HITOTSUBASHI UNIVERSITY

Global COE Hi-Stat Discussion Paper Series 245

Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

Exaggerated Death of Distance: Revisiting Distance Effects on Regional Price Dispersions

Kazuko Kano
Takashi Kano
Kazutaka Takechi

October 2012

Hi-Stat Discussion Paper

Hi-Stat
Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitatchi Tokyo, 186-8601 Japan
http://gcoe.ier.hit-u.ac.jp

Exaggerated Death of Distance:

Revisiting Distance Effects on Regional Price Dispersions

Kazuko Kano

Graduate 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
Hitotsubashi University
2-1 Naka, Kunitachi
Tokyo, 186-8601, Japan
Tel: +81-42-580-8283
Fax: +81-42-580-8283
Email: tkano@econ.hit-u.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: September 26, 2012

Abstract

This paper empirically establishes the significant roles of transport costs in price dispersions across regions. We identify and estimate the iceberg-type distance-elastic transport costs as a parameter of a structural model of cross-regional price differentials featuring product delivery decisions. Utilizing a data set of wholesale prices and product delivery patterns of agricultural products in Japan, our structural estimation approach finds large distance elasticities of the transport costs. The result confirms that geographical barriers are an economically significant contributor to the failures of the law of one price.

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

JEL Classification Number : F11, F14, F41

† *Corresponding author.*

1. Introduction

Recent years have witnessed the increased roles of trade costs in international macroeconomics. Highlighting the microfoundations of international trade patterns and geographical market segmentations with trade costs, careful calibration studies deepen our understanding of puzzling data characteristics in international macroeconomics.[1]

This paper empirically establishes the significant roles of transport costs, which are the major component of trade costs, in price dispersions across regions. Utilizing a data set of price differentials and product delivery patterns across regions, we identify and estimate the distance-elastic transport costs as a parameter of a structural model. The previous reduced-form regression studies treat the data associations between price differential and distance as a proxy of transport costs liberally, as in Rogers and Jenkins (1995), Engel and Rogers (1996), Engel and Rogers (2001), and Crucini et al.(2010). To the contrary, our structural estimation approach econometrically extracts the unobservable size of the transport costs from the reduced-form data associations in our data set.[2] The resulting structural estimate of the distance elasticity of transport costs evaluates an implicit price of the geographical barrier between the segmented markets. Our estimation of a "price of distance," indeed, is the first attempt to parse out structurally different potential contributors to the cross-regional price dispersions.

According to Anderson and van Wincoop (2004), trade costs in general consist of two categories: costs imposed by policies (e.g., tariffs, quotas, and the like) and costs imposed by the environment (e.g., transportation, insurance against various hazards, and time costs). Except for the extensive work by Hummels (1999), the direct measures of both categories are scarce and inaccurate. The empirical task of probing trade costs, therefore, largely relies on indirect econometric inferences from the measurements of equilibrium prices and quantities. Particularly in the field of international macroeconomics, the most common method

[1] Obstfeld and Rogoff (2001) point out trade costs as a central factor to explain six major puzzles in open-economy macroeconomics. Utilizing trade costs to motivate entry and exit behavior of heterogenous firms in export markets, Ghironi and Melitz (2005) provide a microfounded explanation for the Harrod-Balassa-Samuelson effect. Allowing for the distribution of trade costs over goods, Bergin and Glick (2009) endogenously determine the tradedness of goods in a small open-economy model. The resulting endogenous share of non-traded goods in the consumer price index accounts for the empirically observed low volatility in the relative price of non-traded goods. Atkeson and Burstein (2008) show that trade costs are essential to pricing-to-market behaviors of firms with variable markups in an open-economy model of imperfect competitive markets.

[2] Our structural estimate is a cousin of those identified in recent works by Crozet and Koenig (2010) and Balistreri et al.(2011) who use structural gravity models of international trade. Our approach, however, is quite different from theirs.

of inferring trade costs exploits the hypothesis of the law of one price (LOP) because trade costs are recognized to be the main obstacles to the perfect arbitrage of goods across regions. To approach the hypothesis, previous studies scrutinize disaggregate consumer prices, which are surveyed internationally as well as domestically across retail stores. In addition to the well-known violations of the LOP, one of the most robust findings across the previous reduced-form regression exercises is the statistically significant effects of geographical distance on the absolute levels or the times-series variances of the cross-regional price differentials.[3] Because distance is used as a liberal proxy for the transport costs, the empirically significant distance effects in the price differentials are suggestive, but still indecisive, evidence for transport costs as a major contributor to the LOP violations. There are at least three concerns.

The first concern relates to the measurement of transport costs. As argued by Engel and Rogers (1996) and Engel et al.(2005), the dependence of consumer price differentials on the distance observed in the reduced-form regressions is a mixture of several mutually exclusive effects: it reflects not only the transport costs but also other factors such as the geographical differences in the local distributional costs and the heterogeneous markups due to a home bias in preferences. The second concern regards the economic significance of the transport costs in the price differentials. Many of the past studies estimate that the elasticity of the price differential with respect to distance is less than 3%.[4] This small estimate for the distance elasticity of the price differential requires an unrealistically large degree of geographical scattering of sampling points (i.e., retail stores in cities) to explain the observed degree of price dispersions alone.[5] This observation naturally casts doubt on

[3]A not-exhaustive list of studies that conduct gravity-type regressions contains Engel and Rogers (1996), Parsley and Wei (1995), Broda and Weinstein (2008), Engel et al. (2005), Ceglowski (2003), Crucini at al. (2010), and Baba (2007).

[4]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 the barcode-level scanner data of retail prices across Canadian and U.S. cities. Engel et al. (2005) find the distance elasticity of 0.32% with pooled annual panel data distributed by the Economic Intelligence Unit (EIU) that covers retail prices of 100 consumer goods surveyed in 17 Canadian and U.S. cities. Ceglowski (2003) reports 1.6-2.0% estimates for 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 approximately 3% of the distance elasticity after taking into account a border dummy between the two countries.

[5]Because the standard deviation of the absolute value of the log price differential is typically reported at approximately 20% in this literature, we need a standard deviation of the log of distance of 6.66 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 for the geographical scattering of cities. For instance, the standard deviation of the log of distance between two prefectural capital cities in Japan is 0.803 over all of the 1081 city-pairs from 47 prefectures.

defining the transport costs as a main economic source for the cross-regional price dispersions: distance is empirically "dead" as a prime suspect for the commonly observed violations of the LOP.

Lastly, this economically subtle distance effect on the price differentials appears to be sharply inconsistent with the indirect econometric inferences from equilibrium trade volumes. Past empirical studies in international trade unambiguously recognize that distance plays an economically crucial role in determining bilateral trade volumes. Anderson and van Wincoop (2003) estimate a gravity model of bilateral trade volumes and infer that the distance elasticity of transport costs is approximately 20% conditional on a conventional calibration of the elasticity of substitution. Helpman et al.(2008) find that the distance elasticity of bilateral export volumes is approximately 80%, taking account of firms' selections into bilateral export markets with firm heterogeneity in productivity.[6] Importantly, their estimate suggests a 20% distance elasticity of transport costs under the same calibration of the elasticity of substitution as that used in Anderson and van Wincoop (2003). Why is our inference of the distance elasticity of transport costs widely diverse, at between approximately 3% and 20% when using data of equilibrium prices and quantities, respectively? This question is a serious challenge for the students of international economics who admit the importance of trade costs.

We incorporate the above concerns into our inferences on the effects of transport costs on price dispersions. In so doing, we investigate a unique daily data set of wholesale prices of agricultural products in Japan.[7] Following the spirit of Parsley and Wei (1996), we use disaggregate price data within a country to avoid any potential effects of cross-country differences in tax, tariff, quota, and currency on our inference on transport costs. Scrutinizing the information of wholesale prices helps us overcome the first concern: we make our estimate of transport costs immune to the influences of local distributional costs as well as to the local retailers' pricing strategies.[8]

More importantly, there are two outstanding characteristics of our data set. First,

[6] Indeed, this size for the 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.

[7] 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 the prices of agricultural products sold in open-air markets in European countries to test the theoretical implications of his state-dependent pricing model with trade costs.

[8] As pointed out by the editor Charles Engel, our inferences from the wholesale prices are still not immune to the influence of the cross-regional heterogeneity of markups. We empirically control for these effects by regional fixed effects in our estimation exercise.

we can identify the wholesale prices of an identical product at both the producing and the consuming regions. The first characteristic is essential for identifying the transport costs because, as discussed by Anderson and van Wincoop (2004), only when the source region of a product is identified, can the correct information for the transport costs be extracted from the relative prices at the consuming regions to the corresponding source region. The main difficulty that past studies face is the fact that a retail price survey at retail stores rarely provides information on 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.[9]

The second outstanding characteristic of our data set is the provision of information on the daily delivery patterns of an identical product from the source region to the final consuming regions. This data aspect empirically shows us how far a product is delivered from the source region. In this paper, we build a structural model to explain the observed patterns of product delivery, and we claim theoretically that ignoring the underlying choice of delivery might result in a serious bias toward our inference on the role of distance in cross-regional price differentials. Because the price of the product at a consuming region is observed only when a product delivery occurs, an inference drawn only from the 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 transport costs in general increases a price differential but simultaneously deteriorates the probability of delivery.

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 transport costs.[10] We then show that

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

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

the degree of the sample-selection bias depends critically on two structural parameters of the model: the elasticity of transport costs to distance and that of substitution. Our theoretical analysis implies that drawing a correct inference on transport 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 the 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.

We estimate our sample-selection model by FIML using the data for wholesale prices of several selected vegetables. The estimated sample-selection model passes two diagnostic criteria in that it does a fairly good job in replicating the actual delivery patterns of these vegetables and the actual data association of the price differentials with distances. More importantly, the resulting estimates resolve the second and third concerns. We find large estimates for the distance elasticity of transport costs across all of the vegetables relative to the existing estimates in the LOP literature: all of them are over 20% and their average is approximately 24%. The estimate of this paper, therefore, implies an economically significant role of transport costs in cross-regional price differentials. Moreover, this size of the estimate of the distance elasticity of transport costs is fairly consistent with those identified by the international trade models that explain equilibrium trade volumes.

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. Section 3 describes our data set. After reporting the empirical results in section 4, we conclude in section 5.

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 observed in Melitz (2003) and Helpman et al. (2008). In our model, a country consists of distinct consuming regions indexed by $i = 1, 2, \cdots, I$. Each region i is endowed with a representative household who consumes a continuum of agricultural products (such as cabbages, carrots, potatoes, and so on) by purchasing them at the regional wholesale market. Agricultural products are indexed by $l \in [0, 1]$. Each product l is produced in distinct source regions indexed by $j = 1, 2, \cdots, J$. Source region j delivers its product to the wholesale markets in the same region j and in consuming regions $i \neq j$ when the product delivery is expected to be profitable.

The representative household in consuming region i differentiates product l over the distinct source regions with an imperfect degree of substitution. Let x_{ijl} denote the demand of region i for product l that is produced in and delivered from region j. The representative household in region i then earns its utility from consuming product l with the following constant elastic utility function

$$x_{il} = \left[\int_{j \in B_{il}} (\delta_{ijl} x_{ijl})^{\alpha_l} dj \right]^{1/\alpha_l}, \quad 0 < \alpha_l < 1,$$

where B_{il} is the set of source regions that deliver product l to consuming region i. This utility function specific to product l shows 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 $\epsilon_l \equiv 1/(1 - \alpha_l) > 1$. Term δ_{ijl} reflects the household's biased preference for different producing regions: the greater the term δ_{ijl}, 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 delivered from source region j under the price p_{ijl}

$$x_{ijl} = \left(\frac{p_{ijl}}{p_{il}} \right)^{-\epsilon_l} \delta_{ijl}^{\epsilon_l - 1} x_{il}, \tag{1}$$

where $p_{il} \equiv [\int_{j \in B_{il}} (\delta_{ijl} p_{ijl})^{1-\epsilon_l} dj]^{1/(1-\epsilon_l)}$ is the average (i.e., aggregate) price level of product l in consuming region i.

A producer in region j is a monopolistically competitive producer at the wholesale markets in its own region as well as in 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 that minimize 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.

We assume that a producer in a region does not need to bear any transport costs when selling its product at the wholesale market in the same region. Hence, at the wholesale market in region j, the producer of product l, who faces the demand function (1), maximizes profits by charging a markup price. However, if the same producer seeks to sell its product at the wholesale market in distinct consuming region $i \neq j$, two types of delivery costs should be borne by the producer: a fixed cost of serving the market in region i, denoted by $c_j f_{ij}$, and an iceberg-type transport cost, denoted by τ_{ijl}. As in Helpman et al. (2008), no fixed or transport costs are required for a delivery to the local wholesale market: $f_{jjl} = 0$ and $\tau_{jjl} = 1$ for any j. However, a producer in source region j needs to bear positive fixed and transport costs: $f_{ijl} > 0$ and $\tau_{ijl} > 1$ for $i \neq j$. The optimal markup price, then, is

$$p_{ijl} = \tau_{ijl} \frac{c_j a_l}{\alpha_l}. \tag{2}$$

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

$$\pi_{ijl} = (1 - \alpha_l) \left(\frac{\tau_{ijl} c_j}{\alpha_l p_{il}} \right)^{1-\epsilon_l} \theta_{ijl}^{1-\epsilon_l} p_{il} x_{il} - c_j f_{ijl},$$

where $\theta_{ijl} \equiv a_l / \delta_{ijl}$ is the ratio of the productivity level to the producing regional bias. If the producer in region j sells its product l at the local wholesale market, the corresponding monopolistic profit π_{jjl} is always positive. However, delivering the same product to another consuming region i is profitable only if θ_{ijl} is smaller than a threshold $\bar{\theta}_{ijl}$, where $\bar{\theta}_{ijl}$ is defined by the zero profit condition,

$$(1 - \alpha_l) \left(\frac{\tau_{ijl} c_j}{\alpha_l p_{il}} \right)^{1-\epsilon_l} \bar{\theta}_{ijl}^{1-\epsilon_l} p_{il} x_{il} = c_j f_{ijl}. \tag{3}$$

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

$$T_{ijl} = \begin{cases} 1 & \text{if} \quad \theta_{ijl} < \bar{\theta}_{ijl}, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

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

Optimal price (2) implies that a price differential of an identical product between the source and the consuming regions provides a precise identification of transport cost τ_{ijl}. Let q_{ijl} denote the log of the price differential of product l between the producing and the consuming regions j and i: $q_{ijl} \equiv \ln p_{ijl} - \ln p_{jjl}$. Then, optimal price (2) and delivery decision mechanism (4) together yield the price differential equation

$$q_{ijl} = \ln \tau_{ijl}, \quad \text{only if } T_{ijl} = 1. \tag{5}$$

Price differential equation (5) has two important empirical implications. First, transport cost τ_{ijl} can be measured from the corresponding price differential only when we can identify the prices in the source and the consuming regions. This is the argument made by Anderson and van Wincoop (2004) against the conventional approach to measuring trade costs in the literature on regional and cross-country price dispersions. The second implication, however, says that identifying the source and the consuming regions is not enough for a precise estimation of the transport costs. Equation (5) shows that there is an incidental truncation or sample section: we can observe the price differential of product l between the source and the 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 (4). This selection mechanism depends on transport cost τ_{ijl} through the threshold characterized by equation (3). Therefore, transport cost τ_{ijl} in equation (5) could be inconsistently estimated unless we take sample-selection mechanism (4) explicitly into account.

An important caveat of the above identification of transport costs stems from product arbitrage among distinct wholesale markets. With cross-market product arbitrage, a price differential (5) might not be a sufficient statistic for the underlying transport cost because the observed equilibrium price in a consuming region can deviate from the optimal price (2). In this paper, we do not impose a no-arbitrage condition on our data as a restriction of our model *a priori*. However, as we discuss in more detail in the next section, the amount of product transfers across the wholesale markets of agricultural products is quite small relative to the total amount of wholesale transactions in Japan. We interpret this fact as meaning that there is almost no opportunity for product arbitrage in the equilibrium wholesale prices in our data set. We simply control for any possible effects of product arbitrage on the price differentials by adding an i.i.d. zero-mean random error to the price differential equation (5).[11]

2.2. The empirical framework

Given the structural model, we now discuss our empirical framework. Following Helpman et al.(2008), we parametrically specify transport cost τ_{ijl} by

$$\tau_{ijl} = D_{ij}^{\gamma_l} \exp(\mu_l + u_{ijl}), \quad u_{ijl} \sim N(0, \sigma_{u,l}^2),$$

where D_{ij} represents the symmetric distance between regions i and j, γ_l exhibits the distance elasticity of the transport cost, and u_{ijl} is an i.i.d. unobserved region-pair specific part of the

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

transport cost. The positive constant $\mu_l > 0$ confirms that the transport cost always takes a value greater than 1 for all (i, j) pairs. We further assume that the fixed cost of delivery, f_{ijl}, is stochastic due to an i.i.d. unobserved regional-pair specific element v_{ijl}. Just as in Helpman et al. (2008), we exploit a parametric specification of f_{ijl}:

$$f_{ijl} = \exp(\lambda_{jl} + \lambda_{il} - v_{ijl}), \quad v_{ijl} \sim N(0, \sigma^2_{v,l}),$$

where λ_{il} and λ_{jl} are consuming and producing regional specific constants. v_{ijl} is assumed to be uncorrelated with u_{ijl}.

The product delivery choice is characterized as follows. Define a latent variable Z_{ijl}

$$Z_{ijl} = \frac{(1-\alpha_l)\left[\frac{\tau_{ijl}c_j}{\alpha_l p_{il}}\right]^{1-\epsilon_l} \theta_{ijl}^{1-\epsilon_l} p_{il} x_{il}}{c_j f_{ijl}}.$$

According to zero profit condition (3), product l is delivered from region j to region i only if Z_{ijl} is greater than 1. The logarithm of Z_{ijl}, denoted by $z_{ijl} \equiv \log Z_{ijl}$, then is

$$z_{ijl} = \beta_l - (\epsilon_l - 1)\gamma_l d_{ij} + \epsilon_l \ln p_{il} + \ln x_{il} + \xi_{jl} + \lambda_{il} - \varrho_{ijl} + \eta_{ijl}, \quad \eta_{ijl} \sim N(0, \sigma^2_{\eta,l}), \tag{6}$$

where $\beta_l \equiv \ln(1-\alpha_l) + (\epsilon_l - 1)\ln\alpha_l + (1-\epsilon_l)\mu_l + (1-\epsilon_l)\ln a_l$, $\xi_{jl} \equiv -\epsilon_l \ln c_j - \lambda_{jl}$, and $\varrho_{ijl} \equiv (1-\epsilon_l)\ln\delta_{ijl}$. In particular, disturbance η_{ijl} is given as a linear combination of the unobserved components of the transport and the fixed costs, $\eta_{ijl} = (1-\epsilon_l)u_{ijl} + v_{ijl}$, with the variance $\sigma^2_{\eta,l} = (1-\epsilon_l)^2\sigma^2_{u,l} + \sigma^2_{v,l}$. Notice that the delivery of product l occurs from source region j to consuming region i and the corresponding price differential is selected into the sample only when $z_{ijl} > 0$. We thus call equation (6) the selection equation below. Price differential (5), in turn, is rewritten as

$$q_{ijl} = \mu_l + \gamma_l d_{ij} + u_{ijl}, \quad \text{only if } z_{ijl} > 0. \tag{7}$$

Disturbances η_{ijl} and u_{ijl} of the selection and price differential equations (6) and (7) are correlated negatively with the correlation coefficient $\rho_l \equiv \frac{(1-\epsilon_l)\sigma_{u,l}}{\sqrt{(1-\epsilon_l)^2\sigma^2_{u,l}+\sigma^2_{v,l}}} < 0$ because $\epsilon_l > 1$.

Selection and price differential equations (6) and (7) jointly reveal two critical aspects when identifying the distance elasticity of the transport cost, γ_l. First, estimating γ_l respecting only price differential equation (7) might lead to an under-biased inference. Our model explains the joint distribution of the price differential and the distance by two economic mechanisms. The first one, which is captured by price differential equation (7), is an intensive margin effect of distance: the longer the delivery distance is, the wider the price differential between the source and the consuming regions due to the greater transport cost. The second one is an extensive margin effect of distance. A negative correlation between the disturbances of the two equations, η_{ijl} and u_{ijl}, implies that, given the delivery distance between the source and the consuming regions, a product having a higher price at the final destination due to a greater unobservable factor of the transport cost tends not to be delivered to the consuming region. The intensive and the extensive margin effects then mean that only data points with relatively smaller price differentials, which correspond to shorter delivery distances and/or smaller unobservable factors of the transport costs, are likely to be selected into the sample. If we estimate price differential equation (7) with such a truncated sample, we obtain an under-biased estimate of distance elasticity γ_l.[12]

Second, the severity of the under-biasedness depends crucially on the elasticity of substitution, ϵ_l. As shown by the correlation coefficient ρ_l, the degree of the negative correlation between two disturbances u_{ijl} and η_{ijl} relies on the size of the elasticity of substitution, which the model restricts to being less than 1. In particular, notice that there is no correlation under the unit elasticity of substitution. Only in this special case, can we obtain an unbiased estimate of the distance elasticity estimating the price differential equation alone. This is because the distribution of the price differentials becomes independent with the underlying product delivery decision. Moreover, the distance effect on the delivery choice depends on both the distance elasticity and the elasticity of substitution in a nonlinear way. Selection equation (6) shows that the sensitivity of the delivery choice with respect to distance is nonlinearly associated with the two elasticities: if ϵ_l is small, the marginal effect of γ_l 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 transport costs and the elasticity of substitution within a sample-selection model that consists of equations (6) and (7). For this purpose, we conduct a 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_{ijl}, might be correlated with the endogenous variables p_{il} and x_{il} in the RHS of the selection equation. If this is the case, our point estimates of the structural parameters will be biased due to endogeneity.[13] To take into account the potential endogeneity bias, we further incorporate instrumental variables (IVs) into the FIML estimation as follows.[14] Let $\mathbf{y}_i$ denote a bivariate vector that contains $\ln p_{il}$ and $\ln x_{il}$ as its elements: $\mathbf{y}_i \equiv [\ln p_{il} \;\; \ln x_{il}]'$. We assume that vector $\mathbf{y}_i$ is linearly related to

[12] In Appendix A, we analytically describe the possibility of an under-biased estimate of distance elasticity γ_l due to a sample selection.

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

[14] 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.

a vector of exogenous IVs, $\mathbf{s}_i$, up to i.i.d. 2×1 mean zero random vector $\mathbf{e}_i$:

$$\mathbf{y}_i = \Gamma \mathbf{s}_i + \mathbf{e}_i. \tag{8}$$

Endogeneity bias occurs if the error of the selection equation (6), η_{ijl}, is correlated with the errors in equation (8), $\mathbf{e}_i$. More specifically, we assume that the 4×1 random vector of disturbances, $[\mathbf{e}_i' \; u_{ijl} \; \eta_{ijl}]'$, is stochastically governed by a joint normal density with the mean of zero and the 4×4 symmetric positive-definite variance-covariance matrix Ω

$$\Omega = \begin{bmatrix} \Omega_{11} & \varphi_{u,l}' & \varphi_{\eta,l}' \\ \varphi_{u,l} & \sigma_{u,l}^2 & \sigma_{u\eta,l} \\ \varphi_{\eta,l} & \sigma_{u\eta,l} & \sigma_{\eta,l}^2 \end{bmatrix}, \tag{9}$$

where Ω_{11} is a 2×2 matrix, $\varphi_{u,l}$, and $\varphi_{\eta,l}$ are 1×2 row vectors. The non-zero vector $\varphi_{\eta,l}$ characterizes the covariances between the disturbances of selection equation (6) and instrument equation (8) that would lead to potential endogeneity bias. Through our analysis, we presume that there is no correlation between the disturbances of price differential equation (7) and instrument equation (8): $\varphi_{u,l} = [0 \; 0]$.

The accompaning technical appendix shows in detail how our structural sample-selection model consisting of equations (7), (6), (8), and (9) provides the log likelihood function

$$\sum_{i,j} (1 - T_{ijl}) \ln [\Phi(\lambda_{ijl})] + \sum_{i,j} T_{ijl} \ln [\Phi(\kappa_{ijl})] \\ + \sum_{i,j} T_{ijl} \ln [\phi(\omega_{ijl})] - \sum_{i,j} T_{ijl} \ln \sigma_{u,l} + \sum_{i,j} \ln [f(\mathbf{y}_i|\mathbf{s}_i)], \tag{10}$$

where

$$\kappa_{ijl} = \frac{\beta_l - (\epsilon_l - 1)\gamma_l d_{ij} + [\epsilon_l \; 1]\mathbf{y}_i + b_{ijl}^s + \varphi_{\eta,l}\Omega_{11}^{-1}(\mathbf{y}_i - \Gamma \mathbf{s}_i) + \rho_l \sigma_{u,l}^{-1}[q_{ijl} - \mu_l - \gamma_l d_{ij} - b_{ijl}^p]}{(1 - \varphi_{\eta,l}\Omega_{11}^{-1}\varphi_{\eta,l}' - \rho_l^2)^{\frac{1}{2}}},$$

$$\varpi_{ijl} = \frac{q_{ijl} - \mu_l - \gamma_l d_{ij} - b_{ijl}^p}{\sigma_{u,l}},$$

$$\lambda_{ijl} = \frac{\beta_l - (\epsilon_l - 1)\gamma_l d_{ij} + [\epsilon_l \; 1]\mathbf{y}_i + b_{ijl}^s + \varphi_{\eta,l}\Omega_{11}^{-1}(\mathbf{y}_i - \Gamma \mathbf{s}_i)}{(1 - \varphi_{\eta,l}\Omega_{11}^{-1}\varphi_{\eta,l}')^{\frac{1}{2}}},$$

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

Here, constant b_{ijl}^p and b_{ijl}^s control for the regional fixed effects in price differential and selection equations (7) and (6), respectively.[15] We also normalize the selection equation (6)

[15]We also include monthly dummies in the price differential and selection equations to control for seasonality.

by setting the standard deviation of its error term, $\sigma_{\eta,l}$, equal to 1.[16] To maximize the log likelihood function (10) conditional on the observations of the delivery index $\{T_{ijl}\}$, the price differential $\{q_{ijl}\}_{T_{ijl}=1}$, the log of distance $\{d_{ij}\}$, the average price and aggregate transaction of product l in consuming regions $\{p_{il}\}$ and $\{x_{il}\}$, and instruments $\{\mathbf{s}_i\}$, we take a two-step approach to make our computation tractable. 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 (10) and maximize the resulting log likelihood function with respect to the rest of the parameters.[17]

3. 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, and 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 transport costs precisely because the LOP requires the identification of identical goods as its theoretical premise.[18] 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_{ijl} is constructed by subtracting the wholesale price in producing

[16] This standard normalization in a sample-selection model makes ρ_l equal to $(1-\epsilon_l)\sigma_{u,l}$. During estimation, we further impose a restriction that the correlation coefficient is always less than or equal to 1 in the absolute value.

[17] Although reported in the accompaning technical appendix, our Monte Carlo experiments based on the model show us that, given the elasticity of substitution, the degree of sample selection depends positively on the distance elasticity of transport costs. The experiments 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 transport costs and (ii) our FIML estimator successfully identifies the distance elasticity.

[18] The hypothesis of the LOP maintains an identical product with respect to the date of production. Without exact information of the production date, working with daily data is likely to provide a close identification of an identical product in terms of the production date.

prefecture j, p_{jjl}, from that in consuming prefecture i, p_{ijl}.[19] We set $T_{ijl} = 1$ for pair (i, j) if the sample of q_{ijl} is available.[20] The geographical distance between prefectural pair (i, j) is approximated by that between the prefectural governments' head offices placed in the prefectural capital cities. Taking the logarithm of the geographical distance yields variable d_{ij}. Our data set provides the daily aggregate transaction quantity of product l in consuming region i, x_{il}.[21] We are unable to obtain daily data of the aggregate price of product l in the consuming region i, p_{il}. Hence, we use as a proxy of p_{il} the monthly data of the retail price of product l. Moreover, to control for the daily variations in the producing and the consuming prefectures, we include into selection equation (6) the 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 the producing and the consuming regions are important factors for the production of and the demand for agricultural products. Finally, as valid IVs, we use the monthly numbers of regular employees and the scheduled cash earnings in each prefecture in addition to the monthly and consuming-region dummies.

We focus our exercise on 8 selected vegetables: cabbages, carrots, Chinese cabbages (c-cabbage, hereafter), lettuce, shiitake-mushrooms (s-mushroom, hereafter), spinach, potatoes, and welsh onions. 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 cabbages; 1,186 for carrots; 1,001 for c-cabbages; 903 for lettuce; 1,423 for potatoes; 909 for s-mushrooms; 551 for spinach; and 1,115 for welsh onions.

For each product entry l, we count the number of the delivery $T_{ijl} = 1$ and the non-delivery $T_{ijl} = 0$ only for the dates on which the product entry is indeed traded at the wholesale market in producing prefecture j. We identify the product delivery $T_{ijl} = 1$ if the data reports that the source prefecture of the product entry l sold in consuming region i is region j.[22] The first row of panel (b) of the table reports that the total number of both the

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

[20]We also set $T_{jjl} = 1$ whenever we can observe p_{jjl}. We consider such observations as the 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 $T_{jjl} = 1$ to avoid taking the log of zero distance.

[21]Whenever the data set reports that $x_{il} = 0$, we interpolate x_{il} by a very small number of 0.00001 to avoid taking the logarithm of zero.

[22]A problem with this identification would be that we cannot eliminate the possibility of product transfer: a product yielded in a source region is delivered to a consuming region and then transferred to another

delivery and the non-delivery cases over all of the product entries is almost over 190,000 for each vegetable. This is the number of observations 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 approximately 10,000 for each vegetable. Our data set, hence, indicates that the product delivery is quite limited.[23] The third row of panel (b) shows that the mean distance from the producing to the 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), however, conveys that the mean distance over the delivery cases only is much shorter depending on the vegetable, with the minimum number of 2.691 (14.746 km) for s-mushrooms and the maximum of 4.339 (76.630 km) for potatoes. Product delivery, therefore, is localized and concentrated on the local areas neighboring the producing prefectures.

Figure 1 also confirms the locality of the product delivery graphically. Each window of the figure depicts as a contour plot the data frequencies of the product delivery from the producing to the consuming prefectures, which are calculated over all product entries on all traded dates. The horizontal axis represents the producing prefectures and the vertical axis represents the consuming prefectures. The order of the prefectures reflects the geographical positions of the prefectures from the northern most prefecture, Hokkaido, to the southern most, Okinawa. Therefore, two prefectures that are indexed by close integers are indeed geographically close to each other. The brighter the contour line is, the higher the probability of product delivery. The figure then uncovers three facts. First, each vegetable has several dominant producing prefectures that are characterized by vertical contour lines. These main producing prefectures deliver their products to not only the nearby prefectures but also to the other remote prefectures. Second, the data frequencies of product delivery of the main producing prefectures are decreasing in distance. Therefore, even the dominant producers do not deliver their products to the consuming prefectures that are farthest away.[24] Third, the contour lines for the other minor producing prefectures are concentrated on the 45 degree

consuming region. If this case is dominant in our data set, our inference on the 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 cabbages; 6.5% for carrots; 4.9% for c-cabbages; 6.3% for lettuce; 6.0% for potatoes; 3.3% for s-mushrooms; 4.1% for spinach; and 3.9% for welsh onions. These ratios mean that almost all products in our data are directly delivered from the source regions to the consuming regions as their final destinations.

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

[24]An exception is observed in the first producing prefecture, Hokkaido, in the cases of carrots and potatoes.

line. The product delivery of these relatively minor producing prefectures, thus, is highly localized.

The locality of product delivery that Table 1 and Figure 1 together unmask brings us two important implications. First, as observed by Broda and Weinstein (2008) in their barcode data of retail products, the 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 that are idiosyncratic to the product origins. The price differentials across the consuming regions that are generated by these idiosyncratic factors cannot be attributed to transport costs. Hence, given the observed high degree of the regional product clustering, it is crucial to scrutinize the price differentials of a product that shares a source region to correctly infer the role that transport costs play in absolute LOP violations. Second, drawing an inference on the transport costs only from observed price differentials might be subject to a serious sample-selection bias, as we repeatedly claim in this paper.

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 the wholesale prices in the consuming prefectures are higher on average by between 0.3% and 8.1% than those of the producing prefectures. This observation is suggestive for the important role of transport costs in price differentials, as predicted by equation (7). The corresponding standard deviation of the observed log price differential, which is displayed on the second row of panel (c), is approximately 20 %. Our data set, thus, shows the almost same degree of absolute LOP violations as observed in the previous studies (e.g., Crucini et al 2005, and Broda and Weinstein 2008), even after identifying the source regions of products. We also conduct an OLS regression of the observed price differential on the corresponding log distance and the constant for each vegetable. The resulting OLS estimates of the coefficient on the log distance, $\hat{\gamma}_{l\text{OLS}}$, are shown in the third row of panel (c), which are accompanied by the standard errors. All of 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 the price differential is consistent with the estimates that past studies commonly found using different data sets such as in Engel et al. (2005), Broda and Weinstein (2008), and Inanc and Zachariadis (2009).

4. Results

4.1. Results of the FIML estimation

Table 2 summarizes the results of the FIML estimation based on the log likelihood (10). The first and second rows of panel(a) of the table show that the distance elasticity of transport costs, $\hat{\gamma}_{l\mathrm{FIML}}$, is estimated to be positive and statistically significant for each vegetable. The outstanding fact that this row tells us is the large size of the FIML estimates: the mean (over the 8 vegetables) of the estimated distance elasticity is 0.238 with the minimum of 0.210 for cabbages and the maximum of 0.325 for lettuce. According to equation (7), the price differential of a product between the consuming and the producing regions increases by approximately 24% in response to the 100 % stretch in delivery distance when ignoring selection mechanism (6). Compared with the small size of the OLS estimate of the distance elasticity, which is reported to be 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 seriously biased downwards due to the underlying data truncation.

As discussed in section 2, the strength of the observed under bias tightly connects with the elasticity of substitution, 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 8 vegetables. Combined with the large estimate of the distance elasticity of transport costs, the estimated elasticity of substitution implies that the probability of product delivery from the producing to the consuming prefectures depends negatively as well as sensitively on the delivery distance. The point estimate of the correlation coefficient between the unobserved disturbances of price differential equation (7) and selection equation (6), ρ_l, 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_l is estimated to be negative with a high statistical significance: the mean of the estimates of ρ_l over the 8 vegetables is -0.536 with the minimum of -0.684 for welsh onions and the maximum of -0.278 for potatoes. 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 (A.1).

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

4.2. Model validation through diagnostic 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 a model validation, we conduct diagnostic 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_{ijl}, 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 $T_{ijl} = 0$ or 1.[25] To construct the PCPs, we calculate the predicted conditional probabilities of $T_{ijl} = 0$ and $T_{ijl} = 1$ on the observables, $\hat{P}(T_{ijl} = 0|.)$ and $\hat{P}(T_{ijl} = 1|.)$, respectively. Then, if $\hat{P}(T_{ijl} = 0|.) > 0.5$, we recognize that our model predicts $T_{ijl} = 0$, while if $\hat{P}(T_{ijl} = 1|.) > 0.5$, it predicts $T_{ijl} = 1$. The PCP for $T_{ijl} = 0$ (or 1) is calculated as the percentage of the total number of the observations of $T_{ijl} = 0$ (or 1) that are accompanied by $\hat{P}(T_{ijl} = 0|.) > 0.5$ (or $\hat{P}(T_{ijl} = 1|.) > 0.5$). The PCP for either $T_{ijl} = 0$ or 1 is simply derived as a weighted average of the PCPs for $T_{ijl} = 0$ 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 of approximately 0.990 for either $T_{ijl} = 0$ or 1 for all of the vegetables. These results mean 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 $T_{ijl} = 0$ is better than its ability to replicate delivery choice $T_{ijl} = 1$. On the one hand, the high PCPs for $T_{ijl} = 0$ of approximately 0.990 suggest the model's outstanding predictive ability of no delivery choice. On the other hand, the PCPs for $T_{ijl} = 1$ 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 vegetables such as s-mushrooms, spinach, and welsh onions.[26] 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 the data association of the price differentials with the 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, $\hat{\gamma}_{\text{OLS}}$, at

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

[26] The main reason for the model's slightly lower predictive performance for carrots and potatoes is understandable. As observed in Figure 1, the main producing prefecture of these two vegetables, Hokkaido, delivers its products to all other prefectures regardless of the delivery distance. This data aspect is hard to explain with our simple structural model.

approximately 3% on average. The question we ask here is if our sample-selection model predicts this size of the OLS estimate or not.

To perform this diagnosis check, we derive the prediction of the model on price differentials following equation (A.1). Each window of Figure 2 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 of the windows except in the case of carrots. Hence our model successfully predicts the data association of the price differentials with the distances overall, although the actual data show us a much sparser 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 already been 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 the transport cost of our model is estimated at 0.238 by FIML. What is striking is that the sample-selection model with such a large distance elasticity indeed mimics such a small size of the OLS estimate. In this sense, we conclude that our model successfully passes the second diagnostic 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 the price differentials and the distances.

5. Conclusion

As claimed by Anderson and van Wincoop (2004), this paper provides evidence that the "death of distance" is exaggerated, even in the literature of regional price dispersions. Exploiting data on the wholesale price differentials and the delivery patterns of agricultural products in Japan, our structural estimation approach featuring product delivery choice uncovers the high implicit prices of the geographical barriers across regions. The size of the distance elasticity of transport costs that our FIML procedure estimates is fairly consistent with those found in the past empirical studies using gravity-type models of equilibrium trade volumes. The transport costs we infer as distance effects do indeed play economically significant roles in regional price dispersions. The empirical exercise that this paper conducts is the first rigid step to an ambitious goal of international finance — parsing out structurally different potential contributors to the widely documented reduced-form observations of LOP failures in retail prices.

Although this paper intensively scrutinizes the 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 transport costs in regional price dispersions for manufactured non-perishable products. If the transport costs are expensive, the firms might decide to locate their plants close to the consuming markets to economize on transport costs.[27] In this case, because product delivery becomes limited around the local areas that neighbor the plants, the 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 transport costs in regional price dispersions, as in our exercise. This conjecture suggests that there should be a more intensive use of plant level data in the LOP literature.

Finally, it is worth noting a caveat regarding 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 the 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 the price elasticity of demand and markup variable in the 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 the products from other source regions because the market shares of the other source regions change. Given the transport 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 the distance effects biased. We leave this extension to future research.

Appendix A. Analytical description of under-biasedness due to sample selectivity

Taking the expectation of price differential equation (7) conditional on $T_{ijl} = 1$ and other observables yields $E[q_{ijl}|., T_{ijl} = 1] = \mu_l + \gamma_l d_{ij} + E[u_{ijl}|., T_{ijl} = 1]$ where . represents the other observables. Notice that the term $E[u_{ijl}|. , T_{ijl} = 1]$ is related to the conditional expectation $\bar{\eta}_{ijl} \equiv E[\eta_{ijl}|. , T_{ijl} = 1]$ by $E[u_{ijl}|. , T_{ijl} = 1] = \rho_l \frac{\sigma_{u,l}}{\sigma_{\eta,l}} \bar{\eta}_{ijl}$. A consistent estimate of $\bar{\eta}_{ijl}$ is obtained by the inverse Mills ratio $\hat{\bar{\eta}}_{ijl} = \phi[\hat{z}_{ijl}]/\Phi[\hat{z}_{ijl}]$, where $\phi(.)$ and $\Phi(.)$ are the standard normal density

[27] A recent paper by Evans and Hariggan (2005) emphasizes the importance of transport time proxied by distance on the producers' choices of plant location, and it provides empirical evidence that the products where timely delivery is crucial are produced near the final demand markets.

and cumulative distribution functions, respectively. Therefore, we can rewrite price differential equation (7) as

$$q_{ijl} = \mu_l + \gamma_l d_{ij} + \beta_{u\eta}\hat{\hat{\eta}}_{ijl} + e_{ijl}, \tag{A.1}$$

where $\beta_{u\eta} = \rho_l \frac{\sigma_{u,l}}{\sigma_{\eta,l}}$, and e_{ijl} is an i.i.d. error term that satisfies $E[e_{ijl}|., T_{ijl} = 1] = 0$. Our model implies that $\rho_l < 0$. Moreover, the inverse Mills ratio $\hat{\hat{\eta}}_{ijl}$ is increasing in distance because $\hat{\hat{\eta}}_{ijl}$ is a decreasing function of the predicted latent variable $\hat{z}_{ijl}$, which then depends negatively on the distance through selection equation (6). Hence, if we ignore the third term of the RHS of equation (A.1) when estimating the distance elasticity γ_l only through the price differential equation, the resulting estimate could be biased downwards.

Appendix B. Data sources

Wholesale prices:

The data source for 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 following URL: http://www2s.biglobe.ne.jp/ fains/index.html). All of the contents in the data set are surveyed by the Ministry of Agriculture, Forestry, and Fishery (MAFF) for almost all transactions at 55 wholesale markets that are officially opened and operated in the 47 prefectures in Japan on a daily basis.

The data file contains information on the name of the product, the market prices, the name of the production site, the name of the market-place, and the product characteristics. The price reported has three forms: the highest price, the mode price, and the lowest price. Most of the 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 the 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 the prices between the production site and the market place. Thus, we construct the same category products by using the product characteristics and the production site. 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.[28] Because prices depend on detailed characteristics, we use each combination of characteristics to represent the same product.

The coverage of the vegetables traded through the central wholesale markets is substantial in Japan. While currently, the large supermarket and restaurant chains can not only directly purchase agricultural products from producers but can also directly import from foreign producers, the share of agricultural products covered by these markets in the entire set of vegetable transactions is still over 75% in Japan in 2006, according to MAFF. Thus, our data set enables us to approach the population characteristics of transport costs.

[28]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, and it is clean, trimmed, and free from decay and damage by insects. Otherwise, it is ranked as B.

Geographical distance:

The data of distance is provided by the Geographical Survey Institute (GSI) of the Government of Japan. The data are publicly available at 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 following URL: http://www.stat.go.jp/data/kouri/index.htm) from 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 for the number of regular employees and scheduled cash earnings are reported in the Monthly Labor Survey ("*Maitsuki Kinrou Tokei Chosa*") conducted by the Ministry of Health, Labour, and Welfare. The data are available from the website: http://www.mhlw.go.jp/toukei/list/30-1.html.

Acknowledgements

We would like to thank the editor Charles Engel, two anonymous referees, Naohito Abe, Kazumi Asako, Martin Berka, Andrew Bernard, Toni Braun, Dave Cook, Mario Crucini, Alexandre Dmitriev, Wei Dong, Shin-ichi Fukuda, Naoto Jinji, Daiji Kawaguchi, Michael Keane, Junko Koeda, Chul-In Lee, Shiko Maruyama, Yoshiro Miwa, Hideyuki Mizobuchi, Toshi Mukoyama, Jim Nason, Makoto Nirei, Hiroshi Ohashi, Glenn Otto, Ke Pang, John Rogers, Makoto Saito, Yasuyuki Sawada, Shigenori Shiratsuka, Alexandre Skiba, Gregor Smith, Yi-Chan Tsai, Takayuki Tsuruga, Tsutomu Watanabe, Jenny Xu, and the seminar participants in Hitotsubashi, Hong Kong University of Science and Technology, Hosei, Keio, Kyoto, Shanghai University of Finance and Economics, Seoul National, New South Wales, the Institute of Developing Economies of the Japan External Trade Organization, the National Graduate Institute for Policy Study Japan, the Policy Research Institute of the Ministry of Finance of the Government of Japan, the Research Institute of Capital Formation of the Development Bank of Japan, the 2010 Asia Pacific Trade Seminar, the fifth conference of Empirical Investigations in Trade and Investment, the 2011 Canadian Economic Association Meetings, the 2011 North American Summer Meetings of the Econometric Society, the 2011 Spring Meetings of the Japanese Economics Association, the 2010 Summer Workshop on Economic Theory, and the 2010 Tokyo Macro Workshop for their helpful and valuable comments, discussions and suggestions. The first and second authors wish to thank the Kikawada foundation and the Japan Center for Economic Research for financial support. The second and third authors would like to thank the Japan Society for the Promotion of Science for financial support from grants-in-aid for scientific research (numbers 20730205 and 24530270). We are solely responsible for any errors and misinterpretations of this paper.

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," *American Economic Review* 98, 1998 – 2031.

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.

Balistreri, E.J., Hillberry, R.H., Rutherford, T.F., 2011, "Structural estimation and solution of international trade models with heterogeneous firms," *Journal of International Economics* 83, 95 – 108.

Bergin, P.R., Glick, R., 2009, "Endogenous tradability and some macroeconomic implications," *Journal of Monetary Economics* 56, 1086 – 1095.

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 intensity 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.

Crozet, M., Koenig, P., 2010, "Structural gravity equations with intensive and extensive margins," *Canadian Journal of Economics* 43, 41-62.

Crucini, M. J., Shintani, M., Tsuruga, T., 2010, "The law of one price without the border: the role of distance versus sticky prices," *The Economic Journal* 120, 462 – 480.

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., 2001, "Deviations from purchasing power parity: causes and welfare costs," *Journal of International Economics* 55, 29 – 57.

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.

Evans, C.L., Harrigan, J., 2005, "Distance, time, and specialization: lean retailing in general equilibrium," *American Economic Review* 95, 292 – 313.

Ghironi, F., Melitz, M.J., 2005, "International trade and macroeconomic dynamics with heterogeneous firms," *Quarterly Journal of Economics* 120, 865 – 915.

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

Hummels, D., 1999, “Towards a geography of trade costs,” *GTAP working paper* No.17.

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., 2012, “Trade and prices with heterogeneous firms,” *Journal of International Economics* 86, 43-56.

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.

Obstfeld, M., Rogoff, K., 2001, “”The six major puzzles in international macroeconomics: Is there a common cause?” in Bernanke, B.S., Rogoff, K. eds. *NBER Macroeconomics Annual 2000* vol 15, 339 – 412.

Parsley, D. C., Wei, S.-J., 1996, “Convergence to the law of one price without trade barriers or currency fluctuations,” *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.

Rogers, J.H., Jenkins, M., 1995, “Haircuts or hysteresis? Sources of movements in real exchange rates” *Journal of International Economics* 38, 339 – 360.

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 $T_{ijl} = 0$ or 1	369,343	198,129	241,871	239,703	264,280	476,919	466,337	547,272
No. of $T_{ijl} = 1$	15,841	8,395	10,803	11,565	10,921	11,845	15,977	14,874
Mean log distance over $T_{ijl} = 0$ or 1	5.939	6.027	5.938	5.984	6.219	5.930	5.922	5.944
Mean log distance of $T_{ijl} = 1$	3.705	3.990	4.009	3.950	4.339	2.691	3.255	2.943
				(c) Price differential				
Mean log price differential q_{ijl}	0.039	0.075	0.065	0.026	0.081	0.003	0.029	0.016
SD. log price differential q_{ijl}	0.167	0.285	0.227	0.267	0.265	0.127	0.216	0.178
$\hat{\gamma}_{l\text{OLS}}$	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: $\hat{\gamma}_{l\text{OLS}}$ represents the OLS estimate of the coefficient γ_l in the regression specification $q_{ijl} = \mu_l + \gamma_l d_{ij} + u_{ijl}$ where μ_l is constat and u_{ijl} is an OLS disturbance. Note that q_{ijl} is the price differential between consuming and producing regions i and j. "(s.e.)" reports the corresponding standard error.

Table 2: Results of FIML-IVs estimation

	Cabbage	Carrot	C-Cabbage	Lettuce	Potato	S-Mushroom	Spinach	Welsh onion
				(a) Point estimates and s.e.				
$\hat{\gamma}_{l\text{FIML}}$	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)
$\hat{\rho}_l$	-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 $T_{ijl} = 0$ or 1	0.989	0.962	0.990	0.990	0.981	0.994	0.994	0.996
PCP for $T_{ijl} = 0$	0.995	0.976	0.995	0.996	0.997	0.997	0.998	0.999
PCP for $T_{ijl} = 1$	0.856	0.642	0.874	0.865	0.612	0.903	0.902	0.911
$\hat{\gamma}_{l\text{OLS}}$ with predicted q_{ijl}	0.059	0.113	0.062	0.068	0.040	0.018	0.085	0.063
$\hat{\gamma}_{l\text{OLS}}$	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 selection and price differential equations (6) and (7).
Note 2: "Pcp" represents the "percent correctly predicted."

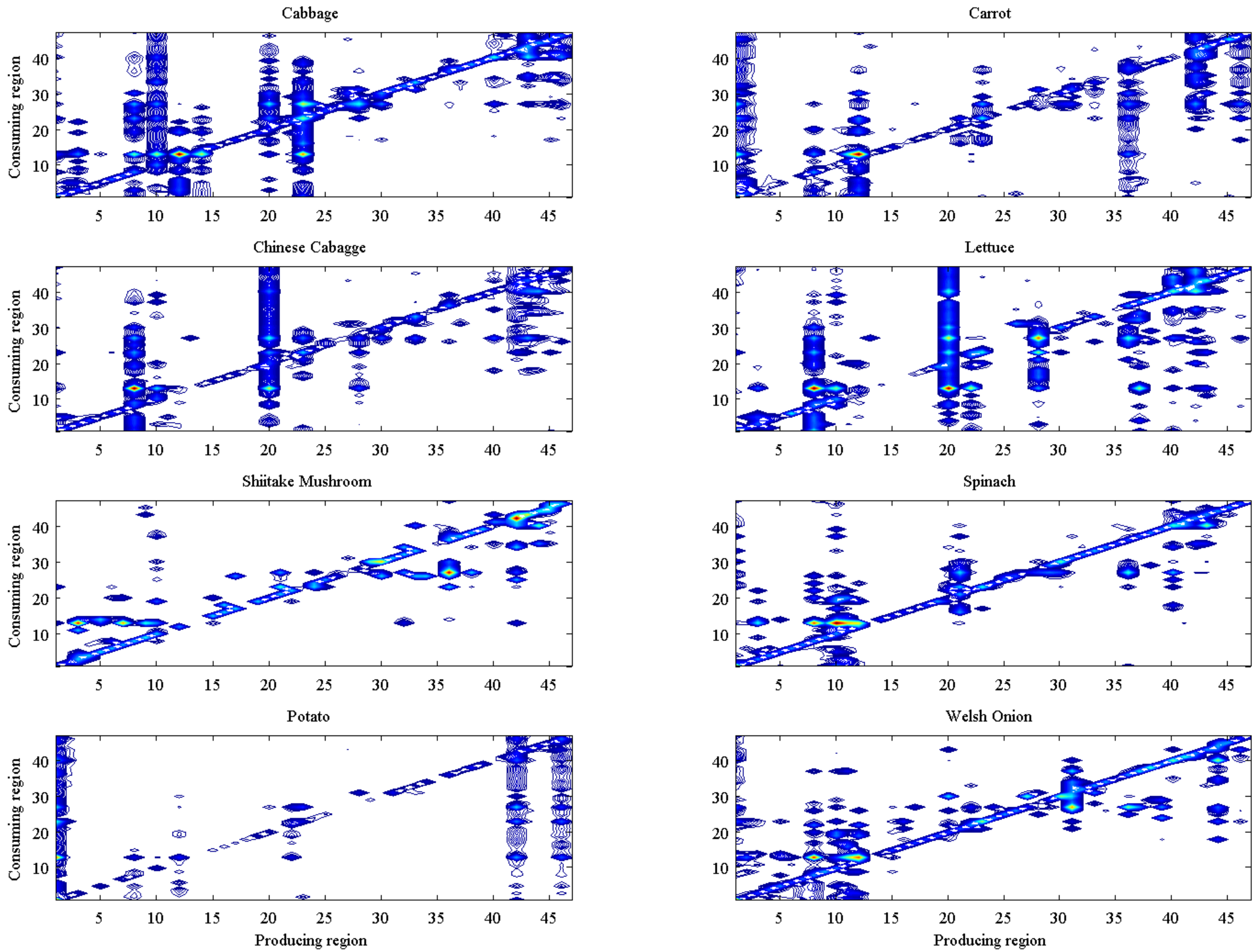

Figure 1: Data Frequencies of Product Delivery. Note. Each window plots data frequencies of product delivery of the corresponding vegetable from producing to consuming prefectures as contour lines. In each window, the horizontal line represents producing prefectures and the vertical line consuming prefectures. The lighter the blue contour line is, the higher the probability of product delivery is. If product delivery is perfectly local, contour lines are concentrated along the 45 degree line.

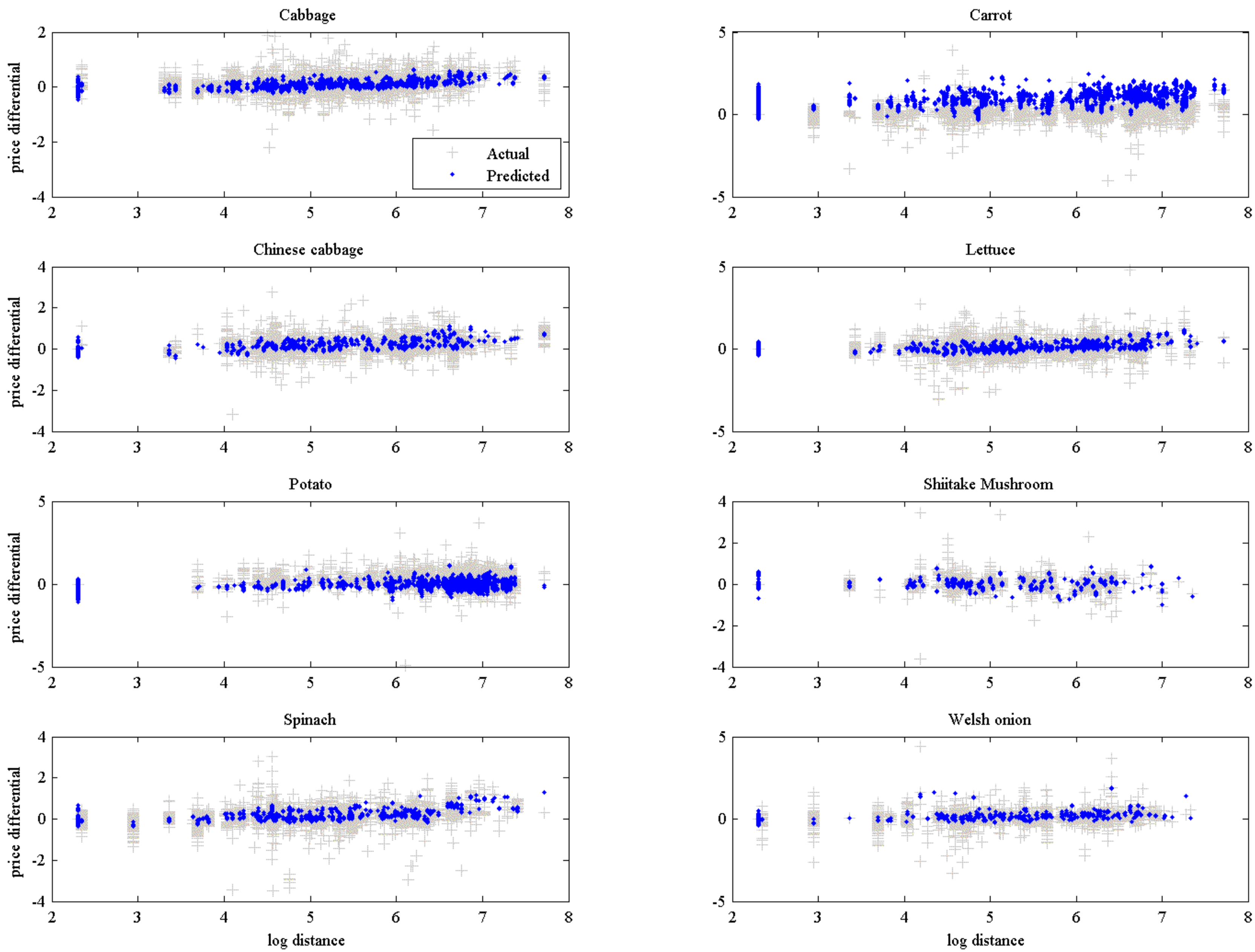

Figure 2: Predicted and Actual Price Differentials. Note. Each window plots the predicted and the actual price differentials on the logarithm of distance for the corresponding vegetable. In each window, the blue dots represent the predicted price differentials, while the gray crosses the actual price differentials.

Appendix:
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
Hitotsubashi University
Naka 2-1, Kunitachi
Tokyo, 186-8601, Japan
Tel: +81-42-580-8283
Fax: +81-42-580-8283
Email: tkano@econ.hit-u.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

May 6, 2012

This appendix is not intended for publication.

Appendix A. A structural sample-selection model with endogenous explanatory variables in selection

A.1. Derivation of the conditional likelihood function

Consider the structural sample-selection model in the main text

$$q_{ij} = \mu + \gamma d_{ij} + b_{ij}^{p} + u_{ij}, \text{ if } T_{ij} = 1, \tag{A.1}$$

$$T_{ij} = \mathbf{1}[\beta - (\epsilon - 1)\gamma d_{ij} + [\epsilon \ 1]\mathbf{y}_i + b_{ij}^{s} + \eta_{ij} > 0] \tag{A.2}$$

$$\mathbf{y}_i = \Gamma \mathbf{s}_i + \mathbf{e}_i. \tag{A.3}$$

To economize notation, we ignore product index l in this appendix. Eqs. (A.1) and (A.2) are the price differential equation and the selection equation, respectively. In this paper, we allow for a possibility that the explanatory variables in the selection equation, $\mathbf{y}_i = [\ln p_i \ \ln x_i]'$, are correlated with the error term of the selection equation, η_{ij}. As shown in eq. (A.3), we assume that $\mathbf{y}_i$ is linearly related to a vector of exogenous variables (i.e., instruments), $\mathbf{s}_i$, up to i.i.d. mean zero random vector $\mathbf{e}_i$. An endogeneity problem then is the case when η_{ij} is correlated with $\mathbf{e}_i$. More specifically, we assume that random vector $\epsilon_{ij} = [\mathbf{e}_i' \ u_{ij} \ \eta_{ij}]'$ is stochastically governed by a joint normal density with the mean of zero and the 4×4 symmetric positive-definite variance-covariance matrix Ω

$$\Omega = \begin{bmatrix} \Omega_{11} & \varphi_u' & \varphi_\eta' \\ \varphi_u & \sigma_u^2 & \sigma_{u\eta} \\ \varphi_\eta & \sigma_{u\eta} & \sigma_\eta^2 \end{bmatrix},, \tag{A.4}$$

where Ω_{11} is an 2×2 matrix, φ_u and φ_η are 1×2 row vectors, respectively. The endogeneity problem then occurs when $\varphi_\eta \neq \mathbf{0}$.

Given the above sample-selection model, we characterize the corresponding conditional likelihood $f(q_{ij}, T_{ij}, \mathbf{y}_i | d_{ij}, \mathbf{s}_i) \equiv f(q_{ij}, T_{ij} = 1, \mathbf{y}_i | d_{ij}, \mathbf{s}_i)^{T_{ij}} f(q_{ij}, T_{ij} = 0, \mathbf{y}_i | d_{ij}, \mathbf{s}_i)^{1-T_{ij}}$. Since q_{ij} is observed only when $T_{ij} = 1$, we can factorize the density $f(q_{ij}, T_{ij} = 1, \mathbf{y}_i | d_{ij}, \mathbf{s}_i)$ as follows:

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

where the second line results from the Bayes' rule and the third line comes from the fact that $f(\mathbf{y}_i | T_{ij} = 1, d_{ij}, \mathbf{s}_i) = f(\mathbf{y}_i | d_{ij}, \mathbf{s}_i)$. When $T_{ij} = 0$, we cannot observe q_{ij}. In this case, we factorize the conditional density $f(q_{ij}, T_{ij} = 0, \mathbf{y}_i | d_{ij}, \mathbf{s}_i)$ as

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

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

To accomplish this task, we first figure out the conditional densities of the error terms, $f(\eta_{ij}|\mathbf{e}_i)$, $f(\eta_{ij}|\mathbf{e}_i,u_{ij})$, and $f(u_{ij}|\mathbf{e}_i)$ by conducting the triangular factorization of the variance-covariance matrix Ω (A.4). We then obtain $\Omega = ADA'$ in which

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

for $H_{22}=\sigma_u^2-\varphi_u\Omega_{11}^{-1}\varphi_u'$, $H_{23}=H_{32}=\sigma_{u\eta}-\varphi_u\Omega_{11}^{-1}\varphi_\eta'$, and $H_{33}=\sigma_\eta^2-\varphi_\eta\Omega_{11}^{-1}\varphi_\eta'$. Let a new random vector $\tilde\epsilon_{ij}\equiv[\tilde{\mathbf{e}}_i{}'\ \tilde u_{ij}\ \tilde\eta_{ij}]'$ denote $A^{-1}\epsilon_{ij}$. The above triangular factorization implies that $\tilde\epsilon_{ij}$ 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:

$$\mathbf{e}_i=\tilde{\mathbf{e}}_i, \tag{A.5}$$
$$u_{ij}=\varphi_u\Omega_{11}^{-1}\tilde{\mathbf{e}}_i+\tilde u_{ij}, \tag{A.6}$$
$$\eta_{ij}=\varphi_\eta\Omega_{11}^{-1}\tilde{\mathbf{e}}_i+H_{32}H_{22}^{-1}\tilde u_{ij}+\tilde\eta_{ij}. \tag{A.7}$$

To derive the conditional density $f(\eta_{ij}|\mathbf{e}_i)$, let $\tilde a_{ij}$ denote $H_{32}H_{22}^{-1}\tilde u_{ij}+\tilde\eta_{ij}$. Notice that $\tilde a_{ij}$ is normally distributed with the mean of zero and the variance of H_{33}.[1] Using this fact, we can rewrite equation (A.7) as

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

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}.

$$\eta_{ij}|\mathbf{e}_i\sim N(\varphi_\eta\Omega_{11}^{-1}\mathbf{e}_i,H_{33}).$$

Following the similar process, we can characterize the conditional densities $f(\eta_{ij}|\mathbf{e}_i,u_{ij})$ and $f(u_{ij}|\mathbf{e}_i)$ from the corresponding conditional distributions

$$\eta_{ij}|\mathbf{e}_i,u_{ij}\sim N(\varphi_\eta\Omega_{11}^{-1}\mathbf{e}_i+H_{32}H_{22}^{-1}(u_{ij}-\varphi_u\Omega_{11}^{-1}\mathbf{e}_i),\ H_{33}-H_{32}H_{22}^{-1}H_{23}),$$

and

$$u_{ij}|\mathbf{e}_i\sim N(\varphi_u\Omega_{11}^{-1}\mathbf{e}_i,\ H_{22}).$$

Now we can characterize conditional probability $\mathbf{P}(T_{ij}=0|\mathbf{y}_i,d_{ij},\mathbf{s}_i)$ as

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

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

where $\Phi(.)$ is the standard normal cumulative distribution and

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

We can show conditional probability $\mathbf{P}(T_{ij} = 1 | q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i)$ as

$$\begin{aligned}
&\mathbf{P}(T_{ij} = 1 | q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i) \\
&= \mathbf{P}(\beta - (\epsilon - 1)\gamma d_{ij} + [\epsilon\ 1]\mathbf{y}_i + b^s_{ij} + \eta_{ij} > 0 | q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\
&= \mathbf{P}(\eta_{ij} > -\beta + (\epsilon - 1)\gamma d_{ij} - [\epsilon\ 1]\mathbf{y}_i - b^s_{ij} | q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\
&= \mathbf{P}(\tilde{\eta}_{ij} > -\beta + (\epsilon - 1)\gamma d_{ij} - [\epsilon\ 1]\mathbf{y}_i - b^s_{ij} - \varphi_\eta \Omega_{11}^{-1}\mathbf{e}_i - H_{32}H_{22}^{-1}(u_{ij} - \varphi_u \Omega_{11}^{-1}\mathbf{e}_i) | q_{ij}, \mathbf{y}_i, d_{ij}, \mathbf{s}_i), \\
&= \Phi(\kappa_{ij}),
\end{aligned}$$

where

$$\kappa_{ij} = \frac{\beta - (\epsilon - 1)\gamma d_{ij} + [\epsilon\ 1]\mathbf{y}_i + b^s_{ij} + (\varphi_\eta - H_{32}H_{22}^{-1}\varphi_u)\Omega_{11}^{-1}(\mathbf{y}_i - \Gamma\mathbf{s}_i) + H_{32}H_{22}^{-1}(q_{ij} - \mu - \gamma d_{ij} - b^p_{ij})}{(H_{33} - H_{32}H_{22}^{-1}H_{23})^{\frac{1}{2}}}.$$

To find out conditional density $f(q_{ij} | \mathbf{y}_i, d_{ij}, \mathbf{s}_i)$, consider conditional distribution

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

where

$$\varpi_{ij} = \frac{q^* - \mu - \gamma d_{ij} - b^p_{ij} - \varphi_u \Omega_{11}^{-1}(\mathbf{y}_i - \Gamma\mathbf{s}_i)}{H_{22}^{\frac{1}{2}}}.$$

We then can construct conditional density $f(q_{ij} | \mathbf{y}_i, d_{ij}, \mathbf{s}_i)$ taking a derivative of $F_{q_{ij}|\mathbf{y}_i, d_{ij}, \mathbf{s}_i}(q^*)$ with respect to q^* and evaluating the result at q_{ij}:

$$f(q_{ij} | \mathbf{y}_i, d_{ij}, \mathbf{s}_i) = H_{22}^{-\frac{1}{2}}\phi(\varpi_{ij})$$

where $\phi(.)$ is the standard normal kernel. Finally, it is straight-forward for us to characterize conditional density $f(\mathbf{y}_i | d_{ij}, \mathbf{s}_i)$. Since the vector of the endogenous explanatory variables, $\mathbf{y}_i$, is always observable regardless of the value of T_{ij}, conditional density $f(\mathbf{y}_i | d_{ij}, \mathbf{s}_i)$ is simply characterized by eq. (A.3) as a Gaussian joint density with the mean of $\Gamma\mathbf{s}_i$ and the variance-covariance matrix of Ω_{11}:

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

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

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

Following the conventional identification exercise with a sample selection model, we normalize the standard deviation of the error term of the selection equation to one: $\sigma_\eta = 1$. We also assume that u_{ij} is uncorrelated with $\mathbf{e}_i$: $\varphi_u = \mathbf{0}$. The likelihood function then turns out to be much simpler:

$$f(q_{ij}, T_{ij}, \mathbf{y}_i | d_{ij}, \mathbf{s}_i) = [\Phi(\kappa_{ij}) \sigma_u^{-1} \phi(\varpi_{ij})]^{T_{ij}} [1 - \Phi(\lambda_{ij})]^{1-T_{ij}} f(\mathbf{y}_i | d_{ij}, \mathbf{s}_i), \tag{A.8}$$

where

$$\kappa_{ij} = \frac{\beta - (\epsilon - 1) d_{ij} + [\epsilon \ 1] \mathbf{y}_i + b_{ij}^s + \varphi_\eta \Omega_{11}^{-1} (\mathbf{y}_i - \Gamma \mathbf{s}_i) + \rho \sigma_u^{-1} (q_{ij} - \mu - \gamma d_{ij} - b_{ij}^p)}{(1 - \varphi_\eta \Omega_{11}^{-1} \varphi_\eta' - \rho^2)^{\frac{1}{2}}},$$
$$\varpi_{ij} = \frac{q_{ij} - \mu - \gamma d_{ij} - b_{ij}^p}{\sigma_u},$$
$$\lambda_{ij} = \frac{\beta - (\epsilon - 1) \gamma d_{ij} + b_{ij}^s + [\epsilon \ 1] \mathbf{y}_i + \varphi_\eta \Omega_{11}^{-1} (\mathbf{y}_i - \Gamma \mathbf{s}_i)}{(1 - \varphi_\eta \Omega_{11}^{-1} \varphi_\eta')^{\frac{1}{2}}},$$

for the correlation coefficient between u_i and η_i, ρ.

A.2. The inverse-Mills ratio

The sample selection model predicts the price differential by the conditional expectation

$$E[q_{ij} | ., T_{ij} = 1] = \mu + \gamma d_{ij} + E[u_{ij} | ., T_{ij} = 1].$$

To understand the conditional expectation of the error term, $E[u_{ij} | ., T_{ij} = 1]$, consider a random vector $\omega_{ij} = [\mathbf{e}_i' \ \eta_{ij} \ u_{ij}]'$ that is stochastically governed by a joint normal density with the mean of zero and the 4×4 symmetric positive-definite variance-covariance matrix Σ

$$\Sigma = \begin{bmatrix} \Omega_{11} & \varphi_\eta' & \varphi_u' \\ \varphi_\eta & \sigma_\eta^2 & \sigma_{u\eta} \\ \varphi_u & \sigma_{u\eta} & \sigma_u^2 \end{bmatrix}.$$

Covariance matrix Σ has a triangular factorization $\Sigma = BQB'$ such that

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

where $J_{22} = \sigma_\eta^2 - \varphi_\eta \Omega_{11}^{-1} \varphi_\eta'$, $J_{23} = J_{32} = \sigma_{u\eta} - \varphi_\eta \Omega_{11}^{-1} \varphi_u'$, and $J_{33} = \sigma_u^2 - \varphi_u \Omega_{11}^{-1} \varphi_u'$. Define a new random vector $\tilde{\omega}_{ij} = [\tilde{\mathbf{e}}_i' \ \tilde{\eta}_{ij} \ \tilde{u}_{ij}]' = B^{-1} \omega_{ij}$. The above triangular factorization implies that new vector $\tilde{\omega}_{ij}$ is normally distributed with the mean of zero and the diagonal variance-covariance matrix of Q. Then, by construction, we can obtain the following system of equations:

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

Under the assumption of $\varphi_u = \mathbf{0}$ and the normalization of $\sigma_\eta = 1$, the above equations imply that

$$u_{ij} = \sigma_{u\eta}(1 - \varphi_\eta \Omega_{11}^{-1} \varphi'_\eta)^{-1}(\eta_{ij} - \varphi_\eta \Omega_{11}^{-1} \mathbf{e}_i) + \tilde{u}_{ij},$$

and conditional expectation $E[u_{ij}|., T_{ij} = 1]$ is

$$\begin{aligned}
E[u_{ij}|., T_{ij} = 1] &= \sigma_{u\eta}(1 - \varphi_\eta \Omega_{11}^{-1} \varphi'_\eta)^{-1} E[\eta_{ij} - \varphi_\eta \Omega_{11}^{-1} \mathbf{e}_i|., T_{ij} = 1], \\
&= \sigma_{u\eta}(1 - \varphi_\eta \Omega_{11}^{-1} \varphi'_\eta)^{-1} E[\tilde{\eta}_{ij}|., T_{ij} = 1]. \qquad \text{(A.9)}
\end{aligned}$$

Conditional expectation $E[\tilde{\eta}_{ij}|., T_{ij} = 1]$ is then given as

$$\begin{aligned}
E[\tilde{\eta}_{ij}|., T_{ij} = 1] &= E[\tilde{\eta}_{ij}|\eta_{ij} > -\beta + (\epsilon - 1)\gamma d_{ij} - [\epsilon\ 1]\mathbf{y}_i - b_{ij}], \\
&= E[\tilde{\eta}_{ij}|\varphi_\eta \Omega_{11}^{-1} \tilde{\mathbf{e}}_i + \tilde{\eta}_{ij} > -\beta + (\epsilon - 1)\gamma d_{ij} - [\epsilon\ 1]\mathbf{y}_i - b_{ij}], \\
&= E[\tilde{\eta}_{ij}|\tilde{\eta}_{ij} > -\beta + (\epsilon - 1)\gamma d_{ij} - [\epsilon\ 1]\mathbf{y}_i - b_{ij} - \varphi_\eta \Omega_{11}^{-1} \tilde{\mathbf{e}}_i], \\
&= E[\tilde{\eta}_{ij}|\tilde{\eta}_{ij} > -\hat{z}_{ij}], \\
&= \phi(\hat{z}^*_{ij})/\Phi(\hat{z}^*_{ij}). \qquad \text{(A.10)}
\end{aligned}$$

where $\hat{z}^*_{ij} = \beta - (\epsilon - 1)\gamma d_{ij} + [\epsilon\ 1]\mathbf{y}_i + b_{ij} + \varphi_\eta \Omega_{11}^{-1}(\mathbf{y}_i - \Gamma \mathbf{s}_i)$. Equation (A.10) is the corresponding inverse-Mills ratio.

Appendix B. A Monte Carlo experiment with a linear economy

In this section, we conduct Monte Carlo experiments based on our model in section 2 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.[2] 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 an identical product under productivity level a_l that is set equal to 1.00. The parameter

[2]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.

of demand function (1), α, 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 and x_i, respectively, both of which we set to 20.00. For simplicity, we ignore the cross-regional variations in the productivity-regional bias ratio θ_{ij} by setting $\delta_{ij} = 1.00$ 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 $(1-\alpha)\alpha^{\epsilon-1}c_j^{-\epsilon}p_i^{\epsilon}x_i$ 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} and latent variable z_{ij} 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} on the log of the distance 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} on the log of the distance d_{ij} using only the samples that are selected with $T_{ij} = 1$. 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 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., $\ln \tilde{p}_i - \ln \tilde{p}_k$ for any two consuming regions i and k, where $\tilde{p}_i$ 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, $|\ln \tilde{p}_i - \ln \tilde{p}_k|$, on the log of the corresponding distance d_{ik}.[3] 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 B.1 depict the contour plots of the probabilities of delivery from producing

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

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 B.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.

Figure B.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.

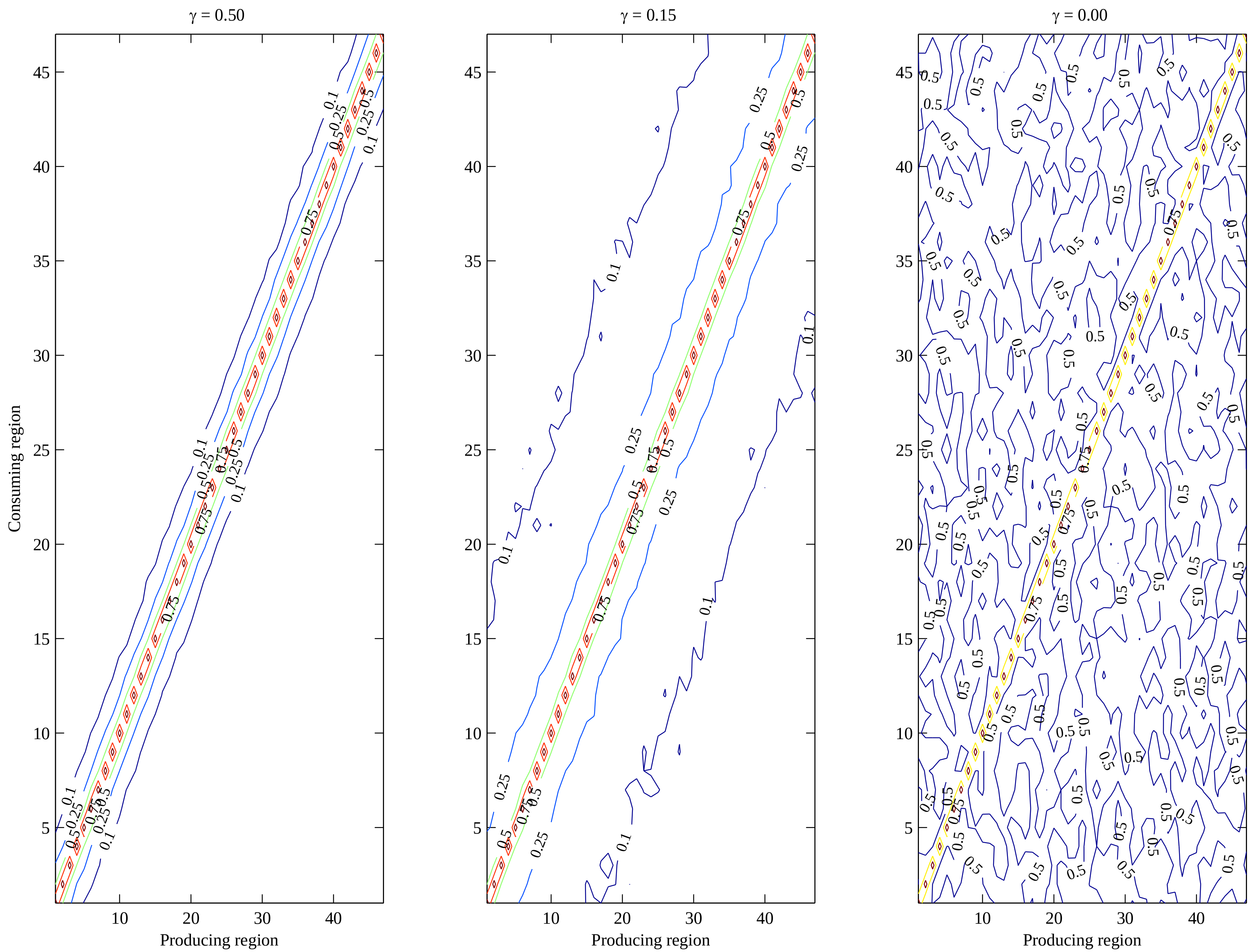

Figure B.1: Simulated probabilities of product delivery

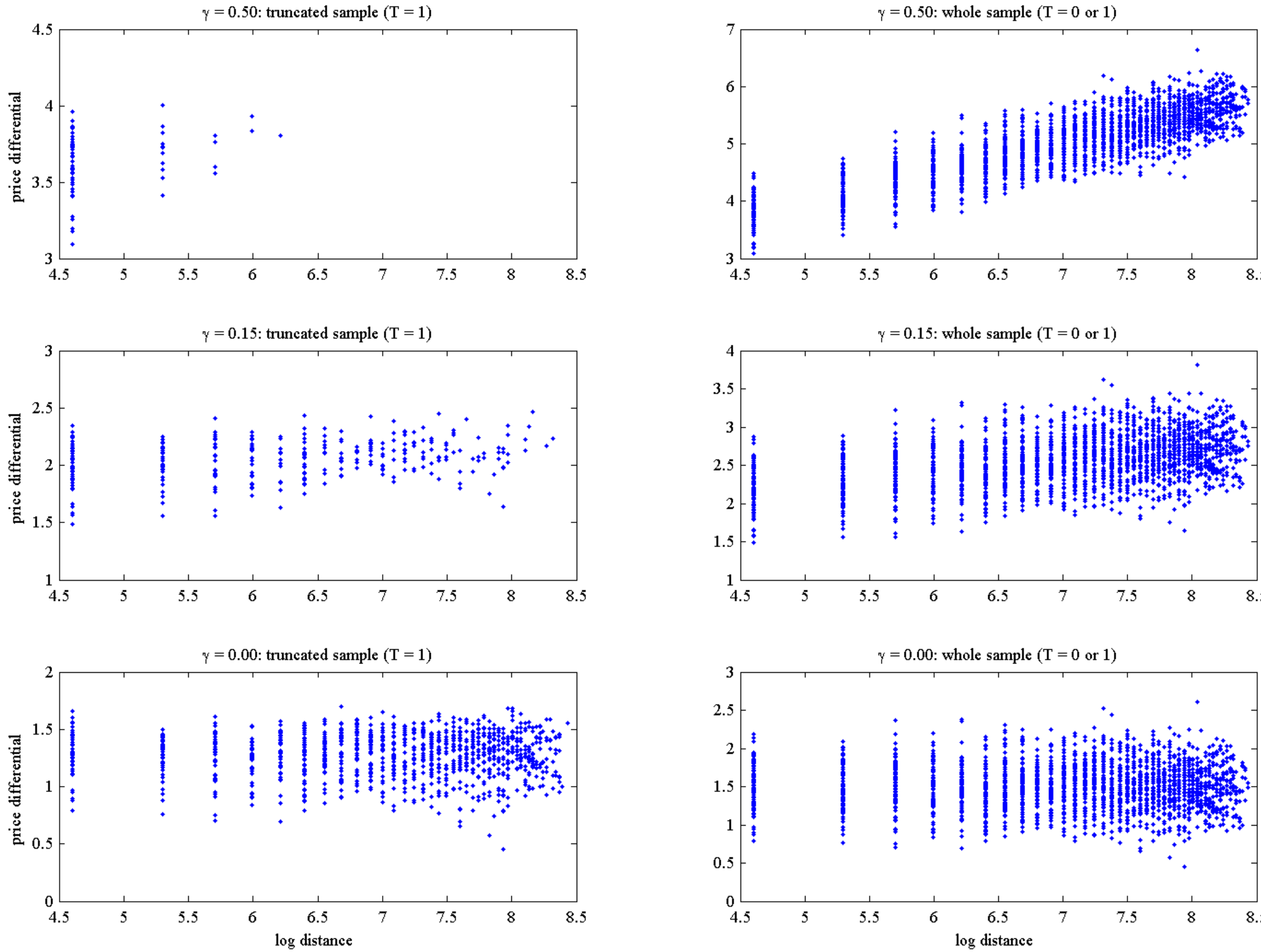

Figure B.2: Simulated price differentials

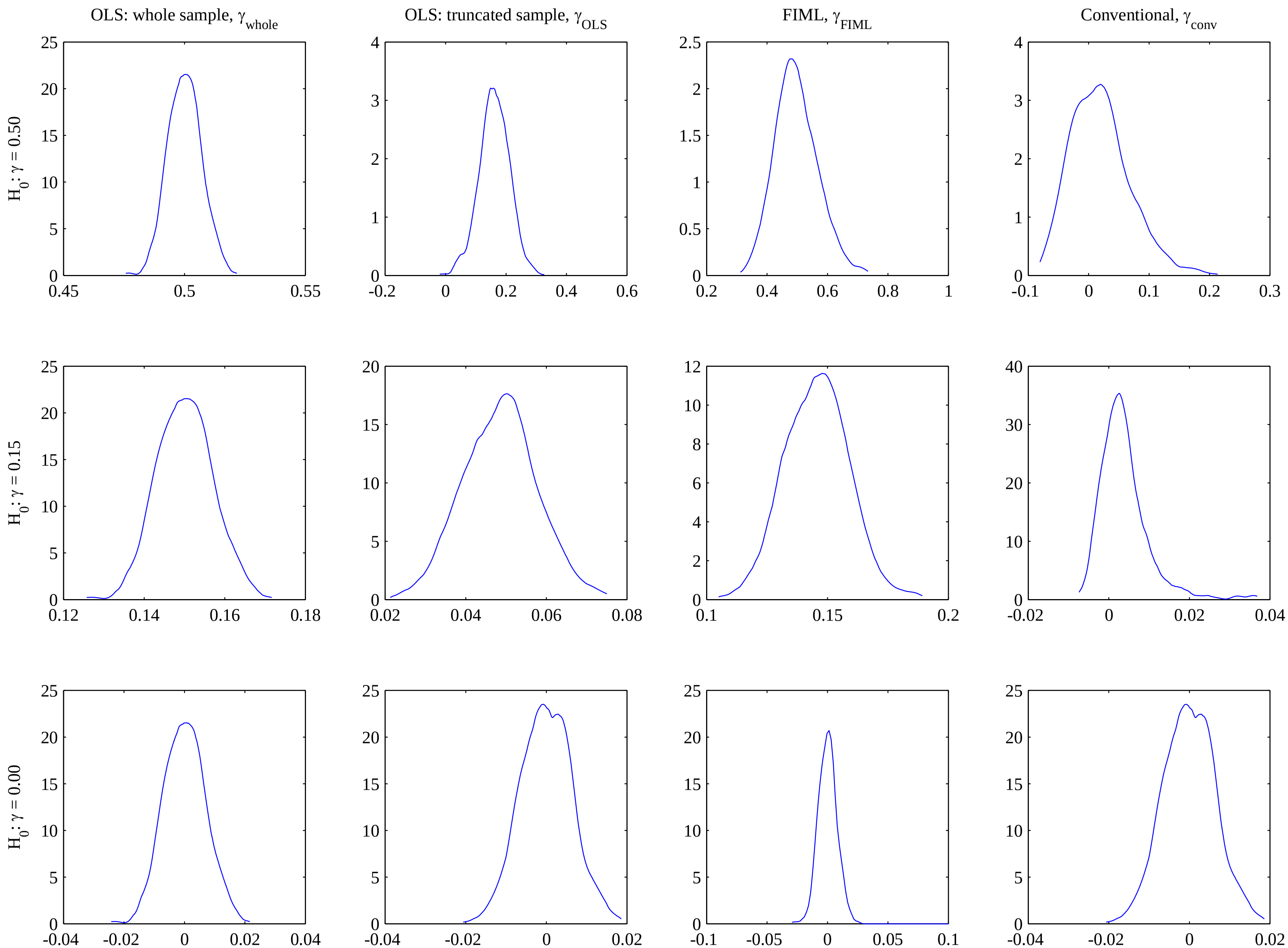

Figure B.3: Kernel-smoothed densities of distance elasticity estimators