14.581: International Trade
— Lecture 6 —
Ricardian Model (Empirics II)

Plan of Today's Lecture

1. Empirical applications of Ricardian assignment models:
   1. Testing Ricardian comparative advantage: Costinot and Donaldson (2012)
   2. Gains from economic integration: Costinot and Donaldson (2016)
2. Conclusion

Introduction

- We now consider empirical applications of particular sort of Ricardian model that one could call an "assignment model" (see Costinot-Vogel, 2015 survey, and a bit more later in this course)
- We will place particular emphasis on settings in which:
  - Each fundamental production unit uses one factor (land). This is of course what makes these Ricardian assignment models.
  - But the observable production units are comprised of many such fundamental production units, each of which is unique (i.e. the type of land is different).
  - Fundamental production units combine as perfect substitutes to generate output at the observable level.

A Key Empirical Challenge

- Suppose that different factors of production specialize in different economic activities based on their relative productivity differences
- Following Ricardo's famous example, if English workers are relatively better at producing cloth than wine compared to Portuguese workers:
  - England will produce cloth
  - Portugal will produce wine
  - At least one of these two countries will be completely specialized in one of these two sectors
- Accordingly—as discussed in Lecture 5—the key explanatory variable in Ricardo's theory, relative productivity, cannot be directly observed

How Can One Solve This Identification Problem?

Existing Approach

- Previous identification problem is emphasized by Deardorff (1984) in his review of empirical work on the Ricardian model of trade
- A similar identification problem arises in labor literature in which self-selection based on CA is often referred to as the Roy model
  - Heckman and Honore (1990): if general distributions of worker skills are allowed, the Roy model has no empirical content
- One potential solution:
  - Make (fundamentally untestable) functional form assumptions about distributions
  - Use these assumptions to relate observable to unobservable productivity,
- Examples:
  - In a labor context: Log-normal distribution of worker skills
  - In a trade context: Fréchet distributions across countries and industries (Costinot, Donaldson and Komunjer, 2012)

How Can One Solve This Identification Problem?

- We'll look at Costinot and Donaldson (2012, 2016) who focus on sector in which scientific knowledge of how essential inputs map into outputs is well understood: **agriculture**
- As a consequence of this knowledge, agronomists predict the productivity of a 'field' (small parcel of land) if it were to grow any one of a set of crops
- In this particular context, we know the productivity of a 'field' in *all* economic activities, not just those in which it is currently employed

Basic Theoretical Environment

- The basic environment is the same as in the purely Ricardian part of Costinot (ECMA, 2009)
- Consider a world economy comprising:
  - $c = 1, ..., C$ countries
  - $g = 1, ..., G$ goods [crops in our empirical analysis]
  - $f = 1, ..., F$ factors of production ['fields', or grid cells, in our empirical analysis]
- Factors are immobile across countries, perfectly mobile across sectors
  - $L_{cf} \geq 0$ denotes the inelastic supply of factor f in country c
- Factors of production are perfect substitutes within each country and sector, but vary in their productivities $A_{cf}^{g} \geq 0$

Cross-Sectional Variation in Output

- Total output of good g in country c is given by

$$Q_c^g = \sum_{f=1}^{F} A_{cf}^g L_{cf}^g$$

- Take producer prices $p_c^g \geq 0$ as given and focus on the allocation that maximizes total revenue at these prices
- Assuming that this allocation is unique, can express output as

$$Q_c^g = \sum_{f \in \mathcal{F}_c^g} A_{cf}^g L_{cf} \tag{1}$$

where $\mathcal{F}_c^g$ is the set of factors allocated to good g in country c:

$$\mathcal{F}_c^g = \{ f = 1, ...F | \, A_{cf}^g / A_{cf}^{g'} > p_c^{g'} / p_c^g \text{ if } g' \neq g \} \tag{2}$$

Data Requirements

- CD (2012; AER P&P)'s test of Ricardo's ideas requires data on:
  - Actual output levels, which we denote by $\widetilde{Q}_c^g$
  - Data to compute predicted output levels, which we denote by Q_c^g
- By equations (1) and (2), we can compute Q_c^g using data on:
  - Productivity, A_{cf}^g, for all factors of production f
  - Endowments of different factors, L_{cf}
  - Producer prices, p_c^g

Output and Price Data

- Output ($\widetilde{Q}_c^g$) and price (p_c^g) data are from FAOSTAT
- Output is equal to quantity harvested and is reported in tonnes
- Producer prices are equal to prices received by farmers net of taxes and subsidies and are reported in local currency units per tonne
- In order to minimize the number of unreported observations, our final sample includes 55 countries and 17 crops
- Since Ricardian predictions are cross-sectional, all data are from 1989

Productivity Data

- Global Agro-Ecological Zones (GAEZ) project run by FAO
  - Used in Nunn and Qian (2011) as proxy for areas where potato could be grown
- Productivity (A^g_{cf}) data for:
  - 154 varieties grouped into 25 crops c (though only 17 are relevant here)
  - All 'fields' f (5 arc-minute grid cells) on Earth
- Inputs:
  - Soil conditions (8 dimensional vector)
  - Climatic conditions (rainfall, temperature, humidity, sun exposure)
  - Elevation, average land gradient.
- Modeling approach:
  - Entirely 'micro-founded' from primitives of how each crop is grown.
  - 64 parameters per crop, each from field and lab experiments.
  - Different scenarios for other human inputs. We use 'mixed, irrigated'

Example: Relative Wheat-to-Sugar Cane Productivity

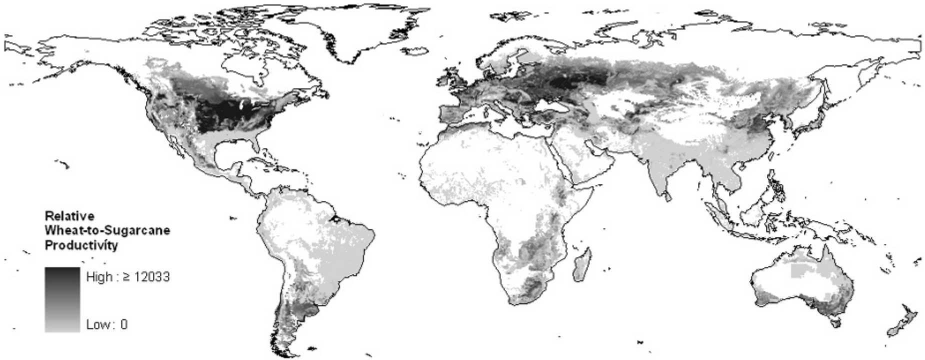


Figure 1: An Example of Relative Productivity Differences. Notes: Ratio of productivity in wheat (in tonnes/ha) relative to productivity in sugarcane (in tonnes/ha). Areas shaded white have either zero productivity in wheat, or zero productivity in both wheat and sugarcane. Areas shaded dark, with the highest value (">12,033"), have zero productivity in sugarcane and strictly positive productivity in wheat. Source: GAEZ project.

Empirical Strategy

- To overcome identification problem highlighted by Deardorff (1984) and Heckman and Honore (1990), CD (2012) follow two-step approach:
  1. We use the GAEZ data to predict the amount of output (Q_c^g) that country c should produce in crop g according to (1) and (2)
  2. We regress observed output ($\widetilde{Q}_c^g$) on predicted output (Q_c^g)
- Like in HOV literature, they consider test of Ricardo's theory of comparative advantage to be a success if:
  - The slope coefficient in this regression is close to unity
  - The coefficient is precisely estimated
  - The regression fit is good
- Compared to HOV literature, CD (2012) estimate regressions in logs:
  - Core of theory lies in how *relative* productivity predict *relative* quantities
  - *Absolute* levels of output are far off because more uses of land than 17 crops

Results

Table 1: Comparison of Actual Output to Predicted Output

Dependent variable:	log (output)				
	(1)	(2)	(3)	(4)	(5)
log (predicted output)	0.212*** (0.057)	0.244*** (0.074)	0.096** (0.038)	0.143** (0.062)	0.273*** (0.074)
sample	all	all	all	major countries	major crops
fixed effects	none	crop	country	none	none
observations	349	349	349	226	209
R-squared	0.06	0.26	0.54	0.04	0.07

Notes: All regressions include a constant. Standard errors clustered by country are in parentheses. ** indicates statistically significant at 5% level and *** at the 1% level.

Recall a Classic Question in this Course: How Large are the Gains from Economic Integration?

- Regions of the world, both across and within countries, appear to have become more economically integrated with one another over time.
- Two natural questions arise:
  1. How large have been the gains from this integration?
  2. How large are the gains from further integration?

How Large are the Gains from Economic Integration?

- Deardorff (1984) identification problem arises again.
- Fundamental challenge lies in predicting how local markets would behave under **counterfactual scenarios** in which they become more or less integrated with rest of the world.
- In a Trade context, counterfactual scenarios typically involve the **reallocation of multiple factors of production** towards different economic activities.
- Hence researcher requires knowledge of **counterfactual productivity** of factors if they were employed in sectors in which producers are currently, and deliberately, not using them.
    - Any study of the gains from economic integration needs to overcome this **identification problem**.

How to Overcome Identification Problem?

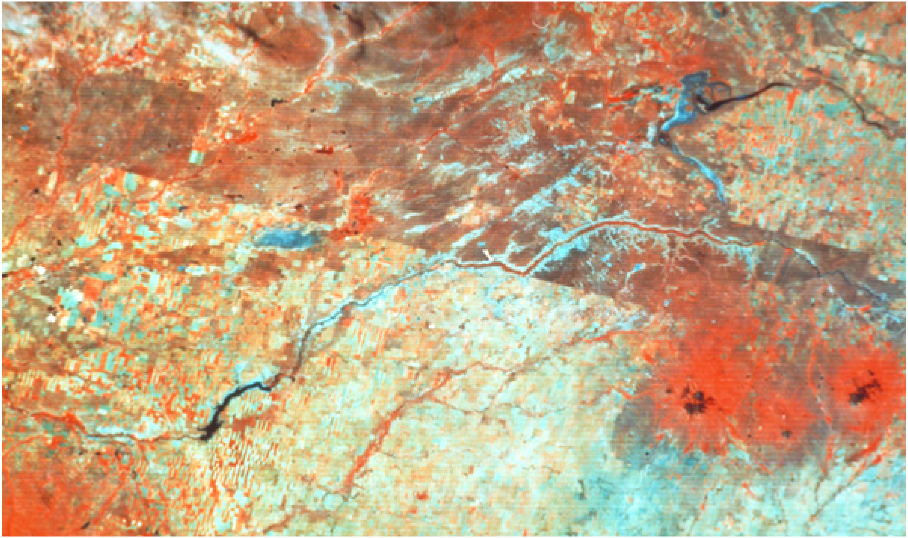
- Four main approaches in the literature:
  - "Reduced form" approach (e.g. Frankel and Romer 1999): knowledge of CF obtained by observing behavior of "similar but open" countries (Lecture 4).
  - "Autarky" approach (e.g. Bernhofen and Brown, 2005): autarky prices, when observed, are useful (Lecture 4).
  - "Sufficient statistic" approach (e.g. Chetty, 2009): knowledge of CF technologies unnecessary (for small changes) because gains from reallocation of production are second-order at optimum.
  - "Structural" approach (e.g. Eaton and Kortum 2002): knowledge of CF obtained by extrapolation based on (untestable) functional forms (Lecture 3).
- Basic idea of CD (2016):
  - Develop new structural approach with weaker need for extrapolation by functional form assumptions by drawing on agronomic knowledge in agricultural sector (from CD2012).

CD (2016): Method

- Consider a panel of $\sim$1,500 U.S. counties from 1880 to 1997.
  - Choose US for long sweep of high-quality, comparable micro-data from important agricultural economy.
- Use Roy/Ricardian model + FAO data to construct PPF in each county.
- Then two steps:
  1. **Measuring Farm-gate Prices:**
     - We combine Census data on output and PPF to infer prices that farmers in local market i appear to have been facing.
  2. **Measuring Gains from Integration:**
     - We compute the spatial distribution of price gaps between U.S. counties and New York/World in each year.
     - We then ask: "For any period t, how much higher (or lower) would the total value of US agricultural output in period t have been if price gaps were those from 1997 rather than those from period t?"

Inferring farm-gate prices

Sometimes effects are clearly visible (eg US-Canada 49th parallel border)



CD (2016): Results

- Farm-gate price estimates look sensible:
  - State-level price estimates correlate well with state-level price data.
- How large have been the gains that arose as counties became increasingly integrated?
  - eg 1880-1920: **2.3 % growth (in agricultural GDP) per year**
  - same order of magnitude as productivity growth in agriculture

A Few Caveats to Keep in Mind

1. FAO data are only available in 2011.
    - Extrapoloation necessary when going back in time.
    - To do so CD (2016) allow unrestricted county-crop-year specific productivity shocks.
2. Highest resolution output data available (from Census) is at county-level.
    - So direct predictions from high-resolution FAO model, pixel by pixel, are not testable.
3. Land (though heterogeneous) is the only factor of production.
    - Should think of land as "equipped" land

Basic Environment

- Many 'local' markets $i \in \mathcal{I} \equiv \{1, ..., I\}$ in which production occurs
- One 'wholesale' market in which goods are sold (New York/World)
- Only factors of production are fields $f \in \mathcal{F}_i \equiv \{1, ..., F_i\}$
  - $V_i^f \geq 0$ denotes the number of acres covered by field f in market i
- Fields can be used to produce multiple goods $k \in \mathcal{K} \equiv \{1, ..., K+1\}$
  - Goods $k = 1, ..., K$ are 'crops'; Good $K+1$ is an 'outside' good
- Total output Q_{it}^k of good k in market i is given by

$$Q_{it}^k = \sum_{f \in \mathcal{F}_i} A_{it}^{fk} L_{it}^{kf}$$

- All fields have same productivity in outside sector: $A_{it}^{fK+1} = \alpha_{it}^{K+1}$

Basic Environment (Continued)

- Large number of price-taking farms in all local markets.
- Profits of farm producing good k in local market i are given by:

$$\Pi_{it}^{k} = p_{it}^{k} \left[\sum_{f \in \mathcal{F}_i} A_{it}^{fk} L_{it}^{kf} \right] - \sum_{f \in \mathcal{F}_i} r_{it}^{f} L_{it}^{fk},$$

  where farm-gate price of good k in local market i is given by:

$$p_{it}^{k} \equiv \bar{p}_{t}^{k} / (1 + \tau_{it}^{k}).$$

- Profit maximization by farms requires:

$$\begin{aligned} p_{it}^{k} A_{it}^{fk} - r_{it}^{f} &\leq 0, \text{ for all } k \in \mathcal{K},\ f \in \mathcal{F}_i, && (3) \\ p_{it}^{k} A_{it}^{fk} - r_{it}^{f} &= 0, \text{ if } L_{it}^{fk} > 0, && (4) \end{aligned}$$

- Factor market clearing in market i requires:

$$\sum_{k \in \mathcal{K}} L_{it}^{fk} \leq V_{i}^{f}, \text{ for all } f \in \mathcal{F}_i. \qquad (5)$$

Competitive Equilibrium

- **Notation:**
  - $\bar{p}_t \equiv (\bar{p}_t^k)_{k \in \mathcal{K}}$ is exogenously given vector of wholesale prices
  - $p_{it} \equiv \left(p_{it}^k\right)_{k \in \mathcal{K}}$ is the vector of farm gate prices
  - $r_{it} \equiv (r_{it}^f)_{f \in \mathcal{F}}$ is the vector of field prices
  - $L_{it} \equiv (L_{it}^{fk})_{k \in \mathcal{K}, f \in \mathcal{F}}$ is the allocation of fields to goods in local market i

Definition

A competitive equilibrium in a local market i at date t is a field allocation, L_{it}, and a price system, (p_{it}, r_{it}), such that conditions (3)-(5) hold.

Two Steps of Analysis

- Recall that CD (2016) break analysis down into two steps:

  1. **Measuring Farm-gate Prices:**
     - Combine data on output (from the Census) and the PPF (from the FAO) to infer the crop prices (p_{it}^k) that farmers in local market i appear to have been facing.

  2. **Measuring Gains from Integration:**
     - Compute price gaps $(1 + \tau_{it}^k)$ as the difference between farm-gate prices and prices in wholesale markets.
     - Then ask how much more productive a collection of local markets i would be under a particular counterfactual 'integration' scenario: all markets i face lower price gaps.

- Now describe how to do these steps in turn.

Measuring Farm-gate Prices

Assumptions about technological change

- The FAO aims for its measures of counterfactual productivity ($\hat{A}^{fk}_{i,2011}$) to be relevant today (ie in 2011). But how relevant are these measures for true technology (A^{fk}_{it}) in, eg, 1880?
- With data on both output and land use, by crop, CD (2016) need only the following assumption:

$$A^{fk}_{it} = \alpha^{k}_{it}\hat{A}^{fk}_{i,2011}, \text{ for all } k = 1, ..., K, \ f \in \mathcal{F}_i.$$

- How realistic is this assumption?
  - The FAO runs model under varied conditions (eg irrigation vs rain-fed).
  - R^2 of $\ln \hat{A}^{fk}_{i,scenario2} - \ln \hat{A}^{fk}_{i,scenario1}$ on crop-county fixed effects is 0.78-0.82.
  - Results are insensitive to using these alternative scenarios.

Measuring Farm-gate Prices

- Dataset contains the following measures, which we assume are related to their theoretical analogues in the following manner:

$$\begin{aligned}
\hat{S}_{it} &= \sum_{k=1}^{K} p_{it}^{k} Q_{it}^{k}, \\
\hat{Q}_{it}^{k} &= Q_{it}^{k}, \text{ for all } k = 1, ..., K, \\
\hat{L}_{it}^{k} &= \sum_{f \in \mathcal{F}_i} L_{it}^{fk}, \text{ for all } k = 1, ..., K, \\
\hat{V}_{i}^{f} &= V_{i}^{f}, \text{ for all } f \in \mathcal{F}_i.
\end{aligned}$$

Definition

Given an observation $X_{it} \equiv [\hat{S}_{it}, \hat{Q}_{it}^{k}, \widehat{L}_{it}^{k}, \hat{V}_{i}^{f}, \hat{A}_{i,2011}^{fk}]$, a vector of productivity shocks and farm gate prices, (α_{it}, p_{it}), is *admissible* if and only if there exist a field allocation, L_{it}, and a vector of field prices, r_{it}, such that (L_{it}, p_{it}, r_{it}) is a competitive equilibrium consistent with X_{it}.

Measuring Farm-gate Prices

Notation

- For any observation X_{it}, we denote:
  - $\mathcal{K}_{it}^{*} \equiv \{k : \hat{Q}_{it}^{k} > 0\}$
  - $\mathcal{A}_{it}^{*} \equiv \{\alpha : \alpha^{k} > 0 \text{ if } k \in \mathcal{K}_{it}^{*}\}$
  - $\mathcal{P}_{it}^{*} \equiv \{p : p^{k} > 0 \text{ if } k \in \mathcal{K}_{it}^{*}\}$
  - $\mathcal{L}_{i} \equiv \left\{L : \sum_{k \in \mathcal{K}} L^{fk} \leq \hat{V}_{i}^{f}\right\}$
  - $L(\alpha_{it}, X_{it}) \equiv \arg\max_{L \in \mathcal{L}_{i}} \min_{k \in \mathcal{K}_{it}^{*}} \left\{\sum_{f \in \mathcal{F}_{i}} \alpha_{it}^{k} \hat{A}_{i,2011}^{fk} L^{fk} / \hat{Q}_{it}^{k}\right\}$

Theorem

For any $X_{it} \in \mathcal{X}$, the set of admissible vectors of productivity shocks and good prices is non-empty and satisfies: (i) if $(\alpha_{it}, p_{it}) \in \mathcal{A}^*_{it} \times \mathcal{P}^*_{it}$ is admissible, then $\left(\alpha^k_{it}\right)_{k \in \mathcal{K}^*_{it}/\{K+1\}}$ is equal to unique solution of

$$\sum_{f \in \mathcal{F}} \alpha^k_{it} \hat{A}^{fk}_{i2011} L^{fk}_{it} = \hat{Q}^k_{it} \text{ for all } k \in \mathcal{K}^*_{it}/\left\{K+1\right\}, \quad (6)$$

$$\sum_{f \in \mathcal{F}_i} L^{fk}_{it} = \hat{L}^k_{it} \text{ for all } k \in \mathcal{K}^*_{it}/\left\{K+1\right\}, \quad (7)$$

with $L_{it} \in L(\alpha_{it}, X_{it})$ and (ii) conditional on $\alpha_{it} \in \mathcal{A}^*_{it}$, $L_{it} \in L(\alpha_{it}, X_{it})$ satisfying (6) and (7), $(\alpha_{it}, p_{it}) \in \mathcal{A}^*_{it} \times \mathcal{P}^*_{it}$ is admissible iff

$$\sum_{k \in \mathcal{K}^*_i/\{K+1\}} p^k_{it} \hat{Q}^k_{it} = \hat{S}_{it},$$

$$\alpha^{k'}_{it} p^{k'}_{it} \hat{A}^{fk'}_{i2011} \leq \alpha^k_{it} p^k_{it} \hat{A}^{fk}_{i2011} \text{ for all } k, k' \in \mathcal{K},\ f \in \mathcal{F}_i, \text{ if } L^{fk}_{it} > 0.$$

Corollary

For almost all $X_{it} \in \mathcal{X}$, $\left(p_{it}^{k}\right)_{k \in \mathcal{K}_{it}^{*}/\{K+1\}}$ *is equal to the unique solution of*

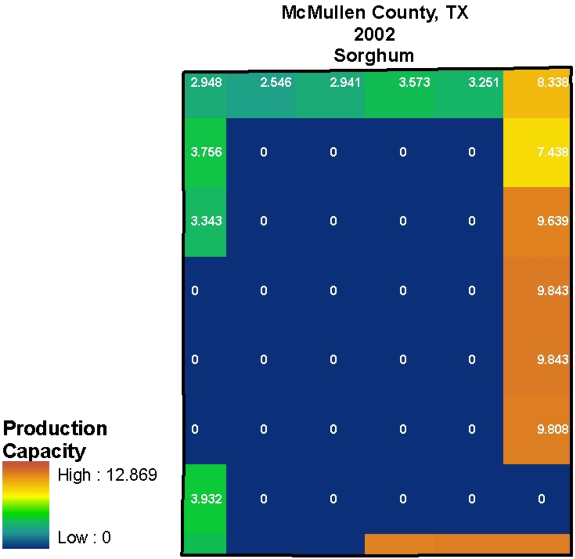
$$\sum_{k \in \mathcal{K}_{i}^{*}/\{K+1\}} p_{it}^{k} \hat{Q}_{it}^{k} = \hat{S}_{it},$$

$$\frac{p_{it}^{k'}}{p_{it}^{k}} = \frac{\alpha_{it}^{k} \hat{A}_{i2011}^{fk}}{\alpha_{it}^{k'} \hat{A}_{i2011}^{fk'}}, \text{ for any } f \in \mathcal{F}_{i} \text{ s.t. } L_{it}^{fk} \times L_{it}^{fk'} > 0,$$

where $\left(\alpha_{it}^{k}\right)_{k \in \mathcal{K}_{it}^{*}/\{K+1\}}$ *and* L_{it} *are as described in previous theorem.*

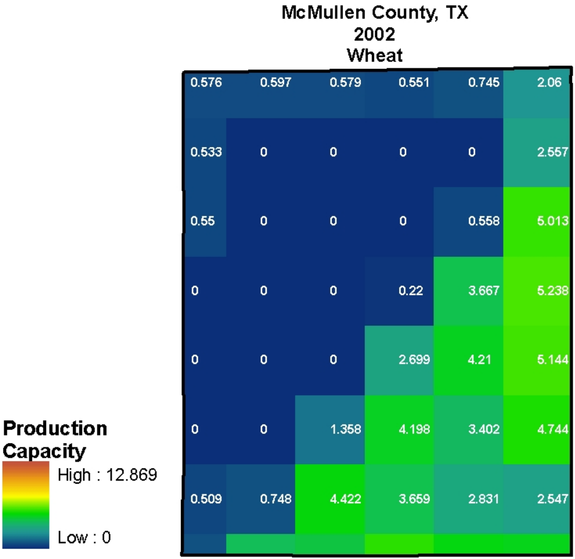
Measuring Farm-gate Prices

In practice, for a county that can be illustrated in 2-dimensions



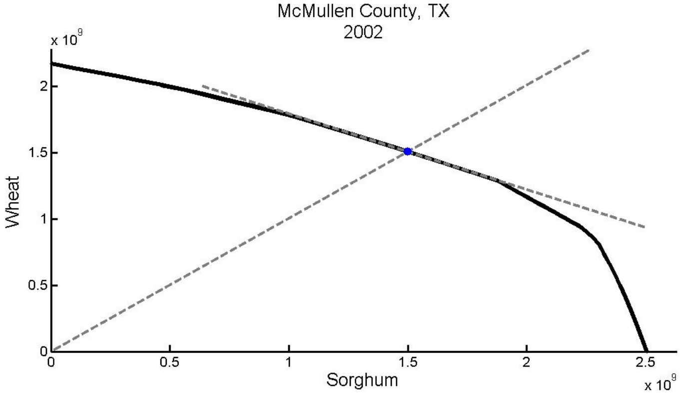
Measuring Farm-gate Prices

In practice, for a county that can be illustrated in 2-dimensions



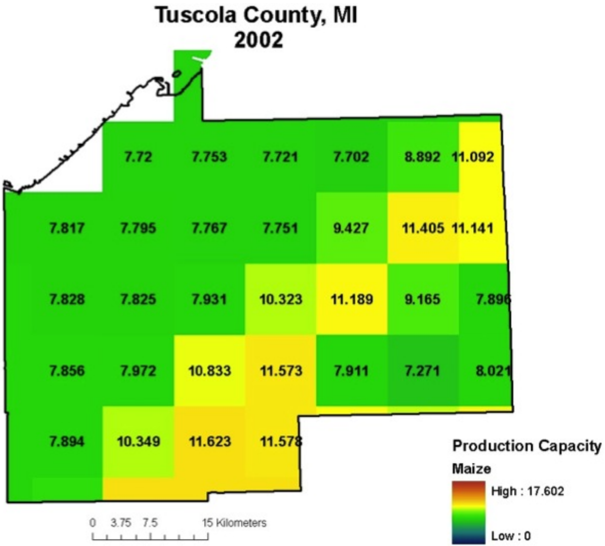
Measuring Farm-gate Prices

In practice, for a county that can be illustrated in 2-dimensions



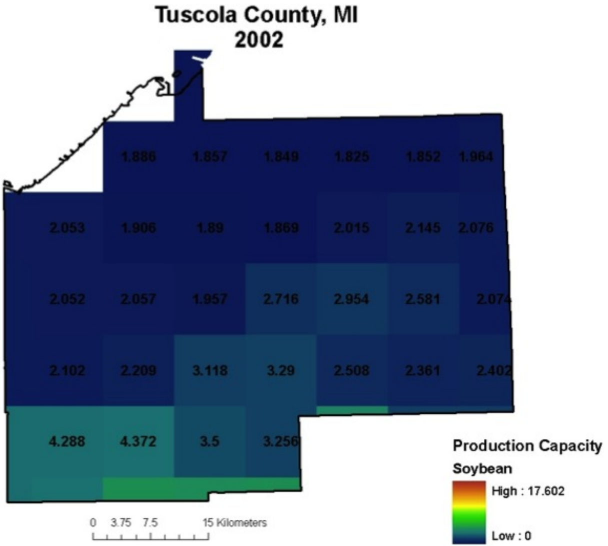
Measuring Farm-gate Prices

In practice, for a county that can be illustrated in 2-dimensions



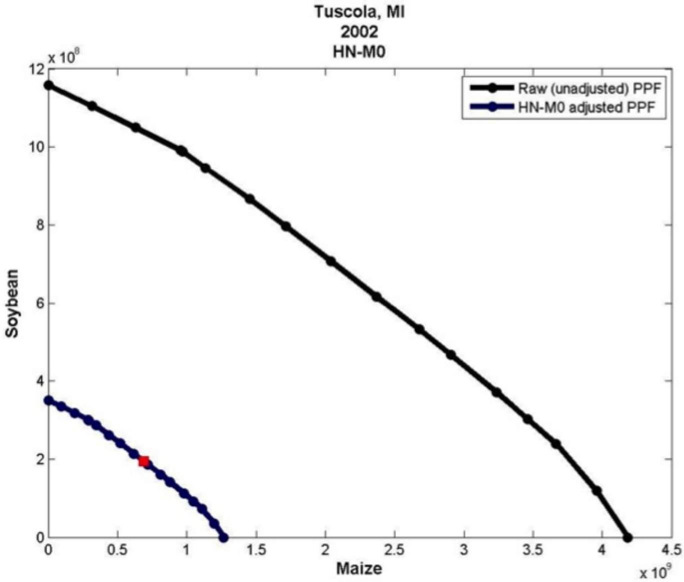
Measuring Farm-gate Prices

In practice, for a county that can be illustrated in 2-dimensions



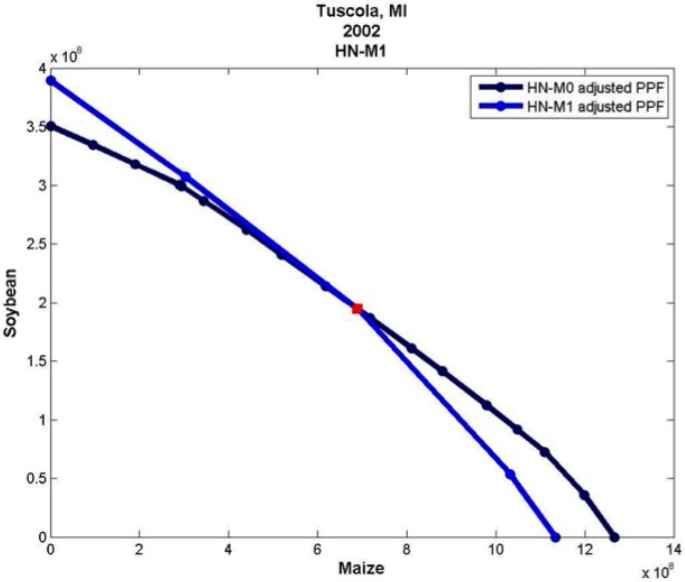
Measuring Farm-gate Prices

In practice, for a county that can be illustrated in 2-dimensions



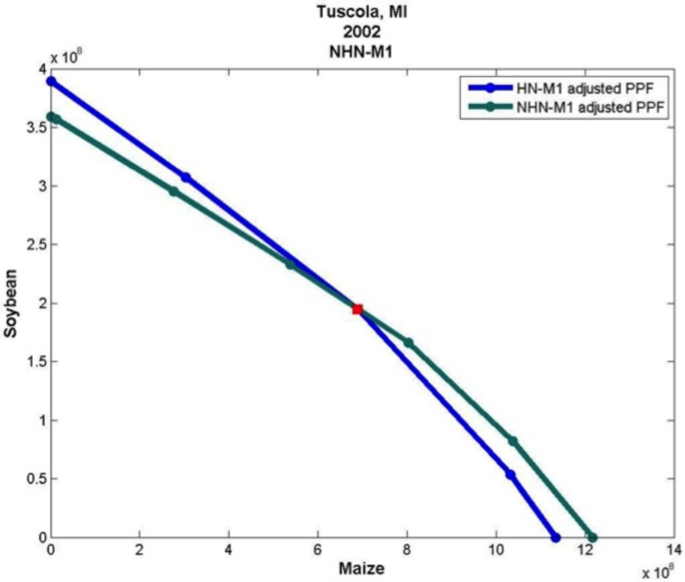
Measuring Farm-gate Prices

In practice, for a county that can be illustrated in 2-dimensions



Measuring Farm-gate Prices

In practice, for a county that can be illustrated in 2-dimensions



Measuring Farm-gate Prices
Computation

- Computation of α_{it} and p_{it} is non-trivial in high dimensional settings like those we consider.
  - For example, median county has $F = 26$ and $K^* = 8$.
  - Hence, $(K^*)^F = 3 \times 10^{23}$ fully specialized allocations to consider just to construct kinks of PPF.
  - Then $\sim$1,500 counties times 16 time periods.
- Theorem 1 is useful in this regard:
  - 'Inner loop': Conditional on α_{it}, farm-gate prices can be inferred by solving a simple linear programming problem.
  - 'Outer loop': α_{it} is relatively low-dimension (K^*).
- Paper develops algorithm that speeds up outer loop (standard algorithms too slow).

- Recall that CD (2016)'s counterfactual question is:

  "For any pair of periods, t and t', how much higher (or lower) would the total value of agricultural output in period t have been if price gaps were those of period t' rather than period t?"

- Let $\left(Q_{it}^k\right)'$ denote counterfactual output level if farmers in market i were facing $\left(p_{it}^k\right)' = \bar{p}_t^k/(1+\tau_{it'}^k)$ rather than $p_{it}^k = \bar{p}_t^k/(1+\tau_{it}^k)$.

Then measure the gains (or losses) from changes in the degree of economic integration as:

$$\Delta\tau_{t,t'}^{I} \equiv \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} \bar{p}_t^k \left(Q_{it}^k\right)'}{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} \bar{p}_t^k \hat{Q}_{it}^k} - 1,$$

$$\Delta\tau_{t,t'}^{II} \equiv \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} \left(p_{it}^k\right)' \left(Q_{it}^k\right)'}{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} p_{it}^k \hat{Q}_{it}^k} - 1.$$

- Using the above framework it is easy to compare the gains from integration (ie $\Delta\tau^{I}_{t,t'}$ and $\Delta\tau^{II}_{t,t'}$) to the gains from pure agricultural technological progress.
- Let $\left(Q^k_{it}\right)''$ denote counterfactual output level if farmers in market i had access to $(\alpha^k_{it})'' = \alpha^k_{it'}$ rather than α^k_{it}, holding prices constant.
- Then compute gains from this change in agricultural technology:

$$\Delta\alpha_{t,t'} \equiv \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} p^k_{it}\left(Q^k_{it}\right)''}{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} p^k_{it}\hat{Q}^k_{it}} - 1,$$

Measuring Gains from Economic Integration

Comments

- $\Delta\tau^{I}_{t,t'}$ and $\Delta\tau^{II}_{t,t'}$ both measure changes in GDP in agriculture in period t if price gaps were those of period t' rather than t.
- But $\Delta\tau^{I}_{t,t'}$ and $\Delta\tau^{II}_{t,t'}$ differ in terms of economic interpretation.
- For $\Delta\tau^{I}_{t,t'}$, we use reference prices to evaluate value of output.
  - Price gaps implictly interpreted as "true" distortions.
  - Similar to impact of misallocations on TFP in Hsieh Klenow (2009).
- For $\Delta\tau^{II}_{t,t'}$, we use local prices to evaluate value of output.
  - Price gaps implicitly interpreted as "true" productivity differences.
  - Similar to impact of trade costs in quantitative trade models

FAO Data: Limitations

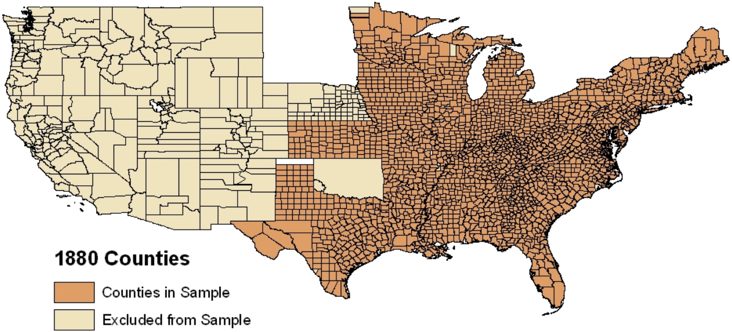
- Potentially realistic farming conditions that do not play a role in the FAO model:
  - Increasing returns to scale in growth of one crop.
  - Product differentiation (vertical or horizontal) within crop categories.
  - Sources of complementarities across crops:
    - Farmers' risk aversion.
    - Crop rotation .
    - Multi-cropping.
- Potentially realistic farming conditions that are inconsistent with CD (2016)'s application of the FAO model:
  - Changing use of non-land factors of production in response to changing prices of those factors. Introduces bias here if:
    - Relative factor prices implicitly used by FAO model differ from those in US 1880-1997,
    - and factor intensities differ across crops (among the crops that a county is growing).
  - Two seasons within a year (eg in some areas, cotton and wheat)

Agricultural Census Data

- Data on actual total output, $\hat{Q}_{it}^{k}$, and land use, $\hat{L}_{it}^{k}$, for:
  - Each crop k (barley, buckwheat, cotton, groundnuts, maize, oats, rye, rice, sorghum, soybean, sugarbeet, sugarcane, sunflower, sweet potato, wheat, white potato).
  - Each US county i (as a whole)
  - Each decade from 1840-1920, then every 5 years from 1950 to 1997.
- Data on total crop sales, $\hat{S}_{it}$, (slightly more than total sales just from our 16 crops) in county.
  - But this data starts in 1880 only.
  - Question asked of farmers changed between 1920 and 1950; comparisons difficult across these years (at the moment).
- Output and sales by county is the finest spatial resolution data available.

US County Borders in 1880

Focus on approximately 1,500 counties from Agricultural Census in 1880.



Price Data

- Key first step of our exercise is estimation of farm-gate prices.
- Natural question: how do those prices correlate with real producer price data?
- Only available producer price data is at the state-level (with unknown sampling procedure within states):
  - 1866-1969: ATICS dataset (Cooley et al, 1977), generously provided by Paul Rhode.
  - 1970-1997: supplemented with data from NASS/USDA website.

Empirical Results

- **Step 1: Measuring Farm-gate Prices**
- **Step 2: Measuring Gains from Integration**
  - How large are these gains?

Gains from Economic Integration: Question

- Recall the counterfactual question of interest:

  How much higher (or lower) would the total value of output across local markets in period t have been if price gaps were those of period t' rather than period t?

- Requires two years, t and t'.
  - For now pick $t' = 1920$ or 1997

Gains from Economic Integration: Procedure

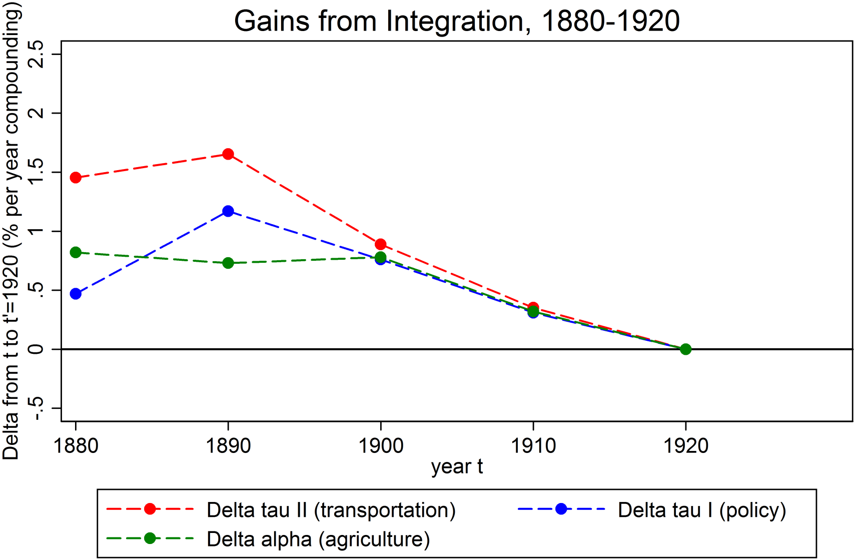
1. Define counterfactual farm-gate prices in year t as: $\left(p_{it}^{k}\right)' = \bar{p}_{t}^{k} / \left(1 + \tau_{it'}^{k}\right)$.

2. Compute counterfactual output levels $\left(Q_{it}^{k}\right)'$.

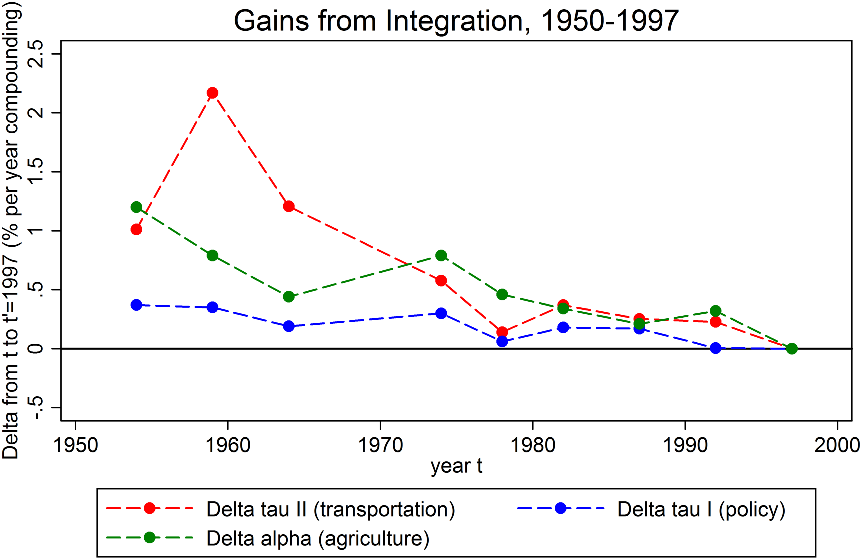
3. Compute gains from counterfactual scenario using:

$$\begin{aligned}
\Delta\tau_{t,t'}^{I} &\equiv \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} \bar{p}_{t}^{k}\left(Q_{it}^{k}\right)'}{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} \bar{p}_{t}^{k}\hat{Q}_{it}^{k}} - 1, \\
\Delta\tau_{t,t'}^{II} &\equiv \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} \left(p_{it}^{k}\right)'\left(Q_{it}^{k}\right)'}{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} p_{it}^{k}\hat{Q}_{it}^{k}} - 1, \\
\Delta\alpha_{t,t'} &\equiv \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} p_{it}^{k}\left(Q_{it}^{k}\right)''}{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} p_{it}^{k}\hat{Q}_{it}^{k}} - 1.
\end{aligned}$$

Gains from Economic Integration: Estimates



Gains from Economic Integration: Estimates



Summary

- CD (2016) have developed a new approach to measuring the gains from economic integration based on Roy/Ricardian model.
- Central to the approach is use of novel agronomic data:
  - Crucially, this source aims to provide *counterfactual productivity data*: productivity of all crops in all regions, not just the crops that are actually being grown there.
- Have used this approach to estimate:
  1. County-level prices for 16 main crops, 1880-1997.
  2. Changes in spatial distribution of price gaps across U.S. counties from 1880 to 1997: estimated gaps appear to have fallen over time.
  3. Gains associated with reductions in the level of these gaps of the same order of magnitude as productivity gains in agriculture

Other Recent Work

1. Adao (2016)
    - Follows and extends many of the insights in Heckman-Honore (1990) about how cross-market variation can identify a Roy model
    - Here, application is to how world commodity price movements over the past 20 years have affected workers (who are assumed to have Roy-like CA) within each municpality (i.e. a local labor market) in Brazil

2. Davis and Dingel (2017)
    - Study the "comparative advantage of cities" in the US using tools of log-supermodularity and monotone comparative statics (related to Costinot (ECMA, 2009))
    - Nice connection of these theoretical ideas to a real dataset

Possible Ideas for Future Research

- Can tools in the empirical matching literature (e.g. Choo-Siow (JPE, 2006), Galichon (various), Agarwal (AER 2015)) be usefully applied to international trade settings?
- The mathematical field of "optimal transport" (see e.g. books by Galichon for economists, or Villiani for mathematicians) offers an extremely general way to think about matching/assignment. Does this field generate new empirical/numerical tools?
- Can new, rich administrative datasets on matching of discrete factors (e.g. workers, parcels of land, buildings, particular pieces of capital) be used to study assignment models in new ways?

Possible Ideas for Future Research

- Are there applications of remote sensing (e.g. satellite) data that would allow a richer test of the land-use predictions of assignment models? Donaldson and Storeygard (JEP, 2016) survey the satellite data literature.
- Are there other settings where scientific/engineering knowledge of the production process can be used like with the FAO GAEZ data?