The Quality of Distance : Quality Sorting, the Alchian-Allen Effect, and Geography

TAKECHI, Kazutaka

The Quality of Distance:

Quality Sorting, the Alchian-Allen Effect, and Geography

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

Abstract

Either quality sorting or the presence of a specific cost (the so-called Alchian-Allen effect) is considered to be the main mechanism for the positive relationship between product quality and the distance to market. However, the reduced-form regressions found in the literature generally fail to reveal which of these two mechanisms is (or even whether both are) the main driving force. In this study, we employ unique Japanese individual goods price data to identify separately the effects of quality sorting and specific costs. Our empirical analysis shows that while high-cost producers produce high-quality goods, as suggested in Baldwin and Harrigan (2011), the quality-sorting mechanism solely is not sufficiently strong to account for the purported positive link between quality and distance. Moreover, we do find that the technology parameter that relates costs to quality is overestimated in the absence of specific costs. On this basis, we confirm that the presence of specific costs is significant, which may generate the positive relationship between quality and distance. We also find that the specific-cost components in transport costs are more distance elastic than any ad valorem components, a finding qualitatively consistent with the trade cost specification in Hummels and Skiba (2004). Finally, our results are robust with respect to various measures of distance and specification.

Key Words : *Quality sorting; Transport costs; Specific costs; Geographic barriers; Producer heterogeneity*

JEL Classification Number : F11, F14, F41

1. Introduction

Are the markets for high-quality goods more remote than for low-quality goods? The response of many studies appears to be in the affirmative (Bastos and Silva (2010), Baldwin and Harrigan (2011), Manova and Zhang (2012), Martin (2012)). However, while this positive relationship between the quality of goods and the distance to market results only from simple observation, it does lead to a more primary concern when evaluating trade models and the specification of trade costs. This is because this empirical relationship is inconsistent with the prediction of standard firm heterogeneity models in the absence of a quality dimension and the specification of iceberg-type trade costs. Therefore, we need to incorporate novel elements into our modelling, in the form of quality sorting and the presence of specific trade costs, to reconcile the available empirical and theoretical evidence.

A quality-sorting mechanism introduces quality into standard firm heterogeneity models. Because high-quality products are also highly profitable, they can overcome the significant trade costs associated with long distances to market. In contrast, in standard firm heterogeneity trade models, as distance increases, only highly productive, hence low-cost, firms can provide supply. Because low-cost producers are able to set lower prices, when measuring quality the average free on board (FOB) price, the FOB price is typically lower in distant markets, which is not what the pattern of observed data suggests. Hence, it is necessary to incorporate quality in a firm-heterogeneity model to account for the supposed positive relationship between quality and distance (Baldwin and Harrigan, 2011).

The presence of specific costs also account for the positive relationship between good quality and distance to market. The relative prices of high quality, and therefore higher-priced goods, are lower in distant markets when there are specific costs in trade. Hence, the relative demand for high-quality goods is also high in these markets. This enables firms producing high-quality goods to ship to these more distant markets, a process referred to as the Alchian–Allen effect (Hummels and Skiba, 2004). Importantly, this change in relative prices does not arise under iceberg-type trade costs.

However, because of data limitations, to our knowledge, the Alchian–Allen and quality-sorting effects have not been jointly analyzed using individual pricing data. In the literature, the FOB price (the unit value) of export goods is regressed on the distance to market. Unit value is then the measure of quality. Unfortunately, in most cases, no data on trade costs are available. Because the Alchian–Allen effects concern the specification of the trade cost function, to identify the impacts of the quality-sorting and Alchian–Allen effects, we need to link quality, trade costs, and distance separately. However, in the absence of trade cost data, we could erroneously attribute variations in quality to distance, not to trade costs. Thus,

with the exception of Hummels (2001) and Hummels and Skiba (2004), such an identification remains undone because trade cost information is usually unavailable. The contribution of this study to the literature is then to analyze the quality-sorting and Alchian–Allen effects jointly and identify these effects separately.

In the recent literature, several studies incorporate specific cost components in trade costs and assess their size and impact. For instance, Irarrazabal et al. (2013) show that the size of specific costs is large and significant, while Khandelwal et al. (2013) use specific costs to model quotas, which affect firm behavior in a different way from an ad-valorem cost reduction. While our study shares a common motivation concerning the impact of specific costs, our focus is slightly different, which is the identification of the impact of distance on ad-valorem and specific costs.

In this paper, we first follow Anderson and van Wincoop's (2004) suggestion for use of the price of production (at the source or origin). The use of source and market price data enables us to measure trade costs because there is actual delivery between these areas. As examples of the use of origin information, Donaldson (2013) uses salt price data in India, Atkin and Donaldson (2014) employ price data in Ethiopia and Nigeria, for which source prices are also available, and Kano et al. (2013) use wholesale vegetable price data in Japan, including a detailed description that allows the identification of identical products in different locations. Because price differentials reflect both ad-valorem and specific costs, it remains necessary to identify these costs separately. Then by utilizing the monotonic relationship between price and quality arising from the optimal price formula, we are able to obtain information on quality and production costs from the price data. Because variable costs consist of ad-valorem trade costs multiplied by production and specific costs, derived production costs enable us to separate ad-valorem costs from specific costs.

There is also an additional identification problem in that if transport is too costly, even high-quality goods may not be supplied to distant markets. This self-selection bias is absent in most of the literature, with the exception of Kano et al. (2013, 2014), and may serve to create an under biased distance effect. To overcome this, we employ unique micro data on agricultural product (vegetable) prices in Japan. As in Kano et al. (2013, 2014), this data set contains market and origin prices, and information on the region where a product is produced. Thus, we can establish product delivery patterns and take into account selection bias arising because of delivery choices.

The analysis in this paper begins with reduced-form regressions as in the existing literature. Our origin price is approximately equivalent to a FOB price in the literature, which is used to measure product quality. Therefore, we first simply regress origin prices on distance to markets and find that our vegetable qualities are also positively associated with

3

the distance to market. We then estimate the structural model to obtain the ad-valorem and specific cost components separately. We use the origin price and markup formula to back out the cost of production and utilize the derived production cost to identify the ad-valorem and specific costs. Our estimations show that the specific cost component is more distance elastic than the ad-valorem component, which is qualitatively consistent with the specification adopted by Hummels and Skiba (2004). The empirical analysis also shows that the technology parameter connecting production costs and quality is positive (high-cost producers produce high-quality goods). However, the magnitude of the increase in quality associated with these costs alone is not sufficient to account for the positive link between quality and distance, suggesting that the quality-sorting effect is weak. The presence of specific costs is then important for a positive relationship between quality and distance. In addition, the size of the technology parameter in the case of no specific costs is higher than when we consider specific costs. This suggests that in the absence of specific costs, the technology parameter is overestimated. Thus, our contribution is to detect not only the relationship between quality and distance, but also the technical relationship between quality and costs.

Existing studies, such as Irarrazabal et al. (2013), have also identified the significance of specific costs. The identification strategy in Irarrazabal et al. (2013) is to utilize the property that the presence of specific costs changes the demand elasticity. To identify this, Irarrazabal et al. (2013) estimate the size of the specific costs relative to the ad-valorem costs using the data variation in FOB (producer) prices and destinations (trade costs). Our study is notable in that we estimate the ad-valorem and specific components separately and then identify how these costs are sensitive to distance. Additionally, we also estimate the elasticity of substitution parameter and thus obtain the key parameters in the heterogeneous-quality model, including the dispersion of productivity, the elasticity of substitution, and the distance elasticity. As these determine the behavior of the heterogeneity model, our estimates then yield a benchmark for evaluating the implications of existing theoretical models.

Of course, our results relate in part to the characteristics of the data employed. In particular, we use price data for agricultural products. Thus, the reason for the rather weak effect of quality sorting in our analysis is that vegetable production is constrained by geographic conditions. While some farmers may produce high-quality goods using superior technology (e.g., greenhouses), farmer productivity is generally not associated with quality rather with costs. Thus, the demand side may matter more. Specific costs make the price of high-quality goods relatively low, creating relatively high demand in remote markets. Hence, the presence of specific costs in our model encourages farmers producing high-quality goods to deliver their product to distant markets.

The remainder of the paper is organized as follows. In Section 2, we discuss the reduced-form regressions representing the relationship between quality and distance. In Section 3, we set up a structural model for our estimations and conduct Monte Carlo exercises to demonstrate the bias in the standard model. Section 4 introduces our data set, and Section 5 details the specification of our model. Section 6 reports the estimation results, and Section 7 provides some robustness checks. In Section 8, we evaluate the welfare improvements associated with the reduction in trade costs using general equilibrium model simulations. The final section concludes the paper.

2. Reduced-Form Relationship

A positive relationship between FOB prices and distance has been obtained in a number of previous studies, including Bastos and Silva (2010), Baldwin and Harrigan (2011), Manova and Zhang (2012), and Martin (2012). This observation motivates the introduction of quality because it is not consistent with a standard firm heterogeneity model. In the standard heterogeneous model, high-productivity firms enter a market with high entry costs (e.g., high transport costs) and can set low prices, so there will be a negative relationship between FOB prices and distance to markets. The introduction of a quality dimension into the firm-heterogeneity model leads to the case where high-FOB-price firms produce high-quality goods, and therefore these firms sell to markets that are more distant. In this section, we conduct similar exercises using regional price data, which contain the price set in the origin market (the production site). After controlling for market-specific effects, the origin prices capture the quality of the product. Therefore, our empirical exercise is comparable to that in the literature.

We use vegetable wholesale price data for Japan. Because our data set includes detailed information about product characteristics, we can compare the prices of identical products. In Japan, vegetables trade in a wholesale market in each prefecture, so we can obtain the price in the production prefecture (the origin price) and the price in the market (the market price). We depict the key observation in the relationship between quality and the distance to market by plotting origin price and distance in Figure 1. We plot the log of distance on the horizontal axis and the log of the origin price on the vertical axis. All figures illustrate a positive relationship between distance and origin price. Thus, there is a positive relationship in our data set.

=== Figure 1 here ===

Next, we report the results of the reduced-form regressions. As in the extant litera-

ture, we regress the price at the source on the distance to the destination:

$$\ln p_{jj} = const + \ln D_{nj} + \eta_{nj}, \tag{1}$$

where p_{jj} is the price in region j, D_{nj} is the distance between origin j and the market n, and η_{nj} is the error term. Using OLS, we find that there is a positive relationship between quality and distance as reported in Table 1 and the fitted lines in the figure. Note that because this regression does not control for region-specific effects, this positive relationship may result from regional shocks. Thus, we conduct the same regressions after including origin- and market-specific effects. The estimates reflecting regional-specific effects also display a positive relationship.

=== Table 1 here ===

As discussed in the literature, several models can explain this positive link. Unfortunately, the results of the reduced-form regressions do not provide us with information about the structural parameters, such as distance elasticity. The purpose of this analysis is then to identify the important structural parameters in quality heterogeneity models.

3.   Model

We adopt a standard monopolistic competition, producer heterogeneity, product quality model following Baldwin and Harrigan (2011). An additional feature is the introduction of specific costs. Assume that there are I regions and in each region there is a continuum of producers whose mass is expressed by N_j.

A Cobb–Douglas CES utility function expresses the preferences of consumers in region n:

$$U_n = \left(\int_{z \in J_n} (c_{nj} q_{nj})^{(\sigma-1)/\sigma} dk\right)^{(\sigma/(\sigma-1))\mu} Z^{1-\mu}, \tag{2}$$

where J_n is a set of products delivered to region I, and Z is the consumption of numeraire goods. With the budget constraint, $Y_n \mu = \int p_{nj}(k) c_{nj}(k)$, the demand function is:

$$c_{nj}(k) = \frac{p_{nj}^{-\sigma}}{q_{nj}^{1-\sigma}} \frac{Y_n \mu}{P_n^{1-\sigma}}, \tag{3}$$

where $P_n = (\int (p_{nj}/q_{nj})^{1-\sigma})^{1/(1-\sigma)}$. This signifies that as the quality of goods improves, consumer demand increases. Quality then acts as a demand shifter in this setting.

We assume that producers produce a differentiated product, face local demand $x_{nj}(z)$, and maximize their profits. On the cost side, producers must pay labor and transportation

costs. The transportation costs consist of ad-valorem and specific costs. Thus, we express profits from market n with:

$$\pi_{nj} = p_{nj}x_{nj} - a_{nj}\tau_{nj}x_{nj} - t_{nj}x_{nj} - f_{nj}, \tag{4}$$

where τ_{nj} is the ad-valorem component, t_{nj} is the specific component in transportation costs, and a is the unit cost. Quality sorting implies that high-cost producers produce high-quality goods. We assume a monotonic relationship between quality and production costs:

$$q = f(a). \tag{5}$$

This is required for us to estimate the quality-sorting model. If the relationship between costs and quality is not monotonic—for example, a U-shaped relationship—we cannot identify the parameter that determines the quality-sorting pattern. We further assume a parametric form of $f(.)$. As in Baldwin and Harrigan (2012), we assume that producers decide their cost level, and the quality of their products is then a function of that cost level:

$$q = a^{1+\theta}. \tag{6}$$

Thus, if $\theta > -1$, then high-cost producers produce high-quality goods. If $\theta > 0$ and specific costs are zero, then high-cost producers will deliver their products to more remote markets than low-cost producers because the rate of quality improvement is greater than that of the increase in cost. This provides the mechanism for quality sorting: high-cost producers produce high-quality goods, so they are more profitable than low-quality producers and hence can reach more costly markets.

Producers facing the local demand function (2) maximize their profits by setting the optimal price in market n:

$$p_{nj} = \frac{\sigma}{\sigma - 1}(\tau_{nj}a + t_{nj}). \tag{7}$$

We assume that there are no interregional transportation costs for within-region trade:

$$p_{jj} = \frac{\sigma a}{\sigma - 1}. \tag{8}$$

Thus, by inverting the above price formula, we can express the cost level of the producer. Using this implied cost enables us to recover the quality level. In our data set, as we can observe the market price and the place of production, we can use the above relationship to identify the specific cost component separately from the ad-valorem component.[1]

---

[1]There is a slight difference between the FOB price and the source price. By definition, FOB price, p_{FOB}, satisfies the following equation: $p_{market} = \tau p_{FOB} + t$. Thus, $p_{FOB} = (\sigma/(\sigma - 1))(a + t/\sigma\tau)$. However, because the source price is the price set for the source market without trade costs, $p_{source} = (\sigma a/(\sigma - 1))$.

With regard to trade costs, the key idea is that by using source and market prices, we can measure trade costs using price data. We normalize interregional trade costs by local trade costs incurred for local delivery; thus, all trade costs are relative to the local cost of delivery. In addition, because price is a monotonic function of production costs, we can replicate costs using price data. Furthermore, given that the price differential function depends on distance and the interaction term between distance and costs, we can also identify the interaction term using the price data.

The price differentials between markets and sources are:

$$\frac{p_{nj}}{p_{jj}} = \tau_{nj} + \frac{1}{a}t_{nj}. \tag{9}$$

Hence, in the price differential equation, while we include the ad-valorem term in the equation directly, the specific component is interacted with the cost term. This serves to identify the ad-valorem and specific terms separately.

The above price differential equation is observed only when there is actual delivery from j to n. Thus, we need to consider the producer's delivery decision. The profit function is:

$$\pi_{nj} = \frac{(\frac{\sigma}{\sigma-1})^{1-\sigma}(\tau_{nj}a + t_{nj})^{1-\sigma}}{q_{nj}^{1-\sigma}}\frac{Y\mu}{\sigma P_n^{1-\sigma}} - f. \tag{10}$$

If profit is positive, there will be delivery from source j to market n. We construct a delivery decision variable, V_{nj}:

$$V_{nj} = [\frac{(\frac{\sigma}{\sigma-1})^{1-\sigma}(\tau_{nj}a + t_{nj})^{1-\sigma}}{q_{nj}^{1-\sigma}}\frac{Y\mu}{\sigma P_n^{1-\sigma}}]/f. \tag{11}$$

If $V > 1$, then there is delivery from j to n. As Irarrazabal et al. (2013) show, because of specific costs, even the lowest-cost producer ($a \approx 0$) earns finite profits. Thus, other than the above condition, there is a further selection condition; i.e., whether producer costs are sufficiently low to obtain profits to cover fixed costs. We assume that this condition holds in order to focus on the entry condition.

To close the general equilibrium model, we can assume that each consumer supplies one unit of labor for production, a numeraire good is produced using the unit of labor, and this is freely traded across regions. This ensures that the wage rate is equal to one and trade balance is attained. However, to focus on the identification of trade costs, we simply analyze individual producer behavior. Regional fixed effects in the estimations capture the general equilibrium effects. For explicit treatment of the general equilibrium effects, we conduct Monte Carlo exercises to reveal how large trade cost reductions increase welfare.

3.1   Illustration of Bias

We conduct Monte Carlo experiments based on the model in the previous section to demonstrate that the estimates using a model without specific costs account for the bias. We create a linear economy geographically separated into 47 regions on the integer line between 1 and 47 sequentially. This linear economy implies that the distance between regions i and j, d_{ij}, is equal to $|j - i|$ with a minimum distance of 1 and a maximum distance of 46.

We assume that the shape of the demand function is common across the regions and characterized by an elasticity of substitution parameter equal to 3.75. Because we focus on estimates using a model with regional fixed effects, each region is also characterized by an aggregate price and aggregate real expenditure, both of which we set to 20.00. For simplicity, we ignore the cross-regional variations in productivity. We assume that in each region, a product is produced with a productivity level equal to 0.99 and a factor cost set to 1. Gaussian random components appear in both the fixed cost and the trade costs. In the fixed costs, the random term has a standard deviation of 0.65. Idiosyncratic random variations in trade costs are captured by the standard deviation, which is 0.25.

In our Monte Carlo experiment, we first draw 100 sets of Gaussian random variables of fixed and trade cost components, u_{ij} and v_{ij} independently from their distributions. We then calculate the price differentials and the selection equation under the hypothesized value of the distance elasticity of trade costs, being 0.3 for the ad-valorem trade cost and 0.5 for the specific trade cost. In each Monte Carlo draw of the true value of the distance elasticity, we then implement our estimations of the distance elasticity. The first is the FIML estimation without specific costs and the second is the FIML with specific costs. By construction, the FIML estimation without specific costs suffers bias caused by misspecification. Because the trade cost associated with the specific component is captured by the ad-valorem component, the distance elasticity of the ad-valorem trade costs will be over biased. Similarly, because the presence of specific costs delivers high-quality goods to distant markets, the elasticity of quality with respect to costs also captures this effect. If this quality elasticity is high, high-quality products are highly profitable, and thus shipped to distant market. With specific costs, the distance elasticity of specific costs correctly estimates this Alchian–Allen effect. However, without specific costs, the positive relationship between quality and the distance to market will be included in the quality elasticity estimates.

Figure 2 reports the nonparametrically smoothed densities of the distance and quality elasticity estimates with the Gaussian kernel. The top panel corresponds to the model with specific cost and the bottom panel to that without specific costs. The figures in the top panel show that the estimates using the true model are consistent and distributed around the underlying true value. However, the figures in the bottom panel reveal that the estimates

using the model without specific costs are subject to sever over bias. As we have argued, while the true ad-valorem distance elasticity is 0.3, the median value of the estimates is 0.536. Similarly, while the true quality elasticity is -0.15, the median is 0.591. Hence, the Monte Carlo exercise confirms the necessity of incorporating a specific cost component for drawing correct inferences on the distance and quality elasticities.

=== Figure 2 here ===

4. Data

We conduct our empirical research using the product-level data. We employ a daily data set of the wholesale prices of agricultural products in Japan, known as "the Daily Wholesale Market Information of Fresh Vegetables and Fruits."[2] This daily market survey reports the wholesale prices and quantities sold of some 120 different fruits and vegetables. We use the 2007 report representing 274 market-opening days.

The main advantage of the data set is in including information about individual product characteristics and a detailed categorization, such that each vegetable is classified by brand, size, grade, and source region. For example, the cabbage category typically includes "cabbage," "red cabbage," and "spring cabbage." Our data set then reports that cabbages of size "6" and grade "syu (excellent)" produced in Aichi Prefecture traded in the Aichi and Tokyo markets on July 1, 2007. As also shown, the price of this type of cabbage is 31.5 yen per kilogram in Aichi and 36.8 yen in Tokyo. Thus, we can calculate the price differential between these two locations, which may reflect the trade costs between two prefectures.

Comparing the prices in different locations to infer trade costs is meaningful if the goods are identical as in the law of one price (LOP) literature, and in fact, the prices of these goods are comparable. As discussed, our data have a high degree of categorization, which is useful for our purpose of assessing our hypothesis. Furthermore, because our data represent information on agricultural products, goods can differ depending on the date of production. However, we do not have exact information on the production date, so we assume that these goods are different when the trading dates are different. Thus, while this represents a slight shortcoming, the information in our data set provides us with the identification of an identical product in terms of many aspects of product characteristics.

The price differential, q_{nj}, that reflects trade costs is obtained by subtracting the wholesale price in source prefecture j, p_{jj}, from that in consuming prefecture n, p_{nj}.[3] The

---

[2] Our data set is identical to that employed in Kano et al. (2013, 2014).

[3] All of the products are sold in markets, but not necessarily in their markets of origin. In this case, when we cannot observe both the market and source prices, we eliminate these product entries.

price in source, p_{jj}, is the price when we observe product l being delivered from the producer to the wholesale market in the source prefecture. If this product is also shipped to market n, then p_{nj} is also observed. Thus, we set $T_{nj} = 1$ for pair (n, j) if we can calculate the price differentials, q_{nj}.

With regard to the distance between regions, we define interprefectural distance as the direct distance between prefectural head offices in the prefectural capital cities. We set the internal distance to 10 km, because the minimum interprefectural distance is 10.4 km (Kyoto–Shiga), and therefore we set the internal distance shorter than the minimum interprefectural distance. In a later section, we use the Head and Mayer (2000) internal distance formula as a robustness check. Natural conditions, not only market conditions, may affect regional prices. For example, preferences and the production of vegetables may change according to the air temperature. Thus, we use daily temperature data for the market and origin to control for these daily variations. As these are exogenous variables, they will also be helpful for identification of our selection models.

=== Table 2 here ===

We focus our exercise on three vegetables; namely, cabbage, Chinese cabbage (c-cabbage, hereafter), and lettuce. As discussed in Section 2, these are the vegetables that are priced higher in the source region and shipped to more distant markets. Table 2 summarizes several descriptive statistics for these products, indicating that each product is highly categorized by product variety, size, and grade. The number of distinct product entries is quite large: 1,207 for cabbage, 1,001 for c-cabbage, and 903 for lettuce. We assume that these products differ when the trading date changes, so to a certain degree, our price differential data are the price differentials of identical products. The average prices are 77.833 yen for cabbage, 61.628 for c-cabbage, and 183.909 for lettuce. There are also market prices in the data. Because we use origin prices to measure quality, Table 1 also reports the prices at the origin. The average origin prices are 67.431, 50.671, and 168.855 for cabbage, c-cabbage, and lettuce, respectively. Thus, market prices outside the origin region are higher than in the origin region. This is primarily because it is costly to ship goods to distant markets. Because we consider the truck transportation market competitive, we do not need to consider markups in the transport sector, unlike Hummels et al (2009). Our purpose is to address how much these price differentials reflect the shipping of high-quality goods to distant markets. Estimating a trade model to identify the key parameters should provide us with an answer to this question.

To understand the behavior of product shipment, we count the number of delivery $T_{ijl} = 1$ and nondelivery $T_{ijl} = 0$ cases. We identify product delivery $T_{ijl} = 1$ if the data report that the source prefecture of product entry l sold in consuming region i is region j. If

11

we observe no market price and only origin prices, then we set $T_{ijl} = 0$. As shown in Table 1, there are some 230,000 delivery and nondelivery cases for each vegetable. This provides the number of observations for our full information maximum likelihood (FIML) estimation. Out of the total number of delivery and nondelivery cases, the number of delivery cases is relatively small, only about 10,000 for each vegetable. Our data set thus suggests that product delivery is quite limited. It is clear that product delivery therefore is quite local and tends to concentrate in the local areas neighboring the producing prefectures. This raises some concern with sample selection. There is an additional concern about these delivery patterns. If products do not ship to markets directly, then the actual delivery distance will be much longer than that between the final market and the origin. This will cause over bias in the distance effects. However, the share of transferred vegetables is low, normally less than 7 percent according to the Ministry of Agriculture, Forestry and Fisheries. Thus, the influence of transit goods in our data is not significant.

As mentioned, we measure product quality with the local price (the price in the source region). Because local shocks affect local market prices, we need to control for such specific effects. If demand shocks occur locally, the price will be higher without any improvement of quality. We consider this by including region-specific effects in our estimations. When supply shocks take place—i.e., an increase in production costs—the price will be also higher. If the cost associated with quality improvement increases, the Baldwin and Harrigan (2011) framework that we employ will capture it. Conversely, source-region-specific effects reflect cost shocks unrelated to quality.

5.   Empirical Specification

In this section, we specify the functional form of the transport cost functions and other elements for estimation. We assume that the ad-valorem and specific components are a function of distance and other factors:

$$\tau_{nj} = D_{nj}^{\gamma_1} \exp(const + \epsilon_{nj}) \tag{12}$$

$$t_{nj} = D_{nj}^{\gamma_2} \exp(const + \epsilon_{nj}). \tag{13}$$

As we specify a monotonic relationship between price and production costs, we can invert this relationship in terms of price and insert it into the trade cost function. For simplicity, we assume that the remaining elements are common to the ad-valorem and specific cost terms.

Then, the log of the price differential equation is:

$$\ln(p_{nj}/p_{jj}) = const + \ln(D_{nj}^{\gamma_1} + \frac{1}{a}D_{nj}^{\gamma_2}) + \epsilon_{nj}$$

$$= const + \ln(D_{nj}^{\gamma_1} + \frac{\sigma - 1}{p_{jj}\sigma}D_{nj}^{\gamma_2}) + \epsilon_{nj}. \tag{14}$$

Thus, using the variations in a (therefore p_{jj}), we separately estimate γ_1 and γ_2.

We estimate the parameter, θ, with the self-selection condition because $q_{nj} = a^{1+\theta} = (p_{jj}(\sigma-1)/\sigma)^{1+\theta}$:

$$\ln V = \ln(\frac{\sigma}{\sigma - 1})^{1-\sigma} + (1 - \sigma)\ln((p_{jj}(\sigma-1)/\sigma)D_{nj}^{\gamma_1} + D_{nj}^{\gamma_2}) + (1 - \sigma)(const + \epsilon_{nj})$$

$$+ \ln(Y_n\mu) + (\sigma - 1)((1 + \theta)(\ln p_{jj} + \ln(\sigma - 1)/\sigma) - \ln\sigma - (1 - \sigma)\ln P_n - f_j. \tag{15}$$

We estimate the system of these nonlinear equations using maximum likelihood.

5.1. Distance elasticity of quality for the threshold producer

Producers choose their product quality level according to market conditions. One of the focuses here is on the relationship between the distance to market and the quality of goods. As discussed earlier, empirical studies generally show that there is a positive relationship between these two variables, such that the model provides us with the signs of the elasticity of quality with respect to the distance to markets for the threshold producer. For the purpose of discussion, let us begin by deriving the elasticity in the case of no specific costs. From the zero-profit condition, the threshold value of cost, a^*, is expressed by:

$$\frac{(\frac{\sigma}{\sigma-1})^{1-\sigma}(\tau_{nj}a^*)^{1-\sigma}}{q_{nj}^{1-\sigma}}\frac{Y\mu}{\sigma P_n^{1-\sigma}} - f = 0. \tag{16}$$

By the implicit function theorem, we obtain the elasticity of costs with respect to distance from:

$$\frac{da^* D_{nj}}{dD_{nj} a^*} = \frac{\gamma_1}{\theta}. \tag{17}$$

Thus, the elasticity of threshold quality (q^*) with respect to distance is:

$$\frac{dq^* D_{nj}}{dD_{nj} q^*} = \frac{(1 + \theta)\gamma_1}{\theta}. \tag{18}$$

If trade cost is an increasing function of distance ($\gamma_1 > 0$) and the speed of quality improvement is relatively high ($\theta > 0$), then this elasticity is positive.

13

In the presence of a specific type cost, the zero-profit condition is:

$$\frac{(\frac{\sigma}{\sigma-1})^{1-\sigma}(\tau_{nj}a^* + t_{nj})^{1-\sigma}}{q_{nj}^{1-\sigma}} \frac{Y_n\mu}{\sigma P_n^{1-\sigma}} - f = 0. \tag{19}$$

Similarly, by the implicit function theorem, the elasticity is:

$$\frac{da^* D_{nj}}{dD_{nj}a^*} = \frac{\gamma_1 D^{\gamma_1}e^{\epsilon_1} + \gamma_2 D^{\gamma_2}e^{\epsilon_2}a^{*-1}}{\theta D^{\gamma_1}e^{\epsilon_1} + (1+\theta)D^{\gamma_2}e^{\epsilon_2}a^{*-1}}. \tag{20}$$

The sign of the above elasticity depends on not only γ_1 and θ, but also γ_2 and $1+\theta$. As long as $\gamma_2 > 0$ and $\theta > -1$, the elasticity will be positive, even if $\theta < 0$. Thus, the presence of specific costs relaxes the condition for the positive relationship between quality and distance.

6. Results: Relationship Between Quality and Distance

In this section, we report our estimation results. To compare our results with previous studies, we conduct our estimations using several different specifications: 1) structural estimation of a simple Melitz (2003) model, 2) structural estimation with a quality model (as in Baldwin and Harrigan (2011)), and 3) structural estimation of a firm-heterogeneity model with quality and specific costs. To compare the results with those in the extant literature, we begin by specifying no quality dimension and no specific costs.

Columns 1, 4, and 7 in Table 3 report the results of a model without quality dimension for cabbage, c-cabbage, and lettuce, respectively. The important parameters are the elasticity of substitution and the elasticity of transport cost with respect to distance. The substitution parameters are 4.957, 4.138, and 3.355 for cabbage, c-cabbage, and lettuce, respectively. These values are reasonable in the context of studies of individual product data. The distance elasticity parameters are, 0.227, 0.325, and 0.343 for cabbage, c-cabbage, and lettuce, respectively. These are also similar to the results in Kano et al. (2013). Thus, the distance effect is larger than those in the LOP literature, and this is because the sample selection problem there is not controlled for as here.

=== Table 3 here ===

We now introduce quality as in Baldwin and Harrigan (2011). The results are in Columns 2, 5, and 8 in Table 3. As shown, the estimates of the distance effect and the elasticity of substitution are almost identical to those without quality (0.228, 0.325, and 0.345 for cabbage, c-cabbage, and lettuce, respectively). The quality parameters turn out to be marginally negative, which suggests that high-cost producers produce high-quality goods. However, the rate of increase of quality is slower than where high-cost producers

14

deliver their products to distant markets. While the earlier reduced-form regressions show a positive link between quality and distance, the results here imply that this is not solely the result of quality sorting. This may also be because we conduct our analysis using daily data for agricultural products over only a single year, such that it may be difficult to improve product quality during the relatively short period.

Finally, we estimate the model incorporating quality and specific costs. Columns 3, 6, and 9 report the results. The distance effects for the ad-valorem term are 0.162, 0.26, and 0.277 for cabbage, c-cabbage, and lettuce, respectively, whereas those for the specific cost term are 0.61, 0.665, and 0.792, respectively. Hummels and Skiba (2004) suggest that ad-valorem trade costs are only tariffs and that specific costs are distance-elastic trade costs. Our results are at least qualitatively consistent with their specification.

The magnitude of the estimates of the quality parameter is also larger than before, being -0.158, -0.204, and -0.193 for cabbage, c-cabbage, and lettuce, respectively. As mentioned, if $\theta > 0$, the model exhibits quality sorting. If $-1 < \theta < 0$, then high-quality goods are produced by high-cost firms, although the increase in quality is not as rapid as the increase in costs. Thus, the positive θ is needed for the quality selection without specific costs. However, as shown in Section 5.1, the positive relationship between quality and distance may arise with specific costs, even if $\theta < 0$. Hence, when we combine the results of the distance effects with the negative values of θ, we conclude that the positive relationship between quality and distance is a consequence of the presence of specific costs. The Alchian–Allen effects are the driving force here.

As we have seen, without taking specific costs into account, the technology parameter is marginally negative. However, this is because the estimation of this parameter is biased without specific costs. If θ is large, high-cost firms produce quite high-quality goods; thus, they ship their products to costly distant markets. However, the reason that high-quality goods are shipped to a distant market may be that consumers have relatively high demand for these goods in costly transport cost markets. Thus, the omitted variable (the specific cost term) will cause the technology parameter to capture this positive demand-side effect between distance and quality. Once we can control for specific costs, we can then identify the true technology parameter. In fact, high costs produce high-quality goods. However, this effect is not strong enough to account on its own for the quality-sorting mechanism in our sample.

Two important parameters other than distance elasticity are the elasticity of substitution and the correlation parameter of the error terms. The elasticity parameters have values between three and six, which is reasonable when using micro data. The absolute values of the correlation parameters are all more than 0.8, suggesting strong correlations.

Thus, sample selection may invoke a serious problem for biased estimates.

7. Quality of Distance Measure

The relationship between quality and distance may also depend on the choice of distance measure. If the quality and distance relationship is sensitive to the measure of distance, our results may not be considered robust. Thus, we specify a different measure of the distance between regions as a robustness check.

In the main analysis, our measure of distance is direct distance. This may differ from actual road distance, which may be a better proxy for transport costs. For example, Ehime Prefecture (the author's hometown) is 666.1 km from Tokyo by direct distance. However, because these prefectures are located on different islands, the actual shipping distance is much longer. In fact, the road distance between Ehime and Tokyo is 853.1 km. Thus, direct distance may cause an over bias in the distance effect. The fact that the distance effect is large may be simply because the actual distance is in fact longer, so each kilometer does not impose a significant burden for suppliers. While other transport modes are available (e.g., air), the most relevant type of transport in our analysis is truck. Thus, a navigation software website (navitime.co.jp) is used to calculate the road distance. Primarily, the distance for the route using only regular roads is calculated. However, where this is not possible, highways are included in the route. In addition, if there is no bridge between the two prefectures, the ferry distance is included.[4]

Columns 2, 5, and 8 in Table 4 report the estimation results using this alternative distance measure. For the most part, and as expected, the effect of distance here is smaller than previously found. However, these are similar to those using direct distance. Thus, the choice of direct or road distance does not represent a serious source of bias in our estimations.

=== Table 4 here ===

With regard to the distance measure, as discussed in the literature, the choice of internal distance may also be important. We now employ the Head–Mayer measure of internal distance: $D_{jj} = 0.376 \times \sqrt{area}$. Figure 3 depicts the same relationship as Figure 1, which is the correlation between the origin price and the distance to market. The results show that the parameter estimates are qualitatively similar to our simple measure of internal distance used previously, as again there is a positive relationship, as depicted by the solid line in the figure. Hence, our results are robust to the choice of internal distance measure.

=== Figure 3 here ===

---

[4]The data on land route distance between prefectures are available on Professor Tsukui's website (www.tiu.ac.jp/~makiko/Japanese/DATA/distance_between_prefectures(land_route).xls).

16

Finally, we adopt a similar specification for the trade cost function as Hummels and Skiba (2004), in which specific costs are increasing in product value. Because transport costs can be high for high-value goods, the trade cost will be:

$$\ln q_{nj} = \tau_{nj} + t_{ij}/a = D_{nj}^{\gamma_1} + p_{jj}^{\beta} D_{nj}^{\gamma_2}/a. \qquad (21)$$

If $\beta < 1$, then the specific transport cost increases as the value of the goods increases but at a slower rate. This again confirms the Alchian–Allen effect. Columns 4, 7, and 10 in Table 4 report the results of the Hummels and Skiba (2004) specification. The parameter values for the distance elasticity and the elasticity of substitution are similar to those for the earlier estimations. The Hummels–Skiba parameters, β, are 0.322 and 0.421 for cabbages and lettuces, respectively. Hence, our results are also consistent with those of the Alchian–Allen effect. For c-cabbage, the Hummels–Skiba parameter is 0.009 and is not significant; hence, our original specification may be the appropriate representation for the specific cost term.

One remark is worth mentioning. While our estimates reveal the large distance effects, these may in fact be the lower bounds of distance elasticity. This is because we exclude the price data, in which there is no information available for local delivery. Because of this, we cannot calculate the price differentials between the origin and the destination. This means that price differential data associated with long distances to market are not included in our analysis, which under biases the distance effect. Consequently, the direction of bias may not weaken our estimation results.

8. Policy Evaluation

How significant is the impact of policies reducing trade costs? To investigate the gains from a trade cost reduction quantitatively, we conduct Monte Carlo exercises using a three-region version of the model. To evaluate the welfare gains, we need to calculate the price indexes numerically. However, this involves some difficulty in the convergence of a model including 47 regions, as revealed in Section 4. Fortunately, as our focus is an illustration of the magnitude of welfare gains, not the replication of the overall Japanese regional gains, we create a single core region located in the middle and two peripheral regions.

We set up Monte Carlo exercises using the program developed by Irarrazabal et al. (2013) and available through Khandelwal's et al. (2013) website. In our model, unit distance is set to 1.5, so the closest region is 3 and the furthest region is 4.5. Because the elasticity for the ad-valorem cost is 0.16 and 0.61 for the specific cost, the specific trade cost

17

is approximately 20 percent higher than the ad-valorem costs for the closest region and 60 percent higher for delivery from one peripheral region to the other. We conduct 150 exercises and calculate the average welfare gains by comparing the real wages across three transition scenarios: 1) the friction to no ad-valorem cost case, 2) the friction to the no-specific-cost case, and 3) the friction to zero friction case.

Table 5 details the average welfare gains as denoted by the percentage increase in each scenario. The first row reveals that the removal of ad-valorem costs increases welfare, but only slightly. In contrast, the reduction in specific costs, which includes a large proportion of trade costs, has a large impact, being 25 percent for the core region and 22 percent for the peripheral regions. Finally, the third row shows that the removal of all trade costs increases welfare by approximately 30 percent in this economy. As expected, when the impact of specific cost removal is large, the magnitude is substantial. Our Monte Carlo experiment thus suggests that specific trade costs are a more severe obstacle to trade than ad-valorem costs, as also shown in Irarrazabal et al. (2013).

=== Table 5 here ===

9. Concluding Remarks

The trade literature uses the iceberg-type trade (or transport) cost function. Under this specification, quality sorting is a mechanism thought to represent quality and the distance to markets. However, it is important to incorporate specific costs in this specification because of the presence of the Alchian–Allen effect. Our study thus attempts to identify the structural parameters of the quality heterogeneity model.

The main empirical test in the literature is the regression of FOB prices (unit values) on distance. Our study extends this analysis using a structural model to reveal whether it is the quality-sorting effect or the Alchian–Allen effect (or both) that drives the relationship between quality and distance. We also estimate the technical parameter that connects cost and quality and take into account selection bias associated with the choice of product delivery. The main findings indicate that specific costs are more distance elastic than ad-valorem costs, and that the presence of specific costs is the key element in the typical empirical observation of a positive link between quality and distance.

While our study reveals the importance of specific costs, further study is required. For example, with CES preferences, monopolistically competitive firms set constant markup prices to all the markets that they serve. However, pricing behavior may differ across markets. In addition, because firms may not pass the increase in production costs on to market prices, the estimation of the distance effect may be biased. Pricing to market behavior also depends

on market competitiveness and the levels of market income (e.g., Lugovskyy and Skiba (2012)). Thus, to take into account the effect of distance fully, we need to incorporate pricing to market behavior. Further research in this area is therefore required.

References

[1] Anderson, J., van Wincoop, E., 2003 Gravity with gravitas: a solution to the border puzzle, American Economic Review 93, 170–192.

[2] Anderson, J., van Wincoop, E., 2004. Trade costs, Journal of Economic Literature 42, 691–751.

[3] Atkin, D., Donaldson, D., 2014. Who's getting globalized? The size and implications of intra-national trade costs, mimeo.

[4] Baldwin, R., Harrigan, J., 2011. Zeros, quality, and space: trade theory and trade evidence, American Economic Journal: Microeconomics, 3, 60–88.

[5] Bastos, P., Silva, J., 2010. The quality of a firm's exports: where you export to matters. Journal of International Economics 82 (2), 99–111.

[6] Donaldson, D., 2014. Railroads of the Raj: estimating the impact of transportation infrastructure, American Economic Review, forthcoming.

[7] Eaton, J., Kortum, S., 2002. Technology, geography, and trade, Econometrica, 1741–1779.

[8] Head, K., Mayer, T., 2000. Non-Europe: the magnitude and causes of market fragmentation in the EU, Weltwirtschaftliches Archiv 136, 284–314.

[9] Helpman, E., Melitz, M., Rubinstein, Y. 2008. Estimating trade flows: trading partners and trading volumes, Quarterly Journal of Economics, 441–487.

[10] Hummels, D. 2001. Toward a geography of trade costs. Working Paper, Purdue University.

[11] Hummels, D., Skiba, A., 2004. Shipping the good apples out? An empirical confirmation of the Alchian–Allen conjecture, Journal of Political Economy 112, 1384–1402.

[12] Hummels, D., Lugovskyy, V., Skiba, A., 2009. The trade reducing effects of market power in international shipping, Journal of Development Economics 89, 84–97.

[13] Irarrazabal, A., Moxnes, A., Opromolla, L.D., 2013. The tip of the iceberg: a quantitative framework for estimating trade costs, NBER Working Paper 19236.

[14] Kano, K., Kano, T., Takechi, K., 2013 Exaggerated death of distance: revisiting distance effects on regional price dispersions, Journal of International Economics 90, 403–413.

[15] Kano, K., Kano, T., Takechi, K., 2014 The price of distance: pricing to market, producer heterogeneity, and geographic barriers, mimeo.

[16] Khandelwal, A. K., P. K. Schott., Wei, S.-J. 2013. Trade liberalization and embedded institutional reform: evidence from Chinese exporters, American Economic Review, 103(6) 2169–2195.

[17] Lugovskyy, V., Skiba, A., 2012. How geography affects quality, mimeo.

[18] Manova, K., Zhang, K., 2012. Export prices across firms and destinations, Quarterly Journal of Economics 127, 379–436.

[19] Martin, J., 2012. Markups, quality, and transport costs, European Economic Review 56, 777–791.

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

	Cabbage	Cabbage	C-Cabbage	C-Cabbage	Lettuce	Lettuce
Distance	0.074	0.007	0.106	0.01	0.046	0.008
	(0.002)	(0.003)	(0.03)	(0.005)	(0.003)	0.004
Num. of Obs.	15841	15841	10803	10803	11565	11565
R squared	0.065	0.494	0.105	0.504	0.019	0.364
Region-Specific Effect	No	Yes	No	Yes	No	Yes

Table 1: Reduced-Form Estimation Results

	Cabbage	C-Cabbage	Lettuce
Average market price	77.833	61.628	183.909
Average local price	67.431	50.671	168.855
Product entry			
No. of varieties	3	4	7
No. of size categories	63	50	71
No. of grade categories	34	50	46
No. of producing prefectures	47	46	43
No. of distinct product entries	1207	1001	903
Data truncation			
No. of $T_n j(l) = 1$ or 0	369343	241871	239703
No. of $T_n j(l) = 1$	15841	10803	11565

Table 2: Summary Statistics

Parameter	Cabbage No Q, No S	Cabbage Q, No S	Cabbage Q, S	C-Cab No Q, N S	C-Cab Q, N S	C-Cab Q, S	Lettuce N Q, N Ss	Lettuce Q, N S	Lettuce Q, S
γ_1	0.227	0.228	0.162	0.325	0.325	0.26	0.343	0.345	0.277
	(0.016)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
γ_2			0.61			0.665			0.792
			(0.003)			(0.004)			(0.005)
θ		-0.041	-0.158		-0.038	-0.204		-0.089	-0.193
		(0.003)	(0.004)		(0.003)	(0.004)		(0.005)	(0.005)
ϵ	4.957	4.966	5.219	4.138	4.149	4.374	3.355	3.363	3.491
	(0.021)	(0.023)	(0.022)	(0.02)	(0.014)	(0.017)	(0.016)	(0.016)	(0.014)
ρ	-0.84	-0.847	-0.847	-0.82	-0.826	-0.829	-0.859	-0.872	-0.87
	(0.0023)	(0.003)	(0.002)	(0.003)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)
Num. of Obs.	369343	369343	369343	241871	241871	241871	239703	239703	239703
Log-likelihood	-21404.133	-21344.762	-20234.094	-21404.133	-21344.762	-20234.094	-22296.627	-22151.746	−21571.669

Table 3: Estimation Results

Parameter	Cab Road Dist	Cab HM Dist	Cab HS Spec	C-Cab Road Dist	C-Cab HM Dist	C-Cab HS Spec	Lettuce Road Dist	Lettuce HM Dist	Lettuce HS Spec
γ_1	0.144	0.222	0.146	0.225	0.36	0.251	0.257	0.337	0.254
	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
γ_2	0.593	0.653	0.462	0.618	0.714	0.647	0.764	0.846	0.563
	(0.003)	(0.004)	(0.007)	(0.004)	(0.005)	(0.01)	(0.005)	(0.006)	(0.012)
θ	-0.173	-0.14	-0.166	-0.213	-0.165	-0.2	-0.206	-0.176	-0.2
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.005)	(0.005)	(0.005)	(0.004)
ϵ	5.272	5.089	5.24	4.404	4.252	4.376	3.509	3.456	3.505
	(0.023)	(0.022)	(0.023)	(0.017)	(0.017)	(0.018)	(0.017)	(0.015)	(0.015)
HS par			0.322			0.009			0.421
			(0.018)			(0.028)			(0.022)
ρ	-0.845	-0.841	-0.847	-0.83	-0.825	-0.829	-0.871	-0.865	-0.87
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Num. of Obs.	369343	369343	369343	241871	241871	241871	239703	239703	239703
Log-likelihood	-19347.496	-23247.471	-20153.728	-13148.843	-15946.852	-13612.545	-20971.987	-23598.581	-21500.773

Table 4: Estimation Results

Welfare Gains (% increase)	Core	Periphery
Friction to No Ad-valorem costs	0.132%	0.07%
Friction to No Specific costs	25.133%	22.131%
Friction to No Friction	30.413%	31%
Size of the trade costs		
τ	1.067	1.192
t	1.281	1.955
t/τ (Specific/Ad-valorem cost)	120.017%	163.974%

Table 5: Average Welfare Gains
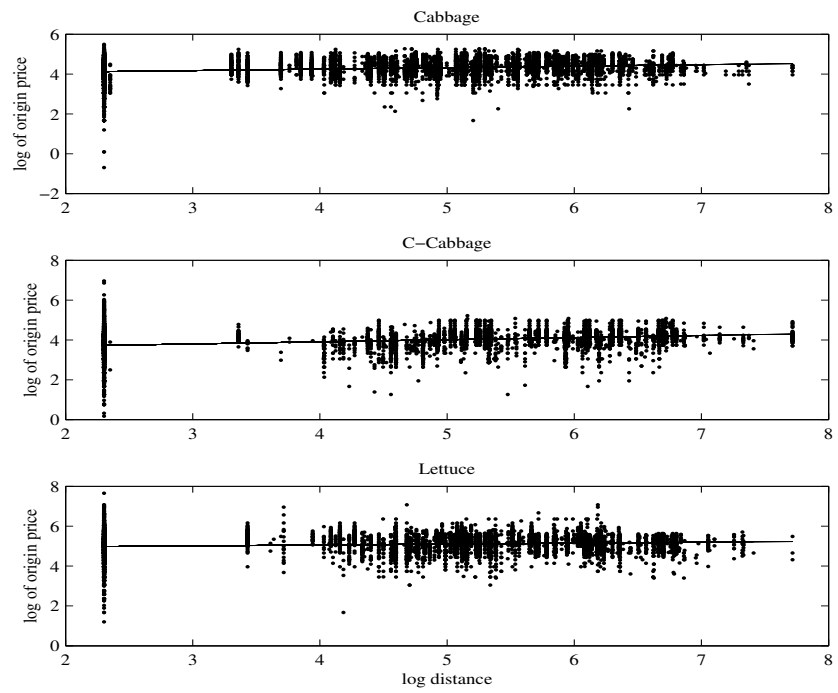
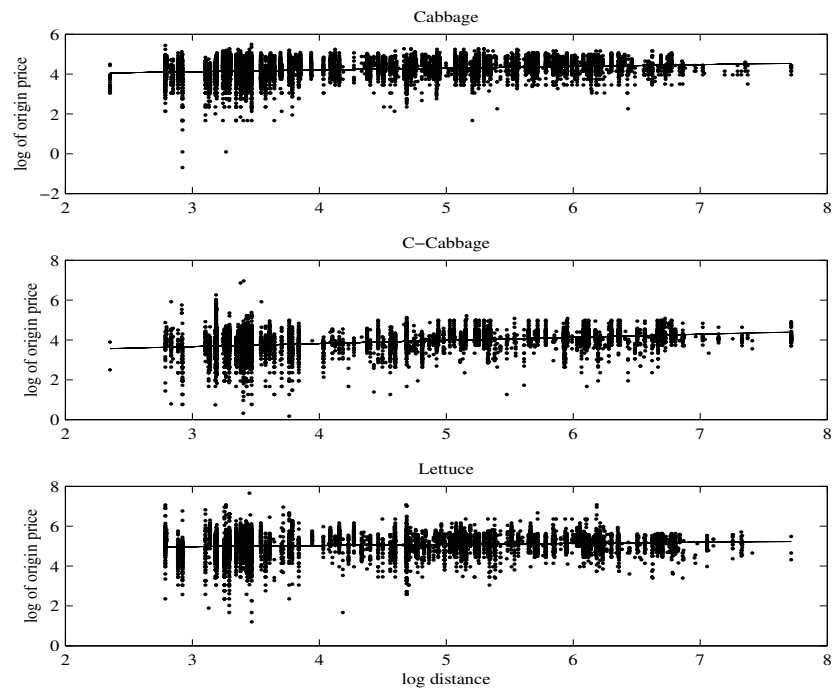Figure 1: Logs of distance and source price relationship

Figure 2: Logs of distance and source price relationship (Head–Mayer internal distance measure)
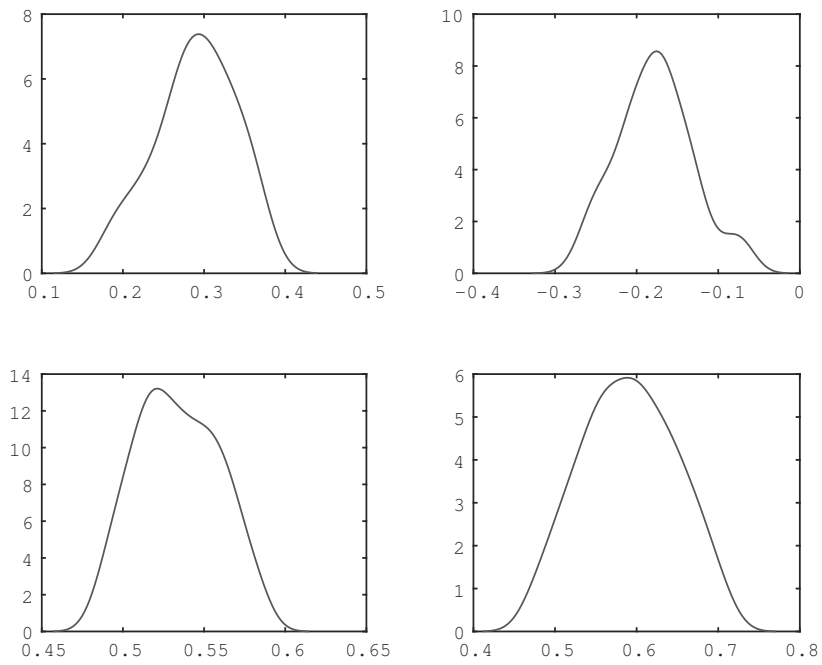
Figure 3: Kernel densities of estimators of distance and quality elasticities