Microdata Analysis of Japanese Farmers' Productivity: Estimating Farm Heterogeneity and Elasticity of Substitution among Varieties

Yuko Akune
Nobuhiro Hosoe

November 2019

Microdata Analysis of Japanese Farmers' Productivity: Estimating Farm Heterogeneity and Elasticity of Substitution among Varieties[†]

November 26, 2019

Yuko Akune (Nihon University)[‡]

Nobuhiro Hosoe (National Graduate Institute for Policy Studies)

Abstract

Modern theories of trade and economic geography pay particular attention to the role of product differentiation under monopolistic competition in manufacturing, while agriculture is considered to produce homogeneous goods. By contrast, agribusiness studies shed light on active entrepreneurs who have high productivity and are engaged in product differentiation by creating new products in their niches. Given these two contradicting views of agriculture, we examine the incidence of farm heterogeneity and product differentiation. This study uses microdata of Japanese farmers to estimate their total factor productivity. We find that heterogeneity is relatively low in the horticulture, grain and soybean (excluding rice), and fruit farming sectors, and high in the livestock sectors. In addition, the degree of product differentiation is relatively high in the livestock sectors, and the elasticity of substitution is as high as three, which is similar to findings in earlier studies on agriculture and manufacturing.

Keywords

Agricultural productivity; firm heterogeneity; product differentiation; total factor productivity

[‡] Corresponding author. Address: 1866 Kameino, Fujisawa, Kanagawa 252-0880, Japan. E-mail: akune.yuko@nihon-u.ac.jp.

1.   Introduction

Agriculture is often described as a homogeneous good sector in many economic analyses, which implies that product differentiation among farmers is low. By contrast, agribusiness studies highlight heterogeneity of productivity among farmers by scale and location and explore how farmers are engaged in product differentiation through breeding for better taste and a longer expiration date, branding based on geographical and traditional identities and producers, marketing for new sales destinations, and timely supply to avoid competition with other sources. Although these activities are well described by a framework of monopolistic competition (Dixit and Stiglitz (1977)), those agribusiness studies do not employ this framework, but mostly conduct descriptive case studies of product development and marketing for Japanese farmers and food companies (Tanaka (2006), Akashi and Tanemura (2006), Nakamura et al. (2011), Ministry of Agriculture, Forestry and Fisheries (MAFF) (2018b)). Empirical studies using statistical data are indeed scant in this field. Only Kano et al. (2013), Takechi (2015), and Ihara et al. (2015) have estimated the elasticity of substitution among varieties for vegetables farming in a new economic geography model (Krugman (1980)).

The modern trade theory initiated by Melitz (2003) examined heterogeneity among firms in analysis of international trade and regional science. This theory shows that product differentiation needs to be considered jointly with variations in productivity. Product

differentiation incurs various pecuniary and nonpecuniary costs, including risks, for production

facilities and breeding, development of cultivation technologies, branding and marketing,

intellectual property right protection, and so on. Only highly productive entrepreneurs are allowed

to access new products to expand their business and increase profits.

Based on those theoretical frameworks, empirical studies have investigated total factor

productivity (TFP) of Japanese manufacturing firms, and its heterogeneity incidence (Wakasugi

et al. (2014) and Mizuno et al. (2012)) and size distributions (Fujimoto and Ishikawa (2011)).

Wakasugi et al. (2014) confirmed that the patterns of productivity, exports, and foreign direct

investment were consistent with the assumptions of modern trade theory. By contrast, agricultural

productivity studies in Japan have very often measured productivity by using single input

productivity, typically, land and labor productivity (e.g., MAFF (2008)). This is partly because

no microdata of Japanese farmers have been available publicly, and partly because of their

practical convenience in estimation. As agriculture employs farmland intensively, a simple

productivity measure such as land productivity has been used extensively. However, omitting

other factor inputs could lead to biases in productivity estimates. To take account of the

contribution of multiple inputs, Hu (1995) and Kunimitsu (2011) employed a Törnqvist

productivity index, and Kondo and Yamamoto (2003) and Yamamoto et al. (2007) used a

Malmquist productivity index. These indexes still have a shortcoming, in that the input shares,

which are used as weights for combining multiple inputs, are assumed to be exogenous. Furthermore, in their approach, the validity of the input shares cannot be tested. An econometric approach allows us to estimate a production function that considers all the inputs, such as capital, labor, and land, to measure TFP, or the Solow residual, as Wakasugi et al. (2014) and Mizuno et al. (2012) did for manufacturing. Studies in many other countries have used this approach in agricultural productivity analysis. For example, the Economic Research Service of the US Department of Agriculture regularly estimated TFP with microdata (Ball et al. (2013) and Shumway et al. (2016)).

These previous studies measured *average* productivity only; they ignored heterogeneity among farmers. A lack of farm-level data necessitated the use of aggregate data for Japan and its subregions for TFP estimation (Hu (1995) and Kunimitsu (2011)). While aggregate data analyses allow us to estimate TFP with a small number of observations, we cannot examine variation within a group. Even if farmers are categorized into several groups, as in Kuroda (2013), the resolution of analysis is still inadequate to identify heterogeneity within agriculture. A wide variety of farmers, including very productive entrepreneurs and smallholders who are aged and almost fully depend on pensions, grow crops for their own use.

Agriculture consists of many subsectors, often characterized by their products. Previous studies have not analyzed agriculture comprehensively, but rather, have focused on particular

farming sectors, such as paddy rice, vegetables, fruit, beef cattle, and so on. The most popular

sector for research has been paddy rice farming (Kondo et al. (2005), Yamamoto et al. (2007),

and Kunimitsu (2014)). These sector-focused studies enable us to control for product- or sector-

specific factors that affect farming activities, such as input prices, production technology, and

weather conditions. However, it is solely a farmer's choice what product to grow with what

production technology and inputs, given his or her farmland. We should assume a general

production function behind the farmer's choice in estimating productivity so as not to cause any

unintended biases because of a priori assumptions about the effect of these factors on farming

activities. From this perspective, Kuroda (2017) adopted a notable approach by using a multiple-

product translog cost function.

Based on the above-mentioned literature review, there has been no productivity study

in Japan allowing for farm heterogeneity in estimation or estimated productivity distributions

indicating the farm heterogeneity incidence. Modern trade theory shows that firm heterogeneity

determines patterns of production and exports. A simplifying assumption for a representative farm

with average productivity may allow us to overlook this key feature in agriculture. In this study,

we use farm-level microdata, *Nogyo Keiei Tokei Chosa* (*Statistical Survey on Farm Management

and Economy*) by MAFF, and estimate production functions to measure TFP in agriculture.

Exploiting the benefit of our microdata, we depict the distributions of TFP and estimate the

elasticity of substitution among varieties to quantify the degree of product differentiation at the subsector level.

This paper proceeds as follows. In Section 2, we present our framework, covering production functions, density functions of productivity, and estimation methods of these functions and the implied elasticity of substitution among varieties. Section 3 shows the estimated production functions and presents the distributions of TFP and the elasticity of substitution for the subsectors. The estimated TFP is compared with the conventional measures of labor and land productivity. Section 4 concludes the paper and discusses the limitations and possible extensions of this study.

2.   Models and Estimation

There are three steps in our model estimation: (1) estimation of production functions for TFP, (2) estimation of the shape parameter of a Pareto distribution of TFP, and (3) estimation of farm size distributions for the elasticity of substitution among differentiated goods. In the first step, we use alternative estimation methods to address the well-known endogeneity bias, and include additional dummy variables to control for sector-specific factors in the production function.

2.1   Production Function

We consider a standard Cobb–Douglas production function (1),

$$Y_{it} = A_{it} K_{it}^{\beta_1} L_{it}^{\beta_2} LD_{it}^{\beta_3} \qquad (1)$$

where Y_{it}, K_{it}, L_{it}, and LD_{it} are the output, capital input, labor input, and land use of farmer i in year t, respectively. β_1, β_2, and β_3 are their input shares ($0 < \beta_f < 1, f = 1,2,3$). We estimate the TFP of farmer i in year t, A_{it}. Log-linearization of (1) yields,

$$y_{it} = \alpha_{it} + \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 ld_{it} \qquad (2)$$

where $y_{it} = lnY_{it}$, $k_{it} = lnK_{it}$, $l_{it} = lnL_{it}$, $ld_{it} = lnLD_{it}$, and $\alpha_{it} = lnA_{it}$. This is estimated with an error term ε_{it},

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 ld_{it} + \varepsilon_{it} \qquad (3)$$

where $\alpha_{it} = \beta_0 + \varepsilon_{it}$. It is widely acknowledged that the endogeneity problem of explanatory variables in panel data estimation must be addressed (Marschak and Andrews (1944), Olley and Pakes (1996), Levinsohn and Petrin (2003), and Mollisi and Rovigatti (2017)). We split the error term ε_{it} in (3) into an observable shock ω_{it} and an unobservable shock ξ_{it},

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 ld_{it} + \omega_{it} + \xi_{it} \qquad (4)$$

In its estimation, we use (log of) real value added produced by farmer i in year t for y_{it}, real capital stock for k_{it}, total work hours for l_{it}, and operating cultivated land area for ld_{it}. To address the endogeneity, we consider the options developed by Olley and Pakes (OP) (1996) and

Levinsohn and Petrin (LP) (2003), and the extensions of these about labor input by Ackerberg, Caves, and Frazer (ACF) (2015). We use the real value of investment for ω_{it} in the OP and OP-ACF methods and intermediate inputs for the LP and LP-ACF methods. With this setup, we estimate the unknown parameters β_0, β_1, β_2, and β_3.

In addition to this standard specification, we add several binary dummy variables to (4) to control for sector-specific effects,

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 ld_{it} + \beta_4 D_i^k k_{it} + \beta_5 D_i^{ld} ld_{it} + \sum_j \beta_{6j} D_{ij} + \omega_{it} + \xi_{it} \quad (5)$$

where D_i^k is a dummy variable for farmer i in capital-intensive sectors, D_i^{ld} is a dummy variable for that in land-intensive sectors, and D_{ij} is a dummy variable for farmer i in farming sectors $j = 1, \dots, 14$. Factor intensity varies widely among farming sectors and determines their susceptibility to particular input-related shocks, such as sunshine duration, temperature, and precipitation. Their effects are controlled for by the coefficient dummy variables, D_k and D_{ld}.[1] Other miscellaneous effects are controlled for by the intercept dummy variables D_j for farming sector j. By estimating (4) or (5), we can compute TFP, $A_{it} = exp(\alpha_{it})$. As mentioned above,

---

[1] These coefficient dummy variables take the value 1 if a sector's capital and land intensity, relative to labor input, exceeds their median values among the 14 farming sectors, respectively. Paddy field, upland field, fruit, dairy, fattening cattle, pig, poultry egg, and miscellaneous farming are categorized as capital-intensive sectors; paddy field, upland field, outdoor-grown vegetables, dairy, breeding cattle, fattening cattle, and miscellaneous farming sectors are categorized as land-intensive sectors. Details are provided in the Appendix. Although we use coefficient dummy variables defined for each of the 14 farming sectors instead of factor-intensity-based dummy variables, we cannot improve our estimation results.

our main data source is *Nogyo Keiei Tokei Chosa* (*Statistical Survey on Farm Management and Economy*) for 2012–2015. This survey consists of several sections, among which we use microdata of *Einou Ruikei Betsu Keiei Toukei* (*Statistics on Management by Farming Type*), which reports farming activities for three management types: individual management, management by corporate organization, and management by community-based farm cooperative (for details, see the Appendix).

2.2   Estimation of Pareto's *k*

Following previous studies, such as Mayer and Ottabiano (2007) and Wakasugi et al. (2014), we assume productivity follows a Pareto distribution. A cumulative distribution function for TFP is assumed as follows,

$$F(TFP_{it}; k) = 1 - \left(\frac{TFP_{it}}{TFP_M}\right)^{-k}, TFP_{it} > TFP_M \qquad (6)$$

where *k* is the shape parameter of a Pareto distribution (hereafter Pareto's *k*), and TFP_M is the minimum value of all TFP estimates, which is also the mode of this distribution. Log-transformation of (6) yields,

$$ln\big(1 - F(TFP_{it})\big) = ln\,B - k\ ln(TFP_{it}) \qquad (7)$$

where $B = TFP_M{}^k$. We estimate Pareto's *k* in (8) by the ordinary least squares (OLS) method.

$$ln\big(1 - F(TFP_{it})\big) = b - k\ ln(TFP_{it}) + e_{it} \qquad (8)$$

where $b = lnB$, and e_{it} is an error term.

2.3   Estimating Elasticity of Substitution among Varieties

Crozet and Koenig (2010) estimated the elasticity of substitution among varieties by using Pareto's k in the trade equations. They assume one farmer produces only one variety, and formulate the following exponential function, which describes the relationship between an individual farmer's productivity and production.

$$X_{it} = \lambda TFP_{it}^{-(k-\sigma+1)} \qquad (9)$$

where X_{it} is cumulative production of farmers with a higher productivity than farmer i in each sector and year t, and σ is the elasticity of substitution among varieties. Log-linearization yields (10) for our estimation.

$$x_{it} = b_0 + b_1 \, ln(TFP_{it}) + s_{it} \qquad (10)$$

where $x_{it} = lnX_{it}$, $b_0 = ln\,\lambda$, $b_1 = -(k - \sigma + 1)$, and s_{it} is an error term. The elasticity of substitution σ can be computed by combining the estimate of k in (7) with that of b_1 in (10).

3.   Estimation Results

3.1   Production Function and TFP

We try four alternative estimation methods that can address the endogeneity problem.

For simplicity, we do not consider any dummy variables in this first step. In the next step, we examine the benefits of factor intensity and subsector dummy variables. Table 1 reports the estimation results of the production function (4), showing the benefits of elaborations with the OP, LP, and ACF methods. Explicitly controlling for the effects of investment or intermediate inputs on output with ω_{it}, OP (1-3) and LP (1-4) yields smaller shares of capital input compared with OLS (1-1). Applying the ACF method that elaborates the assumption for labor input, we can improve estimation of the share parameters for labor and land. The labor input share by OP-ACF (1-5) and LP-ACF (1-6) are closer to that by OLS than to those by simple OP and LP. The fixed effect (FE) model does not perform well. In general, OP-ACF and LP-ACF perform better than others. Thus, in the next step, we examine the effects of the dummy variables while employing these methods.

Table 1: Estimation Results of Production Function (4) without Dummy Variables

	(1-1) OLS	(1-2) FE	(1-3) OP	(1-4) LP	(1-5) OP-ACF	(1-6) LP-ACF
Labor	0.901***	0.485***	0.675***	0.526***	0.928***	0.935***
	(0.0075)	(0.0254)	(0.0127)	(0.0130)	(0.0075)	(0.0079)
Capital	0.139***	0.00789	0.0255**	0.0237**	0.0773***	0.0391***
	(0.0048)	(0.0078)	(0.0112)	(0.0101)	(0.0150)	(0.0101)
Land	0.163***	0.0777***	0.100***	0.0572***	0.206***	0.228***
	(0.0049)	(0.0228)	(0.0089)	(0.0060)	(0.0095)	(0.0161)
Observations	16,099	16,099	15,544	16,099	15,544	16,099

Note: Standard errors are in parentheses. *** and ** indicate significance at the 1% and 5% levels, respectively. The numbers of observations used for OP and OP-ACF are smaller than those for the other methods. OP and OP-ACF use investment data to control for observed productivity shocks. As investment is not carried out in some years in the dataset, these observations are not included in our estimation.

Table 2 reports the estimation results of (5) by the OP-ACF and LP-ACF methods. Compared with (1-5) and (1-6), which have no dummy variables, (2-1) and (2-4) with the farming sector intercept dummy variables show smaller labor shares. Furthermore, (2-2) and (2-5) with the factor intensity coefficient dummy variables show smaller land shares. Incorporating both types of dummy variables, (2-3) and (2-6) show similar share parameter estimates of around 0.8 for labor, 0.1 for capital, and 1.4 for land. Considering the overall performance, we use the LP-

ACF model (2-6) to estimate TFP (Figure 1).[2]

Table 2: Estimation Results of Production Function (5) with Dummy Variables

	(1-5)	(2-1)	(2-2)	(2-3)	(1-6)	(2-4)	(2-5)	(2-6)
	OP-ACF				LP-ACF			
Labor	0.928***	0.808***	0.967***	0.807***	0.935***	0.810***	0.956***	0.799***
	(0.0075)	(0.0045)	(0.0173)	(0.0151)	(0.0079)	(0.0084)	(0.0081)	(0.0097)
Capital	0.0773***	0.0853***	0.0651***	0.105***	0.0391***	0.0526***	0.0216*	0.113***
	(0.0150)	(0.0156)	(0.0193)	(0.0127)	(0.0101)	(0.0098)	(0.0126)	(0.0129)
Land	0.206***	0.255***	0.166***	0.139***	0.228***	0.263***	0.127***	0.148***
	(0.0095)	(0.0054)	(0.0177)	(0.0127)	(0.0161)	(0.0083)	(0.0200)	(0.0117)
Intercept dummy	N	Y	N	Y	N	Y	N	Y
Coefficient dummy	N	N	Y	Y	N	N	Y	Y
Observations	15,544	15,544	15,544	15,544	16,099	16,099	16,099	16,099

Note: Standard errors are in parentheses. *** and * indicate significance at the 1% and 10% levels,

respectively.

TFP is a sophisticated productivity index, taking account of multiple inputs used in

production. Instead, alternative productivity indexes, such as land and labor productivity, are

commonly used for convenience. We compare these alternatives with TFP to examine their

performance. Land or labor productivity is defined as inputs per unit of value added in each farm

(Figure 1). The distribution of land productivity is skewed significantly right (skewness = 0.999)

with a nonsmooth right shoulder. Labor productivity and TFP have left-skewed distributions

---

[2] Summary statistics of TFP estimates are shown in the Appendix.
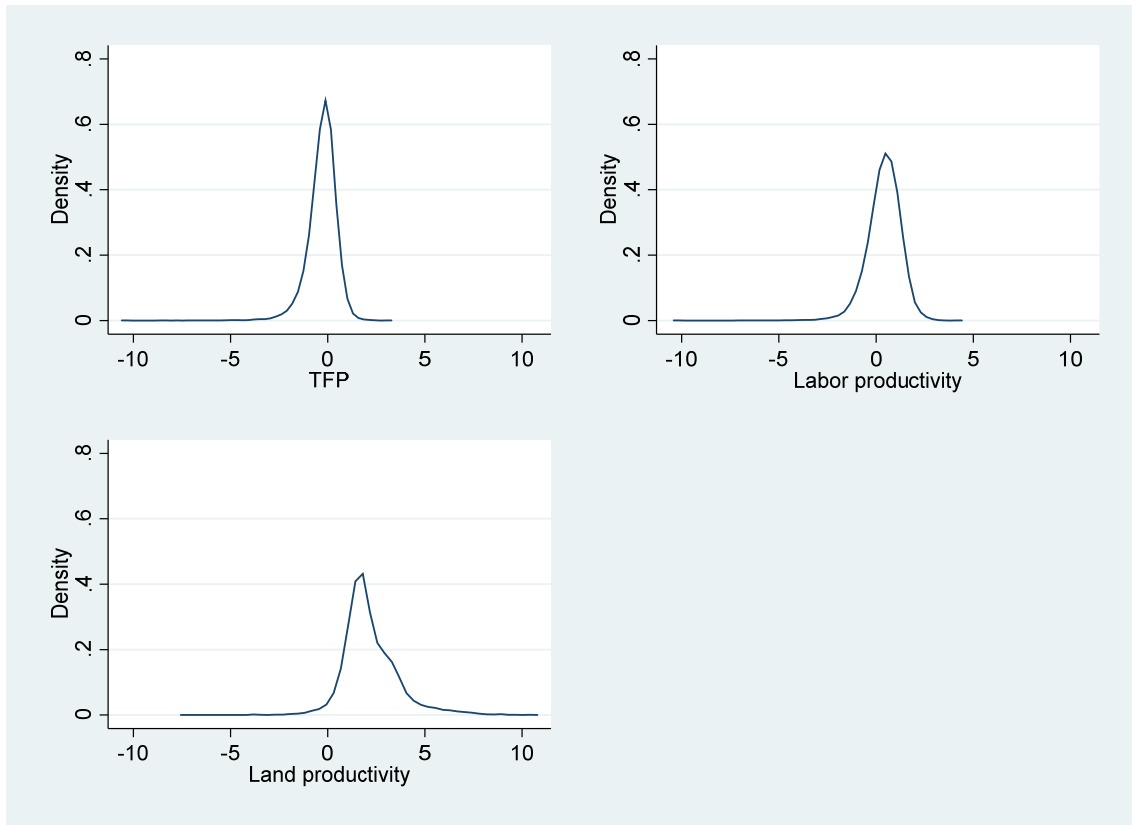
(skewness = –0.953 and –1.583, respectively).

Figure 1: Distributions of TFP, Labor, and Land Productivity

The correlation coefficients among these three productivity measures are presented in Table 3. For the whole agriculture sector, the correlation between TFP and labor productivity is 0.853 and between TFP and land productivity is 0.575. Scrutinizing these correlation coefficients by subsector, we find that TFP exhibits high correlation with these conventional productivity measures: 0.730–0.982 with labor productivity and 0.617–0.860 with land productivity. In all but

the dairy and fattening cattle farming sectors, TFP is more strongly correlated with labor than with land productivity. This implies that labor productivity could be a better proxy for TFP than land productivity, although land productivity is often used to measure productivity in agriculture.

Table 3: Correlations among Three Productivity Measures

	TFP vs.		Labor productivity vs.
	Labor productivity	Land productivity	Land productivity
Agriculture	0.853	0.575	0.377
1. Paddy field	0.938	0.779	0.617
2. Upland field	0.917	0.620	0.315
3. Outdoor-grown vegetables	0.882	0.617	0.385
4. Greenhouse-grown vegetables	0.967	0.655	0.496
5. Fruit	0.982	0.810	0.698
6. Outdoor-grown flowers	0.957	0.769	0.659
7. Greenhouse-grown flowers	0.962	0.830	0.724
8. Dairy	0.730	0.801	0.193
9. Breeding beef cattle	0.906	0.827	0.593
10. Fattening beef cattle	0.881	0.860	0.543
11. Pig	0.981	0.809	0.689
12. Poultry egg	0.957	0.749	0.550
13. Poultry broiler	0.900	0.675	0.417
14. Miscellaneous	0.936	0.758	0.534

The correlation coefficient between labor and land productivity for the agriculture sector is small, at 0.377. It is also small in several subsectors. This implies weak reliability of these single-factor productivity measures. Even if (or because) one measure reports high productivity for a sector, another measure does not necessarily do so. By contrast, TFP is found to be a robust

measure and correlates well with the single-factor productivity measures.

3.2 Pareto's *k* and Elasticity of Substitution among Varieties

Figure 2 shows the TFP distribution for the whole agriculture sample. This figure indicates an obvious inconsistency with the assumption for a Pareto distribution—the minimum value of the distribution must be equal to its mode. The subsample below the mode (0.530) accounts for 25% of the full sample.[3]

---

[3] There are 4,047 observations found below the mode in the full sample of 16,099 observations. Our microdataset is unique, encompassing smallholder samples. Many productivity analyses for Japanese manufacturing (e.g., Wakasugi et al. (2014)) have used the Basic Survey of Japanese Business Structure and Activities by the Ministry of Economy, Trade and Industry, which covers only enterprises with 50 or more employees and whose paid-up capital or investment fund is over 30 million JPY (roughly 270,000 USD). In this sense, they use already-truncated data in their manufacturing analyses, while we use data without any systematic truncation by farm size at the survey design stage.
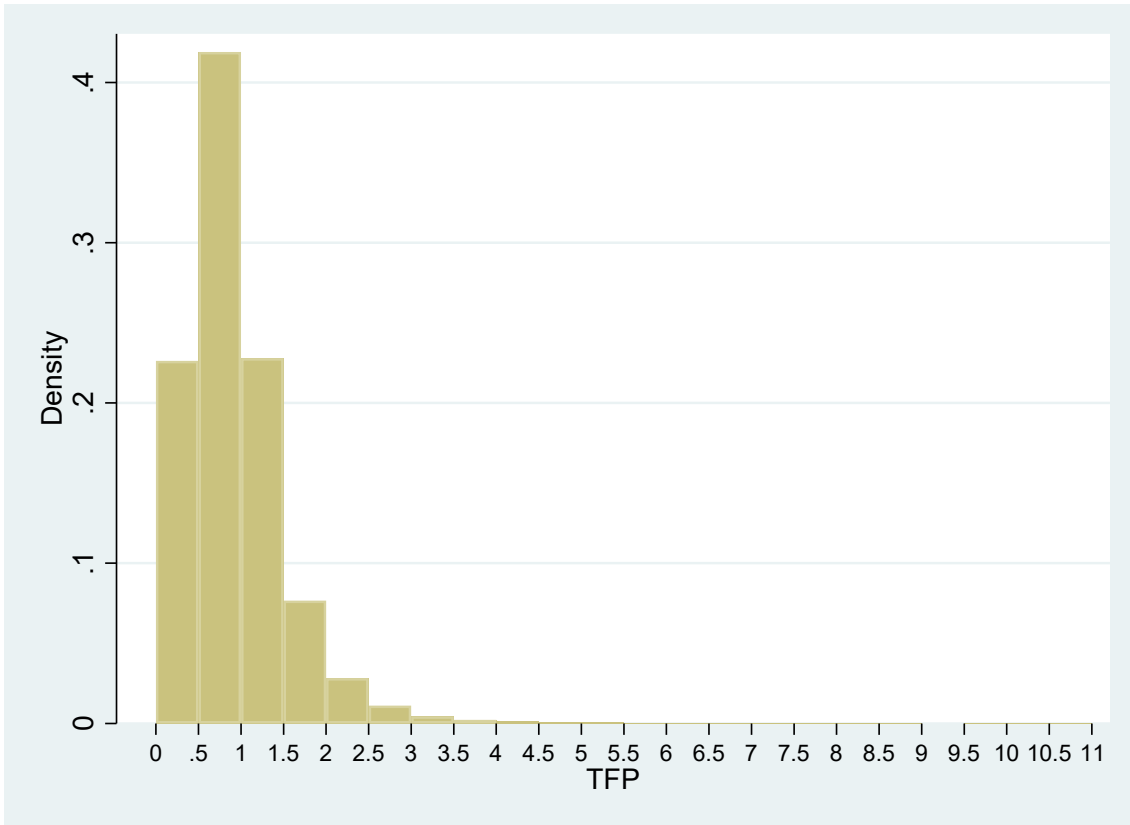
Figure 2: TFP Distribution for the Agriculture Sector

Note: For simplicity of presentation, one observation of paddy field farming (TFP = 22.884) is

dropped.

In Figure 3, the TFP distributions show the unique characteristics of subsectors,

including the different degrees of heterogeneity and modes. For example, the distribution of

fattening cattle farming has the same shape as a Pareto distribution. In the fruit and greenhouse-

grown flower farming sectors, the distributions have many observations below the mode values.

The number of below-mode observations varies widely by sector. As many studies of productivity

analyses and econophysics have reported, the power law is observed only for the upper part of

the right shoulder of the productivity distributions. Therefore, they *visually* choose cutoff points

and focus on the particular part of distributions to measure their regularity (Gabaix (2009)). We

truncate the data at the mode and estimate Pareto's k for the upper part of their distributions. At

the same time, following Crozet and Koenig (2010), we estimate the distributions of cumulative

production, which are not truncated, and compute the elasticity of substitution among varieties σ

using the estimate of k.[4]

---

[4] The summary statistics of TFP and the estimation results of Pareto's k with the full and truncated samples
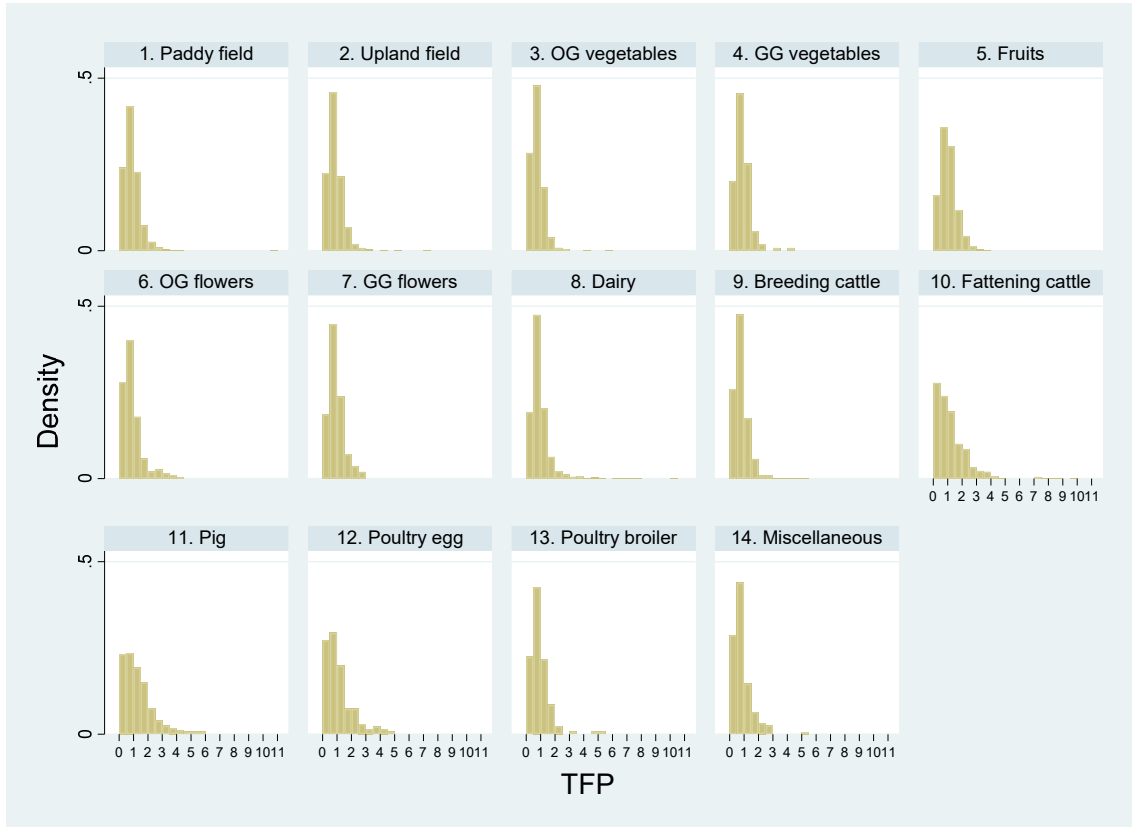 are shown in the Appendix.

Figure 3: TFP Distributions for Farming Sectors

Note: The TFP distribution of paddy field farming is truncated by dropping the largest TFP observation (= 22.884). OG and GG stand for "outdoor-grown" and "greenhouse-grown," respectively.

Table 4 reports the estimates of Pareto's k and the elasticity of substitution among varieties σ. Pareto's k is found to be relatively large for greenhouse-grown vegetables (2.60),

fruit (2.59), greenhouse-grown flowers (2.57), and upland field framing (2.54).[5] Many low

productivity farmers operate in these farming sectors, implying low heterogeneity among farmers.

By contrast, fattening cattle farming (1.17), poultry egg farming (1.34), and poultry broiler

farming (1.65) have greater heterogeneity.

---

[5] The large estimate of Pareto's k for pig farming (2.67) must be interpreted carefully. In its original TFP distribution, only 32% of observations exceed the mode value. Dropping the larger values may result in the underestimation of heterogeneity among farmers. When we use the mode of the distribution of the whole agriculture sector (0.530) as an alternative cutoff, we use 76% of the full sample and obtain an estimate of $k = 1.61$, which is comparable to the estimates for the other livestock farming sectors.

Table 4: Estimates of Pareto's k and Elasticity of Substitution among Varieties σ

| | From Eq. (8) | | From Eq. (10) | | |
	k	R^2	$k - \sigma + 1$	R^2	σ
Agriculture	2.31	(0.94)	0.74	(0.62)	2.58
1. Paddy field	2.45	(0.91)	0.98	(0.59)	2.48
2. Upland field	2.54	(0.93)	0.83	(0.44)	2.71
3. Outdoor-grown vegetables	2.15	(0.88)	0.95	(0.63)	2.20
4. Greenhouse-grown vegetables	2.60	(0.95)	0.30	(0.39)	3.30
5. Fruit	2.59	(0.88)	0.89	(0.55)	2.70
6. Outdoor-grown flowers	1.71	(0.93)	0.69	(0.66)	2.02
7. Greenhouse-grown flowers	2.57	(0.93)	0.75	(0.48)	2.82
8. Dairy	1.90	(0.97)	1.04	(0.78)	1.85
9. Breeding beef cattle	2.11	(0.93)	0.73	(0.50)	2.38
10. Fattening beef cattle	1.17	(0.82)	0.62	(0.54)	1.54
11. Pig	2.67	(0.96)	0.56	(0.58)	3.10
12. Poultry egg	1.34	(0.88)	0.72	(0.75)	1.63
13. Poultry broiler	1.65	(0.89)	1.19	(0.78)	1.47
14. Miscellaneous	1.55	(0.89)	0.65	(0.64)	1.90

Note: All estimates are statistically significant at the 1% level.

The elasticity of substitution among varieties σ is largest for greenhouse-grown vegetable farming (3.30), followed by pigs (3.10), greenhouse-grown flowers (2.82), upland field farming (2.71), and fruit (2.70). The degree of product differentiation is relatively low in these sectors. By contrast, livestock farming sectors, such as poultry broiler (1.47), fattening cattle (1.54), and dairy (1.85), have smaller elasticities of substitution and thus, seem to differentiate their products highly.

Let us compare our elasticity estimates with those in earlier studies. Kano et al. (2013) and Takechi (2015) used wholesale market data for fruits and vegetables to estimate new economic geography models for Japan. Their estimates were about 3–5 for leaf vegetables (cabbage, Chinese cabbage, lettuce, spinach, and green onion) and nearly 2 for root vegetables (carrot and potato). Ihara et al. (2015) also estimated the same model using data from *Seisan Nogyo Syotoku Tokei* (*Statistics of Agricultural Income Produced*) and obtained an elasticity estimate of 2.869 for vegetables. These results are generally consistent with our elasticity estimates of 2.20–3.30 for the vegetables sectors and 2.71 for upland field farming, which includes potato farming. Our estimates for the entire agriculture sector are also similar to the elasticity estimate of 3.79 for US manufacturing by Bernard et al. (2003), which is often used in empirical studies on manufacturing with the Melitz (2003) model.

4.   Concluding Remarks

In this study, we estimated the incidence of farm heterogeneity in the Japanese agricultural sectors by measuring TFP using farm-level microdata. Using the same dataset, we also estimated the elasticity of substitution among varieties to quantify the incidence of product differentiation jointly. Our TFP estimates are a viable productivity indicator, and are well correlated with conventional single-factor productivity measures. The estimated TFP distributions

imply relatively low heterogeneity in the greenhouse horticulture, upland field, and fruit farming sectors, while heterogeneity in the livestock sectors, such as fattening cattle and poultry egg, is found to be high. The elasticities of substitution among varieties for the livestock sectors are relatively low, indicating significant product differentiation.

In agricultural policy making, heterogeneity among farmers needs to be taken into account. Shocks, such as climate and extreme weather conditions and trade regime changes, affect farmers unevenly. Smallholders are often susceptible to negative shocks and need longer periods to recover because of low productivity in the absence of economies of scale. Highly motivated entrepreneurs may exploit the opportunity to expand their business. Measuring their heterogeneity with microdata helps us to form policies tailored to each group.

There are several limitations to our study. In Japan, the productivity of farmers has been improved by research and development activities at local public research laboratories and by spatial spillovers of farming and management skills from nearby entrepreneurs. Using these spatial factors could improve our TFP estimates. While our TFP estimates use aggregate data for 14 sectors, subsector data for commodities such as apples and Japanese *wagyu* beef may reveal different aspects of farm heterogeneity.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015) "Identification Properties of Recent Production Function Estimators," *Econometrica* 83, 2411–2451.

Akashi, J., and M. Tanemura (2006) "Kokusai-kyosoryoku-o-mochi-hajimeta Tohoku-san-nosanbutsu–Aomori-san-ringo-no Chugoku-shiken-yushutsu-no-shokai [Increasing Competitiveness of Made-in-Tohoku Apples–An Export Trial of Aomori-apples]," *JSCE Magazine Civil Engineering* 91, 38–41 (in Japanese).

Ball, E., D. Schimmelpfennig, and S. L. Wang (2013) "Is U.S. Agricultural Productivity Growth Slowing?," *Applied Economic Perspectives and Policy* 35, 435–450.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003) "Plants and Productivity in International Trade," *American Economic Review* 93, 1268–1290.

Cowell, F. A., and E. Flachaire (2015) "Statistical Methods for Distributional Analysis," AMSE Working Papers 1507, Aix-Marseille School of Economics, Marseille, France.

Crozet, M., and P. Koenig (2010) "Structural Gravity Equations with Intensive and Extensive Margins," *Canadian Journal of Economics* 43, 41–62.

Dixit, A., and J. E. Stiglitz (1977) "Monopolistic Competition and Optimum Product Diversity," *American Economic Review* 67, 297–308.

Fujimoto, S., and A. Ishikawa (2011) "Nihon-kigyo-no Uriage-jugyoin-suu-ni-mirareru Beki-

bunpu-oyobi-beki-shisu-no Kankei [Power-laws Observed in Sales and the Number of

Employees of Japanese Firms and the Relations among Indices of Power-laws],"

Journal of Kanazawa Gakuin University (Business Administration, Economics,

Informatics and Natural Sciences) 9, 25–33.

Gabaix, X. (2009) "Power Laws in Economics and Finance," *Annual Review of Economics* 1,

225–293.

Hu, B. (1995) "Nihon-nogyo-no Zenyoso-seisansei-hendo-no Seikaku-to-yoin [The Determinants

of Changes in Total Factor Productivity of Japanese Agriculture, 1960–90]," *Journal of*

Rural Problems 31, 103–111.

Ihara, R., R. Nakamura, and M. Morita (2015) "Kukan-keizaigaku-ni-motozuku Sutoro-koka-no

Kensho–Akashi-kaikyo-ohashi-o-jirei-toshite– [Examining the Straw Effect Based on a

Spatial Economic Theory–The Case of Akashi-kaikyo-Strait Bridge–]," RIETI

Discussion Paper Series 15-J-045 (in Japanese).

Inui, T., Y. Kim, H. Kwon, and K. Fukao (2011) "Seisan-sei-dogaku-to Nihon-no-keizai-seicho:

'Hojin-kigyo-tokei-chosa' Kohyo-deta-ni-yoru Jissho-bunseki [Productivity Dynamics

and Japan's Economic Growth: An Empirical Study with Microdata of *Financial*

Statements Statistics of Corporations by Industry]," RIETI Discussion Paper Series 11-

J-042 (in Japanese).

Kano, K., T. Kano, and K. Takechi (2013) "Exaggerated Death of Distance: Revisiting Distance

Effects on Regional Price Dispersions," *Journal of International Economics* 90, 403–

413.

Kondo, K, Y. Yamamoto, and J. Sasaki (2005) "Gentan-kaishi-ki-iko-ni-okeru Wagakuni-

inasaku-seisansei-keizai-shusoku-bunseki [Convergence of Rice Productivity in Japan

under Rice Acreage Control Programmes]," *Japanese Journal of Farm Management* 43,

45–48 (in Japanese).

Krugman, P. (1980) "Scale Economies, Product Differentiation, and the Pattern of Trade,"

American Economic Review 70, 950–959.

Kunimitsu, Y. (2011) "Nihon-nogyo-ni-okeru Zenyoso-seisan-sei-no Henka-to Eikyo-yoin

[Changes in Total Factor Productivity in the Japanese Agriculture]," *Journal of Rural

Economics Special Issue 2011*, 1–8 (in Japanese).

Kunimitsu, Y. (2014) "Nihon-no Inasau Zenyoso-seisansei-ni-okeru Chiiki-kan-kakusa-no-

Doko: Kukan-keiryo-keizai-moderu-no Tekiyo-ni-yoru-chiken [The Trend of Regional

Gaps in Japanese Rice Total Factor Productivity: Fact Findings from Spatial

Econometric Analysis]," *Journal of Rural Problems* 49, 501–510 (in Japanese).

Kuroda, Y. (2013) *Production Structure and Productivity of Japanese Agriculture*, Palgrave

Macmillan.

Levinsohn, J., and A. Petrin (2003) "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies* 70, 317–341.

Marschak, J., and W. Andrews (1944) "Random Simultaneous Equations and the Theory of Production," *Econometrica* 12, 143–205.

Matsuura, T., K. Hayakawa, and M. Kato (2008) "Mikuro-deta-ni-yoru Seisansei-bunseki-no Kenkyu-doko: Sannyu-taishutsu, Keizai-no Gurobarizeshon-inobeshon-seido-kaikaku-no Eikyo-o-chushin-ni [A Survey of Research on Productivity Analysis Using Microdata–Impact of Entry, Exit, Economic Globalization, Innovation and Systemic Reform]," RIETI Policy Discussion Paper Series 08-P-007 (in Japanese).

Mayer, T., and G. I. Ottaviano (2007) *The Happy Few: The Internationalization of European Firms: New Facts Based on Firm-level Evidence*, Bruegel, Brussels.

Melitz, M. J. (2003) "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity," *Econometrica* 71, 1695–1725.

Melitz, M. J., and G. I. Ottaviano (2008) "Market Size, Trade, and Productivity," *Review of Economic Studies* 75, 295–316.

Ministry of Agriculture, Forestry and Fisheries (MAFF) (2008) *Heisei-20-nendo Shokuryo-nogyo-noson Hakusho [Annual Report on Food, Agriculture and Rural Area in Japan FY 2008]*, The Government of Japan.

MAFF (2018a) "Eino-ruikei-betsu Keiei-tokei-kara-mita Ichi-nogyo-keiei-tai-atari-no Keiei-jokyo (Suikei) (Heisei 25-nen–Heisei 29-nen) [Estimated Farm Management Activities of Average Farmers Based on Statistics on Management by Farming Type for 2013–2017]," The Government of Japan (in Japanese). http://www.maff.go.jp/j/tokei/kouhyou/noukei/einou_kobetu/attach/pdf/index-11.pdf (Accessed on May 13, 2019.)

MAFF (2018b) "Heisei 29-nendo Yushutsu-ni-torikumu Yuryo-jigyosha-hyosho: Jusho-sha-no Torikumi-naiyo [Commendation of Good Practice for Exports by Agri-food Producers in Fiscal 2017: Good Practices of Awardees]," The Government of Japan (in Japanese). http://www.maff.go.jp/j/shokusan/export/torikumi_zirei/attach/pdf/180202-8.pdf (Accessed on November 19, 2018.)

Mizuno, T., A. Ishikawa, S. Fujimoto, and T. Watanabe (2012) "Power Laws in Firm Productivity," *Progress of Theoretical Physics Supplement* 194, 122–134.

Mollisi, V., and G. Rovigatti (2017) "Theory and Practice of TFP Estimation: The Control Function Approach Using Stata," CEIS Working Paper No. 399.

Nakamura, T., A. Maruyama, and Y. Yano (2011) "Kobai-sentaku-kodo-kara-mita Aomori-san-ringo-no Herushinki-shijo-eno Yushutsu-no-kanosei–Finland WINE, FOOD & GOOD LIVING 2009-ni-okeru Anketo-chosa-o-mochiite– [Export Potential of Aomori Apples

in the Helsinki Market–Based on the Questionnaire Survey at Finland WINE, FOOD & GOOD LIVING 2009–]," *Journal of Agricultural Development Studies* 22, 21–32 (in Japanese).

Olley, G. S., and A. Pakes (1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica* 64, 1263–1297.

Shimowatari, T. (2018) *Nihon-no-sanchi-to Yushutsu-sokushin–Nihon-san-nosanbutsu Shokuhin-no-gurobaru-shijo-eno Chosen– [Japanese Farmers and Export Promotion– Challenges of Japanese Agri-food Products in the Global Markets–]*, Tsukuba-shobo Publishing, Tokyo (in Japanese).

Shumway C. R., B. M. Fraumeni, L. E. Fulginiti, J. D. Samuels, and S. E. Stefanou (2016) "U.S. Agricultural Productivity: A Review of USDA Economic Research Service Methods," *Applied Economic Perspectives and Policy* 38, 1–29.

Takechi, K. (2015) "The Quality of Distance: Quality Sorting, Alchian-Allen Effect, and Geography," RIETI Discussion Paper Series 15-E-018.

Tanaka, S. (2006) "Nihon-san Ringo-yushutsu-ni-okeru Sanchi-ryutsu-shutai-no Yakuwari: Aomori-ken-san Ringo-o Jirei-to-shite [The Role of Local Distributers for Apples Produced in Japan for Export: A Case Study of Aomori Prefecture]," *Review of Agricultural Economics Hokkaido University* 62, 141–150 (in Japanese).

Tokui, J., T. Inui, and Y. Kim (2007) "Taika-sareta Gijutsu-shinpo-to Shihon-no Heikin-vinteji

[Embodied Technological Progress and Average Capital Vintage]," RIETI Discussion

Paper 07-J-035 (in Japanese).

Yamamoto, Y., K. Kondo, and J. Sasaki (2007) "Will the Growth Rate of Japan's Rice

Productivity Decline to Zero Percent in the Future?," *Journal of Rural Economics* 79,

154–165.

Wakasugi, R., B. Ito, T. Matsuura, H. Sato, A. Tanaka, and Y. Todo (2014) Features of Japanese

Internationalized Firms: Findings Based on Firm-level Data," in: R. Wakasugi (ed.)

Internationalization of Japanese Firms: Evidence from Firm-level Data, Springer.

Appendix: Details of Data

A.1   Data Sources

We use the microdata of three types of farmers—individual management, management by judicial organization, and management by arbitrary organization—in *Nogyo Keiei Tokei Chosa* (*Statistical Survey on Farm Management and Economy*) for 2012–2015.[6] Using this dataset, we compute value added by subtracting material inputs from total agricultural sales, which are deflated by relevant price indexes reported in *Nogyo Bukka Tokei Chosa* (*Statistical Survey on Prices in Agriculture*). Labor input is measured by man-hours of time of employers, their families, and employees. Capital stock data are fixed capital stock, such as buildings, automobiles, farming machines, plants, and end-of-year cattle stock values (land is considered separately, as explained later). Their book values are converted into market values. Following Inui et al. (2011), we use the following formula to convert the book value of capital stock of farmer i in year y KNB_{it} to its market value K_{it},

$$K_{it} = KNB_{it} \frac{K_t^{AG}}{KNB_t^{AG}}$$

using the ratio of the market value of the total agricultural capital stock K_t^{AG}, obtained from *Nogyo-Shokuryo Kanren-sangyo-no Keizai-keisan* (*Economic Accounts for Agriculture and Food*

---

[6] The full sample contains about 20,300 observations. After dropping observations with missing values, we obtain an unbalanced panel with 16,099 observations. This dataset has 3,193 sample farmers with complete observations for the 4 years.

Related Industries) by MAFF, vis-à-vis its book value KNB_t^{AG}, which is constructed by aggregating the book value of farmer i in farming sector j in year t KNB_{ijt}, weighted by the sampling rate $rsmp_{jt}$ $(KNB_t^{AG} = \sum_i \sum_j KNB_{ijt} \times rsmp_{jt})$.[7]

Land input data are obtained from *Nogyo Keiei Tokei Chosa* (*Statistical Survey on Farm Management and Economy*), covering owned and borrowed land to cultivate agricultural products. The summary statistics are reported in Table A.1.

Table A.1: Summary Descriptive Statistics of the Sample

	Units	Observations	Mean	S.D.	Min	Max
Value-added	thousand JPY	16,099	10,814	22,991	0	626,628
Labor	hours	16,099	66,146	150,517	5	4,188,609
Capital	thousand JPY	16,099	4,744	6,600	18	174,136
Land	ares	16,099	1,353	2,404	1	35,560
Investments	thousand JPY	16,099	2,926	14,817	0	558,073
Intermediates	thousand JPY	16,099	10,418	24,603	26	605,689

Note: Land consists of cultivated land for paddy and upland field farming and pasture. Values are deflated using the relevant price indexes.

The coefficient dummy variables reflect factor intensity. Capital and land inputs relative to labor input are defined as capital and land intensity, respectively. When a sector has capital (or

---

[7] Matsuura et al. (2007) attributed this conversion method with the book value–market value ratios to Tokui et al. (2007). This conversion method was also used by Wakasugi et al. (2014) and Inui et al. (2011).

land) intensity exceeding its median value among the 14 sectors in 2015, it is categorized as a capital- or land-intensive sector, whose dummy variable takes the value 1. Capital-intensive sectors are paddy field farming, upland field farming, fruit, dairy, fattening cattle, pigs, poultry egg, and other sectors. Land-intensive sectors are paddy field farming, upland field farming, outdoor-grown vegetables, dairy, breeding cattle, fattening cattle, and miscellaneous farming sectors (Figure A.1).
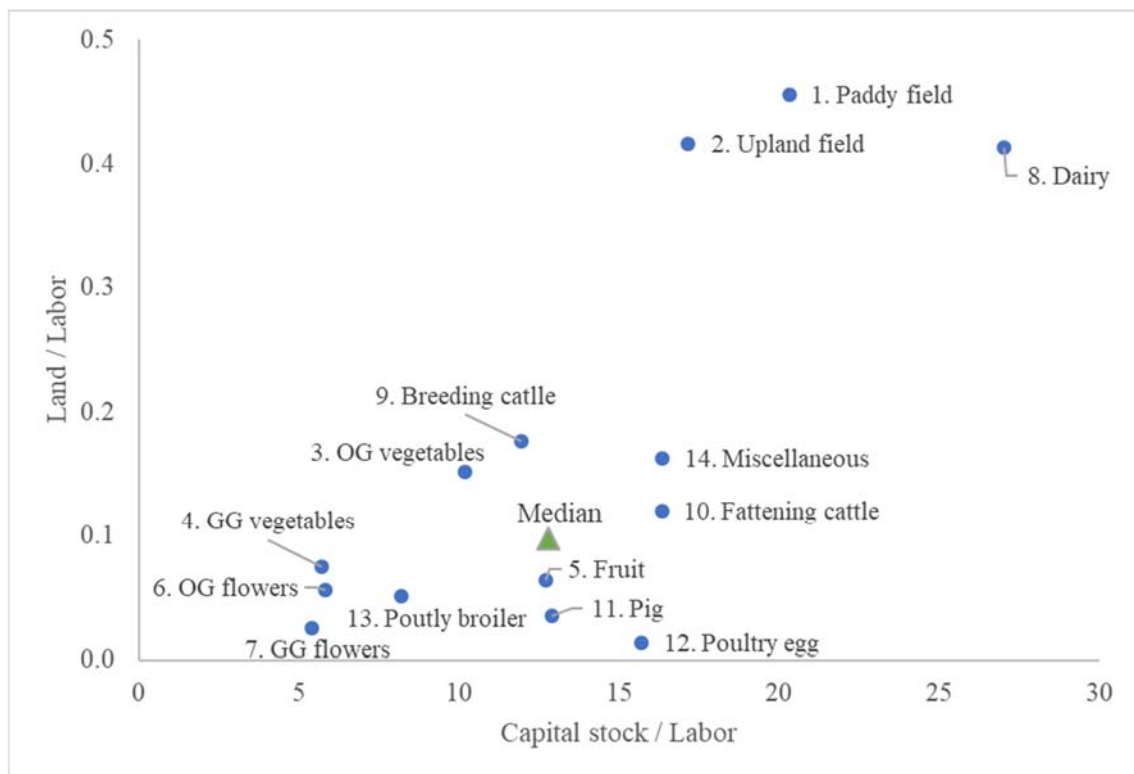

Figure A.1: Capital-labor and Land-labor Ratios in 2015

Note: OG and GG are "outdoor-grown" and "greenhouse-grown," respectively.

A.2   Data Truncation

When we use the full TFP sample, we obtain Pareto's $k = 1.08$ (Table A.2), which looks

too small with a poor fit of $R^2 = 0.63$ (the broken line in Figure A.2).[8] Thus, we truncate the

distribution with a strictly positive cutoff for TFP_M in (6). When the mode of the TFP

distribution is used for this cutoff, we obtain a reasonable estimate of $k = 2.31$ (Table A.2) with a

good fit of $R^2 = 0.94$ (the solid line in Figure A.2). In the literature, the cutoff point is often

visually determined (Gabaix (2009)); few systematic or statistical methods have been applied

(Fujimoto and Ishikawa (2011)).[9] Figure A.2 indicates that the fit could be improved further with

a larger cutoff, leading to a larger estimate of k. Our estimate of k thus implies the lower bound

of possible estimates.

---

[8] Our dataset includes smallholder observations, which are often noisy. Many productivity analyses for Japanese manufacturing (e.g., Wakasugi et al. (2014)) have used the *Basic Survey of Japanese Business Structure and Activities* by the Ministry of Economy, Trade and Industry of Japan, which covers only enterprises with 50 or more employees and whose paid-up capital or investment fund is over 30 million JPY (roughly 270,000 USD). In this sense, they used pretruncated data in their manufacturing analyses, while we use data without any systematic truncation by farm size at the stage of survey design.

[9] Fujimoto and Ishikawa (2011) assumed a joint log-normal distribution and a Pareto distribution and estimated their cutoff point statistically.

Table A.2: Estimates of Pareto's *k* for the Full and Truncated Samples

	Full sample	Trancated sample
lnTFP	-1.0839***	-2.3136***
	(0.0065)	(0.0055)
Constant	-1.3145***	-1.2421***
	(0.0051)	(0.0023)
Observations	16,098	12,051
R2	0.63	0.94
RMSE	0.606	0.2506

Note: Standard errors in parentheses. *** indicates significance at the 1% level.
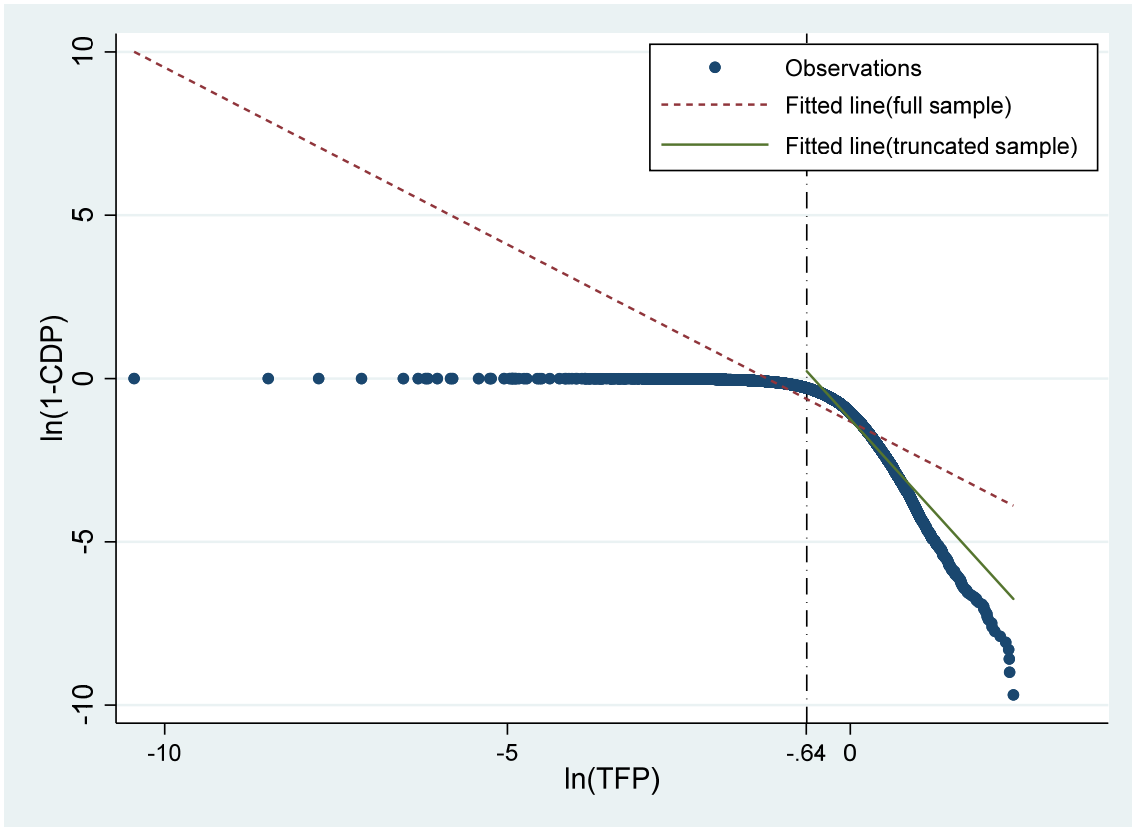
Figure A.2: Pareto's *k* Estimated for the Full and Truncated Samples

Note: The mode is 0.530; log(0.530) = −0.64.

A.3 Summary Statistics of TFP Estimates

Table A.3: Summary Descriptive Statistics of TFP Estimates

	Observations	Mean	S.D.	Min	Max
Full samples	16,099	0.930	0.662	0.00003	22.884
$TFP_{it} \geq TFP_M$	12,052	1.128	0.651	0.53014	22.884
$TFP_{it} < TFP_M$	4,047	0.341	0.133	0.00003	0.5299

Note: The mode (TFP_M) is 0.530.

Table A.4: Summary Descriptive Statistics of TFP Estimates by Farming Sector

|	Full samples						$TFP_{it} \geq TFP_M$		
	Observations	Mean	S.D.	Min	Max	Mode	Observations	Mean	S.D.
1. Paddy field	7,286	0.887	0.597	0.001	22.884	0.530	5,334	1.087	0.575
2. Upland field	2,414	0.872	0.536	0.000	7.012	0.550	1,769	1.064	0.496
3. Outdoor-grown vegetables	606	0.770	0.477	0.007	5.531	0.339	528	0.849	0.458
4. Greenhouse-grown vegetables	369	0.903	0.560	0.000	4.299	0.595	261	1.108	0.542
5. Fruit	2,021	1.046	0.554	0.026	3.608	0.614	1,557	1.238	0.482
6. Outdoor-grown flowers	189	0.933	0.731	0.010	4.217	0.363	159	1.063	0.725
7. Greenhouse-grown flowers	252	0.937	0.534	0.044	2.806	0.637	179	1.154	0.478
8. Dairy	1,448	1.005	0.910	0.021	10.227	0.414	1,273	1.102	0.929
9. Breeding beef cattle	416	0.843	0.602	0.009	5.055	0.407	347	0.963	0.586
10. Fattening beef cattle	396	1.259	1.269	0.027	9.663	0.238	340	1.442	1.279
11. Pig	280	1.312	1.064	0.003	5.881	1.559	90	2.510	1.024
12. Poultry egg	136	1.159	0.928	0.161	4.789	0.298	125	1.240	0.924
13. Poultry broiler	129	0.939	0.745	0.086	5.431	0.318	117	1.009	0.748
14. Miscellaneous	157	0.865	0.673	0.041	5.052	0.244	147	0.912	0.669
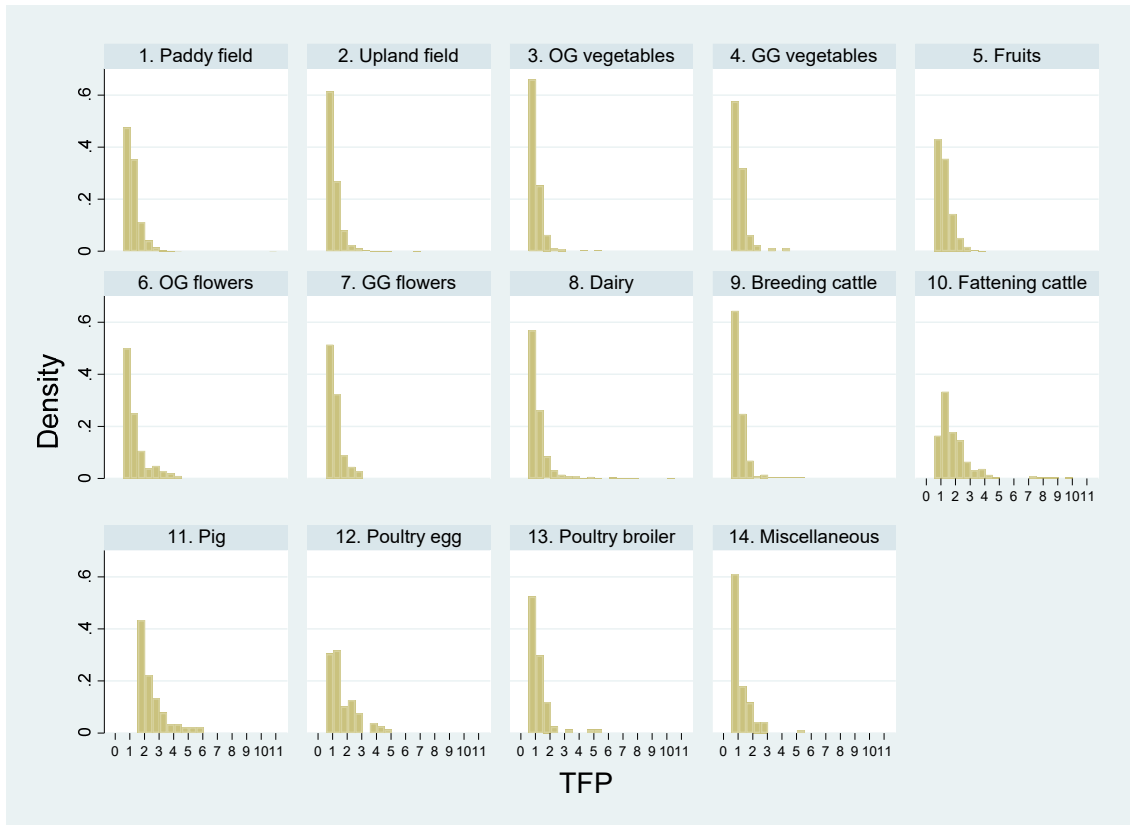
Figure A.3: TFP Distributions Truncated at the Modes

Note: For clarity, the TFP distribution of paddy field farming is truncated by dropping the largest

TFP observation (= 22.884). OG and GG are "outdoor-grown" and "greenhouse-grown,"

respectively.