Misallocation and Product Choice

Stepan Gordeev
UConn

Sudhir Singh
MSU, World Bank

Sep 29, 2023

Stockman

Misallocation and Product Choice

- **misallocation** is a big driver of cross-country income differences

MISALLOCATION AND PRODUCT CHOICE

- **misallocation** is a big driver of cross-country income differences
  - market frictions, distortionary policies → inputs not allocated to most productive uses

- **misallocation** is a big driver of cross-country income differences
  - market frictions, distortionary policies $\rightarrow$ inputs not allocated to most productive uses
- misallocation especially severe & costly in **agriculture** in low-income countries

MISALLOCATION AND PRODUCT CHOICE

- **misallocation** is a big driver of cross-country income differences
  - market frictions, distortionary policies $\rightarrow$ inputs not allocated to most productive uses
- misallocation especially severe & costly in **agriculture** in low-income countries
  - potential **benign** causes: mismeasurement, unobserved late-season shocks

- **misallocation** is a big driver of cross-country income differences
  - market frictions, distortionary policies → inputs not allocated to most productive uses
- misallocation especially severe & costly in **agriculture** in low-income countries
  - potential **benign** causes: mismeasurement, unobserved late-season shocks
  - potential **malign** causes: poor property rights, communal land distribution, min. prices

- **misallocation** is a big driver of cross-country income differences
  - market frictions, distortionary policies → inputs not allocated to most productive uses
- misallocation especially severe & costly in **agriculture** in low-income countries
  - potential **benign** causes: mismeasurement, unobserved late-season shocks
  - potential **malign** causes: poor property rights, communal land distribution, min. prices
- farms grow different crops, most grow multiple

MISALLOCATION AND PRODUCT CHOICE

- **misallocation** is a big driver of cross-country income differences
  - market frictions, distortionary policies $\rightarrow$ inputs not allocated to most productive uses
- misallocation especially severe & costly in **agriculture** in low-income countries
  - potential **benign** causes: mismeasurement, unobserved late-season shocks
  - potential **malign** causes: poor property rights, communal land distribution, min. prices
- farms grow different crops, most grow multiple
  - but misallocation literature uses **single-product** firms using the same production fn

- **misallocation** is a big driver of cross-country income differences
  - market frictions, distortionary policies → inputs not allocated to most productive uses
- misallocation especially severe & costly in **agriculture** in low-income countries
  - potential **benign** causes: mismeasurement, unobserved late-season shocks
  - potential **malign** causes: poor property rights, communal land distribution, min. prices
- farms grow different crops, most grow multiple
  - but misallocation literature uses **single-product** firms using the same production fn

<p align="center">how does heterogeneous product choice affect
aggregate misallocation cost?</p>

DATA

DATA

- Indian farm-crop-level survey

PREVIEW

DATA
- Indian farm-crop-level survey
- estimate crop-level production functions
  - crops are significantly heterogeneous in input intensities

PREVIEW

DATA

- Indian farm-crop-level survey
- estimate crop-level production functions
  - crops are significantly heterogeneous in input intensities

MODEL

PREVIEW

DATA
- Indian farm-crop-level survey
- estimate crop-level production functions
  − crops are significantly heterogeneous in input intensities

MODEL
- multi-product farms facing misallocative distortions

PREVIEW

DATA
- Indian farm-crop-level survey
- estimate crop-level production functions
  - crops are significantly heterogeneous in input intensities

MODEL
- multi-product farms facing misallocative distortions

COMPARED TO THE MULTI-PRODUCT MODEL, STANDARD 1-PRODUCT MODELS:

PREVIEW

DATA
- Indian farm-crop-level survey
- estimate crop-level production functions
  - crops are significantly heterogeneous in input intensities

MODEL
- multi-product farms facing misallocative distortions

COMPARED TO THE MULTI-PRODUCT MODEL, STANDARD 1-PRODUCT MODELS:
- overstate frictions in data

Preview

Data
- Indian farm-crop-level survey
- estimate crop-level production functions
  - crops are significantly heterogeneous in input intensities

Model

- multi-product farms facing misallocative distortions

Compared to the Multi-Product Model, Standard 1-Product Models:
- overstate frictions in data
  - → overstate misallocation if view most dispersion as **benign**

Preview

Data
- Indian farm-crop-level survey
- estimate crop-level production functions
  - crops are significantly heterogeneous in input intensities

Model

- multi-product farms facing misallocative distortions

Compared to the Multi-Product Model, Standard 1-Product Models:
- overstate frictions in data
  - → overstate misallocation if view most dispersion as **benign**
- understate firm expansion if frictions lifted

Preview

Data
- Indian farm-crop-level survey
- estimate crop-level production functions
  - crops are significantly heterogeneous in input intensities

Model

- multi-product farms facing misallocative distortions

Compared to the Multi-Product Model, Standard 1-Product Models:
- overstate frictions in data
  - → overstate misallocation if view most dispersion as **benign**
- understate firm expansion if frictions lifted
  - → understate misallocation if view most dispersion as **malign**

LITERATURE

- AGRICULTURAL MISALLOCATION

Literature

- **Agricultural Misallocation**
  - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing

LITERATURE

- AGRICULTURAL MISALLOCATION
  - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
  - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*

LITERATURE

- ## AGRICULTURAL MISALLOCATION
  - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
  - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
  - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*

LITERATURE

- ## AGRICULTURAL MISALLOCATION
  - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
  - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
  - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*
  - this paper:

LITERATURE

- ## AGRICULTURAL MISALLOCATION
    - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
    - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
    - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*
    - this paper:
        - > crop-specific production functions

LITERATURE

- ## AGRICULTURAL MISALLOCATION
    - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
    - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
    - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*
    - this paper:
        - > crop-specific production functions
        - > multi-crop farms adjusting crop choice to frictions

LITERATURE

- **AGRICULTURAL MISALLOCATION**
  - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
  - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
  - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*
  - this paper:
    - \> crop-specific production functions
    - \> multi-crop farms adjusting crop choice to frictions
- **MULTI-PRODUCT FIRMS & MISALLOCATION**

LITERATURE

- **AGRICULTURAL MISALLOCATION**
  - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
  - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
  - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*
  - this paper:
    - > crop-specific production functions
    - > multi-crop farms adjusting crop choice to frictions

- **MULTI-PRODUCT FIRMS & MISALLOCATION**
  - *Jaef (2018)*, *Wang, Yang (2021)*: manufacturers choose # of products with heterog. TFPs
    - → # of products responds to distortions

LITERATURE

- **AGRICULTURAL MISALLOCATION**
    - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
    - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
    - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*
    - this paper:
        - > crop-specific production functions
        - > multi-crop farms adjusting crop choice to frictions
- **MULTI-PRODUCT FIRMS & MISALLOCATION**
    - *Jaef (2018)*, *Wang, Yang (2021)*: manufacturers choose # of products with heterog. TFPs
        - → # of products responds to distortions
    - this paper:

LITERATURE

- **AGRICULTURAL MISALLOCATION**
  - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
  - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
  - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*
  - this paper:
    - > crop-specific production functions
    - > multi-crop farms adjusting crop choice to frictions
- **MULTI-PRODUCT FIRMS & MISALLOCATION**
  - *Jaef (2018)*, *Wang, Yang (2021)*: manufacturers choose # of products with heterog. TFPs
    - → # of products responds to distortions
  - this paper:
    - > heterogeneous productivities and production functions

LITERATURE

- **AGRICULTURAL MISALLOCATION**
  - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
  - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
  - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*
  - this paper:
    - > crop-specific production functions
    - > multi-crop farms adjusting crop choice to frictions
- **MULTI-PRODUCT FIRMS & MISALLOCATION**
  - *Jaef (2018)*, *Wang, Yang (2021)*: manufacturers choose # of products with heterog. TFPs
    - → # of products responds to distortions
  - this paper:
    - > heterogeneous productivities and production functions
- **agriculture** is perfect to explore interaction of **product choice** and **misallocation**
  - ▶ *why*

Literature

- **Agricultural Misallocation**
  - base framework: *Chen, Restuccia, Santaeulalia-Llopis (2022)* building on *Hsieh, Klenow (2009)* for manufacturing
  - measure misallocation from specific frictions: *Chen (2017)*, *Gottlieb and Grobovsek (2019)*
  - identify overall misallocation from general frictions: *Gollin, Udry (2021)*, *Aragon, Restuccia, Rud (2022)*, *Adamopoulos, Brandt, Leight, Restuccia (2022)*, *Ayerst, Brandt, Restuccia (2023)*
  - this paper:
    - > crop-specific production functions
    - > multi-crop farms adjusting crop choice to frictions
- **Multi-Product Firms & Misallocation**
  - *Jaef (2018)*, *Wang, Yang (2021)*: manufacturers choose # of products with heterog. TFPs
    - → # of products responds to distortions
  - this paper:
    - > heterogeneous productivities and production functions
- **agriculture** is perfect to explore interaction of **product choice** and **misallocation**
  - ▶ *why*
  - but mechanisms apply equally to other sectors

Data

- **India**'s Rural Economic and Demographic Survey (REDS)
  - nationally representative of rural Indian households

- **India**'s Rural Economic and Demographic Survey (REDS)
  - nationally representative of rural Indian households
  - use 2007-08 round, only with crop-plot-level data

- **India**'s Rural Economic and Demographic Survey (REDS)
  - nationally representative of rural Indian households
  - use 2007-08 round, only with crop-plot-level data
- 10,318 plots, 4,803 farmers

- **India**'s Rural Economic and Demographic Survey (REDS)
  - nationally representative of rural Indian households
  - use 2007-08 round, only with crop-plot-level data
- 10,318 plots, 4,803 farmers
  - plot-crop-level inputs, outputs

- **India**'s Rural Economic and Demographic Survey (REDS)
  - nationally representative of rural Indian households
  - use 2007-08 round, only with crop-plot-level data
- 10,318 plots, 4,803 farmers
  - plot-crop-level inputs, outputs
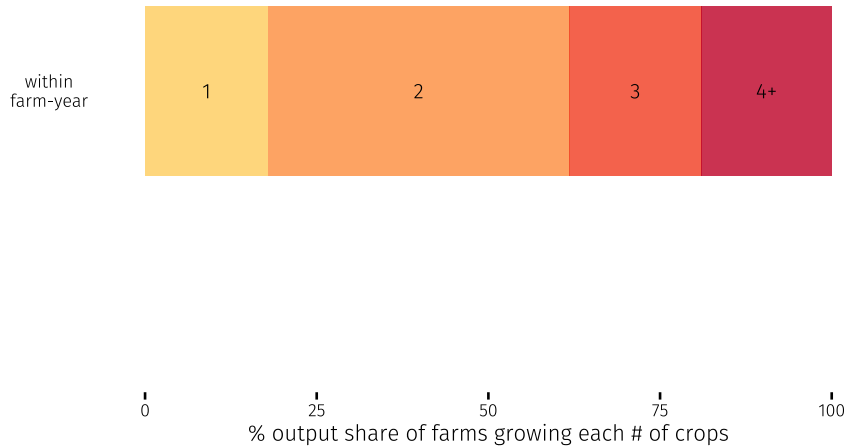 - plot-level physical characteristics

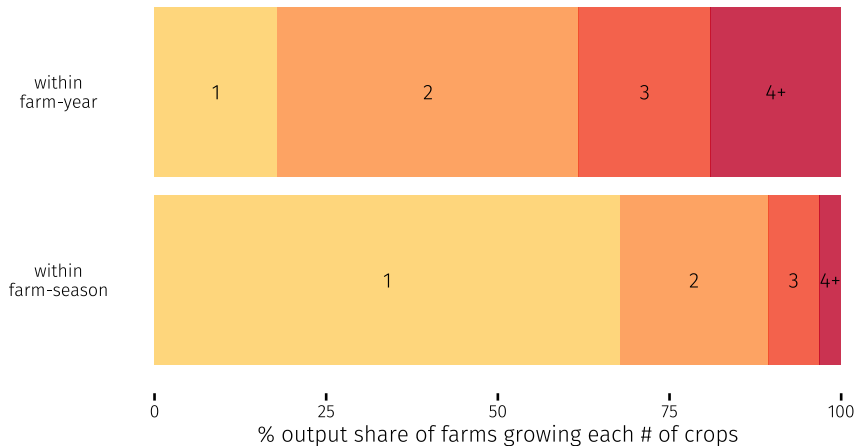# Multi-Product Farms in India

CROP CHOICE IS HETEROGENEOUS
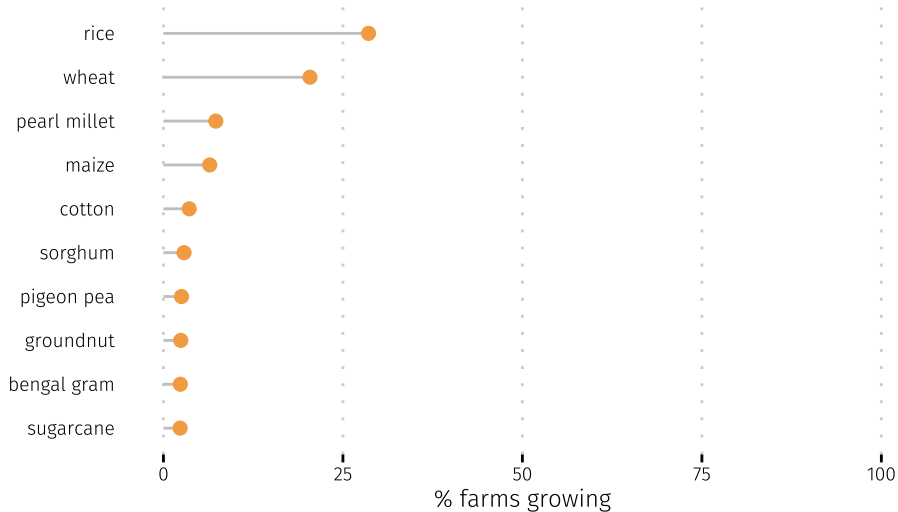
% farms growing

MANY FARMS GROW MULTIPLE CROPS, MAINLY ACROSS SEASONS

within farm-year

1 2 3 4+

% output share of farms growing each # of crops

Many Farms Grow Multiple Crops, Mainly Across Seasons

within farm-year: 1 | 2 | 3 | 4+

within farm-season: 1 | 2 | 3 | 4+

% output share of farms growing each # of crops

- 3 agricultural seasons: Kharif (monsoon), Rabi (winter/spring), Zaid (summer/dry)

Production Functions

ESTIMATE CROP-SPECIFIC PRODUCTION FUNCTIONS

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

Estimate Crop-Specific Production Functions

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

  - for farm f, crop i, season t

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

  — for farm f, crop i, season t
  — y = market value of physical output

ESTIMATE CROP-SPECIFIC PRODUCTION FUNCTIONS

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

  — for farm f, crop i, season t
  — y = market value of physical output
  — l = land input (quality-adjusted)

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

   - for farm f, crop i, season t
   - y = market value of physical output
   - l = land input (quality-adjusted)
      > **common measures**: *land area* ignores quality, *land price* is noisy

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

  - for farm f, crop i, season t
  - y = market value of physical output
  - l = land input (quality-adjusted)
    - **common measures**: *land area* ignores quality, *land price* is noisy
    - **instead**: use random forests to *predict* land price using physical land features
      - *quality index construction*

ESTIMATE CROP-SPECIFIC PRODUCTION FUNCTIONS

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

- for farm f, crop i, season t
- y = market value of physical output
- l = land input (quality-adjusted)
  - **common measures**: *land area* ignores quality, *land price* is noisy
  - **instead**: use random forests to *predict* land price using physical land features
    - *quality index construction*
- x_{labor} = labor input (days)

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

  - for farm f, crop i, season t
  - y = market value of physical output
  - l = land input (quality-adjusted)
    - > **common measures**: *land area* ignores quality, *land price* is noisy
    - > **instead**: use random forests to *predict* land price using physical land features
      - ▶ *quality index construction*
  - x_{labor} = labor input (days)
  - x_{inter} = intermediate inputs (seeds, fertilizer, etc.)
    - ▶ *inputs, output*

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

  - for farm f, crop i, season t
  - y = market value of physical output
  - l = land input (quality-adjusted)
    - > **common measures**: *land area* ignores quality, *land price* is noisy
    - > **instead**: use random forests to *predict* land price using physical land features
      - ▶ *quality index construction*
  - x_{labor} = labor input (days)
  - x_{inter} = intermediate inputs (seeds, fertilizer, etc.)
    - ▶ *inputs, output*
- merge crops into **5 groups** ▶ *list*

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

  - for farm f, crop i, season t
  - y = market value of physical output
  - l = land input (quality-adjusted)
    - > **common measures**: *land area* ignores quality, *land price* is noisy
    - > **instead**: use random forests to *predict* land price using physical land features
      - ▶ *quality index construction*
  - x_{labor} = labor input (days)
  - x_{inter} = intermediate inputs (seeds, fertilizer, etc.)
    - ▶ *inputs, output*
- merge crops into **5 groups** ▶ *list*
- 2SLS at plot-level using **instruments** for inputs adapting *Gollin, Udry (2021)*'s method

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$

  — for farm f, crop i, season t
  — y = market value of physical output
  — l = land input (quality-adjusted)
    > **common measures**: *land area* ignores quality, *land price* is noisy
    > **instead**: use random forests to *predict* land price using physical land features
      ▶ *quality index construction*
  — x_{labor} = labor input (days)
  — x_{inter} = intermediate inputs (seeds, fertilizer, etc.)
    ▶ *inputs, output*
- merge crops into **5 groups** ▶ *list*
- 2SLS at plot-level using **instruments** for inputs adapting *Gollin, Udry (2021)*'s method
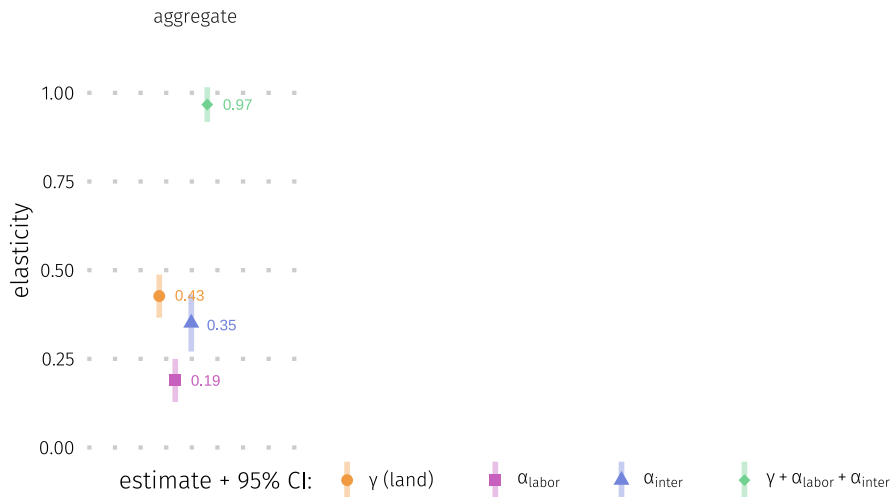 — **idea**: shocks to farm f's plots $k \neq j$ change shadow price of inputs on f's plot j

- production function:

$$y_{f,i,t} = z_{f,i,t} l_{f,i,t}^{\gamma_i} x_{labor,f,i,t}^{\alpha_{labor,i}} x_{inter,f,i,t}^{\alpha_{inter,i}}$$
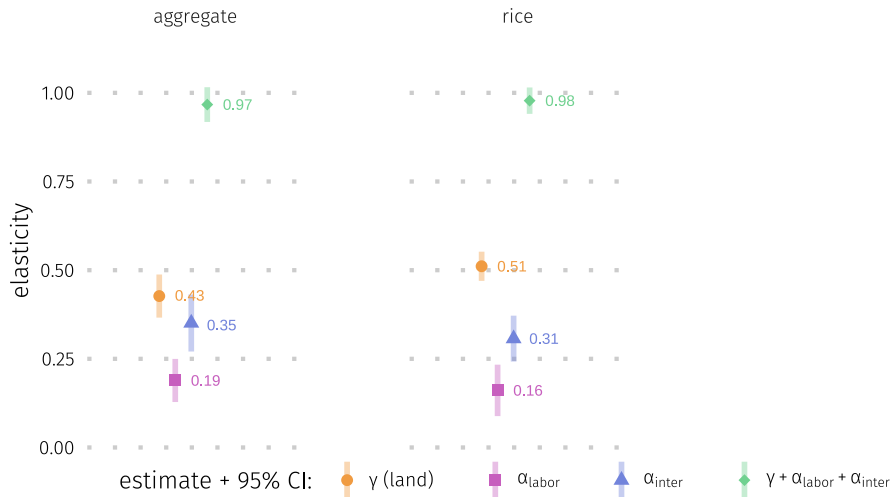
  - for farm f, crop i, season t
  - y = market value of physical output
  - l = land input (quality-adjusted)
    - > **common measures**: *land area* ignores quality, *land price* is noisy
    - > **instead**: use random forests to *predict* land price using physical land features
      - ▶ *quality index construction*
  - x_{labor} = labor input (days)
  - x_{inter} = intermediate inputs (seeds, fertilizer, etc.)
    - ▶ *inputs, output*
- merge crops into **5 groups** ▶ *list*
- 2SLS at plot-level using **instruments** for inputs adapting *Gollin, Udry (2021)*'s method
  - **idea**: shocks to farm f's plots $k \neq j$ change shadow price of inputs on f's plot j
  ▶ *details*

Production Functions are Heterogeneous across Crops

aggregate

elasticity

1.00 — ● 0.97

0.75

0.50 — 0.43

0.25 — 0.35

0.19

0.00

estimate + 95% CI: ● γ (land)   ■ α_{labor} ▲ α_{inter} ◆ $\gamma + \alpha_{labor} + \alpha_{inter}$

▶ *all crops* ▶ *table* ▶ *equality tests*

Production Functions are Heterogeneous across Crops

estimate + 95% CI: ● γ (land) ■ α_labor ▲ α_inter ◆ γ + α_labor + α_inter

▶ *all crops* ▶ *table* ▶ *equality tests*

PRODUCTION FUNCTIONS ARE HETEROGENEOUS ACROSS CROPS
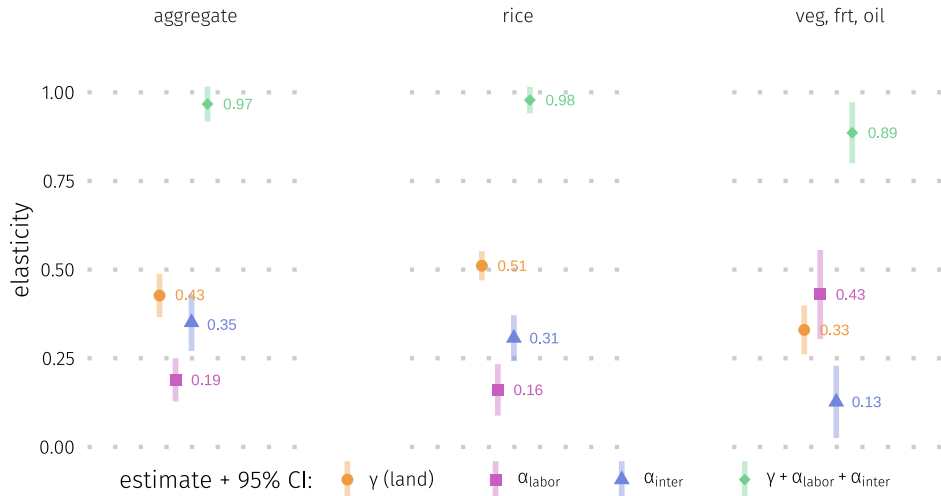
Model

MODEL OBJECTIVES

- OBJECTIVES:

Model Objectives

- OBJECTIVES:
  - model multi-product farm decisions in presence of distortions

Model Objectives

- OBJECTIVES:
    - model multi-product farm decisions in presence of distortions
    - provide a mapping from observable outcomes to unobserved distortions

- OBJECTIVES:
    - model multi-product farm decisions in presence of distortions
    - provide a mapping from observable outcomes to unobserved distortions
    - quantify the aggregate output cost of misallocation induced by distortions

- OBJECTIVES:
    - model multi-product farm decisions in presence of distortions
    - provide a mapping from observable outcomes to unobserved distortions
    - quantify the aggregate output cost of misallocation induced by distortions
- build on models of single-product firm-level misallocation

- OBJECTIVES:
  - model multi-product farm decisions in presence of distortions
  - provide a mapping from observable outcomes to unobserved distortions
  - quantify the aggregate output cost of misallocation induced by distortions
- build on models of single-product firm-level misallocation
  - *Hsieh, Klenow (2009)*: misallocation in **manufacturing**

- OBJECTIVES:
  - model multi-product farm decisions in presence of distortions
  - provide a mapping from observable outcomes to unobserved distortions
  - quantify the aggregate output cost of misallocation induced by distortions
- build on models of single-product firm-level misallocation
  - *Hsieh, Klenow (2009)*: misallocation in **manufacturing**
  - *Chen, Restuccia, Santaeulalia-Llopis (2022)*: misallocation in **agriculture**

Model Objectives

- **objectives:**
  - model multi-product farm decisions in presence of distortions
  - provide a mapping from observable outcomes to unobserved distortions
  - quantify the aggregate output cost of misallocation induced by distortions
- build on models of single-product firm-level misallocation
  - *Hsieh, Klenow (2009)*: misallocation in **manufacturing**
  - *Chen, Restuccia, Santaeulalia-Llopis (2022)*: misallocation in **agriculture**
- add multi-product firms choosing among heterogeneous crops

FARM: PRODUCTION

$$\max \quad \underbrace{p \overbrace{z_f \, l_f^{\ \gamma} \, \Pi_g \left(x_{f,g}^{\ \ \alpha_g} \right)}^{y_f}}_{\text{revenue}} \quad - \quad \underbrace{\sum_{g=1}^{G} r_g \qquad x_{f,g}}_{\text{flex. input costs}}$$

$$\text{s.t.} \qquad l_f \quad = L_f \qquad (\lambda_f)$$

- profit-maximizing farm f: sells output $p y_f$, pays for inputs

FARM: PRODUCTION

$$\max \quad \underbrace{p \, \overbrace{z_f \, l_f{}^{\gamma} \, \Pi_g \left(x_{f,g}{}^{\alpha_g} \right)}^{y_f}}_{\text{revenue}} \quad - \quad \underbrace{\sum_{g=1}^{G} r_g \qquad x_{f,g}}_{\text{flex. input costs}}$$

$$\text{s.t.} \quad l_f \quad = L_f \qquad (\lambda_f)$$

- profit-maximizing farm f: sells output py_f, pays for inputs
- Cobb-Douglas production function with TFP z_f

FARM: PRODUCTION

$$\max \quad \underbrace{p \, \overbrace{z_f \, l_f^{\,\gamma} \, \Pi_g \left(x_{f,g}^{\,\alpha_g} \right)}^{y_f}}_{\text{revenue}} \quad - \quad \underbrace{\sum_{g=1}^{G} r_g \qquad x_{f,g}}_{\text{flex. input costs}}$$

$$\text{s.t.} \quad l_f = L_f \quad (\lambda_f)$$

- profit-maximizing farm f: sells output py_f, pays for inputs
- Cobb-Douglas production function with TFP z_f
- flexible inputs g: labor, intermediates
  - quantity $x_{f,g}$ rented at r_g

$$\max \quad \underbrace{p \overbrace{z_f \, l_f^{\,\gamma} \, \Pi_g \left(x_{f,g}^{\,\alpha_g} \right)}^{y_f}}_{\text{revenue}} \quad - \quad \underbrace{\sum_{g=1}^{G} r_g \qquad x_{f,g}}_{\text{flex. input costs}}$$

$$\text{s.t.} \quad l_f \quad = L_f \qquad (\lambda_f)$$

- profit-maximizing farm f: sells output py_f, pays for inputs
- Cobb-Douglas production function with TFP z_f
- flexible inputs g: labor, intermediates
  - quantity $x_{f,g}$ rented at r_g
- land input l is in fixed supply L_f
  - almost no land market in India

$$\max \quad p \overbrace{z_f \, l_f{}^{\gamma} \, \Pi_g \left(x_{f,g}{}^{\alpha_g} \right)}^{y_f} \quad - \quad \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \qquad x_{f,g}}_{\text{flex. input costs}}$$

$$\underbrace{\phantom{p z_f l_f{}^{\gamma} \Pi_g (x_{f,g}{}^{\alpha_g})}}_{\text{revenue}}$$

$$\text{s.t.} \qquad l_f \quad = L_f \qquad (\lambda_f)$$

- farm-input distortions $\tau_{f,g}$ capture misallocative frictions
  - represented with tax ($\tau_{f,g} > 1$) or subsidy ($\tau_{f,g} < 1$) idiosyncratic to farm f, input g

FARM: DISTORTIONS

$$\max \quad p \underbrace{\overbrace{z_f \, l_f{}^{\gamma} \, \Pi_g \left(x_{f,g}{}^{\alpha_g} \right)}^{y_f}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \quad x_{f,g}}_{\text{flex. input costs}}$$

$$\text{s.t.} \quad l_f = L_f \quad (\lambda_f)$$

- farm-input distortions $\tau_{f,g}$ capture misallocative frictions
  - represented with tax ($\tau_{f,g} > 1$) or subsidy ($\tau_{f,g} < 1$) idiosyncratic to farm f, input g
- **between inputs:** $\tau_{f,labor} > \tau_{f,inter} \rightarrow$ farm f uses "too little" labor

FARM: DISTORTIONS

$$\max \quad \underbrace{p \overbrace{z_f \, l_f^{\,\gamma} \, \Pi_g \left(x_{f,g}^{\,\alpha_g} \right)}^{y_f}}_{\text{revenue}} \quad - \quad \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \qquad x_{f,g}}_{\text{flex. input costs}}$$

$$\text{s.t.} \qquad l_f \qquad = L_f \qquad (\lambda_f)$$

- farm-input distortions $\tau_{f,g}$ capture misallocative frictions
  - represented with tax ($\tau_{f,g} > 1$) or subsidy ($\tau_{f,g} < 1$) idiosyncratic to farm f, input g
- **between inputs:** $\tau_{f,labor} > \tau_{f,inter} \rightarrow$ farm f uses "too little" labor
- **between farms:** $\tau_{1,g} > \tau_{2,g} \; \forall g \rightarrow$ farm 1 is "too small" given its TFP z_f

FARM: DISTORTIONS

$$\max \quad p \overbrace{z_f \, l_f^{\;\gamma} \, \Pi_g \left(x_{f,g}^{\;\alpha_g} \right)}^{y_f} \; - \; \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \quad x_{f,g}}_{\text{flex. input costs}}$$

$$\underbrace{\phantom{p \, z_f \, l_f^{\;\gamma} \, \Pi_g \left(x_{f,g}^{\;\alpha_g} \right)}}_{\text{revenue}}$$

$$\text{s.t.} \quad l_f \quad = L_f \quad (\lambda_f)$$

- farm-input distortions $\tau_{f,g}$ capture misallocative frictions
  - represented with tax ($\tau_{f,g} > 1$) or subsidy ($\tau_{f,g} < 1$) idiosyncratic to farm f, input g
- **between inputs:** $\tau_{f,labor} > \tau_{f,inter} \rightarrow$ farm f uses "too little" labor
- **between farms:** $\tau_{1,g} > \tau_{2,g} \; \forall g \rightarrow$ farm 1 is "too small" given its TFP z_f
- L_f fixed $\rightarrow$ land is also distorted unless distributed to equalize λ_f
  - e.g. lacking property rights, communal land distribution

$$\max \quad \underbrace{p \, \overbrace{z_f \, l_f^{\,\gamma} \, \Pi_g \left(x_{f,g}^{\,\alpha_g} \right)}^{y_f}}_{\text{revenue}} \quad - \quad \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \qquad x_{f,g}}_{\text{flex. input costs}}$$

$$\text{s.t.} \qquad l_f \quad = L_f \qquad (\lambda_f)$$

- farm-input distortions $\tau_{f,g}$ capture misallocative frictions
  - represented with tax ($\tau_{f,g} > 1$) or subsidy ($\tau_{f,g} < 1$) idiosyncratic to farm f, input g
- **between inputs:** $\tau_{f,labor} > \tau_{f,inter}$ $\rightarrow$ farm f uses "too little" labor
- **between farms:** $\tau_{1,g} > \tau_{2,g} \; \forall g$ $\rightarrow$ farm 1 is "too small" given its TFP z_f
- L_f fixed $\rightarrow$ land is also distorted unless distributed to equalize λ_f
  - e.g. lacking property rights, communal land distribution
- distortions extracted from observed input, output choices
  - rationalize all heterogeneity in data ▸ *details*

FARM: MULTIPLE PRODUCTS

$$\max \underbrace{\sum_{i=1}^{N} p_i \overbrace{z_{f,i} l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}^{y_{f,i}}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

- heterogeneous crops $i = 1 \ldots N$

FARM: MULTIPLE PRODUCTS

$$\max \sum_{i=1}^{N} p_i \overbrace{z_{f,i} l_{f,i}{}^{\gamma_i} \Pi_g \left(x_{f,g,i}{}^{\alpha_{g,i}} \right)}^{y_{f,i}} \underbrace{}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

- heterogeneous crops $i = 1 \ldots N$
- l in fixed supply $L_f \rightarrow$ interdependent crop production
  - params of crop i change $\rightarrow \lambda_f$ changes $\rightarrow$ inputs and outputs of crops $-i$ change
  - *Just, Zilberman, and Hochman (1983)*, *Shumway, Pope, Nash (1984)*
  - justifies *Gollin, Udry (2021)* prod. fn. identification

FARM: MULTIPLE PRODUCTS

$$\max \underbrace{\sum_{i=1}^{N} p_i \overbrace{z_{f,i} l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}^{y_{f,i}}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

· heterogeneous crops $i = 1 \ldots N$

· l in fixed supply $L_f \rightarrow$ interdependent crop production
  - params of crop i change $\rightarrow \lambda_f$ changes $\rightarrow$ inputs and outputs of crops $-i$ change
  - *Just, Zilberman, and Hochman (1983)*, *Shumway, Pope, Nash (1984)*
  - justifies *Gollin, Udry (2021)* prod. fn. identification

· farm-input-crop distortions $\tau_{f,g,i}$, $\tau_{f,l,i}$
  - $\rightarrow$ fit observed input ratio heterogeneity across crops within a farm

FARM: FIXED COST

$$\max \underbrace{\sum_{i=1}^{N} p_i z_{f,i} l_{f,i}{}^{\gamma_i} \overbrace{\Pi_g \left(x_{f,g,i}{}^{\alpha_{g,i}} \right)}^{y_{f,i}}}_{\text{revenue}} \;-\; \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}} \;-\; \underbrace{\sum_{i=1}^{N} \omega \cdot \mathbb{1}[y_{f,i} > 0]}_{\text{fixed cost per crop}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

- fixed cost ω per produced crop
  - $\rightarrow$ farms choose **crop set** in addition to **crop mix**
  - $-$ farms don't all produce everything
  - $\rightarrow$ fit observed heterogeneity in crop sets

FARM: CONCAVITY

$$\max \underbrace{\sum_{i=1}^{N}\left(p_i \overbrace{z_{f,i}l_{f,i}{}^{\gamma_i}\Pi_g\left(x_{f,g,i}{}^{\alpha_{g,i}}\right)}^{y_{f,i}}\right)^{\eta}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g\tau_{f,g}\sum_{i=1}^{N}\tau_{f,g,i}x_{f,g,i}}_{\text{flex. input costs}} - \underbrace{\sum_{i=1}^{N}\omega \cdot 1[y_{f,i} > 0]}_{\text{fixed cost per crop}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i}\tau_{f,l,i} = L_f \qquad (\lambda_f)$$

· crop-level concavity term $\eta < 1$

FARM: CONCAVITY

$$\max \underbrace{\sum_{i=1}^{N}\Big(p_i \overbrace{z_{f,i} l_{f,i}^{\gamma_i} \Pi_g \big(x_{f,g,i}^{\alpha_{g,i}}\big)}^{y_{f,i}}\Big)^{\eta}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}} - \underbrace{\sum_{i=1}^{N} \omega \cdot \mathbb{1}[y_{f,i} > 0]}_{\text{fixed cost per crop}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

· crop-level concavity term $\eta < 1$
· captures several motives for farmers to mix crops beyond prod. fn. DRS

$$\max \underbrace{\sum_{i=1}^{N} \left(p_i \overbrace{z_{f,i} l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}^{y_{f,i}} \right)^{\eta}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}} - \underbrace{\sum_{i=1}^{N} \omega \cdot 1[y_{f,i} > 0]}_{\text{fixed cost per crop}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

- crop-level concavity term $\eta < 1$
- captures several motives for farmers to mix crops beyond prod. fn. DRS
    - risk

$$\max \underbrace{\sum_{i=1}^{N} \left(p_i \overbrace{z_{f,i} l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}^{y_{f,i}} \right)^{\eta}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}} - \underbrace{\sum_{i=1}^{N} \omega \cdot 1[y_{f,i} > 0]}_{\text{fixed cost per crop}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

- crop-level concavity term $\eta < 1$
- captures several motives for farmers to mix crops beyond prod. fn. DRS
  - risk
  - subsistence + love of variety

FARM: CONCAVITY

$$\max \underbrace{\sum_{i=1}^{N} \left(p_i \overbrace{z_{f,i} l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}^{y_{f,i}} \right)^{\eta}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}} - \underbrace{\sum_{i=1}^{N} \omega \cdot 1[y_{f,i} > 0]}_{\text{fixed cost per crop}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

- crop-level concavity term $\eta < 1$
- captures several motives for farmers to mix crops beyond prod. fn. DRS
  - risk
  - subsistence + love of variety
  - market power

$$\max \underbrace{\sum_{i=1}^{N} \left(p_i \overbrace{z_{f,i} l_{f,i}{}^{\gamma_i} \Pi_g \left(x_{f,g,i}{}^{\alpha_{g,i}} \right)}^{y_{f,i}} \right)^{\eta}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}} - \underbrace{\sum_{i=1}^{N} \omega \cdot 1[y_{f,i} > 0]}_{\text{fixed cost per crop}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

- crop-level concavity term $\eta < 1$
- captures several motives for farmers to mix crops beyond prod. fn. DRS
  - risk
  - subsistence + love of variety
  - market power
  - farm's problem with η equivalent to subsistence or market power ▶ *details*

FARM: CONCAVITY

$$\max \underbrace{\sum_{i=1}^{N} \left(p_i \overbrace{z_{f,i} l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}^{y_{f,i}} \right)^{\eta}}_{\text{revenue}} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}} - \underbrace{\sum_{i=1}^{N} \omega \cdot 1[y_{f,i} > 0]}_{\text{fixed cost per crop}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

- crop-level concavity term $\eta < 1$
- captures several motives for farmers to mix crops beyond prod. fn. DRS
  - risk
  - subsistence + love of variety
  - market power
  - farm's problem with η equivalent to subsistence or market power ▶ *details*
- calibrated to minimize distortions needed to explain data ▶ *details*

$$\max \sum_{i=1}^{N} \Big(p_i \overbrace{z_{f,i} l_{f,i}{}^{\gamma_i} \Pi_g \big(x_{f,g,i}{}^{\alpha_{g,i}} \big)}^{y_{f,i}} \Big)^{\eta} - \underbrace{\sum_{g=1}^{G} r_g \tau_{f,g} \sum_{i=1}^{N} \tau_{f,g,i} x_{f,g,i}}_{\text{flex. input costs}} - \underbrace{\sum_{i=1}^{N} \omega \cdot 1[y_{f,i} > 0]}_{\text{fixed cost per crop}}$$

$$\underbrace{\phantom{\sum_{i=1}^{N} \Big(p_i z_{f,i} l_{f,i}{}^{\gamma_i} \Pi_g \big(x_{f,g,i}{}^{\alpha_{g,i}} \big) \Big)^{\eta}}}_{\text{revenue}}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} l_{f,i} \tau_{f,l,i} = L_f \qquad (\lambda_f)$$

- crop-level concavity term $\eta < 1$
- captures several motives for farmers to mix crops beyond prod. fn. DRS
  - risk
  - subsistence + love of variety
  - market power
  - farm's problem with η equivalent to subsistence or market power ▶ *details*
- calibrated to minimize distortions needed to explain data ▶ *details*

▶ *solution* ▶ *GE*

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
    - equalize (or reduce) distortions between farms

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
  - equalize (or reduce) distortions between farms
  - prohibit product switching to limit assumptions

REALLOCATION

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
  - equalize (or reduce) distortions between farms
  - prohibit product switching to limit assumptions
  - treat each season as a separate economy

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
  – equalize (or reduce) distortions between farms
  – prohibit product switching to limit assumptions
  – treat each season as a separate economy
- compute counterfactual Δoutput: **reallocation gain**

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
  - equalize (or reduce) distortions between farms
  - prohibit product switching to limit assumptions
  - treat each season as a separate economy
- compute counterfactual Δoutput: **reallocation gain**
  - reflects aggr. TFP gain from better input allocation between farms

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
  - equalize (or reduce) distortions between farms
  - prohibit product switching to limit assumptions
  - treat each season as a separate economy
- compute counterfactual Δoutput: **reallocation gain**
  - reflects aggr. TFP gain from better input allocation between farms
  - captures **misallocation cost**

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
  - equalize (or reduce) distortions between farms
  - prohibit product switching to limit assumptions
  - treat each season as a separate economy
- compute counterfactual Δoutput: **reallocation gain**
  - reflects aggr. TFP gain from better input allocation between farms
  - captures **misallocation cost**
- compare **reallocation gain** between multi-product model and 1-product model
  ▶ *details*

Benchmark Exercise

reallocation gain

BENCHMARK EXERCISE

- equalize cross-farm distortions $\lambda_f, \tau_{f,g}$

reallocation gain

Benchmark Exercise

- equalize cross-farm distortions λ_f, $\tau_{f,g}$
- keep within-farm distortions $\tau_{f,l,i}$, $\tau_{f,g,i}$

reallocation gain

Benchmark Exercise
- equalize cross-farm distortions λ_f, $\tau_{f,g}$
- keep within-farm distortions $\tau_{f,l,i}$, $\tau_{f,g,i}$
  - preserve idiosyncratic product choice motives

reallocation gain

EXERCISES

BENCHMARK EXERCISE
- equalize cross-farm distortions $\lambda_f, \tau_{f,g}$
- keep within-farm distortions $\tau_{f,l,i}, \tau_{f,g,i}$
  - preserve idiosyncratic product choice motives

reallocation gain

multi-product: **294**%
1-product: **212**% (28% ↓)

BENCHMARK EXERCISE

- equalize cross-farm distortions $\lambda_f, \tau_{f,g}$
- keep within-farm distortions $\tau_{f,l,i}, \tau_{f,g,i}$
  - preserve idiosyncratic product choice motives

MATCH LEAST-DISTORTED STATE

reallocation gain

multi-product: **294%**
   1-product: **212%**   (28% ↓)

BENCHMARK EXERCISE

- equalize cross-farm distortions $\lambda_f, \tau_{f,g}$
- keep within-farm distortions $\tau_{f,l,i}, \tau_{f,g,i}$
  - preserve idiosyncratic product choice motives

MATCH LEAST-DISTORTED STATE

- a more conservative and practical estimate

reallocation gain

multi-product: **294**%
1-product: **212**% (28% ↓)

EXERCISES

BENCHMARK EXERCISE

- equalize cross-farm distortions $\lambda_f, \tau_{f,g}$
- keep within-farm distortions $\tau_{f,l,i}, \tau_{f,g,i}$
  - preserve idiosyncratic product choice motives

MATCH LEAST-DISTORTED STATE

- a more conservative and practical estimate
- downscale frictions s.t. their variances match those in the least-distorted state

reallocation gain

multi-product: **294%**
1-product: **212%** (28% ↓)

BENCHMARK EXERCISE

- equalize cross-farm distortions $\lambda_f, \tau_{f,g}$
- keep within-farm distortions $\tau_{f,l,i}, \tau_{f,g,i}$
  - preserve idiosyncratic product choice motives

MATCH LEAST-DISTORTED STATE

- a more conservative and practical estimate
- downscale frictions s.t. their variances match those in the least-distorted state
  - Tamil Nadu has the least dispersed distortions

reallocation gain

multi-product: **294**%
1-product: **212**%   (28% ↓)

EXERCISES

BENCHMARK EXERCISE

- equalize cross-farm distortions $\lambda_f, \tau_{f,g}$
- keep within-farm distortions $\tau_{f,l,i}, \tau_{f,g,i}$
  - preserve idiosyncratic product choice motives

MATCH LEAST-DISTORTED STATE

- a more conservative and practical estimate
- downscale frictions s.t. their variances match those in the least-distorted state
  - Tamil Nadu has the least dispersed distortions

reallocation gain

multi-product: **294%**
1-product: **212%** (28% ↓)

multi-product: **35%**
1-product: **45%** (26% ↑)

EXERCISES

Benchmark Exercise
- equalize cross-farm distortions $\lambda_f, \tau_{f,g}$
- keep within-farm distortions $\tau_{f,l,i}, \tau_{f,g,i}$
  - preserve idiosyncratic product choice motives

Match Least-Distorted State
- a more conservative and practical estimate
- downscale frictions s.t. their variances match those in the least-distorted state
  - Tamil Nadu has the least dispersed distortions

▶ *sensitivity to concavity* ▶ *role of states, seasons*

reallocation gain

multi-product: **294**%
1-product: **212**%   (28% ↓)

multi-product: **35**%
1-product: **45**%   (26% ↑)

Mechanisms

Mechanisms Contributing to 1-Product Model Error

- **Product Heterogeneity as Frictions** ▶ *details*

Mechanisms Contributing to 1-Product Model Error

- **Product Heterogeneity as Frictions** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP DISPERSION** ▶ *details*

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP DISPERSION** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products

Mechanisms Contributing to 1-Product Model Error

- **Product Heterogeneity as Frictions** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP Dispersion** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP DISPERSION** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation
- **RETURNS TO SCALE** ▶ *details*

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP DISPERSION** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation
- **RETURNS TO SCALE** ▶ *details*
  - farms growing products with high returns to scale can expand more

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP DISPERSION** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation
- **RETURNS TO SCALE** ▶ *details*
  - farms growing products with high returns to scale can expand more
  - 1-product model misses returns to scale heterogeneity

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP DISPERSION** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation
- **RETURNS TO SCALE** ▶ *details*
  - farms growing products with high returns to scale can expand more
  - 1-product model misses returns to scale heterogeneity
  - → understates misallocation

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP DISPERSION** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation
- **RETURNS TO SCALE** ▶ *details*
  - farms growing products with high returns to scale can expand more
  - 1-product model misses returns to scale heterogeneity
  - → understates misallocation
- **ENDOGENOUS PRODUCT CHOICE** ▶ *details*

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP DISPERSION** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation
- **RETURNS TO SCALE** ▶ *details*
  - farms growing products with high returns to scale can expand more
  - 1-product model misses returns to scale heterogeneity
  - → understates misallocation
- **ENDOGENOUS PRODUCT CHOICE** ▶ *details*
  - farms can partly mitigate effect of frictions

MECHANISMS CONTRIBUTING TO 1-PRODUCT MODEL ERROR

- **PRODUCT HETEROGENEITY AS FRICTIONS** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP DISPERSION** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation
- **RETURNS TO SCALE** ▶ *details*
  - farms growing products with high returns to scale can expand more
  - 1-product model misses returns to scale heterogeneity
  - → understates misallocation
- **ENDOGENOUS PRODUCT CHOICE** ▶ *details*
  - farms can partly mitigate effect of frictions
  - by growing product less intensive in the distorted input

Mechanisms Contributing to 1-Product Model Error

- **Product Heterogeneity as Frictions** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP Dispersion** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation
- **Returns to Scale** ▶ *details*
  - farms growing products with high returns to scale can expand more
  - 1-product model misses returns to scale heterogeneity
  - → understates misallocation
- **Endogenous Product Choice** ▶ *details*
  - farms can partly mitigate effect of frictions
  - by growing product less intensive in the distorted input
  - 1-product model misses this margin of adjustment

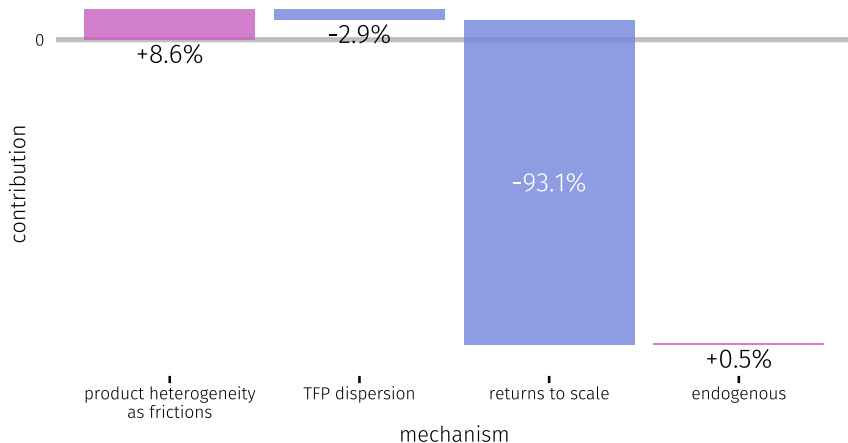# Mechanisms Contributing to 1-Product Model Error

- **Product Heterogeneity as Frictions** ▶ *details*
  - heterogeneous product choice → heterogeneous input mixes
  - 1-product model views heterogeneous input mixes as evidence of frictions
  - → overstates misallocation
- **TFP Dispersion** ▶ *details*
  - 1-product model misses within-farm TFP dispersion across products
  - → understates misallocation
- **Returns to Scale** ▶ *details*
  - farms growing products with high returns to scale can expand more
  - 1-product model misses returns to scale heterogeneity
  - → understates misallocation
- **Endogenous Product Choice** ▶ *details*
  - farms can partly mitigate effect of frictions
  - by growing product less intensive in the distorted input
  - 1-product model misses this margin of adjustment
  - → overstates misallocation

- **benchmark**: single-product model understates gain by **82 pp** (28%)

- **benchmark**: single-product model understates gain by **82 pp** (28%)

Mechanisms Decomposition: Benchmark Reallocation

- **benchmark**: single-product model understates gain by **82 pp** (28%)

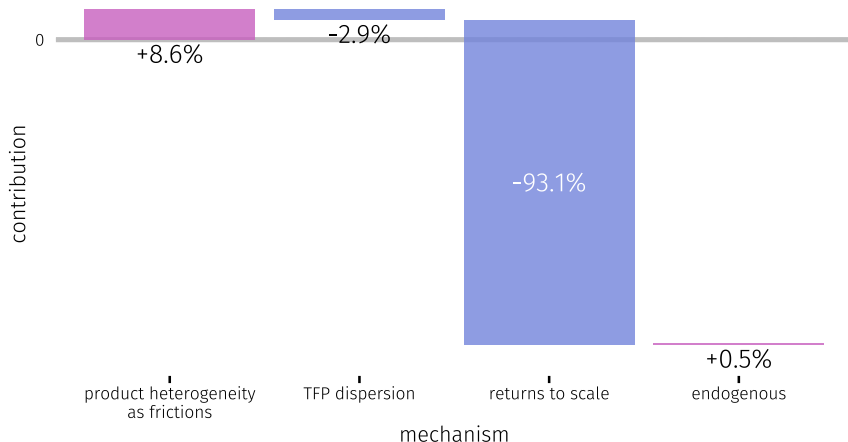- view most dispersion as real frictions to be removed → firms' ability to expand matters most
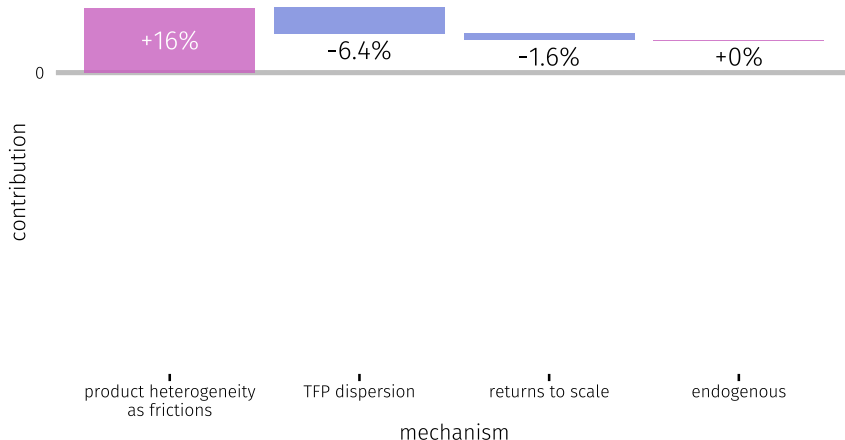 - → single-product model understates misallocation

- "least-distorted state": single-product model overstates gain by **10 pp** (26%)

- "least-distorted state": single-product model overstates gain by **10 pp** (26%)

- "least-distorted state": single-product model overstates gain by **10 pp** (26%)
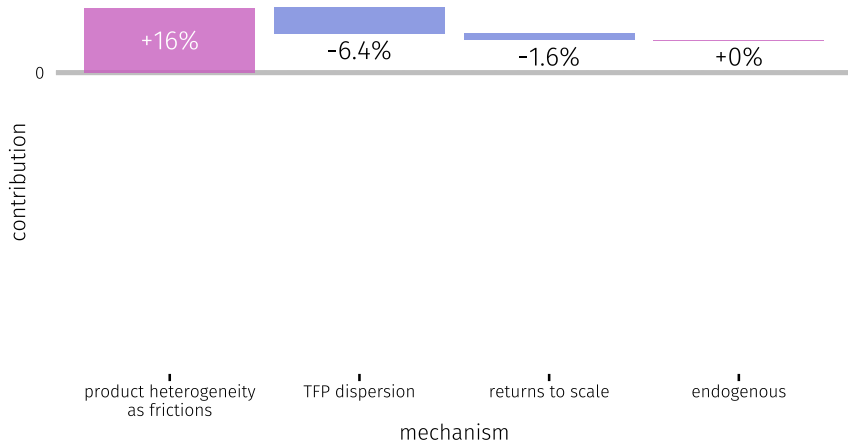
- view most dispersion as mismeasurement/benign → estimation of frictions matters most
  - → single-product model overstates **misallocation** ▶ *details*

CONCLUSION

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products

Conclusion

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India

Conclusion

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:
  - understate misallocation by

Conclusion

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:
    - understate misallocation by
        - > ignoring within-farm productivity heterogeneity

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:
  - understate misallocation by
    - > ignoring within-farm productivity heterogeneity
    - > ignoring returns-to-scale heterogeneity across crops

Conclusion

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:
  - understate misallocation by
    - > ignoring within-farm productivity heterogeneity
    - > ignoring returns-to-scale heterogeneity across crops
  - overstate misallocation by

Conclusion

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:
  - understate misallocation by
    - \> ignoring within-farm productivity heterogeneity
    - \> ignoring returns-to-scale heterogeneity across crops
  - overstate misallocation by
    - \> misinterpreting product heterogeneity as frictions

Conclusion

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:
  - understate misallocation by
    - > ignoring within-farm productivity heterogeneity
    - > ignoring returns-to-scale heterogeneity across crops
  - overstate misallocation by
    - > misinterpreting product heterogeneity as frictions
    - > missing endogenous product choice response

Conclusion

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:
  - understate misallocation by
    - > ignoring within-farm productivity heterogeneity
    - > ignoring returns-to-scale heterogeneity across crops
  - overstate misallocation by
    - > misinterpreting product heterogeneity as frictions
    - > missing endogenous product choice response
- benchmark exercise: 1-product model understates misallocation

Conclusion

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:
  - understate misallocation by
    - > ignoring within-farm productivity heterogeneity
    - > ignoring returns-to-scale heterogeneity across crops
  - overstate misallocation by
    - > misinterpreting product heterogeneity as frictions
    - > missing endogenous product choice response
- benchmark exercise: 1-product model understates misallocation
  - but error depends on exercise of interest

Conclusion

- **existing misallocation estimates**: single aggregate product or multiple homogeneous products
- estimate heterogeneous production functions across crops in India
- multi-product model shows that conventional 1-product models:
  - understate misallocation by
    - > ignoring within-farm productivity heterogeneity
    - > ignoring returns-to-scale heterogeneity across crops
  - overstate misallocation by
    - > misinterpreting product heterogeneity as frictions
    - > missing endogenous product choice response
- benchmark exercise: 1-product model understates misallocation
  - but error depends on exercise of interest
  - no simple correction that can be applied to standard models

WHY AGRICULTURE?

- **agriculture** is the perfect setting to study **endogenous product choice**

- **agriculture** is the perfect setting to study **endogenous product choice**
  - firm-product (farm-crop) inputs and outputs are feasible to measure

back

- **agriculture** is the perfect setting to study **endogenous product choice**
  - firm-product (farm-crop) inputs and outputs are feasible to measure
  - products (crops) are ~homogeneous across firms (farms)

WHY AGRICULTURE?

- **agriculture** is the perfect setting to study **endogenous product choice**
  - firm-product (farm-crop) inputs and outputs are feasible to measure
  - products (crops) are $\sim$homogeneous across firms (farms)
- but mechanism applies to **non-agricultural** settings too
  - $\rightarrow$ relevant for *more* developed countries too

$$\sum_{i \in I_f} \left(\lambda_f^{\frac{\eta \sum_g \alpha_{g,i} - 1}{1 - \eta \sum_g \alpha_{g,i} - \eta \gamma_i}} \right) \left((p_i z_{f,i})^{\eta} \, \eta \left(\frac{\gamma_i}{\tau_{f,l,i}} \right)^{1 - \eta \sum_g \alpha_{g,i}} \Pi_g \left(\frac{\alpha_{g,i}}{r_g \tau_{f,g} \tau_{f,g,i}} \right)^{\eta \alpha_{g,i}} \right)^{\frac{1}{1 - \eta \sum_g \alpha_{g,i} - \eta \gamma_i}} \tau_{f,l,i} = L_f$$

$$x_{f,g,i} = \frac{\alpha_{g,i}}{r_g \tau_{f,g} \tau_{f,g,i}} \left(\frac{\gamma_i}{\lambda_f \tau_{f,l,i}} \right)^{\frac{\eta \gamma_i}{1 - \eta \sum_g \alpha_{g,i} - \eta \gamma_i}} \left((p_i z_{f,i})^{\eta} \, \eta \Pi_h \left(\frac{\alpha_{h,i}}{r_h \tau_{f,h} \tau_{f,h,i}} \right)^{\eta \alpha_{h,i}} \right)^{\frac{1}{1 - \eta \sum_h \alpha_{h,i} - \eta \gamma_i}}$$

$$l_{f,i} = \left((p_i z_{f,i})^{\eta} \, \eta \left(\frac{\gamma_i}{\lambda_f \tau_{f,l,i}} \right)^{1 - \eta \sum_g \alpha_{g,i}} \Pi_g \left(\frac{\alpha_{g,i}}{r_g \tau_{f,g} \tau_{f,g,i}} \right)^{\eta \alpha_{g,i}} \right)^{\frac{1}{1 - \eta \sum_g \alpha_{g,i} - \eta \gamma_i}}$$

back

LIST OF CROPS

Crop list

Rice	Wheat	Other Cereals	Pulses	Oilseeds, Fruits and Vegetables		
				Oilseeds	Vegetables	Fruits /Condiments
		Barley	Black gram	Sesame	Ash gourd	Mango
		Maize	Green peas	Groundnut	Beet root	Papaya
		Sorghum	Pigeon peas	Castor	Bitter gourd	Grapes
		Pearl millet	Horse gram	Sunflower	Bottle gourd	Plum
		Finger millet	Cowpea	Niger	Eggplant	Cardamom
		Others	Kidney bean	Soybean	Board bean	Chilli
			Lentil	Safflower	Cabbage	Cumin
			Chickpeas	Rapseed	Cauliflower	Dill seed
			Others	Linseed	Carrot	Indian mustard
				Others	Potato	Other
					Cucumber	
					Peas	

back

Inputs, Outputs

- labor:
  - family and hired labor days by plots

- labor:
  - family and hired labor days by plots
- land:
  - quality-adjusted land measure

- **labor**:
  - family and hired labor days by plots
- **land**:
  - quality-adjusted land measure
- **intermediate inputs**:
  - expenses on seeds, fertilizer, irrigation, machinery and bullocks, and fuels

- **labor**:
  – family and hired labor days by plots
- **land**:
  – quality-adjusted land measure
- **intermediate inputs**:
  – expenses on seeds, fertilizer, irrigation, machinery and bullocks, and fuels
- **output**:
  – market value of quantity harvested

back

INSTRUMENTAL VARIABLES

- **idea**: shocks to farm f's plots $k \neq j$ change shadow price of inputs on f's plot j

INSTRUMENTAL VARIABLES

- **idea**: shocks to farm f's plots $k \neq j$ change shadow price of inputs on f's plot j
  - $\rightarrow$ serve as instruments for inputs on j

- **idea**: shocks to farm f's plots $k \neq j$ change shadow price of inputs on f's plot j
  - $\rightarrow$ serve as instruments for inputs on j
- **assumptions**:

- **idea**: shocks to farm f's plots $k \neq j$ change shadow price of inputs on f's plot j
  - $\rightarrow$ serve as instruments for inputs on j
- **assumptions**:
  - **observed** shocks to farm f's plot k affect **shadow price** of inputs on f's plot j

back

Instrumental Variables

- **idea**: shocks to farm f's plots $k \neq j$ change shadow price of inputs on f's plot j
  - $\rightarrow$ serve as instruments for inputs on j
- **assumptions**:
  - **observed** shocks to farm f's plot k affect **shadow price** of inputs on f's plot j
  - **observed** shocks affecting input demand on plot k are not correlated with **unobserved** shocks affecting input demand on plot j, conditional on observed shocks to j

back

Instrumental Variables

- **idea**: shocks to farm f's plots $k \neq j$ change shadow price of inputs on f's plot j
  - $\rightarrow$ serve as instruments for inputs on j
- **assumptions**:
  - **observed** shocks to farm f's plot k affect **shadow price** of inputs on f's plot j
  - **observed** shocks affecting input demand on plot k are not correlated with **unobserved** shocks affecting input demand on plot j, conditional on observed shocks to j
  - f-level shocks interacted with plot-level features provide such shocks

Instrumental Variables

- **idea**: shocks to farm f's plots $k \neq j$ change shadow price of inputs on f's plot j
  - $\rightarrow$ serve as instruments for inputs on j
- **assumptions**:
  - **observed** shocks to farm f's plot k affect **shadow price** of inputs on f's plot j
  - **observed** shocks affecting input demand on plot k are not correlated with **unobserved** shocks affecting input demand on plot j, conditional on observed shocks to j
  - f-level shocks interacted with plot-level features provide such shocks
- **2SLS first stage**:

$$M_{j,i,t} = Z_{k \neq j,i,t} + \mu_{j,i,t}$$

  - M = land, labor, intermediates
  - $Z_{k \neq j,i,t}$: instruments **from other plots within farm**
    - $>$ agricultural shocks interacted with plot characteristics
    - $>$ household, community characteristics & shocks interacted with plot characteristics

- common land input measures are problematic

- common land input measures are problematic
  - **plot area**: ignores quality

- common land input measures are problematic
  - **plot area**: ignores quality
  - **plot market price**: likely excessively noisy when land markets are underdeveloped

- common land input measures are problematic
  - **plot area**: ignores quality
  - **plot market price**: likely excessively noisy when land markets are underdeveloped
- instead: predict plot's market price with its physical land features

- common land input measures are problematic
  - **plot area**: ignores quality
  - **plot market price**: likely excessively noisy when land markets are underdeveloped
- instead: predict plot's market price with its physical land features
  - use random forests for prediction

- common land input measures are problematic
  - **plot area**: ignores quality
  - **plot market price**: likely excessively noisy when land markets are underdeveloped
- instead: predict plot's market price with its physical land features
  - use random forests for prediction
- → quality index = predicted price/acre

- common land input measures are problematic
  - **plot area**: ignores quality
  - **plot market price**: likely excessively noisy when land markets are underdeveloped
- instead: predict plot's market price with its physical land features
  - use random forests for prediction
- $\rightarrow$ quality index $=$ predicted price/acre
  - **captures quality**: RF explains $> \frac{1}{2}$ of observed price variation in **test** sample

- common land input measures are problematic
  - **plot area**: ignores quality
  - **plot market price**: likely excessively noisy when land markets are underdeveloped
- instead: predict plot's market price with its physical land features
  - use random forests for prediction
- $\rightarrow$ quality index $=$ predicted price/acre
  - **captures quality**: RF explains $> \frac{1}{2}$ of observed price variation in **test** sample
  - **removes noise**: use predicted values, minimize overfitting $\rightarrow$ captured variation driven by population patterns, not individual mismeasurement

- common land input measures are problematic
  - **plot area**: ignores quality
  - **plot market price**: likely excessively noisy when land markets are underdeveloped
- instead: predict plot's market price with its physical land features
  - use random forests for prediction
- $\rightarrow$ quality index = predicted price/acre
  - **captures quality**: RF explains $> \frac{1}{2}$ of observed price variation in **test** sample
  - **removes noise**: use predicted values, minimize overfitting $\rightarrow$ captured variation driven by population patterns, not individual mismeasurement

▶ *details*

back

Inputs: Quality-Adjusted Land

- procedure

- procedure
  1. split data into **training** sample (2/3) and **test** sample (1/3)

Inputs: Quality-Adjusted Land

- procedure
    1. split data into **training** sample (2/3) and **test** sample (1/3)
    2. estimate OLS & RF on **training** data: **reg log price/acre on land features**

Inputs: Quality-Adjusted Land

- procedure
    1. split data into **training** sample (2/3) and **test** sample (1/3)
    2. estimate OLS & RF on **training** data: **reg log price/acre on land features**
        > physical characteristics: soil type, color, salinity, drainage, ...

- procedure
  1. split data into **training** sample (2/3) and **test** sample (1/3)
  2. estimate OLS & RF on **training** data: **reg log price/acre on land features**
     > physical characteristics: soil type, color, salinity, drainage, ...
     > irrigation access: presence of wells, canals, ...

- procedure
    1. split data into **training** sample (2/3) and **test** sample (1/3)
    2. estimate OLS & RF on **training** data: **reg log price/acre on land features**
        > physical characteristics: soil type, color, salinity, drainage, ...
        > irrigation access: presence of wells, canals, ...
        > OLS includes all 2-way interactions

Inputs: Quality-Adjusted Land

- procedure
    1. split data into **training** sample (2/3) and **test** sample (1/3)
    2. estimate OLS & RF on **training** data: **reg log price/acre on land features**
        > physical characteristics: soil type, color, salinity, drainage, ...
        > irrigation access: presence of wells, canals, ...
        > OLS includes all 2-way interactions
        > RF tuned with k-fold cross-validation

Inputs: Quality-Adjusted Land

- procedure
    1. split data into **training** sample (2/3) and **test** sample (1/3)
    2. estimate OLS & RF on **training** data: reg log price/acre on land features
        > physical characteristics: soil type, color, salinity, drainage, ...
        > irrigation access: presence of wells, canals, ...
        > OLS includes all 2-way interactions
        > RF tuned with k-fold cross-validation
    3. compute Mean Squared Error and R^2 on **test** data
        > to preclude overfitting

Inputs: Quality-Adjusted Land

- procedure
    1. split data into **training** sample (2/3) and **test** sample (1/3)
    2. estimate OLS & RF on **training** data: **reg log price/acre on land features**
        - physical characteristics: soil type, color, salinity, drainage, ...
        - irrigation access: presence of wells, canals, ...
        - OLS includes all 2-way interactions
        - RF tuned with k-fold cross-validation
    3. compute Mean Squared Error and R^2 on **test** data
        - to preclude overfitting
- results

	OLS	RF
MSE	0.61	0.49
R^2	0.39	0.51

Inputs: Quality-Adjusted Land

- procedure
    1. split data into **training** sample (2/3) and **test** sample (1/3)
    2. estimate OLS & RF on **training** data: **reg log price/acre on land features**
        > physical characteristics: soil type, color, salinity, drainage, ...
        > irrigation access: presence of wells, canals, ...
        > OLS includes all 2-way interactions
        > RF tuned with k-fold cross-validation
    3. compute Mean Squared Error and R^2 on **test** data
        > to preclude overfitting
- results

	OLS	RF
MSE	0.61	0.49
R^2	0.39	0.51

- $\rightarrow$ quality index $=$ predicted price/acre

Inputs: Quality-Adjusted Land

- procedure
  1. split data into **training** sample (2/3) and **test** sample (1/3)
  2. estimate OLS & RF on **training** data: reg log price/acre on land features
     > physical characteristics: soil type, color, salinity, drainage, ...
     > irrigation access: presence of wells, canals, ...
     > OLS includes all 2-way interactions
     > RF tuned with k-fold cross-validation
  3. compute Mean Squared Error and R^2 on **test** data
     > to preclude overfitting
- results

	OLS	RF
MSE	0.61	0.49
R^2	0.39	0.51

- $\rightarrow$ quality index $=$ predicted price/acre
- quality-adjusted land $=$ quality index $\times$ plot area

Production Functions: All Crops

estimate + 95% CI: γ (land) α_{labor} α_{inter} $\gamma + \alpha_{labor} + \alpha_{inter}$

back
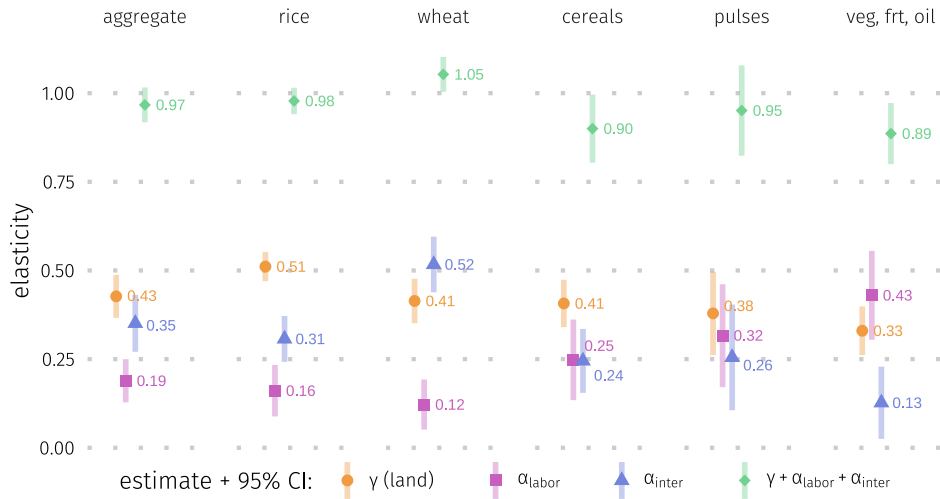
PRODUCTION FUNCTIONS: TABLE

	Aggregate	Rice	Wheat	Other Cereals	Pulses	Veg, Frt, Oil
Land	0.427	0.511	0.414	0.407	0.379	0.330
	(0.031)	(0.021)	(0.032)	(0.034)	(0.060)	(0.035)
Labor	0.189	0.161	0.122	0.248	0.316	0.430
	(0.031)	(0.037)	(0.036)	(0.058)	(0.074)	(0.064)
Intermediates	0.351	0.307	0.517	0.245	0.255	0.127
	(0.041)	(0.033)	(0.040)	(0.046)	(0.076)	(0.052)
Observations	14,705	4,807	3,566	2,779	1,128	2,338
R^2	0.624	0.742	0.713	0.590	0.417	0.572
Village FEs	Y	Y	Y	Y	Y	Y
Season FEs	Y	Y	Y	Y	Y	Y
	First Stage: F statistics					
Land	77.0	62.0	40.3	37.8	15.7	19.3
Labor	49.3	34.7	17.7	25.2	12.9	14.8
Intermediates	35.8	31.7	21.5	19.9	8.9	11.8
K-Paap Wald F statistic	51.1	40.4	16.0	30.8	12.4	12.7

back

Production Functions: Pairwise Equality Test P-Values

γ

	rice	wheat	cereals	pulses
veg, frt, oil	<0.01	0.95	0.62	0.73
pulses	0.05	0.7	1	
cereals	<0.01	0.57		
wheat	<0.01			

α_labor

	rice	wheat	cereals	pulses
veg, frt, oil	<0.01	<0.01	0.08	0.43
pulses	0.01	0.03	0.3	
cereals	0.17	0.33		
wheat	0.65			

α_inter

	rice	wheat	cereals	pulses
veg, frt, oil	<0.01	<0.01	0.08	0.27
pulses	0.53	<0.01	0.94	
cereals	0.35	<0.01		
wheat	<0.01			

back

- for each farm f, need to check all 2^N possible sets of nonzero crops

- for each farm f, need to check all 2^N possible sets of nonzero crops
    1. fix crop set's λ_f by solving one non-linear equation for each farm

back

- for each farm f, need to check all 2^N possible sets of nonzero crops
    1. fix crop set's λ_f by solving one non-linear equation for each farm
    2. compute optimal $\{x_{f,g,i}\}_{g,i}$, $\{l_{f,i}\}_i$, $\{y_{f,i}\}_i$ given λ_f

back

- for each farm f, need to check all 2^N possible sets of nonzero crops
  1. fix crop set's λ_f by solving one non-linear equation for each farm
  2. compute optimal $\{x_{f,g,i}\}_{g,i}$, $\{l_{f,i}\}_i$, $\{y_{f,i}\}_i$ given λ_f
  3. compute optimal profit net of fixed cost
     - ▶ *expressions*

- for each farm f, need to check all 2^N possible sets of nonzero crops
  1. fix crop set's λ_f by solving one non-linear equation for each farm
  2. compute optimal $\{x_{f,g,i}\}_{g,i}$, $\{l_{f,i}\}_i$, $\{y_{f,i}\}_i$ given λ_f
  3. compute optimal profit net of fixed cost
     - ▶ *expressions*
- for each farm, pick profit-maximizing crop set

Potential Sources of Concavity

- subsistence farmer with love of variety

POTENTIAL SOURCES OF CONCAVITY

- subsistence farmer with love of variety
  - farmer minimizes costs of providing utility U:

$$U = \left(\sum_i \varphi_i y_{f,i}^\eta \right)^{\frac{1}{\eta}}$$

POTENTIAL SOURCES OF CONCAVITY

- subsistence farmer with love of variety
  - farmer minimizes costs of providing utility U:

$$U = \left(\sum_i \varphi_i y_{f,i}^{\eta} \right)^{\frac{1}{\eta}}$$

  $\rightarrow$ mix crops to provide a diverse diet

POTENTIAL SOURCES OF CONCAVITY

- subsistence farmer with love of variety
  - farmer minimizes costs of providing utility U:

$$U = \left(\sum_i \varphi_i y_{f,i}^{\eta} \right)^{\frac{1}{\eta}}$$

  $\rightarrow$ mix crops to provide a diverse diet
- crop-level markets are monopolistically competitive

POTENTIAL SOURCES OF CONCAVITY

- subsistence farmer with love of variety
  - farmer minimizes costs of providing utility U:

$$U = \left(\sum_i \varphi_i y_{f,i}^{\eta} \right)^{\frac{1}{\eta}}$$

  $\rightarrow$ mix crops to provide a diverse diet
- crop-level markets are monopolistically competitive
  - intermediate crop aggregator combines farms' varieties:

$$Y_i = \left(\sum_f y_{f,i}^{\eta} \right)^{\frac{1}{\eta}}$$

- subsistence farmer with love of variety
  - farmer minimizes costs of providing utility U:

$$U = \left(\sum_i \varphi_i y_{f,i}^{\eta} \right)^{\frac{1}{\eta}}$$

  $\rightarrow$ mix crops to provide a diverse diet
- crop-level markets are monopolistically competitive
  - intermediate crop aggregator combines farms' varieties:

$$Y_i = \left(\sum_f y_{f,i}^{\eta} \right)^{\frac{1}{\eta}}$$

  $\rightarrow$ mix crops due to downward-sloping demand in each crop

POTENTIAL SOURCES OF CONCAVITY

- subsistence farmer with love of variety
  - farmer minimizes costs of providing utility U:

$$U = \left(\sum_i \varphi_i y_{f,i}^\eta \right)^{\frac{1}{\eta}}$$

  → mix crops to provide a diverse diet
- crop-level markets are monopolistically competitive
  - intermediate crop aggregator combines farms' varieties:

$$Y_i = \left(\sum_f y_{f,i}^\eta \right)^{\frac{1}{\eta}}$$

  → mix crops due to downward-sloping demand in each crop
- both setups produce farm-level FOCs that are identical to the main model

POTENTIAL SOURCES OF CONCAVITY

- subsistence farmer with love of variety
    - farmer minimizes costs of providing utility U:

$$U = \left(\sum_i \varphi_i y_{f,i}^\eta \right)^{\frac{1}{\eta}}$$

  $\rightarrow$ mix crops to provide a diverse diet
- crop-level markets are monopolistically competitive
    - intermediate crop aggregator combines farms' varieties:

$$Y_i = \left(\sum_f y_{f,i}^\eta \right)^{\frac{1}{\eta}}$$

  $\rightarrow$ mix crops due to downward-sloping demand in each crop
- both setups produce farm-level FOCs that are identical to the main model
    - but the appropriate way to define GE may be different

Extracting Distortions

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i} \right)^{\eta}}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i} \right)^{\eta}}{l_{f,i}} = mrpl_{f,i}$$

EXTRACTING DISTORTIONS

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i} \right)^{\eta}}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i} \right)^{\eta}}{l_{f,i}} = mrpl_{f,i}$$

- physical **productivity** implied by production fn.:

$$z_{f,i} = \frac{y_{f,i}}{l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}$$

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i}\right)^{\eta}}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i}\right)^{\eta}}{l_{f,i}} = mrpl_{f,i}$$

- physical **productivity** implied by production fn.:

$$z_{f,i} = \frac{y_{f,i}}{l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}}\right)}$$

- $\rightarrow$ extracted fundamentals rationalize observed dispersion b/w farms

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i}\right)^\eta}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i}\right)^\eta}{l_{f,i}} = mrpl_{f,i}$$

- physical **productivity** implied by production fn.:

$$z_{f,i} = \frac{y_{f,i}}{l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}}\right)}$$

- $\rightarrow$ extracted fundamentals rationalize observed dispersion b/w farms
    frictionless economy

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i}\right)^{\eta}}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i}\right)^{\eta}}{l_{f,i}} = mrpl_{f,i}$$

- physical **productivity** implied by production fn.:

$$z_{f,i} = \frac{y_{f,i}}{l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}}\right)}$$

- $\rightarrow$ extracted fundamentals rationalize observed dispersion b/w farms

  frictionless economy $\quad \Leftrightarrow \quad \tau_{f,g} \tau_{f,g,i} = 1$

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i}\right)^\eta}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i}\right)^\eta}{l_{f,i}} = mrpl_{f,i}$$

- physical **productivity** implied by production fn.:

$$z_{f,i} = \frac{y_{f,i}}{l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}}\right)}$$

- $\rightarrow$ extracted fundamentals rationalize observed dispersion b/w farms

  frictionless economy $\quad \Leftrightarrow \quad \tau_{f,g} \tau_{f,g,i} = 1 \quad \Leftrightarrow \quad mrpg_{f,i} = \overline{mrpg}_i, \; mrpl_{f,i} = \overline{mrpl}_i$

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i} \right)^\eta}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i} \right)^\eta}{l_{f,i}} = mrpl_{f,i}$$

- physical **productivity** implied by production fn.:

$$z_{f,i} = \frac{y_{f,i}}{l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}$$

- $\rightarrow$ extracted fundamentals rationalize observed dispersion b/w farms

    frictionless economy $\quad \Leftrightarrow \quad \tau_{f,g} \tau_{f,g,i} = 1 \quad \Leftrightarrow \quad mrpg_{f,i} = \overline{mrpg}_i, \; mrpl_{f,i} = \overline{mrpl}_i$

    distorted economy

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i} \right)^{\eta}}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i} \right)^{\eta}}{l_{f,i}} = mrpl_{f,i}$$

- physical **productivity** implied by production fn.:

$$z_{f,i} = \frac{y_{f,i}}{l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}$$

- $\rightarrow$ extracted fundamentals rationalize observed dispersion b/w farms

    frictionless economy $\quad \Leftrightarrow \quad \tau_{f,g} \tau_{f,g,i} = 1 \quad \Leftrightarrow \quad mrpg_{f,i} = \overline{mrpg}_i, \; mrpl_{f,i} = \overline{mrpl}_i$

        distorted economy $\quad \Leftrightarrow \quad$ heterog. $\tau_{f,g} \tau_{f,g,i}$

EXTRACTING DISTORTIONS

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i} \right)^{\eta}}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i} \right)^{\eta}}{l_{f,i}} = mrpl_{f,i}$$

- physical **productivity** implied by production fn.:

$$z_{f,i} = \frac{y_{f,i}}{l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}$$

- $\to$ extracted fundamentals rationalize observed dispersion b/w farms

frictionless economy $\quad \Leftrightarrow \quad \tau_{f,g} \tau_{f,g,i} = 1 \quad \Leftrightarrow \quad mrpg_{f,i} = \overline{mrpg}_i, \ mrpl_{f,i} = \overline{mrpl}_i$

distorted economy $\quad \Leftrightarrow \quad$ heterog. $\tau_{f,g} \tau_{f,g,i} \quad \Leftrightarrow \quad$ heterog. $mrpg_{f,i}, \ mrpl_{f,i}$

Extracting Distortions

- unobserved **distortions** map to observed **marginal revenue products**:

$$r_g \tau_{f,g} \tau_{f,g,i} = \frac{\alpha_{g,i} \eta \left(p_i y_{f,i} \right)^\eta}{x_{f,g,i}} = mrpg_{f,i}$$

$$\lambda_f \tau_{f,l,i} = \frac{\gamma_i \eta \left(p_i y_{f,i} \right)^\eta}{l_{f,i}} = mrpl_{f,i}$$

- physical **productivity** implied by production fn.:

$$z_{f,i} = \frac{y_{f,i}}{l_{f,i}^{\gamma_i} \Pi_g \left(x_{f,g,i}^{\alpha_{g,i}} \right)}$$

- $\rightarrow$ extracted fundamentals rationalize observed dispersion b/w farms

$$\text{frictionless economy} \quad \Leftrightarrow \quad \tau_{f,g} \tau_{f,g,i} = 1 \quad \Leftrightarrow \quad mrpg_{f,i} = \overline{mrpg}_i, \ mrpl_{f,i} = \overline{mrpl}_i$$

$$\text{distorted economy} \quad \Leftrightarrow \quad \text{heterog. } \tau_{f,g} \tau_{f,g,i} \quad \Leftrightarrow \quad \text{heterog. } mrpg_{f,i}, \ mrpl_{f,i}$$

▶ *splitting distortions*

Splitting Distortions

- splitting $r_g \tau_{f,g} \tau_{f,g,i}$ into 3 terms is arbitrary from farm f's POV

Splitting Distortions

- splitting $r_g \tau_{f,g} \tau_{f,g,i}$ into 3 terms is arbitrary from farm f's POV
  - and **does not** matter for aggregate misallocation estimate

SPLITTING DISTORTIONS

- splitting $r_g \tau_{f,g} \tau_{f,g,i}$ into 3 terms is arbitrary from farm f's POV
  - and **does not** matter for aggregate misallocation estimate
  - but will matter for *partial* reallocation exercises $\rightarrow$ come up with a sensible split

SPLITTING DISTORTIONS

- splitting $r_g \tau_{f,g} \tau_{f,g,i}$ into 3 terms is arbitrary from farm f's POV
  - and **does not** matter for aggregate misallocation estimate
  - but will matter for *partial* reallocation exercises $\rightarrow$ come up with a sensible split
- split $r_g \tau_{f,g}$ from $\tau_{f,g,i}$:

SPLITTING DISTORTIONS

- splitting $r_g \tau_{f,g} \tau_{f,g,i}$ into 3 terms is arbitrary from farm f's POV
    - and **does not** matter for aggregate misallocation estimate
    - but will matter for *partial* reallocation exercises $\rightarrow$ come up with a sensible split
- split $r_g \tau_{f,g}$ from $\tau_{f,g,i}$:
    - assume $\tau_{f,g,i}$ don't distort f-level demand of g

SPLITTING DISTORTIONS

- splitting $r_g \tau_{f,g} \tau_{f,g,i}$ into 3 terms is arbitrary from farm f's POV
  - and **does not** matter for aggregate misallocation estimate
  - but will matter for *partial* reallocation exercises $\rightarrow$ come up with a sensible split
- split $r_g \tau_{f,g}$ from $\tau_{f,g,i}$:
  - assume $\tau_{f,g,i}$ don't distort f-level demand of g
  - $\rightarrow X_{f,g} = \sum_i x_{f,g,i} = \sum_i \tau_{f,g,i} x_{f,g,i}$

SPLITTING DISTORTIONS

- splitting $r_g \tau_{f,g} \tau_{f,g,i}$ into 3 terms is arbitrary from farm f's POV
  - and **does not** matter for aggregate misallocation estimate
  - but will matter for *partial* reallocation exercises $\rightarrow$ come up with a sensible split
- split $r_g \tau_{f,g}$ from $\tau_{f,g,i}$:
  - assume $\tau_{f,g,i}$ don't distort f-level demand of g
  - $\rightarrow X_{f,g} = \sum_i x_{f,g,i} = \sum_i \tau_{f,g,i} x_{f,g,i}$
- likewise restrict $\tau_{f,l,i}$:
  - assume $\tau_{f,l,i}$ don't distort f-level demand of l

SPLITTING DISTORTIONS

- splitting $r_g \tau_{f,g} \tau_{f,g,i}$ into 3 terms is arbitrary from farm f's POV
  - and **does not** matter for aggregate misallocation estimate
  - but will matter for *partial* reallocation exercises $\rightarrow$ come up with a sensible split
- split $r_g \tau_{f,g}$ from $\tau_{f,g,i}$:
  - assume $\tau_{f,g,i}$ don't distort f-level demand of g
  - $\rightarrow X_{f,g} = \sum_i x_{f,g,i} = \sum_i \tau_{f,g,i} x_{f,g,i}$
- likewise restrict $\tau_{f,l,i}$:
  - assume $\tau_{f,l,i}$ don't distort f-level demand of l
  - $\rightarrow L_f = \sum_i l_{f,i} = \sum_i \tau_{f,l,i} l_{f,i}$

Splitting Distortions

- splitting $r_g \tau_{f,g} \tau_{f,g,i}$ into 3 terms is arbitrary from farm f's POV
  - and **does not** matter for aggregate misallocation estimate
  - but will matter for *partial* reallocation exercises $\rightarrow$ come up with a sensible split
- split $r_g \tau_{f,g}$ from $\tau_{f,g,i}$:
  - assume $\tau_{f,g,i}$ don't distort f-level demand of g
  - $\rightarrow X_{f,g} = \sum_i x_{f,g,i} = \sum_i \tau_{f,g,i} x_{f,g,i}$
- likewise restrict $\tau_{f,l,i}$:
  - assume $\tau_{f,l,i}$ don't distort f-level demand of l
  - $\rightarrow L_f = \sum_i l_{f,i} = \sum_i \tau_{f,l,i} l_{f,i}$
- split of r_g from $\tau_{f,g}$ imposed by GE

ESTIMATE THE ELASTICITY OF SUBSTITUTION

- from consumption FOC:

$$\log\left(\frac{p_i C_i}{\sum_j p_j C_j}\right) = -\log\left(\sum_j \varphi_j^\sigma p_j^{1-\sigma}\right) + (1-\sigma)\log p_i + \sigma \log \varphi_i$$

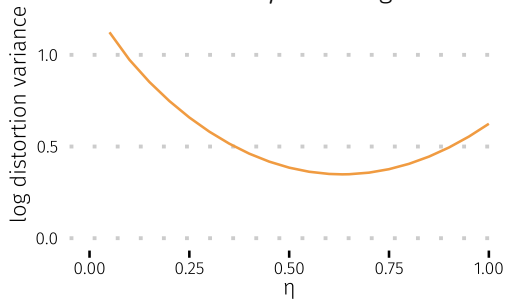## ESTIMATE THE ELASTICITY OF SUBSTITUTION

- from consumption FOC:

$$\log\left(\frac{p_i C_i}{\sum_j p_j C_j}\right) = -\log\left(\sum_j \varphi_j^\sigma p_j^{1-\sigma}\right) + (1-\sigma)\log p_i + \sigma \log \varphi_i$$

- estimate with 2SLS: $\log s_{h,i} = \beta_0 + \beta_1 \log p_{h,i} + \gamma_i + \varepsilon_{h,i}$

Estimate the Elasticity of Substitution

- from consumption FOC:

$$\log \left(\frac{p_i C_i}{\sum_j p_j C_j} \right) = - \log \left(\sum_j \varphi_j^\sigma p_j^{1-\sigma} \right) + (1 - \sigma) \log p_i + \sigma \log \varphi_i$$

- estimate with 2SLS: $\log s_{h,i} = \beta_0 + \beta_1 \log p_{h,i} + \gamma_i + \varepsilon_{h,i}$

	$\log s_{h,i}$
σ	1.699
$\log p_{h,i}$	-0.699
	(0.067)
Observations	40,833
Kleibergen-Paap F stat	230.9

Village-level instruments: Elevation × rain, ruggedness × rain, pucca roads availability

back

- τs reproducing data need to be extreme if η is too high or too low
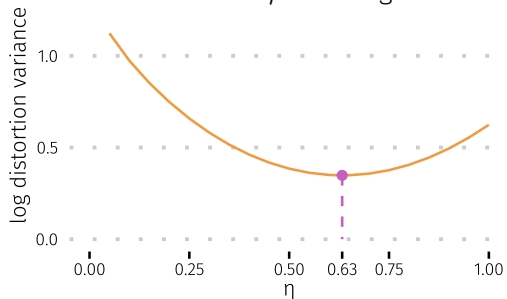
Choose η That Minimizes Implied Dispersion

- τs reproducing data need to be extreme if η is too high or too low

- **low** η: farm, farm-crop output more uniformly distributed in efficient allocation
  - $\rightarrow$ data farm size **"too varied"**, farms mix crops **"too little"** $\rightarrow$ extreme distortions
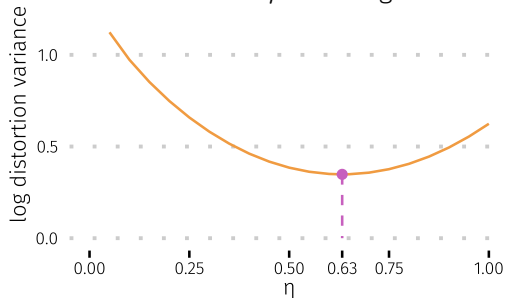
- τs reproducing data need to be extreme if η is too high or too low
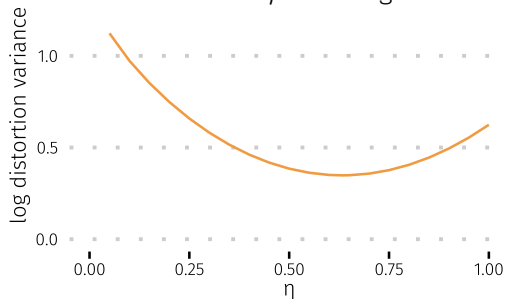
- **low** η: farm, farm-crop output more uniformly distributed in efficient allocation
  - $\rightarrow$ data farm size **"too varied"**, farms mix crops **"too little"** $\rightarrow$ extreme distortions
- **high** η: farm, farm-crop output more dispersed in efficient allocation
  - $\rightarrow$ data farm size **"too uniform"**, farms mix crops **"too much"** $\rightarrow$ extreme distortions

CHOOSE η THAT MINIMIZES IMPLIED DISPERSION

- τs reproducing data need to be extreme if η is too high or too low

- **low** η: farm, farm-crop output more uniformly distributed in efficient allocation
  - $\rightarrow$ data farm size **"too varied"**, farms mix crops **"too little"** $\rightarrow$ extreme distortions
- **high** η: farm, farm-crop output more dispersed in efficient allocation
  - $\rightarrow$ data farm size **"too uniform"**, farms mix crops **"too much"** $\rightarrow$ extreme distortions
- $\rightarrow$ pick η that minimizes distortions required to explain observed output distribution
  - $\rightarrow$ conservative misallocation estimates

CHOOSE η THAT MINIMIZES IMPLIED DISPERSION

- τs reproducing data need to be extreme if η is too high or too low

- **low** η: farm, farm-crop output more uniformly distributed in efficient allocation
    - → data farm size **"too varied"**, farms mix crops **"too little"** → extreme distortions
- **high** η: farm, farm-crop output more dispersed in efficient allocation
    - → data farm size **"too uniform"**, farms mix crops **"too much"** → extreme distortions
- → pick η that minimizes distortions required to explain observed output distribution
    - → conservative misallocation estimates

▶ *details*

back

General Equilibrium

- representative consumer buys crops, sells inputs, receives profit from owned farms

$$\max_{\{C_i\}_{i=1}^N} \left(\sum_i \varphi_i C_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

s.t.

$$\sum_i p_i C_i = \sum_g r_g X_g^{agg} + \Pi$$

General Equilibrium

- representative consumer buys crops, sells inputs, receives profit from owned farms

$$\max_{\{C_i\}_{i=1}^{N}} \left(\sum_i \varphi_i C_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

s.t.

$$\sum_i p_i C_i = \sum_g r_g X_g^{agg} + \Pi$$

- goods markets clear $\forall i$:

$$C_i = \sum_f y_{f,i}$$

back

General Equilibrium

- representative consumer buys crops, sells inputs, receives profit from owned farms

$$\max_{\{C_i\}_{i=1}^N} \left(\sum_i \varphi_i C_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

s.t.

$$\sum_i p_i C_i = \sum_g r_g X_g^{agg} + \Pi$$

- goods markets clear $\forall i$:

$$C_i = \sum_f y_{f,i}$$

- inputs markets clear $\forall g$:

$$\sum_f \sum_i x_{f,g,i} = X_g^{agg}$$

General Equilibrium

- representative consumer buys crops, sells inputs, receives profit from owned farms

$$\max_{\{C_i\}_{i=1}^N} \left(\sum_i \varphi_i C_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

s.t.

$$\sum_i p_i C_i = \sum_g r_g X_g^{agg} + \Pi$$

- goods markets clear $\forall i$:

$$C_i = \sum_f y_{f,i}$$

- inputs markets clear $\forall g$:

$$\sum_f \sum_i x_{f,g,i} = X_g^{agg}$$

- estimate $\sigma = 1.7$ from consumption FOC ▶ *details*

General Equilibrium

- representative consumer buys crops, sells inputs, receives profit from owned farms

$$\max_{\{C_i\}_{i=1}^N} \left(\sum_i \varphi_i C_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

s.t.

$$\sum_i p_i C_i = \sum_g r_g X_g^{agg} + \Pi$$

- goods markets clear $\forall i$:

$$C_i = \sum_f y_{f,i}$$

- inputs markets clear $\forall g$:

$$\sum_f \sum_i x_{f,g,i} = X_g^{agg}$$

- estimate $\sigma = 1.7$ from consumption FOC ▶ *details*

▶ *profits details*

Choose η That Minimizes Implied Dispersion

- farm-crop revenue:

$$p_i y_{f,i} = \left(\underbrace{\left(\frac{1}{\lambda_f \tau_{f,m,i}} \right)^{\gamma_i} \Pi_g \left(\frac{1}{\tau_{f,g} \tau_{f,g,i}} \right)^{\alpha_{g,i}}}_{\text{composite distortion, } dist_{f,i}} \right)^{\frac{1}{1 - \eta(\sum_g \alpha_{g,i} + \gamma_i)}} \underbrace{\left(p_i z_{f,i} \gamma_i^{\gamma_i} \eta^{\sum_g \alpha_{g,i} + \gamma_i} \Pi \left(\frac{\alpha_{g,i}}{r_g} \right)^{\alpha_{g,i}} \right)^{\frac{1}{1 - \eta(\sum_g \alpha_{g,i} + \gamma_i)}}}_{\text{"objective" factors}}$$

back

Choose η That Minimizes Implied Dispersion

- farm-crop revenue:

$$p_i y_{f,i} = \left(\underbrace{\left(\frac{1}{\lambda_f \tau_{f,m,i}} \right)^{\gamma_i} \Pi_g \left(\frac{1}{\tau_{f,g} \tau_{f,g,i}} \right)^{\alpha_{g,i}}}_{\text{composite distortion, } dist_{f,i}} \right)^{\frac{1}{1 - \eta(\sum_g \alpha_{g,i} + \gamma_i)}} \underbrace{\left(p_i z_{f,i} \gamma_i^{\gamma_i} \eta^{\sum_g \alpha_{g,i} + \gamma_i} \Pi \left(\frac{\alpha_{g,i}}{r_g} \right)^{\alpha_{g,i}} \right)}_{\text{"objective" factors}}^{\frac{1}{1 - \eta(\sum_g \alpha_{g,i} + \gamma_i)}}$$

- $Var(\log dist_{f,i})$ needed to match observed output dispersion depends on chosen η

$$\Pi = \sum_f \left[\sum_{i=1}^{N} p_i y_{f,i} - \sum_{g=1}^{G} r_g \sum_{i=1}^{N} x_{f,g,i} \right]$$

- distortions τ, fixed costs ω, concavity η are not reflected in dividends sent to consumer
  - farmers act *as if* frictions they face had monetary representations
  - but these are non-monetary and not added/subtracted from dividends
- equivalent formulation: τs, ωs are monetary taxes/subsidies, administered by consumer
  - show up in dividends and consumer's BC as government revenue/expense
- choice is arbitrary: both formulations (or any mixture) produce identical equilibrium conditions

REALLOCATION

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**

REALLOCATION

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
- equalize (or reduce) distortions τ between farms

Reallocation

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
- equalize (or reduce) distortions τ between farms
- prohibit product set switching

REALLOCATION

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
- equalize (or reduce) distortions τ between farms
- prohibit product set switching
  - $\rightarrow$ don't need productivities, frictions for unobserved farm-product combinations

Reallocation

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
- equalize (or reduce) distortions τ between farms
- prohibit product set switching
  - $\rightarrow$ don't need productivities, frictions for unobserved farm-product combinations
- treat each season as a separate economy
  - to prevent double-booking of inputs in a single season
  - sum aggregate output across seasons

Reallocation

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
- equalize (or reduce) distortions τ between farms
- prohibit product set switching
  - $\rightarrow$ don't need productivities, frictions for unobserved farm-product combinations
- treat each season as a separate economy
  - to prevent double-booking of inputs in a single season
  - sum aggregate output across seasons
- compute counterfactual Δoutput

Reallocation

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
- equalize (or reduce) distortions τ between farms
- prohibit product set switching
  - $\rightarrow$ don't need productivities, frictions for unobserved farm-product combinations
- treat each season as a separate economy
  - to prevent double-booking of inputs in a single season
  - sum aggregate output across seasons
- compute counterfactual Δoutput
  - **reallocation gain** = Δoutput between counterfactual and current

Reallocation

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
- equalize (or reduce) distortions τ between farms
- prohibit product set switching
  - $\rightarrow$ don't need productivities, frictions for unobserved farm-product combinations
- treat each season as a separate economy
  - to prevent double-booking of inputs in a single season
  - sum aggregate output across seasons
- compute counterfactual Δoutput
  - **reallocation gain** = Δoutput between counterfactual and current
  - aggregate inputs are fixed $\rightarrow$ aggregate output $\uparrow$ reflects aggregate TFP $\uparrow$

Reallocation

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
- equalize (or reduce) distortions τ between farms
- prohibit product set switching
    - $\rightarrow$ don't need productivities, frictions for unobserved farm-product combinations
- treat each season as a separate economy
    - to prevent double-booking of inputs in a single season
    - sum aggregate output across seasons
- compute counterfactual Δoutput
    - **reallocation gain** = Δoutput between counterfactual and current
    - aggregate inputs are fixed $\rightarrow$ aggregate output $\uparrow$ reflects aggregate TFP $\uparrow$
    - farm-level TFPs are fixed $\rightarrow$ aggregate TFP $\uparrow$ reflects pure **reallocation gain** or **misallocation cost**

Reallocation

- to quantify **aggregate cost** of distortions, conduct **counterfactual reallocations**
- equalize (or reduce) distortions τ between farms
- prohibit product set switching
  - $\rightarrow$ don't need productivities, frictions for unobserved farm-product combinations
- treat each season as a separate economy
  - to prevent double-booking of inputs in a single season
  - sum aggregate output across seasons
- compute counterfactual Δoutput
  - **reallocation gain** = Δoutput between counterfactual and current
  - aggregate inputs are fixed $\rightarrow$ aggregate output $\uparrow$ reflects aggregate TFP $\uparrow$
  - farm-level TFPs are fixed $\rightarrow$ aggregate TFP $\uparrow$ reflects pure **reallocation gain** or **misallocation cost**
- compare **reallocation gain** between **multi-product** model and **1-product** model
  - ▶ *reallocation exercise details*

Reallocation Exercise Details

- **reallocation 1**: equalize land and labor distortions, crop sets fixed

Reallocation Exercise Details

- **reallocation 1**: equalize land and labor distortions, crop sets fixed
  - farms can't change crop sets, but can change crop ratios

- **reallocation 1**: equalize land and labor distortions, crop sets fixed
  - farms can't change crop sets, but can change crop ratios
  - crop sets fixed $\rightarrow$ extract all needed fundamentals from data

Reallocation Exercise Details

- **reallocation 1**: equalize land and labor distortions, crop sets fixed
  - farms can't change crop sets, but can change crop ratios
  - crop sets fixed $\rightarrow$ extract all needed fundamentals from data
  - set $\tau_{f,g}\tau_{f,g,i} = 1$, $\lambda_f = \bar{\lambda}$

- **reallocation 1**: equalize land and labor distortions, crop sets fixed
  - farms can't change crop sets, but can change crop ratios
  - crop sets fixed $\rightarrow$ extract all needed fundamentals from data
  - set $\tau_{f,g}\tau_{f,g,i} = 1$, $\lambda_f = \bar{\lambda}$
  - solve for market-clearing $\{p_i\}_i$, $\{r_g\}_g$, $\bar{\lambda}$

- **reallocation 1**: equalize land and labor distortions, crop sets fixed
  - farms can't change crop sets, but can change crop ratios
  - crop sets fixed $\rightarrow$ extract all needed fundamentals from data
  - set $\tau_{f,g}\tau_{f,g,i} = 1$, $\lambda_f = \bar{\lambda}$
  - solve for market-clearing $\{p_i\}_i$, $\{r_g\}_g$, $\bar{\lambda}$
- **reallocation 2**: equalize land and labor distortions, allow crop set changes

- **reallocation 1**: equalize land and labor distortions, crop sets fixed
  - farms can't change crop sets, but can change crop ratios
  - crop sets fixed $\rightarrow$ extract all needed fundamentals from data
  - set $\tau_{f,g}\tau_{f,g,i} = 1$, $\lambda_f = \bar{\lambda}$
  - solve for market-clearing $\{p_i\}_i$, $\{r_g\}_g$, $\bar{\lambda}$
- **reallocation 2**: equalize land and labor distortions, allow crop set changes
  - a farm can plant a crop it's not observed growing $\rightarrow$ counterfactual $z_{f,i}$, $\tau_{f,g}$, $\tau_{f,g,i}$ unknown

Reallocation Exercise Details

- **reallocation 1**: equalize land and labor distortions, crop sets fixed
  - farms can't change crop sets, but can change crop ratios
  - crop sets fixed $\rightarrow$ extract all needed fundamentals from data
  - set $\tau_{f,g}\tau_{f,g,i} = 1$, $\lambda_f = \bar{\lambda}$
  - solve for market-clearing $\{p_i\}_i$, $\{r_g\}_g$, $\bar{\lambda}$
- **reallocation 2**: equalize land and labor distortions, allow crop set changes
  - a farm can plant a crop it's not observed growing $\rightarrow$ counterfactual $z_{f,i}$, $\tau_{f,g}$, $\tau_{f,g,i}$ unknown
  - $\rightarrow$ need to parameterize **unconditional** z, τ distributions and calibrate to match observed **conditional** distributions

η \ σ	1.1	1.3	1.5	1.7	1.9	2.1	2.3
0.93	1588	1781	1961	2123	2267	2393	2503
0.83	1300	1472	1635	1787	1924	2047	2157
0.73	762	837	909	977	1040	1098	1152
0.63	277	284	289	294	299	303	306
0.53	139	140	140	141	141	142	142
0.43	99	99	99	99	99	100	100
0.33	79	79	80	80	80	80	80

(a) reallocation gain, %

back

η \ σ	1.1	1.3	1.5	1.7	1.9	2.1	2.3
0.93	1588	1781	1961	2123	2267	2393	2503
0.83	1300	1472	1635	1787	1924	2047	2157
0.73	762	837	909	977	1040	1098	1152
0.63	277	284	289	294	299	303	306
0.53	139	140	140	141	141	142	142
0.43	99	99	99	99	99	100	100
0.33	79	79	80	80	80	80	80

(a) reallocation gain, %

· misallocation estimates are always sensitive to calibrated concavity

back

Benchmark Exercise Sensitivity

η \ σ	1.1	1.3	1.5	1.7	1.9	2.1	2.3
0.93	1588	1781	1961	2123	2267	2393	2503
0.83	1300	1472	1635	1787	1924	2047	2157
0.73	762	837	909	977	1040	1098	1152
0.63	277	284	289	294	299	303	306
0.53	139	140	140	141	141	142	142
0.43	99	99	99	99	99	100	100
0.33	79	79	80	80	80	80	80

(a) reallocation gain, %

· misallocation estimates are always sensitive to calibrated concavity
  — firms can expand grow more easily in reallocation → greater gain

back

Benchmark Exercise Sensitivity

η \ σ	1.1	1.3	1.5	1.7	1.9	2.1	2.3
0.93	1588	1781	1961	2123	2267	2393	2503
0.83	1300	1472	1635	1787	1924	2047	2157
0.73	762	837	909	977	1040	1098	1152
0.63	277	284	289	294	299	303	306
0.53	139	140	140	141	141	142	142
0.43	99	99	99	99	99	100	100
0.33	79	79	80	80	80	80	80

(a) reallocation gain, %

η \ σ	1.1	1.3	1.5	1.7	1.9	2.1	2.3
0.93	+6	−6	−14	−21	−26	−30	−33
0.83	+3	−9	−18	−25	−30	−34	−38
0.73	−38	−43	−48	−52	−55	−57	−59
0.63	−23	−25	−27	−28	−29	−30	−31
0.53	−2	−3	−3	−4	−4	−4	−5
0.43	+2	+2	+2	+2	+2	+1	+1
0.33	+2	+2	+2	+2	+2	+2	+2

(b) single-product model error, %

- misallocation estimates are always sensitive to calibrated concavity
  - firms can expand grow more easily in reallocation → greater gain
- sign and magnitude of single-product model's error also depends on calibration

back

	main	within state	no split by season
multi-product:	294%	107%	314%
1-product:	212%	124%	260%

back

MECHANISM I: PRODUCT HETEROGENEITY AS FRICTIONS

- 1-product model misinterprets crop heterogeneity as frictions
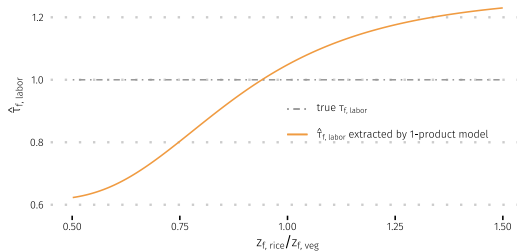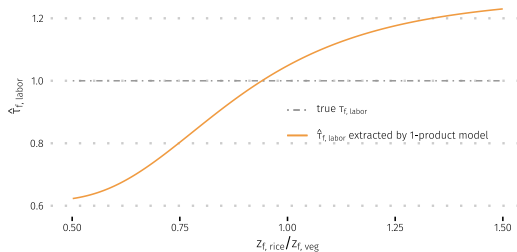
Mechanism I: Product Heterogeneity as Frictions

- 1-product model misinterprets crop heterogeneity as frictions
  - farm 1 draws high z in rice, produces it
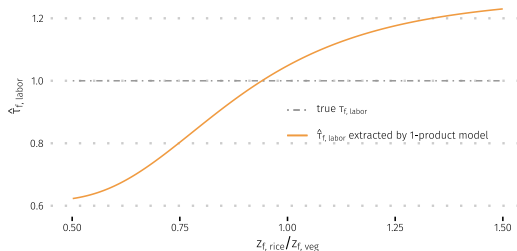
Mechanism I: Product Heterogeneity as Frictions

- 1-product model misinterprets crop heterogeneity as frictions
  - farm 1 draws high z in rice, produces it
  - farm 2 draws high z in vegetables, produces them

- 1-product model misinterprets crop heterogeneity as frictions
  - farm 1 draws high z in rice, produces it
  - farm 2 draws high z in vegetables, produces them
  - assume no frictions $\rightarrow \frac{\alpha_{g,rice}\eta(p_{rice}y_1)^\eta}{x_{1,g}} = \frac{\alpha_{g,veg}\eta(p_{veg}y_2)^\eta}{x_{2,g}}$ in multi-crop model

back

- 1-product model misinterprets crop heterogeneity as frictions
  - farm 1 draws high z in rice, produces it
  - farm 2 draws high z in vegetables, produces them
  - assume no frictions $\rightarrow$ $\frac{\alpha_{g,rice}\eta(p_{rice}y_1)^{\eta}}{x_{1,g}} = \frac{\alpha_{g,veg}\eta(p_{veg}y_2)^{\eta}}{x_{2,g}}$ in multi-crop model
  - 1-product model: $\frac{\alpha_{g,agg}\eta(p_{agg}y_1)^{\eta}}{x_{1,g}} \neq \frac{\alpha_{g,agg}\eta(p_{agg}y_2)^{\eta}}{x_{2,g}} \rightarrow$ imputes frictions

back

Mechanism I: Product Heterogeneity as Frictions

- 1-product model misinterprets crop heterogeneity as frictions
  - farm 1 draws high z in rice, produces it
  - farm 2 draws high z in vegetables, produces them
  - assume no frictions $\rightarrow \dfrac{\alpha_{g,rice}\eta(p_{rice}y_1)^{\eta}}{x_{1,g}} = \dfrac{\alpha_{g,veg}\eta(p_{veg}y_2)^{\eta}}{x_{2,g}}$ in multi-crop model
  - 1-product model: $\dfrac{\alpha_{g,agg}\eta(p_{agg}y_1)^{\eta}}{x_{1,g}} \neq \dfrac{\alpha_{g,agg}\eta(p_{agg}y_2)^{\eta}}{x_{2,g}} \rightarrow$ imputes frictions
  - $\rightarrow$ overstates misallocation

back

Mechanism I: Product Heterogeneity as Frictions

- 1-product model misinterprets crop heterogeneity as frictions
  - farm 1 draws high z in rice, produces it
  - farm 2 draws high z in vegetables, produces them
  - assume no frictions $\rightarrow \frac{\alpha_{g,rice}\eta(p_{rice}y_1)^\eta}{x_{1,g}} = \frac{\alpha_{g,veg}\eta(p_{veg}y_2)^\eta}{x_{2,g}}$ in multi-crop model
  - 1-product model: $\frac{\alpha_{g,agg}\eta(p_{agg}y_1)^\eta}{x_{1,g}} \neq \frac{\alpha_{g,agg}\eta(p_{agg}y_2)^\eta}{x_{2,g}} \rightarrow$ imputes frictions
  - $\rightarrow$ overstates misallocation

- **exercise to isolate**: apply 1-product model to counterfactual reallocation data generated by multi-product model

back

Mechanism II: TFP Dispersion

- 1-product model understates TFP dispersion

- 1-product model understates TFP dispersion
  - a farm has z_L in rice, z_H in vegetables

back

Mechanism II: TFP Dispersion

- 1-product model understates TFP dispersion
  - a farm has z_L in rice, z_H in vegetables
  - severe frictions $\rightarrow$ farm mostly grows rice $\rightarrow$ avg TFP $= z_M < z_H$

back

Mechanism II: TFP Dispersion

- 1-product model understates TFP dispersion
    - a farm has z_L in rice, z_H in vegetables
    - severe frictions $\rightarrow$ farm mostly grows rice $\rightarrow$ avg TFP $= z_M < z_H$
    - multi-crop model: remove frictions $\rightarrow$ farm switches to vegetables $\rightarrow$ avg TFP $= z_H$

back

- 1-product model understates TFP dispersion
    - a farm has z_L in rice, z_H in vegetables
    - severe frictions $\rightarrow$ farm mostly grows rice $\rightarrow$ avg TFP $= z_M < z_H$
    - multi-crop model: remove frictions $\rightarrow$ farm switches to vegetables $\rightarrow$ avg TFP $= z_H$
    - 1-product model: remove frictions $\rightarrow$ avg TFP still z_M

Mechanism II: TFP Dispersion

- 1-product model understates TFP dispersion
  - a farm has z_L in rice, z_H in vegetables
  - severe frictions $\rightarrow$ farm mostly grows rice $\rightarrow$ avg TFP $= z_M < z_H$
  - multi-crop model: remove frictions $\rightarrow$ farm switches to vegetables $\rightarrow$ avg TFP $= z_H$
  - 1-product model: remove frictions $\rightarrow$ avg TFP still z_M
  - $\rightarrow$ 1-product model understates misallocation

- 1-product model understates TFP dispersion
  - a farm has z_L in rice, z_H in vegetables
  - severe frictions $\rightarrow$ farm mostly grows rice $\rightarrow$ avg TFP $= z_M < z_H$
  - multi-crop model: remove frictions $\rightarrow$ farm switches to vegetables $\rightarrow$ avg TFP $= z_H$
  - 1-product model: remove frictions $\rightarrow$ avg TFP still z_M
  - $\rightarrow$ 1-product model understates misallocation
- **exercise to isolate**: treat farm-crops as separate farms for 1-product model

back

Mechanism III: Returns to Scale

- some products have higher returns to scale → some farms grow more in reallocation

Mechanism III: Returns to Scale

- some products have higher returns to scale → some farms grow more in reallocation
- 1-product model misses their greater potential growth

back

- some products have higher returns to scale $\rightarrow$ some farms grow more in reallocation
- 1-product model misses their greater potential growth
- $\rightarrow$ 1-product model understates misallocation

- some products have higher returns to scale → some farms grow more in reallocation
- 1-product model misses their greater potential growth
→ 1-product model understates misallocation
- **exercise to isolate**: rescale input elasticities in multi-product model s.t. all crops have the same returns to scale

back

MECHANISM IV: ENDOGENOUS PRODUCT CHOICE

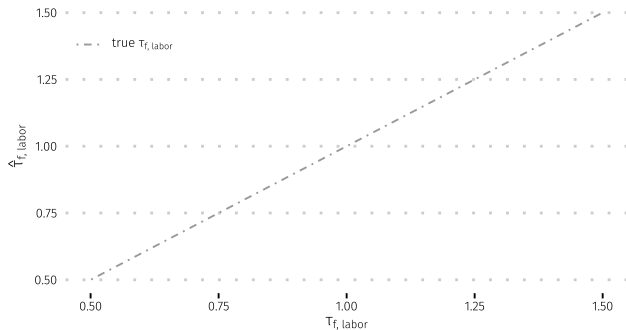· simulated multi-product farm as labor distortion $\tau_{f,labor}$ is varied:
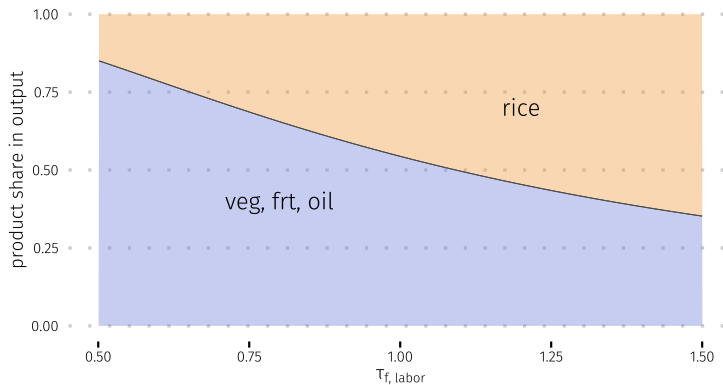
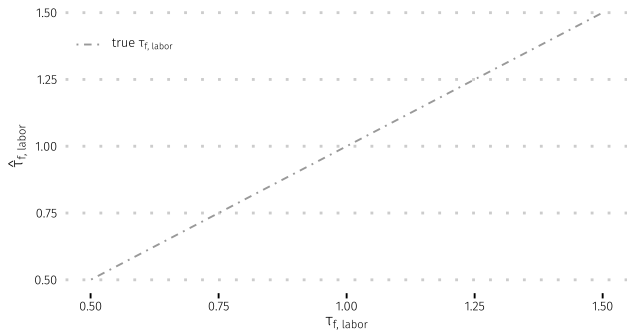· $\tau_{f,labor} \uparrow \ \rightarrow$ shift from labor-intensive vegetables to land-intensive rice

back

Mechanism IV: Endogenous Product Choice

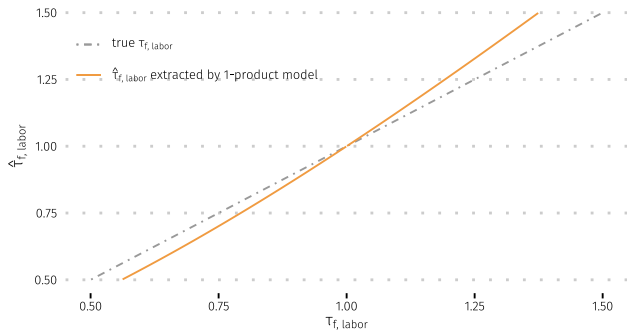- apply **single**-product model to extract frictions from simulated **multi**-product data

MECHANISM IV: ENDOGENOUS PRODUCT CHOICE

- apply **single**-product model to extract frictions from simulated **multi**-product data
- **multi-product model**: optimal product choice response to frictions
 - modest $\tau_{f,labor}$ increase → shift to land-intensive rice → hire even more land rel. to labor

Mechanism IV: Endogenous Product Choice

- apply **single**-product model to extract frictions from simulated **multi**-product data
- **multi-product model**: optimal product choice response to frictions
  - modest $\tau_{f,labor}$ increase $\rightarrow$ shift to land-intensive rice $\rightarrow$ hire even more land rel. to labor
- **1-product model**: high input ratio dispersion $\rightarrow$ infer large heterogeneity in frictions

back

Mechanism IV: Endogenous Product Choice

- apply **single**-product model to extract frictions from simulated **multi**-product data
- **multi-product model**: optimal product choice response to frictions
  - modest $\tau_{f,labor}$ increase $\rightarrow$ shift to land-intensive rice $\rightarrow$ hire even more land rel. to labor
- **1-product model**: high input ratio dispersion $\rightarrow$ infer large heterogeneity in frictions
  - $\rightarrow$ 1-crop model overstates misallocation

back
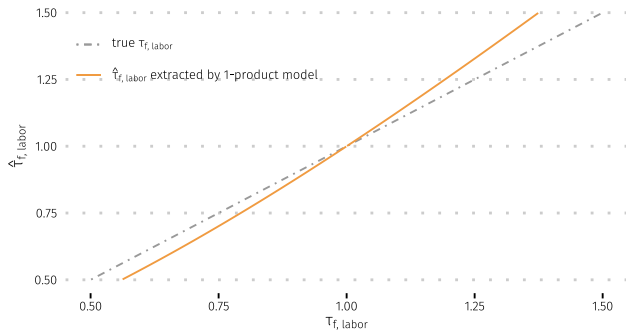
- **exercise to isolate**: prohibit farms in multi-product model to change product choice in counterfactuals
  - keep product sets fixed
  - keep input allocation across crops fixed: farm can choose $\sum_i x_{f,g,i}$ but $x_{f,g,i}$ gets a fixed share of total

back

REMOVE MORE DISTORTIONS → 1-PRODUCT MODEL OVERSTATES

- 1-product error when conducting increasingly expansive reallocations:

P = product heterogeneity as frictions
T = TFP dispersion
R = returns to scale
E = endogenous

— 1-product model error, %

share of maximum reallocation gain

back