

Dark Matter of Trade: How Unobserved Intermediate Goods Shape Supply Chains

Sameer Malik

University of Houston

Fall 2024

Research Question

How do changes in tariffs propagate through production networks, and impact:

- Vertical Specialization
- Input Sourcing Strategy
- Employment and Economic Performance

Motivation



- **Policy Maker:** *Let's boost domestic EV industry!*
- **0% tariffs on direct inputs:** Lithium, Steel, and Aluminum
- **Increase tariffs on material used to produce Battery Casings from 20% to 35% (indirect inputs)**

Motivation



- **Policy Maker:** *Let's boost domestic EV industry!*
 - **0% tariffs on direct inputs:** Lithium, Steel, and Aluminum
 - **Increase tariffs on material used to produce Battery Casings from 20% to 35% (indirect inputs)**
-
- Establishments usually report directly bought inputs and final output(s) produced
 - This misses crucial element of production process, in-house produced goods, or unobserved intermediate goods (UIGS) – **Battery Casings**

Motivation



- **Policy Maker:** *Let's boost domestic EV industry!*
 - **0% tariffs on direct inputs:** Lithium, Steel, and Aluminum
 - **Increase tariffs on material used to produce Battery Casings from 20% to 35% (indirect inputs)**
-
- Establishments usually report directly bought inputs and final output(s) produced
 - This misses crucial element of production process, in-house produced goods, or unobserved intermediate goods (UIGS) – **Battery Casings**
 - **Ignoring tariffs on UIGS can lead to biased estimates of impact of trade policy** since plants adjust their production processes in response to changes in input tariffs

Main Contribution

1. Design an algorithm to construct Unobserved Intermediate Goods Set (UIGS)
2. Develop a new measure for examining impact of trade policy

Main Takeaway:

- By not considering tariffs on UIGS, estimated impact of trade policy is underestimated
- Decline in tariffs on UIGS increases:
 - (a) Vertical Specialization
 - (b) Sourcing of Imported Inputs
 - (c) Employment
 - (d) Sales, Profits, Labor Productivity
 - (e) Total Factor Productivity

Data I: Annual Survey of Industries India

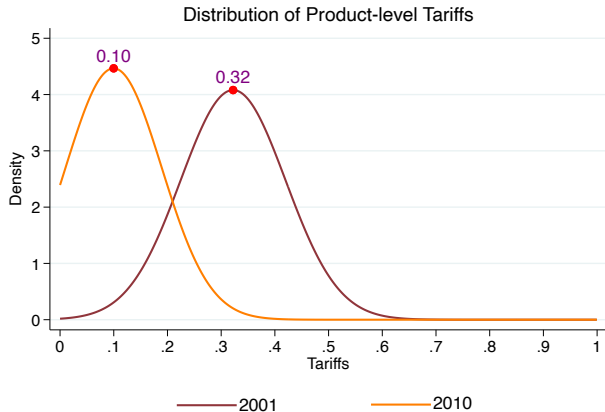
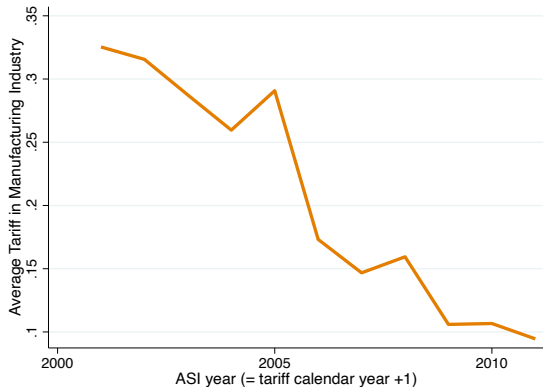
- **Plant-level panel** of formal manufacturing industry from 2001-2010, including:
 - All plants with > 100 employees
 - 1/5 of plants with 20 – 100 employees
- Information on product code, quantity, and unit value for **inputs** and **outputs**, at 5-digit level
- Restrict to single-output plants ($\approx 60\%$ of all plants)
- Using input and output information for 5236 products at HS-6 level (27 million product pairs), construct **Product-level Input-Output Table for India**

Data II: Trade Analysis Information System (TRAINS)

- **Import-Weighted Tariff Rates (in %)**

- Average tariff rates of HS 6-digit products weighted by India's imports from world in same year as tariff.
- Construct concordance from HS-6 to CPC to NPCMS to ASIC-2008 classification
- Merge with ASI plant-level panel data for **2001-10**

Trade Policy: Indian Manufacturing Sector



Unobserved Intermediate Goods Set (UIGS)

- Establishments report directly bought inputs and final output(s) produced
- This misses crucial element of production process, **in-house produced goods**, or **unobserved intermediate goods (UIGS)** – **Battery Casings for EV's**

Unobserved Intermediate Goods Set (UIGS)

- Establishments report directly bought inputs and final output(s) produced
- This misses crucial element of production process, **in-house produced goods**, or **unobserved intermediate goods (UIGS)** – **Battery Casings for EV's**
- **Ignoring tariffs on UIGS can lead to biased estimates of impact of trade policy** since plants adjust their production processes in response to changes in input tariffs

Unobserved Intermediate Goods Set (UIGS)

- Establishments report directly bought inputs and final output(s) produced
- This misses crucial element of production process, **in-house produced goods**, or **unobserved intermediate goods (UIGS)** – **Battery Casings for EV's**
- **Ignoring tariffs on UIGS can lead to biased estimates of impact of trade policy** since plants adjust their production processes in response to changes in input tariffs
- **How can we know intermediate goods a plant is using if it is not reported?**

Unobserved Intermediate Goods Set: Theory I

- Using product-level input-output table: (27 million $\omega - \omega'$ product pairs)

Unobserved Intermediate Goods Set: Theory I

- Using product-level input-output table: (27 million $\omega - \omega'$ product pairs)
- Expenditure on input ω to produce output ω' : $e_{\omega\omega'}$
- Let $\mathcal{D}(\omega)$ be set of products that are downstream to ω , i.e.,

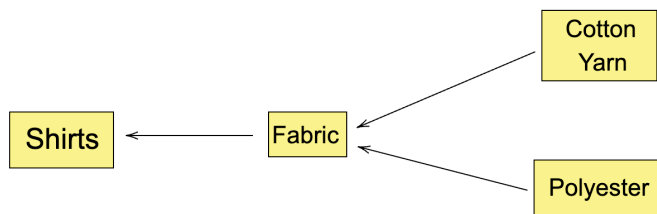
$$\mathcal{D}(\omega) := \{\omega' \mid e_{\omega\omega'} > 0\}$$

- Let $\mathcal{U}(\omega')$ be set of products that are upstream to ω' , i.e.,

$$\mathcal{U}(\omega') := \{\omega \mid e_{\omega\omega'} > 0\}$$

- Then, $\omega' \in \mathcal{D}(\omega)$, and $\omega \in \mathcal{U}(\omega')$

Unobserved Intermediate Goods Set: Illustration I



In this production chain,

$Inputs = \{Cotton\ Yarn, Polyester\}$

$Output = \{Shirts\}$

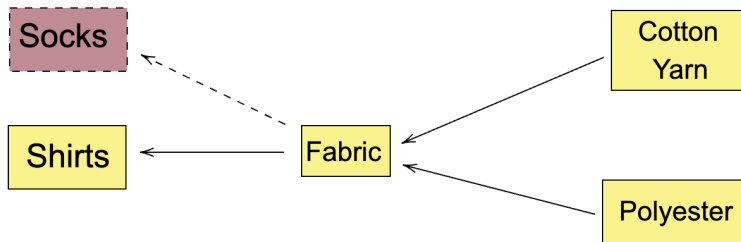
Reported
Output

UIGS

Reported
Inputs

- From I-O table:
- $\mathcal{D}(Cotton\ Yarn) = \{Fabric, Shirts, Socks, Pillowcases\}$,
 $\mathcal{D}(Polyester) = \{Fabric, Shirts, Socks, Car\ Seat\ Covers\}$,
- $[\mathcal{D}(Cotton\ Yarn) \cap \mathcal{D}(Polyester)] \setminus \{Shirts\} = \mathcal{I}_j = \{Fabric, Socks\}$.

Unobserved Intermediate Goods Set: Illustration I



In this production chain,

$Inputs = \{Cotton\ Yarn, Polyester\}$

$Output = \{Shirts\}$

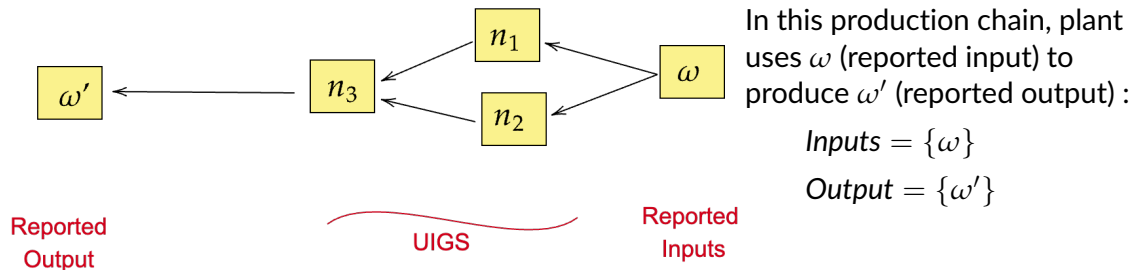
Reported
Output

UIGS

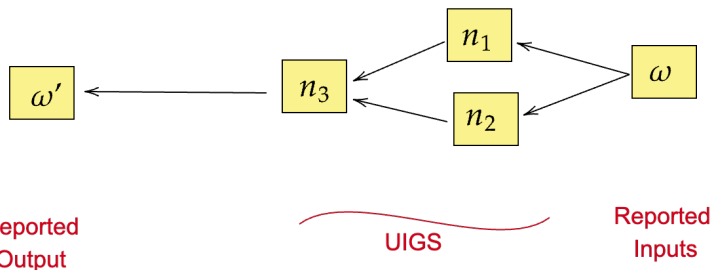
Reported
Inputs

- $[\mathcal{D}(Cotton\ Yarn) \cap \mathcal{D}(Polyester)] \setminus \{Shirts\} = \mathcal{I}_j = \{Fabric, Socks\}$
- $[[\mathcal{D}(Cotton\ Yarn) \cap \mathcal{D}(Polyester)] \setminus \{Shirts\}] \cap \mathcal{U}(Shirts) = \mathcal{I}_j = \{Fabric\}$

Unobserved Intermediate Goods Set: Illustration II



Unobserved Intermediate Goods Set: Illustration II



In this production chain, plant uses ω (reported input) to produce ω' (reported output) :

Inputs = $\{\omega\}$

Output = $\{\omega'\}$

- For $m = 1$ (reported inputs), $\mathcal{I}_j^1 = \{n_1, n_2\}$
- For $m = 2$, $\mathcal{D}(n_1) = \{n_3, \omega'\}$ and $\mathcal{D}(n_2) = \{n_3, \omega'\}$, and so, $\mathcal{I}_j^2 = \{n_3\}$
- Since, $\mathcal{I}_j^3 = \emptyset$, **UIGS**_j is: $\mathcal{I}_j^1 \cup \mathcal{I}_j^2 = \{n_1, n_2, n_3\}$

Unobserved Intermediate Goods Set: Theory II

- For a plant j : (in year t)

Unobserved Intermediate Goods Set: Theory II

- For a plant j : (in year t)
- If it uses inputs $\omega^1, \omega^2, \dots, \omega^k$ to produce output ω' , then,

$$\omega' \in \mathcal{D}(\omega^1), \omega' \in \mathcal{D}(\omega^2), \dots, \omega' \in \mathcal{D}(\omega^k)$$

$$\omega^1, \omega^2, \dots, \omega^k \in \mathcal{U}(\omega')$$

- And,

$$\underbrace{\bigcap_{i=1}^k \mathcal{D}(\omega^i)}$$

Goods that can be produced using this particular set of inputs

Unobserved Intermediate Goods Set: Theory III

- For a plant j , if it uses inputs $\omega^1, \omega^2, \dots, \omega^k$ to produce output ω' , then,

$$\omega' \in \mathcal{D}(\omega^1), \omega' \in \mathcal{D}(\omega^2), \dots, \omega' \in \mathcal{D}(\omega^k)$$

$$\omega^1, \omega^2, \dots, \omega^k \in \mathcal{U}(\omega')$$

- And,

$$\underbrace{\bigcap_{i=1}^k \mathcal{D}(\omega^i) \setminus [\{\omega'\} \cup \mathcal{D}(\omega')]}_{\text{Goods that can be produced using this particular set of inputs, except for output } \omega' \text{ and its downstream products}}$$

Goods that can be produced using this particular set of inputs, except for output ω' and its downstream products

Unobserved Intermediate Goods Set: Theory IV

- For a plant j , if it uses inputs $\omega^1, \omega^2, \dots, \omega^k$ to produce output ω' , then,

$$\omega' \in \mathcal{D}(\omega^1), \omega' \in \mathcal{D}(\omega^2), \dots, \omega' \in \mathcal{D}(\omega^k)$$

$$\omega^1, \omega^2, \dots, \omega^k \in \mathcal{U}(\omega')$$

- And,

$$\underbrace{\left[\bigcap_{i=1}^k \mathcal{D}(\omega^i) \setminus [\{\omega'\} \cup \mathcal{D}(\omega')] \right] \cap \mathcal{U}(\omega')}$$

Goods that can be produced using this particular set of inputs, except output ω' and downstream products, that are used in production of ω'

Unobserved Intermediate Goods Set: Theory V

- For a plant j , if it uses inputs $\omega^1, \omega^2, \dots, \omega^k$ to produce output ω' , then,

$$UIGS_j := \mathcal{I}_j = \bigcup_{m=1}^M (\mathcal{I}_j^m) = \bigcup_{m=1}^M \left(\left[\bigcap_{i=1}^{k^{(m-1)}} \mathcal{D}(\omega^i) \setminus [\{\omega'\} \cup \mathcal{D}(\omega')]\right] \cap \mathcal{U}(\omega') \right)$$

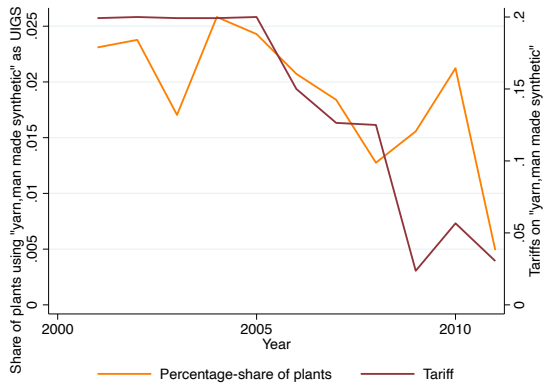
- Where $k^{(0)} = k$, $|\mathcal{I}_j^1| = k^{(1)}$, and $|\mathcal{I}_j^m| = k^{(m)}$
- Start with reported inputs, $\omega^1, \omega^2, \dots, \omega^k$, for $m = 1$ and get \mathcal{I}_j^1 , then, for $m = 2$, consider products in \mathcal{I}_j^1 as inputs, and get \mathcal{I}_j^2, \dots
- Continue this iterative process till $m = M$, if $\mathcal{I}_j^{M+1} = \emptyset$

UIGS Example: India-Made Petrol Car Manufacturing Plant in Panch Mahal District (Gujarat) in 2010

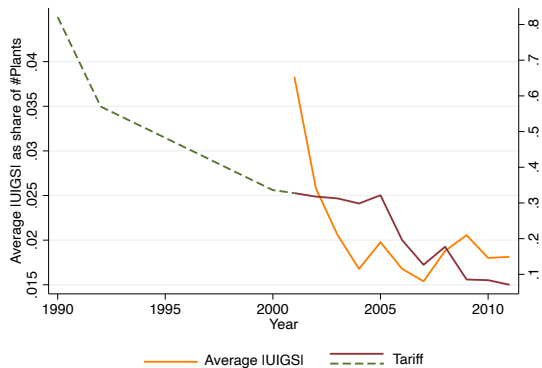
Reported Inputs
colour, ink, artist ink, paint, n.e.c
tyre/tube, others
seat cushion, rubber
rubber & rubber product, n.e.c
iron/steel primary forms, n.e.c
frame, steel - not for doors/windows
engine
general purpose machinery/tools, components, n.e.c
converter
batteries
modules
motor vehicle, others & parts, n.e.c

UIGS
air conditioner
body for motor car
clutch
compressors
cylinder head & block
electronic controller
engine assembly
gear
gear boxes
m.s. bars & rods, angles/- plates/square
meter assembly
wheels

UIGS and Tariffs



Correlation: 0.69



Correlation: 0.5

Empirical Strategy I:

- Total expenditure on product ω to produce output ω' , aggregating for all plants j' in all years t

$$\sum_t \sum_{j'} e_{\omega\omega't}^{j'}$$

- Total expenditure on all input-mix used to produce output ω' , aggregating for all plants j' in all years t

$$\sum_t \sum_{j'} \sum_{\omega} e_{\omega\omega't}^{j'}$$

- Then, **relative importance of product ω in production of output ω'** is given by:

$$z_{\omega\omega'} := \frac{\sum_t \sum_{j'} e_{\omega\omega't}^{j'}}{\sum_t \sum_{j'} \sum_{\omega} e_{\omega\omega't}^{j'}}$$

Empirical Strategy II:

$$\Delta_5 Y_{j\omega't} = \beta_0 + \beta_1 \cdot \sum_{\omega \in \mathcal{I}_{j(t-5)}} z_{\omega\omega'} \cdot \Delta_5 \log(1 + \tau_{\mathcal{I}\omega t}) + \xi_t + \varepsilon_{j\omega't}$$

- $Y_{j\omega't}$ is variable of interest for plant j , producing ω' , in year t
- $\tau_{\mathcal{I}\omega t}$ is tariff on product $\omega \in \mathcal{I}$ in the year t ;
- ξ_t is year fixed effect.
- $z_{\omega\omega'}$ relative importance share depends on input-output pair (and is time invariant)
- Set \mathcal{I} of UIGS varies at plant-year level

Results I: Vertical Specialization I

[Construct Vertical Span, à la Boehm and Oberfield (2023) approach]

► Construction

► Robustness checks

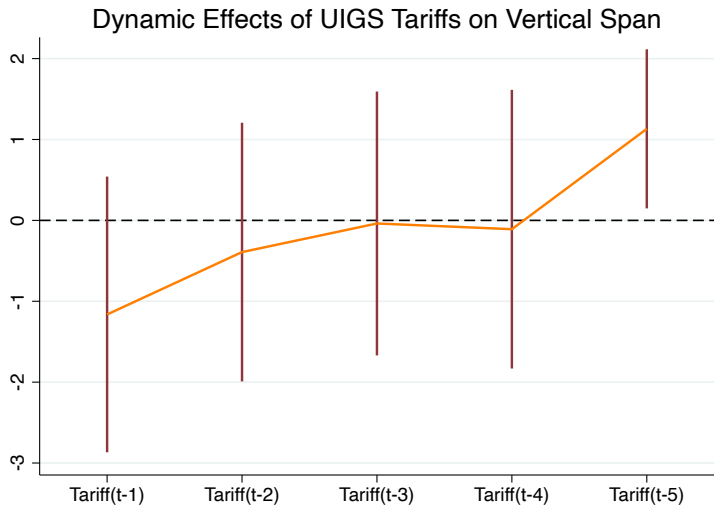
Dependent Variable →	(1) Δ_5 Vertical Span	(2) Δ_5 Vertical Span	(3) Δ_5 Vertical Span
$\sum_{\omega \in \mathcal{I}_{j(t-5)}} z_{\omega\omega'} \cdot \Delta_5 \log(1 + \tau_{\mathcal{I}\omega t})$	0.765*** (0.256)	0.935*** (0.278)	0.955*** (0.344)
Year FE	✓	✓	✓
UIGS Goods	ALL	Share > 1%	Share > 5%
R-squared	0.002	0.002	0.003
Observations	14,362	10,842	5,784

Robust standard errors in parentheses, clustered at the plant level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Results I: Vertical Specialization II

(Adjustments take time)



Results I: Vertical Specialization III

(Small v/s Large Plants)

Dependent Variable →	(1) Δ_5 Vertical Span	(2) Δ_5 Vertical Span
$\sum_{\omega \in \mathcal{I}_j(t-5)} z_{\omega\omega'} \cdot \Delta_5 \log(1 + \tau_{\mathcal{I}\omega t})$	1.134** (0.509)	0.788** (0.323)
Year FE	✓	✓
UIGS Goods	Share > 1%	Share > 1%
Plant Size	Small	Large
R-squared	0.003	0.003
Observations	4,983	5,859

Robust standard errors in parentheses, clustered at the plant level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Results II: Input Sourcing Decisions

► Substitution between Inputs

Dependent Variable →	(1) Log Inputs	(2) Log $\frac{\text{Domestic Inputs}}{\text{Output}}$
$\sum_{\omega \in \mathcal{I}_{j(t-5)}} z_{\omega\omega'} \cdot \Delta_5 \log(1 + \tau_{\mathcal{I}\omega t})$	-1.767*** (0.476)	0.240*** (0.069)
Year FE	✓	✓
UIGS Goods	Share > 1%	Share > 1%
R-squared	0.004	0.006
Observations	11,322	10,536

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Results III: Employment

Dependent Variable →	(1) Δ ₅ Log Managerial Staff	(2) Δ ₅ Log Factory Workers
$\sum_{\omega \in \mathcal{I}_{j(t-5)}} z_{\omega\omega'} \cdot \Delta_5 \log(1 + \tau_{\mathcal{I}\omega t})$	-2.195*** (0.575)	-1.583*** (0.523)
Year FE	✓	✓
UIGS Goods	Share > 1%	Share > 1%
R-squared	0.528	0.034
Observations	10,806	11,249

Robust standard errors in parentheses, clustered at the plant level.

*** p<0.01, ** p<0.05, * p<0.1

Results IV: Economic Performance I

Dependent Variable →	(1) Δ_5 Log Sales	(2) Δ_5 Log Profits	(3) Δ_5 Log Labor Productivity
$\sum_{\omega \in \mathcal{I}_{j(t-5)}} z_{\omega\omega'} \cdot \Delta_5 \log(1 + \tau_{\mathcal{I}\omega t})$	-2.181*** (0.499)	-3.849*** (0.899)	-1.503*** (0.457)
Year FE	✓	✓	✓
UIGS Goods	Share > 1%	Share > 1%	Share > 1%
R-squared	0.008	0.006	0.004
Observations	11,322	7,285	11,309

Robust standard errors in parentheses, clustered at the plant level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Results IV: Economic Performance II

Dependent Variable →	(1) $\Delta_5 \text{ Log Sales}$	(2) $\Delta_5 \text{ Log Sales}$
$\sum_{\omega \in \mathcal{I}_{j(t-5)}} z_{\omega\omega'} \cdot \Delta_5 \log(1 + \tau_{\mathcal{I} \omega t})$		-2.242*** (0.508)
$\Delta_5 \text{ Log Inputs Weighted Lagged Tariff}$	-1.034** (0.496)	-0.999** (0.497)
Year FE	✓	✓
UIGS Goods	–	Share > 1%
R-squared	0.003	0.005
Observations	9,871	9,871

Robust standard errors in parentheses, clustered at the plant level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Results IV: Economic Performance III

[Estimate Production Function using Levinsohn and Petrin (2003) method, with ACF (2015) correction]

► Methodology

Dependent Variable →	$\Delta_5 \widehat{TFP}$
$\sum_{\omega \in \mathcal{I}_{j(t-5)}} z_{\omega\omega'} \cdot \Delta_5 \log(1 + \tau_{\mathcal{I}\omega t})$	-0.116*** (0.034)
Year FE	✓
UIGS Goods	Share > 1%
R-squared	0.002
Observations	10,967

Robust standard errors in parentheses, clustered at the plant level.

*** p<0.01, ** p<0.05, * p<0.1

Conclusion

1. Designed an algorithm to construct Unobserved Intermediate Goods Set (UIGS)
 2. Developed a new measure for examining impact of trade policy
- **Main Result I:**
Decline in tariffs on UIGS increases vertical specialization, sourcing of imported inputs, employment, sales, profits, labor productivity, and TFP of manufacturing plants
 - **Main Result II:**
By not considering tariffs on UIGS, impact of trade policy is underestimated

Thanks!

For comments: smalik9@uh.edu

Vertical Span: Measure of Vertical Specialization

- Constructed at plant-level, à la *Boehm and Oberfield (2023)* approach:
 1. For each output-input pair $(\omega, \hat{\omega})$, construct **vertical distance**: $d_{\omega\hat{\omega}}$
 - Using information from **all plants producing ω from $\hat{\omega}$** , $\forall \hat{\omega}$
 2. For each plant j , construct **vertical span**:
No. of stages b/w inputs coming outside of plant to output going outside of plant

Vertical Span: Intuition for vertical distance $d_{\omega\hat{\omega}}$

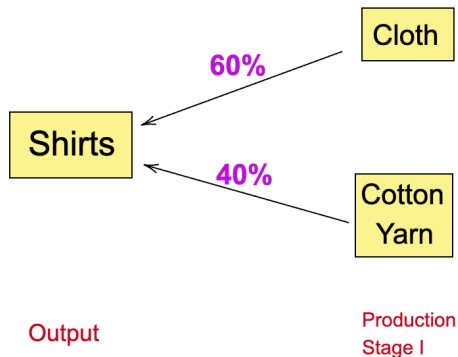


Shirts

Output

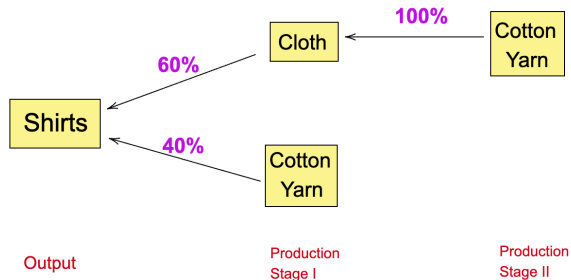
1. Consider **shirts** (ω)

Vertical Span: Intuition for vertical distance $d_{\omega\hat{\omega}}$



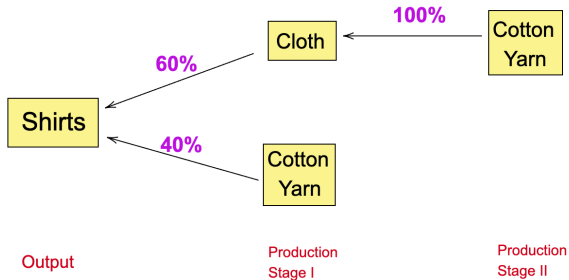
1. Consider **shirts** (ω)
2. Construct materials cost shares of ω for each input ($\hat{\omega}$)

Vertical Span: Intuition for vertical distance $d_{\omega\hat{\omega}}$



1. Consider **shirts** (ω)
2. Construct materials cost shares of ω for each input ($\hat{\omega}$)
3. Recursively construct cost shares of input, inputs' inputs, ...

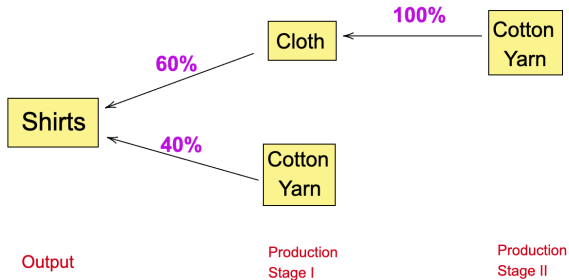
Vertical Span: Intuition for vertical distance $d_{\omega\hat{\omega}}$



1. Consider **shirts** (ω)
2. Construct materials cost shares of ω for each input ($\hat{\omega}$)
3. Recursively construct cost shares of input, inputs' inputs, ...
4. $d_{\omega\hat{\omega}}$ is avg. no. of steps between ω and $\hat{\omega}$, weighted by product of cost shares

- $d_{\text{Shirts} \leftarrow \text{Cloth}} = 1,$
- $d_{\text{Shirts} \leftarrow \text{Yarn}} = (0.4)1 + (0.6)(1.0)2 = 1.6$

Vertical Span: Intuition for vertical distance $d_{\omega\hat{\omega}}$



1. Consider **shirts** (ω)
2. Construct materials cost shares of ω for each input ($\hat{\omega}$)
3. Recursively construct cost shares of input, inputs' inputs, ...
4. $d_{\omega\hat{\omega}}$ is avg. no. of steps between ω and $\hat{\omega}$, weighted by product of cost shares

- $d_{\text{Shirts} \leftarrow \text{Cloth}} = 1,$
- $d_{\text{Shirts} \leftarrow \text{Yarn}} = (0.4)1 + (0.6)(1.0)2 = 1.6$

5. Repeat for 5236 products...

Vertical Span: How far are plant's inputs from output

► [back to results](#)

- For each plant j , construct **vertical span**:

$$\text{span}_j = \sum_{\hat{\omega}} \frac{X_{j\hat{\omega}}}{\sum_{\tilde{\omega}} X_{j\tilde{\omega}}} d_{\omega\hat{\omega}} \quad (1)$$

whereby,

- $X_{j\hat{\omega}}$ is input expenditure of plant j on product $\hat{\omega}$,
- $\sum_{\tilde{\omega}} X_{j\tilde{\omega}}$ is total input expenditure of plant j
- **Weighted average of distance of output from inputs**

Vertical Span: How far are plant's inputs from output

► [back to results](#)

- For each plant j , construct **vertical span**:

$$\text{span}_j = \sum_{\hat{\omega}} \frac{X_{j\hat{\omega}}}{\sum_{\tilde{\omega}} X_{j\tilde{\omega}}} d_{\omega\hat{\omega}} \quad (1)$$

whereby,

- $X_{j\hat{\omega}}$ is input expenditure of plant j on product $\hat{\omega}$,
 - $\sum_{\tilde{\omega}} X_{j\tilde{\omega}}$ is total input expenditure of plant j
 - **Weighted average of distance of output from inputs**
- **Eg:** If plant j (prod. **shirts**) spends **70%** and **30%** of input exp. on cloth and yarn resp.
- $$\text{span}_j = (0.7)(d_{\text{Shirts} \leftarrow \text{Cloth}}) + (0.3)(d_{\text{Shirts} \leftarrow \text{Yarn}}) = (0.7)(1) + (0.3)(1.6) = 1.18$$

Substitution between Imported and Domestic Inputs

$$\text{Log} \left(\frac{\text{Domestic Inputs}}{\text{Imported Inputs}} \right)_{j\hat{\omega}t} = \beta_0 + \beta_1 (\text{Log Input Tariffs})_{\hat{\omega}t} + \zeta_j + \xi_t + \varepsilon_{j\hat{\omega}t}$$

$$\text{Log} \left(\frac{\text{Domestic Inputs}}{\text{Imported Inputs}} \right)_{j\hat{\omega}t} = \beta_0 + \beta_1 (\text{Log Input Tariffs})_{\hat{\omega}(t-1)} + \zeta_j + \xi_t + \varepsilon_{j\hat{\omega}t}$$

- Plant j (producing output ω), using inputs $\hat{\omega}$, in year t

Substitution between Imported and Domestic Inputs

[▶ back to results](#)

Dependent Variable →	(1) $\text{Log} \left(\frac{\text{Domestic Inputs}}{\text{Imported Inputs}} \right)$	(2) $\text{Log} \left(\frac{\text{Domestic Inputs}}{\text{Imported Inputs}} \right)$
Log Input Tariffs	1.107*** (0.232)	
Log Input Tariffs _(t-1)		1.107*** (0.263)
Observations	35,622	33,196
R-squared	0.398	0.401
Plant FE	✓	✓
Year FE	✓	✓

Robust standard errors in parentheses, clustered at the plant level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$