatsheet

The tables summarize input choice and factor payments at a fixed technology $Y = zF(K, L)$. Unless stated otherwise, assume positive factor prices, a smooth constant-returns technology, and an interior cost minimum. Price responses hold output and technology fixed; endowment responses let factor prices adjust to competitive marginal-product payments. All derivatives are local: σ and the shares are evaluated at the initial allocation.

Object	**Mathematical expression**	**Meaning**
Production and constant returns	$Y = zF(K, L)$ $F(tK, tL) = tF(K, L),\ t > 0$	Scaling both inputs by t scales output by t.
Cost function	$C(\overline{Y}, r, w) = \min_{K,L}\{rK + wL\}$ subject to $zF(K, L) \geq \overline{Y}$	Minimum expenditure needed for the target output.
Isocost	$K = \frac{\overline{c}}{r} - \frac{w}{r}L$	Equal-cost bundles; slope $-w/r$ with K on the vertical axis.
Isoquant	$zF(K, L) = \overline{Y}$	Input bundles producing the same output.
MRT and tangency	$\mathrm{MRT}_{L,K} = \frac{F_L}{F_K} = -\left.\frac{dK}{dL}\right\|_{\overline{Y},z}$ $\mathrm{MRT}_{L,K} = \frac{w}{r}$ at the optimum	Capital saved per extra unit of labor; equals w/r at the optimum.
Substitution elasticity	$\sigma = \left.\frac{d\log(K/L)}{d\log \mathrm{MRT}_{L,K}}\right\|_{\overline{Y},z}$ $d\log(K/L) = \sigma\, d\log(w/r)$	Percentage change in K/L per one-percent relative-price change.
Input cost shares	$\alpha_K = \frac{rK}{rK + wL}, \quad \alpha_L = \frac{wL}{rK + wL}$ $\alpha_K + \alpha_L = 1$	Fractions of total cost paid to capital and labor.
Euler and factor payments	$F = KF_K + LF_L$ $r = pzF_K, \quad w = pzF_L$ $rK + wL = pY$	Factor payments exhaust revenue under competition and constant returns.
Individual input responses	$d\log K = \alpha_L \sigma\, d\log(w/r)$ $d\log L = -\alpha_K \sigma\, d\log(w/r)$	Input responses at fixed output and technology.
Relative payments	$\frac{\alpha_K}{\alpha_L} = \frac{rK}{wL} = \frac{K/L}{w/r}$	Relative expenditure combines a quantity ratio and a price ratio.
Marginal cost	$\lambda = \frac{\partial C(\overline{Y}, r, w)}{\partial \overline{Y}}$ $r = \lambda z F_K, \quad w = \lambda z F_L$	Extra cost of a marginal unit of output.

Main questions and answers. A higher w/r means labor is more expensive relative to capital; equivalently, capital is relatively cheaper. For changes in the opposite direction, reverse the local response signs.

Question and what is fixed	Relevant result	Answer
How does the input mix respond to a rise in w/r? Output and technology fixed.	$\dfrac{d\log(K/L)}{d\log(w/r)} = \sigma$	K/L rises. A larger σ means a stronger percentage response.
What happens to the amounts of each input? Output and technology fixed.	$\dfrac{d\log K}{d\log(w/r)} = \alpha_L\sigma$ $\dfrac{d\log L}{d\log(w/r)} = -\alpha_K\sigma$	Capital rises and labor falls. These are fixed-output responses; sales expansion is excluded.
Does cheaper capital raise its income share? Technology fixed; input mix adjusts.	$\dfrac{d\log(\alpha_K/\alpha_L)}{d\log(w/r)} = \sigma - 1$	Yes if $\sigma > 1$; its share falls if $\sigma < 1$ and is locally unchanged if $\sigma = 1$. Quantity substitution competes with the lower relative price.
Does more capital per worker raise its relative return? Technology fixed; factor prices adjust.	$\dfrac{d\log(w/r)}{d\log(K/L)} = \dfrac{1}{\sigma}$	No. The wage rises relative to the rental rate, so capital's relative return r/w falls.
Does more capital per worker raise capital's income share? Technology fixed; factor prices adjust.	$\dfrac{d\log(\alpha_K/\alpha_L)}{d\log(K/L)} = 1 - \dfrac{1}{\sigma}$	Yes if $\sigma > 1$; no change locally if $\sigma = 1$; it falls if $\sigma < 1$. More capital is offset by a lower relative return.
Does tangency determine how much the firm produces? Constant returns.	$\dfrac{F_L}{F_K} = \dfrac{w}{r}$ Input ratio condition.	The output target determines scale. Tangency alone determines only the input ratio.
What if inputs must be used in fixed proportions? Leontief case discussed below; positive prices.	$Y = z\min\left\{\dfrac{K}{a_K}, \dfrac{L}{a_L}\right\}$ $K/L = a_K/a_L, \quad \sigma = 0$	Quantities stay fixed; expenditure shares change. At the kink, smooth tangency and $1/\sigma$ formulas do not apply.

1.9. Historical background: John Hicks

John R. Hicks (1904–1989)

Hicks was a British economist who began his Oxford studies in mathematics before moving to philosophy, politics, and economics. At the London School of Economics he started with descriptive work on labor and industrial relations, then turned increasingly to economic theory. He later worked at Cambridge, Manchester, and Oxford, and shared the 1972 economics Nobel Prize with Kenneth Arrow for work on general equilibrium and welfare theory.

His work connects directly to our questions. In *The Theory of Wages* (1932), substitution and technical change help explain the division of income between labor and capital. As Benjamin Moll's overview shows, Hicks also helped develop compensated demand, the distinction between income and substitution effects, and the IS–LM model. These became standard tools across microeconomics and macroeconomics.

Sources: Hicks's Nobel autobiography (*Nobel Lectures*, Economics 1969–1980); Benjamin Moll, About John Hicks (with links to the original works).

The concept of the elasticity of substitution was introduced by John Hicks in 1932 in his book *The Theory of Wages*. Hicks's motivation lay in questions over the distribution of income between labor and capital and how technological change affects this distribution. His questions are not that different from the ones we have now. Hicks asked

> "Is economic progress likely to raise or lower the proportion of the National Dividend which goes to labour?" Hicks (1932, p.113), Chapter VI, printed p. 113 (PDF p. 126).

He goes on to show that the answer depends on the elasticity of substitution:

> “An increase in the supply of any factor will increase its relative share (i.e., its proportion of the National Dividend) If its "elasticity of substitution" is greater than unity [...] The "elasticity of substitution" is a measure of the ease with which the varying factor can be substituted for others. If the same quantity of the factor is required to give a unit of the product, in any circumstances whatever, then its elasticity of substitution is zero. If all the factors employed are for practical purposes identical, so that the varying factor can be substituted for any co-operating factor without any trouble at all, then the elasticity of substitution is infinite. The case where the elasticity of substitution is unity can only be defined in words by saying that in this case (initially, before any consequential changes in the supply of other factors takes place) the increase in one factor will raise the marginal product of all other factors taken together in the same proportion as the total product is raised.”

What did Hicks have in mind when he was thinking about substitutability?

> “Substitution, in the sense in which we are using it, may take any of three forms:
>
> (a) The change in the relative prices of the factors may lead simply to a shift over from the production of things requiring little of the increasing factor to things requiring more. If capital increases, the commodities in whose production capital had already been used to an extent above the average will become cheaper relatively to others, and presumably, therefore, more of them will be made.

(b) Methods of production already known, but which did not pay previously, may come into use. This form will include, possibly as its most important case, the mere extension of the use of instruments and methods of production from firms where they were previously employed to firms which could not previously afford them.

(c) The changed relative prices will stimulate the search for new methods of production which will use more of the now cheaper factor and less of the expensive one.

Partly, therefore, substitution takes place by a change in the proportions in which productive resources are distributed among existing types of production. But partly it takes place by affording a stimulus to the invention of new types. We cannot really separate, in consequence, our analysis of the effects of changes in the supply of capital and labour from our analysis of the effects of invention."

So, Hicks is also concerned by how innovation can lead to substitution between capital and labor in a way that reduces the share of labor in total income.

"[...] although an invention must increase the total Dividend, it is unlikely at the same time to increase the marginal products of all factors of production in the same ratio [...] then we can classify inventions according as their initial effects are to increase, leave unchanged, or diminish the ratio of the marginal product of capital to that of labour. [The] predominance of labour-saving inventions strikes one as curious. [The reason] is surely that

> which was hinted at in our discussion of substitution. A change in the relative prices of the factors of production is itself a spur to invention, and to invention of a particular kind-directed to economising the use of a factor which has become relatively expensive. [...] where invention is very active, the elasticity of substitution will be high and will remain high. Thus, the relative share of capital will tend to increase; and that of labour to fall."

The conclusion that Hicks arrives at in the 1930s is worrisome, even as it did not come to pass, and it echoes the worries of today as artificial intelligence is developed:

> "[...] a fall in the general level of real wages is really likely to occur as the result of invention only on those rare occasions when invention breaks into a new and extensive field of industry that has long been conservative in its methods. Such "economic revolutions" always cause maladjustment, and social unrest arising from the maladjustment; but it may be useful to point out that in such times the malaise may go deeper. A fall in the equilibrium level of real wages is here a real possibility."

However, Hicks goes on to dispel some of those worries:

> "But it is difficult to feel that this danger is a very pressing one today. The generalised character of technical change is a considerable safeguard against it. Inventive activity usually makes itself felt quickly enough, so that a prolonged failure to adjust technical methods to new circumstances

> is unlikely on a large scale. Our continuous "industrial revolution" protects us from the discontinuous revolutions of the past."

The rest of this course looks at these questions through modern techniques.

1.10. Historical background: Wassily Leontief

Finally, it is worth mentioning the work of Wassily Leontief on production networks. His work greatly influenced how we think about the way production is organized in the economy. From the point of view of this course, his most salient contribution is what we call the Leontief function which captures perfect complements. This function implies an elasticity of substitution of zero ($\sigma = 0$) so that the ratio of inputs does not change with prices. This key feature means that the ratio of inputs is constant and given entirely as a property of production.

Wassily W. Leontief (1906–1999)

Leontief was born in St. Petersburg, studied at the University of Leningrad and the University of Berlin, and moved to the United States in 1931. At Harvard he turned the interdependence of industries into an empirical system of input–output accounts: tables that record which industries supply inputs to other industries and which industries use them. This made the production network of an entire economy measurable rather than leaving it as an abstract general-equilibrium idea.

Leontief received the 1973 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel for developing the input–output method and applying it to important economic problems. His framework became part of modern economic accounting and remains useful for tracing the direct and indirect consequences of changes in demand, technology, trade, and production bottlenecks.

Sources: Nobel Prize, Wassily Leontief facts; U.S. Bureau of Economic Analysis, The Development of BEA's Input–Output Framework.

The Leontief production function. The firm-level version of his fixed-coefficient logic is the Leontief production function

$$Y \;=\; z\min\left\{\frac{K}{a_K},\frac{L}{a_L}\right\}, \qquad a_K > 0, \quad a_L > 0. \tag{1.24}$$

The coefficients a_K and a_L describe a technological recipe: before the productivity multiplier z, one unit of production requires a_K units of capital together with a_L units of labor. Additional capital cannot replace missing labor, and additional labor cannot replace missing capital. One input becomes a bottleneck while any amount of the other input beyond the required proportion does not raise output.

For positive input prices, a firm producing $\overline{Y}$ therefore chooses

$$K \;=\; \frac{a_K\overline{Y}}{z}, \qquad L \;=\; \frac{a_L\overline{Y}}{z}, \qquad C(\overline{Y},r,w) \;=\; \frac{ra_K + wa_L}{z}\overline{Y}. \tag{1.25}$$

The required input ratio is consequently $K/L = a_K/a_L$, independent of w/r. Consequently, the elasticity of substitution is zero:

$$d\log(K/L) \;=\; \underbrace{0}_{\sigma}\, d\log(w/r) \;=\; 0. \tag{1.26}$$

This extreme case makes our earlier comparative statics transparent. At fixed output and technology, a change in relative input prices changes neither conditional labor demand nor conditional capital demand, consistent with (1.18) when $\sigma = 0$. Only expenditures change: (1.13) becomes

$$d\log\frac{\alpha_K}{\alpha_L} \;=\; -\,d\log\frac{w}{r}. \tag{1.27}$$

There is no quantity substitution to offset the direct price effect. For example, cheaper machines lower the cost of the capital required by the recipe, but they do not cause the firm to replace workers with more machines unless the recipe itself changes.

From a firm to a production network. Leontief applied the same fixed-requirements logic to the links among industries. Let x be the vector of gross outputs, let d be final demand, and let the element a_{ij} of the matrix A be the amount of input from industry i required per unit of output in industry j. Each industry's output must cover both intermediate demand from other industries and final demand:

$$x = Ax + d \qquad \longrightarrow \qquad x = (I - A)^{-1} d, \tag{1.28}$$

The *Leontief inverse* $(I - A)^{-1}$ adds the direct and indirect production requirements generated throughout the network. A rise in demand for automobiles, for example, requires not only more automobile production but also more steel, electricity, transportation, and the inputs used to produce each of those inputs.

Input–output tables were used for wartime and postwar planning and became a core component of national economic accounts. Statistical agencies still use supply, use, and requirements tables to measure transactions across industries, while economists use them to study supply-chain propagation, structural change, productivity, trade, environmental demands, and the economy-wide effects of sector-specific shocks. The simplification is also its limitation: the coefficients in A treat production recipes as fixed, so the basic model deliberately sets aside the price-induced substitution that is central to the rest of this course.

I cover input-output networks in another course. Click here for my notes.

2. The Cobb–Douglas technology

The Cobb–Douglas production function is one of the most common functional forms for the production function we use in economics. It is

$$Y = zK^{\alpha}L^{1-\alpha}, \qquad 0 < \alpha < 1. \tag{2.1}$$

However, it is also a restrictive function because it implies that the elasticity of substitution is always equal to one. To see this, consider the marginal products of capital and labor:

$$F_K = \alpha K^{\alpha-1}L^{1-\alpha} = \alpha\frac{Y}{K} \qquad\qquad F_L = (1-\alpha)K^{\alpha}L^{-\alpha} = (1-\alpha)\frac{Y}{L}. \tag{2.2}$$

The MRT is therefore

$$\mathrm{MRT}_{L,K} = \frac{F_L}{F_K} = \frac{1-\alpha}{\alpha}\frac{K}{L}. \tag{2.3}$$

At the cost-minimizing input mix, the MRT must equal w/r. So, we have,

$$\log\frac{K}{L} = \log\frac{\alpha}{1-\alpha} + \log\frac{w}{r}. \tag{2.4}$$

The capital-to-labor ratio is proportional to the relative price of labor and capital. A one-percent increase in w/r leads to a one-percent increase in K/L, regardless of the firm's initial position. From this we see directly that the elasticity of substitution is constant and equal to one:

$$\sigma = \frac{d\log(K/L)}{d\log \mathrm{MRT}_{L,K}} = 1. \tag{2.5}$$

There is another implication (or restriction) of the Cobb–Douglas technology: The input shares are constant. To see this, we can use the optimality condition to write

$$rK = \frac{\alpha}{1-\alpha}wL. \tag{2.6}$$

With this we directly get the input shares:

$$\alpha_K \equiv \frac{rK}{rK + wL} = \frac{\frac{\alpha}{1-\alpha} wL}{\frac{\alpha}{1-\alpha} wL + wL} = \frac{\alpha}{1 - \alpha + \alpha} = \alpha, \tag{2.7}$$

$$\alpha_L \equiv \frac{wL}{rK + wL} = \frac{\frac{1-\alpha}{\alpha} rK}{rK + \frac{1-\alpha}{\alpha} rK} = \frac{1 - \alpha}{1 - \alpha + \alpha} = 1 - \alpha. \tag{2.8}$$

The fact that the input shares are constant is intimately connected to the fact that the elasticity of substitution is constant and equal to one. Recall from our discussion of input shares that

$$d \log \frac{rK}{wL} = (\sigma - 1)\, d \log \frac{w}{r}. \tag{2.9}$$

This is the same as (1.13). When $\sigma = 1$, there is no change in the relative payments to capital and labor as prices change. Similarly, we have that the change in the demand for labor and capital is proportional to the change in relative prices, with the proportionality constant given by the input share of the other input:

$$d \log L = -\alpha\, d \log \frac{w}{r} \qquad d \log K = (1 - \alpha)\, d \log \frac{w}{r}. \tag{2.10}$$

2.1. The direction of technical change

So far, technological change has been summarized by z, which we call *total factor productivity* (TFP). A higher z raises output from any given combination of capital and labor by the same proportion. It is *neutral*: it raises both marginal products proportionally without changing the MRT at a given input mix. But do all technological improvements affect capital and labor in the same way?

Better organizational practices can help workers spend less time coordinating and more time producing, making each hour of labor more productive. Advances in information technology can make computers and robots faster or more accurate, increasing the services provided by each unit of capital. We describe these possibilities as *factor-augmenting* *technological change*. Let A_L measure the productivity of labor and A_K the productivity of capital: $A_L L$ and $A_K K$ are the corresponding effective inputs. For example, a 10 percent increase in A_L lets the same workers provide 10 percent more effective labor, without increasing their hours. These are useful aggregate descriptions of changes in how particular tasks are performed. In Section 5, we will examine those tasks directly and distinguish making an input better at its existing tasks from changing which tasks it performs.

Can Cobb–Douglas capture the distinction between a technology that augments capital and one that augments labor? To isolate the two factor-augmenting terms, initially set the separate neutral multiplier to one and write

$$Y = (A_K K)^{\alpha}(A_L L)^{1-\alpha} = \underbrace{A_K^{\alpha} A_L^{1-\alpha}}_{z} K^{\alpha} L^{1-\alpha}. \tag{2.11}$$

Thus, defining $z = A_K^{\alpha} A_L^{1-\alpha}$ gives exactly our original production function. An increase in either factor-augmenting technology is therefore equivalent to an increase in TFP.

This restriction matters for both input choices and factor rewards. The technology terms cancel from the MRT, so at given factor prices the optimal ratio remains

$$\frac{K}{L} = \frac{\alpha}{1-\alpha}\frac{w}{r}. \tag{2.12}$$

Neither A_K nor A_L induces a change in the chosen physical capital–labor ratio when w/r is held fixed. For a fixed output target, either improvement lets the firm reduce both inputs in the same proportion. This does not mean input demand cannot change: output expansion can also affect how much capital and labor the firm uses.

At fixed quantities of capital and labor, either improvement raises output and both marginal products by the same percentage. Under competitive factor pricing, the real rewards are $r/p = \alpha Y/K$ and $w/p = (1-\alpha)Y/L$, so both rise proportionally as well. The gains are shared in the existing proportions α and $1-\alpha$; they are not necessarily split equally. Even a technology that directly augments capital benefits labor in this sense, and neither factor's income share changes.

Cobb–Douglas therefore rules out technology-driven changes in the input mix at fixed factor prices and in the division of income between capital and labor. We want a model that allows these possibilities when studying technologies that change what workers and machines can do. Section 3 introduces the CES production function, where the elasticity of substitution determines whether a factor-augmenting improvement raises or lowers the relative use and income share of that factor.

2.2. Substitution and scale effects

So far we have held output fixed. That isolates substitution, but it leaves out another important response: cheaper production can lower the output price, increase sales, and raise the scale of production. We now add that channel (and will return to it later on in the course). Suppose the market has isoelastic product demand

$$Y = Bp^{-\varepsilon}, \qquad \varepsilon > 0, \tag{2.13}$$

The key question is:

What happens to the demand for inputs when capital becomes cheaper?

To answer this question we need to know how the output price responds to changes in input prices. We derive the unit cost from the cost minimization problem in (1.4):

$$C(Y; r, w) = Y \underbrace{\frac{1}{z}\left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}}_{\text{per-unit cost of production } c(r,w)}. \tag{2.14}$$

Total cost is proportional to output, so $c(r, w) = C(1; r, w) = C(Y; r, w)/Y$.

We now hold w, z, α, and the demand parameters fixed and change r. Taking logs of unit cost makes its response transparent:

$$\frac{\partial \log c}{\partial \log r} = \alpha. \tag{2.15}$$

Since $p = c(r, w)$, market demand becomes $Y = Bc(r, w)^{-\varepsilon}$. Taking logs and applying the chain rule gives

$$\log Y = \log B - \varepsilon \log c(r, w) \quad \longrightarrow \quad \frac{\partial \log Y}{\partial \log r} = -\varepsilon \frac{\partial \log c}{\partial \log r} = -\varepsilon\alpha. \tag{2.16}$$

A one-percent fall in r therefore lowers unit cost and the output price by approximately α percent, increasing output by $\varepsilon\alpha$ percent.

Finally, the optimality condition for labor coming from cost minimization gives

$$w = (1-\alpha)c\frac{Y}{L} \quad \longrightarrow \quad \log L = \log(1-\alpha) + \log c + \log Y - \log w. \tag{2.17}$$

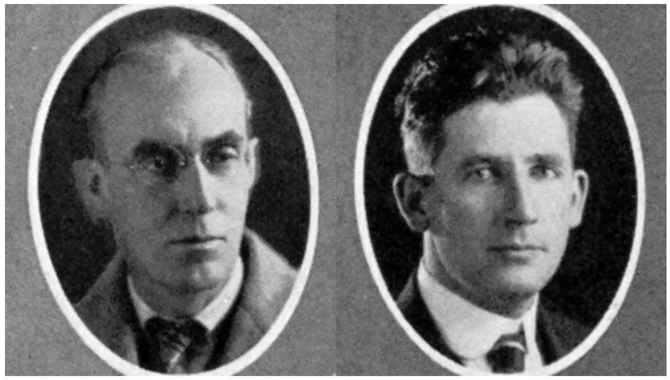
Since w is fixed, differentiating gives the total labor response:

$$\frac{\partial \log L}{\partial \log r} = \frac{\partial \log c}{\partial \log r} + \frac{\partial \log Y}{\partial \log r} = \underbrace{\alpha}_{\text{substitution}} - \underbrace{\varepsilon\alpha}_{\text{scale}} = \alpha(1-\varepsilon). \tag{2.18}$$

The first term is the substitution effect derived above: at a fixed level of output, cheaper capital reduces labor demand. The second term is the scale effect: cheaper production expands output and raises demand for inputs. If $\varepsilon > 1$, the scale effect is larger, so a fall in r raises labor demand. If $\varepsilon < 1$, substitution dominates and labor demand falls.

2.3. Historical background: Cobb–Douglas and factor shares

Charles Wiggins Cobb and Paul Howard Douglas



Charles Wiggins Cobb (1875–1949) was an American mathematician and Amherst College graduate who taught there from 1906 to 1948. He brought mathematical expertise to the collaboration with Douglas; their 1928 paper joined a tractable production function to evidence on U.S. manufacturing. Amherst's archive preserves his manufacturing data and reports alongside his mathematical work.

Source: Amherst College Archives.

Paul Howard Douglas (1892–1976) was an American economist and professor of industrial relations at the University of Chicago. His interest in wages, production, and factor incomes motivated the empirical question behind their joint work. He later represented Illinois in the U.S. Senate (1949–1967), where he chaired the Joint Economic Committee. His career connected economic research with public service.

Sources: U.S. Congress biography; Douglas (1967, pp. 15–16).

Where did the Cobb–Douglas production function come from? It began as an attempt to measure the relationship between production and its inputs. The economist Paul Howard Douglas (1892–1976) from the University of Chicago (later a U.S. Senator) was interested in the regularity of the aggregate labor and capital shares of income in the U.S.. During a 1927 visit to Amherst, he showed the data to Charles Wiggins Cobb (1875–1949), a mathematician. Douglas was looking for a mathematica function that would be consistent with the patterns in the data. Their collaboration let to the 1928 paper Cobb and Douglas (1928), which introduced the Cobb–Douglas production function. See also Amherst College's Charles Wiggins Cobb archival record and the U.S. Congress biography of Paul Howard Douglas.

Douglas's own account of Cobb's mathematical role is in Douglas (1967, pp. 15–16).

The fact that the labor and capital shares of income were approximately constant over time turned out to be quite consequential. Within the competitive two-input model, stable factor shares despite changing relative factor prices are consistent with an elasticity of substitution near one, holding technology fixed. Stable shares alone do not identify the elasticity when technology and other determinants of factor payments also change. A decade after the Cobb–Douglas paper, Keynes (1939) described the stability of labor's share as "*one of the most surprising, yet best-established, facts in the whole range of economic statistics*." The question at that time is the same one we have just studied: how can the division of income remain stable while output, input quantities, and prices change?

Nevertheless, the labor share has declined across many countries and industries since the 1980s. Karabarbounis and Neiman (2014) document this broad pattern and estimate that falling relative investment prices account for roughly half of the decline in their model. Their explanation relies on an elasticity of substitution above one: cheaper capital raises its use enough to increase its share of income. Karabarbounis (2024) provides a recent survey of the evidence and competing explanations.

The task-based approach we will look at later in the course offers a different mechanism. Acemoglu and Restrepo (2018) develop a model in which automation transfers tasks from labor to capital and reduces the labor share, while new tasks can reinstate labor. Acemoglu and Restrepo (2019) connect these mechanisms to evidence on changes in the task content of production. These papers help us distinguish automation from capital becoming cheaper or more productive at tasks it already performs.

Changes in the allocation of production across firms also matter. Autor et al. (2020) find evidence consistent with sales shifting toward highly productive "superstar" firms with low labor shares. In their account, the aggregate share can fall through reallocation toward these firms, without a comparable decline at the typical firm. This cautions against treating all nonlabor income as the competitive return to capital.

3. CES Production and Technological Change

The Cobb–Douglas production function gives us a clean benchmark to study the many issues, but it is too restrictive to study those that depend on the elasticity of substitution. What if capital and labor are easier, or harder, to substitute? We now extend the analysis to a production function that allows for a general elasticity of substitution, but that remains easy to use by imposing that the elasticity of substitution is constant. Accordingly, we call it the *constant-elasticity-of-substitution* (CES) production function. Extending our analysis to the CES production function will also allow us to study the effects of technological change on the relative factor shares of income as we can now distinguish between technological change geared towards a specific factor (this is the *directed technical change* that Hicks referred to).

The CES production function is

$$Y = z\left(\alpha\left(A_K K\right)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\left(A_L L\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \qquad \sigma > 0, \quad \sigma \neq 1. \tag{3.1}$$

where $A_K > 0$ and $A_L > 0$ are factor-augmenting technologies. When $\sigma = 1$ we have the Cobb-Douglas function.

While the function looks more complicated than Cobb–Douglas, it is still easy to use once we realize how to exploit the fact that it has constant returns to scale. The marginal product of capital is

$$\frac{\partial Y}{\partial K} = z\frac{\sigma}{\sigma-1}\left(\alpha\left(A_K K\right)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\left(A_L L\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} \frac{\sigma-1}{\sigma}\alpha A_K^{\frac{\sigma-1}{\sigma}} K^{-\frac{1}{\sigma}}, \tag{3.2}$$

We can greatly simplify this expression by noticing that

$$\frac{1}{\sigma-1} = \left(\frac{\sigma}{\sigma-1}\right)\frac{1}{\sigma}. \tag{3.3}$$

The bracket raised to $\frac{\sigma}{\sigma-1}$ equals Y/z. Keeping the factor z outside the bracket, we therefore obtain

$$\frac{\partial Y}{\partial K} = z\left[\left(\alpha\,(A_K K)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\,(A_L L)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}\right]^{\frac{1}{\sigma}} \alpha A_K^{\frac{\sigma-1}{\sigma}} K^{-\frac{1}{\sigma}}, \tag{3.4}$$

$$= z\left(\frac{Y}{z}\right)^{\frac{1}{\sigma}} \alpha A_K^{\frac{\sigma-1}{\sigma}} K^{-\frac{1}{\sigma}}$$

$$= z^{\frac{\sigma-1}{\sigma}} \alpha A_K^{\frac{\sigma-1}{\sigma}} \left(\frac{K}{Y}\right)^{-\frac{1}{\sigma}}. \tag{3.5}$$

Similarly, the marginal product of labor is

$$\frac{\partial Y}{\partial L} = z^{\frac{\sigma-1}{\sigma}}(1-\alpha) A_L^{\frac{\sigma-1}{\sigma}} \left(\frac{L}{Y}\right)^{-\frac{1}{\sigma}}. \tag{3.6}$$

Finally, the MRT is

$$\mathrm{MRT}_{L,K} = \frac{F_L}{F_K} = \frac{1-\alpha}{\alpha}\left(\frac{A_L}{A_K}\right)^{\frac{\sigma-1}{\sigma}}\left(\frac{L}{K}\right)^{-\frac{1}{\sigma}}. \tag{3.7}$$

We can then get the elasticity of substitution by inverting this equation and applying logarithms:

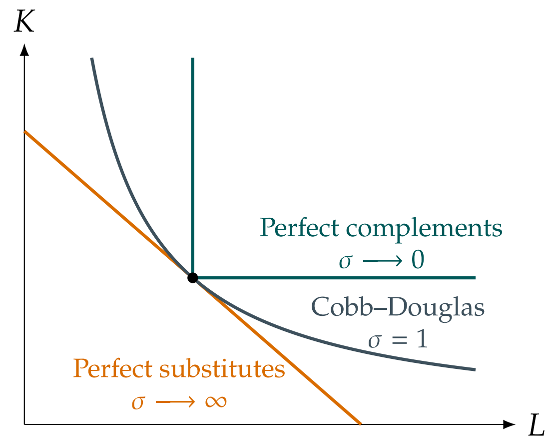
$$\log\frac{K}{L} = \sigma\log\frac{\alpha}{1-\alpha} + (\sigma-1)\log\frac{A_K}{A_L} + \sigma\log\mathrm{MRT}_{L,K}\,. \tag{3.8}$$

The elasticity of substitution is therefore constant and equal to σ. Three limiting cases help us interpret it. Figure 6 compares their isoquants:

(a) $\boldsymbol{\sigma \to \infty}$: The inner and outer exponents both approach one, and the inputs approach perfect substitutes. This gives a linear technology: $Y = z\,(\alpha A_K K + (1-\alpha)A_L L)$.

(b) $\boldsymbol{\sigma = 1}$: The expression in (3.1) is read as a limit and becomes Cobb–Douglas: $Y = z(A_K K)^{\alpha}(A_L L)^{1-\alpha}$.

Figure 6. Isoquants of the three limiting cases of CES production



Notes: For illustration, $\alpha = 1/2$, $z = A_K = A_L = 1$, and output is fixed at $\overline{Y} = 1$. All three isoquants pass through $L = K = 1$: the straight line is tangent to the Cobb–Douglas curve at the kink of the Leontief isoquant.

In this case we cannot distinguish between factor-augmenting technologies and *neutral* technological term z because we can always write $\widetilde{z} = z(A_K)^{\alpha}(A_L)^{1-\alpha}$ and obtain $Y = \widetilde{z}K^{\alpha}L^{1-\alpha}$.

(c) $\boldsymbol{\sigma \rightarrow 0}$: The technology approaches a fixed-proportions (or Leontief) technology. Capital and labor become perfect complements: $Y = z\min\{A_K K, A_L L\}$.

Same as in our general analysis, the elasticity of substitution determines what happens if the relative price of capital and labor changes. At the cost-minimizing input mix, the MRT equals w/r and

$$d\log(K/L) = \sigma\, d\log(w/r). \tag{3.9}$$

So, if capital becomes cheaper, w/r rises and the firm chooses a higher K/L. The value of σ tells us by how much. When $\sigma > 1$, the response is stronger than under Cobb–Douglas. When $\sigma < 1$, it is weaker.

Example 3.1 (The same price change under two technologies). *Suppose r falls by 10 percent and w is fixed. If $\sigma = 2$, the firm raises K/L by approximately 20 percent. If $\sigma = 0.5$, it raises K/L*

by only 5 percent. In both cases the firm uses a more capital-intensive technique; the size of the movement differs.

Where does the weird power in the CES come from?. The CES can look daunting because of the many terms it has. But it does not have to be that complicated. We can see why we write the function the way we do by re-deriving our results starting from a simpler function. Consider the following production function:

$$Y = z\left(K^{\beta} + L^{\beta}\right)^{\frac{1}{\beta}}. \tag{3.10}$$

We have dropped the weights and assumed $A_K = A_L = 1$.

The first order conditions give us

$$r = pz\,(\cdot)^{\frac{1}{\beta}-1} K^{\beta-1} = pz\,(\cdot)^{\frac{1-\beta}{\beta}} K^{\beta-1} = pz\left((\cdot)^{\frac{1}{\beta}}\right)^{1-\beta} K^{\beta-1} = pz^{\beta}\left(\frac{K}{Y}\right)^{\beta-1} \tag{3.11}$$

$$w = pz\,(\cdot)^{\frac{1}{\beta}-1} L^{\beta-1} = pz^{\beta}\left(\frac{L}{Y}\right)^{\beta-1} \tag{3.12}$$

So, optimality requires (by taking the ratio of the two equations)

$$\frac{w}{r} = \mathrm{MRT}_{L,K} = \left(\frac{L}{K}\right)^{\beta-1}. \tag{3.13}$$

From this we also get the ratio of income shares:

$$\frac{\alpha_L}{\alpha_K} = \frac{wL}{rK} = \left(\frac{L}{K}\right)^{\beta}. \tag{3.14}$$

Why do we write the power of the CES as $(\sigma-1)/\sigma$ and not just as β?

The reason is that we do not actually care about β directly. Just like in Section 1, we care about the elasticity of substitution:

$$\sigma = \frac{\partial \log \frac{K}{L}}{\partial \log \frac{w}{r}} . \tag{3.15}$$

We can figure out the elasticity from the optimality condition:

$$\log \frac{w}{r} = (\beta - 1) \log \frac{L}{K} \quad \longrightarrow \quad \log \frac{K}{L} = \frac{1}{1-\beta} \log \frac{w}{r} . \tag{3.16}$$

From this we see that the elasticity of substitution is

$$\sigma = \frac{1}{1-\beta} . \tag{3.17}$$

So, we can replace β in terms of σ:

$$\sigma = \frac{1}{1-\beta} \longrightarrow 1-\beta = \frac{1}{\sigma} \longrightarrow \beta = 1 - \frac{1}{\sigma} \longrightarrow \beta = \frac{\sigma - 1}{\sigma} . \tag{3.18}$$

This gives us the power we use in the CES. Now we can obtain our results directly in terms of the elasticity of substitution without having to translate the level of β into a level of σ.

3.1. Factor-augmenting technological change

As mentioned above, the CES allows us to study what happens when technology changes in a way that favours a particular factor. An increase in A_K makes every physical unit of capital more effective, while an increase in A_L makes every unit of labor more effective. This is different from increasing K or L: it changes what each physical unit can do. It is actually more similar to a change in the price. Being more productive means producing more output with the same physical input, which is (almost) equivalent to having a lower price for that input.

Equation (3.8) lets us separate the price response from the technology response:

$$d\log(K/L) = \sigma \; d\log(w/r) + (\sigma - 1)\, d\log(A_K/A_L). \tag{3.19}$$

The way in which technology affects the relative demand for inputs depends on the elasticity of substitution. Consider an increase in A_K (or a decrease in A_L) with w/r fixed. What happens to the number of physical units of capital relative to labor?

(a) $\boldsymbol{\sigma > 1}$: An increase in A_K raises physical K/L. The firm finds capital sufficiently easy to substitute toward that the productivity gain increases its relative use.

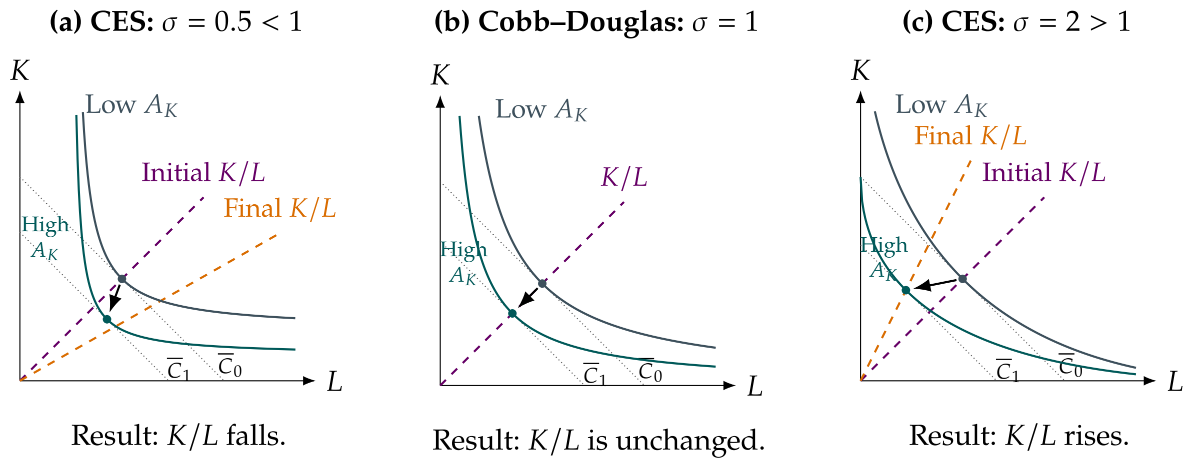
(b) $\boldsymbol{\sigma = 1}$: An increase in A_K does not change physical K/L at fixed prices. This is another special Cobb–Douglas result.

(c) $\boldsymbol{0 < \sigma < 1}$: An increase in A_K lowers physical K/L. Because capital and labor are hard to substitute for each other, each more-effective unit of capital can support the existing labor input with fewer physical units of capital.

The last case is important to understand. When $\sigma < 1$, the firm does not value capital less. Effective capital ($A_K K$) and effective labor ($A_L L$) are hard enough to substitute that each unit of capital now provides more services, allowing the firm to produce with fewer physical units.

Seeing the technology response in the cost-minimization diagram. Figure 7 compares the same increase in capital productivity across the three cases. We hold output, factor prices, and labor productivity fixed, and double A_K. For any given amount of labor, doubling A_K halves the capital needed to produce the same output, so the isoquant shifts inward in all cases. The firm can reach the new isoquant with a lower isocost, but the isocost slope $-w/r$ is unchanged because factor prices have not changed. The dots mark the initial and new cost-minimizing bundles (where MRT $= w/r$).

Figure 7. Capital-augmenting technological change: $A_K \uparrow$



Notes: Each panel holds $\overline{Y} = 1$, $w = r = 1$, $z = A_L = 1$, and $\alpha = 1/2$ fixed while A_K rises from one to two. Blue isoquants show low A_K and teal isoquants high A_K. Dotted lines are isocosts: $\overline{C}_0$ is initial cost and $\overline{C}_1$ is the lower cost after the productivity increase. Arrows connect the initial and new optima. Purple dashed rays mark initial K/L and orange dashed rays final K/L; Cobb–Douglas has a single unchanged ray. In the left panel capital falls proportionally more than labor, in the middle both fall in the same proportion, and in the right labor falls proportionally more than capital.

To see the change in the input mix, compare the rays from the origin through the two tangency points: the slope of each ray is K/L.

(a) With $\sigma < 1$, the new ray is flatter: capital falls proportionally more than labor.

(b) With Cobb–Douglas ($\sigma = 1$), both points lie on the same ray: the two inputs fall in the same proportion.

(c) With $\sigma > 1$, the new ray is steeper: labor falls proportionally more than capital.

The inward shift shows the production-cost saving in all three cases; the change in the ray shows whether the firm also changes its input mix. If we hold production constant this lets the firm produce with fewer inputs (less capital and less labor). If we allow the level of output to adjust we would see that we can now produce more because the lower costs mean lower prices and higher demand. Nevertheless, this higher output will be produced using the new capital to labor ratio K/L.

3.2. Input shares and the elasticity of substitution

As we hinted at in the discussion of the Cobb–Douglas technology, the elasticity of substitution also determines how the division of income between capital and labor responds to changes in relative prices. When $\sigma = 1$, the shares are constant. When $\sigma \neq 1$, they change.

We have already established that optimality involves equating the relative price w/r to the marginal rate of transformation, this gave us from (3.7)

$$\frac{w}{r} = \frac{1-\alpha}{\alpha}\left(\frac{A_L}{A_K}\right)^{\frac{\sigma-1}{\sigma}}\left(\frac{L}{K}\right)^{-\frac{1}{\sigma}} \quad \longleftrightarrow \quad \frac{K}{L} = \left(\frac{\alpha}{1-\alpha}\right)^{\sigma}\left(\frac{A_L}{A_K}\right)^{1-\sigma}\left(\frac{r}{w}\right)^{-\sigma} \tag{3.20}$$

We can use this to get the ratio of input shares. Recall that the input shares are $\alpha_K = \frac{rK}{rK+wL}$ and $\alpha_L = \frac{wL}{rK+wL}$. So their ratio is,

$$\frac{\alpha_K}{\alpha_L} = \frac{rK}{wL} = \left(\frac{r}{w}\right)\left(\frac{K}{L}\right) \tag{3.21}$$

Replacing K/L from the MRT condition gives us the ratio of input shares taking into account how the firm combines capital and labor given prices

$$\frac{\alpha_K}{\alpha_L} = \left(\frac{r}{w}\right)\left(\left(\frac{\alpha}{1-\alpha}\right)^{\sigma}\left(\frac{A_L}{A_K}\right)^{1-\sigma}\left(\frac{r}{w}\right)^{-\sigma}\right) = \left(\frac{\alpha}{1-\alpha}\right)^{\sigma}\left(\frac{A_L}{A_K}\frac{r}{w}\right)^{1-\sigma} \tag{3.22}$$

At $\sigma = 1$, we recover $\alpha_K = \alpha$ and $\alpha_L = 1-\alpha$.

These expressions are useful because they tell us about how inequality in income between input factors changes with relative prices and with technology:

$$\log\left(\frac{\alpha_K}{\alpha_L}\right) = \sigma \log\left(\frac{\alpha}{1-\alpha}\right) + (1-\sigma)\left[\log\left(\frac{r}{w}\right) - \log\left(\frac{A_K}{A_L}\right)\right]. \tag{3.23}$$

Holding technology fixed, the relative-share response is

$$d\log\frac{\alpha_K}{\alpha_L} \;=\; (1-\sigma)\,d\log\frac{r}{w}. \tag{3.24}$$

A fall in r/w raises capital's share when $\sigma > 1$ and lowers it when $\sigma < 1$. A change in technology has the opposite effect as being more productive is equivalent to being cheaper.

$$d\log\frac{\alpha_K}{\alpha_L} \;=\; (\sigma-1)\,d\log\frac{A_K}{A_L}. \tag{3.25}$$

3.3. Technology and factor rewards

We now turn to one of the key questions brought up by Hicks (1932):

How does technological change affect the distribution of income between labor and capital?

To tackle this we need to see how prices respond to technology that favors one input factor over another. Because our question is about technology we are going to hold the physical supplies K and L fixed and let competitive factor prices adjust to equal marginal products. The cost minimization problem implies

$$d\log(w/r) \;=\; d\log \mathrm{MRT}_{L,K} \;=\; -\frac{\sigma-1}{\sigma}\,d\log A_K \;+\; \frac{\sigma-1}{\sigma}\,d\log A_L. \tag{3.26}$$

This gives us our answer: If $\sigma > 1$, capital-augmenting technical change raises the rental price of capital relative to the wage. If $\sigma < 1$, the result reverses. Making capital more productive raises the relative reward to labor because labor and effective capital are hard to substitute for each other. When $\sigma = 1$, all technological change is neutral and does not affect relative factor prices.

Substitutability plays a central role (again). Only when factors are easy to substitute does making one factor more productive raise its relative reward. When factors are hard to substitute, making one factor more productive ends up raising the relative reward to the other factor. In this second case the rewards goes to the factor that is now relatively scarce in effective terms, not to the factor that is now more productive.

What about the level of the wage?. Does a decrease in the relative factor price (w/r) mean the wage itself falls when the rental rate rises by more? The answer under the CES technology is unambiguous: No. In fact, any increase in technology (either capital- or labor-augmenting) raises the level of the wage.

This is consistent with the result we just derived because equation (3.26) concerns the *relative* factor price w/r, not the level of the wage. To get at the effect of technological change in the level of wages we can look at the optimality condition for labor demand

$$\log \frac{w}{p} = \frac{\sigma - 1}{\sigma} \log z + \frac{\sigma - 1}{\sigma} \log A_L - \frac{1}{\sigma} \log L + \frac{1}{\sigma} \log Y. \tag{3.27}$$

From here we can get the effect of a change in A_K on w.

$$\begin{aligned}
\frac{\partial \log \frac{w}{p}}{\partial \log A_K} &= \frac{1}{\sigma} \frac{\frac{\partial Y}{\partial \log A_K}}{Y} = \frac{1}{\sigma} \frac{\frac{\partial Y}{\partial A_K}}{Y} A_K \qquad (3.28)\\
&= \frac{1}{\sigma} \frac{z\left(\alpha \left(A_K K\right)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\left(A_L L\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} \alpha A_K^{-\frac{1}{\sigma}} K^{\frac{\sigma-1}{\sigma}}}{Y} A_K \\
&= \frac{1}{\sigma} \frac{Y^{\frac{1}{\sigma}} \alpha A_K^{\frac{\sigma-1}{\sigma}} K^{-\frac{1}{\sigma}}}{Y} K \\
&= \frac{1}{\sigma} \frac{rK}{pY} = \frac{\alpha_k}{\sigma} > 0
\end{aligned}$$

This applies also to an increase in the level of capital:

$$\frac{\partial \log w}{\partial \log A_K} = \frac{\partial \log w}{\partial \log K} = \frac{\alpha_K}{\sigma} > 0. \tag{3.29}$$

Thus, in this aggregate CES model, both capital deepening (a term used for capital accumulation) and capital-augmenting technical change raise the wage for every $\sigma > 0$. The rental rate may rise even more, causing w/r to fall, but that does not undo the positive effect on the level of w.

Here we reach the limit of the aggregate production function. The results we have do not prove that innovation can never harm workers. Instead, they show that at this level of aggregation the model does not have a margin through which innovation can harm workers. That is why we introduce tasks later on.

3.4. CES Cheatsheet

The formulas below collect the capital–labor CES results. Assume $0 < \alpha < 1$, $\sigma > 0$, positive inputs and productivities, and competitive factor rewards. The output price is p; α is a technology parameter, whereas α_K and α_L are income shares. Read the expressions at $\sigma = 1$ by taking their Cobb–Douglas limits.

Object	**Mathematical expression**
CES production	$Y = F(K, L) = z\left[\alpha(A_K K)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)(A_L L)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$
Cobb–Douglas limit	$Y = z(A_K K)^{\alpha}(A_L L)^{1-\alpha}, \qquad \sigma = 1$
Marginal products	$F_K = z^{\frac{\sigma-1}{\sigma}} \alpha A_K^{\frac{\sigma-1}{\sigma}} \left(\frac{Y}{K}\right)^{\frac{1}{\sigma}}$
	$F_L = z^{\frac{\sigma-1}{\sigma}} (1-\alpha) A_L^{\frac{\sigma-1}{\sigma}} \left(\frac{Y}{L}\right)^{\frac{1}{\sigma}}$
MRT	$\mathrm{MRT}_{L,K} = \frac{F_L}{F_K} = \frac{1-\alpha}{\alpha}\left(\frac{A_L}{A_K}\right)^{\frac{\sigma-1}{\sigma}}\left(\frac{K}{L}\right)^{\frac{1}{\sigma}} = \frac{w}{r}$
Cost-minimizing input ratio	$\frac{K}{L} = \left(\frac{\alpha}{1-\alpha}\right)^{\sigma}\left(\frac{A_K}{A_L}\right)^{\sigma-1}\left(\frac{w}{r}\right)^{\sigma}$
Relative shares	$\frac{\alpha_K}{\alpha_L} = \frac{\alpha}{1-\alpha}\left(\frac{A_K K}{A_L L}\right)^{\frac{\sigma-1}{\sigma}} = \left(\frac{\alpha}{1-\alpha}\right)^{\sigma}\left(\frac{r/A_K}{w/A_L}\right)^{1-\sigma}$
Factor rewards	$w = pF_L = p\alpha_L \frac{Y}{L}, \qquad r = pF_K = p\alpha_K \frac{Y}{K}$ Real rewards are $w/p = F_L$ and $r/p = F_K$.

Main results: input choices, shares, and factor rewards. Each derivative is a local elasticity evaluated at the initial allocation. Technology includes z, A_K, A_L and the weight α; variables not named in an experiment are held fixed. For a decrease in the variable in the denominator, reverse the sign of the response.

Question and what is fixed	**Relevant derivative**	**Answer and role of σ**
How does the input mix respond to relative prices? Technology fixed.	$\dfrac{\partial \log(K/L)}{\partial \log(w/r)} = \sigma$	A higher wage relative to the rental price raises K/L. The response is stronger when σ is larger.
Does better capital raise physical K/L? Factor prices fixed.	$\dfrac{\partial \log(K/L)}{\partial \log A_K} = \sigma - 1$	Yes if $\sigma > 1$; no change if $\sigma = 1$; it lowers K/L if $\sigma < 1$.
Does cheaper capital raise its relative income share? Technology fixed.	$\dfrac{\partial \log(\alpha_K/\alpha_L)}{\partial \log(r/w)} = 1 - \sigma$	A fall in r/w raises capital's share if $\sigma > 1$ and lowers it if $\sigma < 1$. Shares are constant at $\sigma = 1$.
Does better capital raise its relative income share? Factor prices fixed.	$\dfrac{\partial \log(\alpha_K/\alpha_L)}{\partial \log A_K} = \sigma - 1$	Capital's share rises if $\sigma > 1$, falls if $\sigma < 1$, and is unchanged at $\sigma = 1$.
Does more effective capital raise its relative income share? Physical L and labor productivity fixed; rewards adjust.	$\dfrac{\partial \log(\alpha_K/\alpha_L)}{\partial \log(A_K K)} = \dfrac{\sigma - 1}{\sigma}$	Same sign threshold as above, but a different derivative because factor prices now adjust.
Does better capital raise the wage relative to the rental rate? K, L fixed.	$\dfrac{\partial \log(w/r)}{\partial \log A_K} = \dfrac{1 - \sigma}{\sigma}$	w/r falls if $\sigma > 1$, rises if $\sigma < 1$, and is unchanged at $\sigma = 1$.
Does better capital raise the real wage itself? K, L fixed.	$\dfrac{\partial \log(w/p)}{\partial \log A_K} = \dfrac{\alpha_K}{\sigma}$	Yes for every $\sigma > 0$, even when w/r falls. The rental rate can rise faster than the wage.
Does capital accumulation raise the real wage? L and technology fixed.	$\dfrac{\partial \log(w/p)}{\partial \log K} = \dfrac{\alpha_K}{\sigma}$	Yes for every $\sigma > 0$. Capital deepening and better capital have the same local wage elasticity.

3.5. Historical background: why economists introduced CES

Solow: Employment and growth. The story of the CES is rooted in the interest of economist to understand how growth can happen when technological change is biased toward one factor. In particular, Solow's 1956 paper (Solow 1956) on growth theory was motivated by the question of whether saving and investment can keep a growing labor force fully employed. As more workers enter the economy, they need capital to work with. Saving finances investment, but will the resulting capital stock provide enough jobs without leaving machines idle?

The issue economists had is that the growth rate consistent with saving and investment need not equal the growth rate consistent with full employment. What if technology is augmenting capital faster than labor? If capital and labor are hard to substitute for each other, the economy may not be able to use all of the capital and labor that is available. We can see this clearly in the extreme case in which capital and labor are perfect complements and the production function is Leontief (as in the pionering work of Harrod (1939) and Domar (1946)):

$$Y = \min\left\{\frac{K}{A_K}, \frac{L}{A_L}\right\}. \tag{3.30}$$

Production requires fixed amounts $A_K > 0$ of capital and $A_L > 0$ of labor per unit of output. This means that to use all capital and employ all labor, we need $K/A_K = L/A_L = Y$. In particular, the capital–output ratio must remain $K/Y = A_K$.

If investment is such that capital accumulation equals a fraction s of output (the savings rate), then we have that capital grows at rate

$$g_K \equiv \frac{K_{t+1} - K_t}{K_t} = \frac{sY}{K} = \frac{s}{A_K} = g_Y. \tag{3.31}$$

That growth rate must be the same as the growth rate of output g_Y to keep the capital–output ratio constant.

However, full employment requires $Y = L/A_L$, which implies that output and labor must grow at the same rate, equal to the population growth rate n:

$$g_Y = g_L = n. \tag{3.32}$$

Sustained growth with both factors fully employed consequently requires the exact equality

$$\frac{s}{v} = n. \tag{3.33}$$

Why should household saving behavior, the technological capital requirement, and population growth happen to satisfy this equality? There is no reason in this model for them to do so. This result was concerning for Solow because balanced growth rests on a *knife edge*: a small change in s, A_K, or n breaks the compatibility condition.

The answer Solow gave to this problem involved abandoning the benchmark of perfect complements. Under the Leontief technology, a lower wage cannot persuade firms to use more labor per machine. The required ratio $K/L = A_K/A_L$ is technological, regardless of relative factor prices. Solow challenges this restriction:

> "If this assumption is abandoned, the knife-edge notion of unstable balance seems to go with it."

Source and context: Solow (1956, p. 65), *A Contribution to the Theory of Economic Growth*, Introduction, printed pp. 65–66 (PDF pp. 2–3).

Solow replaces the Leontief production function with a constant-returns-to-scale production function $Y = F(K, L) = Lf(k)$, where $k = K/L$ and $f(k) = F(k, 1)$. The same saving and population growth assumptions now give

$$K_{t+1} = sY_t \quad \longrightarrow \quad \frac{K_{t+1}}{L_{t+1}}\frac{L_{t+1}}{L_t} = s\frac{Y_t}{L_t} \quad \longrightarrow \quad k_{t+1}(1+n) = sf(k_t) \tag{3.34}$$

In steady state, a constant capital–labor ratio k^{ss} satisfies $sf(k^{ss}) = (1+n)k^{ss}$. The capital–output ratio can now adjust with k, rather than being fixed at A_K. For example, with the Cobb-Douglas production function we obtain $f(k) = k^{\alpha}$, with $0 < \alpha < 1$. Then

$$k^{ss} = \left(\frac{s}{1+n}\right)^{\frac{1}{1-\alpha}}. \tag{3.35}$$

If $k < k^{ss}$, capital grows faster than labor and k rises. If $k > k^{ss}$, capital grows more slowly than labor and k falls. A change in s or n changes the long-run ratio; it no longer requires that the parameters happen to coincide for there to be long-run growth. Relative wages and rental prices adjust so that firms choose the available input ratio.

So, what do we take from this? We need inputs to be substitutable to obtain *balanced growth*. The Cobb-Douglas with its elasticity of substitution of one is one case. However, and as we saw above, the Cobb-Douglas does not cover every relevant case.

Robert M. Solow (1924–2023)

Solow was an American economist born in Brooklyn. He entered Harvard in 1940, served in the Second World War, and returned to study economics. He joined MIT in 1949 and received the 1987 economics Nobel Prize for his contributions to the theory of economic growth.

His work asked why economies grow and how much of that growth comes from capital accumulation and technological progress. In his 1956 growth model, substitution between capital and labor allowed production to adjust as their relative availability changed. His later collaboration with Arrow, Chenery, and Minhas introduced the CES function. These contributions connect our questions about input substitution to the larger question of how technology changes living standards over time.

Sources: Nobel Prize profile; Solow (1956); Arrow et al. (1961).

Making substitutability an empirical parameter. The CES production function was introduced in 1961 by Kenneth Arrow, Hollis Chenery, Bagicha Minhas, and Robert Solow in Arrow et al. (1961). At the time, applied work was dominated by Leontief technologies with fixed coefficients and Cobb–Douglas production functions. These were not entirely arbitrary benchmarks, but they did constraint results because the elasticity of substitution is fixed at zero for Leontief functions and one for Cobb-Douglas. Arrow, Chenery, Minhas, and Solow wrote:

> "From a mathematical point of view, zero and one are perhaps the most convenient alternatives for this elasticity."

They further argued that it would be better to estimate this elasticity rather than impose its value. Their main argument is that there are many economic questions whose answers depend on the value of the elasticity of substitution:

(a) **Can a growing economy keep both factors fully employed?** This is the problem highlighted by Solow in his previous work, showing that zero substitution between labor and capital is incompatible with long run employment.

(b) **How do a country's factor supplies affect what it trades?** The authors show that the effects of trade depend on this elasticity too, for example implying "*reversals of factor intensities at different factor prices.*" A sector that uses relatively more capital at one wage–rental ratio need not do so at another. Differences in substitution possibilities across sectors can change the ranking and hence conclusions about trade and factor rewards.

(c) **Does labor necessarily receive a constant share of income?** They question the "*supposed constancy of the labor share*." Choosing Cobb–Douglas imposes constant competitive shares. Whether shares are approximately stable, and what explains that stability, are empirical questions rather than reasons to impose the answer in advance.

Source and context: Arrow et al. (1961), printed p. 225 (PDF p. 2), examples (i)–(iii).

At the core of the questions posed by these economists is that we want to attribute differences in output to capital, labor, and technological knowledge. But what if the production function used for that decomposition was wrong? Arrow put it this way:

> "I had speculated that the decomposition might be faulty if the wrong production function were used, but I had done little about it."

Source and context: Arrow (1979), one-page *Citation Classic* retrospective, first two paragraphs of Arrow's account. The statement is dated December 27, 1977 and was published on April 23, 1979.

Kenneth J. Arrow (1921–2017)

Arrow was an American economist born in New York. He studied mathematics at City College and Columbia, moving into economics under Harold Hotelling's influence. Research at the Cowles Commission and RAND helped shape his work on social choice and economic efficiency. He taught at Stanford and Harvard and shared the 1972 economics Nobel Prize with Hicks for contributions to general equilibrium and welfare theory.

His research asked how individual choices fit together, when markets allocate resources efficiently, and how information changes economic decisions. His impossibility theorem identified limits to combining individual preferences into a collective ranking. For this lecture, his collaboration with Chenery, Minhas, and Solow matters because it made the elasticity of substitution a parameter to estimate. His later work on innovation and learning by doing also connects to our study of technology.

Sources: Arrow's Nobel autobiography and 2005 addendum (*Nobel Lectures*, Economics 1969–1980); Nobel Prize profile; Arrow et al. (1961).

4. Examples and Applications of the CES Production Function

4.1. Weak links and capital-augmenting productivity growth

The same CES structure gives us one more result that will matter when we study the introduction of artificial intelligence. We can think of AI as a technology that makes capital more productive at the tasks it already performs. The better the AI the more productive a given level of capital becomes. This is like an increase in A_K in the CES production function. What if AI makes capital *infinitely productive*?

Does infinite productivity growth for machines mean infinite output?

To answer this, we follow Steinsson (2026, Section 7.4) and Jones and Tonetti (2026).

Start from the CES production function,

$$Y = z\left(\alpha \widetilde{K}^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\widetilde{L}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \qquad \widetilde{K} \equiv A_K K, \quad \widetilde{L} \equiv A_L L. \tag{4.1}$$

Think of $\widetilde{K}$ as the output from the collection of tasks currently performed by capital and $\widetilde{L}$ as the output from the collection currently performed by labor. Technology like AI improves the performance of capital at its tasks, increasing $\widetilde{K}$.

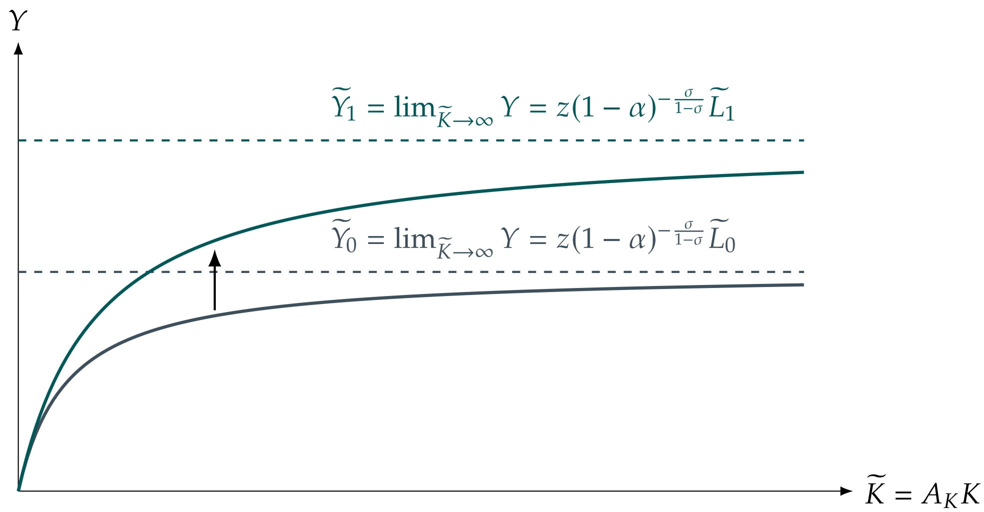
Now consider an extreme experiment. Capital becomes arbitrarily productive at the tasks it already performs, so $\widetilde{K} \to \infty$, while the assignment of tasks stays fixed. However, capital and labor are not very substitutable, so that $0 < \sigma < 1$. So, even though $\widetilde{K}$ grows without bound, aggregate output remains bounded

$$\widetilde{Y} = \lim_{\widetilde{K}\to\infty} Y = z(1-\alpha)^{-\frac{\sigma}{1-\sigma}}\widetilde{L}. \tag{4.2}$$

This is because with $0 < \sigma < 1$ we have $\frac{\sigma-1}{\sigma} < 0$ and $\widetilde{K}^{\frac{\sigma-1}{\sigma}} \to 0$.

Figure 8 shows why better machines alone cannot remove the weak link. At a fixed level of effective labor, more effective capital moves production along a curve toward its horizontal ceiling. Raising effective labor lifts that ceiling: improving the remaining labor tasks, or supplying more labor, relaxes the bottleneck itself. The dashed lines are the weak-link limits in (4.2), not additional production possibilities at finite capital.

Figure 8. Effective labor increase and the weak-link output ceiling



Notes: Both curves use $\sigma = 1/2$, $\alpha = 1/2$, and $z = 1$; effective labor rises from $\widetilde{L}_0 = 1$ to $\widetilde{L}_1 = 1.6$. The blue curve shows initial effective labor $\widetilde{L}_0$ and the teal curve higher effective labor $\widetilde{L}_1$. Each dashed ceiling gives the output limit in (4.2), holding the corresponding effective labor supply fixed. Task assignment is held fixed.

What is this capturing?. The low substitutability between capital and labor means that labor tasks are *weak links* in production (as in a chain that is only as strong as its weakest link). Producing the capital tasks at negligible cost (because of infinte productivity) does not eliminate the need to perform the tasks that remain with labor and are hard to substitute for. In the special case $\sigma = 1/2$, $\widetilde{Y} = \frac{z\widetilde{L}}{1-\alpha}$.

How costly are weak links?. We can measure how costly the weak links are for growth by measuring the gap between current output and the limit output under infinite capital productivity. It turns out that the answer depends only on the capital share (α_K) and the elasticity of substitution (σ), not on the level of technology or the physical supplies of capital and labor. To see this, recall that we can express the capital share as $\alpha_K = \alpha\left(\frac{z\widetilde{K}}{Y}\right)^{\frac{\sigma-1}{\sigma}}$.

Then, from equation (4.2) we obtain

$$\frac{\widetilde{Y}}{Y} = \left(\frac{1}{1-\alpha_K}\right)^{\frac{\sigma}{1-\sigma}}, \qquad 0 < \sigma < 1. \tag{4.3}$$

If the capital tasks that are being improved command a small share of expenditure (low α_K) and σ is low, even an enormous productivity improvement in those tasks can generate only a modest increase in aggregate output.

What happens to input shares?. As capital becomes infinitely productive at its tasks the share of expenditure going to capital and labor can change. When capital and labor are hard to substitute for each other, the share of income going to capital end up falling as its productivity rises, with ever more expenditure going to labor. Holding the task partition fixed,

$$\alpha_K = \alpha \left(\frac{z\widetilde{K}}{Y}\right)^{\frac{\sigma-1}{\sigma}} \longrightarrow 0, \qquad 0 < \sigma < 1. \tag{4.4}$$

The highly productive capital tasks become so cheap that their income share vanishes. This is the logic of Baumol's cost disease (Baumol 1967). When activities are hard to substitute for each other, the activities with slow productivity growth become relatively expensive and account for a growing share of expenditure.

The fixed task partition is an essential qualification in all of these results. Equations (4.2)–(4.4) describe capital becoming better at tasks it already performs. Automation can also expand the set of capital tasks. That extensive-margin change displaces labor from particular tasks and is the subject of lectures to follow.

Weak links: why automating some tasks is not enough

A *weak link* is an essential activity that does not benefit directly from the productivity improvement brought up by AI and other technologies. These activities limit the gains from technological change elsewhere because other activities are poor substitutes for it. For example, faster record keeping does not replace the time needed to help a confused patient through the night; producing more software does not rewire an old building or run a kindergarten classroom. These are among the human activities highlighted by Jones and Tonetti (2026, Section 5.2). The issue is not that they can never be automated, but that they may remain difficult to improve while other tasks become much cheaper.

Weak links limit the growth that can come from productivity and automation. Consider software development, a sector that accounts for 2% of GDP and where AI may eventually automate a large share of the work. Now imagine that AI makes software development infinitely productive ($A_S \longrightarrow \infty$). With an across-task elasticity of $\eta = 1/2$, their fixed-task calculation gives an output multiplier of $1/(1-0.02) \approx 1.0204$: a gain of only about 2%, despite the unlimited improvement in software. The other 98% of sectors not directly affected by AI still limit production.

History illustrates why progress must spread across tasks. This is what it takes for growth to take over. Jones and Tonetti describe combine harvesters replacing manual threshing, automatic elevators replacing elevator operators, software taking over typing and travel-booking tasks, and robots welding and painting cars. These changes extend the range of activities benefiting from improving machines. In the authors' historical counterfactual, freezing the set of automated tasks at its 1950 level while allowing machine productivity to keep improving eliminates 87% of private-business TFP growth over 1950–2023. In their model, automation strengthens the remaining weak links by allowing more tasks to benefit from rapidly improving machines. Their historical counterfactual therefore suggests that improving machines on already-automated tasks would have generated much less growth than combining those improvements with the automation of additional tasks.

Stable factor shares can conceal extensive automation. We can illustrate the two opposing effects with a simple static task example, following Jones and Tonetti (2026, Appendix B.3). Suppose equally weighted tasks combine with elasticity $0 < \sigma < 1$. Capital performs a fraction $\beta \in (0, 1)$ of tasks and labor performs the rest. Within each group, productivity is the same across tasks: ψ_K for capital and ψ_L for labor. Capital is cheaper on the tasks it can perform; the remaining tasks require labor. Their unit costs are r/ψ_K and w/ψ_L, respectively. CES task demand makes expenditure on each task proportional to its unit cost raised to $1-\sigma$. Summing expenditures within each group therefore gives

$$\frac{\alpha_K}{\alpha_L} = \frac{\beta}{1-\beta}\left(\frac{r/\psi_K}{w/\psi_L}\right)^{1-\sigma}. \tag{4.5}$$

Here β measures the fraction of automated tasks, whereas α_K measures their share of expenditure. Holding r, w, and ψ_L fixed, log differentiation separates a change in task coverage from a change in machine productivity:

$$d\log\frac{\alpha_K}{\alpha_L} = \underbrace{\frac{d\beta}{\beta(1-\beta)}}_{\text{more tasks automated}} - \underbrace{(1-\sigma)\,d\log\psi_K}_{\text{better machines}}. \tag{4.6}$$

Both factor shares remain unchanged when

$$\frac{d\beta}{\beta(1-\beta)} = (1-\sigma)\,d\log\psi_K, \qquad d\alpha_K = d\alpha_L = 0. \tag{4.7}$$

Automating more tasks shifts expenditure toward capital, while better machines make the capital tasks cheaper and reduce their expenditure share when tasks are hard to substitute for each other. These effects can exactly offset: stable factor shares can coexist with a larger fraction of tasks performed by machines. This compares static allocations at given factor prices; unchanged shares alone do not imply unchanged task assignments.

William J. Baumol (1922–2017): Why personal services become expensive

Baumol was an American economist who taught at Princeton and New York University, with major contributions to competition, entrepreneurship, and public policy. He was widely regarded by fellow economists as deserving of a Nobel Prize, although he never received one.[a]

[a] See the Princeton obituary and Litan and Hathaway's 2017 assessment.

The cost disease. The issue Baumol wanted to understand is what happens when productivity growth is uneven across sectors of the economy. It was clear that the sector experiencing productivity growth could expand and pay more to its workers. But, what about the sectors that become (relatively) less productive?

The key to Baumol's answer is that the labor market connects the different sectors. So, productivity improvements elsewhere in the economy raise workers' outside earning opportunities. Because of this wages must rise to retain staff even in sectors that are not experiencing productivity growth. This makes the output of these sectors become more expensive relative to goods produced with rapidly improving technology. This is the *cost disease*: Rising productivity can make some goods and services more costly.

Baumol developed the argument with William Bowen in their study of the performing arts and extended it to other services (Baumol and Bowen 1966; Baumol 1967). William Nordaus (Nobel prize 2018) found that industries with slower productivity gains experienced rising relative prices and declining relative real output (Nordhaus 2008). Outside manufacturing, faster productivity gains were also associated with slower increases in employment and hours. This supports the distinction between becoming more expensive and producing relatively more.

Examples.

- *Live music:* a string quartet still requires four musicians for the duration of the performance, as it did in Mozart's time. Recordings reach larger audiences, but provide a different product.
- *Teaching:* individual tutoring, feedback, and classroom interaction require teachers' attention. Increasing class size can raise students per teacher while changing the service each student receives.
- *Personal care:* bedside nursing, helping an older person wash and dress, and a physician's examination require time with a person. Technology can improve particular tasks without eliminating that time.

Connection to weak links. When buyers cannot easily substitute away from these services, their rising relative prices can increase their share of expenditure, just as labor tasks gain expenditure share in our example. If buyers can substitute away, the service may instead contract. These patterns do not imply that all services have stagnant productivity, or that all increases in healthcare and education spending reflect cost disease: quality, demand, and institutions also matter.

4.2. Substitution and the scale of production

If capital becomes cheaper, the firm chooses a higher K/L. Does this mean that it employs fewer workers? A higher ratio tells us about the input mix, but labor demand also depends on how much the firm produces as in our analysis under the Cobb-Douglas production function.

Consider a competitive industry with the technology in (3.1) and market demand for its final good

$$Y = Bp^{-\varepsilon}, \qquad B > 0, \quad \varepsilon > 0. \tag{4.8}$$

Here p is the final-good price and ε is the price elasticity of market demand in absolute value. We hold w, technology, and demand parameters fixed and lower r. Inputs are available to this industry at the given factor prices, so employment can adjust.

Constant returns and perfect competition imply $p = c(r, w)$, the unit cost of production. So, the effect on demand depends on how much costs decrease when r goes down. The cost-minimization conditions are $r = cF_K$ and $w = cF_L$. Substituting the CES marginal products and solving for inputs per unit of output gives

$$\frac{K}{Y} = z^{\sigma-1}\alpha^{\sigma}A_K^{\sigma-1}\left(\frac{c}{r}\right)^{\sigma}, \qquad\qquad \frac{L}{Y} = z^{\sigma-1}(1-\alpha)^{\sigma}A_L^{\sigma-1}\left(\frac{c}{w}\right)^{\sigma}. \tag{4.9}$$

Constant-return to scale imply $c = rK/Y + wL/Y$. So, we can substitute the input-to-output ratios from the cost minimization conditions to get

$$c \;=\; z^{\sigma-1}c^{\sigma}\left[\alpha^{\sigma}\left(\frac{r}{A_K}\right)^{1-\sigma} + (1-\alpha)^{\sigma}\left(\frac{w}{A_L}\right)^{1-\sigma}\right]. \tag{4.10}$$

Dividing by c^{σ} and raising both sides to the power $\frac{1}{1-\sigma}$ gives

$$c(r, w) = \frac{1}{z}\left[\alpha^{\sigma}\left(\frac{r}{A_K}\right)^{1-\sigma} + (1-\alpha)^{\sigma}\left(\frac{w}{A_L}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}. \tag{4.11}$$

Taking logs and differentiating with respect to $\log r$, we obtain

$$\frac{\partial \log c}{\partial \log r} = \frac{\alpha^{\sigma}(r/A_K)^{1-\sigma}}{\alpha^{\sigma}(r/A_K)^{1-\sigma} + (1-\alpha)^{\sigma}(w/A_L)^{1-\sigma}} = \alpha_K. \tag{4.12}$$

The capital cost share therefore measures the sensitivity of unit cost to the rental price. Unlike Cobb–Douglas, this share changes with factor prices; the expression gives the elasticity at the current input mix.

Since $p = c$, final-good demand implies

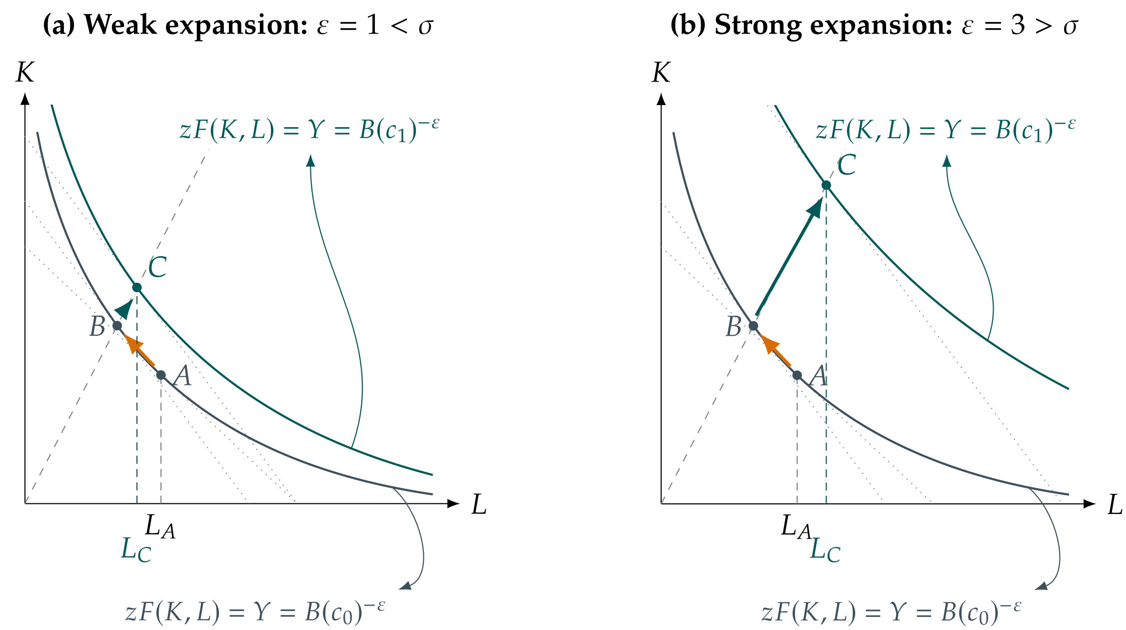
$$\log Y = \log B - \varepsilon \log c \quad \longrightarrow \quad \frac{\partial \log Y}{\partial \log r} = -\varepsilon \alpha_K. \tag{4.13}$$

A one-percent fall in r lowers the output price by approximately α_K percent and expands output by $\varepsilon \alpha_K$ percent. Demand elasticity determines how strongly buyers respond to the lower price.

Does the expansion outweigh substitution away from labor?. Figure 9 separates the two adjustments before we do the algebra. At A, the firm uses the original cost-minimizing bundle. When r falls, the isocost becomes steeper and the fixed-output optimum moves to B: the firm substitutes capital for labor along the original isoquant. Lower unit cost also reduces the output price and increases sales. Production then expands from B to C along the new constant-K/L ray, increasing both inputs. With weak product-demand response, this expansion does not restore the original employment level; with strong response, it more than restores it.

The connection to consumer theory. The movement from A to B is analogous to a compensated substitution effect: we change relative prices while holding the objective level fixed—output here, utility in consumer theory. The movement from B to C resembles the income effect geometrically because it moves to a higher isoquant along an expansion

Figure 9. Substitution and scale effects of cheaper capital



Notes: Both panels use $\sigma = 2$, $\alpha = 1/2$, $z = A_K = A_L = w = 1$, and a rental-price decline from $r = 1$ to $r = 0.7$. Product demand is normalized to give initial output $Y = 1$ in both panels. A and B lie on the original isoquant; C lies on the higher-output isoquant. Dotted isocosts at B and C are parallel. The blue isoquant has output $B(c_0)^{-\varepsilon}$ and the teal isoquant $B(c_1)^{-\varepsilon}$, where $c_1 < c_0$ are final and initial unit costs. Here B in the demand formula is the demand scale parameter, distinct from the point labelled B. The movement $A \longrightarrow B$ is the substitution effect and $B \longrightarrow C$ the scale effect. Final labor is below its initial level in the left panel and above it in the right panel.

path. But its cause is different: the firm has no fixed income budget whose purchasing power has risen. Instead, lower production costs reduce the competitive output price and consumers buy more. The size of this scale effect depends on the elasticity of product demand ε. Under constant returns, expanding output at the new factor prices scales both inputs proportionally, leaving K/L unchanged between B and C. The comparison of ε with σ tells us which movement dominates employment.

Taking logs of (4.9) separates the amount produced from the labor required per unit:

$$\log L = \log Y + (\sigma - 1)\log z + \sigma \log(1 - \alpha) + (\sigma - 1)\log A_L + \sigma \log c - \sigma \log w. \qquad (4.14)$$

The response of labor has two components:

$$\frac{\partial \log L}{\partial \log r} = \left.\frac{\partial \log L}{\partial \log r}\right|_Y + \frac{\partial \log L}{\partial \log Y}\frac{\partial \log Y}{\partial \log r} = \underbrace{\sigma\alpha_K}_{\text{substitution}} - \underbrace{\varepsilon\alpha_K}_{\text{scale}} = \alpha_K(\sigma - \varepsilon). \tag{4.15}$$

Remember that a fall in r is a negative change in $\log r$. Labor demand therefore rises if $\varepsilon > \sigma$, falls if $\varepsilon < \sigma$, and is unchanged if $\varepsilon = \sigma$. In the Cobb–Douglas limit, $\sigma = 1$ and $\alpha_K = \alpha$, recovering $\alpha(1 - \varepsilon)$ in (2.18).

Because w is fixed, $d\log(w/r) = -d\log r$. We can also write

$$\frac{\partial \log L}{\partial \log(w/r)} = \alpha_K(\varepsilon - \sigma), \qquad \frac{\partial \log(K/L)}{\partial \log(w/r)} = \sigma. \tag{4.16}$$

The capital–labor ratio rises as capital becomes cheaper even when total labor demand rises. When $\sigma > 1$, capital and labor are sufficiently substitutable that cheaper capital also lowers labor's cost share, as shown earlier.

A falling labor share and rising employment can therefore occur together: Output expands enough to require more workers despite using less labor per unit of output.

Which outcomes depend on the scale of production?. At given factor prices and technology, *neither factor shares nor prices depend on* ε. The input prices r and w are given, and the output price $p = c(r, w)$ is determined by technology and input prices. Expanding production scales both input demands proportionally, so it does not change their cost shares. Differentiating the share expressions above, with w and technology fixed, gives

$$d\log\alpha_L = (\sigma - 1)\alpha_K \, d\log r, \qquad d\log\alpha_K = -(\sigma - 1)\alpha_L \, d\log r. \tag{4.17}$$

Neither response contains ε. For $\sigma > 1$, cheaper capital lowers labor's share and raises capital's share, whatever the elasticity of final-good demand.

The real wage measured in units of this final good is w/p. Since the nominal wage is fixed and $p = c(r, w)$,

$$d \log(w/p) = -\, d \log c = -\alpha_K \; d \log r. \tag{4.18}$$

Cheaper capital therefore raises this real wage, also independently of ε. This measures purchasing power over the industry's good.

Taking Stock. Hold w and technology fixed, let r change, and use competitive pricing $p = c(r, w)$ with final-good demand $Y = Bp^{-\varepsilon}$, $\varepsilon > 0$. Input quantities and output can adjust. The table separates the change in labor per unit of output from the change in total employment and shows which outcomes depend on the demand elasticity.

Question and what is fixed	Relevant derivative	Answer and role of σ
Does cheaper capital lower costs and expand output?	$\dfrac{\partial \log c}{\partial \log r} = \alpha_K$ $\dfrac{\partial \log Y}{\partial \log r} = -\varepsilon \alpha_K$	Yes for every $\sigma > 0$. The share determines cost pass-through; ε determines output expansion.
Does cheaper capital reduce labor per unit of output?	$\dfrac{\partial \log(L/Y)}{\partial \log r} = \sigma \alpha_K$	Yes for every $\sigma > 0$. This is the substitution effect before allowing output to expand.
Does cheaper capital raise total labor demand?	$\dfrac{\partial \log L}{\partial \log r} = \alpha_K(\sigma - \varepsilon)$	Yes if $\varepsilon > \sigma$; no if $\varepsilon < \sigma$; unchanged if $\varepsilon = \sigma$. The cutoff is ε, not one.
Does cheaper capital lower labor's share?	$\dfrac{\partial \log \alpha_L}{\partial \log r} = (\sigma - 1)\alpha_K$	Yes if $\sigma > 1$; it raises the share if $\sigma < 1$. This result is independent of ε.
Does cheaper capital raise the purchasing power of the fixed wage?	$\dfrac{\partial \log(w/p)}{\partial \log r} = -\alpha_K$	Yes for every $\sigma > 0$, independently of ε. The final good becomes cheaper.

4.3. Heterogeneous labor inputs

One great thing about deriving theoretical results is that many of them can be applied to new settings. Economists have taken advantage of this, using the CES model to understand income differences between different types of workers. We typically refer to these differences as the *skill premium*. The idea is that workers differ in their *skills* leading them to perform different kinds of *tasks* which in turn differ in their productivity.

It will take us some time to give a proper definition of what skills and tasks are, but for now you can think of the differences between the type of jobs carried out mostly by college educated workers and how they differ from some jobs carried out mostly by workers who do not hold a college degree. These jobs differ in the types of skills used and the specific tasks carried out, but, ultimately, we see that college educated workers have, on average, higher incomes. Therefore, the convention has been to separate to workers into *high* and *low* skill groups to reflect their high and low incomes. When we get to discuss the multi-dimensional nature of skills we will see that there is no a priori sense in which some workers are high- or low-skilled, instead, workers differ in the combination of skills they possess and in the tasks they perform. This leads to differences in incomes, but does not reflect (necessarily) "amounts" of skill.

Now, back to the CES. Let H denote *high-skill* labor and L *low-skill* labor. Output comes from employing both types of labor (we will forget about capital for just a while, but it is hiding in plain sight):

$$Y = \left(\beta (A_H H)^{\frac{\sigma-1}{\sigma}} + (1-\beta)(A_L L)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \qquad \sigma > 0, \quad \sigma \neq 1. \tag{4.19}$$

Our objective is to use the insights we developed before to think about how technological change affects different types of workers. To do this we reinterpret the technological terms A_H and A_L as *skill-biased* technological change. That is, changes in technology or in tools (here is where capital is hiding!) that complement one type of labor. The classical example of skill-biased technological change is the introduction of personal computers to the workplace. This technology complements more "high-skilled" workers than "low-skill" workers, increasing their productivity.

The skill premium. Using the same marginal-product steps as in Section 3, and normalizing the output price to one, competitive wages are

$$w_H = \beta A_H^{\frac{\sigma-1}{\sigma}} \left(\frac{H}{Y}\right)^{-\frac{1}{\sigma}}, \qquad w_L = (1-\beta) A_L^{\frac{\sigma-1}{\sigma}} \left(\frac{L}{Y}\right)^{-\frac{1}{\sigma}}. \tag{4.20}$$

Dividing them gives the relative wage, or skill premium:

$$\frac{w_H}{w_L} = \frac{\beta}{1-\beta} \left(\frac{A_H}{A_L}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{H}{L}\right)^{-\frac{1}{\sigma}}. \tag{4.21}$$

Figure 10 reads the skill-premium equation as a relative-demand curve. At unchanged technology, increasing the supply of high-skill workers moves the equilibrium from A to B, lowering their relative wage. When $\sigma > 1$, an increase in A_H/A_L shifts relative demand upward. If this shift is large enough, the final equilibrium C has both more high-skill workers and a higher skill premium than A. Thus a rising premium need not mean that educated workers have become fewer; demand may have risen faster than their supply.

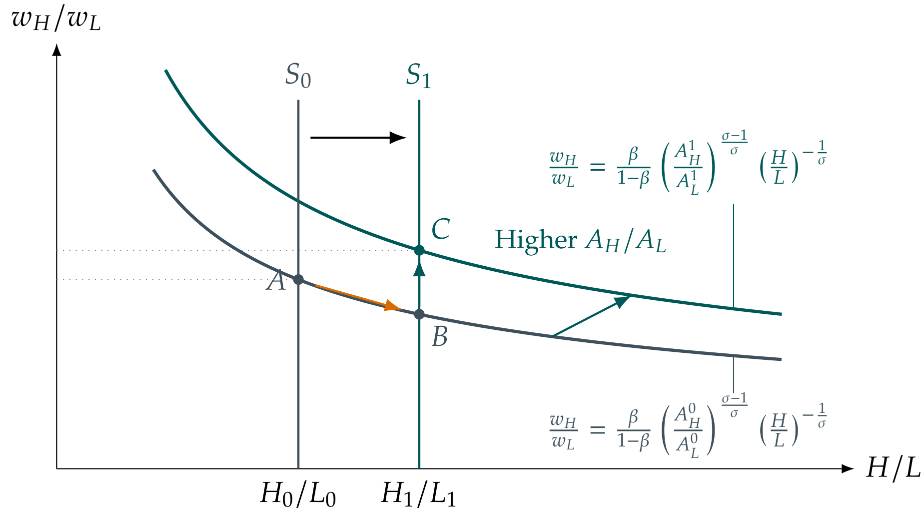
Expressing the skill premium in changes separates it into two forces:

$$\Delta \log \left(\frac{w_H}{w_L}\right) = \frac{\sigma-1}{\sigma} \Delta \log \left(\frac{A_H}{A_L}\right) - \frac{1}{\sigma} \Delta \log \left(\frac{H}{L}\right). \tag{4.22}$$

(a) **Relative supply.** More high-skill workers, $H/L \uparrow$, reduce the skill premium. Skills that become more abundant become relatively cheaper.

(b) **Skill-biased technological change.** When $\sigma > 1$, high-skill-augmenting technology, $A_H/A_L \uparrow$, raises the skill premium. This is because technology shifts demand toward high-skill labor. When $\sigma < 1$ and high- and low-skill labors are hard to substitute,

Figure 10. The race between education and technology



Notes: Vertical lines are relative supplies; downward-sloping curves are relative demands. The demand-curve labels reproduce (4.21), with superscripts 0 and 1 denoting initial and final productivities. The illustration uses $\sigma = 2$ and $\beta = 1/2$. Relative supply rises from $H/L = 1$ to 1.5, while A_H/A_L rises from one to two. The technology shift shown requires $\sigma > 1$.

low-skill labor acts like a *weak link* and skill bias goes down.

This is Tinbergen (1974)'s "race between education and technology." Tinbergen studied how readily graduate workers could be substituted for other workers in production and how this affected their relative earnings. He examined the competing effects of an expanding supply of educated workers and technological change that increased demand for their skills. Education increases the supply of skills; technological change can increase demand for them. The skill premium rises when demand runs ahead of supply and falls when supply runs ahead of demand.

These same forces matter for the differences in income between types of labor. We can look at the ratio of the expenditure in the two labor types the same way we looked at input shares before:

$$\frac{w_H H}{w_L L} = \frac{\beta}{1-\beta}\left(\frac{A_H H}{A_L L}\right)^{\frac{\sigma-1}{\sigma}}. \tag{4.23}$$

Once again, the effect on income differences between input factors depends on whether inputs are easy or hard to substitute, whether σ is higher or lower than one.

Example 4.1 (Supply grows, but the premium also grows). *Suppose the relative supply H/L rises. This is actually what has happened in most countries, as the share of college graduates has increased consistently since the mid twentieth century. By itself this should reduce w_H/w_L. However, the observed skill premium rises instead, equation (4.22) tells us that relative demand must have shifted toward high-skill workers by enough to offset the supply effect.*

Taking Stock. The skill premium responds to relative supplies and relative productivities. The table collects the two results, keeping technology fixed in the supply experiment and physical labor supplies fixed in the technology experiment.

Question and what is fixed	**Relevant derivative**	**Answer and role of σ**
Does greater relative skill supply lower the skill premium? Technology fixed.	$\frac{\partial \log(w_H/w_L)}{\partial \log(H/L)} = -\frac{1}{\sigma}$	Yes for every $\sigma > 0$. A larger σ means a smaller relative wage response.
Does skill-augmenting technology raise the skill premium? H, L fixed.	$\frac{\partial \log(w_H/w_L)}{\partial \log(A_H/A_L)} = \frac{\sigma - 1}{\sigma}$	Yes if $\sigma > 1$; the premium falls if $\sigma < 1$ and is unchanged at $\sigma = 1$.

Testing the model's predictions. We can test if the model is working well to explain the way in which the labor market has transformed by contrasting its predictions with some major trends observed in labor markets:

(a) The college wage premium rose substantially, even as the relative supply of college-educated workers increased.

(b) Real earnings for many workers without a college degree stagnated or fell, particularly for less-educated men.

(c) Employment growth became polarized across occupations: it was strongest in high-wage professional and managerial occupations and in low-wage service occupations, while many middle-wage clerical, production, and administrative occupations contracted.

The first fact fits naturally into Tinbergen's race and the model we have. If the skill premium rose while skilled labor became more abundant, relative demand for skill must have risen even faster. But we know that our model is incompatible with the second fact. Any technological advancement must lead to higher wages for all inputs. This is what we showed in Section 3.3.

The third fact is usually called "job polarization." It refers to a change in the shape of employment across occupations. Employment shares grow near the top and bottom of the occupational wage distribution while shrinking in the middle. The pattern is inconsistent with a strictly monotone mapping from skill rank to employment growth. A one-dimensional skill model does not explain why low-wage service occupations expand while many middle-wage occupations contract.

The routinization hypothesis

The limitations of the CES model call for the development of new theories. One noteworthy theory is the "routinization hypothesis." The key concept that it brings to the table is that how much skill a worker has is not the only thing that matters. We also need to understand but what the worker does in the workplace.

This hypothesis proposes a division of the tasks carried out by workers into two dimensions. One covers *cognitive* and *manual* properties of tasks (think of writing vs moving crates). The second one classifies tasks based on how *routine* they are, that is, how codifiable and repetitive they are. Both cognitive and manual tasks can be routine. For example, Computers and machines are particularly effective at activities that can be described by explicit, stable, and repeatable procedures. These are *routine tasks*. Bookkeeping, quality control, and many production processes can require training and care while still following codifiable rules. Routine does not mean easy or unimportant, but, as we will see more along the course, routine can mean "*automatable*."

	Cognitive	**Manual**
Routine	Bookkeeping, data entry, standard calculations	Repetitive assembly, sorting, predictable machine operation
Non-routine	Problem-solving, persuasion, diagnosis, design	Caregiving, cleaning an unfamiliar space, repairing varied equipment

This division gives us a framework to think about the effects of technological change. Information technology can substitute for routine cognitive and routine manual tasks. At the same time, it can complement non-routine cognitive tasks by making information cheaper to access and process. Many non-routine manual and interpersonal tasks remain difficult to automate because they require physical adaptability, situational judgment, or direct human interaction.

This combination can generate polarization. Technology reduces demand for many routine tasks concentrated in middle-wage occupations, complements abstract tasks concentrated near the top, and leaves demand for many in-person service tasks comparatively intact near the bottom. We will formalize this framework when we develop the task-based model of production in the next Section.

Many researchers have contributed to this literature, but the routinization hypothesis is chiefly associated with the work of David Autor and his coauthors (Autor, Levy, and Murnane 2003). Related contributions include Autor et al. (2006, 2008) on polarization and Autor and Dorn (2013) on low-skill service employment.

Part II

Task-Based Theories of Production

This second part of the course develops the canonical task-based model of production, which allows us to study the effects of automation on labor demand and the distribution of income between labor and capital. The key for us is to distinguish between the *amounts of inputs* used in production and *what they do in production*. This adds a new layer of substitutability between inputs as tasks can be reassigned from one input to another in response to changes in prices or technology. The main sources for this part of the course are the lecture notes and the following references: Acemoglu and Autor (2011), Restrepo (2024), and Acemoglu, Kong, and Restrepo (2025).

5. Task-Based Model of Production

In the previous section we distinguished workers by their skills, but we did not model what they do. We also took capital as an aggregate input for production, while not specifying how it is actually used in the workplace. In practice, the type of technology we covered made one type of labor more productive in every activity, and the production function determined how easily that labor could replace another input. We now open up this production function. Our objective is to describe the activities needed for production and then determine who performs them. We call these activities *tasks*.

This gives us another way to think about substitution. A firm can respond to a higher wage by using less of the tasks performed by the workers whose wage has increased. But it can also assign some of those same tasks to other workers or to machines. The task-based model lets us distinguish these two responses.

So, who (or what) should perform each task? Part of the answer lies in who is the most productive at performing the task, but productivity alone cannot give us the answer. Not because a robot is better than person at doing something we should immediately automate (assign the task to the robot). Whether we assign the task to the person or the robot is also influenced by relative cost (we are back to w/r!). It turns out, Economists have long studied this type of scenario. The most famous case is that of international trade and is the origin the of the principle of *comparative advantage* developed by David Ricardo in the 1800s.

The Ricardian model of trade is still used to understand how production is organized across countries. Modern models of trade are evolutions of the basic framework in Dornbusch, Fischer, and Samuelson (1977). Labor economists have adapted the assignment problem of Ricardian trade to the problem of tasks. In the task model, skill groups (including machines) take the place of countries and tasks take the place of goods: each task goes to the group with the lowest unit cost. Differences in relative productivity across tasks and equilibrium wages jointly determine the assignment boundaries.

5.1. Occupations as task bundles

Before we turn to the mathematical derivations of the canonical task model we will consider a more general setting to understand how tasks help us think about how work is organized. At the core of this is the concept of *occupations*.

An occupations is simple a bundle of tasks performed by a type of worker

While not two workers ever perform the exact same combination of tasks, we can see regularities that lead us to label some of these task-bundles as occupations.

Take the work of a nurse. It involves monitoring patients, administering treatment, keeping records, and communicating with patients and other medical staff. Some of these tasks are similar to the tasks that a doctor performs, but they can differ in their complexity—the degree of skill or knowledge required to perform them. Some of the tasks were actually previously performed by doctors, but the assignment has changed over time.

What, then, determines which activities belong together in a job, and why does that bundle change?

Changes in the skill composition of workers are part of the reason. The supply of nurses and doctors has changed over time, as has the training they each get. But technology also changes. Software that reduces the work needed to keep records changes part of the job. New surgical techniques reduce the time-demands on doctors, but also allow them to perform a wider range of procedures. This, in turn, entails new tasks for nurses.

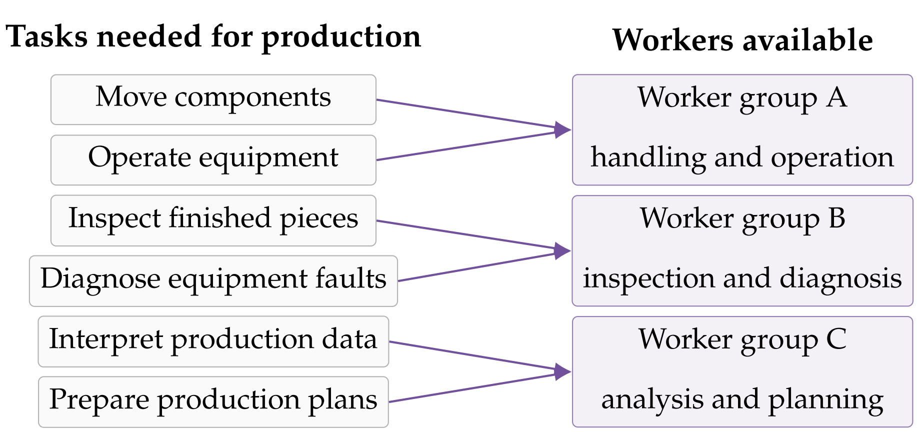
Similar changes take place in manufacturing, where plant operators face industrial robots who can perform task previously assigned to workers, but that require maintenance and allow workers to perform new tasks. Engineers face new software that changes the way they design and organize operation in the plant. An occupation is therefore more than a fixed label attached to a worker. To understand its response to technology, we need to explain how its tasks are bundled in the first place.

The manager's assignment problem. Imagine a manager organizing production in a plant (or a hospital). The plant needs components moved, equipment operated, faults diagnosed, quality checked, and production plans prepared. These activities draw on different combinations of skills. Operating equipment may require manual dexterity as well as attention and judgment; diagnosing a fault may require both practical experience and analytical reasoning.

The problem of the manager is to figure who (or what) should be assigned each task. Part of the problem is that workers also bring different combinations of skills or abilities. Someone adept at handling equipment need not be equally adept at interpreting data, and an excellent analyst need not be the best person to handle a delicate component. There need not be a single ranking from the least to the most skilled worker that applies to every activity. This issue reflects the multidimensional nature of the skills involved in production.

Giving every activity to the person who performs it best in isolation may be impossible: that person's time is limited and is also useful elsewhere. A good assignment uses the available workforce to produce as much as possible, balancing the benefits of a good match against the need to cover the other tasks.

Figure 11. Tasks assigned to worker groups



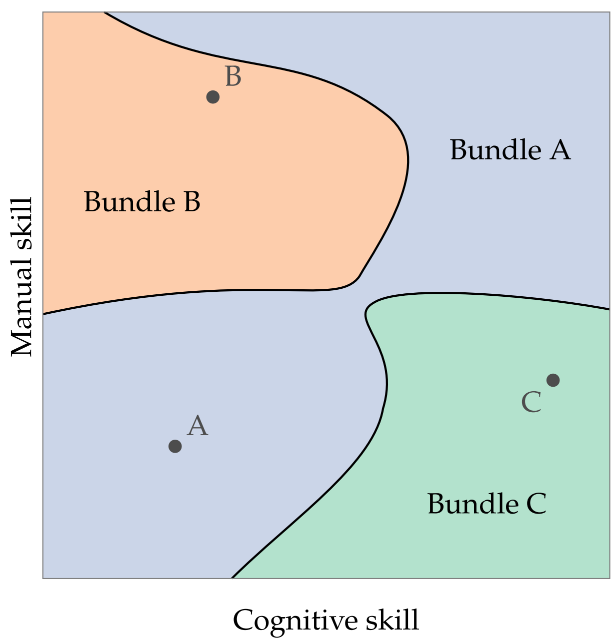
Notes: Each arrow assigns a task to a worker group. The collection of tasks assigned to that group forms its occupation; the task bundles shown are illustrative.

The collection of tasks assigned to each worker group is the group's occupation. Figure

11 gives an example of this. Crucially, the bundle is itself outcome of organizing production, rather than a restriction that the manager must take as given. So, as technology changes so does the assignment. The same worker can keep some tasks and acquire or lose others.

A map of tasks and occupations. We can think of the assignments of tasks to workers more generally, as in Figure 12 originally from Ocampo (2026). For example, think of tasks that require cognitive and manual skills to be performed. There are many tasks, each with a different composition of skills. A location in the square describes the cognitive and manual skills involved in a given task. Similarly, workers differ in the skills they posses. The dots in the square show the skill combinations of three worker types. Occupations are formed by the bundles of tasks assigned to each type of worker.

Figure 12. Occupations as regions of task space



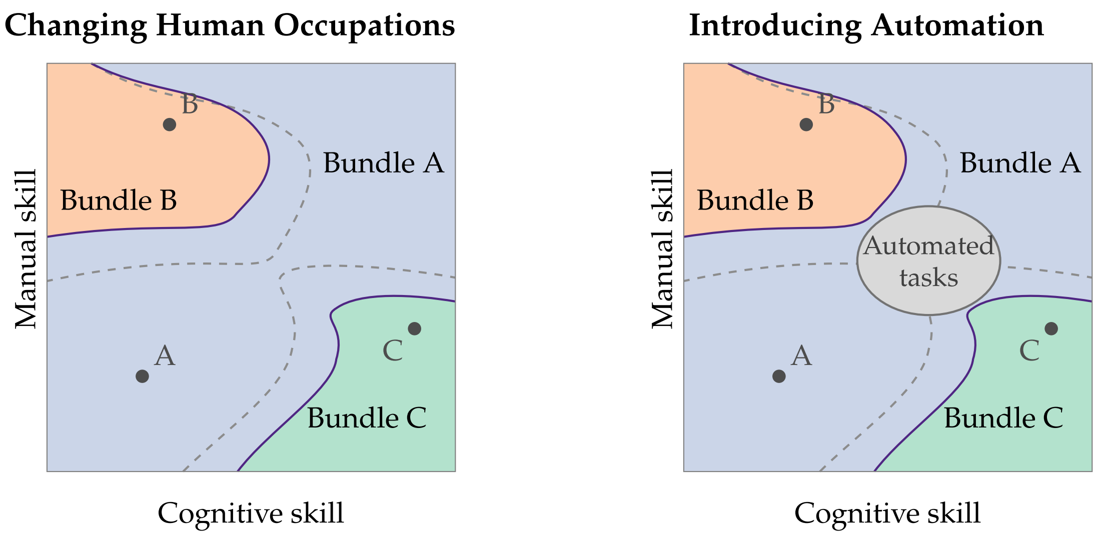
Notes: Locations represent the cognitive and manual skill requirements of tasks. Dots indicate worker skill combinations, and colored regions collect the tasks assigned to each worker type. The map is an illustrative redraw of the framework in Ocampo (2026).

How good an assignment is depends on the *skill mismatch* between workers and the tasks in their occupation and the availability of labor. We can see that as the distance from a task to a worker's skill combination. Technology determines how costly different kinds of mismatch are, and worker availability constrains the assignment; we cannot simply assign every task to the closest worker. It might be that cognitive mismatch matters more than manual mismatch, and having more skill than a task needs may have a different effect from having too little. The boundaries of the bundles summarize the resulting division of work.

Technology can redraw occupations. We will use this framework to think about how technology changes occupations. Software used by engineers, for example, may change which parts of analysis, checking, and planning they perform and which parts are done by other workers. The left panel of Figure 13 illustrates a change in the boundaries between human occupations. The worker skills stay fixed, while some tasks change hands because technology has changed how *skill mismatch* affects production. Changes in training or in the availability of different workers can also alter the assignment by altering workers' skills.

The framework also lets us ask how automation displaces workers from tasks. Self-checkout machines take over some cashier tasks, while trading and accounting software take over some tasks of stockbrokers and accountants. Most notably, industrial robots can take particular positions in an assembly process. These changes can remove part of an occupation's work without removing the whole occupation.

Figure 13. Changes in the division of work



Notes: The left panel shows a change in the boundaries between human occupations; the right adds a gray region assigned to machines. Dashed lines are original boundaries and solid lines illustrative new boundaries. Worker skill locations are held fixed.

In the right panel, tasks in the gray region are performed by machines or software instead of workers. Workers displaced from these tasks may move into other activities, changing the bundles performed by neighboring workers. This introduces an indirect transmission mechanism: workers whose own tasks are not automated can still be affected by the reassignment as highlighted in Ocampo (2026); Acemoglu and Restrepo (2022).

5.2. Canonical task-based model: capital and labor

To set up this model we need to describe two levels of production: how tasks combine into final output and how inputs (capital and labor) produce each task. There are many tasks involved in production, which we capture by having tasks be indexed by $i \in [0, 1]$. The quality or level of a task depends on who or what performs it.

Final output is

$$Y = \left(\int_0^1 y(i)^{\frac{\eta-1}{\eta}} \, di \right)^{\frac{\eta}{\eta-1}}, \qquad \eta > 0, \quad \eta \neq 1, \tag{5.1}$$

where $y(i)$ denote the quality or quantity of task service i and η is the elasticity of substitution *across tasks*. We use η rather than σ because the elasticity between aggregate inputs and tasks need not be the same. When $\eta < 1$, tasks are hard to substitute for each other: producing much more of some activities does not easily compensate for producing little of others. When $\eta > 1$, this substitution is easier. When $\eta \to 1$ we obtain the continuous Cobb–Douglas:

$$\log Y = \int_0^1 \log y(i)\, di. \tag{5.2}$$

Tasks are performed using capita K or labor L, which cost r and w per unit, respectively. The capital used for task i is $k(i)$ and the labor used in that task is $\ell(i)$. The task quality depends on how it is performed and is

$$y(i) = a_K(i)k(i) + a_L(i)\ell(i). \tag{5.3}$$

Here $a_K(i) > 0$ is the productivity of capital in task i and $a_L(i) > 0$ the prodcutivity of labor. These functions encapsulate the mismatch between the skills of the input factor and the demands of the task when being performed.

Crucially, the two inputs are perfect substitutes *within each task*. The substitutability of inputs in production depends on how tasks are combined and assigned, not on their substitutability in a given task.

5.3. Task assignment and comparative advantage

The firm chooses the capital and labor allocated to every task, jointly determining which input performs each task and the quality of the tasks. Taking the rental rate r, and the wage w as given, it solves

$$\min_{\{k(i),\ell(i)\geq 0\}} \int_0^1 (rk(i) + w\ell(i))\, di \quad \text{s.t. } \overline{Y} \leq \left(\int_0^1 (a_K(i)k(i) + a_L(i)\ell(i))^{\frac{\eta-1}{\eta}}\, di \right)^{\frac{\eta}{\eta-1}}. \tag{5.4}$$

To obtain the cost-minimization conditions for task i, notice that an additional unit of $y(i)$ raises final output by

$$\frac{\partial Y}{\partial y(i)} = \left(\frac{Y}{y(i)}\right)^{1/\eta}. \tag{5.5}$$

One additional unit of capital produces $a_K(i)$ units of that service; one additional unit of labor produces $a_L(i)$. Applying the chain rule, the optimality conditions at a positive-output allocation are

$$\lambda\left(\frac{Y}{y(i)}\right)^{1/\eta} a_K(i) \leq r, \qquad \text{with equality if } k(i) > 0, \tag{5.6}$$

$$\lambda\left(\frac{Y}{y(i)}\right)^{1/\eta} a_L(i) \leq w, \qquad \text{with equality if } n(i) > 0. \tag{5.7}$$

If the input its used, then it is demanded so that its marginal contribution to revenue equals its marginal cost (recall that λ is the marginal cost). An unused input cannot contribute more to revenue than it costs.

The principle of comparative-advantage. The term $P(Y/y(i))^{1/\eta}$ is common to both inputs. Consequently, the firm assigns a task to the input that produces more of its service per dollar: capital if $a_K(i)/r > a_L(i)/w$, and labor if the inequality is reversed.

Alternatively, we compare costs. A unit of task output requires $1/a_K(i)$ capital costing $r/a_K(i)$ or $1/a_L(i)$ units of labor costing $w/a_L(i)$. The cheaper input performs the task:

$$\underbrace{\frac{r}{a_K(i)} < \frac{w}{a_L(i)}}_{\text{capital is cheaper}} \quad \longleftrightarrow \quad \underbrace{\frac{a_L(i)}{a_K(i)} < \frac{w}{r}}_{\text{relative productivity versus relative price}}. \tag{5.8}$$

We can make this clearer if we impose an additional assumption that gives the task-index meaning by ordering tasks according to comparative advantage. Suppose that machines are better at low-index tasks and bad at high-index tasks, while humans are the opposite, better at high-index tasks and worse at low-index ones. Then, capital productivity is decreasing and labor productivity is increasing along the task index:

$$a_K'(i) \ < \ 0, \qquad a_L'(i) \ > \ 0. \tag{5.9}$$

These assumptions imply that labor's relative productivity is increasing:

$$q(i) \ \equiv \ \frac{a_L(i)}{a_K(i)}, \qquad q'(i) \ = \ \frac{a_L'(i)a_K(i) \ - \ a_L(i)a_K'(i)}{a_K(i)^2} \ > \ 0. \tag{5.10}$$

Crucially, this assumption is only about the change in productivity, not its level. Moving right makes labor better in absolute terms and capital worse, so labor necessarily becomes better *relative to capital* but not necessarily more productive than capital. We further assume that $q(0) \ < \ w/r \ < \ q(1)$ so that both inputs are used by the firm. If w/r lies below $q(0)$ all tasks go to labor; if it lies above $q(1)$ all go to capital.

The fact that the relative productivity q is increasing means that there is a unique interior cutoff I satisfying

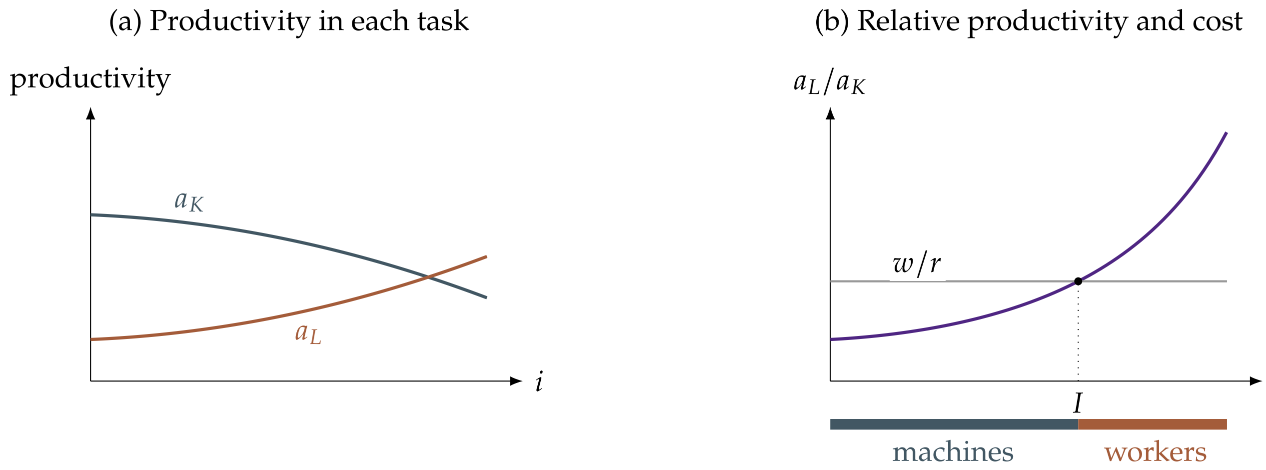
$$q(I) \ = \ \frac{a_L I}{a_K(I)} \ = \ \frac{w}{r} \tag{5.11}$$

Thus $[0, I]$ is the machine bundle, with $q(i) \leq w/R$, and $[I, 1]$ is the labor bundle, with $q(i) \geq w/R$. This is what we see in Figure 14. Tasks to the left of I lie below w/r, so capital is cheaper. Tasks to its right lie above w/r, so labor is cheaper.

Comparative advantage is different from absolute advantage. Capital has an absolute productivity advantage at a task if $a_K(i) > a_L(i)$. But this does not determine the assignment unless $w = r$. The following example helps make this clear. Let $a_K(i) \ = \ 2 - i$, $a_L(i) = 0.5 + i$, and $w/r \ = \ 0.6$. Notice that capital is more productive at all tasks $i < 3/4$. The cost cutoff is

$$\frac{0.5 + I}{2 - I} \ = \ 0.6 \quad \longrightarrow \quad I \ = \ \frac{7}{16} \ = \ 0.4375. \tag{5.12}$$

Figure 14. Optimal Task Assignment



Notes: Decreasing capital productivity and increasing labor productivity imply increasing relative labor productivity. The assignment cutoff compares this ratio with w/r. Prices are held fixed for illustration.

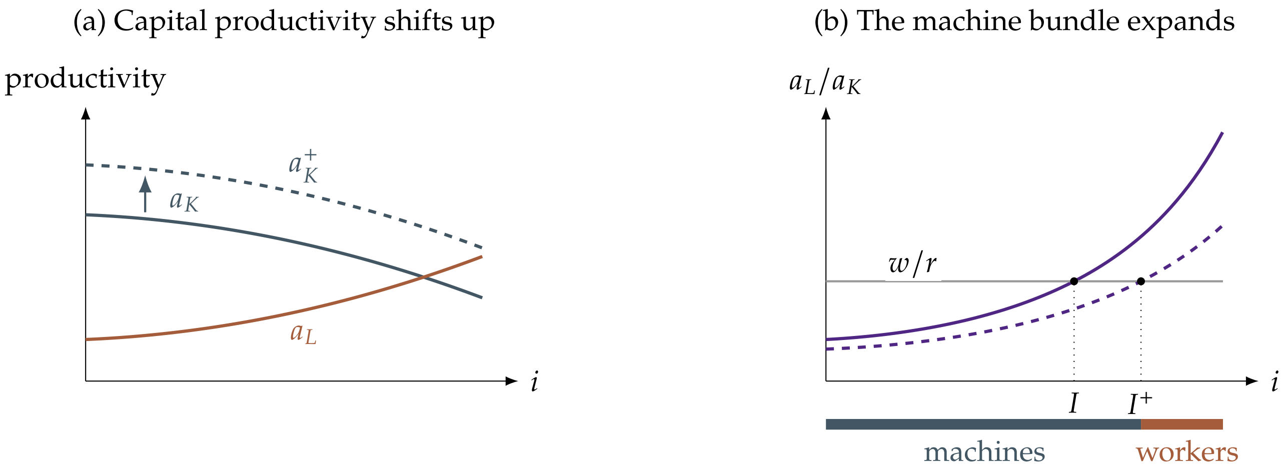
In the optimal assignment, workers perform some tasks at which machines are more productive because workers are sufficiently cheaper.

Now shift capital productivity up to $a_K^+(i) = 3 - i$, holding $w/r = 0.6$. Capital is more productive than labor at *every* task, but labor still performs the highest-index tasks

$$\frac{0.5 + I^+}{3 - I^+} = 0.6 \quad \longrightarrow \quad I^+ = \frac{13}{16} = 0.8125 < 1. \tag{5.13}$$

Absolute advantage everywhere does not eliminate the other input's comparative advantage or its task bundle.

Figure 15. Changes in Absolute Advantage



Notes: Absolute advantage everywhere can coexist with a human task bundle. Dashed curves show higher capital productivity. At unchanged w/r, the cutoff rises from I to I^+; labor still performs the remaining tasks.

5.4. Input demand and input shares

With the assignment at hand we now know *how much* of each task the firm produces. Using equality in (5.6) on the machine bundle and in (5.7) on the human bundle gives

$$y(i) = \begin{cases} Y\left(\dfrac{Pa_K(i)}{r}\right)^{\eta}, & i < I, \\ Y\left(\dfrac{Pa_L(i)}{w}\right)^{\eta}, & i > I. \end{cases} \tag{5.14}$$

where we replace λ for P, the price of the final good.

We also know the demand for capital and labor:

$$\begin{aligned} K &= \int_0^I \frac{y(i)}{a_K(i)}\,di = YP^{\eta}r^{-\eta}\underbrace{\int_0^I a_K(i)^{\eta-1}\,di}_{H_K(I)}, \\ L &= \int_I^1 \frac{y(i)}{a_L(i)}\,di = YP^{\eta}w^{-\eta}\underbrace{\int_I^1 a_L(i)^{\eta-1}\,di}_{H_L(I)}. \end{aligned} \tag{5.15}$$

Where H_K and H_L are productivity indexes that capture how productive capital and labor are, respectively, in the tasks they are assigned to (hence the indexes depend on I).

From productivity schedules to expenditure areas. To connect the integrals to a picture, multiply the input used in each task by its price. Expenditure per unit of task length is

$$\begin{cases} rk(i) = YP^{\eta} r^{1-\eta} a_K(i)^{\eta-1}, & i < I, \\ w\ell(i) = YP^{\eta} w^{1-\eta} a_L(i)^{\eta-1}, & i > I. \end{cases} \tag{5.16}$$

Figure 16 maps the productivity schedules on the left into these expenditure schedules on the right. Their heights depend on productivity, factor prices, output, and the final-good price P, which equals unit cost. The blue area sums $rk(i)$ over machine tasks and is therefore rK; the orange area sums $w\ell(i)$ over human tasks and is wL. Dividing either area by the total area PY gives its factor share.

When $\eta = 1$, the productivity powers vanish and both expenditure schedules have the same height PY. Area shares then equal lengths: $\alpha_K = I$ and $\alpha_L = 1 - I$. For $\eta \neq 1$, task lengths alone do not determine the expenditure shares. In the $\eta = 1/2$ example, tasks with lower productivity have higher expenditure per unit of task length because they are costly and hard to substitute away from. For $\eta > 1$, the productivity exponent is positive and the direction reverses.

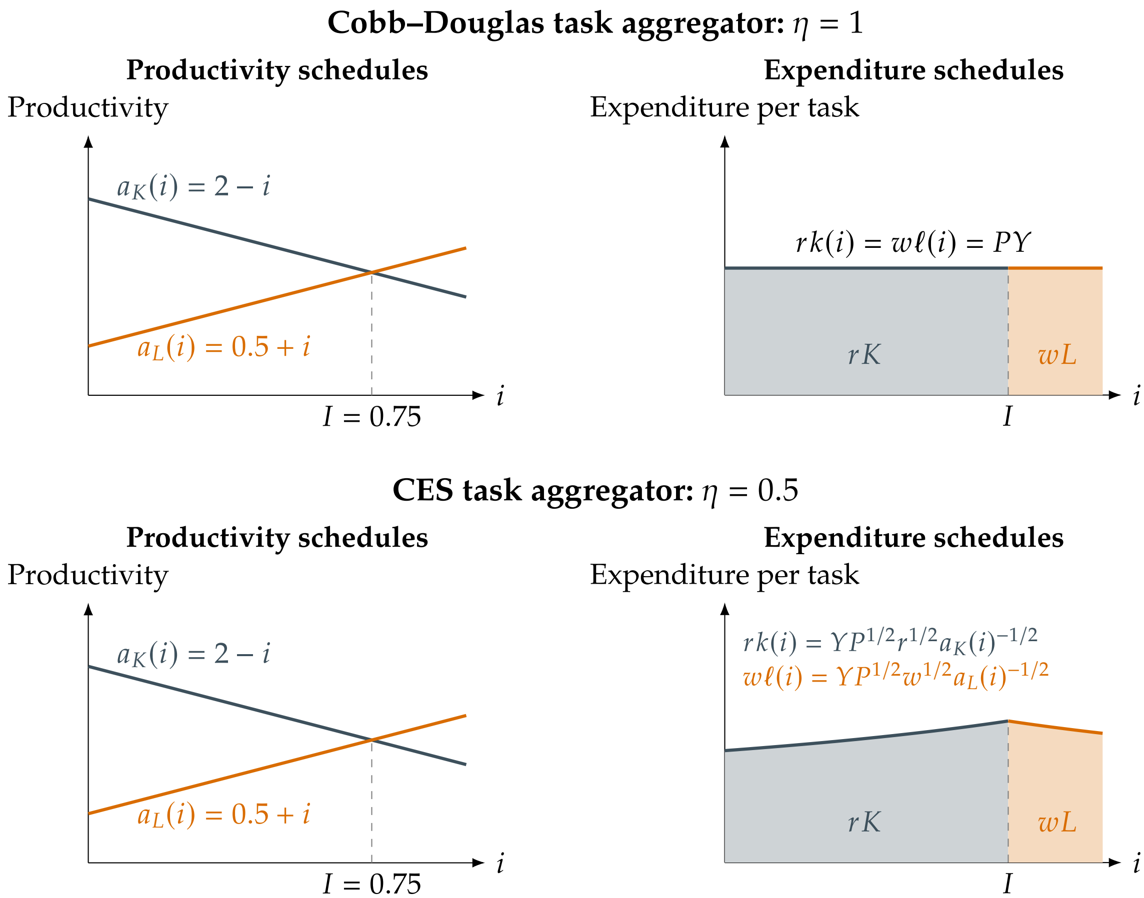
Multiplying input demands by their prices gives input shares:

$$\alpha_K \equiv \frac{rK}{PY} = \left(\frac{r}{P}\right)^{1-\eta} H_K(I), \qquad \alpha_L \equiv \frac{wL}{PY} = \left(\frac{w}{P}\right)^{1-\eta} H_L(I). \tag{5.17}$$

When $\eta = 1$ we aggregate with a Cobb-Douglas technology and optimality conditions give

$$rk(i) = PY \quad (i < I), \qquad wn(i) = PY \quad (i > I). \tag{5.18}$$

Figure 16. From task productivity to factor payments



Notes: Both rows use $a_K(i) = 2 - i$, $a_L(i) = 0.5 + i$, $r = w = Y = 1$, and hence $I = 0.75$. The final-good price P is the unit cost implied by each row's task aggregator; it is not independently held fixed across rows. Right-panel heights follow (5.16); shaded areas are total factor payments.

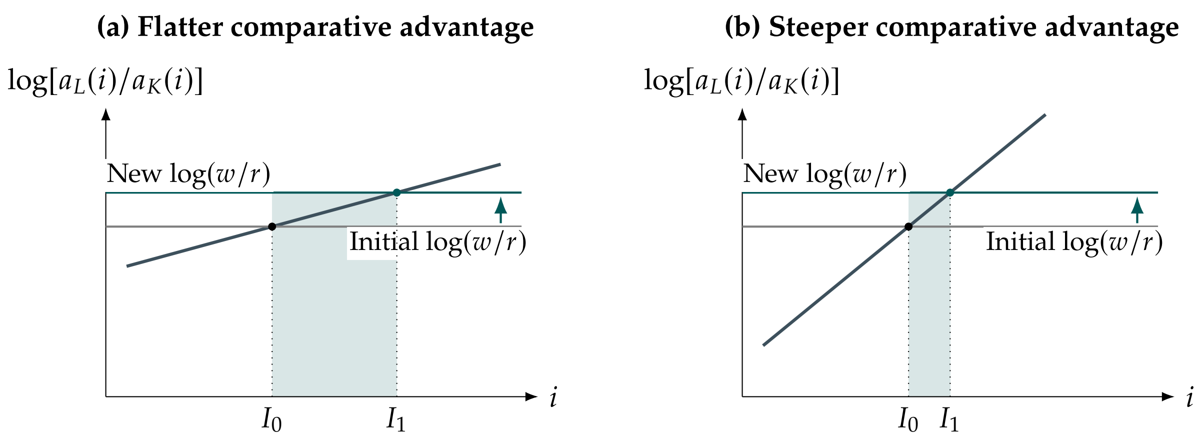
Every equal-length task interval then receives the same input payment. Integrating gives the special result $\alpha_K = I$ and $\alpha_L = 1 - I$.

5.5. Comparative statics of task assignment

We can now look at how the firm's cost-minimizing choices respond to the relative input price w/r. This response is summarized by the change in the assignment cutoff I, which satisfies $a_L(I)/a_K(I) = w/r$. Figure 17 shows what determines the size of this response. A fall in r, with w fixed, raises the horizontal $\log(w/r)$ line by the same amount in both panels.

When log relative productivity changes slowly across nearby tasks, the new intersection is much farther to the right. Many tasks are close substitutes in who can perform them cheaply, so many change hands. A steep schedule instead means sharp differences in comparative advantage, and few tasks switch. The shaded widths measure the change in the assignment cutoff.

Figure 17. Comparative advantage and the reassignment response



Notes: The panels have the same initial cutoff I_0 and the same increase in $\log(w/r)$, but different slopes of log relative productivity at the cutoff. Technology is held fixed. The flatter comparative-advantage schedule in the left panel generates a larger reassigned interval than the steeper schedule in the right panel.

Taking logs and differentiating gives

$$\left[\frac{a_L'(I)}{a_L(I)} - \frac{a_K'(I)}{a_K(I)}\right] dI \;=\; d\log(w/r) \quad\longrightarrow\quad \frac{dI}{d\log(w/r)} = \frac{1}{\dfrac{a_L'(I)}{a_L(I)} - \dfrac{a_K'(I)}{a_K(I)}} > 0. \tag{5.19}$$

The denominator is positive because labor's relative productivity increases with the task index. A higher wage relative to the rental rate therefore expands the machine bundle and contracts the human bundle. The response is larger when comparative advantage changes less sharply around the cutoff.

We can also express this result using the elasticity of each productivity schedule with respect to the task index:

$$\varepsilon_j(i) \;\equiv\; \frac{d\log a_j(i)}{d\log i} \;=\; \frac{i\,a_j'(i)}{a_j(i)}, \qquad j\in\{K,L\}, \qquad \frac{d\log I}{d\log(w/r)} \;=\; \frac{1}{\varepsilon_L(I)-\varepsilon_K(I)}. \tag{5.20}$$

Under our assumptions, $\varepsilon_L(I) > 0$ and $\varepsilon_K(I) < 0$, so the gap in these elasticities is positive.

This result is particularly useful to characterize how the aggregate demand for capital and labor reacts to changes in prices.

How does reassignment affect substitution between the aggregate inputs?

Dividing the input demands in (5.15) cancels their common output and price terms:

$$\frac{K}{L} \;=\; \left(\frac{w}{r}\right)^{\eta}\frac{H_K(I)}{H_L(I)}. \tag{5.21}$$

The limits of integration imply $H_K'(I) = a_K(I)^{\eta-1}$ and $H_L'(I) = -a_L(I)^{\eta-1}$. We use this to get the elasticity of substitution with task assignment:

$$\begin{aligned}\sigma \;\equiv\; \frac{d\log(K/L)}{d\log(w/r)} \;&=\; \underbrace{\eta}_{\text{fixed assignment}} \;+\; \underbrace{\left[\frac{a_K(I)^{\eta-1}}{H_K(I)} + \frac{a_L(I)^{\eta-1}}{H_L(I)}\right]\frac{dI}{d\log(w/r)}}_{\text{task reassignment}}, \qquad (5.22)\\ &=\; \eta + \frac{I}{\varepsilon_L(I)-\varepsilon_K(I)}\left[\frac{a_K(I)^{\eta-1}}{H_K(I)} + \frac{a_L(I)^{\eta-1}}{H_L(I)}\right] \;>\; \eta.\end{aligned}$$

At a fixed assignment, the firm substitutes between the quantities of machine-produced and human-produced tasks with elasticity η. Allowing assignment to change adds a positive response: capital takes over marginal tasks previously performed by labor. Thus aggregate inputs are more substitutable than tasks at an interior cutoff, although σ need not be constant or exceed one when $\eta < 1$.

For example, with Cobb–Douglas task aggregation, $\eta = 1$, we have $H_K(I) = I$ and

$H_L(I) = 1 - I$, so

$$\sigma \;=\; 1 + \frac{1}{(1-I)} \frac{d\log I}{d\log(w/r)}. \tag{5.23}$$

The productivity schedules in Section 6.1, after relabeling its two labor inputs as capital and labor, imply $d\log I/\,d\log(w/r) = \kappa(1-I)$ and hence $\sigma = 1+\kappa$. Section 6.2 instead shows how reassignment can raise an across-task elasticity $\eta < 1$ to an aggregate elasticity of exactly one.

The same reasoning applies to study changes in technology. Write task productivity as

$$a_K(i) = \bar{a}_K \alpha_K(i), \qquad a_L(i) = \bar{a}_L \alpha_L(i), \tag{5.24}$$

where $\bar{a}_K, \bar{a}_L > 0$ scale productivity across all tasks and the fixed schedules $\alpha_K(i), \alpha_L(i)$ describe its task-specific pattern. The cutoff condition becomes

$$\frac{\alpha_L(I)}{\alpha_K(I)} \;=\; \frac{w}{r}\frac{\bar{a}_K}{\bar{a}_L} \quad\longrightarrow\quad \big[\varepsilon_L(I) - \varepsilon_K(I)\big]\, d\log I \;=\; d\log(w/r) + d\log(\bar{a}_K/\bar{a}_L). \tag{5.25}$$

Multiplying a productivity schedule by a constant does not change its elasticity with respect to the task index, so the same $\varepsilon_j(I)$ apply. Holding w/r fixed therefore gives

$$\left.\frac{d\log I}{d\log(\bar{a}_K/\bar{a}_L)}\right|_{w/r} \;=\; \frac{1}{\varepsilon_L(I) - \varepsilon_K(I)} \;>\; 0. \tag{5.26}$$

An increase in $\bar{a}_K$ lowers capital's cost of performing every task and expands the machine bundle. An increase in $\bar{a}_L$ has the opposite effect. For assignment, a proportional increase in $\bar{a}_K/\bar{a}_L$ has the same effect as the same proportional increase in w/r.

Cheaper machines versus more productive machines. Figure 18 separates changing task quantities from changing who performs the tasks. First lower the rental price from r to r/m, where $m > 1$, holding w, technology, and final output fixed. Even if assignment is frozen, the firm changes the relative quantities of machine-produced and human-produced services. Allowing assignment to adjust also transfers a band of tasks from labor to capital, adding the reassignment response in (5.22).

Now compare this with keeping r fixed and multiplying the entire capital-productivity schedule by the same m. At every task, the two experiments give the same machine unit cost:

$$\frac{r/m}{a_K(i)} = \frac{r}{m a_K(i)}. \tag{5.27}$$

Consequently, they give the same assignment, task-service quantities, final-good unit cost, and labor use at the same output target. But they do not give the same physical capital use: producing each machine task service with the better technology requires $1/m$ as much capital as producing it with the cheaper rental price. In terms of local responses to m,

$$\left.\frac{d\log(K/L)}{d\log m}\right|_{r\mapsto r/m} = \sigma, \qquad \left.\frac{d\log(K/L)}{d\log m}\right|_{a_K\mapsto m a_K} = \sigma - 1. \tag{5.28}$$

The fixed-assignment components are respectively η and $\eta - 1$; the additional reassignment component is the same in both experiments. This is why equivalent improvements in the cost of machine services need not have the same effect on the number of machines used.

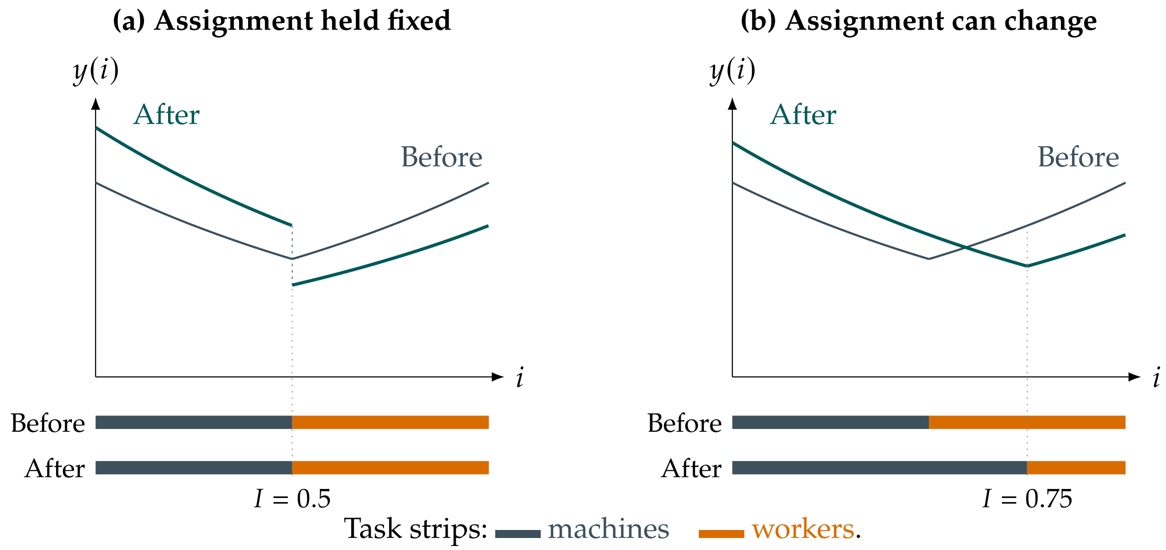
5.6. The link to aggregate production functions

For a given assignment, holding fixed I, we can show that the task-based production function has a CES representation following Jones and Tonetti (2026). The assignment divides tasks between capital and labor. For each task assigned to capital, $i \in [0, I]$, we can write the amount of capital assigned to that task as

$$k(i) \; = \; a_K(i)^{\eta-1}\frac{K}{H_K(I)}. \tag{5.29}$$

Figure 18. Task quantities and task reassignment

The same task-service curves result from $r \longrightarrow r/m$ or $a_K(i) \longrightarrow m a_K(i)$, $m > 1$.



Notes: Task-service quantities change in both panels; only the right panel allows the assignment boundary to move. The example uses $\eta = 1$, $Y = w = r = 1$ initially, $a_K(i) = e^{-i}$, $a_L(i) = e^{i-1}$, and $m = e^{1/2}$. The initial cutoff is $I = 0.5$ and the new unrestricted cutoff is $I = 0.75$. Each panel's output price adjusts to its unit cost. Both experiments give the plotted services, but improved productivity uses fewer physical machines than the rental-price cut.

We can then use this to write capital's total contribution to the task aggregator as

$$\int_0^I [a_K(i)k(i)]^{(\eta-1)/\eta}\, di \;=\; K^{(\eta-1)/\eta} H_K(I)^{1/\eta} \;=\; [A_K(I)K]^{(\eta-1)/\eta}. \tag{5.30}$$

where we define

$$A_K(I) \;=\; H_K(I)^{\frac{1}{\eta-1}}, \qquad A_L(I) \;=\; H_L(I)^{\frac{1}{\eta-1}}. \tag{5.31}$$

We use A_L to follow the same steps and obtain

$$Y \;=\; \left([A_K(I)K]^{\frac{\eta-1}{\eta}} + [A_L(I)L]^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}, \qquad \alpha_K \;=\; \left(\frac{A_K(I)K}{Y}\right)^{\frac{\eta-1}{\eta}}. \tag{5.32}$$

This looks like the CES function from Section 3, but $A_K(I)$ and $A_L(I)$ depend on the task assignment. Holding I fixed gives elasticity η; allowing the cutoff to move adds substitution through reassignment. Thus the aggregate elasticity σ need not equal η. The special cases below show this explicitly. Standard factor-share formulas remain valid at the optimal assignment: the marginal task has equal unit costs, so a small boundary reallocation has no first-order effect on optimized output by itself.

5.7. Tasks cheatsheet

The tables collect the capital–labor task model's definitions and main results. Assume positive productivities, competitive production, $\eta > 0$, and an interior cutoff $0 < I < 1$ with $q'(I) > 0$. The final-good price P equals unit cost; η measures substitution across tasks and σ substitution between aggregate inputs.

Object	Mathematical expression and meaning
Task aggregator	$Y = \left(\int_0^1 y(i)^{\frac{\eta-1}{\eta}}\,di\right)^{\frac{\eta}{\eta-1}}$; at $\eta = 1$, $\log Y = \int_0^1 \log y(i)\,di$. Task services $y(i)$ combine into final output.
Production within a task	$y(i) = a_K(i)k(i) + a_L(i)\ell(i)$. $a_K(i), a_L(i)$ are task productivities; inputs are perfect substitutes within a task.
Comparative advantage and cutoff	$q(i) = \dfrac{a_L(i)}{a_K(i)}, \qquad q(I) = \dfrac{w}{r}$. Capital performs $i < I$, where $r/a_K(i) < w/a_L(i)$; labor performs $i > I$.
Bundle indexes	$H_K(I) = \int_0^I a_K(i)^{\eta-1}\,di, \qquad H_L(I) = \int_I^1 a_L(i)^{\eta-1}\,di$. These sum productivity weights over each input's assigned tasks.
Task-service demand	$y(i) = Y\left(\dfrac{Pa_K(i)}{r}\right)^{\eta}\ (i < I), \qquad y(i) = Y\left(\dfrac{Pa_L(i)}{w}\right)^{\eta}\ (i > I)$. Cheaper task services are used more intensively.
Aggregate inputs	$K = YP^{\eta}r^{-\eta}H_K(I), \qquad L = YP^{\eta}w^{-\eta}H_L(I)$. Hence $\dfrac{K}{L} = \left(\dfrac{w}{r}\right)^{\eta}\dfrac{H_K(I)}{H_L(I)}$.
Factor shares	$\alpha_K = \dfrac{rK}{PY} = \left(\dfrac{r}{P}\right)^{1-\eta}H_K(I), \quad \alpha_L = \dfrac{wL}{PY} = \left(\dfrac{w}{P}\right)^{1-\eta}H_L(I)$. Shares add to one. At $\eta = 1$, $\alpha_K = I$ and $\alpha_L = 1 - I$.
CES representation at fixed assignment	$A_K(I) = H_K(I)^{\frac{1}{\eta-1}}, \quad A_L(I) = H_L(I)^{\frac{1}{\eta-1}}$, for $\eta \neq 1$. $Y = \left([A_K(I)K]^{\frac{\eta-1}{\eta}} + [A_L(I)L]^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}$. Effective factor productivities depend on which tasks each input performs.

Main results: assignment, substitution, and shares. Derivatives are local and evaluated at the initial interior allocation. Unless stated otherwise, productivity schedules are fixed; for input-demand comparisons, final output is held fixed and P adjusts with unit cost.

Question and what is fixed	Relevant relationship	Answer and intuition
Does absolute advantage determine task assignment? Factor prices given.	$\frac{r}{a_K(i)} < \frac{w}{a_L(i)}$	No. Capital gets a task when its cost per unit of service is lower, even if labor is more productive in absolute terms.
How does cheaper capital change assignment? w and technology fixed.	$\frac{dI}{d\log(w/r)} = \frac{q(I)}{q'(I)} > 0$	A fall in r raises w/r and expands the machine bundle. A flatter relative-productivity schedule gives more reassignment.
Does better capital expand its task bundle? Prices fixed; $a_K(i)$ rises proportionally in all tasks.	$\frac{d\log I}{d\log \bar{a}_K} = \frac{1}{\varepsilon_L(I) - \varepsilon_K(I)}$	Yes. It lowers machine task costs. Here $\varepsilon_j(i) = ia_j'(i)/a_j(i)$ measures the slope of a productivity schedule.
How much substitution is possible if assignment is fixed?	$\left.\frac{d\log(K/L)}{d\log(w/r)}\right\|_I = \eta$	Only task quantities adjust. Substitution between aggregate inputs equals substitution across tasks.
What does reassignment add?	$\sigma = \eta + \left[\frac{a_K(I)^{\eta-1}}{H_K(I)} + \frac{a_L(I)^{\eta-1}}{H_L(I)}\right] \frac{dI}{d\log(w/r)}$	$\sigma > \eta$: machines also take over marginal labor tasks. If $\eta < 1$, σ can still be below one; it need not be constant.
Does cheaper capital raise its income share? w and technology fixed; assignment adjusts.	$\frac{d\log(\alpha_K/\alpha_L)}{d\log(w/r)} = \sigma - 1$	Yes if $\sigma > 1$; it falls if $\sigma < 1$. At $\sigma = 1$, the local quantity and price effects offset.
Are cheaper and more productive machines equivalent? Compare $r \mapsto r/m$ with $a_K \mapsto ma_K$, $m > 1$; w, Y fixed.	$\frac{r/m}{a_K(i)} = \frac{r}{ma_K(i)}$ Elasticity $\frac{d\log(K/L)}{d\log m}$: σ for cheaper machines; $\sigma - 1$ for better machines.	Same task costs, assignment, task quantities, and labor use. Better technology needs only $1/m$ as much physical capital as the cheaper-rental experiment.

5.8. The Ricardian roots of task assignment

> *Why should a highly skilled worker leave any task to someone else if that worker can perform every task more quickly?*

This is the same question that motivates David Ricardo's account of international trade: why should a country import anything if it can produce every good with less labor than its trading partner? The answer is that production entails an opportunity cost. Ricardo showed that the principle of *comparative advantage* takes into account the costs and benefits of the different alternatives firms and countries face when organizing production.

Absolute advantage and comparative advantage. A producer has an *absolute advantage* in a good when it needs fewer inputs to produce a unit of that good. It has a *comparative advantage* when its opportunity cost of producing the good is lower. In a two-good, one-factor Ricardian economy, let z_{cg} be the units of labor required to produce one unit of good g in country c. The opportunity cost of one unit of good 1 is b_{c1}/b_{c2} units of good 2: the labor used for good 1 could instead have produced that amount of good 2. Thus country A has a comparative advantage in good 1 relative to country B if

$$\frac{b_{A1}}{b_{A2}} < \frac{z_{B1}}{z_{B2}}. \tag{5.33}$$

Country B then has a comparative advantage in good 2. This can hold even when $z_{A1} < z_{B1}$ and $z_{A2} < z_{B2}$, so that A has an absolute advantage in both.

Trade example: England's cloth and Portugal's wine. Ricardo (1821a, Ch. 7) illustrates the argument with the way England and Portugal traded back then. The key is understanding how relative productivities work in each country. We start from the productivity of labor in each activity (cloth or wine) and each country:

Labor required	One unit of cloth	One unit of wine
England	100	120
Portugal	90	80

Portugal has an absolute advantage in both goods. But one unit of cloth costs England $100/120 = 5/6$ units of wine and costs Portugal $90/80 = 9/8$ units of wine. England therefore has a comparative advantage in cloth. Conversely, one unit of wine costs Portugal $80/90 = 8/9$ units of cloth, compared with $120/100 = 6/5$ in England. Portugal has a comparative advantage in wine.

To see the gains from trade, suppose one unit of cloth trades for one unit of wine, a relative price of 1 strictly between 5/6 and 9/8 units of wine per cloth. England can obtain wine by producing cloth with 100 units of labor and exporting it, instead of producing wine directly with 120. Portugal can obtain cloth by producing wine with 80 units of labor and exporting it, instead of producing cloth directly with 90. Both save labor. This price illustrates mutually beneficial exchange.

Ricardo makes the apparently surprising implication explicit in Ricardo (1821a, Ch. 7, Portugal example); original passage:

> "It would therefore be advantageous for her [Portugal] to export wine in exchange for cloth. This exchange might even take place, notwithstanding that the commodity imported by Portugal could be produced there with less labour than in England. Though she could make the cloth with the labour of 90 men, she would import it from a country where it required the labour of 100 men to produce it, because it would be advantageous to her rather to employ her capital in the production of wine, for which she would obtain more cloth from England, than she could produce by diverting a

> portion of her capital from the cultivation of vines to the manufacture of cloth."

From countries and goods to skills and tasks. Now replace countries with skill groups and goods with tasks. Consider two groups, H and L, and two tasks: preparing an analysis and processing a record. Suppose their hourly productivities are

Task units per hour	Analysis	Record processing
Higher-skill group H	4	8
Lower-skill group L	1	4

Group H is absolutely more productive at both tasks. But one "unit of analysis" costs H two processed records and costs L four. Group H has a comparative advantage in analysis, while L has a comparative advantage in record processing.

The comparative advantage argument becomes a firm's assignment decision through wages. If $w_H = \$60$ and $w_L = \$20$ per hour, the cost per analysis is $60/4 = \$15$ using H and $20/1 = \$20$ using L. The cost per processed record is $60/8 = \$7.50$ using H and $20/4 = \$5$ using L. Firms assign analysis to H and record processing to L.

More generally, the specialization just described requires

$$\frac{a_H(\text{records})}{a_L(\text{records})} < \frac{w_H}{w_L} < \frac{a_H(\text{analysis})}{a_L(\text{analysis})}, \qquad \text{here} \qquad 2 < \frac{w_H}{w_L} < 4. \tag{5.34}$$

For any two groups, our general assignment rule can be written as

$$\frac{w_H}{a_H(i)} < \frac{w_L}{a_L(i)} \quad \longleftrightarrow \quad \frac{a_H(i)}{a_L(i)} > \frac{w_H}{w_L}. \tag{5.35}$$

Relative productivity gives comparative advantage and orders tasks; the relative wage determines the cutoff. Absolute productivity helps determine income levels.

Canonical models based on comparative advantage.

- **Ricardian trade.** Ricardo's two-country example is formalized and extended to a continuum of goods by Dornbusch, Fischer, and Samuelson (1977). Goods are ordered by relative labor requirements, and relative wages determine the specialization boundary. Eaton and Kortum (2002) extend Ricardian trade to many countries with stochastic productivities and geographic trade costs, providing a quantitative account of bilateral trade and gains from trade.

- **Occupational choice and assignment.** In Roy (1951), people choose occupations according to the earnings their different abilities command; selection depends on relative earning opportunities. Assignment models study how matching heterogeneous workers to heterogeneous jobs shapes output and wages; Sattinger (1993) surveys this tradition. Comparative advantage can therefore explain occupational sorting and observed earnings differences.

- **Offshoring and task trade.** Grossman and Rossi-Hansberg (2008) move the international allocation decision inside production: different tasks can be performed in different countries as offshoring costs change. Their framework combines task allocation with factor-endowment forces.

- **Skills, automation, and new tasks.** Acemoglu and Autor (2011) and Acemoglu, Kong, and Restrepo (2025) combine earlier task models with the continuum-of-goods Ricardian framework to determine assignment across skill groups. Acemoglu and Restrepo (2018, 2019) extend task allocation to capital and to the creation of new tasks in which labor has a comparative advantage. These models distinguish displacement from productivity and reinstatement effects.

5.9. David Ricardo and machinery: Capital substituting labor is a longstanding issue

Technological progress and capital accumulation were major concerns during Ricardo's lifetime during the early 19th century, as they are today two centuries later. Industrial machinery and rapid capital accumulation on the hands of a few raised urgent questions about changes the distribution of income among workers, capitalists, and landowners. Ricardo's essay *On Machinery*—added as Chapter 31 of the *Principles*' third edition in 1821—shows us how these topics were discussed back then:

> "In the present chapter I shall enter into some enquiry respecting the influence of machinery on the interests of the different classes of society, a subject of great importance, and one which appears never to have been investigated in a manner to lead to any certain or satisfactory results."

Before this essay, Ricardo had argued that all technological progress would be beneficial to workers. This is the same conclusion we got from our analysis of the aggregate production function. Even if the share of income going to workers could decrease, capital accumulation and better technology would increase real wages and with them the standards of living of workers. He changed his mind on this:

> "My mistake arose from the supposition, that whenever the net income of a society increased, its gross income would also increase; I now, however, see reason to be satisfied that the one fund, from which landlords and capitalists derive their revenue, may increase, while the other, that upon which the labouring class mainly depend, may diminish, and therefore it follows, if I am right, that the same cause which may increase the net revenue of

> the country, may at the same time render the population redundant, and deteriorate the condition of the labourer."

He gives a numerical example of this (in classical Ricardian fashion). He is thinking of a change in technology that diverts part of the labor previously producing food and necessaries towards building machines (capital). The capitalist retains a capital stock worth £20,000 (about 5 Million CAD in today's money). The capitalists invest £7,000 of this capital and that the remaining £13,000 is employed as circulating capital in the support of labour. The value of the goods produced is £15,000. Then, the capitalist decides to construct more machinery (say, data centers). For this they employ half of their labor; the other half produces just as before. Ricardo's argument goes as follows

> "The machine would be worth £7,500, and the food and necessaries £7,500, and, therefore, the capital of the capitalist would be as great as before; for he would have besides these two values, his fixed capital worth £7,000, making in the whole £20,000 capital, and £2,000 profit [the same profit as before]. After deducting this latter sum for his own expenses, he would have a no greater circulating capital than £5,500 with which to carry on his subsequent operations; and, therefore, his means of employing labour, would be reduced in the proportion of £13,000 to £5,500, and, consequently, all the labour which was before employed by £7,500, would become redundant."

Net income (for the capitalist) can be maintained even though the resources employing workers shrink. Ricardo consequently takes workers' concerns seriously:

> "That the opinion entertained by the labouring class, that the employment of machinery is frequently detrimental to their interests, is not founded on prejudice and error, but is conformable to the correct principles of political economy."

But he also stresses the timing and financing of investment:

> "The statements which I have made will not, I hope, lead to the inference that machinery should not be encouraged. To elucidate the principle, I have been supposing, that improved machinery is suddenly discovered, and extensively used; but the truth is, that these discoveries are gradual, and rather operate in determining the employment of the capital which is saved and accumulated, than in diverting capital from its actual employment. [...] But with every increase of capital he would employ more labourers; and, therefore, a portion of the people thrown out of work in the first instance, would be subsequently employed"

Hollander (2019) provides a modern historical perspective on Ricardo's messages.

Our task model makes a related distinction using different machinery. A fall in r from capital accumulation, or a rise in $a_K(i)$ from technological progress, can lower the cost of capital relative to labor in (5.8). If capital is technically capable of the task, assignment can shift toward machines. Lower production costs can expand output and demand for the labor tasks that remain. The balance determines wages and employment.

David Ricardo (1772–1823)

Thomas Phillips, c. 1821. Portrait source (public domain).

Ricardo was a British political economist, successful stockbroker, and Member of Parliament. His *Principles* (1817; third edition 1821) became a foundation of classical economics. He was a close friend of Thomas Robert Malthus and did much to develop economic thinking during the first industrial revolution.

Ricardo introduced many concepts still used in economics. The most famous is, no doubt, comparative advantage, but he also described what we now call *Ricardian equivalence* when thinking about the funding of government spending during the Napoleonic wars. Under strong conditions, replacing current taxes with debt leaves consumption unchanged because households save for the future taxes needed to service the debt; the modern formal benchmark is associated with Barro (1974).

Sources: Ricardo (1821a,b); *Principles*, Ch. 17 (taxes and borrowing); *Letters to Malthus*, ed. Bonar (1887); National Portrait Gallery profile.

6. Examples and Applications of the Task-Based Model

6.1. A special case: CES from the assignment of labor to tasks

This special case shows that the substitutability of tasks and inputs need not coincide. To make the things simple we start with a Cobb–Douglas aggregator for tasks, so $\eta = 1$ as in (5.2), and then we show that the effective elasticity of substitution between inputs is higher than one. The higher elasticity reflects the fact that the assignment also responds to changes in relative prices.

This model was developed in Acemoglu and Zilibotti (2001) in the context of wage inequality between high and low-skilled workers. Accordingly, we consider two types of labor, with supplies L and H. Task production is

$$y(i) \;=\; A_L(1-i)^{\frac{1}{\kappa}} n_L(i) \;+\; A_H i^{\frac{1}{\kappa}} n_H(i), \qquad \kappa > 0. \tag{6.1}$$

where we are giving a specific functional form to the task-worker productivity, so that

$$a_L(i) \;\equiv\; A_L(1-i)^{\frac{1}{\kappa}} \qquad\qquad a_H(i) \;\equiv\; A_H(i)^{\frac{1}{\kappa}} \tag{6.2}$$

Relative productivity of H is $\frac{A_H}{A_L}\left(\frac{i}{1-i}\right)^{\frac{1}{\kappa}}$, which increases with i. So, worker productivity is such that the low-skill workers have comparative advantage at low-index tasks and high-skill workers at high-index tasks

Group L therefore produces tasks below a cutoff I, and group H produces those above it. At the boundary, the two unit costs are equal:

$$\frac{w_L}{A_L(1-I)^{\frac{1}{\kappa}}} \;=\; \frac{w_H}{A_H I^{\frac{1}{\kappa}}} \quad\longrightarrow\quad \frac{I}{1-I} \;=\; \left(\frac{w_H/A_H}{w_L/A_L}\right)^{\kappa}. \tag{6.3}$$

Solving for the lengths of the two task intervals gives

$$I = \frac{(w_H/A_H)^\kappa}{(w_H/A_H)^\kappa + (w_L/A_L)^\kappa}, \qquad 1-I = \frac{(w_L/A_L)^\kappa}{(w_H/A_H)^\kappa + (w_L/A_L)^\kappa}. \tag{6.4}$$

With Cobb–Douglas task demand, every equal-length interval receives the same expenditure. So, integrating over the assigned intervals gives

$$w_L L = I\,PY, \qquad w_H H = (1-I)PY. \tag{6.5}$$

Effective elasticity. Divide the wage bills and then substitute the cutoff condition:

$$\frac{H}{L} = \frac{w_L}{w_H}\frac{1-I}{I} = \frac{w_L}{w_H}\left(\frac{w_L/A_L}{w_H/A_H}\right)^\kappa = \left(\frac{A_H}{A_L}\right)^\kappa \left(\frac{w_H}{w_L}\right)^{-(1+\kappa)}. \tag{6.6}$$

Taking logs, holding technology fixed, and using the same relative-price definition as in our earlier two-input models gives

$$\sigma = -\frac{d\log(H/L)}{d\log(w_H/w_L)} = \underbrace{1}_{\text{across tasks}} + \underbrace{\kappa}_{\text{task reassignment}}. \tag{6.7}$$

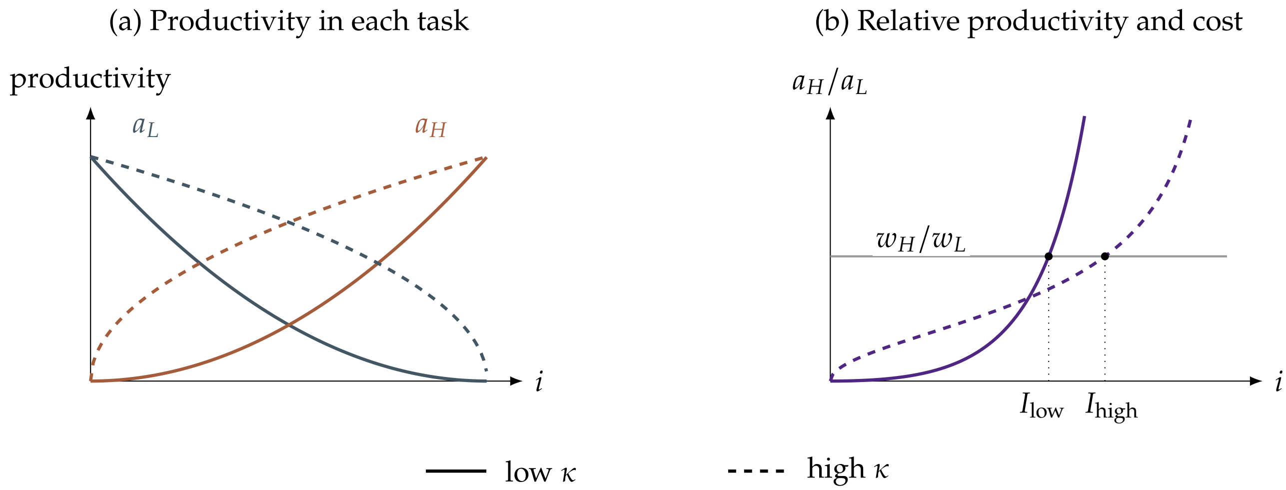
If assignment were fixed, a higher w_H/w_L would reduce demand for H-produced tasks relative to L-produced tasks with unit elasticity. With endogenous assignment, it also raises I: some tasks move from H to L. The aggregate response is larger.

The parameter κ makes comparative advantage change more slowly across tasks. This makes them more substitutable in the margin so a given change in relative wages then moves the boundary farther. Indeed,

$$\frac{d}{di}\log\frac{a_H(i)}{a_L(i)} = \frac{1}{\kappa i(1-i)}. \tag{6.8}$$

Figure 19 compares low κ (solid lines) with high κ (dashed lines). Higher κ raises both

Figure 19. Changes in Productivity and Elasticity κ



Notes: Productivity schedules in (6.2) for low and high κ. Solid curves show low κ and dashed curves high κ in both panels. Higher κ makes log relative productivity change more slowly across tasks.

groups' productivities (panel a); nevertheless, relative productivity does not change monotonically and varies less sharply around the illustrated cutoffs (panel b), making assignment more responsive to relative wages. At the fixed wage ratio shown, where $w_H > w_L$, increasing κ moves the cutoff to the right and expands the low-skill task bundle.

Ultimately, we can recover the output as a CES with elasticity $\sigma = 1 + \kappa$:

$$Y \;=\; e^{-\frac{1}{\kappa}} \left((A_L L)^{\frac{\kappa}{1+\kappa}} + (A_H H)^{\frac{\kappa}{1+\kappa}} \right)^{\frac{1+\kappa}{\kappa}}. \tag{6.9}$$

However, the steps are somewhat cumbersome and do not deliver any additional insight. The derivation involves replacing the labor quantities for each task into the continuous Cobb-Douglas aggregator. The labor quantities satisfy

$$n_L(i) \;=\; \frac{PY}{w_L} \;=\; \frac{L}{I} \quad (i < I), \qquad n_H(i) \;=\; \frac{PY}{w_H} \;=\; \frac{H}{1-I} \quad (i > I). \tag{6.10}$$

which, replaced into the aggregator give

$$\log Y \;=\; \int_0^I \log\left[A_L(1-i)^{\frac{1}{\kappa}}\frac{L}{I}\right] di \;+\; \int_I^1 \log\left[A_H i^{\frac{1}{\kappa}}\frac{H}{1-I}\right] di. \tag{6.11}$$

Solving these integrals delivers the CES representation.

6.2. A special case: Cobb–Douglas from capital–labor assignment

We can also get a Cobb-Douglas out of an environment where tasks are hard to substitute ($\eta < 1$). This is another case in which the task assignment makes inputs more substitutable. For this we follow Hubmer and Restrepo (2026) who are studying automation and changes in the labor share. Therefore, we go back to our notation of capital and labor with task production $y(i) = a_K(i)k(i) + a_L(i)\ell(i)$.

The key to get the substitution right is to choose task productivity appropriately. Hubmer and Restrepo show this is done by setting

$$a_K(i) \;=\; i^{\frac{\gamma_K-1}{\gamma_K(1-\eta)}}(1-i)^{\frac{\gamma_K+1}{\gamma_K(1-\eta)}}, \qquad a_L(i) \;=\; i^{\frac{\gamma_L+1}{\gamma_L(1-\eta)}}(1-i)^{\frac{\gamma_L-1}{\gamma_L(1-\eta)}}, \qquad \gamma_K, \gamma_L > 0. \tag{6.12}$$

The expressions look complicated, but they are chosen so that the task-cost integrals have simple solutions. Dividing the productivity schedules and collecting exponents gives

$$\frac{a_L(i)}{a_K(i)} \;=\; \left(\frac{i}{1-i}\right)^{\frac{\gamma_K+\gamma_L}{(1-\eta)\gamma_K\gamma_L}}. \tag{6.13}$$

So, capital has a comparative advantage in low-index tasks and we are once more looking for a threshold index I, below which capital performs tasks and above which labor performs tasks. This threshold satisfies

$$\frac{w}{r} \;=\; \frac{a_L(I)}{a_K(I)} \;=\; \left(\frac{I}{1-I}\right)^{\frac{\gamma_K+\gamma_L}{(1-\eta)\gamma_K\gamma_L}}. \tag{6.14}$$

Cheaper capital raises, or more expensive labor, raises I, substituting labor for capital through the reassignment of tasks.

The solution for I is more easily express as the ratio $t \equiv I/(1-I)$, so that

$$t = \left(\frac{w}{r}\right)^{\frac{(1-\eta)\gamma_K\gamma_L}{\gamma_K+\gamma_L}}. \tag{6.15}$$

Integrating expenditure over tasks. We know from our previous results that the total demand for capital and labor satisfy (5.15). Multiplying by the respective price of each input gives total input expenditure:

$$rK = YP^\eta r^{1-\eta}\int_0^I a_K(i)^{\eta-1}\,di, \qquad wL = YP^\eta w^{1-\eta}\int_I^1 a_L(i)^{\eta-1}\,di. \tag{6.16}$$

Here, is where the seemingly complicated functional form of the task productivities starts to pay off. Raising the schedules to $\eta-1 = -(1-\eta)$ cancels their denominators:

$$a_K(i)^{\eta-1} = i^{\frac{1}{\gamma_K}-1}(1-i)^{-1-\frac{1}{\gamma_K}}, \qquad a_L(i)^{\eta-1} = i^{-1-\frac{1}{\gamma_L}}(1-i)^{\frac{1}{\gamma_L}-1}. \tag{6.17}$$

We can now solve the integrals directly using the change of variable $u = i/(1-i)$. Then $i = u/(1+u)$, $1-i = 1/(1+u)$, and $di = (1+u)^{-2}\,du$. The factors in $1+u$ cancel in both integrands, and the boundary becomes $u = t$:

$$\int_0^I a_K(i)^{\eta-1}\,di = \int_0^t u^{\frac{1}{\gamma_K}-1}\,du = \gamma_K t^{\frac{1}{\gamma_K}}, \tag{6.18}$$

$$\int_I^1 a_L(i)^{\eta-1}\,di = \int_t^\infty u^{-1-\frac{1}{\gamma_L}}\,du = \gamma_L t^{-\frac{1}{\gamma_L}}. \tag{6.19}$$

Constant shares and unit elasticity. We now show that the results above imply the characteristic properties of a Cobb-Douglas. Substitute these integrals into factor payments and take their ratio:

$$\frac{rK}{wL} = \frac{\gamma_K}{\gamma_L}\left(\frac{r}{w}\right)^{1-\eta} t^{\frac{1}{\gamma_K}+\frac{1}{\gamma_L}}. \tag{6.20}$$

The cutoff condition (6.15) implies $t^{\frac{1}{\gamma_K}+\frac{1}{\gamma_L}} = (w/r)^{1-\eta}$. The price terms cancel, so

$$\frac{rK}{wL} = \frac{\gamma_K}{\gamma_L}, \qquad \alpha_K = \frac{\gamma_K}{\gamma_K+\gamma_L}, \qquad \alpha_L = \frac{\gamma_L}{\gamma_K+\gamma_L}. \tag{6.21}$$

These are constant factor shares, exactly as in our earlier Cobb–Douglas model. Rearranging the payment ratio and taking logs gives

$$\log\frac{K}{L} = \log\frac{\gamma_K}{\gamma_L} + \log\frac{w}{r} \quad \longrightarrow \quad \sigma = \frac{d\log(K/L)}{d\log(w/r)} = 1. \tag{6.22}$$

Taking a step back helps us see what is going on. There are actually two substitution responses taking place when we change relative prices. We can write the capital to labor ratio as

$$\frac{K}{L} = \frac{\gamma_K}{\gamma_L}\left(\frac{w}{r}\right)^{\eta} t^{\frac{1}{\gamma_K}+\frac{1}{\gamma_L}}. \tag{6.23}$$

If I (and hence t) were fixed, the elasticity would be η. Allowing assignment to adjust gives

$$\sigma = \underbrace{\eta}_{\text{across tasks}} + \underbrace{\left(\frac{1}{\gamma_K}+\frac{1}{\gamma_L}\right)\frac{d\log t}{d\log(w/r)}}_{\text{task reassignment}} = \eta + (1-\eta) = 1. \tag{6.24}$$

Tasks are hard to substitute for each other, but the firm can replace the *producer* of a task. Here that additional response is exactly large enough to give unit aggregate elasticity.

6.3. Three inputs: skill assignment, routinization, and polarization

The task model has also been used to study how technology can lead to wage inequality and changes in job prospects for different types of workers. To look at this, we extend the model to three inputs: low-, medium-, and high-skill labor. We will use L, M, H as *skill* *labels*. The task aggregator is the same as in our general case, equation (5.1):

$$Y = \left(\int_0^1 y(i)^{\frac{\eta-1}{\eta}} \, di \right)^{\frac{\eta}{\eta-1}} \tag{6.25}$$

Only task production and the number of assignment boundaries change:

$$y(i) = \sum_{s \in \{L,M,H\}} a_s(i) n_s(i). \tag{6.26}$$

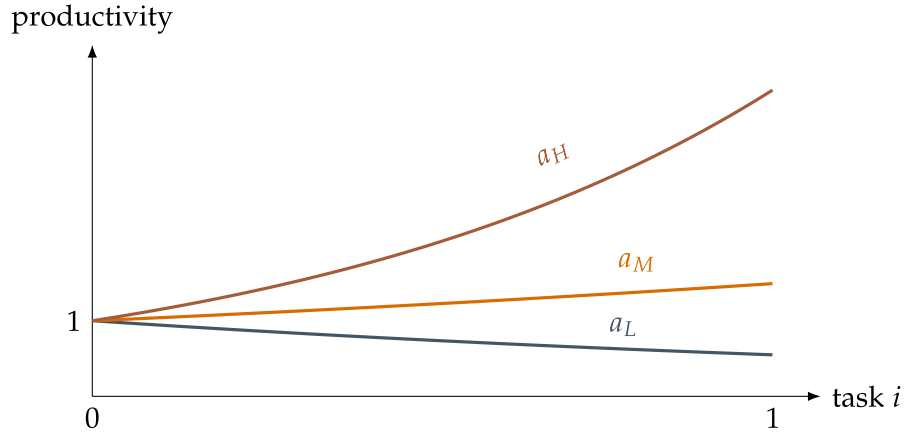
The same optimality conditions give $P(Y/y(i))^{1/\eta} a_s(i) \leq w_s$, with equality when $n_s(i) > 0$. Each task is performed by the type of worker with the highest productivity per dollar, or equivalently the lowest unit cost.

Two relative-productivity curves, two cutoffs. The principle of comparative advantage still applies, but we now have to make comparisons between three inputs. We make the problem easier by ordering groups so that low-skill workers have comparative advantage in low-indexed tasks, medium skill workers in tasks in the middle, and high-skill workers in high-indexed tasks. We get to this ordering by ensuring that when we compare medium and low skilled workers, the comparative advantage of medium skill workers is increasing (just as with labor and capital before). Hence, there is a cutoff task above below which we want to use low-skill labor. Similarly, when hen we compare high and medium skilled workers, the comparative advantage of high skill workers is increasing. This gives rise to a second cutoff, above which we want to use high-skill labor. Medium-skill labor is in the middle of the two cutoffs.

Formally, the task productivity schedules are positive and continuous, with increasing relative schedules as follows

$$\frac{d}{di} \log \left(\frac{a_M(i)}{a_L(i)} \right) > 0, \qquad \frac{d}{di} \log \left(\frac{a_H(i)}{a_M(i)} \right) > 0. \tag{6.27}$$

Figure 20. Task Productivity by Labor Type



Notes: Three skill productivity schedules. High skill has absolute advantage at every $i > 0$. Allocation nevertheless depends on productivity per dollar, so all three groups can perform positive task bundles.

It is the ordering of *relative* productivities that matters. Figure 20 illustrates this distinction: high-skill workers are more productive at every positive-index task, but their relative advantage is strongest at the right end.

At the boundaries, adjacent groups have equal unit costs:

$$\frac{w_L}{a_L(I_L)} = \frac{w_M}{a_M(I_L)} \qquad \longleftrightarrow \qquad \frac{w_M}{w_L} = \frac{a_M(I_L)}{a_L(I_L)}, \tag{6.28}$$

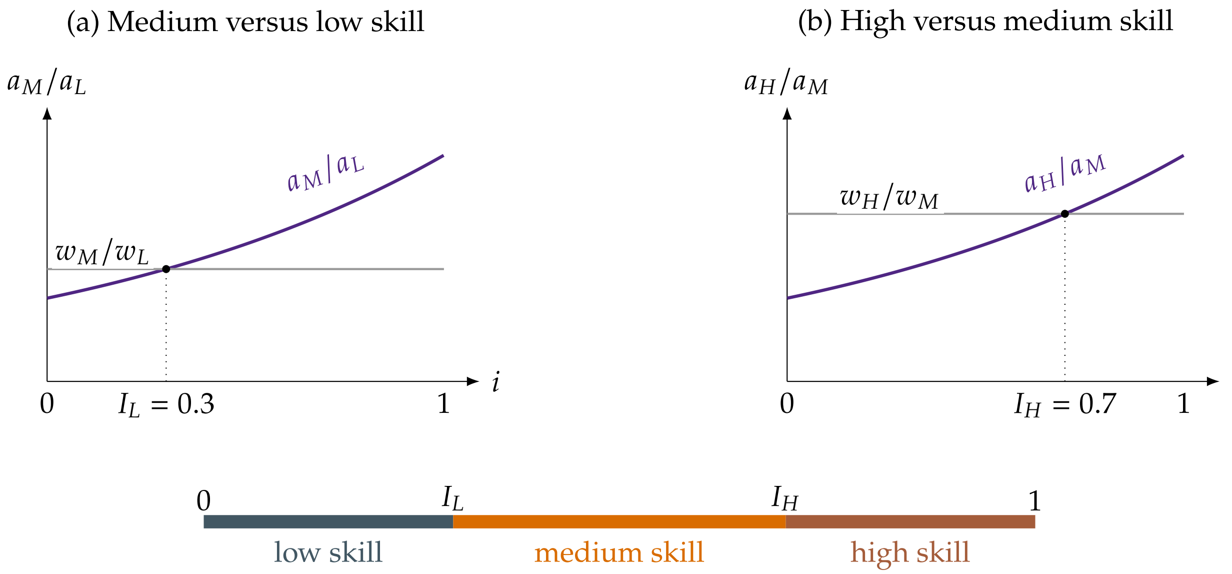
$$\frac{w_M}{a_M(I_H)} = \frac{w_H}{a_H(I_H)} \qquad \longleftrightarrow \qquad \frac{w_H}{w_M} = \frac{a_H(I_H)}{a_M(I_H)}. \tag{6.29}$$

These crossings satisfy $0 < I_L < I_H < 1$ as in Figure 21. Below I_L low skill is cheaper than medium skill, which is cheaper than high skill because $i < I_H$. Between the cutoffs, medium skill beats both neighbors. Above I_H, high skill beats medium skill, which beats low skill.

Equlibrium: Wages and task bundles adjust together. With positive fixed supplies, wages must clear all three labor markets:

$$N_s = \int_{\mathcal{I}_s} \frac{y(i)}{a_s(i)}\, di, \qquad \frac{w_s N_s}{PY} = \left(\frac{w_s}{P}\right)^{1-\eta} \int_{\mathcal{I}_s} a_s(i)^{\eta-1}\, di, \quad s \in \{L, M, H\}. \tag{6.30}$$

Figure 21. Optimal Task Assignment: 3 Bundles - 2 Cutoffs



Notes: Two relative-productivity comparisons generate three task bundles. Ordered cutoffs ensure that the middle group is cheapest between I_L and I_H.

The assignment sets are $\mathcal{I}_L = [0, I_L]$, $\mathcal{I}_M = [I_L, I_H]$, and $\mathcal{I}_H = [I_H, 1]$. The two cutoff conditions and market clearing jointly determine wages and bundles. For Cobb–Douglas task demand, the wage-bill shares simplify to I_L, $I_H - I_L$, and $1 - I_H$.

Routinization and polarization

The three bundles allow us to distinguish nonroutine manual and service activities, routine production and clerical activities, and nonroutine abstract activities. This interpretation is extremely useful to think about how different *occupations* have been affected by technological change in the 2000s.

We can think of non-routine manual and service tasks as being most similar to the tasks in low-paying occupations in retail and services. The fact that these tasks are non-routine makes them hard to automate, think of tasks that require interacting with customers of performing personal services. In the middle we have the traditionally well-paid occupations in manufacturing and some clerical activities, like accounting. These occupations have traditionally commanded a higher wage than the non-routine occupations in the first group, but the fact that their activities are easily codifiable which makes them easy to automate. At the top we have occupations heavy on non-routine abstract tasks. These have traditionally been hard to automate (but likely not for much longer).

The routinization hypothesis, chiefly associated with Autor, Levy, and Murnane (2003), emphasizes that computers can replace codifiable routine tasks while being less effective at many nonroutine manual tasks and complementing abstract work. To represent automation of routine tasks, restore capital as an additional option *alongside* the three labor groups:

$$y(i) = \sum_{s \in \{L,M,H\}} a_s(i) n_s(i) + a_K(i) k(i), \tag{6.31}$$

The set of automated tasks is the set of tasks assigned to capital:

$$\mathcal{A} = \left\{ i : \frac{r}{a_K(i)} < \min_{s \in \{L,M,H\}} \frac{w_s}{a_s(i)} \right\}. \tag{6.32}$$

If machines are particularly cost effective at routine activities, $\mathcal{A}$ can lie inside the middle bundle. If machines get better or cheaper they replace more of these tasks and lower the middle group's expenditure share and labor demand. The demand for the remaining tasks and equilibrium wages also adjust.

This mechanism helps explain *job polarization*: declining employment shares in routine middle-wage occupations alongside rising shares at the low- and high-wage ends, documented by Autor et al. (2006, 2008); Autor and Dorn (2013) develop the connection to low-skill service employment.

7. Automation and New Tasks

We now use the task model to study how technology changes the demand for labor. In Section 5, we determined which tasks are performed by workers and which are performed by machines. The next step is to ask what happens when machines can take over some of the activities previously assigned to workers. This is more easily pictured in the case of manufacturing plants: a robot can move components or operate equipment, while workers continue to inspect finished pieces, diagnose faults, and prepare production plans.

As workers take over tasks, production becomes cheapefacr, but the workers who operated the equipment have lost part of what they used to do. This generates two forces on labor demand. Lower costs allow production to expand, increasing demand for the tasks that remain with workers. At the same time, automation is directly reducing the demand for labor. Our objective is to separate these forces and establish what they imply for employment, the labor share, and wages.

We follow Restrepo (2024, Sections 2.3–2.6), then develop a simple version of Acemoglu and Restrepo (2018)'s race between automation taking over tasks and the creation of new tasks that humans can perform. The treatment of different types of technological change in Acemoglu, Kong, and Restrepo (2025, Section 3) provides a useful companion.

7.1. What changes when a task is automated?

As in previous sections, capital and labor cost r and w per unit, and their productivities in task i are $a_K(i)$ and $a_L(i)$. The firm uses $k(i)$ units of capital and $\ell(i)$ units of labor to produce task i

$$y(i) \;=\; a_K(i)k(i) + a_L(i)\ell(i). \tag{7.1}$$

Tasks are combined with elasticity η as in (5.1).

The key question is how tasks are assigned between capital and labor. We saw that this assignment follows the principle of *comparative advantage*, so that we assign tasks to the input with the lowest effective unit cost. The unit cost incorporates both the direct cost of paying for the capital or labor being used and the productivity that each input has in performing the task. Both margin matter because the assignment changes in a similar way if the input gets cheaper (say, a lower r) or if the input gets more productivity (a higher $a_k(i)$). In both cases we need to pay less to produce the same.

Accordingly, we give the two unit costs short names:

$$c_K^t(i) \equiv \frac{r}{a_K^t(i)}, \qquad c_L(i) \equiv \frac{w}{a_L(i)}, \qquad t \in \{0, 1\}. \tag{7.2}$$

We have added a superscript t to the unit cost of performing tasks with capital to distinguish capital's technology before and after the innovation. If capital cannot perform a task, we set its unit cost to infinity; equivalently, its productivity in that task is zero.

The assignment rule is to use the input with the lower unit cost. The tasks assigned to capital are those being *automated*; we refer to them as $\mathcal{A}$:

$$\mathcal{A}_t \equiv \{i : c_K^t(i) < c_L(i)\}, \tag{7.3}$$

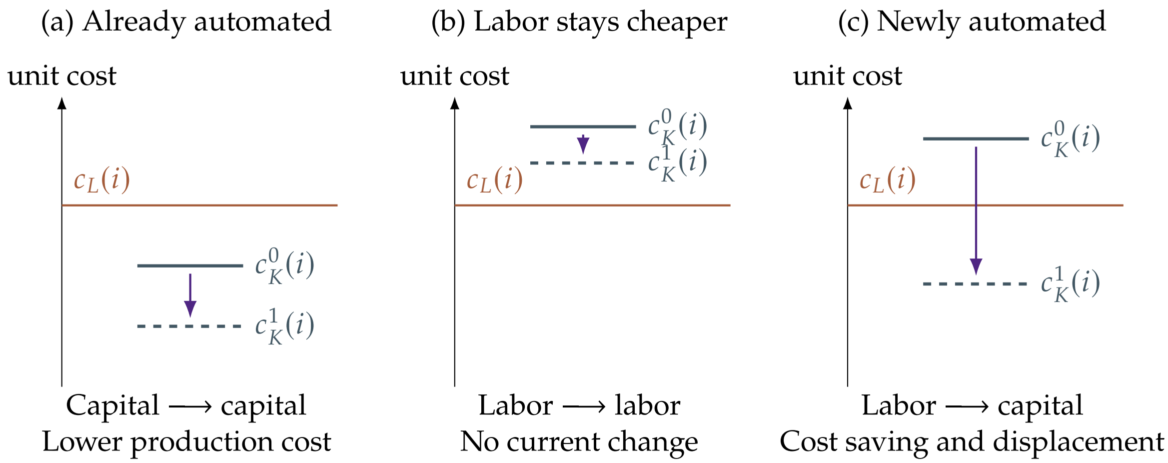
$$\Delta\mathcal{A} \equiv \mathcal{A}_1 \setminus \mathcal{A}_0 = \{i : c_K^1(i) < c_L(i) \leq c_K^0(i)\}. \tag{7.4}$$

When technology changes (going from $t = 0$ to $t = 1$) the set of automated expands tasks in $\Delta\mathcal{A}$ were performed by labor and are now performed by machines. This is the *extensive margin* of automation. The arrival of the capability and its adoption are different: a machine that can perform a task will only be used if doing so makes sense for the firm; only if the unit cost falls below that of labor.

Three possibilities for a better machine. Suppose capital becomes more productive in a task, so $c_K^1(i) < c_K^0(i)$. The effect depends on where these costs lie relative to $c_L(i)$ as we show in Figure 22 (Steinsson 2026, Section 7.3):

(a) **The task was already automated.** If $c_K^1(i) < c_K^0(i) < c_L(i)$, the machine continues to

Figure 22. A Better Machine Can Change Costs, Assignment, or Both



Notes: Each panel compares the cost of producing one task before and after an improvement in capital productivity, holding w, r, and $a_L(i)$ fixed. The red line is labor's unit cost; the solid and dashed blue lines are capital's old and new unit costs. Only panel (c) transfers a task from labor to capital.

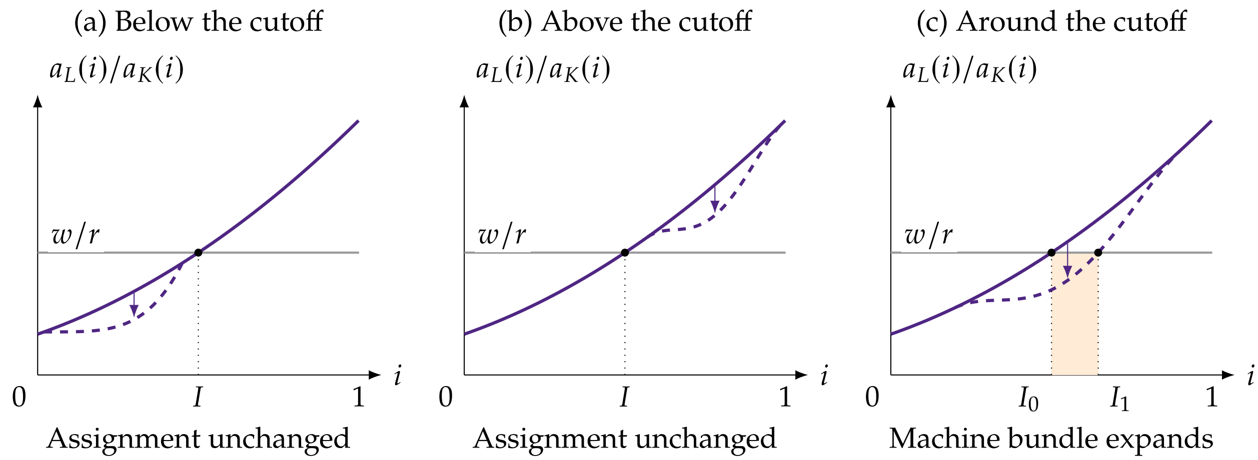
perform it at a lower cost. This improves productivity at the *intensive margin*, without directly removing a task from labor, so there is no *extensive margin* effect. A faster conveyor carrying the same components is an example.

(b) **Labor remains cheaper.** If $c_L(i) < c_K^1(i) < c_K^0(i)$, the firm continues to use labor. The improvement does not yet change the assignment of tasks and so there are no effects. Think of this as the development of LLM models before the release of chatGPT. They were better than previous technology, but were not good enough to perform any task.

(c) **The task switches from labor to capital.** If $c_K^1(i) < c_L(i) \leq c_K^0(i)$, the task enters $\Delta\mathcal{A}$. The innovation both reduces production costs and displaces labor from the task. Think of software taking over the processing of records previously assigned to clerical workers.

Moving the assignment cutoff. When $a_L(i)/a_K(i)$ is increasing, we can summarize assignment with the cutoff I from Section 5.2. In this case, relative labor productivity is higher for higher-indexed tasks and so only low-indexed tasks are automated. A technology that expands the capital bundle from $[0, I_0]$ to $[0, I_1]$ automates $\Delta\mathcal{A} = (I_0, I_1]$.

Figure 23. Local Productivity Improvements and Task Assignment



Notes: The solid curve is the original relative-productivity schedule; the dashed segment shows a local increase in capital productivity at unchanged $a_L(i)$ and w/r. The schedules coincide everywhere outside the affected interval, and both remain increasing. In panel (a), the improvement is confined to tasks already assigned to capital. In panel (b), improved capital remains more costly than labor, so the affected tasks stay with workers. In panel (c), the downward change crosses the wage–rental ratio and moves the cutoff to the right, from I_0 to I_1. The shaded interval contains the newly automated tasks, $\Delta\mathcal{A} = (I_0, I_1]$.

A rise in capital productivity lowers the relative productivity of labor and can end in automation at given w/r. But, as we just discussed, the technology may also improve tasks already assigned to capital, so a single innovation can combine extensive and intensive effects. Figure 23 expands the three cases of Figure 22 to the full relative-productivity schedule. In the first two panels the improvement leaves the cutoff unchanged; in the third, the local downward shift moves the cutoff to the right and expands the machine bundle.

7.2. Displacement and productivity gains

It is clear by now that an increase in the capabilities of capital can harm workers by taking away some of their tasks. However, that only happens if capital becomes a better alternative for production. This can ultimately be good for workers if it allows production to expand, or if workers are hard to substitute in their remaining tasks. We have a tradeoff in our hands. The two key questions that shape this tradeoff are:

How much of workers' existing work moves to machines?

How much cheaper does production become?

We find the answers by measuring the direct displacement of workers, D, and the reduction in unit cost, Γ, that follow automation.

How much of workers' existing work moves to machines? Our previous results show that automation reassigns tasks away from workers, but it does not directly tell us how many workers were displaced because some tasks employ more labor than others. We can measure the *direct displacement* of workers by automation as the fraction of initial employment devoted to tasks that become automated:

$$D \;\equiv\; \frac{\text{initial employment in newly automated tasks}}{\text{total initial employment}}. \tag{7.5}$$

This is a measure of *displacement*, it captures the work transferred to machines before output and wages adjust. It is also the fraction of the initial wage bill paid to workers in those tasks, since all workers receive the same wage.

We can be more precise and link displacement to worker productivity by using the firm's optimal labor demand. From (5.15) we get that labor demand as

$$L \;=\; Y\left(\frac{c}{w}\right)^{\eta} H_L, \quad \text{with} \quad H_L \;\equiv\; \int_{\mathcal{I}_L} a_L(i)^{\eta-1}\, di. \tag{7.6}$$

Here c is the unit cost of the final good; under perfect competition its output price is $P = c$ and the bundle of tasks assigned to labor is $\mathcal{I}_L \subset [0, 1]$. Why does labor productivity enter with the power $\eta - 1$? Higher productivity reduces the workers needed per unit of a task, but it also makes that task cheaper, encouraging the firm to use more of it. The task-demand condition and $\ell(i) = y(i)/a_L(i)$ combine these two effects:

$$\ell(i) \;=\; \frac{Y}{a_L(i)}\left(\frac{c}{w/a_L(i)}\right)^{\eta} \;=\; Y\left(\frac{c}{w}\right)^{\eta} a_L(i)^{\eta-1}, \qquad i \in \mathcal{I}_L. \tag{7.7}$$

At given Y, c, and w, a one-percent increase in task productivity reduces labor per unit of task service by one percent and increases the quantity of that service demanded by approximately η percent. When $\eta > 1$, the quantity response dominates, so more productive labor tasks employ more workers; when $\eta < 1$, the labor saving dominates, so they employ fewer. When $\eta = 1$, the two effects exactly offset. These are comparisons across tasks within the initial allocation, where Y, c, and w are common to every task.

Thus $a_L(i)^{\eta-1}$ weights tasks by the employment they actually account for in the model. The common factor $Y(c/w)^{\eta}$ cancels when we divide employment in newly automated tasks by total initial employment, giving

$$D \;\equiv\; \frac{\displaystyle\int_{\Delta\mathcal{A}} a_L(i)^{\eta-1}\,di}{H_L^0}, \qquad H_L^1 = (1-D)H_L^0, \qquad 0 < D < 1. \tag{7.8}$$

The fall in H_L comes from removing tasks from labor's bundle, not from reducing workers' productivity in the tasks they keep. With Cobb–Douglas tasks, $\eta = 1$, equal-length task intervals employ equal amounts of labor. If automation moves the cutoff from I_0 to I_1, displacement is simply the length of the newly automated interval divided by the length of labor's original interval:

$$D \;=\; \frac{I_1 - I_0}{1 - I_0}. \tag{7.9}$$

How much does automation reduce production costs? Displacement tells us how much work machines take over, but not how much the new technology improves production. For that, we ask how much it costs to produce the same final output before and after automation. The reduction in costs reflects the gains in productivity as tasks are reassigned away from labor and toward capital.

We measure the cost/productivity gain of the firm as the log difference between the initial and final unit costs:

$$\Gamma \ \equiv \ \log \frac{c^0}{c^1} > 0\,, \quad \text{where} \quad c^t \ = \ \left(\int_0^1 \min\{c_L(i), c_K^t(i)\}^{1-\eta}\, di \right)^{\frac{1}{1-\eta}}. \tag{7.10}$$

For a small cost reduction, 100Γ is approximately the percentage fall in unit cost. This fall in costs is translated into a lower price for the firm's output and hence higher sales, increasing the overall scale of production as we have seen in previous sections. That expansion in the scale of production can increase demand for the tasks that remain with workers. We therefore need Γ alongside D to compare the scope for output expansion with the work lost to machines.

The calculation is especially simple with Cobb–Douglas tasks. The log unit cost of the final good is the integral of log task costs, so unchanged task costs cancel when we compare before and after automation:

$$\Gamma \ = \ \int_{\Delta\mathcal{A}} \log \frac{c_L(i)}{c_K^1(i)}\, di, \qquad \eta = 1. \tag{7.11}$$

Thus productivity gains depend both on how many tasks become automated and on how much cheaper machines make each task.

$$\Gamma \ = \ \overbrace{(I_1 - I_0)}^{\text{length of newly automated interval}} \times \underbrace{\text{Average log cost saving per task}}_{\text{log(old task cost/new task cost)}}. \tag{7.12}$$

We need both measures because the same displacement can come with very different productivity gains. A technology can take over a large fraction of workers' tasks while saving little on each one. Acemoglu and Restrepo (2019); Restrepo (2024) call these *so-so automation technologies*: machines are cheap enough to be adopted, but the resulting cost savings are small, resulting in small gains for the economy and potential losses for workers.

7.3. Labor demand and the labor share at given factor prices

We can now answer how much labor demand goes down with automation (holding w, r, and output fixed). From (7.6), the before-and-after ratio of labor per unit of output is

$$\frac{L^1/Y^1}{L^0/Y^0} = (1-D)\left(\frac{c^1}{c^0}\right)^{\eta} \quad \longrightarrow \quad \Delta\log(L/Y) = \log(1-D) - \eta\Gamma < 0. \tag{7.13}$$

There are two forces reducing conditional labor demand:

(a) Labor loses the tasks in $\Delta\mathcal{A}$, captured by displacement D.

(b) The firm changes the quantities of the remaining tasks as automated services become cheaper, captured by cost-savings Γ.

The relevant elasticity here is η, which measures substitution between tasks. With output fixed, cheaper automated services lead the firm to use less of the remaining labor services; a larger η makes this quantity response stronger. This is why the cost-saving term in conditional labor demand is $-\eta\Gamma$. For given D and Γ, a higher η therefore implies a larger decline in labor per unit of output. For a small shock, the expression becomes $\Delta\log(L/Y) \simeq -D - \eta\Gamma$.

Does the labor share also fall?. We can get a useful expression the labor share $\alpha_L = \frac{wL}{PY}$ using $P = c$ and (7.6),

$$\alpha_L = \left(\frac{w}{c}\right)^{1-\eta} H_L. \tag{7.14}$$

At unchanged factor prices, automation therefore gives

$$\Delta\log\alpha_L = \underbrace{\log(1-D)}_{\text{displacement}} + \overbrace{(1-\eta)\Gamma}^{\text{change in task cost shares}} \approx -D + (1-\eta)\Gamma\,. \tag{7.15}$$

Displacement reduces the labor share directly because there is less labor being demanded for production. However, the effect of cost-savings depends on the elasticity of substitution between tasks, η. A reduction in costs means the denominator of the labor share (the total expenditure of the firm) goes down. This increases the labor share. But a reduction in costs creates a further substitution away from labor. That is the $-\eta\Gamma$ term we just found for the change in labor. This gives rise to three (by now) familiar cases:

(a) $\boldsymbol{\eta < 1}$: task quantities respond weakly to relative costs. The remaining labor tasks are hard to substitute away from, so their wage bill falls proportionally less than total cost through the cost-saving channel. The positive term $(1-\eta)\Gamma$ partly offsets displacement, as in our earlier weak-links discussion.

(b) $\boldsymbol{\eta = 1}$: the quantity response exactly offsets the effect of lower total cost on the share. Only the loss of labor tasks changes the labor share: $\alpha_L^1/\alpha_L^0 = 1 - D$.

(c) $\boldsymbol{\eta > 1}$: task quantities respond strongly to relative costs. Substitution toward automated services reduces the remaining labor wage bill proportionally more than total cost, so the cost-saving channel reinforces displacement.

Compare this with better machines in existing capital tasks. Suppose instead that costs fall only within the original capital bundle and assignment stays fixed. Then $D = 0$, and the labor-share response is $(1-\eta)\Gamma$. It is positive when $\eta < 1$, zero when $\eta = 1$, and negative when $\eta > 1$. This reproduces the distinction between more automated tasks and better machines in the weak-links example in Section 3. A broad capital-productivity improvement or a fall in r can also move the assignment cutoff, as Section 5.2 showed. In that case we must account for both margins; capital accumulation does not, by definition, hold assignment fixed.

7.4. Output expansion and employment at adopting firms

Less labor per unit of output need not mean fewer workers in total. We return to the same demand curve used in Sections 2 and 4:

$$Y = BP^{-\varepsilon}, \qquad B > 0, \quad \varepsilon > 0. \tag{7.16}$$

In a competitive industry the price reflects the unit cost, $P = c$. So, the cost reduction Γ raises output by $\Delta \log Y = \varepsilon\Gamma$. Combining this with (7.13) gives

$$\Delta \log L \;=\; \underbrace{\varepsilon\Gamma}_{\text{scale}} \;-\; \overbrace{\eta\Gamma}^{\text{task substitution}} \;+\; \underbrace{\log(1-D)}_{\text{displacement}} \;\approx\; (\varepsilon-\eta)\Gamma - D. \tag{7.17}$$

The increase in the scale of production coming from additional demand for the firm's output introduces a force in favor of more labor. If demand is sufficiently elastic labor demand can increase. The exact condition is

$$(\varepsilon-\eta)\Gamma > -\log(1-D). \tag{7.18}$$

Here we see that $\varepsilon > \eta$ is necessary but no longer sufficient for employment to rise because labor is lost not just to substitution but also to direct displacement D.

7.5. The race between man and machine

We now turn to a central topic at the center of how technological progress and automation affect workers:

> *Automation removes tasks from labor, but technology can also create new activities in which workers have a comparative advantage.*

For example, a new production process might automate record processing while creating work in equipment diagnosis, design, or coordinating the new process. The issue is whether the creation of productive opportunities for workers keeps pace with automation. This section looks into this topic following the work of Acemoglu and Restrepo (2018).

Tasks and the two technology frontiers. Active tasks lie on an interval of length one and production is Cobb-Douglas across tasks:

$$i \in [N-1, N], \qquad \log Y = \int_{N-1}^{N} \log y(i)\, di. \tag{7.19}$$

When $N = 1$, this is the $[0, 1]$ interval from Section 5.

An increase in N introduces new tasks at the top and replaces an equal measure of old tasks at the bottom. We can think of the new tasks as improved versions of old activities. In this way, as new tasks emerge the measure of active tasks stays equal to one.

Technology also makes it possible for some tasks to be performed by capital. We call $\bar{I} \in (N-1, N)$ the *technological automation frontier*. Capital can perform tasks up to $\bar{I}$; tasks above it require labor. This variable $\bar{I}$ tells us what machines able to do and is different from I which refers to the *actual assignment cutoff*, what machines actually do.

Task production is

$$y(i) = \begin{cases} a_K(i)k(i) + a_L(i)\ell(i), & N-1 \le i \le \bar{I}, \\ a_L(i)\ell(i), & \bar{I} < i \le N. \end{cases} \tag{7.20}$$

To simplify the algebra, set $a_K(i) = 1$ and assume $a_L(i) > 0$ is increasing. So, relative productivity $q(i) = a_L(i)/a_K(i) = a_L(i)$ is increasing and capital has a *comparative advantage* at low-indexed tasks and labor has a comparative advantage at high-indexed tasks (that goes to infinity above $\bar{I}$).

At any given w/r, define the unconstrained cost cutoff $\widetilde{I}$ by

$$a_L(\widetilde{I}) = \frac{w}{r}. \tag{7.21}$$

There are two cases.

(a) If $\bar{I} < \widetilde{I}$, machines would be cheaper on additional tasks but the technology is not yet available. The frontier binds, $I = \bar{I}$, and advancing it can displace labor.

(b) If $\widetilde{I} < \bar{I}$, machines are technologically capable of more tasks than firms choose to automate. Advancing $\bar{I}$ alone does nothing locally: assignment remains at $I = \widetilde{I}$.

It is therefore essential to distinguish capability from actual adoption. The optimal assignment satisfies

$$I \;=\; \min\{\bar{I}, \widetilde{I}\}, \qquad \underbrace{[N-1, I]}_{\text{capital tasks}}, \qquad \underbrace{(I, N]}_{\text{labor tasks}}. \tag{7.22}$$

Recovering aggregate production. Define the share of capital tasks as

$$m \;\equiv\; I - (N-1), \qquad 1 - m = N - I, \qquad 0 < m < 1. \tag{7.23}$$

With Cobb–Douglas task aggregation, every equal-length task interval receives the same expenditure. For fixed supplies K and L, the cost-minimizing allocation consequently has

$$k(i) = \frac{K}{m} \quad (i < I), \qquad \ell(i) = \frac{L}{1-m} \quad (i > I). \tag{7.24}$$

Substitute these allocations into (7.19):

$$\log Y = m \log \frac{K}{m} + (1-m) \log \frac{L}{1-m} + \int_I^N \log a_L(i)\, di, \tag{7.25}$$

$$Y = \underbrace{\exp\left(\int_I^N \log a_L(i)\, di\right)}_{\mathcal{A}(I,N)} \left(\frac{K}{m}\right)^m \left(\frac{L}{1-m}\right)^{1-m}. \tag{7.26}$$

The function $\mathcal{A}(I, N)$ collects productivity on labor's assigned tasks.

Production looks like the Cobb–Douglas function in Section 2, but now technology can change its exponents by changing m and productivity reflects how good workers are at their tasks and how many tasks they perform. The factors $m^{-m}(1-m)^{-(1-m)}$ allow us to compare different assignments. They capture the changing quantities of each input per assigned task.

In any case, as in the Cobb–Douglas, the shares m and $1-m$ also link inputs to prices and determine the input shares. Normalize the final-good price to one from this point onward, so w and r are real factor prices. Competitive payments satisfy

$$r = m\frac{Y}{K}, \qquad w = (1-m)\frac{Y}{L}, \qquad \alpha_K = m, \qquad \alpha_L = 1-m = N-I. \tag{7.27}$$

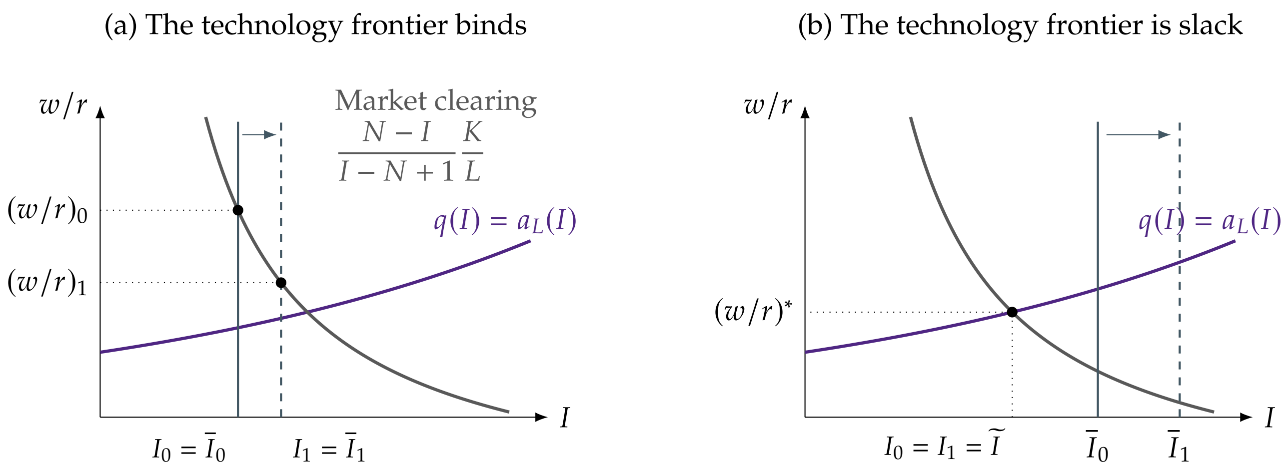
Dividing the factor payments gives the *market-clearing* relationship

$$\frac{w}{r} = \frac{N-I}{I-N+1}\frac{K}{L}. \tag{7.28}$$

This is the relative wage that clear the market at fixed K/L and N. Its right-hand side decreases with I because assigning more tasks to capital increases demand for its fixed supply, raising its rental rate relative to the wage. This curve and the constrained assignment condition (7.22) jointly determine w/r and I.

Figure 24 illustrates both cases. In panel (a), the technology frontier advances and the actual machine bundle expands, while w/r falls along the market-clearing curve. The fall in w/r limits the range of tasks that would be profitable to automate. In panel (b), the cost cutoff is already below the technology frontier, so extra technological capability does not change assignment or relative factor prices. This also connects to Section 5's elasticity

Figure 24. Technological Capability and Equilibrium Assignment



Notes: Factor supplies K, L and the new-task frontier N are fixed. The purple curve is the relative-productivity schedule; the downward-sloping curve imposes factor-market clearing. Vertical lines mark the original and expanded technological frontiers. In panel (a), equilibrium lies on the frontier above $q(I)$: capital would be cheaper on further tasks if the technology existed. In panel (b), the intersection of $q(I)$ and market clearing determines assignment before and after the technological advance.

decomposition: when the frontier binds, local price changes leave assignment fixed and $\sigma = \eta = 1$; when the frontier is slack, reassignment raises σ above one.

7.6. Automation, new tasks, and equilibrium wages

What happens when the automation frontier moves and more tasks are automated?

What happens when new tasks are created thanks to technological progress?

To answer this question we focus on the case in which the automation frontier is binding, $I = \bar{I} < \widetilde{I}$. That is, we would like to automate more tasks but the technology just is not there yet for us to do it. Then, we consider an expansion in the automation frontier ($\bar{I} \uparrow$). This expansion leads to more tasks being automated. We are holding capital (K) and labor (L) fixed and looking at what happens to output and wages.

Automation increases output. Hold K, L, N, and the labor-productivity schedule $a_L(i)$ fixed and increase I. Since $m = I - (N-1)$, we have $\partial m/\partial I = 1$ and $\partial(1-m)/\partial I = -1$. Taking logs of (7.26) turns the product into three terms:

$$\log Y = \int_I^N \log a_L(i)\, di + m \log \frac{K}{m} + (1-m) \log \frac{L}{1-m}.$$

We differentiate each term in turn. Increasing I removes tasks from the bottom of labor's interval, so differentiating the integral with respect to its lower limit gives

$$\frac{\partial}{\partial I} \int_I^N \log a_L(i)\, di = -\log a_L(I).$$

For the capital term, the product rule gives

$$\frac{\partial}{\partial I}\left[m \log \frac{K}{m}\right] = \log \frac{K}{m} + m\left(-\frac{1}{m}\right) = \log \frac{K}{m} - 1.$$

For the labor term, remember that $1-m$ falls as I rises:

$$\frac{\partial}{\partial I}\left[(1-m) \log \frac{L}{1-m}\right] = -\log \frac{L}{1-m} + (1-m)\frac{1}{1-m} = -\log \frac{L}{1-m} + 1.$$

Adding the three derivatives gives us

$$\left.\frac{\partial \log Y}{\partial I}\right|_{K,L,N} = -\log a_L(I) + \log \frac{K}{m} - \log \frac{L}{1-m} = \log \frac{K/m}{a_L(I)L/(1-m)}. \tag{7.29}$$

We can go further by connecting the change in output to the cost savings coming from automation. To do this, we use the factor-price formulas in (7.27):

$$r = \frac{Y}{K/m} \quad \& \quad w = \frac{Y}{L/(1-m)} \quad \longrightarrow \quad \frac{K/m}{a_L(I)L/(1-m)} = \frac{w/a_L(I)}{r}.$$

Because $a_K(i) = 1$, capital's unit cost at the marginal task is r, while labor's unit cost is

$w/a_L(I)$. We can therefore write the output response as the log ratio of the old labor cost to the new machine cost:

$$\left.\frac{\partial \log Y}{\partial I}\right|_{K,L,N} = \log \frac{w/a_L(I)}{r} \;\equiv\; g_I \;>\; 0. \tag{7.30}$$

We know machines would be cheaper if they were available, $r < w/a_L(I)$, because the technological frontier is binding, $I = \bar{I} < \widetilde{I}$,. Expanding the frontier makes this cheaper technique available, so the log cost saving and the output gain are positive. The larger the cost advantage of machines at the newly automated task, the larger the gain. Output rises even though total capital and labor are fixed, because the economy can produce the newly automated tasks with a cheaper technique and reallocate its inputs.

What happens to the wage?. Taking logs of $w = (N - I)Y/L$ gives

$$\left.\frac{\partial \log w}{\partial I}\right|_{K,L,N} = \underbrace{g_I}_{\text{productivity}} - \underbrace{\frac{1}{N - I}}_{\text{displacement}} . \tag{7.31}$$

The two terms have opposite signs. Higher output increases demand for the services that workers still perform. However, the same workers must now be employed in fewer tasks. At each remaining task, the additional labor runs into diminishing returns in final-good production. This puts downward pressure on the marginal product of labor.

The same decomposition follows from the definition of the labor share, $\alpha_L = wL/(PY)$. For homogeneous labor, rearranging and taking log changes gives

$$\frac{w}{P} = \alpha_L \frac{Y}{L} \quad \longrightarrow \quad d\log(w/P) = d\log \alpha_L + d\log(Y/L). \tag{7.32}$$

A falling labor share can therefore coexist with a rising real wage if output per worker rises enough to offset the decline in the share. In the race model, $P = 1$ and L is fixed, so automation raises output per worker at rate g_I while lowering the log labor share at rate

$1/(N-I)$. These are precisely the two terms in (7.31).

A small cost advantage for the new machine gives a small g_I. In that case the displacement term dominates and the equilibrium wage falls, even though output rises. This supplies the channel missing from the capital-augmenting CES experiment in Section 3.3.

What happens to the labor share?. It decreases unambiguously with automation, as does the relative price of labor:

$$\frac{\partial \alpha_L}{\partial I} = -1, \qquad \left.\frac{\partial \log(w/r)}{\partial I}\right|_{K,L,N} = -\frac{1}{N-I} - \frac{1}{I-N+1} < 0. \tag{7.33}$$

New tasks reinstate labor. Now hold the automation frontier fixed and increase N. For now, this increase is given: we first establish its effects, then discuss the incentives to create the new tasks. New tasks move the interval of tasks being performed. An interval of old machine tasks disappears at $N-1$, while an equal interval of new labor tasks enters at N. Labor's bundle becomes longer, and m falls. This is the *reinstatement effect* of new tasks.

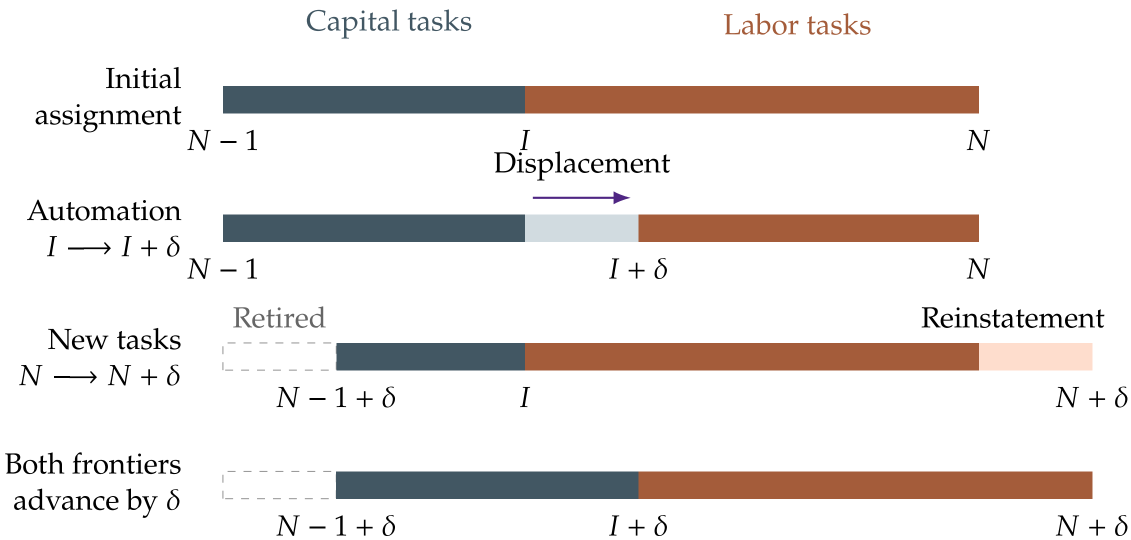
Figure 25 shows the two changes in the technological frontiers separately and together. Automation moves the boundary inside the existing task interval. New-task creation shifts the interval itself. Advancing both frontiers by the same amount preserves the lengths of the capital and labor bundles, even though the activities performed by each input change.

New tasks are the most productive, they expand the limits of what we can do. Accordingly, we assume that the adoption of these new tasks increases productivity (or reduces costs) relative to what we lose from the tasks that disappear at the bottom of the interval. We see this in the effective cost of the new tasks relative to the old tasks:

$$\frac{w}{a_L(N)} < r. \tag{7.34}$$

The new labor task is cheaper than the old machine task it replaces. Combining this

Figure 25. Automation Removes Labor Tasks; New Tasks Reinstate Labor



Notes: Each active interval has length one. Blue denotes tasks assigned to capital and red denotes tasks assigned to labor. Lighter colors identify newly automated tasks or newly created labor tasks; dashed boxes mark retired tasks outside the active interval. Rows two and three are separate changes from the initial assignment. In the final row, both frontiers advance equally, preserving $m = I - N + 1$ and $\alpha_L = N - I$. These comparisons assume that the relevant technology and new-task adoption conditions hold.

condition with the condition for automation at I gives

$$a_L(I) < \frac{w}{r} < a_L(N). \tag{7.35}$$

Labor is relatively unproductive in the tasks being automated and relatively productive in the new tasks being created.

Differentiate (7.26) with respect to N, using $\partial m / \partial N = -1$:

$$\left.\frac{\partial \log Y}{\partial N}\right|_{K,L,I} = \log \frac{a_L(N)L/(1-m)}{K/m} = \log \frac{r}{w/a_L(N)} \equiv g_N > 0. \tag{7.36}$$

The associated wage and share responses are

$$\left.\frac{\partial \log w}{\partial N}\right|_{K,L,I} = \underbrace{g_N}_{\text{productivity}} + \underbrace{\frac{1}{N-I}}_{\text{reinstatement}} > 0, \qquad \frac{\partial \alpha_L}{\partial N} = 1. \tag{7.37}$$

Both effects now favor labor: workers have more tasks to perform, and the new tasks increase the value of production.

The race between the two frontiers. Combining the two changes gives

$$d\alpha_L = dN - dI, \qquad d\log w = g_I\, dI + g_N\, dN + \frac{dN - dI}{N - I}, \qquad K, L \text{ fixed.} \tag{7.38}$$

Automation running ahead of new-task creation lowers the labor share. New-task creation running ahead raises it. Wages add to this the positive productivity effect of moving either technological frontier.

What makes the new-task frontier expand. An increase in N requires research that develops a productive new activity or an improved version of an existing one in which labor has a comparative advantage. In practice, some of the resources of the economy go to research that develops these new modes of production. Researchers can choose between developing automation technologies and creating new tasks. The make this choice based on the prospective profits they receive from the property rights over these new technologies (like patents or the creation of new firms). Three forces determine how much research is directed toward expanding N:

(a) **The productive advantage of the new task.** new labor task is cheaper than the old machine task it replaces ($w/a_L(N) < r$). The higher this productivity gain, the higher the incentive to move this frontier.

(b) **Research capacity and technical difficulty.** Technology also determines how easy it is for researchers to move each frontier. The new-task frontier advances faster when more researchers work on it or when each researcher can develop more new tasks. A scientific advance that makes automation easier can instead attract researchers away from new-task creation. The two directions of innovation therefore compete for research resources.

(c) **Expected returns relative to automation.** Research allocation depends on the present discounted value of profits an innovation can earn. Expected future wages and rental rates matter because they determine how attractive each technology will remain after it is developed. Intellectual property arrangements also matter for how much of the gain inventors can capture. A socially useful new task need not offer the highest private return to research.

These incentives create a feedback from the allocation of tasks to the direction of innovation. In our fixed-K/L comparison, automation lowers w/r, narrowing the advantage of replacing workers and widening the advantage of productive new labor tasks. So, a temporary lead by automation shifts research incentives back toward new-task creation and supports a stable path on which both frontiers advance together. However, a permanent improvement in automation research can lead to a path with a lower labor share, and sufficiently cheap capital can lead to complete automation. Thus the expansion of N depends on productive opportunities, research resources, and the rewards for developing labor-using technologies.

7.7. Automation and new tasks cheatsheet

General Case. Direct effects hold w, r fixed, with $P = c$ and task elasticity $\eta > 0$. The results describe pure automation: only newly automated tasks change technique.

Object	Mathematical expression and meaning
Task costs and adoption	$c_L(i) = \dfrac{w}{a_L(i)}, \qquad c_K(i) = \dfrac{r}{a_K(i)}.$ Automate a feasible task when $c_K(i) < c_L(i)$.
Displacement	$D = \dfrac{\int_{\Delta\mathcal{A}} a_L(i)^{\eta-1}\,di}{H_L^0}, \qquad H_L^1 = (1-D)H_L^0.$ Fraction of initial employment in newly automated tasks $\Delta\mathcal{A}$; H_L^0 sums $a_L(i)^{\eta-1}$ over initial labor tasks.
Cost saving	$\Gamma = \log \dfrac{c^0}{c^1} > 0\,; \qquad \text{if} \quad \eta = 1 \longrightarrow \Gamma = \displaystyle\int_{\Delta\mathcal{A}} \log \frac{c_L(i)}{c_K^1(i)}\,di\,.$ Log fall in final-good unit cost at unchanged factor prices.
Adopter's output	$Y = Bc^{-\varepsilon}, \qquad \Delta \log Y = \varepsilon\Gamma.$ $\varepsilon > 0$ is the elasticity of product demand; B is fixed.

Question and what is fixed	Relevant relationship	Answer and intuition
Does labor per unit of output fall?	$\Delta \log(L/Y)$ $= \log(1-D) - \eta\Gamma < 0$	Yes. Labor loses tasks and faces substitution toward cheaper machine services.
Can employment rise at an adopter? Demand schedule fixed.	$\Delta \log L$ $= \log(1-D) + (\varepsilon - \eta)\Gamma$	Yes if $(\varepsilon - \eta)\Gamma > -\log(1-D)$. Output expansion must offset displacement and substitution.
Does the labor share fall?	$\Delta \log \alpha_L$ $= \log(1-D) + (1-\eta)\Gamma$	Yes. Cost savings partly offset displacement if $\eta < 1$ and reinforce it if $\eta > 1$. At $\eta = 1$, only D matters.

Race Results: definitions and equilibrium. The race model fixes K, L, lets prices adjust, and sets $\eta = 1$, $a_K(i) = 1$, and output price one, with increasing $a_L(i)$.

Object	**Mathematical expression and meaning**
Race-model frontiers and assignment	$i \in [N-1, N], \qquad I = \min\{\bar{I}, \widetilde{I}\}, \qquad a_L(\widetilde{I}) = w/r.$ N: new-task frontier; $\bar{I}$: machine capability; $\widetilde{I}$: unconstrained cost cutoff; I: actual assignment cutoff.
Task shares and production	$m = I - N + 1, \quad \alpha_K = m, \quad \alpha_L = 1 - m = N - I.$ $Y = \exp\left(\int_I^N \log a_L(i)\, di\right)\left(\frac{K}{m}\right)^m \left(\frac{L}{1-m}\right)^{1-m}.$ Capital performs $[N-1, I]$ and labor performs $(I, N]$; $0 < m < 1$.
Factor prices and market clearing	$r = m\frac{Y}{K}, \qquad w = (1-m)\frac{Y}{L}, \qquad \frac{w}{r} = \frac{N-I}{I-N+1}\frac{K}{L}.$ Relative prices adjust to employ the fixed factor supplies.
Marginal productivity gains	$g_I = \log \frac{w/a_L(I)}{r}, \qquad g_N = \log \frac{r}{w/a_L(N)}.$ Log cost savings from automation and new tasks. Both are positive when $a_L(I) < w/r < a_L(N)$.

Race Results: main questions. Derivatives fix K, L and productivity schedules. For the frontier experiments, assume a binding automation frontier and productive new tasks ($g_I, g_N > 0$).

Question and what is fixed	Relevant relationship	Answer and intuition
Does more machine capability always change assignment?	$I = \min\{\bar{I}, \widetilde{I}\}$	Only if binding. If slack, a small rise in $\bar{I}$ changes neither assignment nor prices.
Does automation raise output and wages? Race model; N fixed.	$\frac{\partial \log Y}{\partial I} = g_I > 0$ $\frac{\partial \log w}{\partial I} = g_I - \frac{1}{N-I}$	Output rises. The wage falls if displacement exceeds the productivity gain. Labor's share falls: $\partial \alpha_L / \partial I = -1$.
Can wages rise while labor's share falls?	$\frac{w}{P} = \alpha_L \frac{Y}{L}$	Yes, if output per worker rises enough to offset the falling share; see (7.32).
Do productive new tasks help workers? Race model; I fixed.	$\frac{\partial \log Y}{\partial N} = g_N > 0$ $\frac{\partial \log w}{\partial N} = g_N + \frac{1}{N-I} > 0$	Output and the wage rise. Reinstatement expands labor's bundle: $\partial \alpha_L / \partial N = 1$.
What if both frontiers advance? Race model.	$d\alpha_L = dN - dI$ $d \log w = g_I\, dI + g_N\, dN + \frac{dN - dI}{N-I}$	Equal advances keep the share constant while productivity raises wages. Automation outpacing new tasks lowers the share.
What determines new-task creation?	Productive advantage; research capacity; expected private returns.	Inventors must cover research costs. The two technologies compete for researchers; keeping pace is not automatic.

References

Daron Acemoglu and David H. Autor. Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, volume 4B. Elsevier, 2011. doi: 10.1016/S0169-7218(11)02410-5. URL https://doi.org/10.1016/S0169-7218(11)02410-5.

Daron Acemoglu and Pascual Restrepo. The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6): 1488–1542, 2018. doi: 10.1257/aer.20160696. URL https://doi.org/10.1257/aer.20160696.

Daron Acemoglu and Pascual Restrepo. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30, 2019. doi: 10.1257/jep.33.2.3. URL https://doi.org/10.1257/jep.33.2.3.

Daron Acemoglu and Pascual Restrepo. Tasks, automation, and the rise in U.S. wage inequality. *Econometrica*, 90(5):1973–2016, 2022. doi: 10.3982/ECTA19815. URL https://doi.org/10.3982/ECTA19815.

Daron Acemoglu and Fabrizio Zilibotti. Productivity differences. *The Quarterly Journal of Economics*, 116(2):563–606, 2001. doi: 10.1162/00335530151144104. URL https://doi.org/10.1162/00335530151144104.

Daron Acemoglu, Fredric Kong, and Pascual Restrepo. Tasks at work: Comparative advantage, technology and labor demand. In *Handbook of Labor Economics*, volume 6, pages 1–114. Elsevier, 2025. doi: 10.1016/bs.heslab.2025.08.003. URL https://doi.org/10.1016/bs.heslab.2025.08.003.

Kenneth J. Arrow. Citation classic: Capital-labor substitution and economic efficiency. Current Contents, no. 17, April 1979. URL https://faculty.econ.ucsb.edu/~tedb/Courses/GraduateTheoryUCSB/ArrowCheneryMinhasCiteClassic.pdf. Retrospective dated December 27, 1977. Author's retrospective.

Kenneth J. Arrow, Hollis B. Chenery, Bagicha S. Minhas, and Robert M. Solow. Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics*, 43(3):225–250, 1961. doi: 10.2307/1927286. URL https://doi.org/10.2307/1927286. Original paper.

David Autor, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709, 2020. doi: 10.1093/qje/qjaa004. URL https://doi.org/10.1093/qje/qjaa004.

David H. Autor and David Dorn. The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, 103(5):1553–1597, 2013. doi: 10.1257/aer.103.5.1553. URL https://doi.org/10.1257/aer.103.5.1553.

David H. Autor, Frank Levy, and Richard J. Murnane. The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4):1279–1333, 2003. doi: 10.1162/003355303322552801. URL https://doi.org/10.1162/003355303322552801.

David H. Autor, Lawrence F. Katz, and Melissa S. Kearney. The polarization of the U.S. labor market. *American Economic Review*, 96(2):189–194, 2006. doi: 10.1257/000282806777212620. URL https://doi.org/10.1257/000282806777212620.

David H. Autor, Lawrence F. Katz, and Melissa S. Kearney. Trends in U.S. wage inequality: Re-assessing the revisionists. *Review of Economics and Statistics*, 90(2):300–323, 2008. doi: 10.1162/rest.90.2.300. URL https://doi.org/10.1162/rest.90.2.300.

Robert J. Barro. Are government bonds net wealth? *Journal of Political Economy*, 82(6):1095–1117, 1974. doi: 10.1086/260266. URL https://doi.org/10.1086/260266.

William J. Baumol. Macroeconomics of unbalanced growth: The anatomy of urban crisis. *American Economic Review*, 57(3):415–426, 1967. URL https://cooperative-individualism.org/baumol-william_macroeconomics-of-unbalanced-growth-1967-jun.pdf.

William J. Baumol and William G. Bowen. *Performing Arts: The Economic Dilemma: A Study of Problems Common to Theater, Opera, Music and Dance*. Twentieth Century Fund, New York, 1966. URL https://books.google.com/books?id=fiS0AAAAIAAJ.

Charles W. Cobb and Paul H. Douglas. A theory of production. *The American Economic Review*, 18(1, Supplement):139–165, 1928. URL https://msuweb.montclair.edu/~lebelp/CobbDouglasProdAER1928.pdf. Original paper.

Evsey D. Domar. Capital expansion, rate of growth, and employment. *Econometrica*, 14(2):137–147, 1946. doi: 10.2307/1905364. URL https://doi.org/10.2307/1905364.

Rudiger Dornbusch, Stanley Fischer, and Paul A. Samuelson. Comparative advantage, trade, and payments in a Ricardian model with a continuum of goods. *American Economic Review*, 67(5): 823–839, 1977. URL https://users.econ.umn.edu/~tkehoe/papers/DFS_Ricardo.pdf.

Paul H. Douglas. Comments on the Cobb-Douglas production function. In Murray Brown, editor, *The Theory and Empirical Analysis of Production*, pages 15–22. National Bureau of Economic Research, 1967. URL https://www.nber.org/system/files/chapters/c1474/c1474.pdf. Author's retrospective.

Jonathan Eaton and Samuel Kortum. Technology, geography, and trade. *Econometrica*, 70(5): 1741–1779, 2002. doi: 10.1111/1468-0262.00352. URL https://doi.org/10.1111/1468-0262.00352.

Gene M. Grossman and Esteban Rossi-Hansberg. Trading tasks: A simple theory of offshoring. *American Economic Review*, 98(5):1978–1997, 2008. doi: 10.1257/aer.98.5.1978. URL https://doi.org/10.1257/aer.98.5.1978.

Roy F. Harrod. An essay in dynamic theory. *The Economic Journal*, 49(193):14–33, 1939. doi: 10.2307/2225181. URL https://doi.org/10.2307/2225181.

John R. Hicks. *The Theory of Wages*. Macmillan and Co., Limited, London, 1932. URL https://dspace.gipe.ac.in/xmlui/bitstream/handle/10973/24960/GIPE-009904.pdf?isAllowed=y&sequence=3. Chapter VI, distribution and economic progress. Original book scan.

Samuel Hollander. Retrospectives: Ricardo on machinery. *Journal of Economic Perspectives*, 33(2): 229–242, 2019. doi: 10.1257/jep.33.2.229. URL https://www.aeaweb.org/articles?id=10.1257/jep.33.2.229.

Joachim Hubmer and Pascual Restrepo. Not a typical firm: Capital–labor substitution and firms' labor shares. *American Economic Journal: Macroeconomics*, 18(2):34–71, April 2026. doi: 10.1257/mac.20230325. URL https://www.aeaweb.org/articles?id=10.1257/mac.20230325.

Charles I. Jones and Christopher Tonetti. Past automation and future A.I.: How weak links tame the growth explosion. Working paper, 2026. URL https://www.web.stanford.edu/~chadj/JonesTonetti_Automation.pdf.

Loukas Karabarbounis. Perspectives on the labor share. *Journal of Economic Perspectives*, 38(2): 107–136, 2024. doi: 10.1257/jep.38.2.107. URL https://www.aeaweb.org/articles?id=10.1257/jep.38.2.107.

Loukas Karabarbounis and Brent Neiman. The global decline of the labor share. *The Quarterly Journal of Economics*, 129(1):61–103, 2014. doi: 10.1093/qje/qjt032. URL https://doi.org/10.1093/qje/qjt032.

John Maynard Keynes. Relative movements of real wages and output. *The Economic Journal*, 49(193):

34–51, 1939. URL https://www.hetwebsite.net/het/texts/keynes/keynes1939relative.htm. Transcription with original page markers.

William D. Nordhaus. Baumol's diseases: A macroeconomic perspective. *The B.E. Journal of Macroeconomics*, 8(1):1–39, 2008. doi: 10.2202/1935-1690.1382. URL https://doi.org/10.2202/1935-1690.1382.

Sergio Ocampo. A task-based theory of occupations with multidimensional heterogeneity. Working paper, 2026. URL https://sites.google.com/site/sergiocampod/.

Pascual Restrepo. Automation: Theory, evidence, and outlook. *Annual Review of Economics*, 16:1–25, 2024. doi: 10.1146/annurev-economics-090523-113355. URL https://doi.org/10.1146/annurev-economics-090523-113355.

David Ricardo. *On the Principles of Political Economy and Taxation*. John Murray, London, 3 edition, 1821a. URL https://www.marxists.org/reference/subject/economics/ricardo/tax/index.htm. First edition 1817.

David Ricardo. On machinery. In *On the Principles of Political Economy and Taxation*, chapter 31. John Murray, London, 3 edition, 1821b. URL https://www.marxists.org/reference/subject/economics/ricardo/tax/ch31.htm.

A. D. Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146, 1951. doi: 10.1093/oxfordjournals.oep.a041827. URL https://doi.org/10.1093/oxfordjournals.oep.a041827.

Michael Sattinger. Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31(2):831–880, 1993. URL https://hceconomics.uchicago.edu/sites/default/files/pdf/events/Sattinger_1993_JEL_v31_n2.pdf.

Robert M. Solow. A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1):65–94, 1956. doi: 10.2307/1884513. URL https://doi.org/10.2307/1884513. Original paper.

Jón Steinsson. Production. Lecture notes, University of California, Berkeley, 2026. URL https://jonsteinsson.com/teaching/production.pdf.

Jan Tinbergen. Substitution of graduate by other labour. *Kyklos*, 27(2):217–226, 1974. doi: 10.1111/j.1467-6435.1974.tb01903.x. URL https://doi.org/10.1111/j.1467-6435.1974.tb01903.x.