Exaggerated Death of Distance:

Revisiting Distance Effects on Regional Price Dispersions

Kazuko Kano[†]
Gruduate School of Economics
The University of Tokyo
Hongo 7-3-1, Bunkyo-ku
Tokyo, 113-0033, Japan
Tel: +81-3-5841-1746
Fax: +81-3-5841-1747
Email: kkano@e.u-tokyo.ac.jp

Takashi Kano[‡]
Graduate School of Economics
The University of Tokyo
Hongo 7-3-1, Bunkyo-ku
Tokyo, 113-0033, Japan
Tel: +81-3-5841-5655
Fax: +81-3-5841-5655
Email: tkano@e.u-tokyo.ac.jp

Kazutaka Takechi
Faculty of Economics
Hosei University
4342 Aihara-machi, Machida-shi
Tokyo, 194-0298, Japan
Tel: +81-42-783-2566
Fax: +81-42-783-2611
Email: ktakechi@hosei.ac.jp

Current Draft: April 11, 2011

Abstract

Past studies in the literature of the law of one price (LOP) show statistically significant but economically subtle roles of geographical distance in regional price dispersions. In this paper, we challenge this empirical "death of distance" as a primary source of LOP violations investigating a unique daily data set of wholesale prices of agricultural products in Japan that enables us to identify source regions and observe product-delivery patterns to consuming regions. We build a simple structural model to explain the observed product-delivery patterns and argue that ignoring the underlying delivery choice results in a serious under-bias toward inferences on distance effects on regional price dispersions due to sample selection. Estimating a sample-selection model, on which theoretical restrictions of our structural model are imposed, with data of several agricultural products, we find quite large estimates of the distance elasticity of price differential compared with conventional estimates. This paper, hence, provides evidence that conventional estimates of the distance elasticity could be heavily biased downwards and spuriously underestimate the role transportation costs play in regional price dispersions and LOP violations.

Key Words : *Law of one price; Regional price dispersion; Transportation cost; Geographical distance; Agricultural wholesale price; Sample-selection bias*

JEL Classification Number : F11, F14, F41

[†] We would like to thank Andrew Bernard, Toni Braun, Alexandre Dmitriev, Shin-ichi Fukuda, Naoto Jinji, Daiji Kawaguchi, Michael Keane, Junko Koeda, Chul-In Lee, Hideyuki Mizobuchi, Toshi Mukoyama, Makoto Nirei, Hiroshi Ohashi, Glenn Otto, Yasuyuki Sawada, Shigenori Shiratsuka, Alexandre Skiba, Yi-Chan Tsai, Takayuki Tsuruga, Tsutomu Watanabe, and seminar participants in the 2010 Asia Pacific Trade Seminar, fifth conference of Empirical Investigations in Trade and Investment in Keio University (March 2011), the Institute of Developing Economies of the Japan External Trade Organization, Hosei University, Keio University, Kyoto University, National Graduate Institute for Policy Study, the Policy Research Institute of the Ministry of Fiance of the Government of Japan, the Research Institute of Capital Formation of the Development Bank of Japan, Seoul National University, Summer Workshop on Economic Theory 2010, Tokyo Macro Workshop (July 2010), and University of New South Wales for helpful and valuable discussions and suggestions. The first and second authors wish to thank the Kikawada foundation for financial support. The second and third authors would like to thank the Japan Society for the Promotion of Science for financial supports from grant-in-aids for scientific research (numbers 20730205 and 21730205), respectively. We are solely responsible for any errors and misinterpretations of this paper.

[‡] *Corresponding author.*

1. Introduction

Does an identical good share an identical price across geographically distinct places? Many of recent papers approach this fundamental question of the law of one price (LOP) exploiting micro-level information of retail prices observed across retail stores internationally as well as domestically. Since the seminal works by Parsley and Wei (1996) and Engel and Rogers (1996), one of the most robust findings within the literature of the LOP is a statistically significant effect of geographical distance on statistical properties of cross-regional retail price differentials within a variety of gravity-type regressions. Given economic rationale provided by iceberg-type transportation costs, this robust finding suggests that transportation costs play a statistically significant role in the observed violations of the absolute LOP hindering cross-regional arbitrage of products.

The size of the distance effect that is commonly estimated in this literature, nevertheless, seems economically subtle. Regressing the absolute value of the difference in the log of the retail price of an identical product between two retail stores in two distinct regions on the log of the corresponding geographical distance, many of past studies infer less than about 3 % elasticity of price differential with respect to distance.[1] This means that even when the geographical distance of two distinct cities becomes double, the price differential of a product between the cities increases at best only 3 % on average. Since the standard deviation of the absolute value of the log price differential is typically reported around 20 % in this literature, we need the standard deviation of the log of distance of 6.66 (=0.20/0.03) if we want to explain the observed degree of regional price dispersions only by geographical distance. The required standard deviation of the log of distance, however, is too large to be consistent with the actual degree of geographical scattering of cities.[2] This observation naturally casts doubt on transportation costs, which are approximately measured by distance, as a main economic source of regional price dispersions. In this sense, geographical distance is empirically "dead" as a prime suspect for the commonly observed violations of the LOP.

What is further puzzling is the fact that past empirical studies of international trade unambiguously recognize that geographical distance plays economically crucial roles in determinations of bilateral trade directions and volumes. For example, Anderson and van Wincoop (2003) estimate a gravity model of bilateral trade volumes controlling for multilateral trade resistance and infer the distance elasticity of transportation costs to be around 20 % conditional on a calibrated elasticity of substitution equal to 5. Estimating a gravity model using bilateral export volume data across 183 countries, Helpman et al. (2008) find that the distance elasticity of bilateral export volumes is about 80 % once they take into account firms' selections into bilateral exports as well as firms'

[1] Among a series of past studies, for example, Broda and Weinstein (2008) observe the 1.2 % distance elasticity of the absolute log price differentials within barcode-level scanner data of retail prices across Canadian and the U.S. cities. Engel et al.(2005) find the distance elasticity of 0.32 % with pooled annual panel data distributed by Economic Intelligence Unit (EIU) that covers retail prices of 100 consumer goods surveyed in 17 Canadian and the U.S. cities. Ceglowski (2003) reports 1.6-2.0 % estimates of the distance elasticities of 45 different products across 25 Canadian cities. Baba (2007) scrutinizes Japanese and Korean retail price survey data and estimates less than about 3 % of the distance elasticity after taking into account a border dummy between the two countries.

[2] For instance, the standard deviation of the log of distance between two prefectural capital cities in Japan is 0.803 over all the 1081 city-pairs from 47 prefectures.

heterogeneity in export volumes.[3] Interestingly, their estimate suggests a 20 % distance elasticity of transportation costs once we calibrate the price elasticity of demand equal to 5 as in Anderson and van Wincoop (2003). Since these studies also exploit iceberg-type transportation costs to characterize their gravity equations, the huge discrepancy in terms of the estimated size of the distance elasticity of transportation costs between the above two research agenda — the absolute LOP and the gravity model of international trade — is indeed an empirical challenge students of international economics need to explore profoundly.

In this paper, we tackle this empirical "death of distance" in regional price dispersions. Our contributions are threefold. First, this paper investigates a unique daily data set of wholesale prices of agricultural products in Japan.[4] We follow the spirit of Parsley and Wei (1996) using disaggregate price data within a country to avoid any potential effects of cross-country differences in tax and currency on our inference on transportation costs. Scrutinizing information of wholesale prices helps us make our estimate of transportation costs immune against influences of local distributional costs as well as local retailers' pricing strategies such as temporary sales. More importantly, there are two outstanding characteristics of this data set: (i) we can identify the wholesale prices of an identical product at both producing and consuming regions and (ii) we can also grasp daily delivery patterns of an identical product from the former region to the latter. The first characteristic is essential for identifying transportation costs because, as discussed by Anderson and van Wincoop (2004), only when the source region of a product is identified, correct information of transportation costs could be extracted from relative prices at consuming regions to the corresponding source region. The main difficulty past studies face is in the fact that a retail price survey at retail stores rarely provides information of the source regions of a product and the market prices prevailed in these regions. Our data set, on the other hand, shows us not only in which regions in Japan a variety of fruits and vegetables are produced but also at what wholesale prices these products are sold in their originated regions.[5]

Identification of the source region of a product, however, immediately leads to another fundamental question: how far a product is delivered from the source region? The second outstanding aspect of our data set empirically shows us the answer to this question. As the second contribution of this paper, we build a model to explain the observed patterns of product delivery and claim theoretically that ignoring the underlying choice of delivery might result in a serious under-bias toward

[3]This size of distance effects on export volumes is common in the literature of empirical trade. For example, in their meta analysis based on 1,051 past estimates of distance effects, Disdier and Head (2008) report the average of 0.893.

[4]This is not the first paper that intensively scrutinizes price data of agricultural products in the literature of the LOP and PPP. Midrigan (2007) employs prices of agricultural products sold in open-air markets in European countries to test theoretical implications of his state-dependint pricing model with trade costs.

[5]In a recent paper, Inanc and Zachariadis (2010) identify source regions of products reported in the Eurostat survey in several indirect ways and find around 10 % distance elasticity of price differentials in the 1990 survey. This could be indirect evidence that identification of the origin of a product is essential for inference of transportation costs. A more direct identification of source regions is taken by Donaldson (2010) who scrutinizes a cross-regional data of prices of salt in North India in the British colonial period. In his paper, source regions of salt are identified because salt was produced only in several licensed districts in India. He observes about 24 % distance elasticity of price differentials of salts.

our inference on the role of distance in regional price dispersions. To see this, suppose that transportation costs are unobservable and comprise two components: the one increasing proportionally in geographical distance and the other unobservable. A rise in transportation costs increases the price of the product at a consuming region relative to those of other substitutable products and depresses the corresponding local demand for the product. Given the shape of the demand function, this fall in the local product demand then tends to lower the profitability and, as a result, the probability of delivery to the consuming region from the producing region. Since the price of the product at a consuming region is observed only when a product delivery indeed occurs, an inference drawn only from information of price differentials could be subject to a sample-selection bias due to an incidental truncation. In particular, the direction of the potential bias should be downward because a rise in the unobservable component of transportation costs in general increases a price differential but deteriorates a probability of delivery at the same time.

In this paper, following Melitz (2003) and Helpman et al. (2008), we build a simple structural model of cross-regional product-delivery in which cross-regional price differentials and delivery patterns are jointly determined by the same structure of transportation costs.[6] We then show that the degree of a sample-selection bias depends critically on two structural parameters of the model: the elasticity of transportation costs to distance and that of demand to price. Our theoretical analysis implies that drawing a correct inference on transportation costs requires us to estimate these two elasticities jointly. To do so, we propose a structural sample-selection model, which consists of the price differential and sample-selection equations, imposing nonlinear theoretical restrictions on the joint probability distribution of data. We develop a full information maximum likelihood (FIML) estimator incorporating instrumental variables for the empirical model. Our Monte Carlo experiments based on the model not only show us that, given the price elasticity of demand, the degree of sample selection depends positively on the distance elasticity of transportation costs but also uncover two crucial facts: (i) the standard exercise of regressing price differentials on the corresponding distances provides a heavily downwards-biased estimate of the true distance elasticity of transportation costs and (ii) our FIML estimator successfully identifies the distance elasticity.

Finally, as the third contribution of this paper, we estimate our sample-selection model by FIML using the data of wholesale prices of several selected vegetables. The estimated sample-selection model passes two diagnosis criteria in that it does a fairly good job in replicating the actual delivery patterns of these vegetables as well as the actual data association of price differentials with distances. We find large estimates of the distance elasticity of transportation costs across all the vegetables relative to the existing estimates in the LOP literature: all of them are more than 20 % and their average is about 24 %. Given the 24 % distance elasticity of transportation costs, we need only the standard deviation of the log of distance of 0.833 (=0.2/0.24) if we want to explain only by distance the whole part of the commonly observed standard deviation of the log of price

[6] Closely related to this paper, Johnson (2010) investigates the implications of the model of Helpman et al. (2008) on aggregate sectoral export prices. He exploits model's implications on f.o.b. export prices to improve statistical inferences on the role of firms' heterogeneity in the intensive and extensive margins of international trades. Using f.o.b. export prices, however, makes his econometrics exercise silent about distance effects on price differentials across countries or regions.

differential of 20 %. The estimate of this paper, therefore, implies an economically critical role of transportation costs in regional price dispersions. It is worth noting that this large distance elasticity does not necessarily stem from a particular characteristic of the product category of agricultural products. To prove this, we also conduct the OLS regression exercise without respecting the selection mechanism using our wholesale price data. Interestingly, we obtain the conventional range of the OLS estimate of the distance elasticity about 3 %. This provides evidence that conventional estimates of the distance elasticity could be heavily biased downwards and spuriously underestimate the role transportation costs play in regional price dispersions and LOP violations.[7]

The organization of the rest of this paper is as follows. In the next section, we introduce our model and derive our FIML estimator based on the corresponding sample-selection model. In section 3, we conduct Monte Carlo experiments to check the validity of the FIML estimator. Section 4 describes our data set. After reporting the empirical results in section 5, we conclude in section 6.

2. Model and empirical framework

2.1. A model of cross-regional product delivery

The empirical analysis of this paper relies on a model of monopolistic competitive firms as in Melitz (2003) and Helpman et al. (2008). In this model, a country consists of I distinct regions indexed by $i = 1, 2, \cdots, I$. In each region i, the representative household consumes a continuum of agricultural products indexed by l that takes a value between the closed unit interval, i.e., $l \in [0, 1]$. We assume that the representative household in each region purchases an identical set of agricultural products at the regional wholesale market and raises its utility with a Dixit-Stigliz type constant elastic function $u_i = [\int_0^1 x_i(l)^{\alpha} dl]^{1/\alpha}$ with $0 < \alpha < 1$, where $x_i(l)$ is the consumption level of product l in region i. This utility function implies that the elasticity of substitution across products is $\epsilon = 1/(1-\alpha) > 1$, which is assumed to be common across all regions. Region i's demand function for product l under the average price of product l, $p_i(l)$, is $x_i(l) = [p_i(l)/p_i]^{-\epsilon} x_i$ where $p_i = [\int_0^1 p_i(l)^{1-\epsilon} dl]^{1/(1-\epsilon)}$ represents the consumer price index (CPI) aggregated over all agricultural products in region i and $x_i \equiv u_i$ indicates the indirect utility represented as the aggregate consumption level of agricultural products in the region i.

We assume that each product l can be produced in all regions with an identical production technology discussed below. Producing region j of product l, then, delivers its product to the wholesale markets in the same region j as well as distinct consuming regions $i \neq j$ only if the delivery is profitable. Let $x_i(j, l)$ denote the demand of region i for product l produced in and delivered from region j. The representative household in region i then earns its utility from consuming product l

[7] In other words, the conventional estimator suffers from a low power problem: it cannot distinguish between the null hypotheses of high and low transportation costs.

with the following constant elastic utility function

$$x_i(l) = \left[\int_{j \in B_i(l)} \{\delta_i(j,l) x_i(j,l)\}^{\alpha_l} dj \right]^{1/\alpha_l}, 0 < \alpha_l < 1,$$

where $B_i(l)$ is the set of the producing regions that deliver product l to region i. This utility function specific to product l exhibits that the representative household in region i recognizes product l, if it is produced in different source regions, as different products: the substitution of product l across distinct source regions is imperfect with the constant elasticity $\epsilon_l = 1/(1-\alpha_l)$. Term $\delta_i(j,l)$ reflects the household's biased preference on different producing regions: the greater the term $\delta_i(j,l)$ is, the more the household in region i prefers product l from source region j relative to those from other source regions, *ceteris paribus*. The above utility function then derives region i's demand function for product l produced in region j under the price $p_i(j,l)$

$$x_i(j,l) = \left[\frac{p_i(j,l)}{p_i(l)} \right]^{-\epsilon_l} \delta_i(j,l)^{\epsilon_l - 1} x_i(l), \tag{1}$$

where $p_i(l) = [\int_{j \in B_i(l)} \{\delta_i(j,l) p_i(j,l)\}^{1-\epsilon_l} dj]^{1/(1-\epsilon_l)}$ is the the aggregate price level of product l in region i.

A producer in region j is a monopolistically competitive producer at the wholesale markets in the same region as well as the other regions to deliver. As specified by Helpman et al. (2008), a producer in region j yields a unit of an agricultural product paying costs minimizing a bundle of factor inputs. The marginal cost of producing product l is denoted by $c_j a(l)$, where $a(l)$ measures the number of bundles of factor inputs used per unit output of product l and c_j measures the unit cost of this bundle of factor inputs. Notice that $a(l)$ is product-specific, while c_j is region-specific. This means that the efficient combination of inputs for producing a product is common across regions, while factor costs are different across regions.

At the wholesale market in region j, the producer of product l, who faces the demand function (1), maximizes profits by charging markup price $p_j(j,l) = c_j a(l)/\alpha$. This means that a producer of a region does not need to bear any transportation costs when selling its product at the wholesale market in the same region. On the other hand, if the same producer seeks to sell its product at the wholesale market in region $i \neq j$, two types of delivery costs should be borne by the producer: a fixed cost of serving at the market in region i, denoted by $c_j f_{ij}$, and an "iceberg"-type transportation cost, denoted by τ_{ij}. Hence, as in Helpman et al. (2008), we assume that $f_{jj} = 0$ for any j and $f_{ij} > 0$ for $i \neq j$, and $\tau_{jj} = 1$ for any j and $\tau_{ij} > 1$ for $i \neq j$. The optimal price to set, $p_i(j,l)$, then is

$$p_i(j,l) = \tau_{ij} \frac{c_j a(l)}{\alpha_l}. \tag{2}$$

In this case, the operating profits of delivering product l to region i is

$$\pi_{ij}(l) = (1-\alpha_l) \left[\frac{\tau_{ij} c_j}{\alpha_l p_i(l)} \right]^{1-\epsilon_l} \theta_i(j,l)^{\epsilon_l - 1} p_i(l) x_i(l) - c_j f_{ij},$$

where $\theta_i(j,l)$ is the ratio of the productivity level to the producing regional bias $a(l)/\delta_i(j,l)$. If the producer in region j sells its product l at the regional wholesale market, monopolistic profit $\pi_{jj}(l)$ is always positive because $f_{jj} = 0$ and $\tau_{jj} = 1$. However, delivering the same product to another consuming region i is profitable only if $\theta_i(j,l)$ is smaller than a threshold $\bar{\theta}_{ij}(l)$, where $\bar{\theta}_{ij}(l)$ is defined by the zero profit condition $\pi_{ij}(l) = 0$, or equivalently,

$$(1-\alpha_l)\left[\frac{\tau_{ij}c_j}{\alpha_l p_i(l)}\right]^{1-\epsilon_l}\bar{\theta}_{ij}(l)^{1-\epsilon_l}p_i(l)x_i(l) = c_j f_{ij}. \tag{3}$$

Let $T_{ij}(l)$ denote an indicator function that takes the value of one if there is a delivery of product l from source region j to consuming region i, and the value of zero if there is no delivery. The above determination of the threshold (3), then, implies

$$T_{ij}(l) = \begin{cases} 1 & \text{if} \quad \theta_i(j,l) < \bar{\theta}_{ij}(l), \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Therefore, equations (3) and (4) describe the decision mechanism of a profitable delivery.

Optimal price (2) implies that a price differential of an identical product between source and consuming regions provides a precise identification of transportation cost τ_{ij}. To see this, let $q_{ij}(l)$ denote the log of the price differential of product l between producing and consuming regions j and i: $q_{ij}(l) \equiv \ln p_i(j,l) - \ln p_j(j,l)$. Then, optimal price (2) and delivery decision mechanism (4) together yield the price differential equation

$$q_{ij}(l) = \ln \tau_{ij}, \quad \text{only if } T_{ij}(l) = 1. \tag{5}$$

Price differential equation (5) has two important empirical implications. First, transportation cost τ_{ij} can be measured from the corresponding price differential only when we can identify the prices in the producing and consuming regions. This is exactly the argument of Anderson and van Wincoop (2004) against the conventional approach to measuring transportation costs in the literature of regional and cross-country price dispersions. The second implication, however, says that identifying producing and consuming regions is not enough for estimating transportation costs precisely. Equation (5) shows that there is an incidental truncation or sample section: we can observe the price differential of product l between producing and consuming regions only when the product is indeed delivered from the former region to the latter. Hence, the sample is non-randomly selected by the selection mechanism of (3) and (4). This selection mechanism indeed depends on transportation cost τ_{ij}. Therefore, transportation cost τ_{ij} in equation (5) could be inconsistently estimated unless we can take into account sample-selection mechanism (4) explicitly.

An important caveat of the above identification of transportation costs stems from product arbitrage. With product arbitrage, price differential (5) might not be a sufficient statistic of the underlying transportation cost because the observed equilibrium price in a consuming region can deviate from the optimal price (2). There could be three possibilities of product arbitrage in this model. The first possibility is arbitrage of product l across source regions $j = 1, \cdots, J$ that occurs within each consuming region i. However, because the demand function (1) implies an imperfect degree of product substitution with a constant elasticity ϵ_l, the arbitrage of product l across different producing regions is also imperfect in each consuming region i. The second possibility is the case that product l cropping in source region j is delivered to consuming region i once and then transferred to another consuming region $k \neq i$ that is not delivered from the original source region j: $j \in B_i(l)$ but $j \notin B_k(l)$. There is no profitable product transfer if the inequality $\tau_{ki}\tau_{ij} \geq \tau_{kj}$ holds.[8] Finally the third possibility is the case that product l produced in source region j are delivered to two consuming regions i and k and the arbitrage of product l occurs between delivered consuming regions i and k: $j \in B_i(l)$ and $j \in B_k(l)$. Since product l from source region j is perfectly substitutable between two delivered consuming regions i and k, the standard no-arbitrage band $1/\tau_{ki} \leq p_i(j,l)/p_k(j,l) = \tau_{ij}/\tau_{kj} \leq \tau_{ki}$ can be applicable: if this condition holds, the product arbitrage in the third type does not occurs between delivered consuming regions i and k.

In this paper, we do not impose the no-arbitrage conditions of the second and third types on our data. However, as we discuss more details in section 5, the amount of product transfers across wholesale markets of agricultural products is quite small relative to the total amount of wholesale transactions in Japan. We interpret this fact as almost no opportunity of product arbitrage in equilibrium wholesale prices in our data set. We simply control for any possible effects of product arbitrage on price differentials by adding an i.i.d. zero-mean random error to the price differential equation (5).[9]

2.2. The empirical framework

Given the structural model, we now discuss our empirical framework for estimating transportation cost τ_{ij}. Following Helpman et al.(2008), we specify transportation cost τ_{ij} parametrically by $D_{ij}^{\gamma}\exp(\mu + u_{ij})$ where D_{ij} represents the symmetric distance between regions i and j, and $u_{ij} \sim N(0, \sigma_u^2)$ is an i.i.d. unobserved region-pair specific element of the transportation cost. Positive constant $\mu > 0$ makes it possible that the transportation cost always takes a value greater than 1 for all (i,j) pairs. Price differential (5), then, is

$$q_{ij}(l) = \mu + \gamma d_{ij} + u_{ij}, \quad \text{only if } T_{ij}(l) = 1. \tag{6}$$

The delivery choice of product l from region j to region i is determined by threshold $\bar{\theta}_{ij}(l)$ defined by

[8] This inequality is the direct result of the following two conditions. Notice that the condition for source region j not to deliver its product l to consuming region k is $(1-\alpha_l)(\tau_{kj}c_j/\alpha_l p_k)^{1-\epsilon_l}\theta_k(j,l)^{1-\epsilon_l}p_k x_k < c_j f_{kj}$. Also the condition for source region j takes an option in which it delivers product l to region i and subsequently transfers the product to region k is $(1-\alpha_l)(\tau_{ki}\tau_{ij}c_j/\alpha_l p_k)^{1-\epsilon_l}\theta_k(j,l)^{1-\epsilon_l}p_k x_k \geq c_j f_{kj}$, when we assume the fixed costs are identical between the former and latter delivery options. The two conditions turn out to be $\tau_{ki}\tau_{ij} < \tau_{kj}$, and, hence, no profitable product transfer occurs if the inequality $\tau_{ki}\tau_{ij} \geq \tau_{kj}$ holds. As imposed by Bernard et al.(2003) on their Ricardian model, this well-known "triangular inequality" means that the transportation cost of delivering the product from source region j to consuming region i and then transferring the product further to consuming region k is more expensive than that of delivering the same product from source region j to consuming region k directly.

[9] Atkeson and Burstein (2008) also discuss a possibility of international product arbitrage in their two-country general equilibrium model with imperfect competition and trade costs. They report no role of product arbitrage in their quantitative simulation results. Therefore, our data set shares the same characteristic on product arbitrage as in their simulation exercise.

zero profit condition (3). Define a latent $Z_{ij}(l) \equiv (1-\alpha_l)[\tau_{ij}c_j/\alpha_l p_i(l)]^{1-\epsilon_l}\theta_i(j,l)^{1-\epsilon_l}p_i(l)x_i(l)/c_j f_{ij}$. Product l, then, is delivered from region j to region i only if $Z_{ij}(l) > 1$. We assume that the fixed cost of delivery, f_{ij}, is stochastic due to an i.i.d. unobserved regional-pair specific element v_{ij}. Just as in Helpman et al. (2008), we exploit a parametric specification of f_{ij}: $f_{ij} \equiv \exp(\lambda_j + \lambda_i - v_{ij})$, where $v_{ij} \sim i.i.d.N(0, \sigma_v^2)$ and is uncorrelated with u_{ij}. The log of $Z_{ij}(l)$, $z_{ij}(l)$, is

$$z_{ij}(l) = \beta - (\epsilon_l - 1)\gamma d_{ij} + \epsilon_l \ln p_i(l) + \ln x_i(l) + \xi_j + \lambda_i + \omega_l - \varrho_{ijl} + \eta_{ij}, \tag{7}$$

where $\beta \equiv \ln(1-\alpha_l) + (\epsilon_l - 1)\ln\alpha_l + (1-\epsilon_l)\mu$, $\xi_j \equiv -\epsilon_l \ln c_j - \lambda_j$, $\omega_l \equiv (1-\epsilon_l)\ln a(l)$, $\varrho_{ijl} \equiv (1-\epsilon_l)\ln\delta_i(j,l)$, and $\eta_{ij} \equiv (1-\epsilon_l)u_{ij} + v_{ij} \sim i.i.d.N(0, (1-\epsilon_l)^2\sigma_u^2 + \sigma_v^2)$. Selection equation (7) then implies that $T_{ij}(l) = 1$ only if $z_{ij}(l) > 0$.

Price differential equation (6) and selection equation (7) jointly reveal two critical aspects when identifying the distance elasticity of transportation costs, γ. First, estimating γ respecting only price differential equation (6) might lead to an under-biased inference. To see this, taking the conditional expectation of price differential equation (6) on $T_{ij}(l) = 1$ and other observable yields $E[q_{ij}(l)|., T_{ij}(l) = 1] = \mu + \gamma d_{ij} + E[u_{ij}|., T_{ij}(l) = 1]$ where . represents other observable. Notice that the term $E[u_{ij}|., T_{ij}(l) = 1]$ is related to the conditional expectation $\bar{\eta}_{ij} \equiv E[\eta_{ij}|., T_{ij}(l) = 1]$ by $E[u_{ij}|., T_{ij}(l) = 1] = \rho\frac{\sigma_u}{\sigma_\eta}\bar{\eta}_{ij}$, where ρ is the correlation coefficient between u_{ij} and η_{ij} and $\sigma_\eta = (1-\epsilon)^2\sigma_u^2 + \sigma_v^2$. A consistent estimate of $\bar{\eta}_{ij}$ is obtained by the inverse Mills ratio $\hat{\bar{\eta}}_{ij}(l) = \phi[\hat{z}_{ij}(l)]/\Phi[\hat{z}_{ij}(l)]$ where $\phi(.)$ and $\Phi(.)$ are the standard normal density and cumulative distribution functions, respectively. Therefore, we can rewrite price differential equation (6) as

$$q_{ij}(l) = \mu + \gamma d_{ij} + \beta_{u\eta}\hat{\bar{\eta}}_{ij}(l) + e_{ij}(l), \tag{8}$$

where $\beta_{u\eta} = \rho\frac{\sigma_u}{\sigma_\eta}$, and $e_{ij}(l)$ is an i.i.d. error term satisfying $E[e_{ij}(l)|., T_{ij}(l) = 1] = 0$. Our model implies that given $\epsilon > 1$, the error term in the selection equation, η_{ij}, could be correlated negatively with that in the price differential equation, u_{ij}: $\rho < 0$.[10] Moreover, the inverse Mills ratio $\hat{\bar{\eta}}_{ij}(l)$ is increasing in distance because $\hat{\bar{\eta}}_{ij}(l)$ is a decreasing function of the predicted latent variable $\hat{z}_{ij}(l)$ that then depends negatively on distance through selection equation (7). Hence if we ignore the third term of the RHS of equation (8) when estimating distance elasticity γ only through the price differential equation, the resulting estimate could be biased downwards.

Second, the size of the under bias depends crucially on the price elasticity of demand, ϵ_l. This is because, given the optimal markup price set in a consuming region, selection equation (7) implies that a larger price elasticity leads to a smaller demand for the corresponding product sold in the consuming region and, as a result, lesser profitability of the delivery of the product from the source to the consuming regions. Therefore, the under bias due to the sample selection becomes worse with a larger price elasticity of demand. Moreover, the effect of distance on the delivery choice depends on the distance elasticity of transportation costs as well as the price elasticity of demand in a consuming region in a nonlinear way. This is because given the two elasticities, longer distance of delivery raises the markup price in the consumer region, reduces the regional demand

[10] Because $\eta_{ij} = (1-\epsilon_l)u_{ij} + v_{ij}$ and u_{ij} and v_{ij} are orthogonal, $\rho = \frac{(1-\epsilon_l)\sigma_u}{\sqrt{(1-\epsilon_l)^2\sigma_u^2 + \sigma_v^2}} < 0$, given $\epsilon_l > 1$.

for the product, and, as a result, depresses the profitability of delivery from the producing region. The sensitivity of the choice of delivery with respect to distance is then nonlinearly associated with the two elasticities: if the price elasticity of demand is small, the marginal effect of the distance elasticity of transportation costs on the sensitivity of the delivery choice against distance, i.e., $(\epsilon_l - 1)d_{ij}$, is small, and *vice versa.*

The above empirical implications of our model require that to identify the distance elasticity correctly, we jointly estimate the distance elasticity of transportation costs and the price elasticity of demand within a sample-selection model that consists of equations (6) and (7). For this purpose, we conduct a full information maximum likelihood (FIML) estimation of a sample-selection model on which we impose nonlinear constraints. A concern when implementing the FIML estimation, however, is that the disturbance of the selection equation, η_{ij}, might be correlated with endogenous variables $p_i(l)$ and $x_i(l)$ in the RHS of the selection equation. If this is the case, our point estimates of the structural parameters will be biased due to the endogeneity.[11] To take into account the potential endogeneity bias, we further incorporate instrumental variables (IVs) into the FIML estimation as follows.[12] Let $\mathbf{y}_i$ denote a bivariate vector that contains $\ln p_i(l)$ and $\ln x_i(l)$ as its elements: $\mathbf{y}_i \equiv [\ln p_i(l)\ \ln x_i(l)]'$. We assume that vector $\mathbf{y}_i$ is linearly related to vector of exogenous IVs, $\mathbf{s}_i$, up to i.i.d. 2×1 mean zero random vector $\mathbf{e}_i$:

$$\mathbf{y}_i = \Gamma \mathbf{s}_i + \mathbf{e}_i. \tag{9}$$

Endogeneity bias is the case if the error of the selection equation (7), η_{ij}, is correlated with the errors in equation (9), $\mathbf{e}_i$. More specifically, we assume that the 4×1 random vector of disturbances, $[\mathbf{e}_i'\ u_{ij}\ \eta_{ij}]'$, is stochastically governed by a joint normal density with the mean of zero and the 4×4 symmetric positive-definite variance-covariance matrix Ω

$$\Omega = \begin{bmatrix} \Omega_{11} & \varphi_u' & \varphi_\eta' \\ \varphi_u & \sigma_u^2 & \sigma_{u\eta} \\ \varphi_\eta & \sigma_{u\eta} & \sigma_\eta^2 \end{bmatrix}, \tag{10}$$

where Ω_{11} is an 2×2 matrix, φ_u and φ_η are 1×2 row vectors, respectively. Non-zero vector φ_η characterizes the covariances between the disturbances of selection equation (7) and instrument equation (9) that would lead to potential endogeneity bias. Through our analysis, we presume that there is no correlation between the disturbances of price differential equation (6) and instrument equation (9): $\varphi_u = [0\ 0]$.[13]

[11] We greatly appreciate the suggestion of Mike Keane on the case of endogeneity.

[12] Maximum likelihood methods of limited dependent variable models with endogenous explanatory variables are proposed by, for example, Newey (1987), Rivers and Vuong (1988), and Vella and Verbeek (1999) among past studies.

[13] To understand potential endogeneity bias in our inference of γ, observe as in appendix A that the normality of the disturbances implies that the disturbance of the selection equation η_{ij} is given as a linear projection onto the disturbance of the price differential equation and those of the instrument equation, u_{ij} and $\mathbf{e}_i$

$$\eta_{ij} = \varphi_\eta \Omega_{11}^{-1} \mathbf{e}_i + \rho \sigma_u^{-1} u_{ij} + \tilde{\eta}_{ij},$$

where $\tilde{\eta}_{ij}$ is an i.i.d. disturbance with the mean of zero and the variance of $1 - \varphi_\eta \Omega_{11}^{-1} \varphi_\eta' - \rho^2$ and correlated with

Appendix A shows in detail that our structural sample-selection model consisting of equations (6), (7), (9), and (10) provides the log likelihood function

$$\sum_{i,j}(1 - T_{ij}(l))\ln\left[\Phi\left(\lambda_{ij}\right)\right] + \sum_{i,j}T_{ij}(l)\ln\left[\Phi\left(\kappa_{ij}\right)\right] + \sum_{i,j}T_{ij}(l)\ln\left[\phi\left(\omega_{ij}\right)\right] - \sum_{i,j}T_{ij}(l)\ln\sigma_u + \sum_{i,j}\ln\left[f(\mathbf{y}_i|\mathbf{s}_i)\right], \quad (11)$$

where

$$\kappa_{ij} = \frac{\beta - (\epsilon_l - 1)\gamma d_{ij} + [\epsilon_l\ 1]\mathbf{y}_i + b^s_{ij} + \varphi_\eta \Omega_{11}^{-1}(\mathbf{y}_i - \Gamma\mathbf{s}_i) + \rho\sigma_u^{-1}(q_{ij} - \mu - \gamma d_{ij} - b^p_{ij})}{(1 - \varphi_\eta\Omega_{11}^{-1}\varphi'_\eta - \rho^2)^{\frac{1}{2}}},$$

$$\varpi_{ij} = \frac{q_{ij} - \mu - \gamma d_{ij} - b^p_{ij}}{\sigma_u},$$

$$\lambda_{ij} = \frac{\beta - (\epsilon_l - 1)\gamma d_{ij} + [\epsilon_l\ 1]\mathbf{y}_i + b^s_{ij} + \varphi_\eta\Omega_{11}^{-1}(\mathbf{y}_i - \Gamma\mathbf{s}_i)}{(1 - \varphi_\eta\Omega_{11}^{-1}\varphi'_\eta)^{\frac{1}{2}}},$$

$$f(\mathbf{y}_i|\mathbf{s}_i) = (2\pi)^{-1}|\Omega_{11}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{y}_i - \Gamma\mathbf{s}_i)'\Omega_{11}^{-1}(\mathbf{y}_i - \Gamma\mathbf{s}_i)\right\},$$

Here, constant b^p_{ij} and b^s_{ij} control for regional fixed effects in price differential and selection equations (6) and (7), respectively, and ρ is the correlation coefficient between u_{ij} and η_{ij}: $\rho = (1-\epsilon_l)\sigma_u$.[14] We also normalize the selection equation (7) setting the standard deviation of its error term, σ_η, equal to 1.[15] To maximize the log likelihood function (11) conditional on the observations of the delivery index $\{T_{ij}(l)\}$, the price differential $\{q_{ij}(l)\}_{T_{ij}(l)=1}$, the log of distance $\{d_{ij}\}$, the average price and aggregate transaction of product l in consuming regions $\{p_i(l)\}$ and $\{x_i(l)\}$, and instruments $\{\mathbf{s}_i\}$, we take a two-step approach for the computational purpose. In the first step, we regress endogenous variable vector $\mathbf{y}_i$ on IV vector $\mathbf{s}_i$ by OLS and keep the estimates of Γ and Ω_{11}. In the second step, we then insert the OLS estimates of Γ and Ω_{11} into the log likelihood function (11) and maximize the resulting log likelihood function with respect to the rest of the parameters.

---

neither of u_{ij} or $\mathbf{e}_i$. The selection equation then can be rewritten as

$$z_{ij}(l) = \beta - (\epsilon_l - 1)\gamma d_{ij} + [\epsilon_l\ 1]\mathbf{y}_i + \xi_j + \lambda_i + \omega_l - \varrho_{ijl} + \varphi_\eta\Omega_{11}^{-1}(\mathbf{y}_i - \Gamma\mathbf{s}_i) + \rho\sigma_u^{-1}u_{ij} + \tilde{\eta}_{ij}.$$

Notice that if $\varphi_\eta = [0\ 0]$, the above selection equation turns back to the standard one (7).

Suppose then that we mistakenly ignore the true correlation between η_{ij} and $\mathbf{e}_i$, i.e., $\varphi_\eta = [0\ 0]$. There are two ways in which this ignorance affects estimates of structural parameters. First, since the true coefficient on endogenous vector $\mathbf{y}_i$ is $[\epsilon_l\ 1] + \varphi_\eta\Omega_{11}^{-1}$ in the above selection equation, the estimate of parameter vector $[\epsilon_l\ 1]$ could be biased and so could be the estimate of γ. Second, given the true variance of $\tilde{\eta}_{ij}$, $1 - \varphi_\eta\Omega_{11}^{-1}\varphi'_\eta - \rho^2$, we overestimate ρ^2 because the quadratic form $\varphi_\eta\Omega_{11}^{-1}\varphi'_\eta$ always takes a strictly positive value for the true non-zero φ_η due to the positive definiteness of covariance matrix Ω_{11}. This means that our estimate of a negative correlation between η_{ij} and u_{ij} could be biased upwards and so could be the estimated degree of the sample selection bias. This upper biased degree of the sample selection then might lead to an upper biased estimate of γ.

[14] We also include into price differential and selection equations monthly dummies to control for seasonality.

[15] This standard normalization in a sample-selection model makes the correlation between u_{ij} and η_{ij} equal to $(1-\epsilon_l)\sigma_u$. During estimation, we further impose a restriction that the correlation coefficient $(1-\epsilon_l)\sigma_u$ is always less than or equal to 1 in the absolute value.

3. Monte Carlo experiments with a linear economy

In this section, we conduct Monte Carlo experiments based on our model in the last section to understand the following two questions: (i) what bias does the conventional regression exercise without identifying producing regions and ignoring the sample-selection mechanism introduce into our inference on the distance elasticity γ, and (ii) how much can our FIML estimator correct the bias successfully. To implement the experiments, we assign hypothetical values to the structural parameters of our model as follows.

Consider an economy that is geographically separated into 47 regions. Each region is indexed by an integer between 1 and 47, respectively.[16] The distance between regions i and j, D_{ij}, is equal to $100|i-j|$ with the minimum distance of 100 and the maximum of 4600. Each region yields product l under productivity level $a(l)$ that is set equal to 1.00. The parameter of demand function (1), α_l, is common across the regions and equal to 0.75. This number of α means that the price elasticity of demand is 4.00 and the wholesale price is 33.33 % marked up over the corresponding marginal cost. All the producing regions share the same factor cost c_j of 0.55. Each region is also characterized by the aggregate price and transaction, $p_i(l)$ and $x_i(l)$, respectively, both of which we set to 20.00. For simplicity, we ignore the cross-regional variations in the productivity-regional bias ratio $\theta_i(j,l)$ by setting $\delta_i(j,l) = 1.00$ for all pairs of regions i and j. The fixed cost $f_{ij} = \exp(\lambda_i + \lambda_j - v_{ij})$ is specified as follows. We calibrate the sum of the producing and consuming regional fixed effects, $\lambda_i + \lambda_j$, so that, when $\gamma = 0.00$, the probability of product delivery from source to consuming regions is always equal to 0.50. The resulting fixed effect term $\lambda_i + \lambda_j$ then is $(1-\alpha_l)\alpha_l^{\epsilon_l - 1} c_j^{-\epsilon_l} p_i^{\epsilon_l} x_i$ for all (i,j) pairs. The Gaussian random component in the fixed cost, v_{ij}, has the standard deviation of $\sigma_v = 0.30$. We set the constant term of the transportation cost μ to 1.50 and allow for idiosyncratic random variations in the transportation cost setting the standard deviations of the random component of the transportation cost, σ_u, to 0.30. Finally, in our experiments, we admit no possibility of endogeneity bias simply setting $\varphi_\eta = [0\ 0]$.

In our Monte Carlo experiments, we first draw 1000 sets of Gaussian random variables u_{ij} and v_{ij} independently from their distributions. We then calculate price differential $q_{ij}(l)$ and latent variable $z_{ij}(l)$ following equations (6) and (7) under one of the three hypothetical values of γ, 0.00, 0.15, and 0.50. In each Monte Carlo draw with each true value of γ, we then implement four different estimations of γ. The first one is the simple OLS regression of price differential $q_{ij}(l)$ on the log of the distance $\ln d_{ij}$ using the whole synthetic samples regardless of $T_{ij} = 0$ or 1. By construction, this OLS estimator, denoted by $\hat{\gamma}_{\text{whole}}$, is consistent and, hence, should be distributed around the hypothetical true value. The second one is the OLS regression of the price differential $q_{ij}(l)$ on the log of the distance $\ln d_{ij}$ using only the samples that are selected with $T_{ij}(l) = 1$. This second OLS estimator, denoted by $\hat{\gamma}_{\text{OLS}}$, suffers from sample-selection bias. Therefore, we expect to observe that the distribution of $\hat{\gamma}_{\text{OLS}}$ is biased against the true value. The third estimation is with the FIML estimator we introduce in section 3. This estimator, denoted by $\hat{\gamma}_{\text{FIML}}$, should

[16] This assumption of the linear economy might be the most relevant for an island country with a long-narrow arc shape like Japan that consists of 47 prefectures.

correct potential bias due to sample selection. Finally, to explain the fourth estimator, consider the price differential between two consuming regions *without identifying producing regions*, i.e., $\ln \tilde{p}_i(l) - \ln \tilde{p}_k(l)$ for any two consuming regions i and k, where $\tilde{p}_i(l)$ denotes the price of product l in consuming region i. The OLS estimator conventional in the literature of the absolute LOP, which is denoted by $\hat{\gamma}_{\text{conv}}$, then is constructed by regressing the absolute value of the price differential between consuming regions i and k, $|\ln \tilde{p}_i(l) - \ln \tilde{p}_k(l)|$, on the log of the corresponding distance $\ln d_{ik}$.[17] Comparing the distribution of $\hat{\gamma}_{\text{conv}}$ with that of $\hat{\gamma}_{\text{whole}}$, we can understand the degree of bias the conventional regression exercise suffers from on the inference of γ.

We first observe how the size of γ affects delivery choice. The left, middle, and right windows of Figure 1 depict the contour plots of the probabilities of delivery from producing regions to consumption regions for the cases of $\gamma = 0.50$, 0.15, and 0.00, respectively. In each window, the contour lines represent the combinations of the producing and consuming regions that have an identical delivery probability. The left window shows that with the large distance elasticity of $\gamma = 0.50$, the product delivery is profitable only locally. This is obvious from the fact that all contour lines are parallel to the 45 degree line and the equiprobability bands, which are constructed by two contour lines with the same probability, are very narrow and always include the 45 degree line. This shape of the contour plot implies that the product delivery occurs only to consuming regions neighboring source regions closely. The middle window then exhibits that the equiprobability bands become much wider with the smaller distance elasticity of $\gamma = 0.15$. Hence, in this model, a larger distance elasticity creates geographical clustering of products based on different source regions. This is clearer if we set $\gamma = 0.00$. As displayed in the right window, the equiprobability lines with the delivery probability of 0.50 are almost randomly placed over the whole window: the product delivery occurs with the 50 % chance even between the producing and consuming regions that are farthest apart each other.

Figure 2 depicts simulated price differentials against the corresponding logs of distances. The first, second, and third rows of the figure are for the cases with $\gamma = 0.50$, 0.15, and 0.00, respectively. In each row, the first column reports the simulated samples conditional on the choice of delivery $T_{ij} = 1$, while the second column plots the whole samples regardless of delivery choice $T_{ij} = 0$, or 1. The two windows in the first row reveal severe data truncation under $\gamma = 0.50$. Although the whole samples of the simulated price differentials have a clear positive association with the logs of distances, the underlying selection mechanism is so strong that the observed samples are concentrated only on local areas surrounding source regions with short range delivery. The association of the observed price differentials with the logs of distances then becomes quite vague. The second and third rows prove that the sample selection turns out to be weaker when γ becomes smaller to 0.15 and 0.00.

[17] For each Monte Carlo draw, the price of product l that is sampled in consuming region i, $\tilde{p}_i(l)$, is constructed as follows. For each consuming region i, we obtain the set of the truncated prices that are delivered from producing regions $S_i(l) = \{p_i(j, l) | j \in B_i(l)\}$. This set $S_i(l)$ includes the prices of product l that can be sampled as the representative price in consuming region i, $\tilde{p}_i(l)$. We uniformly draw 100 prices from this set $S_i(l)$ and take the average over them to construct $\tilde{p}_i(l)$.

Figure 3 reports the densities of the four different estimators of γ that are nonparametrically smoothed with the Epanechnikov kernel. The first row corresponds to the case with $\gamma = 0.50$; the second the case with $\gamma = 0.15$; and the third the case with $\gamma = 0.00$. The first column plots the smoothed densities of $\hat{\gamma}_{\text{whole}}$; the second $\hat{\gamma}_{\text{OLS}}$; the third $\hat{\gamma}_{\text{FIML}}$; and the fourth $\hat{\gamma}_{\text{conv}}$. The three windows in the first column show that $\hat{\gamma}_{\text{whole}}$ is consistent and distributed around the underlying true value. The three windows in the second column, however, uncover that $\hat{\gamma}_{\text{OLS}}$ is subject to severe downward bias. On the one hand, as displayed in the first and second rows in the second column, $\hat{\gamma}_{\text{OLS}}$ is distributed far left from the corresponding true value when γ is set to either 0.50 or 0.15. On the other hand, as shown in the third row of the second column, $\hat{\gamma}_{\text{OLS}}$ is consistent and distributed around the true value if $\gamma = 0.00$. Therefore, a positive distance elasticity generates the data truncation that causes the OLS estimates to be biased downwards. The three windows in the third column clearly reveal that $\hat{\gamma}_{\text{FIML}}$ is consistent and distributed around the underlying true value. The most striking fact from the three windows in the fourth column is that $\hat{\gamma}_{\text{conv}}$ performs the worst among the other estimators. In the first and second rows for the cases of $\gamma = 0.50$ and 0.15, $\hat{\gamma}_{\text{conv}}$ is distributed with the means of 0.019 and 0.003, respectively, and even far left from the corresponding density of $\hat{\gamma}_{\text{OLS}}$. This is the evidence that the conventional regression exercise without identifying producing regions suffers from the worst under-bias toward an inference on γ among all the other estimators.

The Monte Carlo experiments of this section, therefore, confirm the necessity of identifying producing regions and taking into account the sample-selection mechanism to draw a correct inference on the distance elasticity of transportation costs. The proposed FIML estimator can correctly identify the true values of the distance elasticity with synthetic data generated from our structural model.

4. Data and descriptive statistics

In this paper, we investigate a daily data set of the wholesale prices of agricultural products in Japan — the Daily Wholesale Market Information of Fresh Vegetables and Fruits. The details of our data set are provided in appendix B. This daily market survey covers the wholesale prices of 120 different fruits and vegetables. Each agricultural product is further categorized by different varieties, sizes, grades, as well as producing prefectures. Hence, for example, the data set reports the wholesale prices at 6 different wholesale markets of the "Dansyaku (Irish Cobbler equivalent)" variety of potato of size "L" with grade "Syu (excellent)" that was produced in "Hokkaido" prefecture on September 7, 2007. This high degree of categorization is ideal for our purpose of approaching the absolute LOP rigorously and inferring transportation costs precisely because the LOP requires to identify identical goods as its theoretical premise at the first place. This daily market survey has been recorded since 1976. In this paper, we scrutinize the 2007 survey that reports the market transactions on 274 market opening days.

Price differential $q_{ij}(l)$ is constructed by subtracting the wholesale price in producing prefecture j, $p_j(j,l)$, from that in consuming prefecture i, $p_i(j,l)$.[18] We set $T_{ij}(l) = 1$ for pair (i,j) if the sample of $q_{ij}(l)$ is available.[19] The geographical distance between prefectural pair (i,j) is approximated by that between the prefectural head offices placed in the prefectural capital cities. Taking the log of the geographical distance yields variable d_{ij}. Our data set provides the daily aggregate transaction level of product l in consuming region i, $x_i(l)$.[20] We are unable to obtain daily data of the aggregate price of product l in the consuming region i, $p_i(l)$. Hence, we use as a proxy of $p_i(l)$ the monthly data of the retail price of product l. Moreover, to control for daily variations in producing and consuming prefectures, we include into selection equation (7) daily temperature data in both of the two prefectures as other explanatory variables. This inclusion of the regional temperatures as determinants of delivery choice comes from our prior belief that the temperatures in producing and consuming regions are important factors for productions of and demands for agricultural products. Finally, as valid IVs, we use the monthly variations of the numbers of regular employees and scheduled cash earnings in each prefecture besides the monthly and consuming-region dummies as well.

We focus our exercise on eight selected vegetables: cabbage, carrot, Chinese cabbage (c-cabbage, hereafter), lettuce, shiitake-mushroom (s-mushroom, hereafter), spinach, potato, and welsh onion. Table 1 summarizes several descriptive statistics for these products. Panel (a) of the table shows that each product is highly categorized by product varieties, sizes, and grades. The number of distinct product entries is quite large; 1,207 for cabbage; 1,186 for carrot; 1,001 for c-cabbage; 903 for lettuce; 1,423 for potato; 909 for s-mushroom; 551 for spinach; and 1,115 for welsh onion.

For each product entry l, we count the number of delivery $T_{ij}(l) = 1$ and non-delivery $T_{ij} = 0$ only for the dates on which the product entry is indeed traded at the wholesale market in producing prefecture j. We identify product delivery $T_{ij}(l) = 1$ if the data reports that the source prefecture of product entry l sold in consuming region i is region j.[21] The first row of panel (b) of the table reports that the total number of both delivery and non-delivery cases all over the product entries is almost beyond 180,000 for each vegetable. This is the number of observations

[18] For some products, we cannot find the wholesale prices in producing prefectures, although we can observe those prices in consuming prefectures. In this case, because we cannot construct the price differentials between producing and consuming prefectures, we drop the data of these product entries from our investigation.

[19] We also set $T_{jj}(l) = 1$ whenever we can observe $p_j(j,l)$. We consider this case that product l is delivered from the producer to the wholesale market in the producing prefecture. We attach the minimum distance of 10.00km to the samples with $T_{jj}(l) = 1$ to avoid taking the log of zero distance.

[20] Whenever the data set reports that $x_i(l) = 0$, we interpolate $x_i(l)$ by a very small number of 0.00001 to avoid taking the log of zero.

[21] A problem of this identification would be that we cannot eliminate the possibility of product transfer: a product yielded in a source region is once delivered to a consuming region and then transferred to another consuming region. If this is the dominant case in our data set, our inference on distance effects might be biased. However, according to the Ministry of Agriculture, Forestry, and Fishery, the amount of product transfers across the wholesale markets is very small relative to the total amount of wholesale transactions in Japan. For example, in 2007, the ratio of product transfers to the total wholesale transactions is 4.8 % for cabbage; 6.5 % for carrot; 4.9 % for c-cabbage; 6.3 % for lettuce; 6.0 % for potato; 3.3 % for s-mushroom; 4.1 % for spinach; and 3.9 % for welsh onion. This means that almost all products in our data are directly delivered from source regions to consuming regions as final destinations.

for our FIML estimation. Out of the total number of delivery and non-delivery cases, the number of delivery cases is relatively small, as exhibited in the second row of panel (b): it is about 10,000 for each vegetable. Our data set, hence, indicates that product delivery is quite limited.[22] The third row of panel (b) shows that the mean distance from producing to consuming prefectures over all delivery and non-delivery cases is about 6.00 in the logarithmic term (or 403.428 km) and almost identical across the vegetables. The fourth row of panel (b), on the other hand, conveys that the mean distance over delivery cases only is much shorter depending on vegetables with the minimum number of 2.691 (14.746km) for s-mushroom and the maximum of 4.339 (76.630km) for potato. Product delivery, therefore, is localized and concentrated on local areas neighboring around producing prefectures.

Figure 4 also confirms the locality of product delivery graphically. As in those of Figure 1, each window of Figure 4 depicts as a contour plot the data frequencies of product delivery from producing to consuming prefectures that are calculated over all product entries on all traded dates. The horizontal axis represents producing prefectures and the vertical axis consuming prefectures. The order of prefectures reflects the geographical positions of the prefectures from the most north prefecture, Hokkaido, to the most south one, Okinawa. Therefore, two prefectures that are indexed by close integers are indeed geographically close to each other. Then, the brighter the contour line is, the higher the probability of product delivery is. The figure then uncovers three facts. First, each vegetable has several dominant producing prefectures that are characterized by vertical contour lines. This means that these main producing prefectures deliver their products to not only nearby prefectures but also other remote prefectures. Second, the data frequencies of product delivery of the main producing prefectures are decreasing in distance. Therefore, even dominant producers do not deliver their products to consuming prefectures farthest away.[23] Third, the contour lines for other minor producing prefectures are concentrated on the 45 degree line. The product delivery of these relatively minor producing prefectures, thus, is highly localized.

The locality of product delivery that Table 1 and Figure 4 unmask together brings us two important implications. First, as observed by Broda and Weinstein (2008) in their barcode data of retail products, agricultural products in our data set are segmented and clustered geographically. Even in the same vegetable category, products that are sold in two distinct prefectures far away from one another come from different sources and the corresponding wholesale prices might be affected by regional factors idiosyncratic to the product origins. Price differentials across consuming regions that are generated by such idiosyncratic factors cannot be attributed to transportation costs. Hence, given the observed high degree of regional product clustering, it is crucial to scrutinize price differentials of a product that shares a source region in order to infer the role transportation costs play in absolute LOP violations correctly. Second, drawing an inference on transportation costs only from observed price differentials might be subject to a serious sample-selection bias, as we repeatedly claim in this paper.

[22] This observation echoes the findings of recent researches on the extent of firms' participation to export. For instance, Bernard and Jensen (2004) report that only a small portion of the U.S. manufacturing plants export their products.

[23] Exception is observed in the first producing prefecture, Hokkaido, in the cases of carrot and potato.

The mean of the observed log price differential is reported on the first row of panel (c) of Table 1. The positive numbers reported in the first row imply that wholesale prices in consuming prefectures are on average higher by between 0.3 % and 8.1 % than those of producing prefectures. This observation is suggestive for an important role of transportation costs in price differentials, as predicted by equation (6). The corresponding standard deviation of the observed log price differential, which is displayed on the second row of panel (c), is around 20 %. Our data set, thus, shows the almost same degree of absolute LOP violations as observed in past studies (e.g., Crucini et al. 2005, and Broda and Weinstein 2008), even after identifying source regions of products. We also conduct an OLS regression of the observed price differential on the corresponding log distance and constant for each vegetable. The resulting OLS estimates of the coefficient on the log distance, $\hat{\gamma}_{\mathrm{OLS}}$, are shown in the third row of panel (c), which are accompanied by the standard errors. All the point estimates are positive with values between the minimum of 0.007 and the maximum of 0.051 at any conventional statistical significance levels. This range of the estimated distance elasticity of price differential is consistent with the estimates past studies commonly find using different data sets such as in Engel et al. (2005), Broda and Weinstein (2008), and Inanc and Zachariadis (2009).

5. Results

5.1. Results of FIML estimation

Table 2 summarizes the results of the FIML estimation based on the log likelihood (11). The first and second rows of panel(a) of the table shows that the distance elasticity of transportation costs, $\hat{\gamma}_{\mathrm{FIML}}$, is estimated positive and statistically significant for each vegetable. The outstanding fact this row tells us is the large size of the FIML estimates: the mean (over the eight vegetables) of the estimated distance elasticity is 0.238 with the minimum of 0.210 for cabbage and the maximum of 0.325 for lettuce. According to equation (6), the price differential of a product between consuming and producing regions rises by about 24 % in response to the 100 % stretch in delivery distance when ignoring selection mechanism (7). Compared with the small size of the OLS estimate of the distance elasticity, which is reported between 0.008 and 0.051 in the last row of Table 1, this large size of the FIML estimate implies that the OLS estimate is biased downwards seriously due to the underlying data truncation.

As discussed in section 2, the strength of the observed under bias tightly connects with the price elasticity of demand, ϵ_l. As reported in the third and fourth rows of panel (a) of Table 2, ϵ_l is estimated sensibly and significantly: the mean of the point estimate of ϵ_l is 3.132 over the eight vegetables. Combining with the large estimate of the distance elasticity of transportation costs, the estimated price elasticity of demand implies that the probability of product delivery from producing to consuming prefectures depends negatively as well as sensitively on delivery distance. The point estimate of the correlation coefficient between the unobserved disturbances of price differential equation (6) and selection equation (7), ρ, then provides empirical evidence that sample-selection bias does matter. As displayed in the fifth and sixth rows of panel(a) of Table 2, ρ is estimated negative with high statistical significance: the mean of the estimates of ρ over the eight vegetables is -0.536 with the minimum of -0.684 for welsh onion and the maximum of -0.278 for potato. This highly negative correlation between the unobserved disturbances in the two equations is the fundamental source for the under bias in the OLS estimate of the distance elasticity in the price differential equation, as shown in equation (8).

In summary, our FIML estimates of the sample-selection model reveal dual roles geographical distance plays in regional price differentials. Distance creates a large price gap between consuming and producing regions. At the same time, distance significantly affects choice of product delivery from the latter to the former regions. As a result, price differentials are not randomly sampled and, especially, their observations are concentrated on local areas neighboring producing regions. This concentration of the observations within relatively short distance conceals the actual size of the underlying distance elasticity of transportation costs and makes the OLS estimates biased downwards.

5.2. Model validation through diagnosis checks

The above FIML estimates of the three structural parameters depend on the identification provided by our structural sample-selection model. Therefore, the relevance of the estimates relies on the empirical validity of our model. As model validation, we conduct diagnosis checks of our model with respect to two important aspects of the actual data: the pattern of product delivery and the association of price differentials with delivery distances.

If our sample-selection model is reliable, it should explain the pattern of product delivery, $T_{ij}(l)$, that is actually observed in our data. To check the ability of our model to mimic the product delivery pattern in the data, we calculate the percents correctly predicted (PCPs) measures for $T_{ij}(l) = 0$ or 1.[24] To construct the PCPs, we calculate the predicted conditional probabilities of $T_{ij}(l) = 0$ and $T_{ij}(l) = 1$ on the observables, $\hat{P}(T_{ij}(l) = 0|.)$ and $\hat{P}(T_{ij}(l) = 1|.)$, respectively. Then if $\hat{P}(T_{ij}(l) = 0|.) > 0.5$, we recognize that our model predicts $T_{ij}(l) = 0$, while if $\hat{P}(T_{ij}(l) = 1|.) > 0.5$, it predicts $T_{ij}(l) = 1$. The PCP for $T_{ij}(l) = 0$ (or 1) then is calculated as the percentage of the total number of the observations of $T_{ij}(l) = 0$ (or 1) that are accompanied by $\hat{P}(T_{ij}(l) = 0|.) > 0.5$ (or $\hat{P}(T_{ij}(l) = 1|.) > 0.5$). The PCP for either $T_{ij}(l) = 0$ or 1 is simply derived as a weighted average of the PCPs for $T_{ij}(l) = 0$ and 1.

The results of the PCPs are summarized in the first, second, and third rows of panel (b) of Table 2. As shown in the first row, our sample-selection model yields high PCPs around 0.990 for either $T_{ij}(l) = 0$ or 1 for all the vegetables. This means that the model is fairly successful in replicating the observed pattern of product delivery overall. In particular, as implied by the PCPs reported in the second and third rows of panel (b), the model's ability to replicate no delivery choice $T_{ij}(l) = 0$ is better than that to replicate delivery choice $T_{ij}(l) = 1$. On the one hand, the high PCPs for $T_{ij}(l) = 0$ around 0.990 suggest the model's outstanding predictive ability of no delivery choice. On the other hand, the PCPs for $T_{ij}(l) = 1$ are lower than those of no delivery choice with the mean of 0.820. The model does a good job in predicting the delivery choice especially for some

[24] Wooldridge (2002) discusses the PCP for model validation of probit models.

vegetables such as s-mushroom, spinach, and welsh onion.[25] We confirm through this diagnosis criterion that the model's predictive ability for the pattern of product delivery is remarkable.

The second diagnosis criterion is data association of price differentials with delivery distances. As observed in the last row of Table 1, the OLS regression of the former on the latter in actual data yields the estimate of the distance elasticity, $\hat{\gamma}_{\mathrm{OLS}}$, around 3% on average. The question we ask here is if our sample-selection model predicts this size of the OLS estimate or not.

To do this diagnosis check, we derive the prediction of the model on price differentials following equation (8). Each window of Figure 5 plots the resulting predicted price differentials (blue dots) as well as the data counterparts (gray crosses) against the corresponding log distances for each vegetable. The blue dots are distributed inside the cloud made of the gray crosses in all the windows except the case of carrot. This means that our model successfully predicts the data association of price differentials with distances overall, although the actual data show us a much sparse joint distribution between the two variables. The fourth row of panel (b) of Table 2 reports the OLS estimate of regressing the predicted price differentials on the corresponding distances. For comparison, we also display in the last row of the panel the OLS estimate with the actual data that has been already reported in Table 1. The model's prediction on the OLS estimate is close to but slightly larger than its actual data counterpart: the cross-vegetable average of the predicted OLS estimate is 0.063 whereas that with the actual data is 0.033. It is important, however, to remember that the distance elasticity of transportation costs of our model is estimated 0.238 by FIML. What is striking is that the sample-selection model with such a large distance elasticity of transportation costs indeed mimics such a small size of the OLS estimate. In this sense, we conclude that our model successfully passes the second diagnosis check, although we fully understand that there is still an unexplained gap between the model's prediction and the actual data with respect to the observed joint distribution of price differentials and distances.

6. Conclusion

As claimed by Anderson and van Wincoop (2004) in their introduction, the "death of distance" is indeed exaggerated even in the literature of regional price dispersions. In this paper, we try to revive and rejuvenate transportation costs, which are measured by geographical distance, as a main driving force of absolute LOP violations. To do so, we identify producing regions and take into account sample selectivity due to the underlying choice of product delivery in our data set of daily wholesale prices of agricultural products in Japan. After estimating our structural sample-selection model by FIML using data of price differentials and delivery patterns, we find that the estimated distance elasticity of transportation costs is so large that this paper successfully fills the reported huge gap between the two fields of international economics — international finance and trade — with respect to inferences on distance effects.

[25] The main reason for the model's slightly lower predictive performance for carrot and potato is understandable. As observed in Figure 4, the main producing prefecture of these two vegetables, Hokkaido, delivers its products to all other prefecture regardless of delivery distance. This data aspect is hard to explain by our simple structural model.

Although this paper intensively scrutinizes data aspects of agricultural products, the main arguments in this paper are also applicable to other products. For instance, identifying in which plant products are manufactured and taking into account the underlying location choice of plants could be crucial for correct inferences on the role of transportation costs in regional price dispersions for manufactured non-perishable products. This is because, if transportation costs are expensive, firms might decide to locate their plants close to consuming markets to economize transportation costs. In this case, because product delivery becomes limited around local areas neighboring plants, observations of the price of a product sharing an identical plant will be truncated. The resulting sample selectivity then leads to a biased inference on the role of transportation costs in regional price dispersions as in our exercise. This conjecture suggests more intensive use of plant level data in the absolute LOP literature.

Finally, it is worth noting a caveat against our inferences that depend on the implications of the highly stylized structural model. An obvious limitation of our structural inferences stems from the model's assumption of monopolistically competitive firms facing regional demand functions with a constant elasticity. To figure out historical movements of the relative PPP of the United States, a recent paper by Atkeson and Burstein (2008) emphasizes the importance of richer market structures that make price elasticity of demand and markup variable in market shares. If this is the case, the delivery choice of a source region to its wholesale market should have a non-negligible impact on the price elasticities of demand for products from other source regions because the market shares of other source regions change. Given transportation costs, this change in the sensitivity of demand then might affect the product delivery choices of the other source regions. This mechanism potentially makes our inferences on distance effects biased. We want to leave this extension to future research.

Appendix A. Derivation of log likelihood function (11)

Given the sample-selection model with IVs that consists of equations (6), (7), (9), and (10), we derive the corresponding likelihood $f(q_{ij}, T_{ij}, \mathbf{y}_i | d_{ij}, \mathbf{s}_i)$. Since q_{ij} is observable only when $T_{ij} = 1$, we can factorize the likelihood as follows:

$$\begin{aligned} f(q_{ij}, T_{ij} = 0, \mathbf{y}_i | d_{ij}, \mathbf{s}_i) &= f(T_{ij} = 0, \mathbf{y}_i | d_{ij}, \mathbf{s}_i), \\ &= \mathbf{P}(T_{ij} = 0 | \mathbf{y}_i d_{ij}, \mathbf{s}_i) f(\mathbf{y}_i | d_{ij}, \mathbf{s}_i), \end{aligned}$$

and

$$\begin{aligned} f(q_{ij}, T_{ij} = 1, \mathbf{y}_i | d_{ij}, \mathbf{s}_i) &= f(q_{ij} | T_{ij} = 1, \mathbf{y}_i, d_{ij}, \mathbf{s}_i) f(T_{ij} = 1, \mathbf{y}_i | d_{ij}, \mathbf{s}_i), \\ &= \frac{P(T_{ij} = 1 | q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i) f(q_{ij} | \mathbf{y}_i, d_{ij}, \mathbf{s}_i)}{\mathbf{P}(T_{ij} = 1 | \mathbf{y}_i, d_{ij}, \mathbf{s}_i)} f(T_{ij} = 1, \mathbf{y}_i | d_{ij}, \mathbf{s}_i), \\ &= \mathbf{P}(T_{ij} = 1 | q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i) f(q_{ij} | \mathbf{y}_i, d_{ij}, \mathbf{s}_i) f(\mathbf{y}_i | d_{ij}, \mathbf{s}_i). \end{aligned}$$

Hence, our task is to characterize conditional densities $f(\mathbf{y}_i | d_{ij}, \mathbf{s}_i)$, $f(q_{ij} | \mathbf{y}_i, d_{ij}, \mathbf{s}_i)$, $\mathbf{P}(T_{ij} = 0 | \mathbf{y}_i, d_{ij}, \mathbf{s}_i)$, and $\mathbf{P}(T_{ij} = 1 | q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i)$.

Since endogenous explanatory vector $\mathbf{y}_i$ is always observable regardless of the value of T_{ij}, conditional distribution $f(\mathbf{y}_i | d_{ij}, \mathbf{s}_i)$ is simply characterized by equation (9) as a Gaussian joint density with the mean of $\Gamma \mathbf{s}_i$ and the variance-covariance matrix of Ω_{11}:

$$f(\mathbf{y}_i|d_{ij}, \mathbf{s}_i) = f(\mathbf{y}_i|\mathbf{s}_i) = (2\pi)^{-1}|\Omega_{11}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \Gamma\mathbf{s}_i)'\Omega_{11}^{-1}(\mathbf{y}_i - \Gamma\mathbf{s}_i)\right\}.$$

Characterizing the rest of the conditional densities requires us to figure out the conditional densities $f(\eta_{ij}|\mathbf{e}_i)$, $f(\eta_{ij}|\mathbf{e}_i, u_{ij})$, and $f(u_{ij}|\mathbf{e}_i)$, respectively. To do so, we conduct the triangular factorization of the variance-covariance matrix Ω to obtain $\Omega = ADA'$, in which

$$A = \begin{bmatrix} I_n & 0 & 0 \\ \varphi_u \Omega_{11}^{-1} & 1 & 0 \\ \varphi_\eta \Omega_{11}^{-1} & H_{32}H_{22}^{-1} & 1 \end{bmatrix}, \quad \text{and} \quad D = \begin{bmatrix} \Omega_{11} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} \\ \mathbf{0}_{1,2} & H_{22} & 0 \\ \mathbf{0}_{1,2} & 0 & H_{33} - H_{32}H_{22}^{-1}H_{23} \end{bmatrix},$$

where $H_{22} = \sigma_u^2 - \varphi_u \Omega_{11}^{-1}\varphi_u'$, $H_{32} = \sigma_{u\eta} - \varphi_u \Omega_{11}^{-1}\varphi_\eta'$, $H_{33} = \sigma_\eta^2 - \varphi_\eta \Omega_{11}^{-1}\varphi_\eta'$, and $H_{32} = H_{23}$, as shown in Hamilton (1994). Define a new random vector $\tilde{\epsilon}_{ij} = [\tilde{\mathbf{e}}_i{}' \ \tilde{u}_{ij} \ \tilde{\eta}_{ij}]' = A^{-1}\epsilon_{ij}$. The above triangular factorization implies that new vector $\tilde{\epsilon}_{ij}$ is normally distributed with the mean of zero and the diagonal variance-covariance matrix of D. Then, by construction, we can obtain the following system of equations.

$$\begin{aligned} \mathbf{e}_i &= \tilde{\mathbf{e}}_i, \\ u_{ij} &= \varphi_u \Omega_{11}^{-1}\tilde{\mathbf{e}}_i + \tilde{u}_{ij}, \\ \eta_{ij} &= \varphi_\eta \Omega_{11}^{-1}\tilde{\mathbf{e}}_i + H_{32}H_{22}^{-1}\tilde{u}_{ij} + \tilde{\eta}_{ij}. \end{aligned}$$

To derive the conditional density $f(\eta_{ij}|\mathbf{e}_i)$, define a new random variable $\tilde{a}_{ij} = H_{32}H_{22}^{-1}\tilde{u}_{ij} + \tilde{\eta}_{ij}$. Notice that random variable $\tilde{a}_{ij}$ is normally distributed with the mean of zero and the variance of H_{33}.[26] We can obtain

$$\eta_{ij} = \varphi_\eta \Omega_{11}^{-1}\mathbf{e}_i + \tilde{a}_{ij}$$

Since $\tilde{a}_{ij}$ is orthogonal to $\mathbf{e}_i = \tilde{\mathbf{e}}_i$, the above equation implies that the conditional distribution of η_{ij} on $\mathbf{e}_i$ is normal with the mean of $\varphi_\eta \Omega_{11}^{-1}\mathbf{e}_i$ and the variance of H_{33}: $\eta_{ij} \sim N(\varphi_\eta \Omega_{11}^{-1}\mathbf{e}_i, H_{33})$. Similarly, we can characterize conditional densities $f(\eta_{ij}|\mathbf{e}_i, u_{ij})$ and $f(u_{ij}|\mathbf{e}_i)$ by the corresponding conditional distributions $\eta_{ij} \sim N(\varphi_\eta \Omega_{11}^{-1}\mathbf{e}_i + H_{32}H_{22}^{-1}(u_{ij} - \varphi_u \Omega_{11}^{-1}\mathbf{e}_i),\ H_{33} - H_{32}H_{22}^{-1}H_{23})$ and $u_{ij} \sim N(\varphi_u \Omega_{11}^{-1}\mathbf{e}_i,\ H_{22})$, respectively.

Conditional mass probability $\mathbf{P}(T_{ij} = 0|\mathbf{y}_i, d_{ij}, \mathbf{s}_i)$, then, is

$$\begin{aligned} &\mathbf{P}(T_{ij} = 0|\mathbf{y}_i, d_{ij}, \mathbf{s}_i) \\ &= \mathbf{P}(\eta_{ij} \leq -\beta + (\epsilon - 1)\gamma d_{ij} - [\epsilon\ 1]\mathbf{y}_i - b_{ij}|\mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\ &= \mathbf{P}(\tilde{a}_{ij} \leq -\beta + (\epsilon - 1)\gamma d_{ij} - [\epsilon\ 1]\mathbf{y}_i - b_{ij} - \varphi_\eta \Omega_{11}^{-1}\mathbf{e}_i|\mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\ &= 1 - \Phi(\lambda_{ij}) \end{aligned}$$

where $\lambda_{ij} = \frac{\beta - (\epsilon-1)\gamma d_{ij} + [\epsilon\ 1]\mathbf{y}_i + b_{ij} + \varphi_\eta \Omega_{11}^{-1}(\mathbf{y}_i - \Gamma\mathbf{s}_i)}{(\sigma_\eta^2 - \varphi_\eta \Omega_{11}^{-1}\varphi_\eta')^{\frac{1}{2}}}$. Conditional mass probability $\mathbf{P}(T_{ij} = 1|q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i)$ is

$$\begin{aligned} &\mathbf{P}(T_{ij} = 1|q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i) \\ &= \mathbf{P}(\eta_{ij} > -\beta + (\epsilon - 1)\gamma d_{ij} - [\epsilon\ 1]\mathbf{y}_i - b_{ij}|q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\ &= \mathbf{P}(\tilde{\eta}_{ij} > -\beta + (\epsilon - 1)\gamma d_{ij} - [\epsilon\ 1]\mathbf{y}_i - b_{ij} - \varphi_\eta \Omega_{11}^{-1}\mathbf{e}_i - H_{32}H_{22}^{-1}(u_{ij} - \varphi_u \Omega_{11}^{-1}\mathbf{e}_i)|q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\ &= \Phi(\kappa_{ij}), \end{aligned}$$

[26] The variance of $\tilde{a}_{ij}$ is $(H_{32}H_{22}^{-1})^2\sigma_{\tilde{u}}^2 + \sigma_{\tilde{\eta}}^2 = (H_{32}H_{22}^{-1})^2 H_{22} + H_{33} - H_{32}H_{22}^{-1}H_{23} = H_{33}$.

where $\kappa_{ij} = \frac{\beta-(\epsilon-1)\gamma d_{ij}+[\epsilon\ 1]\mathbf{y}_i+b_{ij}+(\varphi_\eta-H_{32}H_{22}^{-1}\varphi_u)\Omega_{11}^{-1}(\mathbf{y}_i-\Gamma\mathbf{s}_i)+H_{32}H_{22}^{-1}(q_{ij}-\mu-\gamma d_{ij})}{(H_{33}-H_{32}H_{22}^{-1}H_{23})^{\frac{1}{2}}}$. Finally, to characterize conditional density $f(q_{ij}|\mathbf{y}_i, d_{ij}, \mathbf{s}_i)$, consider the conditional distribution

$$
\begin{aligned}
F_{q_{ij}|\mathbf{y}_i,d_{ij},\mathbf{s}_i}(q^*) &= \mathbf{P}(q_{ij} < q^*|\mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\
&= \mathbf{P}(u_{ij} < q^* - \mu - \gamma d_{ij}|\mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\
&= \mathbf{P}(\tilde{u}_{ij} < q^* - \mu - \gamma d_{ij} - \varphi_u \Omega_{11}^{-1}\mathbf{e}_i|\mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\
&= \Phi(\varpi_{ij})
\end{aligned}
$$

where $\varpi_{ij} = \frac{q^*-\mu-\gamma d_{ij}-\varphi_u\Omega_{11}^{-1}(\mathbf{y}_i-\Gamma\mathbf{s}_i)}{H_{22}^{\frac{1}{2}}}$. Therefore, we can obtain conditional density $f(q_{ij}|\mathbf{y}_i, d_{ij}, \mathbf{s}_i)$ by taking the derivative of $F_{q_{ij}|\mathbf{y}_i,d_{ij},\mathbf{s}_i}(q^*)$ with respect to q^*: $f(q^*|\mathbf{y}_i, d_{ij}, \mathbf{s}_i) = H_{22}^{-\frac{1}{2}}\phi(\varpi_{ij})$.

Summarizing the above characterization of the conditional densities, we can derive the likelihood $f(q_{ij}, T_{ij}, \mathbf{y}_i|d_{ij}, \mathbf{s}_i)$ as

$$f(q_{ij}, T_{ij}, \mathbf{y}_i|d_{ij}, \mathbf{s}_i) = [\Phi(\kappa_{ij})H_{22}^{-\frac{1}{2}}\phi(\varpi_{ij})]^{T_{ij}}[1 - \Phi(\lambda_{ij})]^{1-T_{ij}} f(\mathbf{y}_i|d_{ij}, \mathbf{s}_i).$$

We normalize the standard deviation of the disturbance of the selection equation to one: $\sigma_\eta = 1$. Under the case that $\varphi_u = [0,\ 0]$, i.e, u_{ij} is orthogonal to each of the elements of $\mathbf{e}_i$, the likelihood turns out to be equation (11) in the log form.

Appendix B. Data sources

Wholesale prices:

The data source of wholesale prices is the Daily Wholesale Market Information of Fresh Vegetables and Fruits ("*Seikabutsu Hinmokubetsu Shikyo Joho*" in Japanese). The data set is distributed by the Center of Fresh Food Market Information Service ("*Zenkoku Seisen Syokuryohin Ryutsu Joho Senta*" with the URL: http://www2s.biglobe.ne.jp/ fains/index.html). All contents in the data set are surveyed by the Ministry of Agriculture, Forestry, and Fishery (MAFF) for almost all transactions at 55 wholesale markets officially opened and operated in the 47 prefectures in Japan on a daily basis.

The data file contains information on name of product, market prices, name of production cite, name of market place, and product characteristics. The price reported has three forms: the highest price, the mode price, and the lowest price. Most markets record all three prices, but several markets report only the highest and the lowest prices or only the mode price. Thus, we construct our price variable by averaging these price variables. We use the mode price when only the mode price is available. The transaction unit of each product is also reported. To obtain same unit for each product, we divide the price by the number of the transaction unit.

We need to control for product characteristics to examine prices between production site and market place. Thus, we construct same category product by using product characteristics and production cite. The product characteristics are: brand name, size of products, and grade of products. The size is coded by categorical variables, such as large, medium, and small. The grade is also measured by the categorical variables, such as A, B or superior.[27] Because prices depend on detailed characteristics, we take each combination of characteristics to have the same product.

The coverage of vegetables traded through the central wholesale markets is substantial in Japan. While nowadays large supermarket and restaurant chains can not only directly purchase agricultural products from producers but also directly import from foreign producers, the share of agricultural products covered

[27] For example, according to the guideline document of Yamanashi prefecture, spinach is classified as grade A under the following conditions: it is of one type and no mixture of types affects the appearance; it is clean, trimmed, and free from decay and damages by insects. Otherwise, it is ranked as B.

by these markets in the whole vegetable transactions is still more than 75 % in Japan in 2006, according to MAFF. Thus, our data set enable us to approach the population moments of transportation costs.

Geographical distance:

The data of distance is provided by the Geographical Survey Institute (GSI) of the Government of Japan. The data is publicly available in the GSI website (http://www.gsi.go.jp/kokujyoho/kenchokan.html).

Retail prices :

The monthly data of the retail price of product l is reported in the Retail Price Survey ("*Kouri Bukka Tokei Chosa*" with the URL: http://www.stat.go.jp/data/kouri/index.htm) the Ministry of Internal Affairs and Communication conducts.

Daily temperatures:

The daily temperature data are reported by the Japan Meteorological Agency. We download the data from the website: http://www.data.jma.go.jp/obd/stats/etrn/index.php.

Regular employees and scheduled cash earnings:

The monthly data of the numbers of regular employees and scheduled cash earnings are reported in the Monthly Labour Survey ("*Maitsuki Kinrou Tokei Chosa*") the Ministry of Health, Labour, and Welfare conducts. The data is available at the URL: http://www.mhlw.go.jp/toukei/list/30-1.html.

References

Anderson, J.E., van Wincoop, E., 2003, "Gravity with gravitas: a solution to the border puzzles," *American Economic Review* 93, 170 − 192.

Anderson, J.E., van Wincoop, E., 2004, "Trade costs," *Journal of Economic Literature* 42, 691−751.

Atkeson, A. and Burstein, A., 2008, "Pricing-to-market, trade costs, and international relative prices," mimeo.

Baba, C., 2007, "Price dispersion across and within countries: the case of Japan and Korea," *Journal of the Japanese and International Economies* 21, 237 − 259.

Bernard, A.B., Eaton, J., Jensen, J.B., Kortum, S., 2003, "Plants and productivity in international trade," *American Economic Review* 93, 1268 − 1290.

Bernard, A.B., Jensen, J.B., 2004, "Entry, expansion, and indensity in the US export boom 1987-1992," *Review of International Economics* 12, 662 − 675.

Broda, C., Weinstein, D.E., 2008, "Understanding international price differences using barcode data," *NBER working paper* 14017.

Crucini, M. J., Telmer, C.I., Zachariadis, M., 2005, "Understanding European real exchange rates," *American Economic Review* 95, 724 − 738.

Crucini, M. J., Telmer, C.I., Zachariadis, M., 2005, "Price dispersions: the role of borders, distance, and location," mimeo.

Ceglowski, J., 2003, "The law of one price: intranational evidence for Canada," *Canadian Journal of Economics* 36, 373 – 400.

Disdier, A.-C., Head, K., 2008, "The puzzling persistence of the distance effect on bilateral trade," *Review of Economics and Statistics* 90, 37 – 48.

Donaldson, D., 2010, "Railloads of the Raj: estimating the impact of transportation infrastructure," *NBER working paper* 16487.

Engel, C., Rogers, J.H., 1996, "How wide is the border?" *American Economic Review* 86, 1112 – 1125.

Engel, C., Rogers, J.H., Wang, S.-Y., 2005, "Revisiting the boarder: an assessment of the law of one price using very disaggregate consumer price data," in Driver, R., Sinclair, P., Thoenissen, C., eds., *Exchange Rates, Capital Flows and Policy*, Routledge, NY.

Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Helpman, E., Meitz, M., Rubinstein, Y., 2008, "Estimating trade flows: trading partners and trading volumes," *The Quarterly Journal of Economics* 123, 441 – 487.

Inanc, O., Zachariadis, M., 2010, "The importance of trade costs in deviations from the law of one price: estimates based on the direction of trade," *Economic Inquiry*, forthcoming.

Johnson, R., 2010, "Trade and prices with heterogeneous firms," mimeo, Department of Economics, Dartmouth College.

Melitz, M., 2003, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica* 71, 1695 – 1725.

Midrigan, V., 2007, "International price dispersion in state-dependent pricing models," *Journal of Monetary Economics* 54, 2231 – 2250.

Newey, W.K., 1987, "Efficient estimation of limited dependent variable models with endogenous explanatory variables," *Journal of Econometrics* 36, 231 – 250.

Parsley, D. C., Wei, S.-J., 1996, "Convergence to the law of one price without trade barriers or currency fluctuations," *The Quarterly Journal of Economics* 114, 1211 – 1236.

Rivers, D., Vuong, Q.H., 1988, "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics* 39, 347 – 366.

Vella, F., Verbeek, M., 1999, "Two-step estimation of panel data models with censored endogenous variables and selection bias," *Journal of Econometrics* 90, 239 – 263.

Wooldridge, J.M., 2002, *Econometric Analysis of Cross Section and Panel Data* MIT Press, Cambridge.

Table 1: Descriptive Statistics of Data

	Cabbage	Carrot	C-Cabbage	Lettuce	Potato	S-Mushroom	Spinach	Welsh Onion
(a) Product entry								
No. of varieties	3	10	4	7	10	1	4	11
No. of size categories	63	62	50	71	50	74	17	103
No. of grade categories	34	66	50	46	93	55	85	58
No. of producing prefectures	47	46	46	43	47	44	47	46
No. of wholesale markets	47	47	47	47	47	47	47	47
No. of distinct product entries	1,207	1,186	1,001	903	1,423	909	551	1,115
(b) Data truncation								
No. of $T_{ij}(l) = 0$ or 1	369,343	198,129	241,871	239,703	264,280	476,919	466,337	547,272
No. of $T_{ij}(l) = 1$	15,841	8,395	10,803	11,565	10,921	11,845	15,977	14,874
Mean log distance over $T_{ij}(l) = 0$ or 1	5.939	6.027	5.938	5.984	6.219	5.930	5.922	5.944
Mean log distance of $T_{ij}(l) = 1$	3.705	3.990	4.009	3.950	4.339	2.691	3.255	2.943
(c) Price differential								
Mean log price differential $q_{ij}(l)$	0.039	0.075	0.065	0.026	0.081	0.003	0.029	0.016
SD. log price differential $q_{ij}(l)$	0.167	0.285	0.227	0.267	0.265	0.127	0.216	0.178
$\hat{\gamma}_{\mathrm{OLS}}$	0.033	0.051	0.042	0.022	0.037	0.008	0.044	0.033
(s.e.)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Note 1: $\hat{\gamma}_{\mathrm{OLS}}$ represents the OLS estimate of the coefficient γ in the regression specification $q_{ij}(l) = \mu + \gamma d_{ij} + u_{ij}(l)$ where μ is constat and $u_{ij}(l)$ is an OLS disturbance. Note that $q_{ij}(l)$ is the price differential between consuming and producing regions i and j. "(s.e.)" reports the corresponding standard error.

Table 2: Results of FIML Estimation

	Cabbage	Carrot	C-Cabbage	Lettuce	Potato	S-Mushroom	Spinach	Welsh onion
(a) Point estimates and s.e.								
$\hat{\gamma}_{\text{FIML}}$	0.210	0.312	0.304	0.325	0.256	0.303	0.302	0.256
(s.e.)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\epsilon}_l$	3.907	1.819	3.435	2.876	1.919	3.576	3.521	4.004
(s.e.)	(0.012)	(0.010)	(0.016)	(0.013)	(0.012)	(0.041)	(0.013)	(0.014)
$\hat{\rho}$	-0.629	-0.313	-0.646	-0.691	-0.278	-0.395	-0.656	-0.684
(s.e.)	(0.003)	(0.004)	(0.002)	(0.002)	(0.003)	(0.006)	(0.003)	(0.002)
log likelihood	-36434.658	-192952.779	-11670.878	-53524.355	-144737.779	-448473.896	-29364.250	99351.534
No. of observations	369,343	198,129	241,871	239,703	264,280	476,919	466,337	547,272
(b) Diagnosis check								
PCP for $T_{ij}(l) = 0$ or 1	0.989	0.962	0.990	0.990	0.981	0.994	0.994	0.996
PCP for $T_{ij}(l) = 0$	0.995	0.976	0.995	0.996	0.997	0.997	0.998	0.999
PCP for $T_{ij}(l) = 1$	0.856	0.642	0.874	0.865	0.612	0.903	0.902	0.911
$\hat{\gamma}_{\text{OLS}}$ with predicted $q_{ij}(l)$	0.059	0.113	0.062	0.068	0.040	0.018	0.085	0.063
$\hat{\gamma}_{\text{OLS}}$ with actual data	0.033	0.051	0.042	0.022	0.037	0.008	0.044	0.033

Note 1: The log likelihood of the FIML estimation is given by equation (11). Each estimation includes monthly dummies, consuming prefectural dummies, and producing prefectural dummies both in price differential and selection equations (9) and (10).
Note 2: "Pcp" represents the "percent correctly predicted."

Figure 1: Simulated Probabilities of Product Delivery

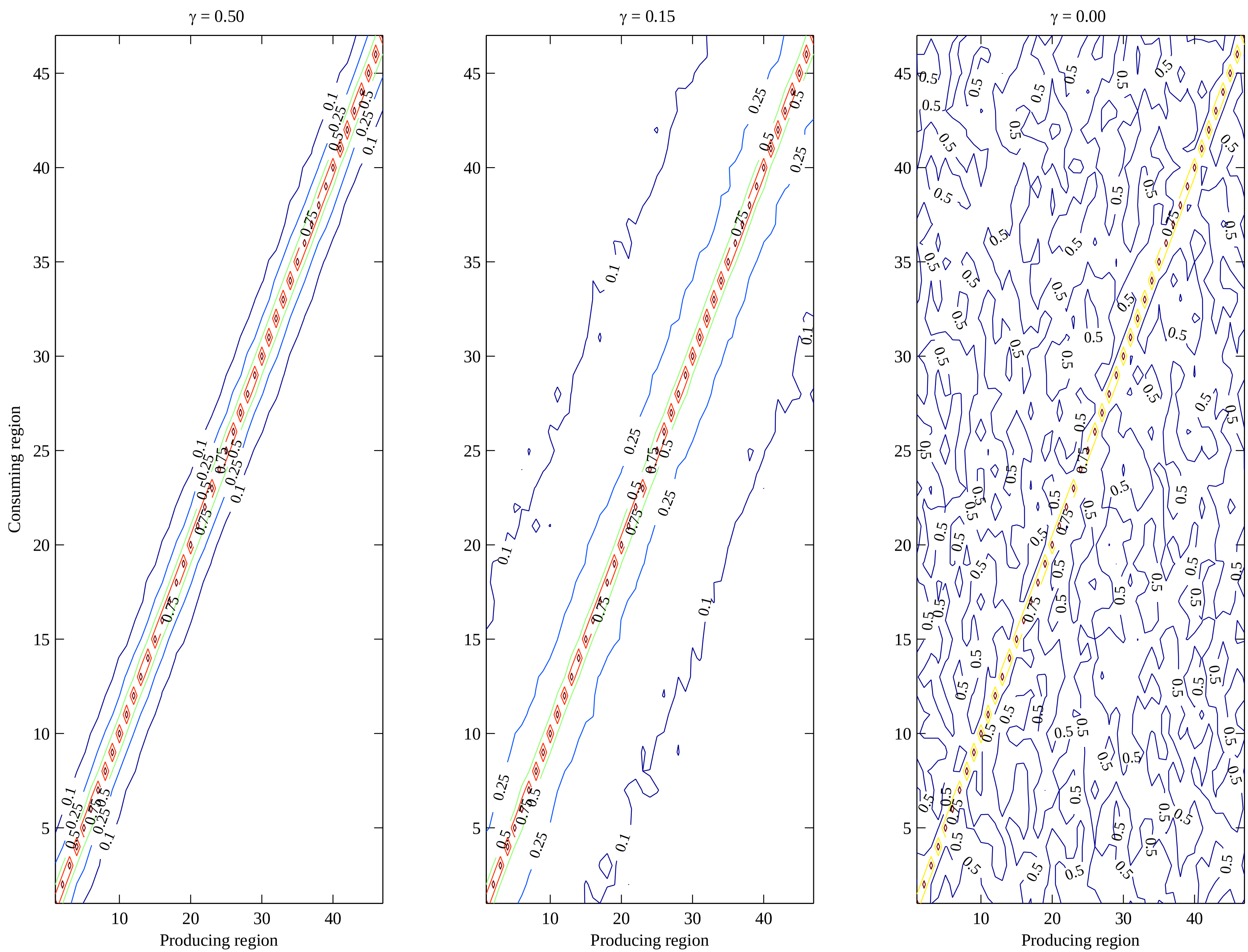

Figure 2: Simulated Price Differentials

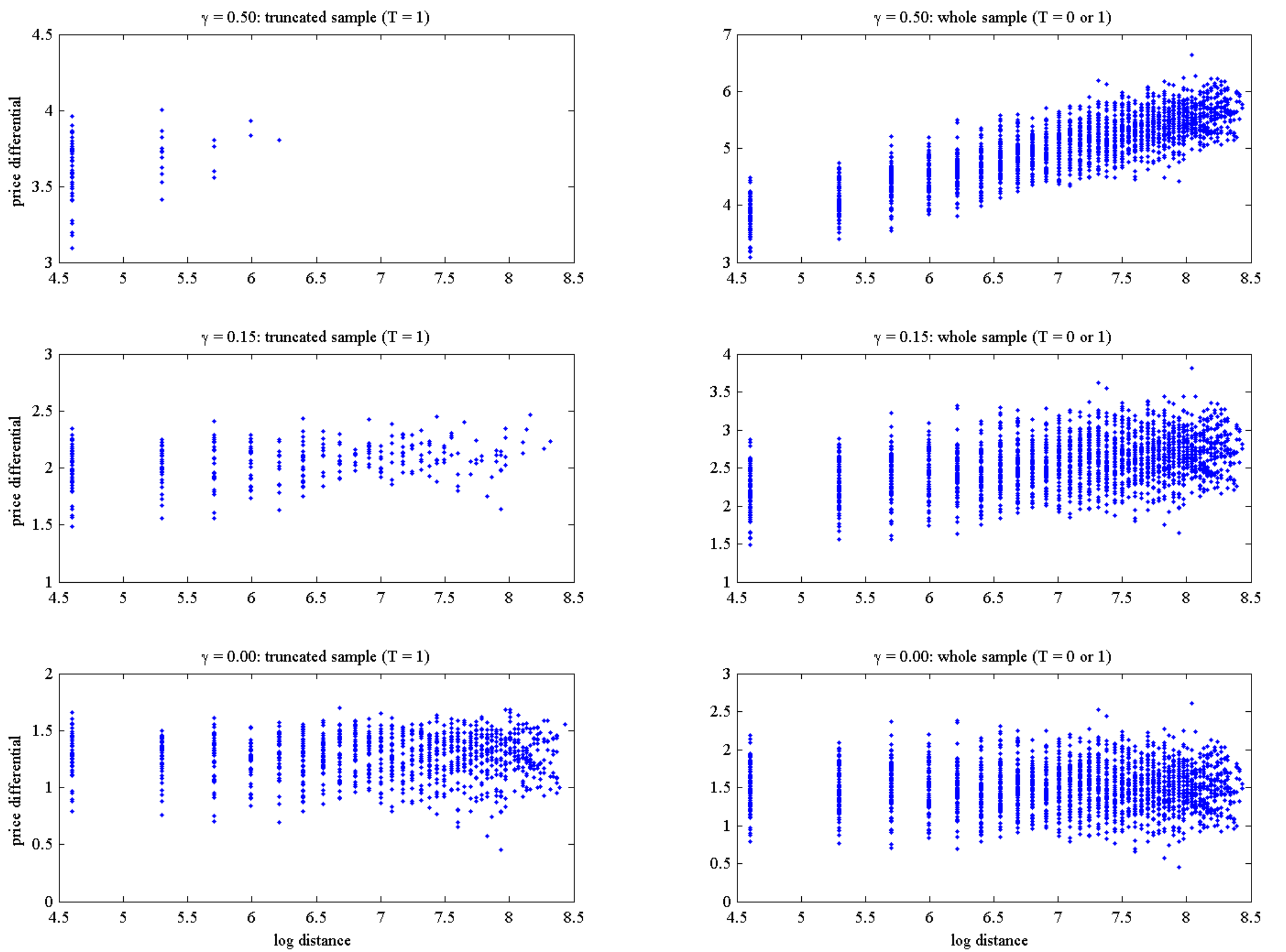

Figure 3: Kernel-Smoothed Densities of Estimators of Distance Elasticity

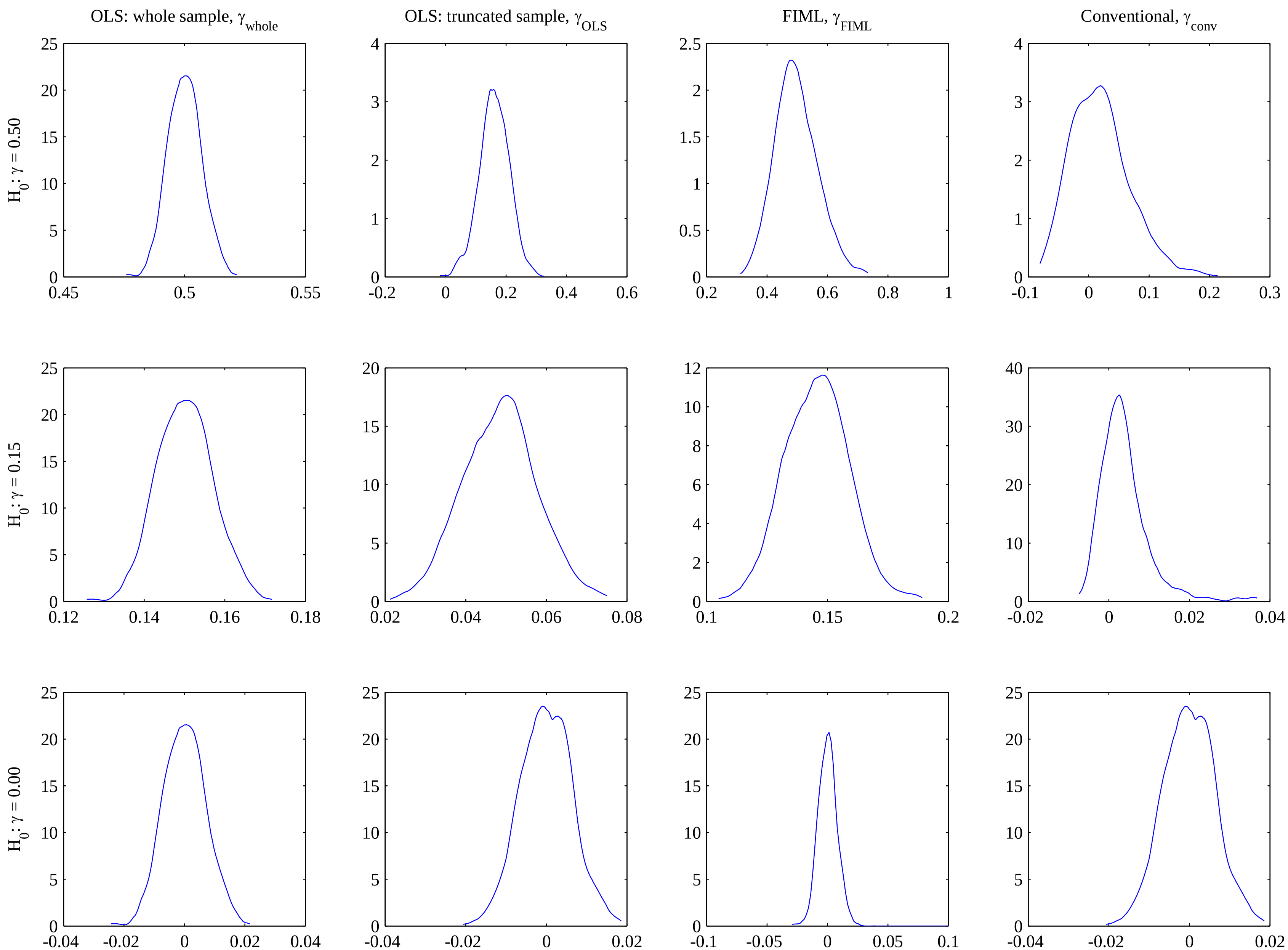

Figure 4: Data Probabilities of Product Delivery

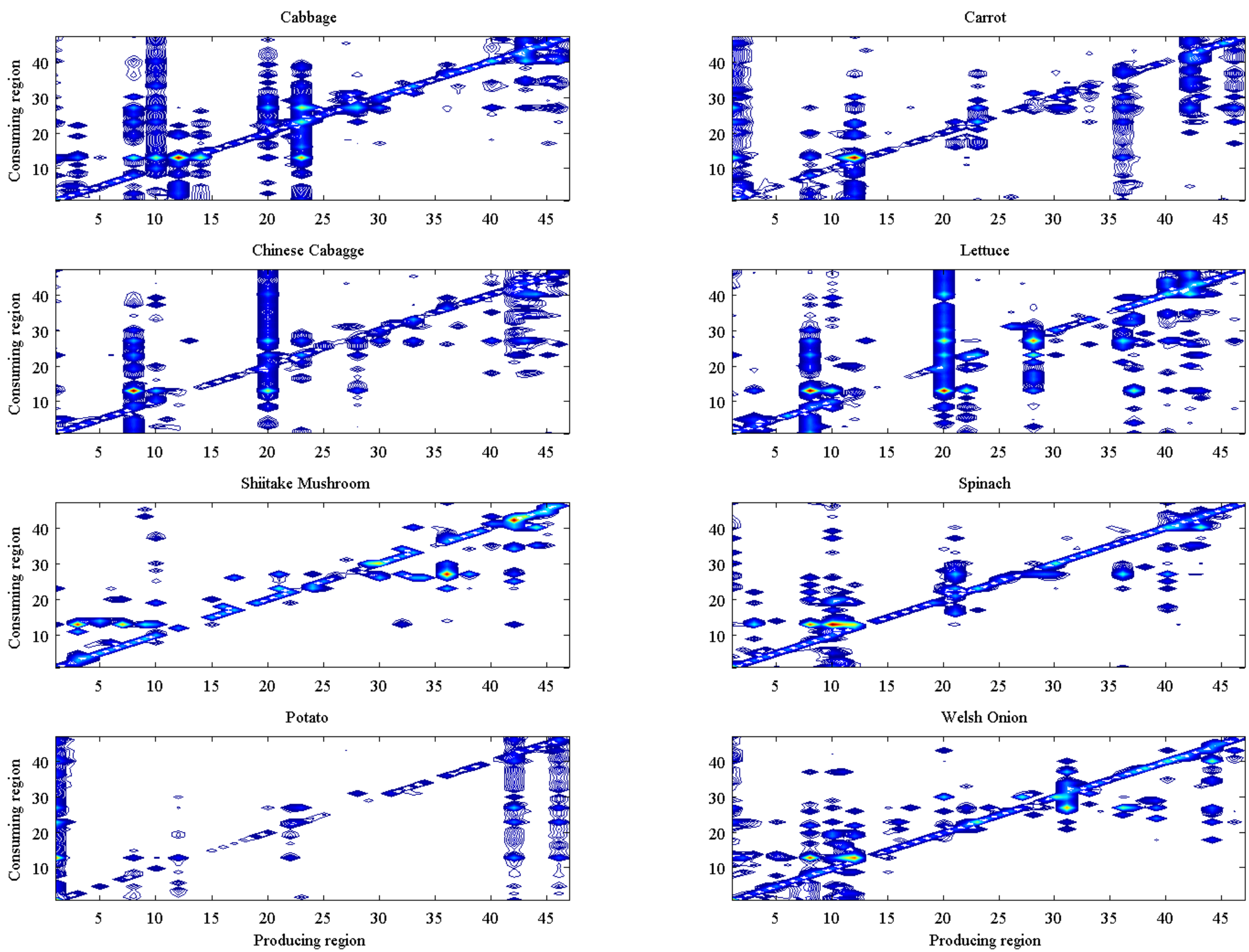

Figure 5: Predicted and Actual Price Differentials