Nutrition Demand, Subsistence Farming, and Agricultural Productivity

Stepan Gordeev
University of Rochester

Nutrition Demand, Subsistence Farming, and Agricultural Productivity

- Low-income countries dominated by unproductive agriculture
  - Malawi: **76%** of employment, **23%** of GDP in agriculture
  - low agricultural productivity critical for cross-country income differences
- Subsistence farming is common
  - Malawi: $\frac{3}{4}$ of households cultivate own land, **11%** sell most of the output

What are the implications of subsistence farming for aggregate agricultural productivity?

- what drives the production choices of subsistence farmers?

own nutritional needs + trade frictions

$\Downarrow$

farm targets family nutrition demand

$\Downarrow$

product choice $\neq$ comparative advantage

$\Downarrow$

agricultural productivity $\downarrow$

Preview

- **Model**:
  - farm-operating households
  - heterogeneous agricultural products
  - explicit caloric needs
- **Data**:
  - survey of households in Malawi
- **Farm Behavior in Model & Data**:
  - smallest farms focus on calories $\rightarrow$ specialize in staples
  - medium farms shift to dietary diversity $\rightarrow$ diversify production
  - large farms shift to manufactured goods $\rightarrow$ produce & sell marketable goods
- **Aggregate Productivity**:
  - trade costs $\downarrow$ s.t. farm commercialization 16% $\rightarrow$ 50%:

  $\rightarrow$ aggregate agricultural productivity $\uparrow$ 42%
    - $\sim$ half due to improved farm product choice
    - smallest farmers gain the most

Literature

economics literature:

Subsistence Farming &
Agricultural Productivity

- *Gollin and Rogerson (2014), Rivera-Padilla (2020), Sotelo (2020), Kebede (2020)*
  - region/village-level subsistence
  - remote regions trade little
  - → subsistence depresses agricultural productivity

nutrition literature:

Subsistence Farming &
Nutrition

- *Jones (2017), Sibhatu et al. (2015)*
  - smallholder farm biodiversity related to dietary diversity
  - especially with poor market access
  - → farm characteristics matter for nutritional outcomes

This Paper:

- explore farm-level subsistence, document scale-dependent product choice
- propose nutrition demand as explanation
- explicitly model caloric needs, explore role in aggr. agricultural productivity

DATA

Data

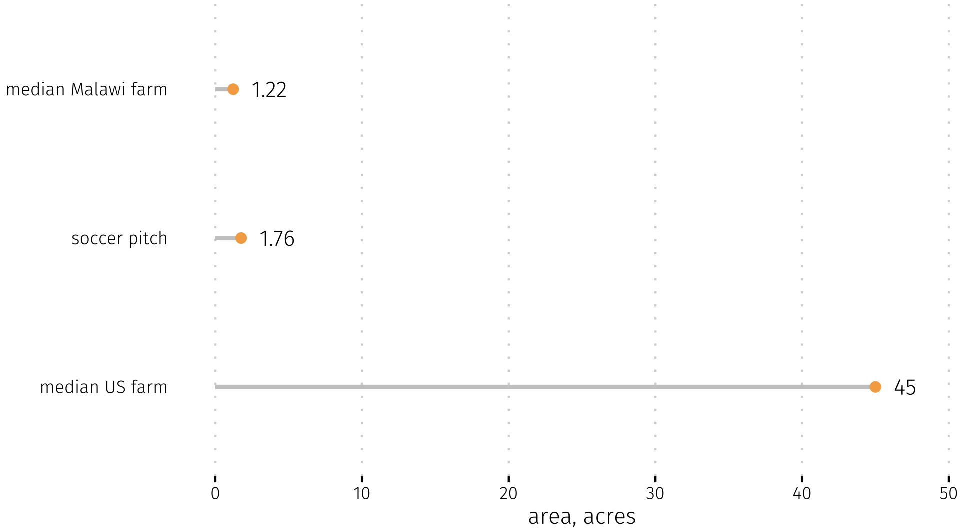
- Survey of households in Malawi (2016–2017)
  - government + World Bank
  - nationally representative, 12.5k HHs
  - 79% operate a farm ← **my sample**
- Detailed data on HH characteristics and activity
  - characteristics of family members → HH **kcal & nutrient requirements**
  - food consumed (past week) → HH **kcal & nutrient intakes**
  - agricultural inputs & outputs → HH **output & sales**
  - employment and non-farm enterprises → HH **non-farm income**

  ▶ *food* ▶ *output*
- Rescale HH kcal intake, output, income by HH kcal requirement

  → "per capita" measures, weighted by energy needs

Subsistence

Malawian Farms are Small



Semi-Subsistence is Common

- Farming is important for HH **consumption**:
  - foods self-produced (count): 24%
  - kcal self-produced: 36%
- $\frac{1}{2}$ of farms **sell none** of their output
  - avg share of output sold: 16%

Model

Model Objectives

- Is farm subsistence important for aggregate productivity?
  - quantitative counterfactual experiments
- Model that can generate HH-level subsistence
  - farm-operating HH
  - consumption and production decisions are non-separable
  - preferences to generate realistic food consumption
- Compare predicted farm behavior to data
  - extract predictions on farm product choice

Model: HH Problem

$$\max \left((1-\varphi_m) \left(\sum_{i=1}^{n} \varphi_i c_{h,i}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}\frac{\gamma-1}{\gamma}} + \varphi_m c_{h,m}^{\frac{\gamma-1}{\gamma}} \right)^{\frac{\gamma}{\gamma-1}} - \underbrace{f\left(\sum_{i=1}^{n} c_{h,i} k_i,\ K_{req,h} \right)}_{\text{kcal deviation penalty}}$$

$$\sum_{i=1}^{n} \frac{x_{h,i}}{z_{h,i}} \leq L_h$$

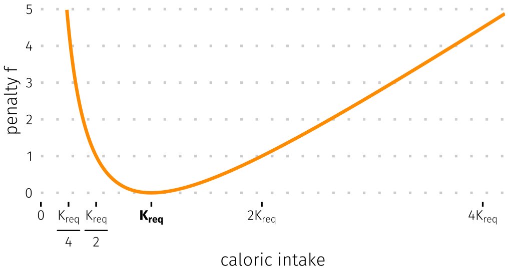
$$\sum_{i=1}^{n} x_{h,i}^{p} p_i d + p_m c_{h,m} \leq \sum_{i=1}^{n} x_{h,i}^{s} \frac{p_i}{d} + wN_h$$

- HH h consumes n foods $\{c_{h,i}\}_{i=1}^{n}$ and a manufactured good $c_{h,m}$
  - foods differ in kcal content k_i
- HH prefers $\underbrace{\sum_{i=1}^{n} c_{h,i} k_i}_{\text{caloric intake}} \approx \underbrace{K_{req,h}}_{\text{caloric requirement}}$
- For each good i, HH h can $\begin{cases} \text{produce } x_{h,i} \text{ with productivity } z_{h,i} \\ \text{purchase } x_{h,i}^{p} \text{ or sell } x_{i}^{s} \text{ at } p_i \text{ with trade cost } d > 1 \end{cases}$

Caloric Deviation Penalty f

- Caloric deviation penalty fn (▶ *properties*):

$$f\left(\sum_i c_{h,i}k_i, K_{req,h}\right) = \psi \left(\frac{\sum_i c_{h,i}k_i - K_{req,h}}{K_{req,h}}\right)^2 \frac{K_{req,h}}{\sum_i c_{h,i}k_i}$$



- As $\psi \to 0$, preferences $\to$ pure CES

Calibration: Agricultural Goods

- 6 agricultural goods commonly produced and consumed
  - ▶ *list*

Agricultural goods are heterogeneous in

1. taste weight φ_i
   - consumption FOC → estimable expression → identify $\{\varphi_i\}_i$, σ
     - ▶ *estimation*
2. kcal density k_i
   - food composition tables
3. distribution of productivity across households $z_{h,i}$
   - GAEZ: crop-location-level attainable yield predicted by local soil and climate
   - fit a cross-household lognormal distribution for each crop in GAEZ

HH Heterogeneity

1. Non-farm income wN_h distribution
   - lognormal + mass at $wN_h = 0$
2. Land L_h distribution
   - lognormal
   - mean: target avg $\frac{\text{kcal intake}}{\text{kcal requirement}}$ to ensure realistic scale
   - variance: target output value variance
3. Good productivity $z_{h,i}$ distributions
   - HHs take independent draws for each good

General Equilibrium: Manufacturer

- Representative competitive manufacturer:

$$\max \; p_m Y_m - wN$$

s.t.

$$Y_m = z_m \underbrace{N}_{\sum_h N_h}$$

  - normalize $p_m = z_m = 1$
  - zero profits $\to w = 1$

General Equilibrium: Agricultural Goods

- Solve for agricultural prices $\{p_i\}_i$ s.t. edible good markets clear:

$$\frac{1}{d}\sum_h x^s_{h,i} = d\sum_h x^p_{h,i} \quad \forall i$$

- Tobacco market doesn't need to clear
  - data: tobacco accounts for 60% of Malawi's exports
  - tobacco traded internationally at exogenous $\bar{p}_t$
  - some manufactured good is imported to balance the trade:

$$\underbrace{\bar{p}_{\text{tobacco}}\left(\frac{1}{d}\sum_h x^s_{h,\text{tobacco}} - d\sum_h x^p_{h,\text{tobacco}}\right)}_{\text{tobacco exports}} = \underbrace{p_m\left(\sum_h c_{h,m} - Y_m\right)}_{\text{manuf. good imports}}$$

Estimation

parameter		value	moment/source	**data** moment	**model** moment
Distributions					
$\mathbb{E}(\log L_h)$		-15	avg $K_{in,h}/K_{req,h}$	1.036	0.902
$V(\log L_h)$		1.5	$V(\log \text{output}_h)$	1.528	1.385
$P(N_h = 0)$		0.112	$P(\text{non-farm income}_h = 0)$	0.112	0.117
$V(\log N_h \mid N_h > 0)$		2.103	$V(\log \text{non-farm income}_h)$	2.103	1.924
Parameters					
σ	*(EoS across foods)*	0.75	estimated	—	—
γ	*(EoS between food & manuf.)*	1	—	—	—
d	*(agricultural trade cost)*	1.75	avg share sold	0.159	0.203
ψ	*(kcal deviation penalty)*	0.5	output elasticity of K_{in}	0.091	0.124
Good characteristics					
φ_m	*(manuf. taste weight)*	0.5	$\frac{\text{aggr. non-farm income}}{\text{aggr. farm output}}$	1.539	1.632
$\bar{p}_{\text{tobacco}}/p_{\text{maize}}$		5.4	aggr. tobacco output share	0.091	0.094

▶ *solution strategy*

Farm Behavior in Model and Data

Calories Skew Consumption: Model

- Consider the problem of a household
- Suppose $\psi > 0$ (benchmark), $\sum_i c_i k_i < K_{req}$

$$\underbrace{MU_i^{CES}(c_i)}_{\downarrow} - \underbrace{k_i f_1\left(\sum_i c_i k_i,\ K_{req}\right)}_{\leq 0,\ \downarrow} = MC_i \qquad (c_i \text{ FOC})$$

- $k_i = 0 \ \rightarrow\ k_i f_1\left(\sum_i c_i k_i,\ K_{req}\right) = 0$
- $k_i > 0 \ \rightarrow\ k_i f_1\left(\sum_i c_i k_i,\ K_{req}\right) \downarrow \ \rightarrow\ MU_i^{CES} \downarrow \ \rightarrow\ c_i \uparrow$
- When energy intake $<$ requirement, consume more efficient sources of calories

- Why *these* non-homothetic preferences?
  - endogenous predictions on allocation across agricultural goods
  - produces structural transformation forces *across* sectors and *within* agriculture
  - $\rightarrow$ predictions on kcal-diversity tradeoff in consumption and on farm product choice

- Extreme poverty limit:

$$\max \left((1-\varphi_m) \left(\sum_{i=1}^{n} \varphi_i c_{h,i}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}\frac{\gamma-1}{\gamma}} + \varphi_m c_{h,m}^{\frac{\gamma-1}{\gamma}} \right)^{\frac{\gamma}{\gamma-1}} - f\left(\sum_{i=1}^{n} c_{h,i} k_i,\ K_{req,h} \right)$$

$$\Big\downarrow \; L_h, wN_h \to 0$$

$$\max \sum_{i=1}^{n} c_{h,i} k_i$$

- In the limit, goods are perfect substitutes → consume at most 2:
  - $\textbf{arg max}_i\, k_i/p_i$ (good with lowest price of 1 kcal)
  - $\textbf{arg max}_i\, k_i z_{h,i}$ (good with highest kcal productivity)

Farm Size ↑ → Shift From Dietary Energy to Diversity: Model & Data

	log kcal intake			food diversity ▶ *def* ▶ *nutrients*		
	model: $\psi = 0$	model: benchmark	data	model: $\psi = 0$	model: benchmark	data
log output	0.663	0.124	0.091***	-0.082	0.428	0.395***
	(0.002)	(0.001)	(0.005)	(0.002)	(0.002)	(0.034)
log non-farm income	0.354	0.084	0.063***	0.041	0.396	0.857***
	(0.004)	(0.001)	(0.004)	(0.002)	(0.002)	(0.033)
N	29,168	33,613	8,674	29,168	33,613	8,675
Adj. R^2	0.919	0.393	0.063	0.085	0.762	0.131

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

- $\psi = 0$ (pure CES): relative consumptions invariant to size/income
  - kcal intake ↑ proportionally to total shadow income, diversity constant
- $\psi > 0$ (benchmark): reallocate resources from calories to diversity as size/income ↑
  - kcal intake ↑ little with total shadow income, diversity ↑

Farm Sales are Specialized: Model & Data

- Model: each farm sells at most one good
  - the revenue-maximizing one: $\arg\max_i p_i z_{h,i}$
  - can produce more goods for own consumption
- Data: sales are specialized compared to overall production
  - **69%** sell just 1 good, only **9%** produce just 1 good
  - on avg, top good accounts for **91%** in sales but **67%** in output

Lower Trade Costs $\longrightarrow$ All Specialize: Model & Data

- Model: $d \downarrow \rightarrow$ specialize production
  – below some cutoff $\tilde{d}_h$, HH h only produces the revenue-maximizing good
- Data:
  – HHs with better market access specialize production
  ▶ *table*

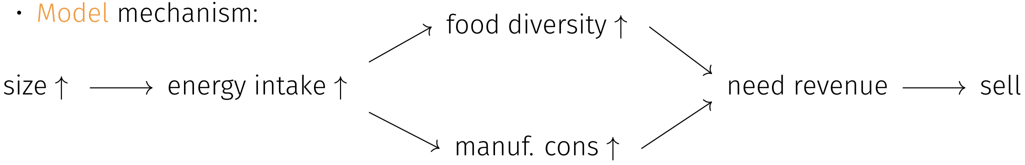
Large Farms Sell More: Model & Data

- Larger farms are more active sellers:

output quartile	sold output share		fraction sellers	
	model	data	model	data
1	<1%	13%	<1%	14%
4	67%	31%	>99%	77%

  – Pure CES Model: no scale dependence in selling behavior

- Model mechanism:



- What's missing from the model? *Risk*
  – volatile harvest/prices → specializing in cash crops raises starvation risk

  → risk should smooth the size-selling relationship

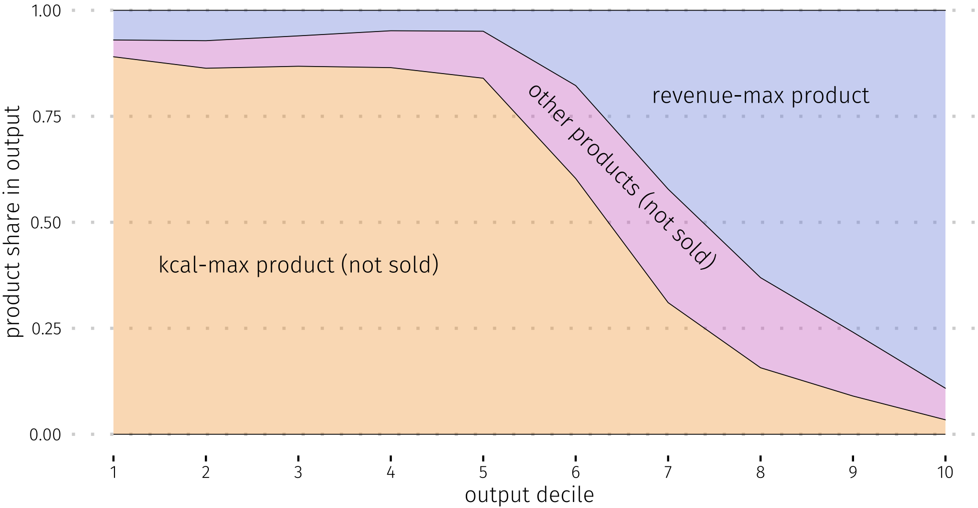
Smallest Farms Specialize Production: Model

- Extreme poverty limit: production converges to perfect specialization

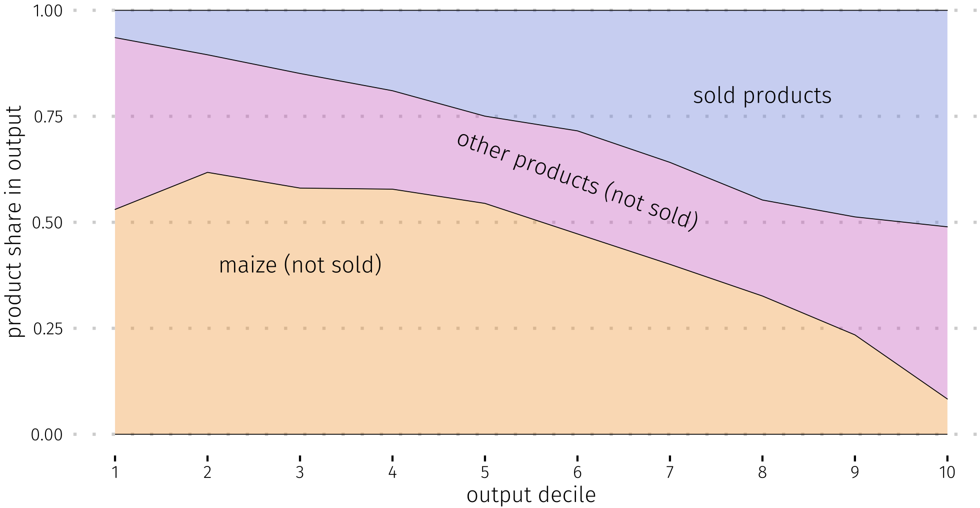
$$\lim_{L_h, wN_h \to 0} (\#\text{ goods produced}) = 1$$

  - maximize kcal intake $\rightarrow$ specialize consumption $\rightarrow$ specialize production
- There is a cutoff trade cost $\bar{d}_h$:
  - ▶ $\bar{d}_h$
  - if $d < \bar{d}_h$, produce the most revenue-productive good: $\arg\max_i p_i z_{h,i}$
  - if $d > \bar{d}_h$, produce the most kcal-productive good: $\arg\max_i k_i z_{h,i}$

Larger Farms Diversify, Shift to Marketable Goods: Model



Larger Farms Diversify, Shift to Marketable Goods: Data



▶ *diversity* ▶ *product frequency*

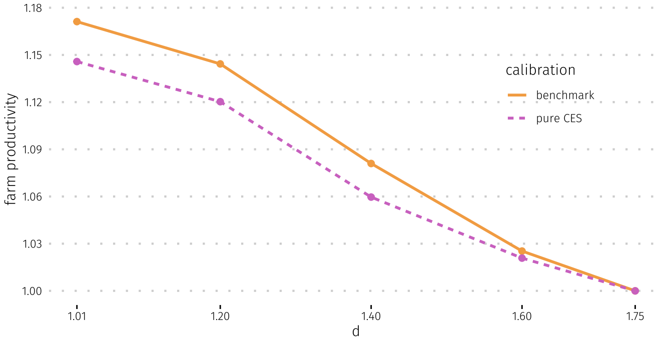
Aggregate Productivity

Aggregate Counterfactuals

- Model: trade costs $\rightarrow$ nutrition demand influences product choice
- How relevant is it for aggregate agricultural productivity?
- Conduct counterfactual reductions in domestic agricultural trade costs ($d \downarrow$)
  - measure effect on productivity

Trade Costs ↓ → Farm Productivity ↑

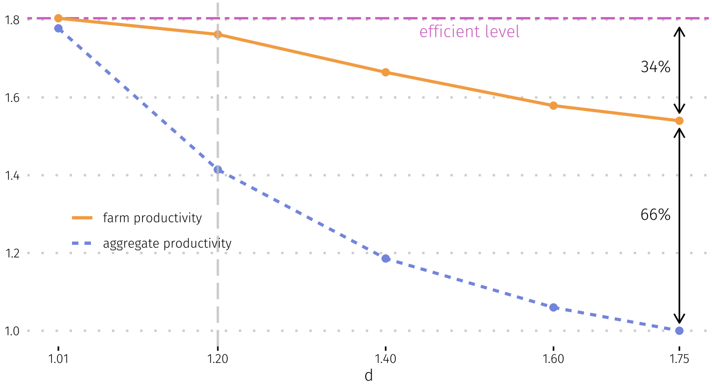
- Total "farm-gate" productivity: $\frac{\sum_h \sum_i x_{h,i} p_i}{\sum_h L_h}$
  - does not count mechanical losses from d
  - deflate using chain-weighting between consecutive d values



- $d \downarrow$ → farmers choose products they are productive at → farm productivity ↑

Trade Costs ↓ → Aggregate Productivity ↑

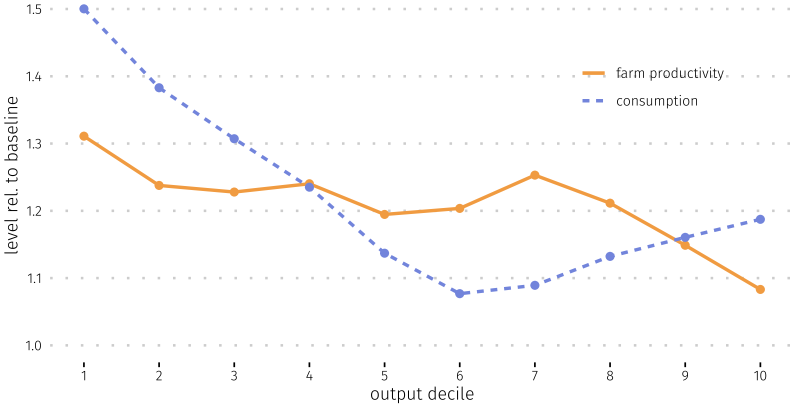
- Compare "farm-gate" production to final consumption ("aggregate productivity")
  - farm production only accounts for product choice changes
  - final consumption also accounts for mechanical losses from d



- $d \to 1$: aggr. productivity ↑ **78%** ($\frac{1}{3}$ due to product choice)
- $d \downarrow$ s.t. avg share sold 16% → 50%: aggr. productivity ↑ **42%** ($\frac{1}{2}$ due to product choice)

Trade Costs ↓ → Heterogeneous Effects in Farm Size

- $d \downarrow$ s.t. avg share sold 16% → 50%:
  - farm productivity: small ↑ the most, large ↑ the least
  - consumption: small ↑ the most, medium ↑ the least



Conclusion

Conclusion

- Subsistence farmer nutrition demand $\rightarrow$ subsistence farm production decisions
  - smallest farms specialize in calorie consumption & production
  - medium farms diversify diet & production
  - largest farms become market-oriented
- Farm production decisions $\rightarrow$ aggregate agricultural productivity $\downarrow$
  - trade costs $\downarrow$ s.t. farms just leave subsistence $\rightarrow$ agric. productivity $\uparrow$
  - half because improved product choice, half because less is lost to trade cost

Future Research

- Analyze government programs targeting smallholder farmers
  - smallholder farmer support is central to public policy in poor countries
  - e.g. encourage staples, biodiversity, or cash crops?
  - framework well suited for predicting nutritional, economic outcomes
- Interaction of nutrition with labor choice
  - allocate labor between own farm, other farms, and non-agricultural sector
  - study structural transformation between sectors and within agriculture jointly

Food

- Food consumption
  - HH-product consumption in past week
- Food composition
  - product nutritional contents from Malawian and Tanzanian food composition tables
  - → obtain HH-level calorie and nutrient intakes
- Nutritional needs
  - kcal requirement for each individual (age, sex) from FAO's Human Energy Requirements
  - nutrient daily recommended allowances from Dietary Guidelines for Americans
  - → obtain HH-level calorie and nutrient requirements

back

Output & Income

- Farm output
  - HH-product output in past year
  - sales if any
  - total farm output: quantities weighted by median sale price
- Non-farm income
  - income from employment and non-farm enterprises

back

$f\left(\sum_i c_i k_i, K_{req}\right)$ Properties

Properties:

1. $f(bK_{in}, bK_{req}) = f(K_{in}, K_{req})$ (homogeneity of deg. 0)
2. $f(bK_{req}, K_{req}) = f\left(\frac{K_{req}}{b}, K_{req}\right)$ (symmetry around K_{req} in ratios)
3. $\min_{K_{in}>0} f(K_{in}, K_{req}) = f(K_{req}, K_{req}) = 0$ (minimum and zero if eat K_{req})
4. $f_{11}(K_{in}, K_{req}) = \frac{2\psi K_{req}}{K_{in}^3} > 0$ (convex in intake)

back

Agricultural Goods Used in Calibration

- Selected goods:
  1. maize
  2. pigeonpea
  3. groundnut
  4. tomato
  5. soybean
  6. tobacco
- These goods account for, on average,
  - 70% of HH output market value
  - 43% of HH food consumption market value

back

Estimation of EoS and Taste Weights

- Consumption FOC $\rightarrow$ if i is produced by h, then

$$\log c_{h,i} \approx \underbrace{\gamma(\log p_m - \log \frac{\varphi_m}{1-\varphi_m})}_{\text{constant}}$$

$$\underbrace{+\log c_{h,m} - \sigma \log \lambda_h + \gamma \log \mu_h + \frac{\gamma - \sigma}{\sigma - 1} \log \left(\sum_{j=1}^{n} \varphi_j^{\sigma} \left(\eta_{h,j} + f_1 \left(\sum c_{h,j} k_j,\ K_{req,h} \right) k_j \right)^{1-\sigma} \right)}_{\text{HH-produced FE}}$$

$$+ \underbrace{\sigma \log \varphi_i}_{\text{good FE}} + \sigma \underbrace{\log z_{h,i}}_{X_{1,h,i}} - \underbrace{k_i z_{h,i}}_{X_{2,h,i}} \cdot \sigma \underbrace{\frac{f_1 \left(\sum_j c_{h,j} k_j, K_{req,h} \right)}{\lambda_h}}_{\text{HH-produced FE2}}$$

- Analogous expression if i is purchased by h ($X_{1,h,i} = \log p_i d_h$, $X_{2,h,i} = \frac{k_i}{p_i d_h}$)

$\rightarrow$ Can estimate food taste weights $\{\varphi_i\}_i$, elasticity of substitution across foods σ

back

Solution Strategy

- Each HH needs to be solved as an independent problem
- Simulation
  - 500 HH types take independent draws form N_h and $\{z_i\}_i$ distributions
  - within each type, approximate the L_h distribution using 80 sub-types

  $\rightarrow$ solve economy with 40,000 HHs

back

Cutoff Trade Cost $\bar{d}$

$$\bar{d}_h = \sqrt{\frac{\max_i p_i z_{h,i}}{\min_i p_i / k_i \cdot \max_i k_i z_{h,i}}}$$

back

Production Diversity

- Production Diversity = Inverse Simpson Index

$$\text{Production Diversity}_h = \left(\sum_{i=1}^{n} \left(\frac{\text{output}_{h,i}}{\sum_{j=1}^{n} \text{output}_{h,j}} \right)^2 \right)^{-1}$$

where n is the total number of agricultural products, $\text{output}_{h,i}$ is the market value of product i produced by h's farm.

- Simpson Index: sum of squared product output shares within farm's output
  - same as HHI
  - interpretation: probability that two random dollars of output come from the same product
- Inverse Simpson Index $= \frac{1}{SI}$, commonly used in measuring species diversity

back

Food Diversity

- Food Diversity = Inverse Simpson Index

$$\text{Food Diversity}_h = \left(\sum_{i=1}^{n} \left(\frac{\text{food quantity}_{h,i} \times \text{median purchase price}_i}{\sum_{j=1}^{n} \text{food quantity}_{h,j} \times \text{median purchase price}_j} \right)^2 \right)^{-1}$$

  where h is the HH index, n is the total number of distinct foods in the dataset.
- Simpson Index: sum of squared food shares within HH's consumption
  - same as HHI
  - interpretation: probability that two random dollars of (shadow) food expenditure come from the same product
- Inverse Simpson Index $= \frac{1}{SI}$, commonly used in measuring species diversity

back

Nutrient Richness

	NRF9		NRF9.3	
	(1)	(2)	(3)	(4)
log output	17.046***	5.695***	−13.296***	−13.400***
	(0.964)	(0.724)	(3.326)	(3.358)
log non-farm income	10.285***	2.441***	−7.257**	−7.305**
	(0.792)	(0.603)	(3.898)	(3.548)
log kcal intake		124.025***		0.550
		(2.282)		(26.234)
N	8,675	8,674	8,675	8,674
Adj. R^2	0.054	0.451	0.002	0.002

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

- NRF9: sum of daily intakes (relative to recommended level) of 9 nutrients
- NRF9.3: subtracts the relative excessive consumption of 3 disqualifying nutrients

back

Lower Trade Costs → All Specialize: Data

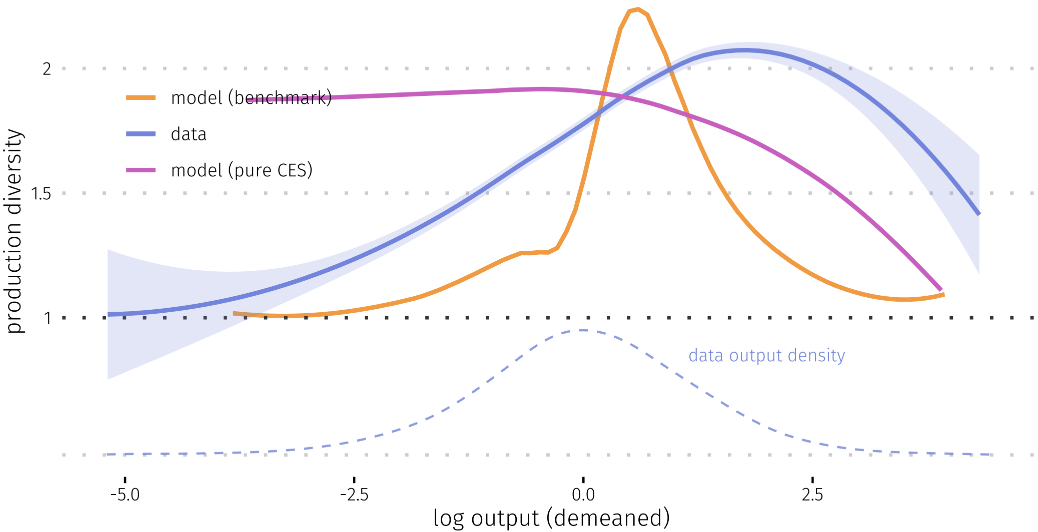
	production diversity	
sold output share	−0.044***	
	(0.016)	
$\mathbb{1}$[good mkt access]		−0.164***
		(0.018)
N	4,042	8,675
Adj. R^2	0.025	0.099

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

Note. Controls: log output, log non-farm income.

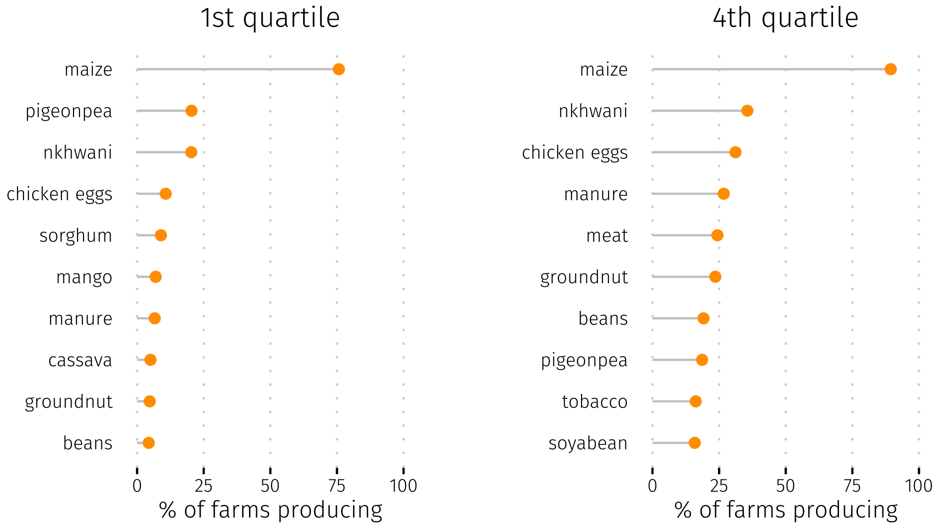
back

Larger Farms Diversify, Largest Specialize: Model & Data



▶ production diversity back

Product Frequency by Size: Data



back