# Input-output

**A crash course**

Dr. Rowan Shi

May 12, 2023

This document is meant to be a reference about input-output. Covers first the basics of input-output, limited to what's essential. Then, moves on to how these are typically represented in data. Finally, a theoretical treatment of the IO model and its key intuitions (not necessary, but conceptually interesting).

## Contents



## 1 Basics

Input-output at its core is the idea that an industry (say, textile manufacturing) requires *input materials* (say, from machinery manufacturing) as part of its production process. For example, to manufacture textiles and clothing, one might expect we need heavy factory machinery that is produced by the machinery manufacturing industry. Here, the *input* is manufactured machinery and the *output* is manufactured textiles. The input is sometimes also called an *intermediate good* or an intermediate.

In practice, production processes can be quite complex (and growing increasingly more so!). For example, the textile manufacturing industry might use inputs from multiple other industries, not just machinery manufacturing. In particular,

- the output industry usually heavily uses its *own* production as input (textiles production using textiles is a good example)
- the output industry could use inputs from all industries in the economy

Countries are generally motivated to track these usage patterns to understand their economy and how production takes place. This data is often called the *input-output table* (io table for short). It generally records, for each *output* industry, the total value of intermediates used from each *input* industry. This data is *bilateral* in the sense that every observation corresponds to an industry *pair* (the input industry and the output industry).

As an example of an input-output table, consider (in millions of dollars):

| | Output industry | | Total |
|---|---|---|---|
| | Textiles | Machinery | |
| Textiles | 1 | 0 | 1 |
| Machinery | 0.25 | 2 | 2.25 |
| Total | 1.25 | 2 | 3.25 |

In this economy, there are two industries: textiles and machinery. The columns of the table represent output industries (the good being produced). The rows represent the input industries (the source of the intermediate). We can see that textile production (the first column) used $1M of textiles as input and $0.25M of machinery as input. The total value of inputs to textile production was $1.25M. Similarly, machinery production (the second column) used no input from textiles, but $2M of machinery. Reading across the rows, we can see that textiles produced $1M of inputs for textile production, and no inputs for machinery production. In total, it produced $1M worth of intermediates. On the other hand, machinery produced $0.25M worth of inputs for textile production, and $2M worth of inputs for machinery production. A key outcome, of course, is that total inputs used is equal to total inputs produced.

# 2 Data practicalities

Input-output tables collected by government agencies are more complex than the simple example above. There are usually many more industries. They also tend to include more than just intermediate inputs, but also final demand and non-intermediate inputs.

## 2.1 Non-bilateral information

Recall that the columns correspond to the good being produced. Reading down a single column is like reading a recipe: we see all the inputs that industry used to produce its output. Beyond intermediate goods, usually an industry will need additional factors to produce its output. Labour is a key input (and, increasingly, capital). National accounts data generally include these to paint a full picture of everything needed to produce a certain good. These appear as additional rows, usually at the bottom of the table. Consider the table below. It now includes an additional line for labour, which describes the total "value of labour" used in the production of each good. For textiles in the first column, we can see that it used $3M of labour (in other words, textile workers were paid $3M). In total, the textile industry spent $4.25M on costs — $1.25 of which was on intermediate inputs, and $3M of which was on labour. In the same way, we can read the machinery column to find that it paid $2.75M in wages. At the end of the day, workers in this economy were compensated $5.75M in total (reading the labour row across). In general, the contributions of the factors that aren't intermediates is called "value added" (i.e., this industry took some raw materials, which are the intermediate inputs, then *added value* via labour and other transformations).

| | Intermediate | | Final | Total |
|---|---|---|---|---|
| | Textiles | Machinery | | |
| Textiles | 1 | 0 | 3.25 | 4.25 |
| Machinery | 0.25 | 2 | 2.5 | 4.75 |
| Labour | 3 | 2.75 | | 5.75 |
| Total spent | 4.25 | 4.75 | 5.75 | 14.75 |

Recall also that the rows correspond to who uses a certain industry's products. We said above that, of the produced machinery, $0.25M of it went to the textile industry to make garments, and $2M of it went back into the machinery industry to produce more machinery. Of course, machinery could also have other uses. In particular, we might expect that households want machinery (and textiles) too. These uses are *final demand* — final demand in contrast to *intermediate demand*. Final demand is therefore usually added as an additional column to the input-output table. Reading across the machinery row, we can see that, in addition to the $2.25M of machinery used as intermediate input, and additional $2.5M of machinery goes to households. The total demand for machinery in this economy was thus $4.75M.

In practice, there are usually more non-intermediate inputs than just labour. Real IO tables generally have several additional rows at the bottom, capturing the various additional costs to production. These may be other factors, such as non-labour value added, or costs like taxes. In the same way, there are usually multiple final demand categories, each with its own column in the full IO table. These may include domestic household demand, government or public demand, and exports.

## 2.2 Industry encoding

To represent industries in an easily digestible format, many IO tables use certain industry encodings. These are encoding systems that assign a number to each industry. For example, in our example, we could encode the textile industry as industry 1 and the machinery industry as industry 2. You may encounter an IO table without text, but just industry encodings:

| | Intermediate | | Final | Total |
|---|---|---|---|---|
| | 1 | 2 | | |
| 1 | 1 | 0 | 3.25 | 4.25 |
| 2 | 0.25 | 2 | 2.5 | 4.75 |
| Labour | 3 | 2.75 | | 5.75 |
| Total spent | 4.25 | 4.75 | 5.75 | 14.75 |

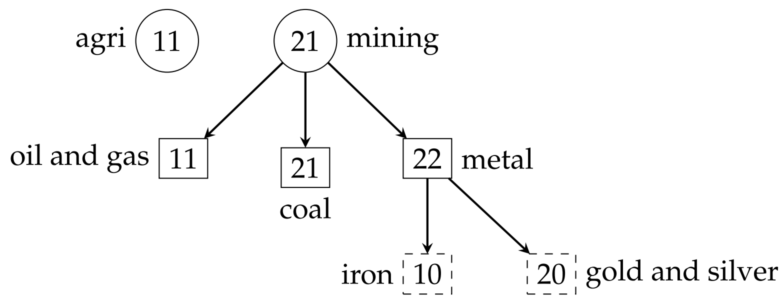
along with a separate file or sheet that lists the industries and the corresponding code:

| Description | Code |
|---|---|
| Textiles | 1 |
| Machinery | 2 |

There are several standardised encodings, including NAICS, NACE, SIC, and so on. If the IO table is using these encodings, it may not provide the second table, since it's assumed you have it yourself. You can easily look up these encodings to see which industry corresponds to which code, but be careful — each of these has gone through several revisions, where industries have been recoded, so it's important to get the correct version or year.

**Levels of aggregation** Depending on how detailed your data is, an industry encodings may be very fine or very coarse. The NAICS encoding is a good example. At the highest level, there are encodings like Mining, Agriculture, Construction, and so on. However, within each of these high level industries, there is also further classification: within Mining (code 21), there are the subindustries "Coal mining" (code 2121), "Metal ore mining" (2122), and others. NAICS in turn breaks down metal ore mining into "Iron" (212210), "gold and silver" (212220), and more. If your data is extremely detailed, you might get values at extremely fine levels of classification.

The standard way to deal encoding subindustries is to read the encodings left to right. NAICS follows this convention, along with most systems.




When reading the encoding for Mining of Iron (212210),

- the first two digits (21) correspond to the highest level of classification, Mining (the circle)
- the middle two digits (22) correspond to the first level of subclassification, Metals (the solid rectangle); the full encoding for the classification mining of metals is 2122
- the last two digits (10) corresponding to the deepest level of subclassification, Iron (the dashed rectangle)

For this reason, you may see phrases like "industries at the 2-digit level", "industries at the 4-digit level", and so on.

**Crosswalks** If you have several datasets with encoded industries, they may use different encoding standards. If that's the case, it's common to attempt to harmonise them. With the common encodings, crosswalks that translate one to another tend to exist online. However, the crosswalks sometimes that need to be done by hand, especially for uncommon industry encodings.

### 2.3 Wide vs. long format

Sometimes, it can be helpful to reshape the data from the wide format (the tables above, usually how IO tables are presented) to a "long" format. This format usually has three key variables: the input type, the output type, and the value associated with that input-output pair. In the long format, our table above would look something like

| input | output | value |
|---|---|---:|
| 1 | 1 | 1 |
| 1 | 2 | 0 |
| 1 | final | 3.25 |
| 2 | 1 | 0.25 |
| 2 | 2 | 2 |
| 2 | final | 2.5 |
| labour | 1 | 3 |
| labour | 2 | 2.75 |

Especially with long data, it may be useful to separate out the bilateral data from the non-bilateral observations:

| input | output | value |
|---|---|---:|
| 1 | 1 | 1 |
| 1 | 2 | 0 |
| 2 | 1 | 0.25 |
| 2 | 2 | 2 |

| industry | labour | final |
|---|---|---:|
| 1 | 3 | 3.25 |
| 2 | 2.75 | 2.5 |

## 3 Model and multipliers

Though it may be easy to observe that industries use intermediate inputs during production in practice, what are the theoretical implications of this production pattern? The Leontief model is a simple way to represent this type of economic activity and draw out its implications.

## 3.1 Multipliers

Let's take a quick moment to discuss multipliers. It's likely you've seen this concept in macroeconomics: if you give the household a dollar, they might spend a fraction of it (saving the rest). This fraction is the *marginal propensity to consume*. However, because a dollar spent must be spent on something, one household's expenditure is another household's income. Let's say the MPC is 0.25, so that for every dollar earned, the household spends a quarter. If the government spends \$1 on something (say, civic services) a civil servant somewhere earns \$1; then,

- they spend \$0.25 on something (say, a cookie) — then, someone (a baker) will earn \$0.25
- the baker will in turn spend a quarter of that, i.e. $0.25^2$ on something (say, a dog treat) — then, a dog treat maker will earn that amount
- the dog treat maker will then spend a quarter of that, i.e. $0.25^3$, and so on

So ultimately, the initial dollar spent by the government generates more than a dollar of overall additional expenditure. Let's let $x$ represent the total amount of stimulated spending. As we follow this sequence of events, we can add up the total additional expenditure in this economy generated from the first dollar given:

$$x = 1 + 0.25 + 0.25^2 + 0.25^3 + \dots$$

How can we solve for the actual value of $x$ if it's a sum that goes on forever? There's some theoretical math that goes into the actual formal proof, but the general idea is to multiply both sides by 0.25:

$$\begin{aligned} x &= 1 + 0.25 + 0.25^2 + 0.25^3 + \dots \\ 0.25x &= \phantom{1 +{}} 0.25 + 0.25^2 + 0.25^3 + 0.25^4 + \dots \end{aligned}$$

So subtracting the second line from the first,

$$\begin{aligned} x - 0.25x &= 1 \\ (1 - 0.25)x &= 1 \\ x &= \frac{1}{1 - 0.25} = (1 - 0.25)^{-1} = \frac{4}{3} > 1 \end{aligned}$$

So in this example, the multiplier is 4/3.

In general, if $\alpha$ is the MPC, you can work out that the multiplier will be $(1-\alpha)^{-1}$. The larger the MPC ($\alpha$), the larger the multiplier. For this reason, macroeconomists are very interested in knowing the value of the MPC — it determines how impactful fiscal policy will be.

## 3.2 One industry model

Before we consider multiple industries, let's start with the simple example of a single industry. Let's call it "machinery". Machinery uses labour and other machinery to produce more machinery. Let's suppose we observed that machinery employed \$2.75M of labour and used \$2M of machinery inputs to produce \$4.75M of machinery in total. In what follows, we'll use millions of dollars as our units.

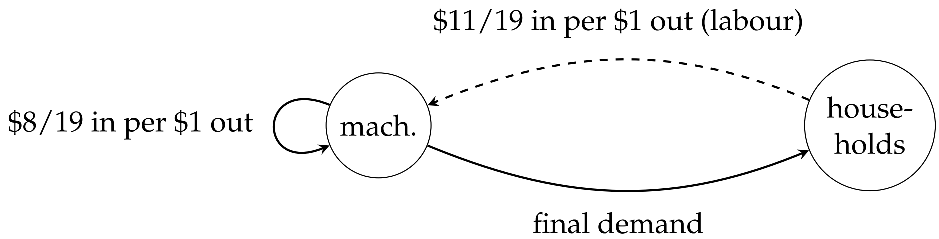
Let's suppose that machinery is constant returns to scale. That means that, for every dollar of machinery produced, we need

$$\frac{2.75}{4.75} = \frac{11}{19}$$

dollars of labour and

$$\frac{2}{4.75} = \frac{8}{19}$$

dollars of machinery inputs. Sometimes, these are called the "labour share" (of production costs) and the "intermediates share". Represented in a diagram, the economy looks like:




Without the dashed line, this picture is quite similar to those from the first week of ECN230. The only thing we've added here is that machinery probably employs workers, alongside using intermediate inputs.

Let's introduce some notation. Let $p$ be the unit price of machinery, $q$ be the total machinery produced, $w$ the wage, and $L$ the total number of hours worked. We know that the total value of machinery used as *input* is

$$\frac{8}{19}\frac{\$\text{ input}}{\$\text{ output}}(\$pq\text{ output}) = \$\frac{8}{19}\text{ input}$$

We can account for the total value of machinery produced in a familiar way:

$$\begin{aligned}
\text{mach. produced} &= \text{mach. demanded} \\
&= \text{mach. for input} + \text{mach. for final consumption} \\
pq &= \frac{8}{19}pq + \text{mach. for final consumption} \\
\left(1 - \frac{8}{19}\right) pq &= \text{mach. for final consumption}
\end{aligned}$$

So far, we've captured the solid arrows from our diagram of the economy. At this point in ECN230, we would hope for the problem to give us a value for the final demand. Then, we can calculate the total value of machinery produced.

We can actually do better in this case. In particular, this time we've fleshed out the household side of the economy. They demand machinery for final consumption, and earn labour income. Since there's nothing else for households to spend money on (no other goods and no way to save), we know

$$\text{final demand for mach.} = \text{household income} = \text{labour income} = wL$$

Since the final demand for machinery must equal the supply of machinery for final demand, we ultimately know that

$$\text{mach. for final consumption} = wL$$

Substituting in,

$$\begin{aligned}
\left(1 - \frac{8}{19}\right) pq &= wL \\
pq &= \left(1 - \frac{8}{19}\right)^{-1} wL > wL
\end{aligned}$$

so the total value of production is equal to the total value of labour, multiplied by $(1 - \frac{8}{19})^{-1}$. Note that total production is the value of labour, multiplied a multiplier! Let's discuss these two components in turn, starting with the fraction.

If you refer back to Section 3.1, this fraction is very similar to our macro multiplier. The only difference is that the MPC from the macro multiplier has now been replaced by the intermediates share. It isn't a coincidence: it is the same sort of intuition. For every dollar of machinery input produced, we needed some machinery to begin with (to use as input). To make things more concrete, let's put some numbers to the concept. Let's say we know we produced \$1 of machinery output.

- to have produced that \$1 of machinery output, we know that we had to put $\$\frac{11}{19}$ of machinery production in as input.
- in turn, to have previously produced that $\$\frac{11}{19}$ of machinery to use as input, it *itself* needed machinery input; to be precise, $\$ \left(\frac{11}{19}\frac{11}{19}\right)$ worth
- in turn, to have previously produced that $\$ \left(\frac{11}{19}\right)^2$ of machinery to use as input, we needed $\$ \left(\frac{11}{19}^2\frac{11}{19}\right)$ worth of machinery input

So that \$1 of machinery output actually means we had to have produced more than a dollar! Adding up all the machinery we must have done,

$$1 + \frac{11}{19} + \left(\frac{11}{19}\right)^2 + \left(\frac{11}{19}\right)^3 \cdots = \frac{1}{1 - \frac{11}{19}} = \left(1 - \frac{11}{19}\right)^{-1}$$

When the intermediates share is high, the multiplier is high. Sometimes, this production pattern is called "roundabout production", because what's produced just turns around and goes back into the industry an intermediate for further production.

Well then, how much machinery total have we produced in this economy? We know we need to meet final demand, which is $wL$. To have produced $wL$ worth of final demand machinery, we need to have produced $wL\frac{1}{1-\frac{11}{19}}$ worth of machinery total. This intuition is the one we discussed in ECN230: given some total amount that will be diverted for household consumption, more needs to be produced to order to finally come up with the amount we dedicate to households.

We can think about this intuition from the other angle. In our economy, the only real fundamental factor is labour. The other factor of production, intermediates, are the final good itself. Roundabout production thus *magnifies* the total production of our fundamental factor: When we produce machinery, some of it goes back into the production of more machinery inputs! In a way, the industry feeds itself.

$$\underbrace{pq}_{\text{value of production}} = \left(1 - \frac{11}{19}\right)^{-1} \underbrace{wL}_{\text{value of labour}}$$

You can think about it as getting a lot of bang for your buck with labour. For each hour worked, you produce machinery, which can then be used as an input to produce more machinery, which can then be used as an input to produce more machinery, and so on. When the labour share is low (i.e. when the intermediates share is high), this effect is magnified. One hour of labour is very productive, because production is nearly "hands-off" — it mostly needs inputs rather than labour. In this way, the machinery industry mostly runs itself, and doesn't need much labour to produce a dollar of output. From this perspective, the higher the intermediates share (and thus the higher the multiplier), the more each hour of labour produces in the end. Given the important role for the intermediates share, let's call it $\alpha_{\text{mach.}} = \alpha_m$ in what follows.

Even though we only have one industry in this simple example, it already draws out one of the key intuitions of the Leontief economy: this multiplication or magnifications. This intuition persists even as we add more industries.

### 3.3 Two industries

Now let's add back in our textiles industry with our machinery industry. From our IO table, we know that textiles produced \$4.25M in total, for which it needed \$1M in textile input, \$0.25M in machinery input, and \$3M in labour. Then, the labour share in textiles is

$$\frac{3}{4.25} = \frac{12}{17}$$

and the intermediates share in textiles is

$$\frac{1.25}{4.25} = \frac{5}{17} \equiv \alpha_{\text{text.}} = \alpha_t$$

The machinery industry remains as before, with an intermediates share of $\frac{11}{19}$ and labour share of $\frac{9}{19}$.

With two industries, things are now a bit more complicated, since each industry has two sources of inputs. The machinery industry is as straightforward as before: it only uses machinery as input. For textiles, it uses both textiles and machinery as input. Out of the total \$1.25M spend on inputs, \$1M came from textiles itself and \$0.25M came from machinery. In other words,

$$\frac{1}{1.25} = \frac{4}{5} \equiv \eta_{\text{text.} \to \text{text.}} = \eta_{tt}$$

of input costs were on textile intermediates and

$$\frac{0.25}{1.25} = \frac{1}{5} \equiv \eta_{\text{mach.} \to \text{text.}} = \eta_{mt}$$

of input costs were on inputs from machinery (e.g. looms). These are sometimes called "input-output" shares.

We can also say that, among the total costs of textile production, a share

$$\frac{1}{4.25} = \frac{4}{17} = \underbrace{\frac{4}{5}}_{\eta_{\text{text.} \to \text{text.}}} \underbrace{\frac{5}{17}}_{\alpha_{\text{text.}}} = \eta_{tt}\alpha_t$$

was spent on textile inputs and

$$\frac{0.25}{4.25} = \frac{1}{17} = \underbrace{\frac{1}{5}}_{\eta_{\text{mach.} \to \text{text.}}} \underbrace{\frac{5}{17}}_{\alpha_{\text{text.}}} = \eta_{mt}\alpha_t$$

Similarly, for the production of machinery, a share

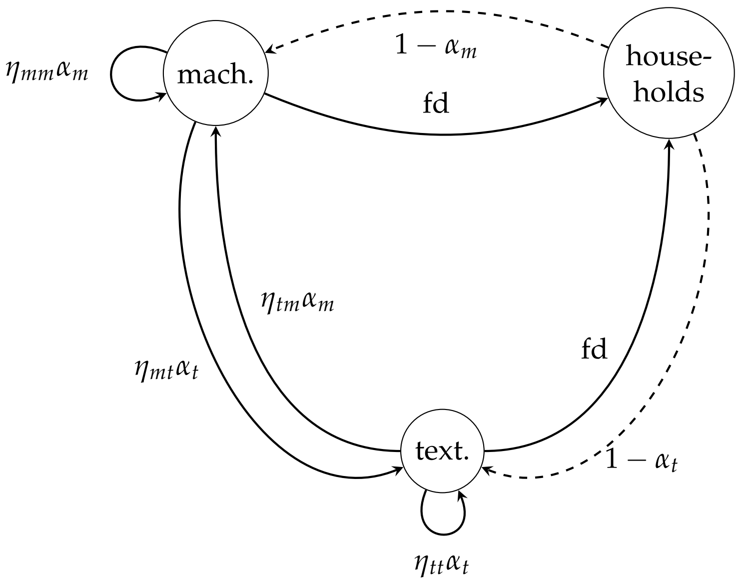
$$\frac{2}{2} = 1 = \eta_{mm}$$

of total inputs come from machinery; and a share

$$\frac{0}{2} = 0 = \eta_{tm}$$

comes from the textiles industry.

Now, we can draw the picture as before, where fd is final demand for each industry:



where we now have the lower portion representing the textiles industry.

Let's let the total value of textiles production be $t$ — it represents the total quantity of textiles produced times the unit price, to save on notation. Similarly, let's let the total value of machinery produced be $m$. We can account for the total produced textiles as before.

$$\begin{aligned}
\text{text. produced} &= \text{text. demanded} \\
&= \text{text. for input} + \text{text. for final consumption} \\
t &= \eta_{tt}\alpha_t t + \eta_{tm}\alpha_m m + \mathrm{fd}_t
\end{aligned}$$

Similarly,

$$m = \eta_{mm}\alpha_m m + \eta_{mt}\alpha_t t + \mathrm{fd}_m$$

So we can write the Leontief economy in matrix notation!

$$\begin{aligned}
t - \eta_{tt}\alpha_t t - \eta_{tm}\alpha_m m &= \mathrm{fd}_t \\
m - \eta_{mm}\alpha_m m - \eta_{mt}\alpha_t t &= \mathrm{fd}_m \\
\begin{bmatrix} (1-\eta_{tt}\alpha_t) & -\eta_{tm}\alpha_m \\ -\eta_{mt}\alpha_t & (1-\eta_{mm}\alpha_m) \end{bmatrix} \begin{bmatrix} t \\ m \end{bmatrix} &= \begin{bmatrix} \mathrm{fd}_t \\ \mathrm{fd}_m \end{bmatrix} \\
\left(\mathbf{I} - \begin{bmatrix} \eta_{tt}\alpha_t & \eta_{tm}\alpha_m \\ \eta_{mt}\alpha_t & \eta_{mm}\alpha_m \end{bmatrix}\right) \begin{bmatrix} t \\ m \end{bmatrix} &= \begin{bmatrix} \mathrm{fd}_t \\ \mathrm{fd}_m \end{bmatrix} \\
\left(\mathbf{I} - \begin{bmatrix} \eta_{tt} & \eta_{tm} \\ \eta_{mt} & \eta_{mm} \end{bmatrix} \begin{bmatrix} \alpha_t & 0 \\ 0 & \alpha_m \end{bmatrix}\right) \begin{bmatrix} t \\ m \end{bmatrix} &= \begin{bmatrix} \mathrm{fd}_t \\ \mathrm{fd}_m \end{bmatrix}
\end{aligned}$$

and finally:

$$\underbrace{\begin{bmatrix} t \\ m \end{bmatrix}}_{\text{production}} = \underbrace{\left(\mathbf{I} - \begin{bmatrix} \eta_{tt} & \eta_{tm} \\ \eta_{mt} & \eta_{mm} \end{bmatrix} \begin{bmatrix} \alpha_t & 0 \\ 0 & \alpha_m \end{bmatrix}\right)^{-1}}_{\text{multiplier}} \begin{bmatrix} \mathrm{fd}_t \\ \mathrm{fd}_m \end{bmatrix}$$

This matrix equation is the matrix version of what we got in our single-industry economy. On the left-hand side is total production (in the two industries). On the right hand side is the matrix version of the multiplier, $(\mathbf{I} - \boldsymbol{\eta\alpha})^{-1}$, multiplied by final demand. If we know how consumers will spend their total income, we can one again link final demand with the total value of labour, the fundamental factor. For example, in our previous tables, workers earned \$5.75M total in labour income, \$3.25M of which was spent on textiles and

\$2.5M of which was spent on machinery. If we assume that workers will always spend

$$\frac{3.25}{5.75} = \frac{13}{23} = \beta_t$$

of their income on textiles and

$$\frac{2.5}{5.75} = \frac{10}{23} = \beta_m$$

of their income on machinery, then we can rewrite the total final production on each good in terms of the total value of labour:

$$\underbrace{\begin{bmatrix} t \\ m \end{bmatrix}}_{\text{production}} = \underbrace{\left( \mathbf{I} - \begin{bmatrix} \eta_{tt} & \eta_{tm} \\ \eta_{mt} & \eta_{mm} \end{bmatrix} \begin{bmatrix} \alpha_t & 0 \\ 0 & \alpha_m \end{bmatrix} \right)^{-1}}_{\text{multiplier}} \begin{bmatrix} \beta_t \\ \beta_m \end{bmatrix} wL$$

The main observation is that the central intuition of the one-industry model still applies. The total value of labour is multiplied by the intermediates shares in the two industries. The key difference with multiple industries is that there are two intermediates shares: one for each industry. These are pre-multiplied by the input-output shares, which effectively calculate a weighted average of the two intermediates shares.