How Large Are the Gains from Economic Integration?
Theory and Evidence from Agricultural Markets

Arnaud Costinot    Dave Donaldson

MIT and NBER

October 2011
How Large are the Gains from Economic Integration in Agriculture?

Regions of the world—both across and within countries—appear to be separated by significant price distortions and barriers to trade.

These distortions and barriers appear to be substantially larger in agriculture than in manufacturing.

Two natural questions arise:

1. What would be the effects of a particular agricultural trade reform package (such as the Doha round)?

2. How large are the additional gains from complete integration (i.e., free trade)?
How Large are the Gains from Economic Integration?

- Fundamental challenge lies in predicting how local markets would behave under **counterfactual scenarios** in which they become more 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 to be employed in sectors in which producers are currently, and deliberately, choosing not to use them.

  - Any study of the gains from economic integration needs to overcome this fundamental **identification problem**.
Two main approaches in the trade literature:

- “Reduced form” approach (e.g. Frankel Romer 1999): knowledge of counterfactual technology obtained by observing behavior of “similar but open” countries.

- “Structural” approach (e.g. Eaton Kortum 2002; LINKAGE model): knowledge of counterfactual technology obtained by extrapolation based on (untestable) functional forms.

Basic idea of this paper:

- Develop new structural approach with weaker need for extrapolation by functional form assumptions.

- Exploit fact that, for agriculture, scientific knowledge of crop production provides data on both factual and counterfactual productivity.

- Goal: to study how trade liberalization will affect agriculture-related outcomes of particular interest in development context: poverty, hunger, nutrition.
Factual and Counterfactual Technologies (GAEZ Data)

Example: Rice
Factual and Counterfactual Productivities (GAEZ Data)
Example: Wheat
Factual and Counterfactual Productivities (GAEZ Data)

Example: Sugarcane
Factual and Counterfactual Productivities (GAEZ Data)

Example: Ratio of Rice Yield to Wheat Yield
What Do We Propose to Do? Theory and Estimation

- Exploit **extremely rich global agricultural data** available from the FAO and World Bank:
  - Virtually all countries around the world (plus much sub-national data).
  - Data by crop on: output, area cultivated, producer prices, trade flows, trade barriers and other distortions.
  - Detailed, comparable household consumption surveys from 36 countries (World Bank LSMS data).

- Estimate model of global agricultural economy:
  - **Technology:** Combine simple variation of Roy/Ricardian model and GAEZ data to construct PPF in each county. Adjust for non-neutral technological differences and outside land use. Extends work in Costinot and Donaldson (2011) on US history.
  - **Tastes:** Flexible, non-homothetic EASI (Lewbel and Pendakur, 2009) demand system with intra-country heterogeneity.
  - **Trade barriers:** Estimate from variation in producer prices; decompose total trade costs into policy barriers and transport costs.
What Do We Propose to Do? Using the Estimated Model

- Test GE restrictions within dataset as much as possible.
  - These could be cross-sectional restrictions within an equilibrium or comparative statics in the past.

- Compute general equilibrium consequences of reductions in trade barriers.
  - Distributional consequences (especially net producers vs net consumers within countries).
  - Poverty levels.
  - Health/nutrition-related outcomes (since we can convert food availability into its nutritive content).
Roadmap of the Talk

1. Data

2. Theory and Estimation
   - Technology
   - Tastes
   - Trade Costs

3. A Few Preliminary Empirical Results

4. A Potential Application to Study of Climate Change

5. Concluding Remarks
Roadmap of the Talk

1. Data

2. Theory and Estimation
   - Technology
   - Tastes
   - Trade Costs

3. A Few Preliminary Empirical Results

4. A Potential Application to Study of Climate Change

5. Concluding Remarks
Global Agro-Ecological Zones (GAEZ v 3.0, 2011) project run by the Food and Agriculture Organization (FAO)

- Used in Nunn and Qian (2011) to study impact of the potato.

Predicted productivity, $A^c_i (f)$, for:
- 154 varieties grouped into 25 crops $c$.
- All ‘fields’ $f$ (5 arc-minute grid cells) on Earth. 156 of these ‘fields’ in median sub-national region in our dataset.

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 levels of other human inputs. We use ‘mixed, irrigated’ to start with.
GAEZ Data: Limitations

- Potentially realistic farming conditions that do not play a role in the GAEZ 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 our (current) application of the GAEZ 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 GAEZ model differ from those in place around the world,
    - and factor intensities differ across crops (among the crops that a region is growing).
  - Two seasons within a year.
Other Data Sources

- **FAOStat Agricultural Database**: Output ($Q_{ij}^c$), cultivated area ($L_{ij}^c$), producer prices ($p_{ij}^c$), bilateral trade ($X_{ij}^c$)

- **World Bank Agricultural Distortions Database**: Domestic and cross-border agricultural policy distortions.

- **Agricultural Market Access Database (AMAD)**: Detailed enumeration of trade policy barriers (tariffs and NTBs) by HS8 agricultural product lines.

- **World Bank LSMS Household Surveys**: Consistent, comparable household consumption (and in some cases, ag. production) surveys for thousands of households in 36 developing countries.

  Typically available for:
  - 182 countries (and thousands of sub-national ‘admin 2’ regions).
  - Each of 25 crops in GAEZ study (and many other ag. activities).
  - Much of 1990-2009
Roadmap of the Talk

1. Data

2. Theory and Estimation
   - Technology
   - Tastes
   - Trade Costs

3. A Few Preliminary Empirical Results

4. A Potential Application to Study of Climate Change

5. Concluding Remarks
Basic Environment: Technology

- ‘Local’ markets \( i \in \mathcal{I} \equiv \{1,\ldots,I\} \) in which production occurs
  - In our empirical analysis, a local market will be a region of a country

- Destination markets \( j \in \mathcal{J} \equiv \{1,\ldots,J\} \) in which goods are sold

- Only factors of production are fields \( f \in \mathcal{F} \equiv \{1,\ldots,F\} \)
  - \( L_i(f) \geq 0 \) denotes the number of acres covered by field \( f \) in market \( i \)

- Fields can be used to produce multiple crops \( c \in \mathcal{C} \equiv \{1,\ldots,C\} \)

- Total output \( Q_i^c \) of crop \( c \) in market \( i \) is given by
  \[
  Q_i^c = \sum_{f \in \mathcal{F}} A_i^c(f) \cdot L_i^c(f)
  \]
Large number of price-taking farms in all local markets.

Profits of farm producing crop $c$ in $i$ and selling it to $j$ are given by:

$$\Pi_{ij}^c = \left[ \bar{p}_j^c / (1 + \tau_{ij}^c) \right] \left[ \sum_{f \in \mathcal{F}} A_i^c (f) L_i^c (f) \right] - \sum_{f \in \mathcal{F}} r_i (f) L_i^c (f),$$

Profit maximization by farms requires:

$$p_i^c A_i^c (f) - r_i (f) \leq 0, \text{ for all } c \in \mathcal{C}, f \in \mathcal{F},$$

$$p_i^c A_i^c (f) - r_i (f) = 0, \text{ if } L_i^c (f) > 0,$$

where the farm-gate price of crop $c$ in local market $i$ is given by:

$$p_i^c \equiv \max_{j \in \mathcal{J}} \left\{ \bar{p}_j^c / (1 + \tau_{ij}^c) \right\}.$$  

Factor market clearing in market $i$ requires:

$$\sum_{c \in \mathcal{C}} L_i^c (f) \leq L_i (f), \text{ for all } f \in \mathcal{F}. $$
Majority of the world’s land is used on activities other than our 25 crops (e.g., manufacturing, livestock, forestry, residential, recreational, orchards, other crops).

To allow for this, suppose land can also be used to produce a ‘manufacturing’ good:

$$Q_i^m = \sum_{f \in \mathcal{F}} A_i^m L_i^m(f), \quad (4)$$

Rest of the model is unchanged.

Key difference between agriculture and manufacturing for our purposes:

- we do not observe $\hat{Q}_i^m$ and $\hat{A}_i^m$
- we observe instead $\hat{L}_i^m = \sum_{f \in \mathcal{F}} L_i^m(f)$
Our dataset contains the following measures, which we assume are related to their theoretical analogues in the following manner:

\[ \hat{p}_i^c = p_i^c, \text{ for all } c \in C, \ i \in I \]  \hspace{1cm} (5)

\[ \hat{Q}_i^c = Q_i^c, \text{ for all } c \in C, \ i \in I \] \hspace{1cm} (6)

\[ \hat{L}_i (f) = L_i (f), \text{ for all } f \in F, \ i \in I. \]  \hspace{1cm} (7)

\[ \hat{L}_i^c = L_i^c, \text{ for all } c \in C \cup \{m\}, \ i \in I \] \hspace{1cm} (8)

\[ \hat{A}_i^c (f) = \alpha_i^c A_i^c (f), \text{ for all } c \in C, f \in F, \ i \in I. \] \hspace{1cm} (9)

**Key Estimation Challenge:** How to recover unobservable technology shifters, $\alpha_i^c$?
Theorem

Given data on \( \{ \hat{L}_i^c, \hat{A}_i^c (f) \} \) and on either \( \{ \hat{p}_i^c \} \) or \( \{ \hat{Q}_i^c \} \), there exists a (generically) unique set of technology shifters \( \alpha_i^c \) that satisfy equations (1)-(9).

Comment: If the data on \( \{ \hat{p}_i^c \} \) is used for estimation then data on \( \{ \hat{Q}_i^c \} \) can be used for testing, and vice versa.
Estimation of Technology
China, rice
Estimation of Technology
China, sugarcane
Estimation of Technology
China, slice of 25-dimensional unadjusted ppf, through rice-sugarcane dimension
Estimation of Technology

Brazil, rice
Estimation of Technology
Brazil, sugarcane
Estimation of Technology
Brazil, slice of 25-dimensional unadjusted ppf, through rice-sugarcane dimension
Roadmap of the Talk

1. Data

2. Theory and Estimation
   - Technology
   - Tastes
   - Trade Costs

3. A Few Preliminary Empirical Results

4. A Potential Application to Study of Climate Change

5. Concluding Remarks
Lewbel and Pendakur (AER 2009) propose an ‘EASI’ demand system that has a number of attractive features:

- Like AIDS (Deaton and Muellbauer, 1980), demand is both second-order flexible and non-homothetic.
- Like AIDS, estimation is via GMM/ML but (typically very successful) approximations can be estimated linearly.
- Unlike AIDS, Engel curves need not be linear (and can take any polynomial/spline shape of rank up to the number of goods).
- Unlike AIDS, error terms in budget share estimation equation have a structural interpretation (as an unobserved component of the agents’ utility functions).
Basic Environment: Tastes

- Budget shares \((w_{ik}^c)\) for good \(c\) by household \(k\) in country \(i\) are given by:

\[
w_{ik}^c = \sum_{r=0}^{R} b_r (y_{ik})^r + \sum_{l=1}^{L} (C_{lc}z_{ik}^l + D_{lc}z_{ik}^l y_{ik}) + \sum_{l=1}^{L} \sum_{h \in C \cup \{m\}} A_{hlc} z_{ik}^l p_{ik}^h + \sum_{h \in C \cup \{m\}} B_{hc} p_{ik}^h y_{ik} + \varepsilon_{ik}^c
\]

- Where
  - \(y_{ik}\) is a measure of log total real expenditures (which can be approximated by \(y_{ik} = x_{ik} - \sum_{c \in C \cup \{m\}} p_{ik}^c w_{ik}^c\), where \(x\) is log nominal total expenditure).
  - \(z_{ik}^l\) is the \(l\)th element in a potential \(L\)-dimensional vector of household characteristics.
  - \(p_{ik}^h\) is the log price
Estimation of such a demand system is straightforward, subject to the following caveats:

- Presence of $w_{ik}^c$ in $y_{ik}$ presents endogeneity problem (though Lewbel and Pendakur, 2009 find this to be very small).

- Traditional problem of estimating demand curves: the need for exogenous variation in prices. Natural instruments appear here due to the supply-side information from GAEZ.

- Model is linear under the approximation (ie $y_{ik} = x_{ik} - \sum_{c \in C \cup \{m\}} p_{ik}^c w_{ik}^c$) but not otherwise. Lewbel and Pendakur (2009) find that, as with AIDS applications, the approximation is very accurate. But non-linear estimation (GMM/ML) of exact model is of course feasible.

- Estimation of demand parameters (ie $b_r$, $C_{lc}$, $D_{lc}$, $A_{hlc}$ and $B_{hc}$) could be done pooled across countries, or separately by country.
Roadmap of the Talk

1. Data

2. **Theory and Estimation**
   - Technology
   - Tastes
   - Trade Costs

3. A Few Preliminary Empirical Results

4. A Potential Application to Study of Climate Change

5. Concluding Remarks
Two types of trade costs ($\tau_{ij}^c = T_{ij}^c + t_{ij}^c$):

- $t_{ij}^c$: policy costs (e.g., tariffs, NTB tariff equivalents) that generate revenue (redistributed lump sum?)
- $T_{ij}^c$: ‘transportation’ costs that use up resources (modeled as iceberg).
Estimating Trade Costs

- Our strategy for estimating these costs:
  - \( t_{ij}^{C} \): Directly observed (World Bank Agricultural Distortions data, AMAD).
  - \( T_{ij}^{C} \): Estimate \( \tau_{ij}^{C} \) and then use \( T_{ij}^{C} = \tau_{ij}^{C} - t_{ij}^{C} \)

- How to estimate \( \tau_{ij}^{C} \)?
  - We observe \( p_{i}^{C} \) in many (but by no means all) countries around the world.
  - Free arbitrage means that \( p_{j}^{C} - p_{i}^{C} = \tau_{ij}^{C} \) if \( X_{ij}^{C} > 0 \). (Compared to typical manufacturing study, these are presumably relatively narrowly defined goods.)
  - Use places where \( p_{j}^{C} - p_{i}^{C} = \tau_{ij}^{C} \) to estimate \( \tau_{ij}^{C} = f(\text{Distance, etc})_{ij}^{C} \).
  - Note that this is testable since \( p_{j}^{C} - p_{i}^{C} \leq \tau_{ij}^{C} = f(\text{Distance, etc})_{ij}^{C} \) if \( X_{ij}^{C} = 0 \).
Roadmap of the Talk

1. Data

2. Theory and Estimation
   - Technology
   - Tastes
   - Trade Costs

3. A Few Preliminary Empirical Results

4. A Potential Application to Study of Climate Change

5. Concluding Remarks
Now briefly discuss two results from an extremely modest and rough (!) first look at the data from 2001.

1. Testing the technology: how well does predicted output correlate with real output data?
   - For now, just a simple first pass with neutral technology shifters.

2. Estimating transport costs: does the dependence on distance look sensible?
### Results: Testing Technology Side of Model

<table>
<thead>
<tr>
<th>Dep. variable:</th>
<th>log(quantity produced in model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All crops</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>log (quantity produced in data)</td>
<td>0.795***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. *** indicates statistically significant at 0.1% level, * at 5% level.
## Results: Testing Technology Side of Model

<table>
<thead>
<tr>
<th>Dep. variable:</th>
<th>log(quantity produced in model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rice (1)</td>
</tr>
<tr>
<td></td>
<td>Rye (2)</td>
</tr>
<tr>
<td></td>
<td>Sorghum (3)</td>
</tr>
<tr>
<td></td>
<td>Soybean (4)</td>
</tr>
<tr>
<td></td>
<td>Sugar beet (5)</td>
</tr>
<tr>
<td></td>
<td>Sunflower (6)</td>
</tr>
<tr>
<td>log (quantity produced in data)</td>
<td>0.831*** (0.027)</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. *** indicates statistically significant at 0.1% level, * at 5% level.
## Results: Testing Technology Side of Model

<table>
<thead>
<tr>
<th>Dep. variable:</th>
<th>log(quantity produced in model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sweet potato</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>log (quantity produced in data)</td>
<td>0.720*** (0.038)</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. *** indicates statistically significant at 0.1% level, * at 5% level.
## Results: Trade Cost Estimation
Using typical ‘gravity regression’ determinants of trade costs, from CEPII dataset

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>log (bilateral price gap), trading pairs only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
</tr>
<tr>
<td>log (bilateral distance)</td>
<td>0.103***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>Contiguous</td>
<td>-0.066**</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
</tr>
<tr>
<td>Common official language</td>
<td>0.087**</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
</tr>
<tr>
<td>Former colonial relationship</td>
<td>-0.139***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
</tr>
</tbody>
</table>

Notes: Regressions control for other language and colonial variables in CEPII dataset. Standard errors corrected for clustering at country pair level in parentheses. *** indicates statistically significant at 0.1% level, ** at 1% level.
Roadmap of the Talk

1. Data

2. Theory and Estimation
   - Technology
   - Tastes
   - Trade Costs

3. A Few Preliminary Empirical Results

4. A Potential Application to Study of Climate Change

5. Concluding Remarks
Models of climate change portend a grim future for many specific crops in specific locations in the world.

- Climatologists and agronomists make these predictions.

But how will economic living standards suffer under climate change?

- Answer depends of course on the costs of adaptation, such as:
  - Adoption of irrigation.
  - Invention/adoption of drought-resistant varieties.
But 3 simple and potentially important types of adaptation are built into the model we have just discussed:

1. **Farmers will change (perhaps at a cost) what they grow.** In many locations, some crops will suffer and others will benefit.

2. **Prices will change.** Depends on global aggregation of crop- and location-specific productivity changes (e.g., most crops in Northern Europe expected to be helped). Depends also on barriers to trade.

3. **Consumers will substitute.**

Studying features 1 and 2 requires data/model with both factual and counterfactual technologies built in.
Roadmap of the Talk

1. Data

2. Theory and Estimation
   - Technology
   - Tastes
   - Trade Costs

3. A Few Preliminary Empirical Results

4. A Potential Application to Study of Climate Change

5. Concluding Remarks
We have proposed a new approach to estimating the gains from economic integration in world agriculture.

Central to our 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.

- In our opinion, such data is an essential input into the construction of counterfactual scenarios in Trade.

Other central features:

1. Agricultural sector is important (especially in development context), offers extremely rich data (output, prices, consumption, trade), and probably presents fewer complications (eg vertical and horizontal product differentiation, endogenous technological change) than manufacturing.

2. Estimate model on micro-data.

3. Emphasis on testing equilibrium relationships in model.