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

February, 2012

Abstract

In this paper we develop a new strutural approach to measuring the gains from economic integration based on a Ricardian model in which heterogeneous factors of production are allocated to multiple sectors in multiple local markets based on comparative advantage. We implement our 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 units for all crops, not just those that are actually being grown. Using this new approach we find that the long-run gains from economic integration among US agricultural markets have been substantial.

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.

The standard approach in the international trade literature consists in estimating or calibrating fully specified models of how countries behave under any trading regime. Eaton and Kortum (2002) is the most influential application of this approach. 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 untestable 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 these agronomic predictions about factual and counterfactual technologies with an assignment model in which heterogeneous fields are allocated to multiple crops in multiple local markets based on comparative advantage.

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 experiencing large changes in transportation costs. 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—analogous to countries in a standard trade model. Each county is endowed with many 'fields' of arable land.[1] 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,

[1] While 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.

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 2000.

Our empirical analysis relies on one key identifying assumption: the pattern of comparative advantage of fields across crops within counties is stable over time. That is, if agronomists predict that a field is 10% more productive at producing wheat than corn in 2000 compared to another field in the same county, then we assume that is 10% more productive in all prior years, though productivity levels are free to vary across crops, counties and years. Under this assumption, we first demonstrate how one can combine modern GAEZ data and historical Census data to identify the spatial distribution of crop prices and crop-specific productivity shocks across U.S. counties over time. The basic idea is to find the vector of crop-specific productivity shocks such that the predictions of our assignment model exactly match total output per crop as well as the total acres of land allocated to each crop in each county. Using classical results in general equilibrium theory, we first provide mild sufficient conditions—which we will be able to test in the data—under which such a vector of shocks exists and is unique. Having identified productivity shocks, we then back out the vector of crop prices that, according to the model, must have supported this allocation as an equilibrium outcome. The difference or wedge between "local" crop prices, estimated from the model at the county level, and "world" prices, observed in historical data, finally give us a measure of trade costs between each U.S. county and the rest of the world from 1880 to 2002.

In order to quantify the gains from economic integration, we focus on the following counterfactual question: "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 trade costs were those of period t' rather than period t?" In our assignment model, our estimates of historical productivity shocks and the modern GAEZ data provide sufficient information to construct the production possibility frontier associated with each U.S. county at any point in time. Thus at this point, addressing the previous counterfactual question reduces to solving a simple linear programming problem. We find that the gains from economic integration, measured in this way, are substantial. For example, the estimated gains from 1880-1920 amount to 148% of 1880 agricultural output and those from 1950-1997 amount to 98% of 1950 output.

Another attractive feature of our new structural approach is that it allows us to estimate simultaneously trade costs and productivity shocks. Thus we can compare—using the same theoretical framework and the same data sources—how the gains from economic integration compare to productivity gains in agriculture over that same period. Formally, we 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 crop-and-county-specific productivity shocks were those of period t' rather than period t? Answering this question again boils down to solving a simple linear programming problem. We find that the gains from trade cost reductions are similar in magnitude to those of agricultural productivity improvements.

In the existing 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 Fieler (2011), Ossa (2011), Levchenko and Zhang (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 county-level in the United States over the extended time period that we consider.

Our paper 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.

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 the data that feeds into our analysis. Section 4 presents our main empirical results and Section 5 explores the robustness of those results. Finally, Section 6 concludes. All formal proofs can be found in the Appendix.

2 Theoretical Framework

In this section we describe the theoretical framework that we use throughout our analysis. Our approach can be broken into two steps. First, we show (using Theorem 1 and Corollary 1 in Section 2.2 below) that available data on aggregate (ie county-wide) production and land allocation can be used, along with agronomic data on the productivity of fields across crops, to infer the farm-gate prices that farmers appear to be facing, as well as unobserved shocks to farmers' productivity, for any crop, county and year in our sample. Second, we describe (in Section 2.3 below) how these estmated farm-gate prices can be used to infer the costs that farmers appear to be facing to trade goods, as well as how to calculate the gains (or losses) that would obtain under a counterfactual change to these trade costs.

2.1 Endowments, Technology, and Market Structure

Our theoretical framework is a comparative advantage-based assignment model, as in Costinot (2009). At any date t we consider an economy with multiple local markets indexed by $i \in \mathcal{I} \equiv \{1, ..., I\}$—in which production occurs—and one wholesale market—in which goods are sold. In our empirical analysis, local markets will be U.S. counties. In each local market, the only factors of production are different types of land or fields indexed by $f \in \mathcal{F}_i \equiv \{1, ..., F_i\}$. We denote by $V_i^f \geq 0$ the number of acres covered by field f in market i. Fields can be used to produce multiple goods indexed by $k \in \mathcal{K} \equiv \{1, ..., K+1\}$. In our empirical analysis, goods $1, ..., K$ will be crops (of which there are 16 in our sample), whereas good $K+1$ will be an outside good. We think of the outside good as manufacturing, forestry, or any agricultural activity (such as livestock production) that does not correspond to the crops included in our dataset.

Fields are perfect substitutes in the production of each good, but vary in their exogenously-given productivity per acre, $A_{it}^{fk} > 0$. Total output Q_{it}^k of good k in market i at date t is given by

$$Q_{it}^k = \sum_{f \in \mathcal{F}_i} A_{it}^{fk} L_{it}^{fk}, \tag{1}$$

where $L_{it}^{fk} \geq 0$ denotes the endogenous number of acres of field f allocated to good k in market i at date t. Note that A_{it}^{fk} may vary both with f and k. Thus although fields are perfect substitutes in the production of each good, some fields may have a comparative as well as absolute advantage in producing particular goods.

All goods are produced by a large number of price-taking farms in local markets and then shipped to the wholesale market. The profits of a representative farm producing good k in a local market i at date t are given by

$$\Pi_{it}^{k} = \left(\frac{\bar{p}_t^k}{1+\tau_{it}^k}\right)\left(\sum_{f\in\mathcal{F}} A_{it}^{fk} L_{it}^{fk}\right) - \sum_{f\in\mathcal{F}} r_{it}^{f} L_{it}^{fk},$$

where $\bar{p}_t^k$ is the selling price of good k in the wholesale market; $\tau_{it}^k \geq 0$ is an iceberg trade cost associated with shipping good k from i to the wholesale market; and r_{it}^f is the rental rate per acre of field f in local market i at date t. We denote by $p_{it}^k \equiv \bar{p}_t^k / \left(1+\tau_{it}^k\right)$ the farm-gate price of good k in market i at date t. Profit maximization by farms further requires

$$p_{it}^k A_{it}^{kf} - r_{it}^f \leq 0, \text{ for all } k \in \mathcal{K},\ f \in \mathcal{F}_i, \tag{2}$$

$$p_{it}^k A_{it}^{kf} - r_{it}^f = 0, \text{ for all } k \in \mathcal{K},\ f \in \mathcal{F}_i \text{ such that } L_{it}^{fk} > 0. \tag{3}$$

Local factor markets are segmented (ie land cannot move). Thus factor market clearing requires

$$\sum_{k\in\mathcal{K}} L_{it}^{fk} = V_i^f, \text{ for all } f \in \mathcal{F}_i. \tag{4}$$

We leave goods market clearing conditions unspecified, thereby treating the wholesale market as a small open economy. Formally, $\bar{p}_t \equiv (\bar{p}_t^k)_{k\in\mathcal{K}}$ is exogenously given. In the remainder of this paper we denote by $p_{it} \equiv \left(p_{it}^k\right)_{k\in\mathcal{K}}$ the vector of farm gate prices, $r_{it} \equiv (r_{it}^f)_{f\in\mathcal{F}}$ the vector of field prices, and $L_{it} \equiv (L_{it}^{fk})_{k\in\mathcal{K}, f\in\mathcal{F}}$ the allocation of fields to goods in local market i. Armed with this notation, we formally define a competitive equilibrium as follows.

Definition 1 *A competitive equilibrium in a local market i at date t is a field allocation, L_{it}, and a price system, (p_{it}, r_{it}), such that conditions* (2)-(4) *hold.*

2.2 Measuring Local Prices and Productivity Shocks

In this section, we describe how we use theory and data to infer measures of local prices and productivity across time and space. We separate this description into two parts. The first and most important part focuses on "non-zero" crops, i.e. crops for which we observe production in a given location at a particular point in time, while the second one deals with "zero" crops, i.e. crops for which we do not.

"Non-zero" crops. Our dataset contains, for each local market $i \in \mathcal{I}$, year t, and crop $k \in \mathcal{K}/\left\{K+1\right\}$,historical measures of each of the following variables: (i) total farms' sales, $\hat{S}_{it}$, (ii) total output per crop, $\hat{Q}_{it}^k$, (iii) total acres of land allocated to each crop, $\widehat{L}_{it}^k$, as well as (iv) total acres of land covered by each field, $\hat{V}_i^f$. Throughout our empirical analysis, we assume (in Assumption A1) that none of these variables is subject to measurement error—that is, that these variables in the model (written without hats) equal their equivalents in the data (written with hats).

A1. *In all local markets $i \in \mathcal{I}$ and at all dates t, we assume that*

$$\begin{aligned} \textstyle\sum_{k\in\mathcal{K}/\{K+1\}} p_{it}^k Q_{it}^k &= \hat{S}_{it}, & (5)\\ Q_{it}^k &= \hat{Q}_{it}^k, \text{ for all } k \in \mathcal{K}/\{K+1\}, & (6)\\ \textstyle\sum_{f\in\mathcal{F}_i} L_{it}^{fk} &= \widehat{L}_{it}^k, \text{ for all } k \in \mathcal{K}/\{K+1\}, & (7)\\ V_i^f &= \hat{V}_i^f, \text{ for all } f \in \mathcal{F}_i. & (8) \end{aligned}$$

By contrast, we do not have access to historical productivity measures. Instead we have access to measures of productivity per acre, $\hat{A}_{i2011}^{fk}$, for each field in each market if that field were to be allocated to the production of crop k in 2011—the agronomists who have assembled the GAEZ project data in 2011 aim for it to be relevant to contemporaneous farmers, not those in the distant past. Since we only have access to these measures at one point in time, we assume (in Assumption A2) that the true productivity A_{it}^{fk} is equal to measured productivity (ie $\hat{A}_{i2011}^{fk}$) times some crop-and-market-and-year specific productivity shock (denoted by α_{it}^k).

A2. *In all local markets $i \in \mathcal{I}$ and at all dates t, we assume that*

$$A_{it}^{fk} = \alpha_{it}^k \hat{A}_{i2011}^{fk}, \text{ for all } k \in \mathcal{K}/\{K+1\}, f \in \mathcal{F}_i. \tag{9}$$

The key restriction imposed by Equation (9) on the structure of local productivity shocks is that they do not affect the pattern of comparative advantage across fields. If field f_1 is deemed to be relatively more productive than field f_2 at producing crop k_1 than k_2 in 2011, $\hat{A}_{i2011}^{f_1k_1}/\hat{A}_{i2011}^{f_1k_2} > \hat{A}_{i2011}^{f_2k_1}/\hat{A}_{i2011}^{f_2k_2}$, then we assume that it must have been relatively more productive in all earlier periods, $A_{it}^{f_1k_1}/A_{it}^{f_1k_2} > A_{it}^{f_2k_1}/A_{it}^{f_2k_2}$. In addition, since we do not have any productivity data in the outside sector, we assume (in Assumption A3) that in any given local market, all fields have a common (unknown) productivity in the outside sector.

A3. *In all local markets $i \in \mathcal{I}$ and at all dates t, we assume that*

$$A_{it}^{f,K+1} = \alpha_{it}^{K+1}, \text{ for all } f \in \mathcal{F}_i. \tag{10}$$

Now let $\mathcal{K}_{it}^* \equiv \{k \in \mathcal{K}/\{K+1\} : \hat{Q}_{it}^k > 0\} \cup \{K+1\}$ denote the set of crops with strictly positive output in local market i at date t plus the outside good. In our empirical analysis we will restrict attention to local markets and dates (i,t) such that the following restriction holds.

A4. *For any $N \geq 2$, there does not exist a sequence $\{k_n\}_{n=1,\ldots,N+1} \in \mathcal{K}_{it}^*$ and a sequence*

$\{f_n\}_{n=1,\dots,N} \in \mathcal{F}_i$ such that (i) $k_1 = k_{N+1}$ and $k_n \neq k_{n'}$ for all $n' \neq n$, $n \neq 1$, $n' \neq 1$, (ii) $f_1 \neq f_N$ and $f_n \neq f_{n+1}$ for all n, and (iii) measured productivity in local market i satisfies $\hat{A}_{i2011}^{f_1 k_1} \neq \hat{A}_{i2011}^{f_N k_1}$ and $\prod_{n=1}^{N} \left(\hat{A}_{i2011}^{f_n k_n} / \hat{A}_{i2011}^{f_n k_{n+1}} \right) = 1$, with the convention $\hat{A}_{i2011}^{f K+1} \equiv 1$ for all $f \in \mathcal{F}_i$.

Assumption A4 is a mild technical restriction, which we will be able to test county-by-county and year-by-year. In the case of $N = 2$, it simply states that there do not exist two distinct goods, k_1 and k_2, and two distinct fields, f_1 and f_2, such that $\hat{A}_{i2011}^{f_1 k_1} / \hat{A}_{i2011}^{f_1 k_2} = \hat{A}_{i2011}^{f_2 k_1} / \hat{A}_{i2011}^{f_2 k_2}$. In other words, the pattern of comparative advantage across goods is strict. Intuitively, without such a restriction and its generalization to $N > 2$, we cannot identify crop-and-market-and-year specific productivity shock since changes in observed output conditional on changes in the land allocation may not only reflect true productivity changes, but also the reallocation across crops of fields with the same pattern of comparative advantage, but different absolute advantage.

From now on we refer to $X_{it} \equiv \left[\hat{S}_{it}, \hat{Q}_{it}^k, \hat{L}_{it}^k, \hat{V}_{if}, \hat{A}_{i2011}^{fk} \right]_{k \in \mathcal{K}/\{K+1\}, f \in \mathcal{F}_i}$ as an *observation* for market i at date t and to $\mathcal{X}$ as the set of observations such that Assumptions A1-A4 hold. We denote by $\mathcal{A}_{it}^* \equiv \{\alpha \in \mathbb{R}_+^{K+1} : \alpha^k > 0 \text{ if } k \in \mathcal{K}_{it}^*\}$ and $\mathcal{P}_{it}^* \equiv \{p \in \mathbb{R}_+^{K+1} : p^k > 0 \text{ if } k \in \mathcal{K}_{it}^*\}$ the set of productivity shocks and prices, respectively, that could be consistent with an observation X_{it}. We also denote by $\mathcal{L}_i \equiv \{L \in \mathbb{R}_+^{(K+1) \times F_i} : \sum_{k \in \mathcal{K}} L^{fk} \leq \hat{V}_i^f$ for all $f \in \mathcal{F}_i\}$ the set of feasible allocations of fields to crops. Finally, for any $\alpha_{it} \in \mathcal{A}_{it}^*$, we let $L(\alpha_{it}, X_{it}) \equiv \arg\max_{L \in \mathcal{L}_i} \min_{k \in \mathcal{K}_{it}^*} \{\sum_{f \in \mathcal{F}_i} \alpha_{it}^k \hat{A}_{i2011}^{fk} L^{fk} / \hat{Q}_{it}^k\}$, with the convention $\hat{Q}_{it}^{K+1} \equiv 1$ for all i and t. As we formally establish in the Appendix, $L(\alpha_{it}, X_{it})$ corresponds to the set of efficient allocations that, conditional on a vector of productivity shocks α_{it}, are consistent with relative output levels observed in the data. This set of allocations will play a crucial role in our analysis.

Before stating our main theoretical result, we introduce the following definition.

Definition 2 *Given $X_{it} \in \mathcal{X}$, a vector of productivity shocks and good prices $(\alpha_{it}, p_{it}) \in \mathcal{A}_{it}^* \times \mathcal{P}_{it}^*$ is* admissible *if and only if there exists a field allocation, L_{it}, and a vector of field prices, r_{it}, such that (L_{it}, p_{it}, r_{it}) is a competitive equilibrium consistent with X_{it}.*

Put differently, a vector of unobservable productivity shocks and good prices (α_{it}, p_{it}) is *admissible* if, given these unobserved variables, the observed variables in X_{it} are compatible with perfect competition. The next theorem characterizes the set of admissible vectors of productivity shocks and good prices.

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

$$\begin{aligned} \textstyle\sum_{f\in\mathcal{F}} \alpha_{it}^k \hat{A}_{i2011}^{fk} L_{it}^{fk} &= \hat{Q}_{it}^k \text{ for all } k \in \mathcal{K}_{it}^* / \{K+1\}, \quad (11) \\ \textstyle\sum_{f\in\mathcal{F}_i} L_{it}^{fk} &= \hat{L}_{it}^k \text{ for all } k \in \mathcal{K}_{it}^* / \{K+1\}, \quad (12) \end{aligned}$$

where $L_{it} \in L(\alpha_{it}, X_{it})$*; and (ii) conditional on* $\alpha_{it} \in \mathcal{A}_{it}^*$ *and* $L_{it} \in L(\alpha_{it}, X_{it})$ *satisfying Equations* (11) *and* (12)*,* $(\alpha_{it}, p_{it}) \in \mathcal{A}_{it}^* \times \mathcal{P}_{it}^*$ *is admissible if and only if*

$$\begin{aligned} \textstyle\sum_{k\in\mathcal{K}_i^*/\{K+1\}} p_{it}^k \hat{Q}_{it}^k &= \hat{S}_{it}, \quad (13) \\ \alpha_{it}^{k'} p_{it}^{k'} \hat{A}_{i2011}^{fk'} &\leq \alpha_{it}^k p_{it}^k \hat{A}_{i2011}^{fk} \text{ for all } k,k' \in \mathcal{K},\ f \in \mathcal{F}_i,\ \text{if } L_{it}^{fk} > 0. \quad (14) \end{aligned}$$

The proof of Theorem 1 builds on four classical results in general equilibrium theory: the First Welfare Theorem; the Second Welfare Theorem; the existence of a competitive equilibrium; and the uniqueness of this equilibrium in an endowment economy under the Gross-Substitute Property. The main argument can be sketched as follows.

By the First and Second Welfare Theorems, a land allocation L is part of a competitive equilibrium consistent with relative output levels, $\hat{Q}_{it}^k/\hat{Q}_{it}^{k'}$ for all $k,k' \in \mathcal{K}_{it}^*$, if and only if $L \in L(\alpha_{it}, X_{it}) \equiv \arg\max_{L\in\mathcal{L}_i} \min_{k\in\mathcal{K}_{it}^*} \{\sum_{f\in\mathcal{F}} \alpha_{it}^k \hat{A}_{i2011}^{fk} L^{fk}/\hat{Q}_{it}^k\}$. Thus to find a vector of admissible productivity shocks and good prices, one can start by finding a vector $\left(\alpha_{it}^k\right)_{k\in\mathcal{K}_{it}^*}$, up to a normalization, such that Equations (11) and (12) hold. Mathematically, the problem of finding $\left(\alpha_{it}^k\right)_{k\in\mathcal{K}_{it}^*}$ such that Equation (12) holds is akin to the problem of proving the existence and uniqueness of a vector of competitive prices in an endowment economy, with the observed allocation $\hat{L}_{it}^k$ playing the role of the exogenous endowments. Since a version of the Gross-Substitute Property holds in our environment, such an $\left(\alpha_{it}^k\right)_{k\in\mathcal{K}_{it}^*}$ exists and is unique, up to a normalization. The overall productivity level can then be chosen so that the land allocation L not only matches relative output levels, but also absolute levels. This is the idea behind Equation (11). Once productivity shocks have been identified and the associated equilibrium allocation has been constructed, Condition (14) directly derives from the zero-profit conditions (2) and (3). Finally, given relative prices, the overall price level can be chosen so that total sales in equilibrium are equal to total sales in the data. That is the idea behind Equation (13).

According to Theorem 1, productivity shocks are identified whenever a crop is produced. By contrast, Theorem 1 allows, in principle, for a large number of admissible good prices. For almost all observations $X_{it} \in \mathcal{X}$, however, this is not so. Namely, admissible good prices for crops that are produced must also be unique. To see this, note that for almost all observations, the output vector associated with the equilibrium allocation is not colinear to a vertex of the Production Possibility Frontier (PPF). This is illustrated in Figure 1,

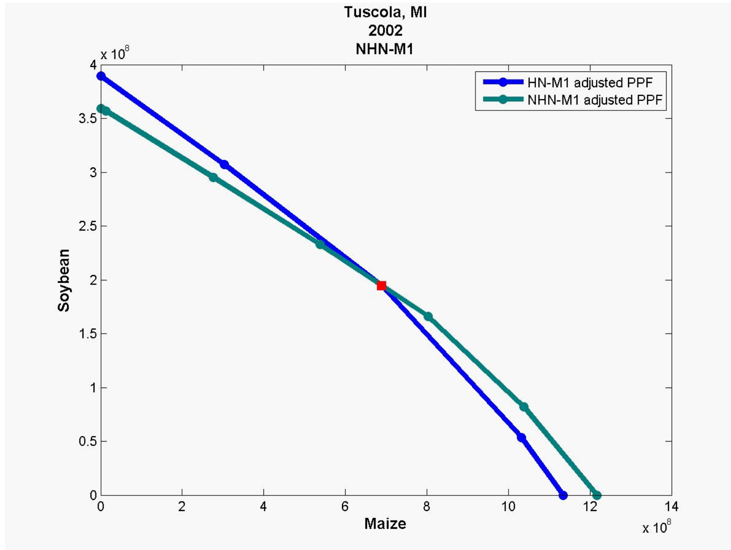

Figure 1: Production possibility frontiers (PPF) for one county and year in our dataset in which only two crops were produced. The blue line ("HN-M1-adjusted") illustrates the PPF after adjustment for selection into the outside good, but before adjustment for productivity shocks; the green line ("NHN-M1 adjusted") illustrates the PPF after additional adjustment for productivity shocks. The red dot (where the two lines cross) is the county's actual production point in this year.

which corresponds to one of the approximately 20,000 county-years in our dataset, Tuscola, MI, in 2002; this county-year is chosen because it is a rare case in which a county-year is producing only two non-zero crops. Whenever we are in such a situation, for any pair of crops $k, k' \in \mathcal{K}^*_{it}/\{K+1\}$, there exists a field $f \in \mathcal{F}_i$ such that $L^{fk}_{it} \times L^{fk'}_{it} > 0$. Condition (14) therefore implies that the relative price of these two crops is uniquely determined by $p^{k'}_{it}/p^k_{it} = \alpha^k_{it}\hat{A}^{fk}_{i2011}/\alpha^{k'}_{it}\hat{A}^{fk'}_{i2011}$. The overall level can then be computed using Equation (13). We can state the following corollary to Theorem 1.

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

$$\sum_{k\in\mathcal{K}^*_i/\{K+1\}} p^k_{it}\hat{Q}^k_{it} = \hat{S}_{it}, \tag{15}$$

$$\frac{p^{k'}_{it}}{p^k_{it}} = \frac{\alpha^k_{it}\hat{A}^{fk}_{i2011}}{\alpha^{k'}_{it}\hat{A}^{fk'}_{i2011}}, \text{ for any } f \in \mathcal{F}_i \text{ such that } L^{fk}_{it} \times L^{fk'}_{it} > 0, \tag{16}$$

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

Theorem 1 and Corollary 1 have two attractive features. First, they imply that for any "non-zero" crop k, i.e. any crop with strictly positive output in county i at date t, the productivity shock, α^k_{it}, and the local price, p^k_{it}, are almost always identified—and in our dataset, they always will be. Second, they imply that conditional on a vector of productivity shocks,

α_{it}, the problem of solving for prices, p_{it}, is a linear program.[2] Since our dataset includes approximately 1,500 counties over 12 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 16 crops and 26 fields—it is therefore possible to characterize the set of unknowns (i.e. α_i and p_i) in each county in a short period of time using standard software packages.

"Zero" crops. Theorem 1 and Corollary 1 only provide information about crops that are produced in a local market at a given date. This is intuitive. For $k \notin \mathcal{K}_{it}^*$, we know that output is zero, but since the amount of resources allocated to these crops is also zero, we do not know whether this outcome reflects low prices or low productivity levels. For our counterfactual exercises, however, we will need to take a stand on what productivity shocks and prices were for "zero" crop, i.e. crops that were not produced. To fill this gap between theory and data, we first assume that whenever a crop is not produced in a given county-year, the productivity shock is equal to the national average of observed productivity shocks for that crop (ie shocks for counties where that crop is produced) in that year.

A5. *In all local markets $i \in \mathcal{I}$ and at all dates t, we assume that*

$$\alpha_{it}^k = \frac{1}{N_t^k} \sum\nolimits_{j \in \mathcal{I}: k \in \mathcal{K}_{jt}^*/\{K+1\}} \alpha_{jt}^k, \text{ for all } k \notin \mathcal{K}_{it}^*, \tag{17}$$

where N_t^k is the number of markets with positive output of crop k in period t.

Second we use the fact that conditional on α_{it}^k, Inequality (14) in Theorem 1 provides an upper-bound on the relative price of crops with zero ouput. Formally, Theorem 1 implies that if $k \notin \mathcal{K}_{it}^*$, then p_{it}^k must be bounded from above by $\tilde{p}_{it}^k/\alpha_{it}^k$, where $\tilde{p}_{it}^k \equiv \min_{k' \in \mathcal{K}_{it}^*/\{K+1\}} \left\{ p_{it}^{k'} \alpha_{it}^{k'} \min_{f: L^{fk'}>0} \left\{ \hat{A}_{i2000}^{fk'} / \hat{A}_{i2000}^{fk} \right\} \right\}$. In our baseline counterfactual exercises, we simply assume that this upper-bound is binding.

A6. *In all local markets $i \in \mathcal{I}$ and at all dates t, we assume that*

$$p_{it}^k = \tilde{p}_{it}^k / \alpha_{it}^k, \text{ for all } k \notin \mathcal{K}_{it}^*, \tag{18}$$

where α_{it}^k is given by Equation (17).

Section 5 discusses the sensitivity of our results to these two assumptions.

[2] Solving for productivity shocks is only slighlty more complicated since it requires looking for the fixed point of a function that itself is defined as the solution of a linear program.

2.3 Measuring Aggregate Gains from Economic Integration and Productivity Improvements in Agriculture

The goal of Section 2.2 above was to infer the unkown farm-gate prices (p_{it}^k) and productivity shocks (α_{it}^k) that prevailed in each local market i, year t and crop k in our dataset. We have described a procedure by which these unknowns can be identified from aggregate (ie county-wide) data on farmers' choices about what to grow using how much land. We now turn to the second stage of our analysis, in which we aim to measure the gains from a counterfactual rise in economic integration.

In order to measure gains from economic integration, we first need to estimate trade costs and how they vary over time. For any crop $k \in \mathcal{K}/\{K+1\}$, Corollary 1 and Assumption A6 provide measures of p_{it}^k. Thus using our model, we can estimate trade costs using

$$\tau_{it}^k = \frac{\bar{p}_t^k}{p_{it}^k} - 1, \text{ for all } k \in \mathcal{K}/\{K+1\}. \tag{19}$$

Given these measures of trade costs, we then estimate 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 crops produced in period t have been if trade costs were those of period t' rather than period t?" The counterfactual equilibrium that we construct to address this question has two features that are worth emphasizing.

First, and most importantly, we assume that crop producers in market i at date t maximize profits facing the counterfactual prices, $\left(p_{it}^k\right)' = \bar{p}_t^k / \left(1 + \tau_{it'}^k\right)$, rather than the observed prices $p_{it}^k = \bar{p}_t^k / \left(1 + \tau_{it}^k\right)$, where trade costs at both dates are computed using Equation (19). Second, we assume that the overall land allocation to crops is the same in the observed equilibrium and the counterfactual equilibrium. Namely, if $L^{f,K+1}$ acres of a field f are allocated to the outside good in the initial equilibrium, then $L^{f,K+1}$ acres remain allocated to the outside good in the counterfactual equilibrium. This implies that our measure of the gains from economic integration will abstract from any reallocation from the outside good to crops and vice versa. Given our lack of information about the outside good, in general, and the trade costs that it might face at different points in time, in particular, we believe that this is the right approach. The only role of the outside good in our paper is to solve for endogenous sorting of fields into the economic activities for which we have data, i.e. crops.

Let $\left(Q_{it}^k\right)'$ denote the counterfactual output level of crop k in market i at date t. Using this notation, we measure the gains (or losses) from changes in the degree of economic integration between two periods t and $t' > t$ as:

$$\Delta_{tt'}^{\tau} \equiv \frac{\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} \left(p_{it}^{k}\right)' \left(Q_{it}^{k}\right)'}{\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} p_{it}^{k} \hat{Q}_{it}^{k}} - 1. \tag{20}$$

By construction, $\Delta_{tt'}^{\tau}$ measures how much larger (or smaller) the total value of output across crops would have been in period t if trade costs were those of period t' rather than those of period t. It is important to note that in Equation (20), we use local prices to evaluate output both in the original and the counterfactual equilibrium. This is consistent with the view that differences in local crop prices reflect "true" technological considerations: farmers face the "right" prices, but local prices are lower because of transportation costs. One can therefore think of $\Delta_{tt'}^{\tau}$ as a measure of aggregate productivity gains in the transportation sector, broadly defined, between t and t'.[3]

We follow the same approach to estimate the gains (or losses) from productivity changes in agriculture. Namely, for any pair of periods t and $t' > t$, we ask: "How much higher (or lower) would the total value of output across local markets in period t have been if productivity shocks were those of period t' rather than period t?" The answer to this question provides an aggregate measure of productivity changes in agriculture between these two periods. In line with the previous counterfactual exercise, we construct the counterfactual equilibrium under the assumption that the overall land allocation to crops is the same as in the initial equilibrium. Let $\left(Q_{it}^{k}\right)''$ denote the counterfactual output level of crop k in market i at date t if farms in this market were maximizing profits facing the counterfactual productivity shocks $\left(\alpha_{it}^{k}\right)'' = \alpha_{it'}^{k}$ rather than the true productivity shocks α_{it}^{k}. Using this notation, we measure the gains (or losses) from productivity changes in agriculture between two periods t and $t' > t$ as:

$$\Delta_{tt'}^{\alpha} \equiv \frac{\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} p_{it}^{k} \left(Q_{it}^{k}\right)''}{\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} p_{it}^{k} \hat{Q}_{it}^{k}} - 1. \tag{21}$$

In the rest of this paper, we implement the strutural approach described in this section in the context of U.S. agricultural markets from 1880 to 2002.

3 Data

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

[3] We discuss alternative interpretations of price gaps in Section 5.

3.1 Modern Productivity Data

The first and most novel data source that we make use provides measures of productivity (i.e. $\hat{A}_i^c(f)$ 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).[4] 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. In our baseline results we use the scenario that relates to 'mixed inputs, with possible irrigation.' We come back to other scenarios in Section 5.

Finally it is important to emphasize that while the GAEZ project 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 project 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; the average productivity among fields that produce a crop in any given market and time period is endogenous and conditioned on the set of fields who endogenously produce that crop at prevailing prices.

[4] This 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.

3.2 Historical Output, Area and Sales Data

The second set of data on which we draw contains records of actual output by crop, $\hat{Q}_i^c$, area of land harvested under each crop, $\hat{L}_i^c$, and the nominal value of total sales of all crops taken together, $\hat{S}_i$, in each U.S. county from 1880-2002.[5] These measures come from the Census of Agriculture that began in 1840 and has been digitized in Haines (2005).[6] The Census was conducted decadally until 1950 and then roughly once every five years thereafter; however, the 1930 and 1940 data are not available in digital form (for all counties in our sample). 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').

An important consideration in using the data on total crop sales (in order to construct $\hat{S}_i$) is that farmers from 1880-1920 were asked to report the total value of crops produced (which is appropriate for our analysis), whereas from 1950 onwards farmers were asked to report the value of crops actually sold. For this reason in our preliminary use of the data here we simply avoid making comparisons across the 1920-1950 period in which the two proxies for $\hat{S}_i$ differ. 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 Historical Price Data

A final source of data that we use is actual data on observed producer (i.e. farm-gate) prices. While price data is not necessary for our analysis, below we perform some simple tests of our exercise by comparing farm-gate 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 farm-gate 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

[5] We refer to the years in our dataset by the yeasr in which the corresponding Census was published (eg 1880) rather than the year in which farmers were enumerated (1879).

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

[7] This 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).

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

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.

4 Empirical Results

This section presents preliminary estimates of the gains from economic integration within US crop agriculture from 1880-2002. Before presenting these estimates, however, we first discuss (in the next subsection to follow below) the empirical plausibility of our estimates of farm gate prices, which are central to our analysis.

4.1 Do Estimated Farmgate Prices Look Sensible?

The first step of our analysis uses Corollary 1 to estimate the local price for each of our 16 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 from 1880 to 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 farm-gate 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 obtained for zero-output crops 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 find a positive and statistically significant correlation between the two price series, with a coefficient that varies between 0.69 and 1.05 depending on the specification.[9] While most of the coefficient estimates are below one (the result that would obtain if price estimates agreed perfectly 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

[9] We 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 0.58 to 0.89.

Table 1: Correlation between estimated prices and price data

Dep. variable:	estimated price (from model)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
observed producer price (from real world)	0.810***	0.713***	0.680***	0.692***	1.049***	0.842***	0.804***	0.835***
	(0.0186)	(0.0215)	(0.0336)	(0.0395)	(0.00553)	(0.0217)	(0.0258)	(0.0506)
Logs or levels	Levels	Levels	Levels	Levels	Logs	Logs	Logs	Logs
constant	No	Yes	Yes	Yes	No	Yes	Yes	Yes
crop fixed effect	No	No	Yes	Yes	No	No	Yes	Yes
year fixed effect	No	No	No	Yes	No	No	No	Yes
R-squared	0.408	0.285	0.370	0.431	0.929	0.352	0.614	0.714
observations	2,766	2,766	2,766	2,766	2,766	2,766	2,766	2,766

Notes: Robust standard errors, clustered at the state-level, in parentheses. *** indicates statistically significant at 0.1% level. Columns (1) and (5) report uncentered R-squared values.

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.

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; recall from Section 2.3 above that we model the local farm gate price as $p_{it}^c = \bar{p}_t^c / (1 + \tau_{it}^c)$.

To construct reference prices we make two extreme assumptions that will be relaxed in future work. First, we assume that all of the counties in our sample (of Eastern U.S. counties) were trading at least some of their output with one major agricultural wholesale market, that in New York City. This implies that the New York City price can be used as the reference price (since free arbitrage would ensure that, under this assumption, $p_{it}^c = \bar{p}_t^c / (1 + \tau_{it}^c)$ always holds). Second, we assume that trade costs within New York state were small (relative to the costs of trading at longer distances) such that we can obtain a preliminary estimate of

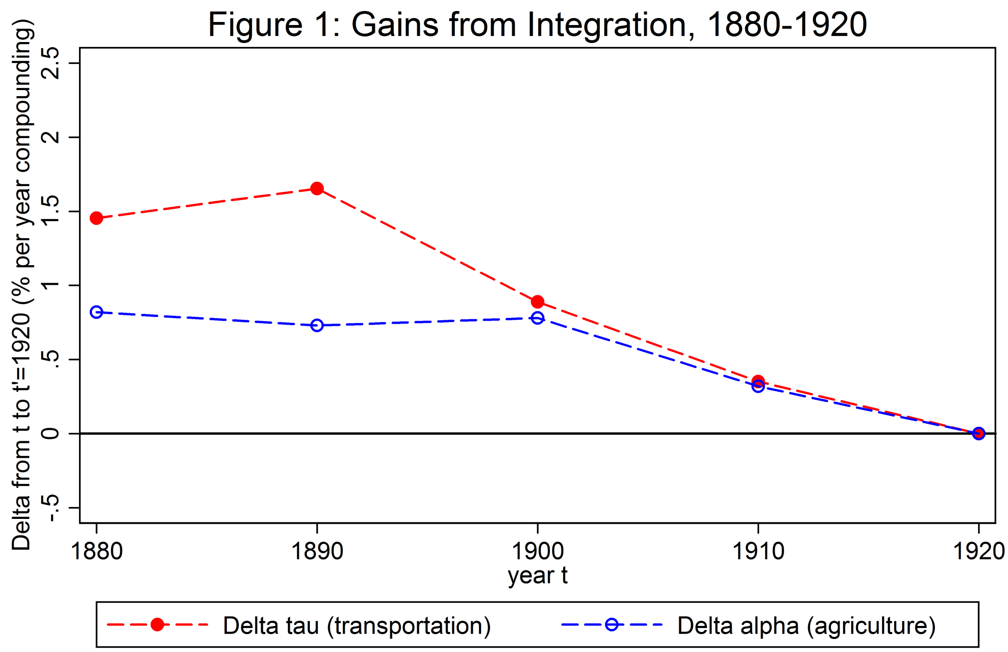

Figure 2: The gains (or losses) from a counterfactual change in economic integration in which trade costs (price gaps) and agricultural productivity levels from year t are replaced with those from year $t' = 1920$. The red line (solid dots) plots $\Delta^{\tau}_{tt'}$(gains when price gaps are interpreted as pure transport costs) and the blue line (open dots) plots $\tilde{\Delta}^{\alpha}_{tt'}$(gains from agricultural technological change holding price gaps constant), but where the reported numbers expressed as an annual compounding growth rate from year t to year t'.

the reference price, the New York City wholesale market price, from the New York state farm-gate price which is available in the Cooley et al (1977) and USDA state-level price dataset.

Armed with data on reference prices, $\bar{p}^c_t$, for each year and crop we follow the procedure in Section 2.3 above to compute the gains from various counterfactual scenarios of economic integration. As described in Section 3.2 the sales data (used to construct $\hat{S}_i$) cannot be used—without auxiliary information that we aim to incorporate in future work—to draw comparisons between years prior to 1920 and years after 1950. For this reason we simply describe two counterfactual scenarios here: that for the gains from 1880-1920 (i.e. t corresponds to years from $1880 - 1910$ and t' corresponds to 1920) and that for the gains from 1950-1997 (i.e. $t = 1950$ to 1997 and $t' = 2002$).

Our results are presented in Figure 1 (for 1880-1920) and Figure 2 (for 1950-2002). Figure 1 plots (in red) the value of $\Delta^{\tau}_{tt'}$ by year t for the case of $t' = 1920$ and Figure 2 does the same for $t' = 2002$. In both cases $\Delta^{\tau}_{tt'}$ is expressed in *compounding annual growth rates from year t to t'* to aid interpretation across different lengths of time period (from t to t').

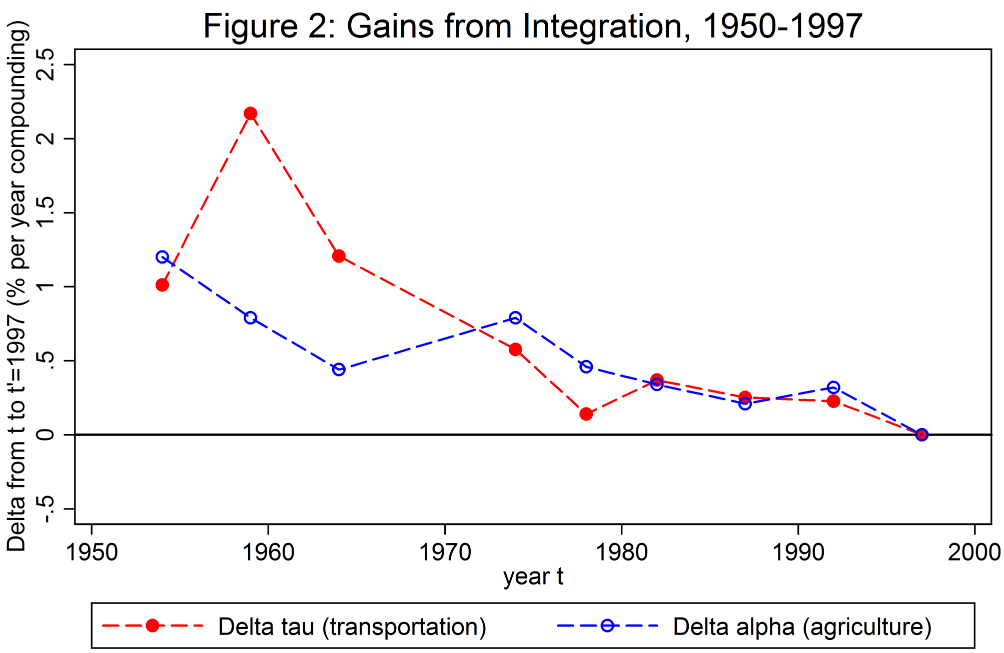

Figure 3: The gains (or losses) from a counterfactual change in economic integration in which trade costs (price gaps) and agricultural productivity levels from year t are replaced with those from year $t' = 2002$. The red line (solid dots) plots $\Delta^{\tau}_{tt'}$(gains when price gaps are interpreted as pure transport costs) and the blue line (open dots) plots $\tilde{\Delta}^{\alpha}_{tt'}$(gains from agricultural technological change holding price gaps constant), but where the reported numbers expressed as an annual compounding growth rate from year t to year t'.

We find that, according to the formula in equation (20), $\Delta_{t=1880,t'=1920} = 1.48$ and $\Delta_{t=1950,t'=2002} = 0.98$. These estimates imply that substantial inter-spatial price differences have existed within the United States, but that these differences have become smaller over time (both before and after the second World War). Further, the gains from economic integration that have accrued as spatial price gaps have fallen are substantial—for example, equal to 2.3% compounding per annum growth from 1880-1920 or 1.3% from 1950-2002.

To put these estimates in context we compare them, in Figures 1 and 2, to the growth of productivity within the agricultural sector; that is, we plot (in blue) the value of $\Delta\alpha_{tt'}$ for each year t. It is straightforward to do so within our framework because we have estimated technology shifters, $\alpha^c_{i,t}$, by county, crop and year. If we consider instead a counterfactual scenario in which year t technology is replaced with year t'technology (holding wedges fixed at their year t levels) we find that the per annum gains, according to Equation (20) are 1.50% from 1880-1920 and 1.53% from 1950-2002. That is, gains from economic integration are on the same order of magnitude as the gains from pure agricultural productivity growth.

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 important robustness checks. Nevertheless these preliminary results strike us as both encouraging and plausible.

5 Robustness

To measure the gains from economic integration, we have proceeded in two steps. First, we have made assumptions on the production functions of various crops and how they vary over time to infer the distribution of price gaps between local markets and wholesale markets over time. Most notably, we have assumed that land was the only factor of production (Equation 1) and that the pattern of comparative advantage across heterogeneous parcels of land was stable over time (Assumption A2). Second, we have interpreted the price gaps inferred from our model as transportation costs between local and wholesale markets (Equation 20). We now discuss the sensitivity of our estimates of the gains from economic integration to these two sets of assumptions.

5.1 Alternative Interpretations of Price Gaps

For now, let us take the price gaps estimated in Section 4 at face value. There are two separate issues associated with the interpretation of these price gaps. The first issue is whether the price gaps represent true transportation costs, in which case the prices used to estimate the value of output should be local prices as in Equation (20), or whether the price gaps represent distortions broadly defined, in which case the prices used to estimate the value of output should be the prices in the wholesale market. The second issue is: assuming that price gaps reflect true transportation costs, which transportation costs do they measure? In the model, local markets are assumed to export output to the wholesale market. Thus price gaps must reflect the cost of shipping cost from the local market to the wholesale market. In practice, however, local markets may be exporting to other local markets or even importing crops from the wholesale market.

Price gaps as distortions. To address the first issue, we consider the following alternative measure of the gains from economic integration:

$$\tilde{\Delta}_{tt'}^{\tau} \equiv \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} \bar{p}_t^k \left(Q_{it}^k\right)'}{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} \bar{p}_t^k \hat{Q}_{it}^k} - 1. \tag{22}$$

Like in Section 5, $\left(Q_{it}^k\right)'$ denotes the counterfactual output level of crop k in market i at date t when crop producers in market i at date t maximize profits facing the counterfactual prices, $\left(p_{it}^k\right)' = \bar{p}_t^k / \left(1+\tau_{it'}^k\right)$, rather than the observed prices $p_{it}^k = \bar{p}_t^k / \left(1+\tau_{it}^k\right)$. The only difference between Equations (20) and (22) is that Equation (22) evaluates output using prices in the wholesale market rather than local prices. As alluded to above, the implicit assumption underlying $\tilde{\Delta}_{tt'}^{\tau}$ is that differences in local crop prices reflect "true" distortions. In order to maximize total agricultural revenue in the United States, local farmers should be maximizing profits taking the reference prices $\bar{p}_t^k$ as given, but because of various local policy reasons, they do not. This alternative measure of the gains from economic integration is close in spirit to the measurement of the impact of misallocations on TFP in Hsieh and Klenow (2009). In Equation (22) τ_{it}^k is interpreted as a "wedge," i.e. a tax, that may vary across crops over space.

Figures 3 and 4 report the gains from economic integration using this alternative measure (ie $\tilde{\Delta}_{tt'}^{\tau}$) alongside the estimates seen earlier in Figures 1 and 2 (ie based on). It is clear that, as expected, an interpretation of price gaps as pure transportation costs (that consume resources in shipping) leads to larger estimated gains than an interpretation of these price gaps as pure policy distortions (that redistribute revenue lump-sum) But even the pure policy distortions interpretation of price gaps suggests that the gains from economic integration have been significant.

Price gaps as directed trade costs. To address the second issue, the ideal approach would consist in combining the price gaps inferred from the model with data on trade flows from and to local markets. Such data, unfortunately, are unavailable. We therefore settle on the following approach: we propose an alternative assumption about the direction of trade flows that also is consistent with our measures of prices and explore how this alternative

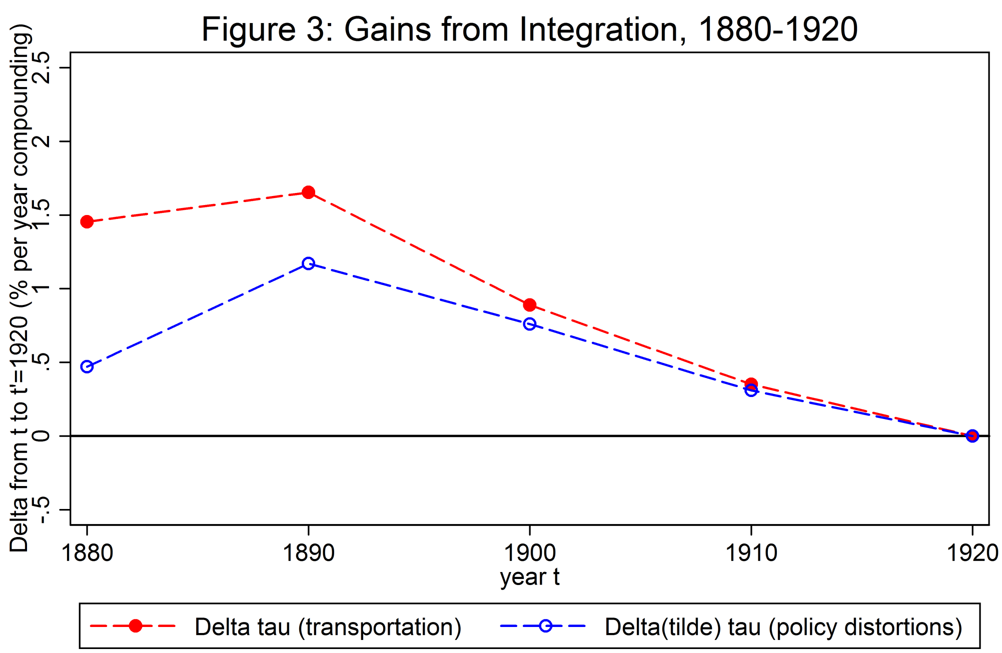

Figure 4: The gains (or losses) from a counterfactual change in economic integration in which trade costs (price gaps) from year t are replaced with those from year $t' = 1920$. The red line (solid dots) plots $\Delta^{\tau}_{tt'}$(gains when price gaps are interpreted as pure transport costs) and the blue line (open dots) plots $\tilde{\Delta}^{\tau}_{tt'}$(gains when price gaps are interpreted as pure policy distortions that are redistributed lump sum), but where the reported numbers expressed as an annual compounding growth rate from year t to year t'.

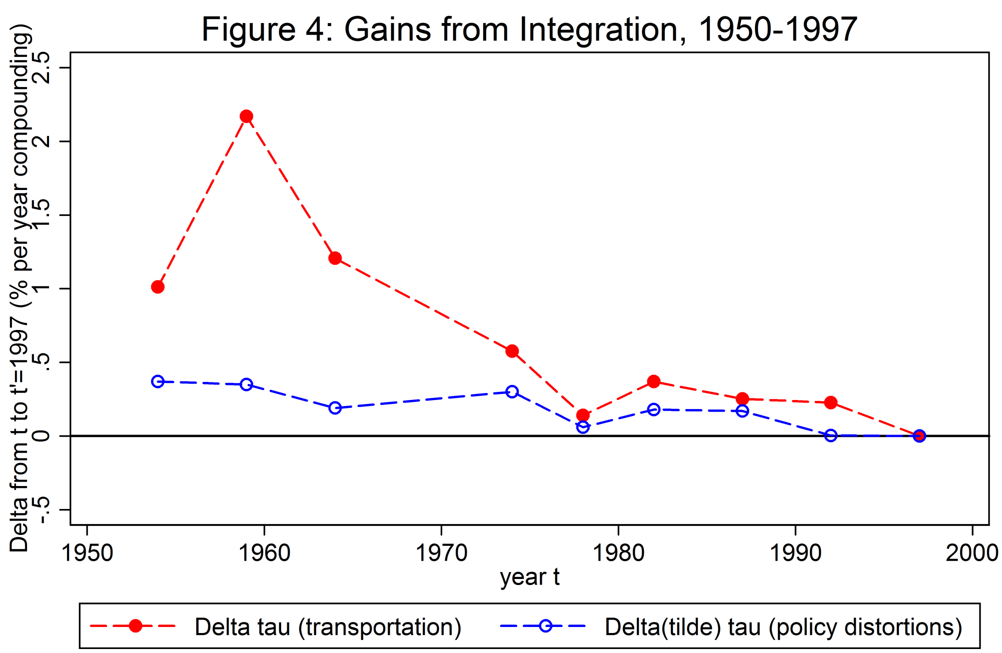

Figure 5: The gains (or losses) from a counterfactual change in economic integration in which trade costs (price gaps) from year t are replaced with those from year $t' = 2002$. The red line (solid dots) plots $\Delta^{\tau}_{tt'}$(gains when price gaps are interpreted as pure transport costs) and the blue line (open dots) plots $\tilde{\Delta}^{\tau}_{tt'}$(gains when price gaps are interpreted as pure policy distortions that are redistributed lump sum), but where the reported numbers expressed as an annual compounding growth rate from year t to year t'.

assumption affects our measures of the gains from economic integration.

In our baseline results, we have assumed that all local markets export to the wholesale market. Under this assumption, local prices should always be lower than prices in the wholesale market. In the data, they are not. Thus our baseline results implicitly rely on the assumption that there is measurement error in the prices of the wholesale market, which we find quite reasonnable. The idea is that if some local prices appear to be higher, it is only because the true wholesale price is lower than what is observed in the data.[10] An alternative approach would be to assume that prices in the wholesale market are measured without error. Under the maintained assumption that local markets are trading with the wholesale market, a higher local price should now be interpreted as a measure of exporting cost from the wholesale market to the local market. Compared to our baseline results, a very high local price should be interpreted as a very high trade cost, not a very low trade cost subject to high measurement error.

Formally, let $\mathcal{I}_t^{k-} \equiv \left\{i \in \mathcal{I} | p_{it}^k < \bar{p}_t^k\right\}$ and $\mathcal{I}_t^{k+} \equiv \left\{i \in \mathcal{I} | p_{it}^k \geq \bar{p}_t^k\right\}$ denote the set of counties with local prices lower and higher, respectively, than the price of crop k in the local market. Compared to Section 2.3, we now propose to measure trade costs as

$$\tau_{it}^k = \frac{\bar{p}_t^k}{p_{it}^k} - 1, \text{ for all } k \in \mathcal{K}/\left\{K+1\right\} \text{ and } i \in \mathcal{I}_t^{k-}, \tag{23}$$

$$\tau_{it}^k = \frac{p_{it}^k}{\bar{p}_t^k} - 1, \text{ for all } k \in \mathcal{K}/\left\{K+1\right\} \text{ and } i \in \mathcal{I}_t^{k+}. \tag{24}$$

Conditional on this new mapping between price gaps and trade costs, we then measure the gains from economic integration as

$$\Delta_{tt'}^{-\tau} \equiv \frac{\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}^{k-}} \left(p_{it}^k\right)' \left(Q_{it}^k\right)'}{\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}^{k-}} p_{it}^k \hat{Q}_{it}^k} - 1,$$

where $\left(Q_{it}^k\right)'$ again denotes the counterfactual output level when crop producers face the counterfactual prices, $\left(p_{it}^k\right)' = \bar{p}_t^k / \left(1 + \tau_{it'}^k\right)$, rather than the observed prices $p_{it}^k = \bar{p}_t^k / \left(1 + \tau_{it}^k\right)$. Note that $\Delta_{tt'}^{-\tau}$ implictly assumes that (i) changes in trade costs do not revert the direction of trade flows and (ii) trade costs are symmetric, i.e. the cost of shipping crops from the local markets to the wholesale market is the same as the cost of shipping crops from the wholesale market to the local markets.

[10] As long as measurement error is constant across crops and over time, measurement error does not affect the measure of the gains from economic integration given by Equation (20).

5.2 Alternative Forms of Technological Change

Over the time period that we consider, 1880-2002, U.S. agriculture has experienced dramatic technological change, from increased irrigation and mechanization to the adoption of fertilizers and hybrid seeds. While our dataset provides an unusually rich picture of differences in productivity across fields over space, it does not contain any information about how the previous technological changes may have affected productivity over time. In order to estimate local prices at various point in time, our empirical strategy therefore relies on one key identifying assumption: the pattern of comparative advantage of fields across crops within counties is stable over time (Assumption A2). The goal of this subsection is to explore the sensitivity of our results to this assumption.

To do so, we take advantage of another attractive feature of the GAEZ data mentioned in Section 3: the predicted productivity of a given parcel of land is available under various scenarios. For our baseline resuts, we have chosen the scenario that relates to 'mixed inputs, with available irrigation' for all years. One might expect that scenarios with lower levels of non-land inputs are a better description of U.S. agriculture in the earlier years of our sample. If so, the question is: Would allowing for that particular form of technological change—i.e., variations in GAEZ scenarios over time—have large effects on our results? The short answer is no.

The main reason behind the roubustness of our results is that Assumption A2 holds reasonably well across GAEZ scenarios in the sense that the relative productivity of fields across crops is fairly stable. Namely, if we regress the predicted productivity of field f for crop k in county i under each available alternative scenario on a crop-county-scenario fixed effect, we find that the R-squared ranges from 0.78-0.82. In other words, changes in the pattern of comparative advantage across GAEZ scenarios only account for approximately 20% of the variation in productivity across crops and fields. Since our empirical strategy is only sensitive to changes in the pattern of comparative advantage, assuming different GAEZ scenarios over time leads to very small differences in our estimates of the gains from economic integration.

5.3 Other Factors of Production

Although the only factors of production in our baseline model are heterogeneous fields, it should be clear that—fortunately—our results do not hinge on the assumption that land is the only factor of production in the world. Instead, our baseline results implicitly rely on the assumption that each crop is produced using land and other factors of production in a similar Leontieff fashion over time. Specifically, the baseline results in Section 4 above rely on the assumption that productivity in the GAEZ data, $\hat{A}^{fk}_{i2011}$, can be interpreted as the productivity of "equipped" land and that the time variation in land "equipment" does not violate Assumption A2.

In this subsection, we generalize our approach to allow for some substitution between factors of production and for factor intensity to vary over time and space. Formally, we assume the production for each crop k in a local market i at date t is now given by

$$Q_{it}^{k} = \sum_{f \in \mathcal{F}} A_{it}^{fk} \left(L_{it}^{fk}\right)^{\beta_{it}} \left(N_{it}^{fk}\right)^{1-\beta_{it}},$$

where $\beta_{it} \in [0,1]$ measures the land intensity of crop k and $N_{it}^{k} \geq 0$ denotes the number of workers producing that cropy. In the same way that land was interpreted as equipped land in our baseline model, labor should now be interpreted as equipped labor.

Let w_{it} denote the wage in county i at date t. The profits of a representative farm producing crop k in a local market i and selling it to the wholesale market are now given by

$$\Pi_{it}^{k} = \left[\bar{p}_{t}^{k} / \left(1+\tau_{it}^{k}\right)\right] \left(\sum_{f \in \mathcal{F}} A_{it}^{fk} \left(L_{it}^{fk}\right)^{\beta_{it}} \left(N_{it}^{fk}\right)^{1-\beta_{it}}\right) - \sum_{f \in \mathcal{F}} \left(r_{it}^{f} L_{it}^{fk} + w_{it} N_{it}^{fk}\right). \quad (25)$$

Compared to Section 2.1, cost minimization by farms now requires

$$N_{it}^{fk} = \frac{1-\beta_{it}}{\beta_{it}} \frac{r_{it}^{f}}{w_{i_t}} L_{it}^{fk}.$$

Subsituting for the optimal input mix in Equation (25), we can rearrange farmer's profit function as

$$\Pi_{it}^{k} = p_{it}^{k} \sum_{f \in \mathcal{F}} \left[A_{it}^{fk} - \frac{\left(r_{it}^{f}\right)^{\beta_{it}} \left(w_{i_t}\right)^{1-\beta_{it}}}{\left(\beta_{it}\right)^{\beta_{it}} \left(1-\beta_{it}\right)^{1-\beta_{it}}}\right] \left(\frac{1-\beta_{it}}{\beta_{it}} \frac{r_{it}^{f}}{w_{i_t}}\right)^{1-\beta_{it}} L_{it}^{fk}.$$

In line with Section 2.1, profit maximization by farms therefore requires

$$p_{it}^{k} A_{it}^{fk} - c_{it}^{f} \leq 0, \text{ for all } c \in \mathcal{C},\ f \in \mathcal{F}, \quad (26)$$

$$p_{it}^{k} A_{it}^{fk} - c_{it}^{f} = 0, \text{ for all } c \in \mathcal{C},\ f \in \mathcal{F} \text{ such that } L_{it}^{fk} > 0. \quad (27)$$

where $c_{it}^{f} = \left(r_{it}^{f}\right)^{\beta_{it}} \left(w_{it}\right)^{1-\beta_{it}} \Big/ \left(\beta_{it}\right)^{\beta_{it}} \left(1-\beta_{it}\right)^{1-\beta_{it}}$. Factor market clearing in market i still requires

$$\sum_{k \in \mathcal{K}} L_{i}^{fk} \leq V_{i}^{f}, \text{ for all } f \in \mathcal{F}_{i}. \quad (28)$$

Finally, we assume that the wage w_{it} is exogenously pinned down by labor market conditions in counties or economic activities outside of our dataset. Definition 1 generalizes to this new environment in a straightforward manner.

In order to implement our structural approach in this more general environment, we use historical data on an additional variable, also available from the U.S. Census of Agriculture: $\hat{\beta}_{it}$, the average labor intensity in county i at date t computed as the ratio of total farm sales to total labor expenditure. In line with Section 2.2, we assume that this variable is not subject to measurement error: $\beta_{it} = \hat{\beta}_{it}$. Given this new information, and without any risk of confusion, we now refer to an observation for market i at date t as $X_{it} \equiv \left[\hat{w}_{it}, \hat{\beta}_{it}, \hat{S}_{it}, \hat{Q}^k_{it}, \hat{L}^k_{it}, \hat{V}_{if}, \hat{A}^{fk}_{i2011}\right]_{k\in\mathcal{K}/\{K+1\}, f\in\mathcal{F}}$. Finally, we let $\mathcal{Z}_i \equiv \{(L,N) \in \mathcal{L}_i \times \mathbb{R}_+^{(K+1)\times F_i}\}$ denote the set of feasible allocations in county i and let

$$Z(\alpha_{it}, X_{it}) \equiv \arg\max_{(L,N)\in\mathcal{Z}_i} \left\{ \min_{k\in\mathcal{K}^*_{it}} \{\textstyle\sum_{f\in\mathcal{F}_i} \alpha^k_{it} \hat{A}^{fk}_{i2011} \left(L^{fk}_{it}\right)^{\beta^k_{it}} \left(N^{fk}_{it}\right)^{1-\beta^k_{it}} / \hat{Q}^k_{it}\} - w_{it} \sum_{k\in\mathcal{K}^*_{it}} \sum_{f\in\mathcal{F}_i} N^{fk}_{it} \right\} \tag{29}$$

denote the the set of efficient allocations that, conditional on a vector of productivity shocks, are consistent with relative output levels observed in the data. Using this notation, our main theorem generalizes as follows.

Theorem 2 *For any $X_{it} \in \mathcal{X}$, the set of admissible vectors of productivity shocks and good prices is non-empty and satisfies the two following properties: (i) if $(\alpha_{it}, p_{it}) \in \mathcal{A}^*_{it} \times \mathcal{P}^*_{it}$ is admissible, then the vector $\left(\alpha^k_{it}\right)_{k\in\mathcal{K}^*_{it}/\{K+1\}}$ is equal to the unique solution of*

$$\textstyle\sum_{f\in\mathcal{F}_i} \alpha^k_{it} \hat{A}^{fk}_{i2011} \left(L^{fk}_{it}\right)^{\beta_{it}} \left(N^{fk}_{it}\right)^{1-\beta_{it}} = \hat{Q}^k_{it} \text{ for all } k \in \mathcal{K}^*_{it}/\{K+1\}, \tag{30}$$

$$\textstyle\sum_{f\in\mathcal{F}} L^{fk}_{it} = \hat{L}^k_{it} \text{ for all } k \in \mathcal{K}^*_{it}/\{K+1\}, \tag{31}$$

where $(L_{it}, N_{it}) \in Z(\alpha_{it}, X_{it})$; and (ii) conditional on $\alpha_{it} \in \mathcal{A}^*_{it}$ and $(L_{it}, N_{it}) \in Z(\alpha_{it}, X_{it})$ satisfying Equations (11) and (12), $(\alpha_{it}, p_{it}) \in \mathcal{A}^*_{it} \times \mathcal{P}^*_{it}$ is admissible if and only if

$$\textstyle\sum_{k\in\mathcal{K}^*_i/\{K+1\}} p^k_{it} \hat{Q}^k_{it} = \hat{S}_{it}, \tag{32}$$

$$\alpha^{k'}_{it} \hat{A}^{fk'}_{i2011} p^{k'}_{it} \leq \alpha^k_{it} \hat{A}^{fk}_{i2011} p^k_{it} \text{ for all } k,k' \in \mathcal{K},\ f \in \mathcal{F}_i,\ \text{if } L^{fk}_{it} > 0. \tag{33}$$

The broad logic behind Theorem 2 is the same as in Section 2.2. The only key difference is that once we have solved for a vector of productivity shocks and an efficient allocation that matches the land allocation observed in the data, we now need to take into account the fact that differences in output levels across crops reflect both differences in total factor productivity, $\alpha^k_{it} \hat{A}^{fk}_{i2011}$, and differences in labor allocation, $\left(N^{fk}_{it}\right)^{1-\beta_{it}}$. This is the idea behind Equation (30). Since there are no differences in factor intensity across crops—β_{it} does not vary with k—all other conditions are unchanged.

Armed with Theorem 2, we can estimate trade costs and the gains from economic integration in the exact same way as we did in Section 2.3. These results are to be computed.

6 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. Our restimates imply that the gains from integration amount US counties from 1880-2002 have been substantial.

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.

Fieler, 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 (2005), "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.

A Proofs

For notational convenience, and without any risk of confusion, we drop all market and time indices, i and t, from the subsequent proofs. For any $\alpha \in \mathcal{A}^*$, we let $z(\alpha, X)$ denote the "excess demand" for goods in $\mathcal{K}$:

$$z(\alpha, X) \equiv \{(z^1, ..., z^{K+1}) : z^k = \textstyle\sum_{f\in\mathcal{F}} L^{fk} - \hat{L}^k, \text{ for some } L \in L(\alpha, X)\},$$

with the convention $\hat{L}^{K+1} \equiv \sum_{f\in\mathcal{F}} \hat{V}^f - \sum_{k\in\mathcal{K}^*/\{K+1\}} \hat{L}^k$. Before establishing Theorem 1, we establish a number of preliminary results.

A.1 $z(\alpha, X)$ is single-valued

The goal of this section is to show that $z(\alpha, X)$ is single-valued. Throughout our proofs, we repeatedly use the fact that, by Assumption A2, if $L \in L(\alpha, X)$, then

$$L \in \arg\max_{\tilde{L}\in\mathcal{L}} \min_{k\in\mathcal{K}^*} \left\{ \textstyle\sum_{f\in\mathcal{F}} A^{fk} \tilde{L}^{fk} / \hat{Q}^k \right\}. \tag{34}$$

We first use a version of the second welfare theorem to show that any allocation in $L(\alpha, X)$ is associated with a competitive equilibrium.

Lemma 1 (Competitive Prices) *For any $X \in \mathcal{X}$ and any $\alpha \in \mathcal{A}^*$, if $L \in L(\alpha, X)$, then there exist $p \in \mathcal{P}^*$ and $r \in \mathbb{R}^F_+$ such that (L, p, r) is a competitive equilibrium.*

Proof. Let $Q \equiv \{Q^k\}_{k\in\mathcal{K}}$ denote the production vector associated with $L \in L(\alpha, X)$. By definition of $L(\alpha, X)$, Q is efficient. To see this, note that if there were $Q' \geq Q$ with $Q' \neq Q$, then we would have $\sum_{f\in\mathcal{F}} A^{fk} L'^{fk} \geq \sum_{f\in\mathcal{F}} A^{fk} L^{fk}$ for all k, with strict inequality for some $k_0 \in \mathcal{K}$. Thus starting from Q', we could reallocate a small amount of at least one field from k_0 to all other goods in $\mathcal{K}^*$. By construction, the new allocation L'' would be such that $\min_{k\in\mathcal{K}^*} \left\{ \sum_{f\in\mathcal{F}} A^{fk} L''^{fk} / \hat{Q}^k \right\} > \min_{k\in\mathcal{K}^*} \left\{ \sum_{f\in\mathcal{F}} A^{fk} L^{fk} / \hat{Q}^k \right\}$, which contradicts $L \in L(\alpha, X)$. Since Q is efficient and production functions are linear—which implies that the production set is convex—Proposition 5.F.2 in Mas-Colell et al. (1995) implies the existence of non-zero price vectors $p \in \mathbb{R}^{K+1}_+$ and $r \in \mathbb{R}^F_+$ such that Conditions (2) and (3) are satisfied. Furthermore, p must be such that $p^k > 0$ for all $k \in \mathcal{K}^*$. To see this note that if there exists $k_0 \in \mathcal{K}^*$ such that $p^{k_0} = 0$, then equation (3) implies $r^f = 0$ for some f; in turn, condition (2) implies $p^k = 0$ for all $k \in \mathcal{K}$; and finally, equation (3) implies $r^f = 0$ for all f, contradicting the fact that (p, r) is non-zero. Finally, since $L \in L(\alpha, X)$ implies $L \in \mathcal{L}$, Equation (4) is satisfied as well, which concludes our proof. ■

The next Lemma establishes joint properties of any pair of elements of $L(\alpha, X)$, which we will later use to establish that $z(\alpha, X)$ is single-valued.

Lemma 2 *For any $X \in \mathcal{X}$, any $\alpha \in \mathcal{A}^*$, and any pair of allocations L, $L' \in L(\alpha, X)$, let $\Delta L \equiv L - L'$. If there exist $f \in \mathcal{F}$ and two goods $k \neq k' \in \mathcal{K}^*$ such that $\Delta L^{fk} \neq 0$ and $\Delta L^{fk'} \neq 0$, then $A^{fk}/A^{fk'} = p^{k'}/p^k = p'^{k'}/p'^k$, where p and p' are competitive prices associated with L and L', respectively.*

Proof. Consider two allocations L, $L' \in L(\alpha, X)$. By Lemma 1, we know that there exist (p, r) and (p', r') such that (L, p, r) and (L', p', r') are competitive equilibria. Let us introduce the following notation. First let $\underline{\mathcal{K}}_0 = \bar{\mathcal{K}}_0 = \mathcal{K}^*$ and for $n \geq 1$, let $\underline{\mathcal{K}}_n$ and $\bar{\mathcal{K}}_n$ be such that

$$\begin{aligned}
\underline{\mathcal{K}}_n &= \arg\min_{k \in \bar{\mathcal{K}}_{n-1}} \left\{p'^k / p^k\right\}, \\
\bar{\mathcal{K}}_n &= \bar{\mathcal{K}}_{n-1} / \underline{\mathcal{K}}_n.
\end{aligned}$$

By construction, there exists $n \geq 1$ such that $\{\underline{\mathcal{K}}_1, ..., \underline{\mathcal{K}}_n\}$ is a partition of $\mathcal{K}^*$. Second for any subset $\tilde{\mathcal{K}} \subset \mathcal{K}^*$, let $L^f\left(\tilde{\mathcal{K}}\right) \equiv \sum_{k \in \tilde{\mathcal{K}}} L^{fk}$ and $L'^f\left(\tilde{\mathcal{K}}\right) \equiv \sum_{k \in \tilde{\mathcal{K}}} L'^{fk}$. Third let $\mathcal{F}_n$ and $\mathcal{F}'_n$ denote the subset of fields such that $L^f\left(\bar{\mathcal{K}}_n\right) > 0$ and $L'^f\left(\bar{\mathcal{K}}_n\right) > 0$, respectively.

We will first show by iteration that for all $n \geq 0$, (i) $L^f\left(\underline{\mathcal{K}}_n\right) = L'^f\left(\underline{\mathcal{K}}_n\right)$ and $L^f\left(\bar{\mathcal{K}}_n\right) = L'^f\left(\bar{\mathcal{K}}_n\right)$ for all $f \in \mathcal{F}$, and (ii) $\mathcal{F}_n = \mathcal{F}'_n$. For $n = 0$, this is trivial since L, $L' \in L(\alpha, X)$ implies $L^f\left(\mathcal{K}^*\right) = L'^f\left(\mathcal{K}^*\right) = \hat{V}^f$ and $\mathcal{F}_n = \mathcal{F}'_n = \mathcal{F}$. Now suppose that this it is true for $n \geq 0$ and let us show that is true for $n + 1$. If $\mathcal{F}_n = \mathcal{F}'_n = \emptyset$, this is trivial again since $L^f\left(\underline{\mathcal{K}}_{n+1}\right) = L'^f\left(\underline{\mathcal{K}}_{n+1}\right) = 0$, $L^f\left(\bar{\mathcal{K}}_{n+1}\right) = L'^f\left(\bar{\mathcal{K}}_{n+1}\right) = 0$, and $\mathcal{F}_{n+1} = \mathcal{F}'_{n+1} = \emptyset$. Thus suppose that $\mathcal{F}_n = \mathcal{F}'_n \neq \emptyset$. We proceed in two steps.

Step 1: *For all $f \in \mathcal{F}$, $L'^f\left(\bar{\mathcal{K}}_{n+1}\right) \geq L^f\left(\bar{\mathcal{K}}_{n+1}\right)$.*

First note that if $f \notin \mathcal{F}_n$, then $L'^f\left(\bar{\mathcal{K}}_{n+1}\right) = L^f\left(\bar{\mathcal{K}}_{n+1}\right) = 0$. Thus the above inequality holds. Now consider $f \in \mathcal{F}_n$. We proceed by contradiction. Suppose that $L'^f\left(\bar{\mathcal{K}}_{n+1}\right) < L^f\left(\bar{\mathcal{K}}_{n+1}\right)$. Since $L^f\left(\bar{\mathcal{K}}_n\right) = L'^f\left(\bar{\mathcal{K}}_n\right)$, this implies $L'^f\left(\underline{\mathcal{K}}_{n+1}\right) > L^f\left(\underline{\mathcal{K}}_{n+1}\right) \geq 0$. Thus there must be $k_1 \in \underline{\mathcal{K}}_{n+1}$ such that $L'^{fk_1} > 0$. By Conditions (2) and (3), this further implies $p'^{k_1} A^{fk_1} \geq p'^k A^{fk}$ for all $k \in \mathcal{K}^*$, which can be rearranged as

$$\frac{p'^{k_1}}{p'^k} \geq \frac{A^{fk}}{A^{fk_1}} \text{ for all } k \in \mathcal{K}^*.$$

Since $p'^{k_1}/p^{k_1} \leq p'^k/p^k$ for all $k \in \bar{\mathcal{K}}_n$, with strict inequality for $k \in \bar{\mathcal{K}}_{n+1}$, this implies

$$\frac{p^{k_1}}{p^k} \geq \frac{A^{fk}}{A^{fk_1}} \text{ for all } k \in \bar{\mathcal{K}}_n, \text{ with strict inequality if } k \in \bar{\mathcal{K}}_{n+1}.$$

Together with Conditions (2) and (3), the previous series of inequalities implies $L^f\left(\bar{\mathcal{K}}_{n+1}\right) = 0$, which contradicts $L^f\left(\bar{\mathcal{K}}_{n+1}\right) > L'^f\left(\bar{\mathcal{K}}_{n+1}\right) \geq 0$.

Step 2: *For all $f \in \mathcal{F}$, $L'^f\left(\bar{\mathcal{K}}_{n+1}\right) = L^f\left(\bar{\mathcal{K}}_{n+1}\right)$, $L^f\left(\underline{\mathcal{K}}_{n+1}\right) = L'^f\left(\underline{\mathcal{K}}_{n+1}\right)$, and $\mathcal{F}_{n+1} = \mathcal{F}'_{n+1}$.*

First note that if $f \notin \mathcal{F}_n$, then $L'^f\left(\bar{\mathcal{K}}_{n+1}\right) = L^f\left(\bar{\mathcal{K}}_{n+1}\right) = 0$ and $L^f\left(\underline{\mathcal{K}}_{n+1}\right) = L'^f\left(\underline{\mathcal{K}}_{n+1}\right) = 0$. Thus the two previous equations hold. Now consider $f \in \mathcal{F}_n$. Suppose that there exists $f \in \mathcal{F}_n$ such that $L'^f\left(\bar{\mathcal{K}}_{n+1}\right) > L^f\left(\bar{\mathcal{K}}_{n+1}\right)$. By Step 1, we know that $L'^{f'}\left(\bar{\mathcal{K}}_{n+1}\right) \geq L^{f'}\left(\bar{\mathcal{K}}_{n+1}\right)$ for all $f' \in \mathcal{F}$. By assumption, we also know that $L^{f'}\left(\bar{\mathcal{K}}_n\right) = L'^{f'}\left(\bar{\mathcal{K}}_n\right)$ for all $f' \in \mathcal{F}$. We must therefore have $L'^{f'}\left(\underline{\mathcal{K}}_{n+1}\right) \leq L^{f'}\left(\underline{\mathcal{K}}_{n+1}\right)$ with strict inequality for some f'. This implies that the ratio of output of at least one good in $\bar{\mathcal{K}}_{n+1}$ to another good in $\underline{\mathcal{K}}_{n+1}$ cannot be the same under L and L', which contradicts L,$L' \in L(\alpha, X)$. Thus, we must have $L'^f\left(\bar{\mathcal{K}}_{n+1}\right) = L^f\left(\bar{\mathcal{K}}_{n+1}\right)$ for all $f \in \mathcal{F}$, $L'^f\left(\bar{\mathcal{K}}_{n+1}\right) = L^f\left(\bar{\mathcal{K}}_{n+1}\right)$, which also implies $L^f\left(\underline{\mathcal{K}}_{n+1}\right) = L'^f\left(\underline{\mathcal{K}}_{n+1}\right)$ and $\mathcal{F}_{n+1} = \mathcal{F}'_{n+1}$.

Now note that to establish Lemma 2 , it suffices to show that if there exist $f_0 \in \mathcal{F}$ and $k_0 \in \mathcal{K}^*$ such that $\Delta L^{f_0 k_0} \neq 0$, then $p^{k_0} A^{f_0 k_0} = r^{f_0}$ and $p'^{k_0} A^{f_0 k_0} = r'^{f_0}$. Note that $\Delta L^{f_0 k_0} \neq 0$ implies that either $L^{f_0 k_0} > 0$ or $L'^{f_0 k_0} > 0$. Without loss of generality, suppose that $L^{f_0 k_0} > 0$. In this case $p^{k_0} A^{f_0 k_0} = r^{f_0} = \max_{k \in \mathcal{K}^*}\left\{p^k A^{f_0 k}\right\}$ follows from the profit maximization condition (3). Thus we need to establish that $p'^{k_0} A^{f_0 k_0} = r'^{f_0} = \max_{k \in \mathcal{K}^*}\left\{p'^k A^{f_0 k}\right\}$. Let n_0 be such that $k_0 \in \underline{\mathcal{K}}_{n_0}$. By the previous result we know that if $L^{f_0 k_0} > 0$, then there must be $k'_0 \in \underline{\mathcal{C}}_{n_0}$ such that $L'^{f_0 k'_0} > 0$, which implies $p'^{k'_0} A^{f_0 k'_0} = \max_{k \in \mathcal{K}^*}\left\{p'^k A^{f_0 k}\right\}$. Let us now show that $p'^{k_0} A^{f_0 k_0} \geq p'^{k'_0} A^{f_0 f'_0}$. We know that

$$p^{k_0} A^{f_0 k_0} = \max_{k \in \mathcal{K}^*}\left\{p^k A^{f_0 k}\right\} = \max_{k \in \underline{\mathcal{K}}_{n_0}}\left\{p^k A^{f_0 k}\right\}.$$

For all $k \in \underline{\mathcal{K}}_{n_0}$, we also know that $p'^k = \theta p^k$, where $\theta \equiv \min_{k \in \bar{\mathcal{K}}_{n_0 - 1}}\left\{p'^k / p^k\right\}$. Multiplying the previous expression by θ, we therefore obtain $p'^{k_0} A^{f_0 k_0} \geq p'^{k'_0} A^{f_0 k'_0}$. Together with $p'^{k'_0} A^{f_0 k'_0} = \max_{k \in \mathcal{K}^*}\left\{p'^k A^{f_0 k}\right\}$, this implies $p'^{k_0} A^{f_0 k_0} = r'^{f_0} = \max_{k \in \mathcal{K}^*}\left\{p'^k A^{f_0 k}\right\}$, which concludes the proof of Lemma 2. ■

We are now ready to establish that $z(\alpha, X)$ is single-valued.

Lemma 3 (Single-Valued) *For any $X \in \mathcal{X}$ and any $\alpha \in \mathcal{A}^*$, $z(\alpha, X)$ is single-valued.*

Proof. We proceed by contradiction. Suppose that there exist $z, z' \in z(\alpha, X)$ such that $z \neq z'$. By definition of $z(\alpha, X)$, there must be $L, L' \in L(\alpha, X)$ such that $L \neq L'$ and $\sum_{f \in \mathcal{F}} L^{fk} \neq \sum_{f \in \mathcal{F}} L'^{fk}$ for some $k \in \mathcal{K}^*$. By Lemma 1, we know that there exist vectors of prices, (p, r) and (p', r'), such that (L, p, r) and (L', p', r') are competitive equilibria. In addition to the property established in Lemma 2, $\Delta L \equiv L - L'$ must satisfy the two following properties.

Property 1: *If there exist $f \in \mathcal{F}$ and $k \in \mathcal{K}^*$ such that $\Delta L^{fk} \neq 0$, then there exists $k' \neq k \in \mathcal{K}^*$ such that $\Delta L^{fk'} \neq 0$.*

Property 1 directly derives from the fact that if $L, L' \in L(\alpha, X)$, then we must have $\sum_{k \in \mathcal{K}^*} L^{fk} = \sum_{k \in \mathcal{K}^*} L'^{fk} = V_f$ for all $f \in \mathcal{F}$.

Property 2: *If there exist $f \in \mathcal{F}$ and $k \in \mathcal{K}^*$ such that $\Delta L^{fk} \neq 0$, then there exists $f' \neq f$ in $\mathcal{F}$ such that $\Delta L^{f'k} \neq 0$.*

Property 2 directly derives from the fact that if $L, L' \in L(\alpha, X)$, then we must have $\sum_{f \in \mathcal{F}} A^{fk} L^{fk} = \sum_{f \in \mathcal{F}} A^{fk} L'^{fk}$ for all $k \in \mathcal{K}^*$.

The rest of the proof of Lemma 3 proceeds as follows. Since $L \neq L'$, there exist $k_1 \in \mathcal{K}^*$ and $f_1 \in \mathcal{F}$ such that $\Delta L^{f_1 k_1} \neq 0$. This further implies the existence of $k_2 \neq k_1$ such that $\Delta L^{f_1 k_2} \neq 0$, by Property 1, and the existence of $f_2 \neq f_1$ such that and $\Delta L^{f_2 k_2} \neq 0$, by Property 2. By iteration, we can therefore construct an infinite sequence of goods $\{k_n\}_{n \geq 1}$ and an infinite sequence of fields $\{f_n\}_{n \geq 1}$ such that $k_{n+1} \neq k_n$ and $f_{n+1} \neq f_n$. Since the number of goods in $\mathcal{K}^*$ is finite, there must be $M < N$ such that $k_n \neq k_{n'}$ for all $n \neq n'$,

$n,n' \in \{M,...,N\}$ but $k_M = k_{N+1}$. By construction, this subsequence is such that $\Delta L^{f_n k_n} \neq 0$ and $\Delta L^{f_n k_{n+1}} \neq 0$ for all $n \in \{M,...,N\}$. Thus Lemma 2 implies

$$\prod_{n=M}^{N} \frac{A^{f_n k_n}}{A^{f_n k_{n+1}}} = \prod_{n=M}^{N} \frac{p^{k_n}}{p^{k_{n+1}}} = 1, \tag{35}$$

where the second equality comes from the fact that $k_M = k_{N+1}$. By Assumptions A2 and A3 (using the convention $\hat{A}^{fK+1} \equiv 1$), we also know that

$$\prod_{n=M}^{N} \frac{\hat{A}^{f_n k_n}}{\hat{A}^{f_n k_{n+1}}} = \prod_{n=M}^{N} \frac{A^{f_n k_n}}{A^{f_n k_{n+1}}} \prod_{n=M}^{N} \frac{\alpha^{k_{n+1}}}{\alpha^{k_n}} = \prod_{n=M}^{N} \frac{A^{f_n k_n}}{A^{f_n k_{n+1}}}. \tag{36}$$

By Equations (35) and (36), we have therefore constructed a sequence of goods and fields $\{k_M,...,k_{N+1}\}$ and $\{f_M,...,f_N\}$ such that (i) $k_M = k_{N+1}$, $k_n \neq k_{n'}$ for all $n \neq n'$, $n,n' \in \{M,...,N\}$ and $\hat{Q}^{k_n} > 0$ whenever $k_n \in \mathcal{K}/\{K+1\}$ and (ii) $f_n \neq f_{n+1}$ for all n, and measured productivity satisfies

$$\prod_{n=M}^{N} \frac{\hat{A}^{f_n k_n}}{\hat{A}^{f_n k_{n+1}}} = 1. \tag{37}$$

There are two possible cases.

Case 1: *There exists $n_0 \in \{M,...,N-1\}$ such that $\hat{A}^{f_{n_0} k_{n_0+1}} \neq \hat{A}^{f_{n_0+1} k_{n_0+1}}$.*

In this case we can rearrange Equation (37) as

$$\begin{aligned} &\prod_{n=M}^{N} \frac{\hat{A}^{f_n k_n}}{\hat{A}^{f_n k_{n+1}}} \\ &\quad = \left(\frac{\hat{A}^{f_{n_0+1} k_{n_0+1}}}{\hat{A}^{f_{n_0+1} k_{n_0+2}}}\right)\left(\frac{\hat{A}^{f_{n_0+2} k_{n_0+2}}}{\hat{A}^{f_{n_0+2} k_{n_0+3}}}\right)...\left(\frac{\hat{A}^{f_N k_N}}{\hat{A}^{f_N k_{N+1}}}\right)\left(\frac{\hat{A}^{f_M k_M}}{\hat{A}^{f_M k_{M+1}}}\right)...\left(\frac{\hat{A}^{f_{n_0} k_{n_0}}}{\hat{A}^{f_{n_0} k_{n_0+1}}}\right) = 1. \end{aligned}$$

If $f_M \neq f_N$, $\{k_{n_0+1}, k_{n_0+2}, ..., k_N, k_M, ..., k_{n_0}\}$ and $\{f_{n_0+1}, f_{n_0+2}..., f_N, f_M, ..., f_{n_0}\}$ violate Assumption A4. If $f_M = f_N$, then $\hat{A}^{f_N k_{N+1}} = \hat{A}^{f_M k_M}$. Thus $\{k_{n_0+1}, k_{n_0+2}, ..., k_N, k_{M+1}, ..., k_{n_0}\}$ and $\{f_{n_0+1}, f_{n_0+2}..., f_N, f_{M+1}, ..., f_{n_0}\}$ violate Assumption A4.

Case 2: $\hat{A}^{f_n k_{n+1}} = \hat{A}^{f_{n+1} k_{n+1}}$ *for all $n \in \{M,...,N-1\}$.*

In this case starting from $L_1 \equiv L'$, we construct a new allocation L_2 as follows. Without loss of generality, assume that $L^{f_M k_M} - L_1^{f_M k_M} < 0$. Thus the same arguments as in the proofs of Property 1 and Property 2 imply that $\{k_M,...,k_{N+1}\}$ and $\{f_M,...,f_N\}$ are such that $L^{f_n k_n} - L_1^{f_n k_n} < 0$ and $L^{f_n k_{n+1}} - L_1^{f_n k_{n+1}} > 0$ for all $n \in \{M,...,N\}$. We set L_2 such that

$$\begin{aligned} L_2^{f_n k_n} &= L_1^{f_n k_n} - \min_{n \in \{M,...,N\}} \left| L^{f_n k_n} - L_1^{f_n k_n} \right|, \text{ for all } n \in \{M,...,N\}, \\ L_2^{f_n k_{n+1}} &= L_1^{f_n k_{n+1}} + \min_{n \in \{M,...,N\}} \left| L^{f_n k_n} - L_1^{f_n k_n} \right|, \text{ for all } n \in \{M,...,N\}, \\ L_2^{fk} &= L_1^{fk}, \text{ otherwise.} \end{aligned}$$

By construction the new allocation $L_2 \in \mathcal{L}$. Furthermore, L_2 satisfies

$$\begin{aligned} \textstyle\sum_{f \in \mathcal{F}} L_2^{fk} &= \textstyle\sum_{f \in \mathcal{F}} L'^{fk}, \text{ for all } k \in \mathcal{K}, &(38)\\ \textstyle\sum_{f \in \mathcal{F}} A^{fk} L_2^{fk} &= \textstyle\sum_{f \in \mathcal{F}} A^{fk} L'^{fk}, \text{ for all } k \in \mathcal{K}, &(39)\end{aligned}$$

where the second equality uses $\hat{A}^{f_n k_{n+1}} = \hat{A}^{f_{n+1} k_{n+1}}$ for all $n \in \{M, ..., N-1\}$. Since $L_2 \in \mathcal{L}$, Equation (39) implies $L_2 \in L(\alpha, X)$.

Starting from L_2, we can therefore follow the same procedure as above to create a new sequence of goods and fields satisfying conditions (i) and (ii) as well as Equation (37). Either the new sequence falls into Case 1, which violates Assumption A4, or it falls into Case 2, in which case we can construct $L_3 \in L(\alpha, X)$ satisfying Equation (38) in the exact same way we have just constructed L_2. We can iterate the following process. Since after each iteration $j \geq 1$, there is one less field-good pair (f, k) such that $L - L_j \neq 0$, there must be a sequence that falls into Case 1 after a finite number of iterations. Otherwise, given the finite number of fields and goods, there would be an allocation L_j satisfying Equation (38) such that $L_j = L$, which contradicts $\sum_{f \in \mathcal{F}} L^{fk} \neq \sum_{f \in \mathcal{F}} L'^{fk}$ for some $k \in \mathcal{K}^*$. Thus Assumption A4 must be violated, which establishes that $z(\alpha, X)$ is single valued. ■

A.2 Existence of Admissible Productivity Shocks

The goal of this section is to show that there exists $\alpha \in \mathcal{A}^*$ such that Equation (12) holds. Since $z(\alpha, X)$ is single-valued for any $X \in \mathcal{X}$ by Lemma 3, we slightly abuse notation from now on and also denote by $z(\alpha, X)$ the unique element of $z(\alpha, X)$. The next Lemma establishes properties of $z(\alpha, X)$ that parallel the properties of excess demand functions in Proposition 17.B.2 in Mas-Colell et al. (1995).

Lemma 4 *For any $X \in \mathcal{X}$, the excess demand function $z(\alpha, X)$, defined for all $\alpha \in \mathcal{A}^*$, satisfies the following properties:*

(i) *$z(\cdot, X)$ is continuous;*

(ii) *$z(\cdot, X)$ is homogeneous of degree zero;*

(iii) *$\sum_{k \in \mathcal{K}} z^k(\alpha, X) = 0$ for all α;*

(iv) *For a sequence $\{\alpha_n\} \in \mathcal{A}^*$, $\alpha_n \to \alpha$, where $\alpha \neq 0$, $\alpha^{k_1} > 0$ for some $k_1 \in \mathcal{K}^*$, and $\alpha^{k_2} = 0$ for some $k_2 \in \mathcal{K}^*$, there exists $N > 0$ such that for any $k \in \mathcal{K}^*$ satisfying $\alpha^k > 0$, $z^k(\alpha_n, X) < 0$ for all $n \geq N$.*

Proof. Condition (i) derives from the definition of $L(\alpha, X)$ and the Maximum Theorem. Conditions (ii) directly derives from the definition of $L(\alpha, X)$. Condition (iii) derives from the fact that $L(\alpha, X)$ must satisfy $\sum_{f \in \mathcal{F}} \sum_{k \in \mathcal{K}} L^{fk}(\alpha, X) = \sum_{f \in \mathcal{F}} \hat{V}^f$. We now turn to Condition (iv).

Without loss of generality, consider a sequence $\{\alpha_n\} \in \mathcal{A}^*$ such that $\alpha_n \to \alpha$, with $\alpha \neq 0$, $\alpha^{k_1^*} > 0$, and $\alpha^{k_2^*} = 0$, for $k_1, k_2 \in \mathcal{K}^*$. Since $\alpha_n \to \alpha$, for any $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, there exists

$N_{\varepsilon_1,\varepsilon_2}$ such that $\alpha_n^{k_1} > \alpha^{k_1} - \varepsilon_1$ and $\alpha_n^{k_2} < \varepsilon_2$ for all $n \geq N_{\varepsilon_1,\varepsilon_2}$. Let $\tilde{A}^{fk} \equiv \hat{A}^{fk}/\hat{Q}^k$. Using this notation, $L(\alpha_n, X)$ is given by

$$L(\alpha_n, X) = \arg\max_{\tilde{L}\in\mathcal{L}} \min_{k\in\mathcal{K}^*} \{\alpha_n^k \sum_{f\in\mathcal{F}} \tilde{A}^{fk}\tilde{L}^{fk}\}.$$

For all n, if $L_n \in L(\alpha_n, X)$, we must have

$$\sum_{f\in\mathcal{F}} \alpha_n^{k_1} \tilde{A}^{fk_1} L_n^{fk_1} = \sum_{f\in\mathcal{F}} \alpha_n^{k_2} \tilde{A}^{fk_2} L_n^{fk_2}.$$

This implies

$$\min_{f\in\mathcal{F}} \left\{ \min_{k\in\mathcal{K}^*} \{\tilde{A}^{fk}\} \right\} \sum_{f\in\mathcal{F}} \alpha_n^{k_1} L_n^{fk_1} \leq \max_{f\in\mathcal{F}} \left\{ \max_{k\in\mathcal{K}^*} \{\tilde{A}^{fk}\} \right\} \sum_{f\in\mathcal{F}} \alpha_n^{k_2} L_n^{fk_2},$$

and in turn, for all $n \geq N_{\varepsilon_1,\varepsilon_2}$,

$$\min_{f\in\mathcal{F}} \left\{ \min_{k\in\mathcal{K}^*} \{\tilde{A}^{fk}\} \right\} (\alpha^{k_1} - \varepsilon_1) \sum_{f\in\mathcal{F}} L_n^{fk_1} < \max_{f\in\mathcal{F}} \left\{ \max_{k\in\mathcal{K}^*} \{\tilde{A}^{fk}\} \right\} \varepsilon_2 \sum_{f\in\mathcal{F}} L_n^{fk_2}. \tag{40}$$

Let $\hat{V} \equiv \sum_{f\in\mathcal{F}} \hat{V}^f$. We must have

$$\sum_{f\in\mathcal{F}} L_n^{fk_1} + \sum_{f\in\mathcal{F}} L_n^{fk_2} \leq \hat{V}. \tag{41}$$

By Inequality 40 and 41, for all $n \geq N_{\varepsilon_1,\varepsilon_2}$, we therefore have

$$\sum_{f\in\mathcal{F}} L_n^{fk_1} < \frac{\max_{f\in\mathcal{F}} \left\{ \max_{k\in\mathcal{K}^*} \{\tilde{A}^{fk}\} \right\} \varepsilon_2}{\min_{f\in\mathcal{F}} \left\{ \min_{k\in\mathcal{K}^*} \{\tilde{A}^{fk}\} \right\} (\alpha^{k_1} - \varepsilon_1) + \max_{f\in\mathcal{F}} \left\{ \max_{k\in\mathcal{K}^*} \{\tilde{A}^{fk}\} \right\} \varepsilon_2} \hat{V}.$$

Now let us set ε_1 and ε_2 such that $\varepsilon_1 = \alpha^{k_1}/2$, and $\varepsilon_2 = \frac{\min_{f\in\mathcal{F}} \left\{ \min_{k\in\mathcal{K}^*} \{\tilde{A}^{fk}\} \right\} \alpha^{k_1} \hat{L}^{k_1}}{2 \max_{f\in\mathcal{F}} \left\{ \max_{k\in\mathcal{K}^*} \{\tilde{A}^{fk}\} \right\} (\hat{V} - \hat{L}^{k_1})}$. The previous inequality simplifies into $\sum_{f\in\mathcal{F}} L_n^{fk_1} < \hat{L}^{k_1}$. At this point, we have established the existence of $N^{k_1} \equiv N_{\varepsilon_1,\varepsilon_2}$ such that $z^1(\alpha_n, X) < 0$ for all $n \geq N^{k_1}$. To conclude, let $\mathcal{K}^+ \equiv \{k \in \mathcal{K}^* : \alpha^k > 0\}$ and let $N \equiv \max_{k\in\mathcal{K}^+}\{N^k\}$. By construction, for all $n \geq N$, we must have $z^k(\alpha_n, X) < 0$ for all $k \in \mathcal{K}^*$ satisfying $\alpha^k > 0$. ■

The next lemma uses the properties of the excess demand function established in Lemma 4 to establish the existence of a unique vector of productivity shocks such that any $L \in L(\alpha, X)$ matches the land allocation observed in the data. The proof is almost identical to the proof of Proposition 17.C.1 in Mas-Colell et al. (1995). The only difference between the present proof and the one given in Mas-Colell et al. (1995) is that our excess demand function does not satisfy two of the five conditions invoked in Proposition 17.C.1, namely Conditions (iii) and (v). Nevertheless, as we demonstrate below, Conditions (iii) and (iv) in Lemma 4 are sufficient for the same argument to go through.

Lemma 5 (Existence) *For any $X \in \mathcal{X}$, there exists $\alpha \in \mathcal{A}^*$ such that*

$$\textstyle\sum_{f\in\mathcal{F}} L^{fk} = \hat{L}^k \text{ for all } k \in \mathcal{K},$$

where $L \in L(\alpha, X)$.

Proof. Throughout this proof we use the following notation. We let $\{k_1^*, ..., k_{K^*}^*\}$ denote the set of goods with positive output; for any $\alpha \in \mathbb{R}_+^{K+1}$, we let $\underline{\alpha} \equiv \left(\alpha^{k_1^*}, ..., \alpha^{k_{K^*}^*}\right)$ denote the vector of productivity shocks associated with these goods; and for any $\alpha \in \mathcal{A}^*$, $\underline{\alpha} \in \mathbb{R}_{++}^{K^*}$, and $j = 1, ..., K^*$, we let

$$\underline{z}^j(\underline{\alpha}, X) \equiv \textstyle\sum_{f\in\mathcal{F}} L^{fk_j^*} - \hat{L}^{k_j^*}, \text{ for some } L \in L(\alpha, X),$$

denote the excess demand for the j-th good in $\mathcal{K}^*$ as a function of $\underline{\alpha}$ only. The properties of $z(\alpha, X)$ easily transfer to $\underline{z}(\underline{\alpha}, X) \equiv [\underline{z}^j(\underline{\alpha}, X)]_{j=1,...,K^*}$. By Conditions (i) and (ii) in Lemma 4, $\underline{z}(\cdot, X)$ is continuous and homogeneous of degree zero. Furthermore, since $z^k(\alpha, X) = 0$ for all $k \notin \mathcal{K}^*$, Condition (iii) in Lemma 4 implies $\sum_{j=1}^{K^*} \underline{z}^j(\underline{\alpha}, X) = 0$. Finally, Condition (iv) in Lemma 4 implies that for a sequence $\{\underline{\alpha}_n\} \in \mathbb{R}_{++}^{K^*}$, $\underline{\alpha}_n \to \underline{\alpha}$, where $\underline{\alpha} \neq 0$, $\alpha^{k_{j_1}^*} > 0$ for some $j_1 \in \{1, ..., K^*\}$, and $\alpha^{k_{j_2}^*} = 0$ for some $j_2 \in \{1, ..., K^*\}$, there exists $N > 0$ such that for any $j \in \{1, ..., K^*\}$ satisfying $\alpha^{k_j^*} > 0$, $\underline{z}^j(\underline{\alpha}_n, X) < 0$ for all $n \geq N$.

We first show that for any $X \in \mathcal{X}$, there exists $\underline{\alpha}^* \in \mathbb{R}_{++}^{K^*}$ such that

$$\underline{z}(\underline{\alpha}^*, X) = 0. \tag{42}$$

Following Mas-Colell et al. (1995), we define

$$\Delta \equiv \left\{\underline{\alpha} \in \mathbb{R}_+^{K^*} : \textstyle\sum_{j=1}^{K^*} \underline{\alpha}^j = 1\right\}.$$

Since $\underline{z}(\cdot, X)$ is homogeneous of degree zero, we can restrict our search for a solution of Equation (42) to $\underline{\alpha}$ in Δ. Note that $\underline{z}(\cdot, X)$ is well defined only for $\underline{\alpha}$ in the set

$$Interior\ \Delta \equiv \{\underline{\alpha} \in \Delta : \underline{\alpha}^j > 0 \text{ for all } j = 1, ..., K^*\}.$$

To establish existence of a solution of equation (42) in Interior Δ, the strategy in Mas-Colell et al. (1995) consists in constructing an upper hemicontinuous correspondence f from Δ to Δ such that solutions of (42) correspond to fixed points of f and applying Kakutani's fixed-point theorem. We reproduce the key steps here and highlight the two minor points at which the present proof differs from the original one.

Step 1: *Construction of the fixed-point correspondence for $\underline{\alpha} \in Interior\ \Delta$.*

For $\underline{\alpha} \in Interior\ \Delta$, we define

$$f(\underline{\alpha}) \equiv \{\gamma \in \Delta : \underline{z}(\underline{\alpha}, X) \cdot \gamma \geq \underline{z}(\underline{\alpha}, X) \cdot \gamma' \text{ for all } \gamma' \in \Delta\}.$$

Since for any $\gamma \in \Delta$, the sum of its coordinates is equal to 1, we must have

$$f(\underline{\alpha}) = \{\gamma \in \Delta : \gamma^j = 0 \text{ if } \underline{z}^j(\underline{\alpha}, X) < \max_{j'=1,...,K^*}\{\underline{z}^{j'}(\underline{\alpha}, X)\}\} \tag{43}$$

At this point, Mas-Colell et al. (1995) use the fact that by Walras' Law, if $\underline{z}(\underline{\alpha}, X) \neq 0$, then there must be $\underline{z}^j(\underline{\alpha}, X) > 0$ and $\underline{z}^{j'}(\underline{\alpha}, X) < 0$ for some $j \neq j'$. Here, we do not have $\underline{\alpha} \cdot \underline{z}(\underline{\alpha}, X) = 0$. However, $\sum_{j=1}^{K^*} \underline{z}^j(\underline{\alpha}, X) = 0$ also implies that if $\underline{z}(\underline{\alpha}, X) \neq 0$, then there must be $\underline{z}^j(\underline{\alpha}, X) > 0$ and $\underline{z}^{j'}(\underline{\alpha}, X) < 0$ for some $j \neq j'$. Thus, for such an $\underline{\alpha}$, any $\gamma \in f(\underline{\alpha})$ has $\gamma^{j'} = 0$, and then $f(\underline{\alpha}) \subset$ *Boundary* $\Delta = \Delta/$*Interior* Δ.

Step 2: *Construction of the fixed-point correspondence for $\underline{\alpha} \in$ Boundary Δ.*

This step is exactly the same as in Mas-Colell et al. (1995). For $\underline{\alpha} \in$ *Boundary* Δ, let

$$f(\underline{\alpha}) \equiv \{\gamma \in \Delta : \underline{\alpha} \cdot \gamma = 0\} = \{\gamma \in \Delta : \gamma^j = 0 \text{ if } \underline{\alpha}^j > 0\}. \tag{44}$$

Because $\underline{\alpha}^j = 0$ for some j, we have $f(\underline{\alpha}) \neq \emptyset$. Also note that with this construction, we cannot have $\underline{\alpha} \in f(\underline{\alpha})$, because $\underline{\alpha} \cdot \underline{\alpha} > 0$. In other words, no $\underline{\alpha} \in$ *Boundary* Δ can be a fixed point.

Step 3: *If $\underline{\alpha}^*$ is a fixed point of $f(\cdot)$, then $\underline{\alpha}^* \in \mathbb{R}_{++}^{K^*}$ and $\underline{z}(\underline{\alpha}^*, X) = 0$.*

This step is also exactly the same as in Mas-Colell et al. (1995). Take $\underline{\alpha}^* \in f(\underline{\alpha}^*)$. By Step 2, we must have$\underline{\alpha}^* \in$ *Interior* $\Delta \subset \mathbb{R}_{++}^{K^*}$. By Step 1, $z(\underline{\alpha}^*, X) \neq 0$ implies $\underline{\alpha}^* \in$ *Boundary* Δ. Thus if $\underline{\alpha}^* \in f(\underline{\alpha}^*)$, then $\underline{\alpha}^* \in \mathbb{R}_{++}^{K^*}$ and $\underline{z}(\underline{\alpha}^*, X) = 0$.

Step 4: *The fixed-point correspondence $f(\cdot)$ is convex-valued and upper hemicontinuous.*

Following Mas-Colell et al. (1995), note that, both when $\underline{\alpha} \in$ *Interior* Δ and when $\underline{\alpha} \in$ *Boundary* Δ, $f(\underline{\alpha})$ equals a level set of a linear function defined on the convex set Δ, so $f(\cdot)$ is convex.

To establish upper hemicontinuity, consider sequences $\underline{\alpha}_n \to \underline{\alpha}$, $\gamma_n \to \gamma$ with $\gamma_n \in f(\underline{\alpha}_n)$ for all n. We have to show that $\gamma \in f(\alpha)$. There are two cases: $\underline{\alpha} \in$ *Interior* Δ or $\underline{\alpha} \in$ *Boundary* Δ. Suppose first that $\underline{\alpha} \in$ *Interior* Δ, then $\alpha_n \in \mathbb{R}_{++}^{K^*}$ for n sufficiently large. From $\gamma_n \cdot \underline{z}(\underline{\alpha}_n, X) \geq \gamma' \cdot \underline{z}(\underline{\alpha}_n, X)$ for all $\gamma' \in \Delta$ and the continuity of $\underline{z}(\cdot, X)$, we get $\gamma \cdot \underline{z}(\underline{\alpha}, X) \geq \gamma' \cdot \underline{z}(\underline{\alpha}, X)$ for all γ'. Thus $\gamma \in f(\underline{\alpha})$.

Now suppose that $\underline{\alpha} \in$ *Boundary* Δ. Take any j such that $\underline{\alpha}^j > 0$. We want to show that for n sufficiently large we have $\gamma_n^j = 0$ and therefore $\gamma^j = 0$, which establishes that $\gamma \in f(\underline{\alpha})$. The argument differs slightly from the proof in Mas-Colell et al. (1995) since our excess demand function does not satisfy $\max\left\{z^1(\underline{\alpha}_n), ..., z^{K^*}(\underline{\alpha}_n)\right\} \to \infty$. Because $\underline{\alpha}^j > 0$, there must exist $M > 0$ such that $\underline{\alpha}_n^j > 0$ for all $n \geq M$. By the counterpart of Condition (iv) in Lemma 4 described above, since $\underline{\alpha}^j > 0$ and $\underline{\alpha}^{j'} = 0$ for some $j' \in \{1, ..., K^*\}$, there must also exist $N > 0$ such that $\underline{z}^j(\underline{\alpha}_n, X) < 0$ for all $n \geq N$. Since $\sum_{j'=1}^{K^*} \underline{z}^{j'}(\alpha, X) = 0$, this implies $\underline{z}^j(\underline{\alpha}_n, X) < \max_{j'=1,...,K^*}\{\underline{z}^{j'}(\underline{\alpha}_n, X)\}$. Now consider $n \geq \max(M, N)$. If $\underline{\alpha}_n \in$ *Boundary* Δ, then $\underline{\alpha}_n^j > 0$, $\gamma_n \in f(\underline{\alpha}_n)$, and Equation (44) imply $\gamma_n^j = 0$. If, instead, $\underline{\alpha}_n \in$ *Interior* Δ, then $\underline{z}^j(\underline{\alpha}_n, X) < \max_{j'=1,...,K^*}\{\underline{z}^{j'}(\underline{\alpha}_n, X)\}\}$, $\gamma_n \in f(\underline{\alpha}_n)$, and Equation (43) imply $\gamma_n^j = 0$ as well. Thus $f(\cdot)$ is upper hemicontinuous.

Step 5: *A fixed point exists.*

The final step is exactly the same as in Mas-Colell et al. (1995). By Kakutani's fixed-point theorem, a convex-valued, upper hemicontinuous correspondence from a non-empty, compact, convex set into itself has a fixed point. Since Δ is non-empty, convex, and compact, and since $f(\cdot)$ is a convex-valued upper hemicontinuous correspondence from Δ to Δ, we conclude that there exists $\underline{\alpha}^* \in \Delta$ with $\underline{\alpha}^* \in f(\underline{\alpha}^*)$. By Step 3, we have $\underline{\alpha}^* \in \mathbb{R}_{++}^{K^*}$ and $\underline{z}(\underline{\alpha}^*, X) \equiv 0$.

To conclude the proof of Lemma 5, take any $\alpha \in \mathcal{A}^*$ such that $\left(\alpha^{k_1^*}, ..., \alpha^{k_{K^*}^*}\right) = \underline{\alpha}^*$. By construction of $\underline{\alpha}^*$, we have

$$\textstyle\sum_{f \in \mathcal{F}} L^{fk} = \hat{L}^k \text{ for all } k \in \mathcal{K}^*,$$

where $L \in L(\alpha, X)$. Given our definition of $L(\alpha, X)$, if $k \notin \mathcal{K}^*$, then $\sum_{f \in \mathcal{F}} L^{fk} = 0$. Thus we trivially have

$$\textstyle\sum_{f \in \mathcal{F}} L^{fk} = \hat{L}^k \text{ for all } k \notin \mathcal{K}^*,$$

which completes the proof of Lemma 5. ■

A.3 Uniqueness of Admissible Productivity Shocks

The next lemma shows that the excess demand for goods in $\mathcal{K}^*$ has the gross substitute property.

Lemma 6 (Gross Substitute) *If $\alpha, \alpha' \in \mathcal{A}^*$ are such that $\alpha'^{k_0} > \alpha^{k_0}$ for some $k_0 \in \mathcal{K}^*$ and $\alpha'^k = \alpha^k$ for all $k \neq k_0$ in $\mathcal{K}^*$, then $z^k(\alpha', X) > z^k(\alpha, X)$ for all $k \neq k_0$ in $\mathcal{K}^*$.*

Proof. Consider α and α' such that $\alpha'^{k_0} > \alpha^{k_0}$ for some $k_0 \in \mathcal{K}^*$ and $\alpha'^k = \alpha^k$ for all $k \neq k_0$ in $\mathcal{K}^*$. Take $L \in L(\alpha, X)$ and $L' \in L(\alpha', X)$. Throughout this proof we let $\tilde{A}^{fk} \equiv \hat{A}^{fk}/\hat{Q}^k$. Using this notation, we therefore have

$$\begin{aligned} L &= \arg\max_{\tilde{L} \in \mathcal{L}} \min_{k \in \mathcal{K}^*} \{\alpha^k \sum_{f \in \mathcal{F}} \tilde{A}^{fk} \tilde{L}^{fk}\}, \\ L' &= \arg\max_{\tilde{L} \in \mathcal{L}} \min_{k \in \mathcal{K}^*} \{\alpha'^k \sum_{f \in \mathcal{F}} \tilde{A}^{fk} \tilde{L}^{fk}\}, \end{aligned}$$

where $\mathcal{L} \equiv \left\{L \in (\mathbb{R}^+)^{(K+1) \times F} : \sum_{k \in \mathcal{K}} L^{fk} \leq \hat{V}^f \text{ for all } f \in \mathcal{F}\right\}$. We need to show that $\sum_{f \in \mathcal{F}} L'^{fk} > \sum_{f \in \mathcal{F}} L^{fk}$ for all $k \neq k_0$ in $\mathcal{K}^*$. We follow a guess and verify strategy.

Consider the allocation L'' such that

$$\begin{aligned} L''^{f(k)k} &= L^{f(k)k} + \varepsilon^k, \text{ for all } k \in \mathcal{K}^*, \, k \neq k_0, \\ L''^{fk_0} &= L^{fk_0} - \textstyle\sum_{k: f(k) = f} \varepsilon^k, \text{ for all } f \in \mathcal{F}, \\ L''^{fk} &= L^{fk}, \text{ otherwise.} \end{aligned}$$

where $\{f(k)\}_{k\neq k_0}$ and $\{\varepsilon_k\}_{k\neq k_0}$ are chosen such that

$$f(k) = \arg\max_{f\in\mathcal{F}}\left\{\tilde{A}^{fk}/\tilde{A}^{fk_0} : L^{fk_0} > 0\right\}, \tag{45}$$

$$\varepsilon^k = \frac{\alpha'^{k_0}-\alpha^{k_0}}{\alpha^k\tilde{A}^{f(k)k}}\frac{\sum_{f\in\mathcal{F}}\tilde{A}^{fk_0}L^{fk_0}}{1+\alpha'^{k_0}\sum_{k'\neq k_0}\left(\frac{\tilde{A}^{f(k')k_0}}{\alpha^{k'}\tilde{A}^{f(k')k'}}\right)}. \tag{46}$$

Since $k_0 \in \mathcal{K}^*$, Assumption A4 implies that $f(k)$ exists and is unique for all $k \in \mathcal{K}^*$, $k \neq k_0$. In addition, since $\alpha'^{k_0} > \alpha^{k_0}$, we have $\varepsilon^k > 0$ for all $k \in \mathcal{K}^*$, $k \neq k_0$, which implies $\sum_{f\in\mathcal{F}} L''^{fk} > \sum_{f\in\mathcal{F}} L^{fk}$ for all $k \in \mathcal{K}^*$, $k \neq k_0$. First note that if $\alpha'^{k_0} - \alpha^{k_0}$ is small enough, then $L''^{fk_0} \geq 0$ for all f since, by construction, $\sum_{k:f(k)=f}\varepsilon^k > 0$ only if $L^{fk_0} > 0$. We now restrict ourselves to such a situation. The rest of the argument proceeds by contradiction. Suppose that for any $\delta > 0$, there exists $\alpha'^{k_0} - \alpha^{k_0} \in (0,\delta)$ such that $L'' \neq L'$. Since $L' \in L(\alpha', X)$, the same argument as in the proof of Properties 1 and 2 in Lemma 3 imply

$$\sum_{k\in\mathcal{K}^*} L'^{fk} = \hat{V}^f,\text{ for all } f \in \mathcal{F}, \tag{47}$$

$$\sum_{f\in\mathcal{F}}\alpha'^k\tilde{A}^{fk}L'^{fk} = \sum_{f\in\mathcal{F}}\alpha'^{k'}\tilde{A}^{fk'}L'^{fk'},\text{ for all } k,k' \in \mathcal{K}^*. \tag{48}$$

Since $L \in L(\alpha, X)$, the same conditions hold for L with productivity shocks given by $\{\alpha^k\}$. By construction of $\{\varepsilon_k\}_{k\neq k_0}$, this further implies

$$\sum_{k\in\mathcal{K}^*} L''^{fk} = \hat{V}^f,\text{ for all } f \in \mathcal{F}, \tag{49}$$

$$\sum_{f\in\mathcal{F}}\alpha'^k\tilde{A}^{fk}L''^{fk} = \sum_{f\in\mathcal{F}}\alpha'^{k'}\tilde{A}^{fk'}L''^{fk'},\text{ for all } k,k' \in \mathcal{K}^*. \tag{50}$$

Let $\Delta L = L' - L'' \neq 0$. By Equations (47)-(50), we therefore have

$$\sum_{k\in\mathcal{K}^*}\Delta L^{fk} = 0,\text{ for all } f \in \mathcal{F} \tag{51}$$

$$\sum_{f\in\mathcal{F}}\alpha'^k\tilde{A}^{fk}\Delta L^{fk} = \sum_{f\in\mathcal{F}}\alpha'^{k'}\tilde{A}^{fk'}\Delta L^{fk'},\text{ for all } k,k' \in \mathcal{K}^*. \tag{52}$$

Since $L' \in L(\alpha', X) \neq L''$, we know that

$$\min_{k\in\mathcal{K}^*}\{\sum_{f\in\mathcal{F}}\alpha'^k\tilde{A}^{fk}L'^{fk}\} > \min_{k\in\mathcal{K}^*}\{\sum_{f\in\mathcal{F}}\alpha'^k\tilde{A}^{fk}L''^{fk}\}.$$

By Equations (48), (50), and (52), we therefore have $\min_{k\in\mathcal{K}^*}\{\sum_{f\in\mathcal{F}}\alpha'^k\tilde{A}^{fk}\Delta L^{fk}\} > 0$, which implies $\min_{k\in\mathcal{K}^*}\{\sum_{f\in\mathcal{F}}\alpha^k\tilde{A}^{fk}\Delta L^{fk}\} > 0$, and in turn,

$$\min_{k\in\mathcal{K}^*}\{\sum_{f\in\mathcal{F}}\alpha^k\tilde{A}^{fk}(L^{fk}+\Delta L^{fk})\} > \min_{k\in\mathcal{K}^*}\{\sum_{f\in\mathcal{F}}\alpha^k\tilde{A}^{fk}L^{fk}\}, \tag{53}$$

where we have used the fact that $\sum_{f\in\mathcal{F}}\alpha^k\tilde{A}^{fk}L^{fk} = \sum_{f\in\mathcal{F}}\alpha^{k'}\tilde{A}^{fk'}L^{fk'}$. Since $L \in L(\alpha, X)$, Inequality (53) implies $(L + \Delta L) \notin \mathcal{L}$. Thus for any $\delta > 0$, there exists $\alpha'^{k_0} - \alpha^{k_0} \in (0,\delta)$ such that (*i*) $L'' + \Delta L = L' \in \mathcal{L}$; (*ii*) $\sum_{k\in\mathcal{K}^*}\Delta L^{fk} = 0$ for all $f \in \mathcal{F}$; and (*iii*) $(L+\Delta L) \notin \mathcal{L}$.

As $\alpha'^{k_0} - \alpha^{k_0}$ converges to zero, L' and L'' must both converge to L, and in turn, ΔL must converge to zero. Thus conditions (i)-(iii) require the existence of (f_1, k_1) such that $L''^{f_1 k_1} > 0$ and $L^{f_1 k_1} = 0$. By construction of L'', this further requires $f_1 = f(k_1)$. Since $k_1 \in \mathcal{K}^*$, we know that there exists $f \neq f_1 = f(k_1)$ such that $L^{f k_1} > 0$. By equation (45), $f_1 = f(k_1)$ and $f \neq f_1$ must satisfy

$$\frac{\tilde{A}^{f_1 k_1}}{\tilde{A}^{f_1 k_0}} > \frac{\tilde{A}^{f k_1}}{\tilde{A}^{f k_0}}. \tag{54}$$

Now starting from L, consider the following reallocation. Take $\varepsilon \in \left(0, L^{f_1 k_0}\right)$ acres of field f_1 from good k_0 and reallocate them to good k_1. Then take $\eta \in \left(\varepsilon \tilde{A}^{f_1 k_0}/\tilde{A}^{f k_0}, \varepsilon \tilde{A}^{f_1 k_1}/\tilde{A}^{f k_1}\right)$ acres of field f from good k_1 and reallocate them to good k_0. Finally, take $\varepsilon \tilde{A}^{f_1 k_1}/\tilde{A}^{f k_1} - \eta$ acres of field f from good k_1 and reallocate them uniformly to all other goods in $\mathcal{K}^*$. Since $L^{f_1 k_0} > 0$, $L^{f k_1} > 0$ and Inequality (54) holds, such a reallocation is feasible. Furthermore, the change in the output of good k_0 is equal to $-\tilde{A}^{f_1 k_0}\varepsilon + \eta \tilde{A}^{f k_0} > 0$; the change in the output of good k_1 is equal to $\tilde{A}^{f_1 k_1}\varepsilon - \eta \tilde{A}^{f k_1} > 0$; and the change in all other goods is strictly positive. This contradicts $L \in L(\alpha, X)$.

At this point we have shown that there exists $\delta > 0$ such that if $\alpha'^{k_0} - \alpha^{k_0} \in (0, \delta)$, then $L'' = L'$. By construction, we have $\sum_{f \in \mathcal{F}} L''^{fk} > \sum_{f \in \mathcal{F}} L^{fk}$ for all $k \in \mathcal{K}^*$, $k \neq k_0$. Thus we have established that Lemma 6 holds for small changes, $\alpha'^{k_0} - \alpha^{k_0} \in (0, \delta)$. Since this is true for any initial value of α, Lemma 6 must hold for large changes as well. ∎

Using the fact that the gross substitute property holds, one can establish the uniqueness, up to a normalization, of the vector of productivity shocks for goods in $\mathcal{K}^*$ such that $L(\alpha, X)$ matches the observed land allocation.

Lemma 7 (Uniqueness) *For any $X \in \mathcal{X}$, there exists at most one vector $\left(\alpha^k\right)_{k \in \mathcal{K}^*} \in \mathbb{R}_{++}^{K^*}$, up to a normalization, such that*

$$\textstyle\sum_{f \in \mathcal{F}} L^{fk} = \hat{L}^k \text{ for all } k \in \mathcal{K}^* / \{K+1\},$$

where $L \in L(\alpha, X)$.

Proof. In order to establish Lemma 7, it is sufficient to show that if $z^k(\alpha, X) = z^k(\alpha', X) = 0$ for all $k \in \mathcal{K}$, then there must be $\mu > 0$ such that $\alpha^k = \mu \alpha'^k$ for all $k \in \mathcal{K}^*$. The argument is the same as in Proposition 17.F.3 in Mas-Colell et al. (1995). One can proceed by contradiction. Suppose that there exist α and α' such that $z^k(\alpha, X) = z^k(\alpha', X) = 0$ for all $k \in \mathcal{K}$, but there does not exist $\lambda > 0$ such that $\alpha^k = \lambda \alpha'^k$ for all $k \in \mathcal{K}^*$. Since $z(\cdot, X)$ is homogeneous of degree zero, we must therefore α'' such that (i) $z^k(\alpha'', X) = 0$ for all $k \in \mathcal{K}$ and (ii) $\left(\alpha''^k\right)_{k \in \mathcal{K}^*} \neq \left(\alpha^k\right)_{k \in \mathcal{K}^*}$, $\alpha''^k \geq \alpha^k$ for all $k \in \mathcal{K}^*$, and $\alpha''^{k_0} = \alpha^{k_0}$ for some $k_0 \in \mathcal{K}^*$. Now consider lowering α'' to obtain α in $K^* - 1$ steps, lowering (or keeping unaltered) the productivity shock associated with each good $k \neq k_0 \in \mathcal{K}^*$ one at a time. By Lemma 6 and property (ii), the excess demand for good k_0 never increases and strictly decreases in at least one step since $\left(\alpha''^k\right)_{k \in \mathcal{K}^*} \neq \left(\alpha^k\right)_{k \in \mathcal{K}^*}$. This implies $z^{k_0}(\alpha, X) < z^{k_0}(\alpha'', X)$, which contradicts property (i). ∎

A.4 Proof of Theorem 1

We now use the previous intermediary results to establish the proof of Theorem 1.

Proof of Theorem 1. We proceed in four steps.

Step 1: *For all $X \in \mathcal{X}$, there exists $\alpha \in \mathcal{A}^*$ such that Equations* (11)-(12) *hold for $L \in L(\alpha, X)$. Furthermore, if there exists another $\alpha' \in \mathcal{A}^*$ such that Equations* (11)-(12) *hold for $L' \in L(\alpha', X)$, then $\alpha^k = \alpha'^k$ for all $k \in \mathcal{K}^*$.*

By Lemmas 5, we know that there exists $\alpha_0 \in \mathcal{A}^*$ such that Equation (12) holds for $L_0 \in L(\alpha_0, X)$. Now pick $k_0 \in \mathcal{K}^* / \{K+1\}$ and consider $\alpha = \lambda \alpha_0$ and $L \in L(\alpha, X)$, where $\lambda \equiv \hat{Q}^{k_0} / \left[\sum_{f \in \mathcal{F}} \alpha_0^{k_0} \hat{A}^{fk_0} L_0^{fk_0}\right]$. By construction, we have

$$\sum_{f \in \mathcal{F}} \alpha^{k_0} \hat{A}^{fk_0} L^{fk_0} = \lambda \sum_{f \in \mathcal{F}} \alpha_0^{k_0} \hat{A}^{fk_0} L_0^{fk_0} = \hat{Q}^{k_0},$$

where the first equality uses the fact that $L(\cdot, X)$ is homogeneous of degree zero by Lemma 4 Condition (ii). Since $L(\alpha, X) = \arg\max_{L \in \mathcal{L}} \min_{k \in \mathcal{K}^*} \left\{\sum_{f \in \mathcal{F}} \alpha^k \hat{A}^{fk} L^{fk} / \hat{Q}^k\right\}$, we know that

$$\frac{\sum_{f \in \mathcal{F}} \alpha^k \hat{A}^{fk} L^{fk}}{\hat{Q}^k} = \frac{\sum_{f \in \mathcal{F}} \alpha^{k_0} \hat{A}^{fk_0}(f) L^{fk_0}}{\hat{Q}^{k_0}} \text{ for all } k \in \mathcal{K}^*.$$

The two previous expressions imply that Equation (11) holds as well. Thus we have constructed $\alpha \in \mathcal{A}^*$ such that Equations (11)-(12) hold. Now suppose that there exists another $\alpha' \in \mathcal{A}^*$ such that Equations (11)-(12) hold for $L' \in L(\alpha', X)$. Since Equation (12) holds, Lemma 7 implies the existence of $\mu > 0$ such that $\alpha'^k = \mu \alpha_0^k$ for all $k \in \mathcal{K}^*$. Since Equation (11) holds as well, the previous argument implies $\mu = \lambda$, and so, $\alpha^k = \alpha'^k$ for all $k \in \mathcal{K}^*$.

Step 2: *Conditional on $\alpha \in \mathcal{A}^*$ and $L \in L(\alpha, X)$ satisfying Equations* (11)-(12), *the set of prices $p \in \mathcal{P}^*$ such that Conditions* (13)-(14) *hold is non-empty.*

By Lemma 1, we know that for any $X \in \mathcal{X}$ and $\alpha \in \mathcal{A}^*$, if $L \in L(\alpha, X)$, then there exist $p_0 \in \mathcal{P}^*$ and $r_0 \in \mathbb{R}_+^F$ such that (L, p_0, r_0) is a competitive equilibrium. Thus this must also be true for $\alpha \in \mathcal{A}^*$ and $L \in L(\alpha, X)$ satisfying Equations (11)-(12). By Definition 1, (L, p_0, r_0) must satisfy Conditions (2)-(3). Combining this observation with Assumptions A2 and A3, we have therefore found $p_0 \in \mathcal{P}^*$ such that Condition (14) holds. To conclude, consider $p = \mu p_0$, where $\mu \equiv \hat{S} / \left[\sum_{k \in \mathcal{K}_i^* / \{K+1\}} p_0^k \hat{Q}^k\right]$. By construction, $p \in \mathcal{P}^*$ and p satisfies Conditions (13)-(14).

Step 3: *For any $X \in \mathcal{X}$, if a vector of productivity shocks and good prices $(\alpha, p) \in \mathcal{A}^* \times \mathcal{P}^*$ are such that there exists $L \in L(\alpha, X)$ satisfying Conditions* (11)-(14), *then it is admissible.*

By Definitions 1 and 2, we want to show that one can construct a vector of field prices, r, and an allocation of fields, L, such that conditions (2)-(7) hold. A natural candidate for the allocation is $L \in L(\alpha, X)$ such that Equations (11)-(14) hold. The fact that Equation (4) holds for allocation L is immediate from the fact that $L \in L(\alpha, X)$. The fact that Equations (5)-(7) hold derives from Equations (11)-(13), together with the fact that if $k \notin \mathcal{K}^*$, then $\sum_{f \in \mathcal{F}} A^{fk} L^{fk} = \sum_{f \in \mathcal{F}} L^{fk} = 0$. Let us now construct the vector of field prices, r, such that

$r^f = \max_{k\in\mathcal{K}} p^k \alpha^k \hat{A}^{fk}$ for all $f \in \mathcal{F}$. By Assumptions A2 and A3, we can rearrange the previous expression as

$$r^f = \max_{k\in\mathcal{K}} p^k A^{fk}.$$

This immediately implies Inequality (2). To conclude, note that by Condition (14), together with Assumptions A2 and A3, we must have $p^k A^{fk} = \max_{k'\in\mathcal{K}} p^{k'} A^{fk'} = r^f$ if $L^{fk} > 0$. Thus Equation (3) holds as well.

Step 4: *For any $X \in X$, if a vector of productivity shocks and good prices $(\alpha, p) \in \mathcal{A}^* \times \mathcal{P}^*$ is admissible then there exist $\tilde{\alpha} \in \mathcal{A}^*$ satisfying $\tilde{\alpha}^k = \alpha^k$ for all $k \neq K+1$, $\tilde{p} \in \mathcal{P}^*$ satisfying $\tilde{p}^k = p^k$ for all $k \neq K+1$, and $L \in L(\tilde{\alpha}, X)$ such that Conditions* (11)-(14) *hold.*

By Definitions 1 and 2, if $(\alpha, p) \in \mathcal{A}^* \times \mathcal{P}^*$ is admissible given $X \in \mathcal{X}$, then there exist r and L such that conditions (2)-(9) hold. Equations (6) and (7) imply Equations (11) and (12), respectively. Conditions (2) and (3)—together with Assumptions A2 and A3—imply Condition (14), with our convention $\hat{A}^{fK+1} = 1$. Finally, Equations (5) and (6) imply Equation (13). Furthermore, since good $K+1$ only appears in Condition (14), Conditions (11)-(13) must also hold for any $\tilde{\alpha} \in \mathcal{A}^*$ satisfying $\tilde{\alpha}^k = \alpha^k$ for all $k \neq K+1$ and any $\tilde{p} \in \mathcal{P}^*$ satisfying $\tilde{p}^k = p^k$ for all $k \neq K+1$ provided that $\tilde{\alpha}^{K+1}\tilde{p}^{K+1} = \alpha^{K+1}p^{K+1}$. Thus all we need to establish at this point is that one can construct $\tilde{\alpha} \in \mathcal{A}^*$ such that $\tilde{\alpha}^k = \alpha^k$ for all $k \neq K+1$ and $L \in L(\tilde{\alpha}, X)$.

We follow a guess and verify strategy. Let us construct $\tilde{\alpha} \in \mathcal{A}^*$ such that $\tilde{\alpha}^k = \alpha^k$ for all $k \neq K+1$ and $\tilde{\alpha}^{K+1} = \hat{Q}^{K+1} / \sum_{f\in\mathcal{F}} \hat{A}^{fK+1} L^{fK+1}$, with our convention $\hat{Q}^{K+1} = \hat{A}^{fK+1} = 1$. By conditions (2)-(4), L is a feasible allocation that maximizes total profits. Thus, by the First Welfare Theorem (Mas-Colell et al. Proposition 5.F.1), L must be a solution of

$$\max_{\tilde{L}\in\mathcal{L}} \; \sum\nolimits_{f\in\mathcal{F}} \alpha^{K+1} \hat{A}^{fK+1} \tilde{L}^{fK+1}$$
$$\sum\nolimits_{f\in\mathcal{F}} \alpha^k \hat{A}^{fk} \tilde{L}^{fk} \geq \hat{Q}^k, \text{ for all } k \neq K+1.$$

By construction of $\tilde{\alpha}$, we therefore also have

$$\max_{\tilde{L}\in\mathcal{L}} \; \sum\nolimits_{f\in\mathcal{F}} \tilde{\alpha}^{K+1} \hat{A}^{fK+1} \tilde{L}^{fK+1} \tag{P}$$
$$\sum\nolimits_{f\in\mathcal{F}} \tilde{\alpha}^k \hat{A}^{fk} \tilde{L}^{fk} \geq \hat{Q}^k, \text{ for all } k \neq K+1. \tag{55}$$

The rest of the proof proceeds by contradiction. Suppose that there exists $L' \in \mathcal{L}$ such that

$$\min_{k\in\mathcal{K}^*} \left\{ \sum\nolimits_{f\in\mathcal{F}} \tilde{\alpha}^k \hat{A}^{fk} L'^{fk} / \hat{Q}^k \right\} > \min_{k\in\mathcal{K}^*} \left\{ \sum\nolimits_{f\in\mathcal{F}} \tilde{\alpha}^k \hat{A}^{fk} L^{fk} / \hat{Q}^k \right\}. \tag{56}$$

By Equation (6) and the definition of $\tilde{\alpha}$, we know that $\min_{k\in\mathcal{K}^*} \left\{ \sum_{f\in\mathcal{F}} \tilde{\alpha}^k \hat{A}^{fk} L^{fk} / \hat{Q}^k \right\} = 1$. Thus Inequality (56) implies

$$\sum\nolimits_{f\in\mathcal{F}} \tilde{\alpha}^k \hat{A}^{fk} L'^{fk} > \hat{Q}^k, \text{ for all } k \in \mathcal{K}^*. \tag{57}$$

Since $L'^{fk} \geq 0$ for all $k \in \mathcal{K}$, $f \in \mathcal{F}$, we also trivially have

$$\sum_{f\in\mathcal{F}} \tilde{\alpha}^k \hat{A}^{fk} L'^{fk} \geq \hat{Q}^k, \text{ for all } k \notin \mathcal{K}^*. \tag{58}$$

By Inequalities (57) and (58), L' satisfies constraint (55). By Inequality (57) evaluated at $k = K+1$, we also have

$$\sum_{f\in\mathcal{F}} \tilde{\alpha}^{K+1} \hat{A}^{fK+1} L'^{fK+1} > \sum_{f\in\mathcal{F}} \tilde{\alpha}^{K+1} \hat{A}^{fK+1} L^{fK+1},$$

which contradicts the fact that L is a solution of (P). Thus there cannot be $L' \in \mathcal{L}$ such that Inequality (56) holds, which implies that $L \in L(\tilde{\alpha}, X)$. This completes the proof of Step 4.

For any $X \in \mathcal{X}$, Steps 1 and 2 imply the existence of $(\alpha, p) \in \mathcal{A}^* \times \mathcal{P}^*$ and $L \in L(\alpha, X)$ such that Conditions (11)-(14) hold. By Step 3, for any $X \in \mathcal{X}$, there therefore exists a vector of productivity shocks and good prices $(\alpha, p) \in \mathcal{A}^* \times \mathcal{P}^*$ that is admissible. So the the set of admissible productivity shocks and prices in $\mathcal{A}^* \times \mathcal{P}^*$ is non-empty.

To conclude, we need to show that Properties (i) and (ii) hold. We start with Property (i). Suppose that there exist $(\alpha_1, p_1) \in \mathcal{A}^* \times \mathcal{P}^*$ and $(\alpha_2, p_2) \in \mathcal{A}^* \times \mathcal{P}^*$ that are admissible and satisfy $\alpha_1^{k_0} \neq \alpha_2^{k_0}$ for some $k_0 \in \mathcal{K}^* / \{K+1\}$. By Step 4, there must exist $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ such that for $j = 1, 2$, $\tilde{\alpha}_j^k = \alpha_j^k$ for all $k \neq K+1$ and there exists $L_j \in L(\tilde{\alpha}_j, X)$ satisfying Equations (11)-(12). By Step 1, however, there exists a unique α such that $L \in L(\alpha, X)$ satisfies Equations (11)-(12). Thus we must have $\alpha_1^k = \alpha_2^k$ for all $k \neq K+1$, which contradicts $\alpha_1^{k_0} \neq \alpha_2^{k_0}$. This establishes Property (i).

We now turn to Property (ii). Suppose that $\alpha \in \mathcal{A}^*$ and $L \in L(\alpha, X)$ satisfy Equations (11)-(12). By Step 3, if Conditions (13)-(14) hold, then $(\alpha, p) \in \mathcal{A}^* \times \mathcal{P}^*$ is admissible. Conversely, if $(\alpha, p) \in \mathcal{A}^* \times \mathcal{P}^*$ is admissible, then Step 4 implies the existence of $\tilde{\alpha}$ and $\tilde{L} \in L(\tilde{\alpha}, X)$ such that Conditions (11)-(14) hold. Since $\alpha \in \mathcal{A}^*$ and $\tilde{\alpha} \in \mathcal{A}^*$ are both such that there exist $L \in L(\alpha, X)$ and $\tilde{L} \in L(\tilde{\alpha}, X)$ satisfying Equations (11)-(12), Step 1 implies $\alpha = \tilde{\alpha}$. Thus $(\alpha, p) \in \mathcal{A}^* \times \mathcal{P}^*$ are such that Conditions (13)-(14) hold. This establishes Property (ii) and completes the proof of Theorem 1. ∎