

How Large are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-2002

Arnaud Costinot
MIT and NBER

Dave Donaldson
MIT, NBER and CIFAR

PRELIMINARY AND INCOMPLETE

August 15, 2011

Abstract

In this paper we develop a new approach to measuring the gains from economic integration based on a Roy-like assignment model in which heterogeneous factors of production are allocated to multiple sectors in multiple local markets. We implement this approach using data on crop markets in 1,500 U.S. counties from 1880 to 2002. Central to our empirical analysis is the use of a novel agronomic data source on predicted output by crop for small spatial units. Crucially, this dataset contains information about the productivity of all units for all crops, not just those that are actually being grown. Using this new approach we estimate: *(i)* the spatial distribution of price wedges across U.S. counties from 1880 to 2002; *(ii)* the gains associated with changes in the level of these wedges over time; and *(iii)* the further gains that could obtain if all wedges were removed and gains from market integration were fully realized.

1 Introduction

How large are the gains from economic integration? Since researchers never observe markets that are both closed and open at the same time, the fundamental challenge in answering this question lies in predicting how local markets, either countries or regions, would behave under counterfactual scenarios in which they suddenly become more or less integrated with the rest of the world.

One prominent response to this challenge, the “reduced form” approach, argues that knowledge about such counterfactual scenarios can be obtained by comparing countries that are more open to countries that are more closed (or analogously, by comparing a country to itself before and after its level of trade openness changes).¹ Frankel and Romer (1999) is a well-known example. The implicit assumption about counterfactual scenarios embodied in the reduced form approach is that currently open economies would behave, were they to be less open, in exactly the same manner as countries that are currently less open are currently behaving. An unsettled area of debate in this literature concerns the credibility of this counterfactual assumption. A potential weakness of this approach is that it does not estimate policy invariant parameters of economic models, thereby limiting its relevance for policy and welfare analysis.

Another prominent approach, the “structural” approach, aims to estimate or calibrate fully specified models of how countries behave under any trading regime. Eaton and Kortum (2002) is the most influential application of this approach in the international trade literature. A core ingredient of such models is that there exists a set of technologies that a country would have no choice but to use if trade were restricted, but which the country can choose not to use when it is able to trade. Estimates of the gains from economic integration, however defined, thereby require the researcher to compare factual technologies that are currently being used to inferior, counterfactual technologies that are deliberately not being used and are therefore unobservable to the researcher. This comparison is typically made through the use of functional form assumptions that allow an extrapolation from observed technologies to unobserved ones.

The goal of this paper is to develop a new structural approach with less need for extrapolation by functional form assumptions in order to obtain knowledge of counterfactual scenarios. Our basic idea is to focus on agriculture, a sector of the economy in which scientific knowledge of how essential inputs such as water, soil and climatic conditions map into outputs is uniquely well understood. As a consequence of this knowledge, agronomists are able to predict—typically with great success—how productive a given parcel of land (a ‘field’) would be were it to be used to grow any one of a set of crops. Our approach combines

¹We use the terms “reduced form” and “structural” in the same sense as, for example, Chetty (2009).

these agronomic predictions about factual and counterfactual technologies with a Roy-like assignment model in which heterogeneous fields are allocated to multiple crops in multiple local markets.

We implement our approach in the context of U.S. agricultural markets from 1880 to 2002—a setting with an uncommonly long stretch of high-quality, comparable micro-data from an important agricultural economy. Our dataset consists of approximately 1,500 U.S. counties which we treat as separate local markets that may be segmented by barriers to trade—analogue to countries in a standard trade model. Each county is endowed with many ‘fields’ of arable land.² At each of these fields, a team of agronomists, as part of the Food and Agriculture Organization’s (FAO) Global Agro-Ecological Zones (GAEZ) project, have used high-resolution data on soil, topography, elevation and climatic conditions, fed into state-of-the-art models that embody the biology, chemistry and physics of plant growth, to predict the quantity of yield that each field could obtain if it were to grow each of 17 different crops. In our Roy-like assignment model, these data are sufficient to construct the production possibility frontier associated with each U.S. county, which is the essential ingredient required to perform any counterfactual analysis.

Counterfactual analysis in our paper formally proceeds in two steps. First we combine the productivity data from the GAEZ project with data from the decadal Agricultural Census on the total amount of output of each crop in each county. Under perfect competition, we demonstrate how this information can be used to infer local crop prices in each county and time period as the solution of a simple linear programming problem. Second, armed with these estimates of local crop prices, we compute the spatial distribution of price ‘wedges’ across U.S. counties from 1880 to 2002 and ask: “For any pair of periods, t and t' , how much higher (or lower) would the total value of agricultural output across U.S. counties in period t have been if wedges were those of period t' rather than period t ?” The answer to this counterfactual question will be our measure of the gains (or losses) from changes in the degree of economic integration in U.S. agricultural markets over time.

A natural concern with our use of the FAO-GAEZ data centers on its reliability for our purposes—namely, as a reflection of the state of relative productivity levels across fields and crops within any given county and year (the absolute level of a county’s productivity relative to that predicted in the FAO-GAEZ data is irrelevant for our exercise). For example, this data source would be unsuitable for our purposes if technological change has been biased across crops (for example, if irrigation became more affordable over time and some crops benefit more from irrigation than other crops). As a response to this concern we develop in Section 2.5 below an alternative procedure that is robust to general, unknown crop-specific

²While we use the term ‘fields’ to describe the finest spatial regions in our dataset, fields are still relatively large spatial units. For example, the median U.S. county contains 26 fields.

technological change (or indeed crop-specific errors in the FAO-GAEZ model) as long as this technological change (or model error) affects all fields in a county equally. Doing so requires additional information that was also collected in the US agricultural census, namely the amount of land devoted to each crop (not just the physical quantity of output). That is, we show that, as long as the pattern of comparative advantage (across fields and crops) within a county-year is preserved by technological change, the allocation of fields to crops can still be solved for (even with unknown crop-specific technological change) using data on land use by crop and the resulting equilibrium allocation can then be used to correct estimated local crop prices for crop-specific technological change.

A second potential concern with our estimates is that they relate only to the 17 crops covered by the FAO-GAEZ project—a set of activities that does not span the set of all uses of land in the United States. More specifically, at prices prevailing in a given county and time period, the optimal use of land may not be in one of the 17 crops at all. With this concern in mind Section 2.4 below develops an extension to our baseline model in which there is an outside use of land whose price and productivity level (in each county and year) are unknown. We show that, just as in the case of crop-specific technological change described above, with data on the total amount of land being used to grow the sum of the 17 crops the estimation of prices and productivity of each of these 17 crops is not jeopardized by potential outside uses of land.

Relation to the Literature. In the trade literature, most structural work aimed at quantifying the gains from market integration is based on the seminal work of Eaton and Kortum (2002). A non-exhaustive list of recent quantitative papers building on Eaton and Kortum’s (EK) approach includes Dekle, Eaton, and Kortum (2008), Chor (2010), Donaldson (2010), Waugh (2010), Ramondo and Rodriguez-Clare (2010), Caliendo and Parro (2010), Costinot, Donaldson, and Komunjer (2011), and Fielor (2011). The EK approach can be sketched as follows. First, combine data on bilateral imports and trade costs to estimate the elasticity of import demand (most often through a simple gravity equation). Second, use functional forms in the model together with elasticity of import demand to predict changes in real GDP associated with a counterfactual change in trade costs; see Arkolakis, Costinot, and Rodriguez-Clare (2011).

Our approach, by contrast, focuses entirely on the supply-side of the economy. First we combine data on output and productivity to estimate producer prices, and in turn, trade costs. Second we use the exact same data to predict the changes in nominal GDP associated with a counterfactual change in trade costs. As emphasized above, the main benefit of our approach is that it weakens the need for extrapolation by functional form assumptions. The main cost of our approach—in addition to the fact that it applies only to agriculture—is that it only allows us to infer production gains from trade. In order to estimate consumption gains

from trade, we would also need consumption data, which is not available at the US county level over our extended time period.

Our proposal is related more broadly to work on the economic history of domestic market integration; see e.g. Shiue (2002) and Keller and Shiue (2007). Using market-level price data this body of work typically aims to estimate the magnitude of deviations from perfect market integration. Our approach, by contrast, first estimates market-level prices (and hence can be applied in settings, like ours, where price data is not available), and then goes beyond the previous literature by estimating the magnitude of the production efficiency gains that would occur if market integration improved.

A final area of research to which this paper relates is the recent macro literature on aggregate productivity losses due to misallocation of production, e.g., Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). In this literature misallocation is defined as an instance in which the value of the marginal product of a factor is not equalized across productive units (typically firms) with deviations from such equality, i.e. wedges, identifying the extent of distortions. Although we will interpret price wedges as physical trade costs throughout this proposal—which means that the allocation is efficient given the transportation technology—our approach could easily be extended to investigate the importance of misallocation of land across counties. Since our approach builds on detailed information about technology, it also has the potential to reduce the role of functional form assumptions both when identifying misallocation and when measuring its aggregate productivity consequences.

The rest of this paper is organized as follows. Section 2 introduces our theoretical framework, describes how to measure local prices, and in turn, how to measure the gains from economic integration. Section 3 describes our data. Section 4 presents our main empirical results. Section 6 concludes. All formal proofs can be found in the appendix.

2 Theoretical Framework

2.1 Endowments, Technology, and Market Structure

Our theoretical framework is a Roy-like assignment model, as in Costinot (2009). We consider an economy with multiple local markets indexed by $i \in \mathcal{I} \equiv \{1, \dots, I\}$. In our empirical analysis, a local market will be a US county. In each market, the only factors of production are different types of land or fields indexed by $f \in \mathcal{F} \equiv \{1, \dots, F\}$. $L_i(f) \geq 0$ denotes the number of acres covered by field f in market i . Fields can be used to produce multiple crops indexed by $c \in \mathcal{C} \equiv \{1, \dots, C\}$. Fields are perfect substitutes in the production of each crop, but vary in their productivity per acre, $A_i^c(f) > 0$. Total output Q_i^c of crop c in market i is

given by

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

where $L_i^c(f) \geq 0$ denotes the number of acres of field f allocated to crop c in market i . Note that $A_i^c(f)$ may vary both with f and c . Thus although fields are perfect substitutes in the production of each crop, some fields may have a comparative as well as absolute advantage in producing particular crops.

All crops are produced by a large number of price-taking farms in all local markets. The profits of a representative farm producing crop c in market i are given by

$$\Pi_i^c = p_i^c \sum_{f \in \mathcal{F}} A_i^c(f) L_i^c(f) - r_i(f) L_i^c(f),$$

where p_i^c and $r_i(f)$ denote the price of crop c and rental rate per acre of field f in market i , respectively. Profit maximization by farms requires

$$p_i^c A_i^c(f) - r_i(f) \leq 0, \text{ for all } c \in \mathcal{C}, f \in \mathcal{F}, \quad (1)$$

$$p_i^c A_i^c(f) - r_i(f) = 0, \text{ if } L_i^c(f) > 0 \quad (2)$$

Local factor markets are segmented. Thus factor market clearing in market i requires

$$\sum_{c \in \mathcal{C}} L_i^c(f) \leq L_i(f), \text{ for all } f \in \mathcal{F}. \quad (3)$$

We leave good market clearing conditions unspecified, thereby treating each local market as a small open economy. In the rest of this paper we denote by $p_i \equiv (p_i^c)_{c \in \mathcal{C}}$ the vector of crop prices, $r_i \equiv [r_i(f)]_{f \in \mathcal{F}}$ the vector of field prices, and $L_i \equiv [L_i^c(f)]_{c \in \mathcal{C}, f \in \mathcal{F}}$ the allocation of fields to crops in market i .

2.2 Measuring Local Crop Prices

Our dataset contains measures of total farms' sales, \hat{S}_i , total output per crop, \hat{Q}_i^c , as well as total acres of land covered by field f , $\hat{L}_i(f)$, for all local markets. Throughout our empirical analysis, we assume that none of these variables is subject to measurement error:

$$\hat{S}_i = \sum_{c \in \mathcal{C}} p_i^c Q_i^c, \quad (4)$$

$$\hat{Q}_i^c = Q_i^c, \text{ for all } c \in \mathcal{C}, \quad (5)$$

$$\hat{L}_i(f) = L_i(f), \text{ for all } f \in \mathcal{F}. \quad (6)$$

Our dataset also contains measures of productivity per acre, $\hat{A}_i^c(f)$, for each field in each market if that field were to be allocated to the production of crop c . Since we only have

access to these measures at one point in time, 2002, we assume that measured productivity is equal to the true productivity, $A_i^c(f)$, times some market specific error term:

$$\hat{A}_i^c(f) = \alpha_i A_i^c(f), \text{ for all } c \in \mathcal{C}, f \in \mathcal{F}. \quad (7)$$

We refer to $X_i \equiv \left[\hat{Q}_i^c, \hat{L}_i(f), \hat{S}_i, \hat{A}_i^c(f) \right]_{c \in \mathcal{C}, f \in \mathcal{F}}$ as an ‘‘observation’’ for market i . Unless otherwise specified, we assume from now on that all observations satisfy Equations (4)-(7).

Before stating our main theoretical result, it is useful to introduce two definitions.

Definition 1 *A vector of crop prices, p_i , is competitive conditional on an observation X_i if and only if there exist a vector of field prices, r_i , and an allocation of fields to crops, L_i , such that Equations (1)-(7) hold.*

Broadly speaking, Definition 1 states that a vector of crop prices, p_i , is competitive conditional on observation X_i if there exists a competitive equilibrium in market i such that: (i) the allocation of fields to crops is consistent with X_i ; and (ii) crop prices are equal to p_i .

Definition 2 *An allocation L_i is efficient conditional on an observation X_i if and only if it is a solution of the following planning problem*

$$\max_{\tilde{L}_i} \min_{c \in \mathcal{C}_i^*} \left\{ \sum_{f \in \mathcal{F}} \hat{A}_i^c(f) \tilde{L}_i^c(f) / \hat{Q}_i^c \right\} \quad (\text{P})$$

$$\sum_{c \in \mathcal{C}} \tilde{L}_i^c(f) \leq \hat{L}_i(f), \text{ for all } f \in \mathcal{F}, \quad (8)$$

$$\tilde{L}_i^c(f) \geq 0, \text{ for all } c \in \mathcal{C}, f \in \mathcal{F}, \quad (9)$$

where $\mathcal{C}_i^* \equiv \{c \in \mathcal{C} | \hat{Q}_i^c > 0\}$ denotes the set of crops with positive output in market i .

Formally, Definition 2 states that an allocation L_i is efficient conditional on an observation X_i if and only if (i) it is consistent with X_i ; and (ii) it maximizes the utility of a representative agent with Leontieff preference whose consumption coincide with output levels observed in the data. One can show that if L_i was *not* efficient in the sense of Definition 2, then there would be *no* error term α_i such that L_i is consistent with X_i and lies on the PPF of market i , hence our choice of terminology. This observation is at the core of the following theorem.

Theorem 1 *A vector of crop prices, p_i , is competitive conditional on X_i if and only if there exists an efficient allocation L_i conditional on X_i such that for any pair of crops $c, c' \in \mathcal{C}$, the relative price of the two crops satisfies*

$$p_i^{c'} / p_i^c = \hat{A}_i^c(f) / \hat{A}_i^{c'}(f), \text{ if } L_i^c(f) > 0 \text{ and } L_i^{c'}(f) > 0, \quad (10)$$

$$p_i^{c'} / p_i^c \leq \hat{A}_i^c(f) / \hat{A}_i^{c'}(f), \text{ if } L_i^c(f) > 0 \text{ and } L_i^{c'}(f) = 0, \quad (11)$$

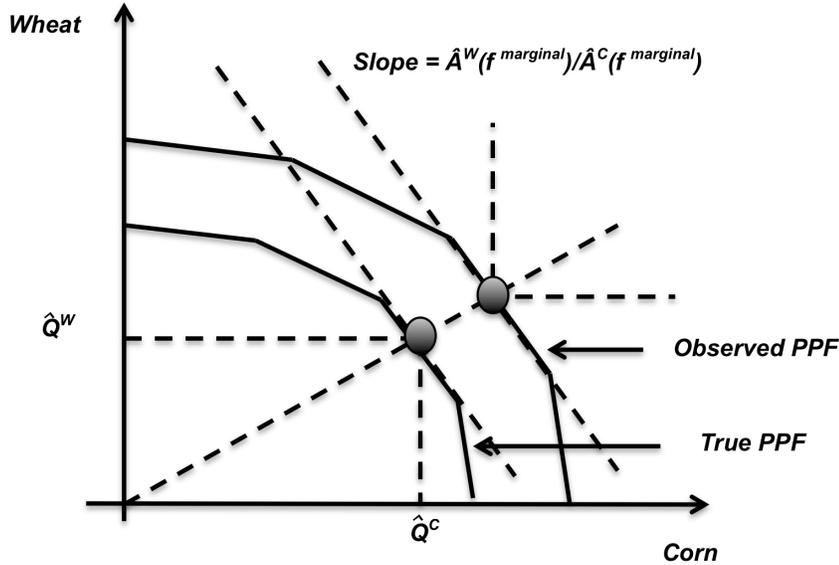


Figure 1: Measuring Local Crop Prices

with nominal prices such that $\sum_{c \in C_i^*} p_i^c \hat{Q}_i^c = \hat{S}_i$.

Theorem 1 characterizes the set of competitive prices associated with any observation. The logic behind Theorem 1 can be sketched as follows. By the zero profit condition, for any pair of crops that are produced in equilibrium, relative prices are equal to the inverse of the relative productivity of the “marginal” field, i.e. the field involved in the production of both crops in equilibrium. By the First and Second Welfare Theorems, the identity of the marginal field in a competitive equilibrium can be recovered from a planning problem.

Our procedure is illustrated in Figure 1 for a market producing two crops, corn and wheat, using four types of fields. From data on endowments and productivity, one can construct the Production Possibility Frontier (PPF). Knowing the shape of the PPF, information about relative output levels of the two crops can then be used to infer the slope of the PPF at the observed vector of output, and in turn, the relative price of the two crops produced in this market.

According to Theorem 1, one can infer the set of competitive crop prices by solving a simple linear programming problem. Since our dataset includes approximately 1,500 counties over 17 decades, this is very appealing from a computational standpoint. In spite of the high-dimensionality of the problem we are interested in—the median U.S. county in our dataset features 17 crops and 26 fields and hence 17^{26} or almost 10^{31} possible perfectly specialized allocations of crops to fields (and the equilibrium allocation will not be perfectly specialized)—it is therefore possible to characterize the set of competitive crop prices in each county in a very limited amount of time using standard software packages.

Finally, note that for any pair of goods produced in a given local market, the relative price is generically unique. Non-uniqueness may only occur if the observed vector of output is colinear with a vertex of the production possibility frontier associated with observed productivity levels and endowments. Not surprisingly, this situation will never occur in our empirical analysis.

2.3 Measuring the Gains from Economic Integration

Before measuring the gains from economic integration, one first needs to take a stand on how to measure economic integration across markets. Intuitively, the extent of economic integration should be related to differences in local crop prices. For one thing, we know that if crop markets were perfectly integrated, then crop prices should be the same across markets. To operationalize that idea, we introduce the following definition.

Definition 3 *For any pair of crops $c, c' \in \mathcal{C}$ in any period t , we define the extent of economic integration between market i and some reference market as the percentage difference or “wedge,” $\tau_{it}^c \equiv p_i^c/p_{it}^c - 1$, between the price of crop c in the two markets.*

Armed with Definition 3, one can then estimate the gains (or losses) from changes in the degree of economic integration across markets between two periods t and $t' > t$ by answering the following counterfactual question:

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

Formally, let $(Q_{it}^c)'$ denote the counterfactual output level of crop c in market i in period t if farms in this market were maximizing profits facing the counterfactual prices $(p_{it}^c)' = p_i^c/(1 + \tau_{it}^c)$ rather than the true prices $p_{it}^c = p_i^c/(1 + \tau_{it}^c)$. Using this notation, we express the gains (or losses) from changes in the degree of economic integration between two periods t and $t' > t$ as:

$$\Delta_{tt'} \equiv \frac{\sum_{i \in \mathcal{I}} \sum_{c \in \mathcal{C}} (p_{it}^c)' (Q_{it}^c)'}{\sum_{i \in \mathcal{I}} \sum_{c \in \mathcal{C}} p_{it}^c \hat{Q}_{it}^c} - 1. \quad (12)$$

By construction, $\Delta_{tt'}$ measures how much larger (or smaller) GDP in agriculture would have been in period t if wedges were those of period t' rather than those of period t .

In Equation (12) we use local prices both in the original and the counterfactual equilibrium. The implicit assumption underlying $\Delta_{tt'}$ is that differences in local crop prices reflect “true” technological considerations. Under this interpretation of wedges, farmers face the “right” prices, but producer prices are lower in local markets than in the reference market because of the cost of shipping crops. This metric is in the same spirit as the measurement

of the impact of trade costs in quantitative trade models; see e.g. Eaton and Kortum (2002) and Waugh (2010).³

A few comments are in order. First, it should be clear that our strategy only allows us to identify wedges relative to some reference prices. This implies that measurement error in reference prices may affect our estimates of the gains from economic integration. We come back to this important issue in the next sections. Second, our strategy only allows us to identify the production gains from economic integration. Since we do not have consumption data, our analysis will remain silent about any consumption gains from economic integration.

2.4 Extension I: Non-Agricultural Land Use

The first of our two extensions allows for non-agricultural land uses, e.g. forests, services, or manufacturing. For expositional purposes, we simply refer to such activities in short as “manufacturing.” Crop production is as described in Section 2.1. But unlike in our baseline model, land can also be used to produce a composite manufacturing good according to

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

where $L_i^m(f) \geq 0$ denotes the number of acres of field f allocated to manufacturing. The key difference between agriculture and manufacturing is that land productivity is assumed to be constant across fields in manufacturing: A_i^m is independent of f .

Zero-profit conditions and factor market clearing conditions (1)-(3) continue to hold as described in Section 2.1, but now for all sectors $j \in \mathcal{C} \cup \{m\}$. In terms of measurement, the key difference between agriculture and manufacturing is that instead of having access to a quantity index for our composite manufacturing good as well as a measure of productivity, we only observe the total acres of land devoted to manufacturing activities:

$$\hat{L}_i^m = \sum_{f \in \mathcal{F}} L_i^m(f). \tag{13}$$

Equations (4)-(7) are unchanged. We now refer to $Y_i \equiv \left[\hat{Q}_i^c, \hat{L}_i(f), \hat{A}_i^c(f), \hat{L}_i^m \right]_{c \in \mathcal{C}, f \in \mathcal{F}}$ as an “observation” for market i and assume that all observations now satisfy Equations (4)-(7) and (13). In this environment our definition of competitive prices and efficient allocations can be generalized as follows.

³An alternative metric might use the reference prices both in the original and the counterfactual equilibrium. In this case the implicit assumption underlying $\Delta_{tt'}$ would be that differences in local crop prices reflect “true” distortions. In order to maximize welfare—whatever the underlying preferences of the U.S. representative agent may be—farmers should be maximizing profits taking the reference prices p_i^c as given, but because of various policy reasons, they do not. Such an alternative welfare metric would be in the spirit of the measurement of the impact of misallocations on TFP in Hsieh and Klenow (2009).

Definition 1(M) A vector of crop prices, p_i , is competitive conditional on an observation Y_i if and only if there exist a vector of field prices, r_i , and an allocation of fields to crops, L_i , such that Equations (1)-(7) and (13) hold.

Definition 2(M) An allocation L_i is efficient conditional on an observation Y_i if and only if it is a solution of the following planning problem

$$\max_{\tilde{L}_i} \min_{c \in C_i^*} \left\{ \sum_{f \in \mathcal{F}} \hat{A}_i^c(f) \tilde{L}_i^c(f) / \hat{Q}_i^c \right\} \quad (\text{P-M})$$

$$\sum_{j \in \mathcal{C} \cup \{m\}} \tilde{L}_i^j(f) \leq \hat{L}_i(f), \text{ for all } f \in \mathcal{F}, \quad (14)$$

$$\tilde{L}_i^j(f) \geq 0, \text{ for all } j \in \mathcal{C} \cup \{m\}, f \in \mathcal{F}, \quad (15)$$

$$\sum_{f \in \mathcal{F}} \tilde{L}_i^m(f) \geq \hat{L}_i^m. \quad (16)$$

where $C_i^* \equiv \{c \in \mathcal{C} | \hat{Q}_i^c > 0\}$ denotes the set of crops with positive output in market i .

Compared to Definitions 1 and 2, the previous definitions require competitive prices and efficient allocations to be consistent with the observed allocation of land to non-agricultural activities. Note also that since Definition 1 (M) only applies to the vector of crop prices, it only requires conditions (1) and (2) to hold for all $c \in \mathcal{C}$ rather than all sectors $j \in \mathcal{C} \cup \{m\}$. Modulo this change of definitions, Theorem 1 remains unchanged.

2.5 Extension II: Crop-and-County Specific Productivity Shocks

Our second extensions allows for a weakening of the assumption that the FAO's predictions are right—and right in all years t —up to a scalar. More specifically, we now relax Equation (7) and allow for crop-and-market specific productivity shocks:

$$\hat{A}_i^c(f) = \alpha_i^c A_i^c(f). \quad (17)$$

In order to infer what is now a vector of error terms $\alpha_i \equiv (\alpha_i^c)_{c \in \mathcal{C}}$ for each local market, we use an extra piece of information contained in our dataset: the total acres of land allocated to crop c in county i , \hat{L}_i^c . Since we have access to this measure in all periods, we again assume that it is not subject to measurement error:

$$\hat{L}_i^c = \sum_{f \in \mathcal{F}_m} L_i^c(f). \quad (18)$$

We now refer to $Z_i \equiv \left[\hat{Q}_i^c, \hat{L}_i(f), \hat{S}_i, \hat{A}_i^c(f), \hat{L}_i^c \right]_{c \in \mathcal{C}, f \in \mathcal{F}}$ as an “observation” for market i and assume that all observations now satisfy equations (4)-(6) and (17)-(18). In this environment, we introduce the following extensions of Definitions 1 and 2.

Definition 1(C) A vector of crop prices, p_i , is competitive conditional on an observation Z_i if and only if there exist a vector of field prices, r_i , and an allocation of fields to crops, L_i , such that Equations (1)-(6) and (18)-(17) hold.

Definition 2(C) An allocation L_i is efficient conditional on an observation Z_i if and only if it is a solution of

$$L_i = \arg \max_{\tilde{L}_i} \min_{c \in \mathcal{C}_i^*} \left\{ \frac{\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) \tilde{L}_i^c(f)}{\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f)} \right\} \quad (\text{P-C})$$

$$\sum_{c \in \mathcal{C}} \tilde{L}_i^c(f) \leq \hat{L}_i(f), \text{ for all } f \in \mathcal{F}, \quad (19)$$

$$\tilde{L}_i^c(f) \geq 0, \text{ for all } c \in \mathcal{C}, f \in \mathcal{F}, \quad (20)$$

$$\sum_{f \in \mathcal{F}} \tilde{L}_i^c(f) = \hat{L}_i^c, \text{ for all } c \in \mathcal{C}. \quad (21)$$

where $\mathcal{C}_i^* \equiv \{c \in \mathcal{C} | \hat{Q}_i^c > 0\}$ denotes the set of crops with positive output in market i .

Compared to Definitions 1 and 2, three differences are worth noting. First, like in Section 2.4, competitive prices and efficient allocations need to be consistent with an additional observation, here the allocation of fields to crops. Second, observed output levels no longer enter explicitly the definition of efficient allocations. The reason is that given any allocation L_i , one can always choose crop-and-market specific α_i^c productivity shocks such that the allocation is consistent with observed output levels. Namely, one would simply set $\alpha_i^c = \sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f) / \hat{Q}_i^c$. Third, efficient allocations are no longer given by the solution of a simple linear programming problem. Instead they correspond to the solution of a fixed point problem (that still involves linear programming).

For simplicity, suppose that all crops are being produced in a given county.⁴ Then modulo this change of definitions, Theorem 1 generalizes as follows.

Theorem 1(C) A vector of crop prices, p_i , is competitive conditional on Z_i if and only if there exists an efficient allocation L_i conditional on Z_i such that for any pair of crops $c, c' \in \mathcal{C}_i^*$, the relative price of the two crops satisfies

$$p_i^{c'} / p_i^c = \hat{A}_i^c(f) \alpha_i^{c'} / \hat{A}_i^{c'}(f) \alpha_i^c, \text{ if } L_i^c(f) > 0 \text{ and } L_i^{c'}(f) > 0, \quad (22)$$

$$p_i^{c'} / p_i^c \leq \hat{A}_i^c(f) \alpha_i^{c'} / \hat{A}_i^{c'}(f) \alpha_i^c, \text{ if } L_i^c(f) > 0 \text{ and } L_i^{c'}(f) = 0, \quad (23)$$

with $\alpha_i^c = \sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^{*c}(f) / \hat{Q}_i^c$ and nominal prices such that $\sum_{c \in \mathcal{C}_i^*} p_i^c \hat{Q}_i^c = \hat{S}_i$.

As hinted above, Theorem 1(C) is less useful than Theorem 1 since it is harder to solve the fixed point problem in Definition 2(C) than the linear programming problem in Definition

⁴The other cases simply require making additional assumptions on prices and/or productivity changes for crops that are not produced in a given county.

2.

3 Data

Our analysis draws on three main sources of data: predicted productivity by field and crop (from the FAO-GAEZ project); aggregate county-level data (from the US Agricultural Census) on output by crop, cultivated area by crop, and total sales of all crops; and data on reference prices. We describe these here in turn.

3.1 Productivity Data

The first and most novel data source that we make use provides measures of productivity (ie $\hat{A}_i^c(f)$ in the model above) by crop c , county i , and field f . These measures comes from the Global Agro-Ecological Zones (GAEZ) project run by the Food and Agriculture Organization (FAO).⁵ The GAEZ aims to provide a resource that farmers and government agencies can use (along with knowledge of prices) to make decisions about the optimal crop choice in a given location that draw on the best available agronomic knowledge of how crops grow under different conditions.

The core ingredient of the GAEZ predictions is a set of inputs that are known with extremely high spatial resolution. This resolution governs the resolution of the final GAEZ database and, equally, that of our analysis—what we call a ‘field’ (of which there are 26 in the median U.S. county) is the spatial resolution of GAEZ’s most spatially coarse input variable. The inputs to the GAEZ database are data on an eight-dimensional vector of soil types and conditions, the elevation, the average land gradient, and climatic variables (based on rainfall, temperature, humidity, sun exposure), in each ‘field’. These inputs are then fed into an agronomic model—one for each crop—that predicts how these inputs affect the ‘microfoundations’ of the plant growth process and thereby map into crop yields. Naturally, farmers’ decisions about how to grow their crops and what complementary inputs (such as irrigation, fertilizers, machinery and labor) to use affect crop yields in addition to those inputs (such as sun exposure and soil types) over which farmers have very little control. For this reason the GAEZ project constructs different sets of productivity predictions for different scenarios of farmer inputs. For now we use their scenario that relates to ‘mixed inputs, with possible irrigation’ but in future work we will explore how our results change across such scenarios.

⁵This database has been used by Nunn and Qian (2011) to obtain predictions about the potential productivity of European regions in producing potatoes, in order to estimate the effect of the discovery of the potato on population growth in Europe.

Finally we wish to emphasize that while the GAEZ has devoted a great deal of attention to testing their predictions on knowledge of actual growing conditions (e.g. under controlled experiments at agricultural research stations) the GAEZ does not form its predictions by estimating any sort of statistical relationship between observed inputs around the world and observed outputs around the world. Indeed, the model outlined above illustrates how inference from such relationships could be misleading.

3.2 Output, Area and Sales Data

The second set of data on which we draw comprises county-level data from the US Agricultural Census in every decade from 1880-2002 (Haines, 2010).⁶ Of interest to us are records of actual output of each crop (\hat{Q}_i^c) the amount of land cultivated in each crop (\hat{L}_i^c) and the total value of sales (in contemporary currency units) obtained from all crops (\hat{S}_i). We use only the approximately 1,500 counties that reported agricultural output data in 1880.⁷ Although the total output of each crop in each decade in each county is known, such measures are not available for spatial units smaller than the county (such as the ‘field’, f).

3.3 Price Data

A final source of data that we use is actual data on observed producer (ie ‘farm gate’) prices. While price data is not necessary for our analysis, below we perform some simple tests of our exercise by comparing producer price data to the predicted prices that emerge from our exercise. Unfortunately, the best available price data is at the state-, rather than the county-, level. Indeed, if county-level producer price data were available the first step of our empirical analysis below, that in which we estimate local prices, would be unnecessary.

The state-level price data we use comes from two sources. First, we use the Agricultural Time Series-Cross Section Dataset (ATICS) from Cooley, DeCanio and Matthews (1977), which covers the period from 1866 (at the earliest) to 1970 (at the latest).⁸ Second, we have extracted all of the post-1970 price data available on the USDA (NASS) website so as to create a price series that extends from 1880 to 2002.

⁶While the Agricultural Census began in 1840 it was not until 1880 that the question on value of total sales was added. For this reason we begin our analysis in 1880.

⁷This figure is approximate because the exact set of counties is changing from decade to decade due to redefinitions of county borders. None of our analysis requires the ability to track specific counties across time so we work with this unbalanced panel of counties (although the exact number varies only from 1,447 to 1,562).

⁸We are extremely grateful to Paul Rhode for making a copy of this data available to us.

4 Empirical Results

This section presents preliminary empirical results, based—for now—only on output data from 1880, 1900, 1920, 1950, 1974 and 2002. We present only baseline estimates—that is, estimates that do not pursue either of the extensions (to allow for a non-agricultural good, and to allow for crop-specific technological change) outlined in Sections 2.4 and 2.5 above.

4.1 Local Crop Price Estimates

The first step of our analysis uses Theorem 1 to estimate the local price for each of our 17 crops (or upper bound on each crop that is not grown) in each of our approximately 1,500 counties, in each of our sample years (ie 1880, 1900, 1920, 1950, 1974 and 2002).

Having done this, we first ask how well these estimated prices correspond to actual price data. The procedure we follow here is not intended to be a formal test of our model (and the underlying agronomic model used by GAEZ). As mentioned before, the best producer price data available is at the state-level, whereas our price estimates are free to vary at the county-level. Our goal here is more modest. We simply aim to assess whether the price estimates emerging from our model bear any resemblance to those in the data.

In order to compare our price estimates to the state-level price data we therefore simply compute averages across all counties within each state, for each crop and year. (We do not use the price estimates that are only upper bounds on prices in calculating these averages.) We then simply regress our price estimates on the equivalent prices in the data (without a constant), year by year (on all years in our sample after the start of the Cooley et al (1977) price data, 1866), pooling across crops and states.

Table 1 contains the results of these simple regressions. In all cases we see a positive and statistically significant correlation between the two price series, with a coefficient that varies between 0.38 and 0.88.⁹ While all coefficients are less than one (as one would expect if price estimates agreed well with price data) this is unsurprising given that the regressor, actual price data, is mismeasured from our perspective because it constitutes a state-level average of underlying price observations whose sampling procedure is unknown. Given this, we consider the results in Table 1 to be encouraging. Our procedure for estimating local prices had nothing to do with price data at all—its key inputs were data on quantities and technology. But reassuringly there is a robust correlation between our price estimates and price estimates in real data.

Having examined the relationship between our local crop price estimates and outside data

⁹We have also looked at the correlation between relative (ie across crops, within state-years) price estimates and price data by running regressions across all (unique and non-trivial) such crop pairs for which data are available. The coefficients are again positive, statistically significant, and range from 0.30 to 0.81.

Table 1: Correlation between estimated prices and price data

Dependent variable:	estimated price (from model)					
	1880	1900	1920	1950	1974	2002
	(1)	(2)	(3)	(4)	(5)	(6)
observed producer price	0.876 (0.351)**	0.561 (0.247)**	0.668 (0.311)**	0.382 (0.158)**	0.712 (0.338)**	0.488 (0.110)**
observations	261	312	337	470	235	190

Notes: Results from a regression of estimated local crop prices, averaged within states, on state-level producer price data from Cooley et al (1977) and NASS. Robust standard errors in parentheses. ** indicates statistically significant at the 1% level.

on producer crop prices, we now explore the extent to which our estimated prices appear to have anything systematic to do with space. That is, do counties that are close to one another have more similar prices? Do prices decline for counties that are further and further away from major agricultural wholesale destination markets such as Chicago? While one would expect transportation costs to distort prices over space such that the answer to these questions is in the affirmative, it is entirely possible that other types of distortions (such as producer subsidies) would generate price dispersion that has no systematic correlation with distance.

Table 2 explores the extent to which proximate counties tend to have more similar prices. In all years the correlation between price gaps (here, the absolute value of the gap in log prices) and log distance is positive and statistically significant. While this is indicative of a spatial relationship in our price estimates, the coefficient estimates presented in Table 2 should not be interpreted as estimates of the structural relationship between trade costs and distance. By a free arbitrage argument, the gap in prices for a good between two markets, our dependent variable here, is only equal to the trade cost (for that good) between these two markets when the two markets are actually trading some of that good between them. For the bulk of our county pairs this will not be the case and in such settings the price gap only puts a lower bound on the trade cost (ie if trade costs were lower than the price gap then arbitrage opportunities would exist). Hence little can be said about the magnitude of the coefficients in Table 2 (for example, it is not clear how we should expect them to change over time). That said, to the extent that there exist transportation and related costs that separate markets spatially, and more so at a greater distance, we find it encouraging that our price estimates are uncovering such spatial relationships even though the spatial location of any county was not used in the construction of our price estimates.

The final analysis of our local crop price estimates that we conduct here is to evaluate whether there is a price gradient across counties with respect to their proximity to major

Table 2: Proximity of counties and local crop price estimates

Dependent variable:	absolute value of [$\log(\text{price in county } i) - \log(\text{price in county } j)$]					
	1880	1900	1920	1950	1974	2002
	(1)	(2)	(3)	(4)	(5)	(6)
$\log(\text{distance from county } i \text{ to county } j)$	0.0949 (0.0003)**	0.1014 (0.0006)**	0.0808 (0.0005)**	0.0767 (0.0002)**	0.0541 (0.0006)**	0.0742 (0.0008)**
observations	7,631,840	7,446,840	7,105,686	6,543,440	6,018,522	5,296,872

Notes: Results from a regression of the absolute value of the gap in log prices, between any (unique, non-trivial) pair of counties, and the log distance between those counties. Robust standard errors in parentheses. ** indicates statistically significant at 1% level.

agricultural wholesale markets. This is what one would expect if each county is trading at least some quantity of a crop to its nearest wholesale market, and if such trading occurs at a cost that depends on distance. For now we simply take three major wholesale markets in our sample area, Chicago, New Orleans and New York, and assign each county to the nearest of these three locations.¹⁰ We then estimate the following regression:

$$\ln p_{it}^c = \alpha_t^{c,M} + \beta_t \ln d_i + \varepsilon_{it}, \quad (24)$$

where d_i denotes the distance from county i to its nearest major wholesale market, and $\alpha_t^{c,M}$ is a separate fixed effect for each crop times each of the three major wholesale markets. These fixed effects are necessary (and important for what we do below) because they correspond to the crop-specific price at each wholesale market and year.

The results from these regressions (estimates of the coefficient β_t for each year t) are presented in Table 3. As in Table 2 we find a statistically significant correlation between prices and distance, but here the coefficient has a different interpretation, and captures more economic meaning. As long as some amount of the produce of a crop in a county is traded with its nearest major wholesale market then the coefficient here identifies both the direction of trade (here a negative sign indicates that goods are being sent to these wholesale markets, as would be expected), and, under a standard free arbitrage assumption, the extent to which increased distance increases the cost of trading (ie the parameter β_t above). In light of this, our results are encouraging since the coefficients are all negative, as would be expected, and precisely estimated. They are also revealing: the coefficients are falling in absolute value over time, from 0.142 in 1880 to 0.036 in 2002. This suggests a large—75 percent—decline in the cost of trading goods, per unit distance, over this 122 year period.

¹⁰This list of major wholesale markets can of course be enriched in future work. We aim to use a list that is crop- and year-specific, and that allows for far more wholesale markets than only the largest three.

Table 3: Local crop price estimates and proximity to major markets

Dependent variable:	log (price level)					
	1880	1900	1920	1950	1974	2002
	(1)	(2)	(3)	(4)	(5)	(6)
log (distance to closest major market)	-0.142 (0.012)**	-0.104 (0.015)**	-0.086 (0.024)**	-0.024 (0.008)**	-0.041 (0.010)**	-0.036 (0.011)**
observations	4,836	4,431	4,250	3,846	3,758	3,711

Notes: Results from a regression of the log price of each crop in each county on the log distance of that county to its nearest major wholesale market (taken to be Chicago, New Orleans or New York). Fixed effects are included for each crop times each major wholesale market. Standard errors clustered by county are in parentheses. ** indicates statistically significant at 1% level.

In the next section we estimate the gains from this large reduction in trade costs. We take seriously the distance-trade cost coefficients in each year from Table 3 and use these estimates of the structural cost of trading goods over space as the metric for how much economic integration has taken place over time (the key ingredient for estimating the gains of this heightened integration).

4.2 Gains from Economic Integration

We now turn to our preliminary estimates of the gains from economic integration. As discussed in Section 2.3 above, we formulate these gains as the answer to the following counterfactual question: “How much higher (or lower) would the total value of output across local markets in period t have been if ‘wedges’ were those of period t' rather than period t ?” Given our ability to construct the PPF for each county using the GAEZ productivity data, answering this question is straightforward once we know the prices that would prevail in each county under this counterfactual scenario. In order to formulate those prices, however, we are required to take a stand on the reference price to use in period t .

To construct reference prices we take an empirical approach inspired by the regressions that underpinned Table 3 above. Under the assumption that each county is trading each crop to its nearest ‘major’ wholesale market, the regression in Equation (24) above identifies three separate reference prices for each time period and crop, i.e. the price prevailing at each of the three major wholesale markets. Specifically, each market’s price is identified as $e^{\alpha_t^{c,M}}$, and we use the coefficient estimates $\alpha_t^{c,M}$ to calculate our reference prices in this manner.

Our final requirement in answering the above counterfactual question is a measure of the wedge for each county, crop and year. The method described in Section 2.3 above proposed that each county’s wedge be simply the gap between its price level and the reference price. While this has the attraction of simplicity, it is vulnerable to noisy price estimates. We

therefore use the fitted values from Equation (24) above (now run separately for each crop so as to estimate a separate distance coefficient β_t^c by crop and year) instead, both to construct factual price levels (that are hence smoothed over space) and to construct the counterfactual prices required to evaluate counterfactual scenarios. For example, we use the estimate $\beta_{t'}^c$ to construct the fitted values for the counterfactual prices in year t —this amounts to asking how the economy in year t would respond if it faced the counterfactual coefficient on distance from year t' (i.e. $\beta_{t'}^c$) rather than the factual coefficient on distance from year t (i.e. β_t^c).

Following this procedure we find that the gains, according to the formula in Equation (12), from moving in 1880 to 2002 distance costs are large: a 94 percent increase in the total value of agricultural output. This is considerably higher than standard estimates of the static gains from trade (for example, those in Bernhofen and Brown (2005)'s study of Japanese exit of autarky, in which the estimated gains are no more than nine percent). But this difference should not be surprising: the formula in Equation (12) implicitly captures productivity gains in the transportation sector, whereas standard estimates of the gains from trade do not.

Our preliminary estimates of gains in other years are smaller, as is to be expected from the lower estimates of β_t that we obtained in Table 3 for those years. For example, the gain from moving 1974 to 2002 distance costs is nine percent. There remains much to be done in exploring these estimates further—breaking them down by region, exploring their robustness to alternative methods for obtaining reference prices and estimating wedges, and implementing the potentially important extensions in Sections 2.4 and 2.5 above. Nevertheless these preliminary results strike us as both encouraging and plausible.

5 Concluding Remarks

In this paper we have developed a new approach to measuring the gains from economic integration based on a Roy-like assignment model in which heterogeneous factors of production are allocated to multiple sectors in multiple local markets. We have implemented this approach using data on crop markets in approximately 1,500 U.S. counties from 1880 to 2002. Central to our empirical analysis is the use of a novel agronomic data source on predicted output by crop for small spatial units. Crucially, this dataset contains information about the productivity of all spatial units for all crops, not just the endogenously selected crop that farmers at each spatial have chosen to grow in some equilibrium. Using this new approach we have estimated (i) the spatial distribution of price wedges across U.S. counties in 1880 and 2002 and (ii) the gains associated with changes in the level of these wedges over time.

A Proofs

Proof of Theorem 1. The proof of Theorem 1 proceeds in two steps.

Step 1: If p_i is competitive conditional on X_i , then there exists an efficient allocation L_i conditional on X_i such that Equations (10) and (11) hold and price levels are such that $\sum_{c \in \mathcal{C}_i^*} p_i^c \hat{Q}_i^c = \hat{S}_i$.

By Definition 1, if p_i is competitive conditional on X_i , then there exist r_i and L_i such that conditions (1)-(7) hold. Equations (4) and (5) immediately imply $\sum_{c \in \mathcal{C}_i^*} p_i^c \hat{Q}_i^c = \hat{S}_i$. Conditions (1) and (2) further imply

$$\begin{aligned} p_i^{c'} / p_i^c &= \hat{A}_i^c(f) / \hat{A}_i^{c'}(f), \text{ if } L_i^c(f) > 0 \text{ and } L_i^{c'}(f) > 0, \\ p_i^{c'} / p_i^c &\leq \hat{A}_i^c(f) / \hat{A}_i^{c'}(f), \text{ if } L_i^c(f) > 0, L_i^{c'}(f) = 0. \end{aligned}$$

Thus conditions (10) and (11) hold. Let us now check that if r_i and L_i are such that conditions (1)-(7) hold, then L_i necessarily is efficient conditional on X_i in the sense of Definition 2. By conditions (1)-(3), L_i is a feasible allocation that maximizes total profits. Thus the First Welfare Theorem (Mas-Colell et al. Proposition 5.F.1) implies that L_i must be a solution of

$$\begin{aligned} &\max_{\tilde{L}_i} \sum_{f \in \mathcal{F}} \left[\hat{A}_i^{c_i^*}(f) / \alpha_i \right] \tilde{L}_i^{c_i^*}(f) \\ &\sum_{f \in \mathcal{F}} \left[\hat{A}_i^c(f) / \alpha_i \right] \tilde{L}_i^c(f) \geq \hat{Q}_i^c, \text{ for all } c \neq c_i^*, \\ &\sum_{c \in \mathcal{C}} \tilde{L}_i^c(f) \leq \hat{L}_i(f), \text{ for all } f \in \mathcal{F}, \\ &\tilde{L}_i^c(f) \geq 0, \text{ for all } c \in \mathcal{C}, f \in \mathcal{F}, \end{aligned}$$

where we have arbitrarily chosen c_i^* such that $c_i^* \equiv \min \{c \in \mathcal{C} | \hat{Q}_i^c > 0\}$. Since Equation (5) holds for $c = c_i^*$, we must have $\hat{Q}_i^{c_i^*} = \sum \left[\hat{A}_i^{c_i^*}(f) / \alpha_i \right] L_i^{c_i^*}(f)$, which can be rearranged as $\alpha_i = \left[\sum \hat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right] / \hat{Q}_i^{c_i^*}$. Accordingly, L_i is a solution of

$$\max_{\tilde{L}_i} \sum_{f \in \mathcal{F}} \hat{A}_i^{c_i^*}(f) \tilde{L}_i^{c_i^*}(f) \tag{P'}$$

$$\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) \tilde{L}_i^c(f) \geq \left(\hat{Q}_i^c / \hat{Q}_i^{c_i^*} \right) \sum_{f \in \mathcal{F}} \hat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f), \text{ for all } c \in \mathcal{C}, \tag{25}$$

$$\sum_{c \in \mathcal{C}} \tilde{L}_i^c(f) \leq \hat{L}_i(f), \text{ for all } f \in \mathcal{F}, \tag{26}$$

$$\tilde{L}_i^c(f) \geq 0, \text{ for all } c \in \mathcal{C}, f \in \mathcal{F}, \tag{27}$$

This establishes that L_i satisfies all the constraints of (P). In order to demonstrate that L_i is a solution of (P), we now proceed by contradiction. Suppose that there exists L'_i satisfying all the constraints in (P) such that

$$\min_{c \in \mathcal{C}_i^*} \left[\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i'^c(f) \right] / \hat{Q}_i^c > \min_{c \in \mathcal{C}_i^*} \left[\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f) \right] / \hat{Q}_i^c. \tag{28}$$

Since Inequality (25) is necessarily binding at a solution of (P') , we therefore have

$$\left[\sum_{f \in \mathcal{F}} \widehat{A}_i^c(f) L_i^c(f) \right] / \widehat{Q}_i^c = \left[\sum_{f \in \mathcal{F}} \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right] / \widehat{Q}_i^{c_i^*}, \text{ for all } c \in \mathcal{C}_i^*. \quad (29)$$

Combining Inequality (28) and Equation (29), we obtain

$$\left[\sum_{f \in \mathcal{F}} \widehat{A}_i^c(f) L_i'^c(f) \right] / \widehat{Q}_i^c > \left[\sum_{f \in \mathcal{F}} \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right] / \widehat{Q}_i^{c_i^*}, \text{ for all } c \in \mathcal{C}_i^*. \quad (30)$$

Since $L_i'^c(f) \geq 0$ for all $c \in \mathcal{C}$, $f \in \mathcal{F}$, we also trivially have

$$\sum \widehat{A}_i^c(f) L_i'^c(f) \geq \left(\widehat{Q}_i^c / \widehat{Q}_i^{c_i^*} \right) \sum \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f), \text{ for all } c \notin \mathcal{C}_i^*. \quad (31)$$

By Inequalities (30) and (31), L_i' satisfies all the constraints in (P') . By Inequality (30) evaluated at $c = c_i^*$, we also have

$$\left[\sum_{f \in \mathcal{F}} \widehat{A}_i^{c_i^*}(f) L_i'^{c_i^*}(f) \right] / \widehat{Q}_i^{c_i^*} > \left[\sum_{f \in \mathcal{F}} \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right] / \widehat{Q}_i^{c_i^*}.$$

This contradicts the fact that L_i is a solution of (P') . Thus L_i is efficient conditional on X_i in the sense of Definition 2. This completes the proof of Step 1.

Step 2: *If there exists an efficient allocation L_i conditional on X_i such that Equations (10) and (11) hold and price levels are such that $\sum_{c \in \mathcal{C}_i^*} p_i^c \widehat{Q}_i^c = \widehat{S}_i$, then p_i is competitive conditional on X_i .*

By assumption, the observation X_i is such that Equations (4)-(7) hold. Thus by Definition 1, we only need to show that one can construct a vector of field prices, r_i , and an allocation of fields, L_i , such that conditions (1)-(3) hold as well.

A natural candidate for the allocation is L_i solution of (P) such that conditions (10) and (11) hold. By Definition 2, such a solution exists since there exists an efficient allocation L_i conditional on X_i such that conditions (10) and (11) hold. The fact that Inequality (3) holds for allocation L_i is immediate from Equation (6) and Inequality (8). Let us now construct the vector of field prices, r_i , such that for

$$r_i(f) = \max_{c \in \mathcal{C}} p_i^c \widehat{A}_i^c(f) \widehat{Q}_i^{c_i^*} / \left[\sum_{f \in \mathcal{F}} \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right], \quad (32)$$

with $c_i^* \equiv \min \left\{ c \in \mathcal{C} \mid \widehat{Q}_i^c > 0 \right\}$. Since Equation (5) holds for $c = c_i^*$, Equation (7) implies $\alpha_i = \left[\sum_{f \in \mathcal{F}} \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right] / \widehat{Q}_i^{c_i^*}$. Thus we can rearrange Equation (32) as

$$r_i(f) = \max_{c \in \mathcal{C}} p_i^c A_i^c(f).$$

This immediately implies Inequality (1). To conclude, all we need to show is that $p_i^c A_i^c(f) = r_i(f)$ if $L_i^c(f) > 0$. We proceed by contradiction. Suppose that there exist $c \in \mathcal{C}$ and $f \in \mathcal{F}$ such that $L_i^c(f) > 0$ and $p_i^c A_i^c(f) < \max_{c' \in \mathcal{C}} p_i^{c'} A_i^{c'}(f)$. By Equation (7), this can

be rearranged as $p_i^c \widehat{A}_i^c(f) < \max_{c' \in \mathcal{C}} p_i^{c'} \widehat{A}_i^{c'}(f)$. Now consider $c_0 = \arg \max_{c' \in \mathcal{C}} p_i^{c'} \widehat{A}_i^{c'}(f)$. By construction, we have

$$p_i^{c_0} / p_i^c > \widehat{A}_i^c(f) / \widehat{A}_i^{c_0}(f),$$

which contradicts either condition (10), if $L_i^{c_0}(f) > 0$, or condition (11), if $L_i^{c_0}(f) = 0$. This completes the proof of Step 2.

Theorem 1 directly derives from Steps 1 and 2. **QED.** ■

Proof of Theorem 1(M). The proof of Theorem 1(M) is similar to the previous proof.

Step 1: If p_i is competitive conditional on Y_i , then there exists an efficient allocation L_i conditional on Y_i such that Equations (10) and (11) hold and price levels are such that $\sum_{c \in \mathcal{C}_i^*} p_i^c \widehat{Q}_i^c = \widehat{S}_i$.

By Definition 1 (M), if p_i is feasible conditional on Z_i , then there exist r_i and L_i such that (1)-(7) and (13) hold. Equations (4) and (5) immediately imply $\sum_{c \in \mathcal{C}_i^*} p_i^c \widehat{Q}_i^c = \widehat{S}_i$. Conditions (1) and (2) further imply

$$\begin{aligned} p_i^{c'} / p_i^c &= \widehat{A}_i^c(f) / \widehat{A}_i^{c'}(f), \text{ if } L_i^c(f) > 0 \text{ and } L_i^{c'}(f) > 0, \\ p_i^{c'} / p_i^c &\leq \widehat{A}_i^c(f) / \widehat{A}_i^{c'}(f), \text{ if } L_i^c(f) > 0, L_i^{c'}(f) = 0. \end{aligned}$$

Thus conditions (10) and (11) hold. Let us now check that if r_i and L_i are such that conditions (1)-(7) hold, then L_i necessarily is efficient conditional on X_i in the sense of Definition 2 (M). By conditions (1)-(3), L_i is a feasible allocation that maximizes total profits. Thus the First Welfare Theorem (Mas-Colell et al. Proposition 5.F.1) implies that L_i must be a solution of

$$\begin{aligned} &\max_{\widetilde{L}_i} \sum_{f \in \mathcal{F}} \left[\widehat{A}_i^{c_i^*}(f) / \alpha_i \right] \widetilde{L}_i^{c_i^*}(f) \\ &\sum_{f \in \mathcal{F}} \left[\widehat{A}_i^c(f) / \alpha_i \right] \widetilde{L}_i^c(f) \geq \widehat{Q}_i^c, \text{ for all } c \neq c_i^*, \\ &\sum_{f \in \mathcal{F}} A_i^m \widetilde{L}_i^m(f) \geq Q_i^m, \\ &\sum_{j \in \mathcal{C} \cup \{m\}} \widetilde{L}_i^j(f) \leq \widehat{L}_i(f), \text{ for all } f \in \mathcal{F}, \\ &\widetilde{L}_i^j(f) \geq 0, \text{ for all } j \in \mathcal{C} \cup \{m\}, f \in \mathcal{F}, \end{aligned}$$

where we have arbitrarily chosen c_i^* such that $c_i^* \equiv \min \{c \in \mathcal{C} | \widehat{Q}_i^c > 0\}$. Since Equation (5) holds for $c = c_i^*$, we must have $\widehat{Q}_i^{c_i^*} = \sum \left[\widehat{A}_i^{c_i^*}(f) / \alpha_i \right] L_i^{c_i^*}(f)$, which can be rearranged $\alpha_i = \left[\sum \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right] / \widehat{Q}_i^{c_i^*}$. By Equation (13) we must also have $Q_i^m = A_i^m \widehat{L}_i^m$. Thus we

can rearrange the previous program as

$$\max_{\tilde{L}_i} \sum_{f \in \mathcal{F}} \hat{A}_i^{c_i^*}(f) \tilde{L}_i^{c_i^*}(f) \quad (\text{P-M}') \quad (32)$$

$$\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) \tilde{L}_i^c(f) \geq \left(\hat{Q}_i^c / \hat{Q}_i^{c_i^*} \right) \sum_{f \in \mathcal{F}} \hat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f), \text{ for all } c \in \mathcal{C}, \quad (33)$$

$$\sum_{f \in \mathcal{F}} \tilde{L}_i^m(f) \geq \hat{L}_i^m, \quad (34)$$

$$\sum_{j \in \mathcal{C} \cup \{m\}} \tilde{L}_i^j(f) \leq \hat{L}_i(f), \text{ for all } f \in \mathcal{F}, \quad (35)$$

$$\tilde{L}_i^j(f) \geq 0, \text{ for all } j \in \mathcal{C} \cup \{m\}, f \in \mathcal{F}. \quad (36)$$

This establishes that L_i satisfies all the constraints of $(P - M)$. In order to demonstrate that L_i is a solution of $(P - M)$, we now proceed by contradiction. Suppose that there exists L'_i satisfying all the constraints in $(P - M)$ such that

$$\min_{c \in \mathcal{C}_i^*} \left[\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i'^c(f) \right] / \hat{Q}_i^c > \min_{c \in \mathcal{C}_i^*} \left[\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f) \right] / \hat{Q}_i^c. \quad (37)$$

Since Inequality $(P - M')$ is necessarily binding at a solution of $(P - M')$, we therefore have

$$\left[\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f) \right] / \hat{Q}_i^c = \left[\sum_{f \in \mathcal{F}} \hat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right] / \hat{Q}_i^{c_i^*}, \text{ for all } c \in \mathcal{C}_i^*. \quad (38)$$

Combining Inequality (37) and Equation (38), we obtain

$$\left[\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i'^c(f) \right] / \hat{Q}_i^c > \left[\sum_{f \in \mathcal{F}} \hat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right] / \hat{Q}_i^{c_i^*}, \text{ for all } c \in \mathcal{C}_i^*. \quad (39)$$

Since $L_i'^c(f) \geq 0$ for all $c \in \mathcal{C}$, $f \in \mathcal{F}$, we also trivially have

$$\sum \hat{A}_i^c(f) L_i'^c(f) \geq \left(\hat{Q}_i^c / \hat{Q}_i^{c_i^*} \right) \sum \hat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f), \text{ for all } c \notin \mathcal{C}_i^*. \quad (40)$$

By Inequalities (39) and (40), L'_i satisfies all the constraints in $(P - M')$. By Inequality (39) evaluated at $c = c_i^*$, we also have

$$\left[\sum_{f \in \mathcal{F}} \hat{A}_i^{c_i^*}(f) L_i'^{c_i^*}(f) \right] / \hat{Q}_i^{c_i^*} > \left[\sum_{f \in \mathcal{F}} \hat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) \right] / \hat{Q}_i^{c_i^*}.$$

This contradicts the fact that L_i is a solution of $(P - M')$. Thus L_i is efficient conditional on Y_i in the sense of Definition 1 (M). This completes the proof of Step 1.

Step 2: *If there exists an efficient allocation L_i conditional on Y_i such that Equations (10) and (11) hold and price levels are such that $\sum_{c \in \mathcal{C}_i^*} p_i^c \hat{Q}_i^c = \hat{S}_i$, then p_i is competitive conditional on Y_i .*

By assumption, the observation Y_i is such that Equations (4)-(7) and (13) hold. Thus by Definition 1 (M), we only need to show that one can construct a vector of field prices, r_i , and an allocation of fields, L_i , such that conditions (1)-(3) hold as well.

A natural candidate for the allocation is L_i solution of $(P - M)$ such that conditions

(10) and (11) hold. By Definition 2 (M), such a solution exists since there exists an efficient allocation L_i conditional on Y_i such that conditions (10) and (11) hold. The fact that Inequality (3) holds for allocation L_i is immediate from Equation (6) and Inequality (14). Let us now construct the vector of field prices, r_i , such that for

$$r_i(f) = \max_{c \in \mathcal{C}} p_i^c \widehat{A}_i^c(f) \widehat{Q}_i^{c^*} / [\sum_{f \in \mathcal{F}} \widehat{A}_i^{c^*}(f) L_i^{c^*}(f)], \quad (41)$$

with $c_i^* \equiv \min \{c \in \mathcal{C} | \widehat{Q}_i^c > 0\}$. Since Equation (5) holds for $c = c_i^*$, Equation (7) implies $\alpha_i = [\sum_{f \in \mathcal{F}} \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f)] / \widehat{Q}_i^{c_i^*}$. Thus we can rearrange Equation (41) as

$$r_i(f) = \max_{c \in \mathcal{C}} p_i^c A_i^c(f).$$

This immediately implies that Inequality (1) holds for all $c \in \mathcal{C}$. To conclude, all we need to show is that $p_i^c A_i^c(f) = r_i(f)$ if $L_i^c(f) > 0$. We proceed by contradiction. Suppose that there exist $c \in \mathcal{C}$ and $f \in \mathcal{F}$ such that $L_i^c(f) > 0$ and $p_i^c A_i^c(f) < \max_{c' \in \mathcal{C}} p_i^{c'} A_i^{c'}(f)$. By Equation (7), this can be rearranged as $p_i^c \widehat{A}_i^c(f) < \max_{c' \in \mathcal{C}} p_i^{c'} \widehat{A}_i^{c'}(f)$. Now consider $c_0 = \arg \max_{c' \in \mathcal{C}} p_i^{c'} \widehat{A}_i^{c'}(f)$. By construction, we have

$$p_i^{c_0} / p_i^c > \widehat{A}_i^c(f) / \widehat{A}_i^{c_0}(f),$$

which contradicts either condition (10), if $L_i^{c_0}(f) > 0$, or condition (11), if $L_i^{c_0}(f) = 0$. This completes the proof of Step 2.

Theorem 1(M) directly derives from Steps 1 and 2. **QED.** ■

Proof of Theorem 1(C). The proof of Theorem 1(C) is similar to the previous proofs.

Step 1: If p_i is competitive conditional on Z_i , then there exists an efficient allocation L_i conditional on Z_i such that Equations (22) and (23) hold and price levels are such that $\sum_{c \in \mathcal{C}_i^*} p_i^c \widehat{Q}_i^c = \widehat{S}_i$.

By Definition 1(C), if p_i is competitive conditional on Z_i , then there exist r_i and L_i such that conditions (1)-(6) and (18)-(17) hold. Equations (4) and (5) immediately imply $\sum_{c \in \mathcal{C}_i^*} p_i^c \widehat{Q}_i^c = \widehat{S}_i$. Conditions (1) and (2) further imply

$$\begin{aligned} p_i^{c'} / p_i^c &= \widehat{A}_i^c(f) \alpha_i^{c'} / \widehat{A}_i^{c'}(f) \alpha_i^c, \text{ if } L_i^c(f) > 0 \text{ and } L_i^{c'}(f) > 0, \\ p_i^{c'} / p_i^c &\leq \widehat{A}_i^c(f) \alpha_i^{c'} / \widehat{A}_i^{c'}(f) \alpha_i^c, \text{ if } L_i^c(f) > 0, L_i^{c'}(f) = 0. \end{aligned}$$

Thus conditions (22) and (23) hold. Let us now check that if r_i and L_i are such that conditions (1)-(6) and (18)-(17) hold, then L_i necessarily is efficient conditional on Z_i in the sense of Definition 2(C). By conditions (1)-(3), L_i is a feasible allocation that maximizes total profits. By Equation (18), it also satisfies $\sum_{f \in \mathcal{F}} \widetilde{L}_i^c(f) = \widehat{L}_i^c$. Thus, by the First Welfare

Theorem (Mas-Colell et al. Proposition 5.F.1), L_i must be a solution of

$$\begin{aligned} & \max_{\tilde{L}_i} \sum_{f \in \mathcal{F}} \left[\hat{A}_i^{c_i^*}(f) / \alpha_i^{c_i^*} \right] \tilde{L}_i^{c_i^*}(f) \\ & \sum_{f \in \mathcal{F}} \left[\hat{A}_i^c(f) / \alpha_i^c \right] \tilde{L}_i^c(f) \geq \hat{Q}_i^c, \text{ for all } c \neq c_i^*, \\ & \sum_{c \in \mathcal{C}} \tilde{L}_i^c(f) \leq \hat{L}_i(f), \text{ for all } f \in \mathcal{F}, \\ & \tilde{L}_i^c(f) \geq 0, \text{ for all } c \in \mathcal{C}, f \in \mathcal{F}, \\ & \sum_{f \in \mathcal{F}} \tilde{L}_i^c(f) = \hat{L}_i^c, \text{ for all } c \in \mathcal{C}. \end{aligned}$$

where we have arbitrarily chosen c_i^* such that $c_i^* \equiv \min \left\{ c \in \mathcal{C} \mid \hat{Q}_i^c > 0 \right\}$. Since Equation (5) holds for all c , we must have $\hat{Q}_i^c = \sum \left[\hat{A}_i^c(f) / \alpha_i^c \right] L_i^c(f)$, which can be rearranged as $\alpha_i^c = \left[\sum \hat{A}_i^c(f) L_i^c(f) \right] / \hat{Q}_i^c$ for all $c \in \mathcal{C}_i^*$. Accordingly, L_i is a solution of

$$\max_{\tilde{L}_i} \sum_{f \in \mathcal{F}} \hat{A}_i^{c_i^*}(f) \tilde{L}_i^{c_i^*}(f) \quad (\text{P-C}') \tag{41}$$

$$\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) \tilde{L}_i^c(f) \geq \sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f), \text{ for all } c \neq c_i^*, \tag{42}$$

$$\sum_{c \in \mathcal{C}} \tilde{L}_i^c(f) \leq \hat{L}_i(f), \text{ for all } f \in \mathcal{F}, \tag{43}$$

$$\tilde{L}_i^c(f) \geq 0, \text{ for all } c \in \mathcal{C}, f \in \mathcal{F}, \tag{44}$$

$$\sum_{f \in \mathcal{F}} \tilde{L}_i^c(f) = \hat{L}_i^c, \text{ for all } c \in \mathcal{C}. \tag{45}$$

This establishes that L_i satisfies all the constraints of $(P - C)$. In order to demonstrate that L_i is a solution of $(P - C)$, we now proceed by contradiction. Suppose that there exists L'_i satisfying all the constraints in $(P - C)$ such that

$$\min_{c \in \mathcal{C}_i^*} \left\{ \frac{\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i'^c(f)}{\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f)} \right\} > \min_{c \in \mathcal{C}_i^*} \left\{ \frac{\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f)}{\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f)} \right\} = 1. \tag{46}$$

Inequality (46) immediately implies

$$\sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i'^c(f) > \sum_{f \in \mathcal{F}} \hat{A}_i^c(f) L_i^c(f) \text{ for all } c \in \mathcal{C}_i^*. \tag{47}$$

Since L_i is such that Equation (5) holds for all c , we also know that $L_i^c(f) = 0$ for all $f \in \mathcal{F}, c \notin \mathcal{C}_i^*$. Since $L_i'^c(f) \geq 0$ for all $c \in \mathcal{C}, f \in \mathcal{F}$, we also trivially have

$$\sum \hat{A}_i^c(f) L_i'^c(f) \geq \sum \hat{A}_i^c(f) L_i^c(f), \text{ for all } c \notin \mathcal{C}_i^*. \tag{48}$$

By Inequalities (47) and (48), L'_i satisfies all the constraints in $(P - C')$. By Inequality (47)

evaluated at $c = c_i^*$, we also have

$$\sum_{f \in \mathcal{F}} \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f) > \sum_{f \in \mathcal{F}} \widehat{A}_i^{c_i^*}(f) L_i^{c_i^*}(f).$$

This contradicts the fact that L_i is a solution of $(P - C')$. Thus L_i is efficient conditional on Z_i in the sense of Definition 2(C). This completes the proof of Step 1.

Step 2: *If there exists an efficient allocation L_i conditional on Z_i such that Equations (22) and (23) hold and price levels are such that $\sum_{c \in \mathcal{C}_i^*} p_i^c \widehat{Q}_i^c = \widehat{S}_i$, then p_i is competitive conditional on Z_i .*

By assumption, the observation Z_i is such that Equations (4)-(6) and (18)-(17) hold. Thus by Definition 1(C), we only need to show that one can construct a vector of field prices, r_i , and an allocation of fields, L_i , such that conditions (1)-(3) hold as well.

A natural candidate for the allocation is L_i solution of $(P - C)$ such that conditions (22) and (23) hold. By Definition 2(C), such a solution exists since there exists an efficient allocation L_i conditional on Z_i such that conditions (22) and (23) hold. The fact that Inequality (3) holds for allocation L_i is immediate from Equation (6) and Inequality (19). Let us now construct the vector of field prices, r_i , such that for

$$r_i(f) = \max_{c \in \mathcal{C}_i^*} p_i^c \widehat{A}_i^c(f) \widehat{Q}_i^c / [\sum_{f \in \mathcal{F}} \widehat{A}_i^c(f) L_i^c(f)], \quad (49)$$

with $c_i^* \equiv \min \{c \in \mathcal{C} | \widehat{Q}_i^c > 0\}$. Since Equation (5) holds for all c , Equation (17) implies $\alpha_i^c = [\sum_{f \in \mathcal{F}} \widehat{A}_i^c(f) L_i^c(f)] / \widehat{Q}_i^c$ for all $c \in \mathcal{C}_i^*$. Thus we can rearrange Equation (49) as

$$r_i(f) = \max_{c \in \mathcal{C}_i^*} p_i^c A_i^c(f).$$

Under the assumption $\mathcal{C} = \mathcal{C}_i^*$, this immediately implies Inequality (1). To conclude, all we need to show is that $p_i^c A_i^c(f) = r_i(f)$ if $L_i^c(f) > 0$. We proceed by contradiction. Suppose that there exist $c \in \mathcal{C}$ and $f \in \mathcal{F}$ such that $L_i^c(f) > 0$ and $p_i^c A_i^c(f) < \max_{c' \in \mathcal{C}_i^*} p_i^{c'} A_i^{c'}(f)$. By Equation (17), this can be rearranged as $p_i^c \widehat{A}_i^c(f) / \alpha_i^c < \max_{c' \in \mathcal{C}} p_i^{c'} \widehat{A}_i^{c'}(f) / \alpha_i^{c'}$. Now consider $c_0 = \arg \max_{c' \in \mathcal{C}} p_i^{c'} \widehat{A}_i^{c'}(f) / \alpha_i^{c'}$. By construction, we have

$$p_i^{c_0} / p_i^c > \widehat{A}_i^c(f) \alpha_i^{c_0} / \widehat{A}_i^{c_0}(f) \alpha_i^c,$$

which contradicts either condition (22), if $L_i^{c_0}(f) > 0$, or condition (23), if $L_i^{c_0}(f) = 0$. This completes the proof of Step 2.

Theorem 1(C) directly derives from Steps 1 and 2. **QED.** ■

References

- Arkolakis, C., A. Costinot and A. Rodriguez-Clare (2011): "New Trade Models, Same Old Gains?," *American Economic Review* (forthcoming).
- Bernhofen, D. and J. Brown (2005): "An Empirical Assessment of the Comparative Advantage Gains from Trade: Evidence from Japan," *American Economic Review*, 95(1), pp. 208-225.
- Caliendo, L. and F. Parro (2010): "Estimates of the Trade and Welfare Effects of NAFTA," Unpublished manuscript, Yale University.
- Chetty, Raj (2009), "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, pp. 451-488.
- Chor, D. (2010): "Unpacking Sources of Comparative Advantage: A Quantitative Approach," *Journal of International Economics*, 82(2), pp. 152-167.
- Cooley, T., S. De Canio and M. Matthews (1977): "An Agricultural Time Series-Cross Section Data Set," NBER Working Paper No. 197.
- Costinot, Arnaud (2009), "An Elementary Theory of Comparative Advantage," *Econometrica*, 77(4), pp. 1165-1192.
- Costinot, A., D. Donaldson and I. Komunjer (2011): "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies* (forthcoming).
- Dekle, R., J. Eaton and S. Kortum (2008): "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, 55, pp. 511-540.
- Donaldson, D. (2010): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," Unpublished manuscript, MIT.
- Eaton, Jonathan and Samuel Kortum (2002), "Technology, Geography and Trade," *Econometrica*, 70(5), pp. 1741-1779.
- Fielor, A. (2011): "Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation," *Econometrica*, 79(4), 1069-1101.
- Frankel, Jeffrey A. and David Romer, "Does Trade Cause Growth?," *American Economic Review*, 89(3), pp. 379-399.
- Hsieh, Chang-Tai and Pete Klenow (2009), Misallocation and Manufacturing TFP in China and India, *Quarterly Journal of Economics*, 124(4), pp. 1403-1448.
- Haines, Michael R (2010), "Inter-university Consortium for Political and Social Research, Historical, Demographic, Economic, and Social Data: The United States, 1790-2002," [Computer file].
- Keller, W. and C.H. Shiue (2008): "Markets in China and Europe on the Eve of the Industrial Revolution," *American Economic Review*, 97(4), pp. 1189-1216.

- Nunn, Nathan and Nancy Qian (2011), "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment," *Quarterly Journal of Economics*, 126(2), pp. 593-650.
- Ramondo and Rodriguez-Clare (2010), "Trade, Multinational Production, and the Gains from Openness", unpublished manuscript, Penn State University.
- Shiue, C.H. (2005): "Transport Costs and the Geography of Arbitrage in Eighteenth-Century China," *American Economic Review*, 92(5), pp. 1406-1419.
- Waugh, Michael (2010), "International Trade and Income Differences," *American Economic Review*, 100(5), pp. 2093-2124.