RIETI Discussion Paper Series 18-E-017

Markup and Market Size: Evidence from Japan (Revised)

KONDO, Keisuke
RIETI

Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 18-E-017
First Draft: March 2018
Revised: May 2020
Revised: December 2023

Markup and Market Size: Evidence from Japan*

Keisuke KONDO†
RIETI & Kobe University

Abstract

This study empirically investigates how market size affects markups in the Japanese manufacturing sector. Recently developed models on monopolistic competition with endogenous price-cost markups show that markups in larger markets are lower because competition is stronger. This study proposes a new empirical approach to identify effective geographical range of market competition that affect markups in the tradable goods sector. The approach in this study is novel because market size is measured as the market potential within the threshold distance from 100 km until 1,000 km. This study finds both the size of market in closer proximity to the production location and the size of the distant market affect markups, suggesting that manufacturing establishments face stronger competition in geographically wider markets.

JEL classification: D24, F14, L11, L60, R32
Keywords: Markup, Productivity, Market size, Firm size

* I thank Kozo Kiyota, Toshiyuki Matsuura, Shota Araki, Hirokazu Ishise, Masayuki Morikawa, Yasusada Murata, Atsushi Nakajima, Yoshiyuki Nakazono, Akira Sasahara, Yasuyuki Todo, Jens Wrona, Makoto Yano, and participants of the RIETI DP Seminar, the 8th European Meeting of the Urban Economics Association, the 2018 Spring Meeting of the Japanese Economic Association, the 77th Annual Meeting of the Japan Society of International Economics, and the seminar in the Yokohama City University for their useful comments. I am indebted to Lionel Nesta for his special lecture, "Estimation of firm level markups," at Keio University in July 2015. Naturally, any remaining errors are my own. This study uses microdata (questionnaire information) from the Census of Manufacture (Ministry of Economy, Trade and Industry) and the Economic Census for Business Activity (Ministry of Internal Affairs and Communications and Ministry of Economy, Trade and Industry) with permission under the Statistics Act of Japan. This study is an outcome of the research conducted under the "RIETI Data Management Project." This research uses the establishment-level panel converter table developed by the aforementioned project. This paper was previously circulated under the title "Markups, city size, and exports: Evidence from Japan" in the RIETI Discussion Paper Series.

† Research Institute of Economy, Trade and Industry (RIETI): 1-3-1 Kasumigaseki, Chiyoda-ku, Tokyo, 100-8901, Japan. (E-mail: kondo-keisuke@rieti.go.jp)

1 Introduction

Monopolistic competition model developed by Dixit and Stiglitz (1977) has become a cornerstone of economic theory. Price-setting behavior is the key difference between monopolistic and perfect competition. Monopolistic competition theory defines that firms charge markups to maximize their profits; this is supported by the empirical economic literature (e.g., Hall, 1986, 1988). Although a standard monopolistic competition model exhibits uniform productivity across firms and constant markups, recent theoretical studies in the urban and international economics literature incorporate both firm heterogeneity in productivity and endogenous markups. The markup estimation approach developed by De Loecker and Warzynski (2012) bridges the gap between theoretical and empirical studies on this topic.[1]

The novelty of this study is to uncover how the geographical range of markets affects markups in the tradable goods sector. In empirical studies, considering the geographical boundary of markets is important because the manufacturing goods are tradable across markets. In the literature of international trade, some studies distinguish between domestic and export markets to exploit data advantages, such as De Loecker and Warzynski (2012), Lu and Yu (2015), Bellone et al. (2016), De Loecker et al. (2016), and Georgiev (2018). Although either the metropolitan area or the commuting zone is often used in the literature on urban economics, selling in multiple regional markets should be considered with regard to market competition.[2] This study addresses this issue, by proposing a market potential approach with distance boundary.

The theoretical literature on urban economics has studied the underlying mechanism of firms' price-setting and the market competition. According to both Melitz and Ottaviano (2008) and Combes

[1]In the literature on industrial organization, the markup estimation has been studied to understand the competitiveness of markets. The markup is generally defined as the ratio of the price and marginal cost of production. Focusing on firms' production side, Hall (1986, 1988) develops an econometric model for markup estimation. In his framework, the markup is estimated as a coefficient parameter in the regression. Thus, the estimated markups are constant at the industry level and have no variation across firms within industries. De Loecker and Warzynski (2012) extend the work of Hall (1986, 1988) in terms of the firms' cost-minimization problem and combine it with production function estimation to propose a markup estimation approach, which considers cross-firm variation. In this approach, a markup is estimated, as well as the production function estimation. Because the literature on total factor productivity (TFP) includes seminal studies on production function estimation, such as Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015), research on markup estimation can be effectively extended from productivity analysis. In turn, the markup can also be estimated from the consumer demand perspective. De Loecker and Scott (2016) provide an interesting comparison in markup estimation between production and demand approaches.

[2]For example, Behrens et al. (2017) examines firm selection in a different model setting with quantitative simulations using the U.S. MSA-level data. They find that additional productivity gains from agglomeration economies are small. In other words, the strength of the competition from pro-competitive effects in large cities plays an important role.

et al. (2012), stronger competition in larger market decreases firms' markups, resulting in a negative correlation between markups and market size. Firm factors also affect markups. Theoretically, Melitz and Ottaviano (2008) show that firms with higher productivity charge higher markups. This prediction is essentially identical to the findings presented in recent theoretical studies by Behrens et al. (2014, 2017) and Bellone et al. (2016). An important implication of these theoretical studies is that markups are determined by a number of firm-level and market competition-related factors. Thus, this study elucidates the market competition factors on markups after controlling for firm/establishment-level factors.

Using the establishment-level panel dataset in the Japanese manufacturing sector, this study finds that manufacturing establishments face market competition not only in the production location's neighboring markets but also in geographically wider markets. Further, it is shown that larger markets strengthen competition, leading to lower markups, as shown by Hottman (2016) in the U.S. retail sector and Lu et al. (2014) in the Chinese manufacturing sector.

These findings draw important inferences for market competition and firm selection (i.e., stronger competition in larger markets forces less productive firms to exit and, as a result, more productive firms operate in such locations) by relating spatial productivity differences with spatial markup differences. In the literature on urban economics, Combes et al. (2012) extend the framework of Melitz and Ottaviano (2008) and provide a new empirical approach that distinguishes agglomeration economies from firm selection for the spatial differences in productivity. They conclude that selection has no explanation power for spatial productivity differences. Kondo (2016) also obtains similar results using the Japanese manufacturing establishment-level data. However, defining the boundary of the market for tradable sectors in an empirical setting is challenging because the production location is not necessarily identical to the demand location. Whereas the literature on agglomeration economies tends to focus on supply-side factors, such as input-output linkage, labor market pooling, and knowledge spillover, within the metropolitan area or commuting zones of a production location, market competition in the tradable goods sector is not restricted to local markets. As discussed by Accetturo et al. (2018), market competition in distant markets is not considered in Combes et al. (2012). As a conclusion, this study emphasizes the importance of considering market potential as an alternative measure of market size.

The remainder of this paper is organized as follows. Section 2 explains the estimation strategy. Section 3 summarizes the dataset of Japanese manufacturing establishments. Section 4 discusses the estimation results. Finally, Section 5 presents the conclusions.

2 Empirical Strategy

The empirical strategy consists of two steps. In the first step, this study estimates area factors of markups by regressing establishment-level markups on establishment factors with area dummies. In the second step, it is investigated whether market size explains the variations in markups estimated in the regression analysis from the first step.

2.1 Extracting Area Factors from Establishment-Level Markups

To investigate the market size effect on markups, this study first extracts area factors from the establishment-level markups following the two-step method suggested by Combes and Gobillon (2015). The first-step regression is given by

$$\begin{aligned} \log(\hat{\mu}_{iast}) = \gamma_1 \log(\widehat{\mathrm{TFP}}^{\omega}_{it}) + \gamma_2 \log(\widehat{\mathrm{TFP}}^{\varepsilon}_{it}) + \gamma_3 \mathrm{Multi}_{it} + \gamma_4 \log(\mathrm{Labor}_{it}) \\ + \gamma_5 \log(\mathrm{Capital}_{it}) + \gamma_6 \mathrm{Export}_{it} + \eta_{at} + \nu_s + u_{iast}, \end{aligned} \tag{1}$$

where $\log(\hat{\mu}_{it})$ is the logarithm of estimated markup for establishment i in area a in sector s and year t, $\log(\widehat{\mathrm{TFP}}^{\omega}_{it})$ and $\log(\widehat{\mathrm{TFP}}^{\varepsilon}_{it})$ are the logarithm of TFP (measuring input efficiency and other efficiency, respectively), Multi_{it} is a dummy variable for a multi-establishment firm, $\log(\mathrm{Labor}_{it})$ is the logarithm of employment, $\log(\mathrm{Capital}_{it})$ is the logarithm of financial capital, Export_{it} is the export dummy which takes the value of 1 if establishment i exports in year t and 0 otherwise, $\boldsymbol{\gamma} = (\gamma_1\ \gamma_2\ \gamma_3\ \gamma_4\ \gamma_5\ \gamma_6)'$ is the parameter vector of these corresponding variables, η_{at} is a dummy variable of cross-term between area a and year t, ν_s is a dummy variable for sector s, and u_{it} is an error term.

Based on findings from De Loecker et al. (2016), this study introduces two types of TFP (cost-saving productivity related to inputs and productivity related to other factors) into the regression. Traditionally, TFP is calculated as the residuals of a production function regression. A seminal paper in the economics literature by Olley and Pakes (1996) proposes a structural estimation approach to distinguish productivity factors within the residuals. This idea is followed by Levinsohn and Petrin (2003) and Ackerberg et al. (2015). In the literature on international trade, De Loecker et al. (2016) examine the influence of trade reform on markups and find that, whereas decreasing output tariff leads to lower markups because of stronger competition, a reduction in the input tariff allows firms to charge higher markups by achieving cost saving. That is, trade reform has two opposite channels of effects on markups through output and input prices, which is examined in this study by considering two types of TFP.[3]

The first-step regression controls for productivity, the existence of multiple establishments, establishment size (number of employees), capital, and export status. As studied by Melitz and Ottaviano (2008), more productive firms can charge higher markups. As studied by Kugler and Verhoogen (2012), large establishments may charge higher markups than small establishments, because of higher quality goods. As pointed out by De Loecker and Warzynski (2012), exporting firms show higher markups. These factors are excluded from the aggregate markups estimated at the establishment level.

This first-step regression clarifies the extent to which establishment factors explain variations in markups and the parameter estimate $\hat{\eta}_{at}$ captures locational factors of markups. The estimated area-year factors of markups are used for the analysis in the next step.

2.2 Markup Variations and Market Size

The second-step regression using the area-year factors, $\hat{\eta}_{at}$, estimated in the first-step regression (1) is given by

$$\hat{\eta}_{at} = \psi^{d\,\mathrm{km}} \log(\mathrm{MP}_{at}^{d\,\mathrm{km}}) + \pi_p + \tau_t + e_{at}, \tag{2}$$

where $\mathrm{MP}_{at}^{d\mathrm{km}}$ is the market size variable (i.e., market potential) within d km from the production location, $\psi^{d\,\mathrm{km}}$ captures the market-size elasticity of the markup, π_p is the prefecture dummy, τ_t is the year dummy, and e_{at} is an error term. The prefecture dummy is introduced to control for area fixed effects. For example, the minimum wage, which affects markups, is stipulated at the prefecture level. Note that the observations are weighted by the number of establishments observed in each area to avoid over-evaluation of areas with small number of establishments.

The market size is measured by the market potential, which is calculated as a distance-weighted sum of market demand. The novelty of the approach of this study is to introduce geographical boundary into the market potential. The market potential within d km from the center of area a in

[3]By contrast, firm heterogeneity as defined by Melitz (2003) considers the input efficiency in production and not the positive externality in production. In the empirical literature on markups, for example, Bellone et al. (2016) use only TFP related with input efficiency. See Appendix A for more details.

year t is calculated as follows:

$$\mathrm{MP}_{at}^{d\,\mathrm{km}} = \sum_{b=1}^{N(d)} Y_{bt} D_{ab}^{-1} \quad \text{for} \quad D_{ab} < d\,\mathrm{km} \tag{3}$$

where Y_{at} is total income in area a in year t, D_{ab} is the great-circle distance between areas a and b, d is the threshold distance of market boundary, and $N(d)$ is the number of areas covered within d km from the location of city hall of area a. Note that the standard market potential is defined when $d = \infty$, and $N(d = \infty)$ indicates the number of all municipalities within the country. This new approach uncovers marginal effects of geographical extension of market area that affect markups.[4]

The parameter $\psi^{d\,\mathrm{km}}$ is used to simulate the spatial variations in markups among different markets. The basic formula for quantification is given by $(\mathrm{MP}_a^{d\,\mathrm{km}}/\mathrm{MP}_b^{d\,\mathrm{km}})^{\hat{\psi}^{d\,\mathrm{km}}} - 1$, which means that the spatial percentage change in markup between areas a and b depends on the market-size ratio $\mathrm{MP}_a^{d\,\mathrm{km}}/\mathrm{MP}_b^{d\,\mathrm{km}}$ and market-size elasticity of the markup $\hat{\psi}^{d\,\mathrm{km}}$. It is suggested that manufacturing establishments face tougher competition in distant markets if the spatial percentage change in markups monotonically increases in threshold distance d.

2.3 Instrumental Variables Method

The market-size elasticity of markup $\psi^{d\,\mathrm{km}}$ might be estimated with bias using the standard OLS method. To obtain a consistent estimate, this study controls for the endogeneity bias. The monopolistic competition model with an endogenous markup, such as in Behrens et al. (2014), reveals that a larger market shows a higher wage, higher average productivity, lower average markup, and higher welfare than a smaller market. Whereas larger markets lead to lower markups, the lower markups also lead to a larger market size through the in-migration channel related to higher wages. If this reverse causal relation is not controlled for, the magnitude is underestimated because of the positive effects of the latter channel. Another possibility of bias arises from the measurement errors because researchers do not know the exact information on where establishments sell their products. A simplified market potential approach may lead to attenuation bias.

To address these endogeneity issues, this study depends on the instrumental variable (IV) esti-

[4]Donaldson and Hornbeck (2016) propose a market access approach to examine the historical impact of railroads on the U.S. economy. As a robustness check, they consider a measure of market access that is limited to counties beyond 100 miles of each county to control for local shocks that affect land values. This study takes an extensive view of market access to capture the impact of market size on markups. This study considers a measure of market potential that is limited to areas within each 100 km radius (from 100 km to 1,000 km), which clarifies an effective market range that affects markups.

mation using the long-lagged variable suggested by Ciccone and Hall (1996). The IV approach is identical to that proposed in the literature on urban wage premium (e.g., Combes and Gobillon, 2015). Many previous studies, such as Combes et al. (2010) and de la Roca and Puga (2017), confirm that including the long-lagged city-size variable as an IV helps control for endogeneity. The present study makes the novel proposition of incorporating the population potential in 1930 for market potential as follows:

$$\mathrm{PP}_{a,1930} = \sum_{b=1}^{N} P_{a,1930} \exp(-\delta \times D_{ab}), \tag{4}$$

where $P_{a,1930}$ is the population in area a in 1930, δ is the distance-decay parameter ($\delta = 0.05$ is used in this study). Note that this specification takes an exponential distance weighting instead of an inverse distance weighting to avoid the inverse of 0 km. This is necessary because some areas share the same locations in the present study, given the administrative unit inconsistency.

3 Dataset and Estimated Markups

This study uses data on the Japanese manufacturing sector obtained from confidential datasets of the Census of Manufacture (CM), which is conducted annually by the Ministry of Economy, Trade and Industry.[5] The dataset ranges from 2001 to 2019. Note that the CM has started surveying the export status of establishments since 2001.[6]

The CM includes two forms of questionnaires: Form A (*Kou*), which includes establishments with 30 or more employees, and Form B (*Otsu*), which includes establishments with 29 or fewer employees. The data on capital stock are only available for Form A. Thus, this study uses the datasets of Form A to estimate establishment-level markups and TFP.

Regarding the markup and TFP estimation, the value-added is used as a dependent variable. In this case, the value-added is calculated as the total production minus the total materials, fuel, and energy consumed, as well as the subcontracting expenses for production outsourcing. Labor is defined as the total hours worked in a year. Using the average hours worked in a year in the manufacturing sector, which are taken from the Monthly Labour Survey (Ministry of Health, Labour and Welfare), the total annual hours worked are calculated by multiplying the annual number of

[5] In 2012 and 2016, the CM was integrated into the Economic Census for Business Activity (ECBA). The 2012 and 2016 ECBA surveyed annual economic activity in the previous year (i.e., 2011 and 2015, respectively), and the survey was jointly conducted by the Ministry of Economy, Trade and Industry and the Ministry of Internal Affairs and Communications.

[6] The Online Appendix provides additional estimation results in 1986–2000 although the export dummy is not included. The export status is surveyed from 2001.

workers by the hours worked.[7] Capital stocks are measured as end-of-year book values using the perpetual inventory method. Energy consumption is used as a proxy of material demand for productivity shocks, which are unobserved by the econometricians but observable to establishment i. Total wage payments are also surveyed and directly observed in the data. All nominal values of outputs, intermediate inputs, energy consumption, capital stocks, and wage payments are deflated by each price index. Finally, the deflators of output price (2011=100), input price (2011=100), and investment price (2010=100) are constructed by price indices available from the Bank of Japan, and all monthly price indices are averaged yearly.[8] Wage payments are deflated by the output price index.[9]

This study constructs a municipal panel dataset from 1985 to 2016 to consider municipal merges during this study period. The reference date for geographical information is October 1, 2018, at which point the total number of municipalities is 1,741. Tokyo's 23 wards are counted individually. The cities designated by a government ordinance (*Seirei Shitei Toshi*) are counted as cities (*shi*). The corresponding cities are Sapporo-shi, Sendai-shi, Saitama-shi, Chiba-shi, Yokohama-shi, Kawasaki-shi, Sagamihara-shi, Niigata-shi, Shizuoka-shi, Hamamatsu-shi, Nagoya-shi, Kyoto-shi, Osaka-shi, Sakai-shi, Kobe-shi, Okayama-shi, Hiroshima-shi, Kitakyushu-shi, and Fukuoka-shi, Kumamoto-Shi. Municipal data are aggregated based on the municipal unit at the reference date.[10]

The market size variable is calculated as the market potential based on municipal panel data. First, yearly data on income at the municipality level is taken from the Survey of Municipal Taxation (*Shi-Cho-Son Kazei Jokyoto no Shirabe*) (Ministry of Internal Affairs and Communications). Income data at the municipal level during 1975–2014 is available from the website of the Cabinet Office of Japan.[11] Data during 2010–2016 are available from the website of the Ministry of Internal Affairs and

[7] Since 2001, the CM has categorized workers into regular and non-regular workers. This study merges the average hours worked for each type of worker at the industry level with data from the Monthly Labour Survey (Ministry of Health, Labour and Welfare) to calculate the hours worked for regular and non-regular workers during 2001–2019. The logarithm of the sum of the hours worked for regular and non-regular workers is used as an explanatory variable.

[8] The deflators of output and input prices at the two-digit industry level are constructed using the Input-Output Price Index of the Manufacturing Industry by Sector (2011 base) from the Bank of Japan (URL: `https://www.boj.or.jp/en/statistics/pi/iopi_2011/index.htm/`). The deflator of investment price is constructed using the "Index by Stage of Demand and Use" contained in the "Corporate Goods Price Index (2010 base)" (URL: `https://www.boj.or.jp/en/statistics/pi/cgpi_2010/index.htm/`; Data code: `PR01'PRCG10\_18K1020005`).

[9] In the markup and TFP estimation, this study trims the outliers of the variables as follows. The lowermost and uppermost 0.5 percent of value-added, capital stock, energy consumption, and total wage payments are trimmed as outliers. The uppermost 1 percent of labor is trimmed as outliers. In addition, the labor share $\hat{\alpha}^L_{it}$ is trimmed to between 0 and 1. This study excludes establishments not observed four times or more during the years of the empirical analysis.

[10] See e-Stat, the portal site for Japan's official statistics, for changes in statistical area codes (URL: `http://www.e-stat.go.jp/SG1/hyoujun/initialize.do`).

[11] `https://www5.cao.go.jp/keizai-shimon/kaigi/special/future/keizai-jinkou_data.html` (as of January 19, 2017)

Communications.[12]

The inter-municipal distance is required to calculate the market potential. The market potential of 1,741 municipalities is calculated using the `spgen` command developed by Kondo (2017) on Stata, wherein the great circle distance is calculated based on the latitude and longitude of the city hall location of each municipality.

One might consider whether the location information of products each establishment offers for sale is available. However, this information is not available in the present study. Instead, this study proposes a market potential approach that is fundamentally based on the gravity model of trade and that can be used in a general situation. The sales area of establishment i in municipality a is assumed to extend to within the d km range, with transportation costs increasing with the distance. Although the standard market potential does not limit the geographical range of markets (i.e., $d = \infty$), as shown in Figure 1, this study introduces market potential from 100 km to 1,000 km ($d = 100, 200, \ldots, 1000, \infty$). This approach clarifies the marginal market size effects with respect to the geographical range of the market by checking how the parameter estimate changes with distance.

[Figure 1]

Historical population data at the city (*Shi*) and district (*Gun*) levels for IV are obtained the 1930 Population Census (Ministry of Internal Affairs and Communications). Note that the geographical unit in 1930 is based on the district (*Gun*) for small municipalities, such as town (*Cho* or *Machi*) and village (*Son* or *Mura*). In addition, following de la Roca and Puga (2017), this study also uses the mean altitude of areas, which is calculated based on the 500 meters by 500 meters grid square data on "Altitude and Inclination" in Ministry of Land, Infrastructure, Transport and Tourism (2020). Some missing data on mean altitude are complemented by the data offered by Zaiki et al. (2005). The validity of the IVs is examined using a weak instrument test and an overidentification test.

Table 1 presents the descriptive statistics of the variables used in the empirical analysis. Note that the lowermost and uppermost 1 percent of the estimated markup and TFP distributions are trimmed as outliers. During 2001–2019, the mean of the markups was 1.69, and the median was 1.56.[13]

[12] `https://www.soumu.go.jp/main_sosiki/jichi_zeisei/czaisei/czaisei_seido/ichiran09.html` (as of September 7, 2019)

[13] The markup estimation of this study basically follows the Stata program code developed by De Loecker and Warzynski (2012). This study slightly modifies their code to include observations of the initial year. In their Stata program code, these are dropped since the lagged variables are used in the TFP estimation procedure. Although observations of the initial year are not used in the parameter estimation, TFP and markups in the initial year are computed using the estimated parameters. Appendix A offers detailed explanations for markup and TFP estimation.

[Table 1]

Figure 2 compares the markup distributions for all sectors between large and small markets. The solid and dashed lines represent the markup distributions of markets with above- and below-median market potentials, respectively. Clearly, the markup distribution in large markets shifts toward left and is less dispersed than that in small markets, as shown in Figure 2, which means that the average markup is lower in large markets. Importantly, a majority of establishments charge higher markups than 1 (the markup must be 1 at the long-term equilibrium under the conditions of a perfect competition economy and constant returns to scale). Figure 3 presents the markup distributions for each sector. Large heterogeneity exists across sectors and is investigated using regression analysis.

[Figures 2–3]

4 Estimation Results

4.1 Higher Markups for More Productive and Large Establishments

Table 2 presents the estimation results of regression (1). The baseline regression estimation results are provided in Column (1) and, as a robustness check to control for unobserved establishment factors, the fixed-effect regression estimation results are in Column (2). The area-year dummies η_{at} estimated in Column (1) are used in the next step. Note that the aim of this regression is to extract area factor from the total markup.

Table 2 shows that the input efficiency component in TFP increases the markup, suggesting that efficient technology related to inputs enables firms to charge high markups because firms can efficiently produce output. This suggestion is consistent with Melitz (2003) and Bellone et al. (2016). This study finds that the other productivity components decrease markups. A number of possible explanations exist for this result. For example, productivity factors resulting from the skilled workers will simultaneously increase the share of labor through higher wages, which reduces the markups from Equation (A.3) in Appendix B. One suggestion is that, in this body of literature, distinguishing efficient input-related technology from the other productivity factors is important.

This study finds that large firms with multiple establishments tend to charge higher markups, as noted in Column (1). However, the dummy for multiple establishment is not significant in fixed effect estimation, as noted in Column (2). The estimation results on employment size are consistent with

Atkin et al. (2015), who investigate cross-firm variations in markups and costs in Pakistan's soccer ball industry. They find that larger firms charge higher markups. In addition, Kugler and Verhoogen (2012) use data from the Colombian manufacturing industry and show that large plants charge higher output prices, which also implies that they charge higher markups. Kugler and Verhoogen (2012) discuss that higher output prices reflect higher product quality. The authors also note that larger plants also use inputs with higher prices.

An interesting finding is that the export dummy estimate is significantly negative, which is different from that of De Loecker and Warzynski (2012) and consistent with Bellone et al. (2016). This study confirms a markup premium for exporters in the regression without control variables; However, the coefficient of the export dummy becomes negative when control variables are added. A further discussion on markup premiums for exporters is provided in the Online Appendix.

This study finds that large, highly productive establishments can charge higher markups, which agrees with recent empirical studies, such as De Loecker and Warzynski (2012) and Bellone et al. (2016). However, using a firm-level dataset of Japanese manufacturers in 1994–2012, Dobbelaere and Kiyota (2018) show that productivity and firm size are not relevant as markup determinants. This difference may have arisen from the difference of data source. Dobbelaere and Kiyota (2018) use firm-level data, whereas this study uses establishment-level data. Another possibility may be related with the productivity measurement, as previously discussed. In the present study, these gaps remain unresolved. The next step is to investigate how market size explains variations in area factors $\hat{\eta}_{at}$ of the markups.

[Table 2]

4.2 Larger Markets Lead to Lower Markups

Figure 4 shows a visual relation between area-mean markups and market size as measured by market potential. The area-mean markups are the estimated η_{at} in Column (1) of Table 2, which is centered on the mean. It is evident that the larger markets show lower markups. This relationship is further examined by regression analysis.

Table 3 presents the OLS estimation results based on regression (2) with different geographical market ranges. The market-size elasticity on markup is significant at the 1 % level and −0.054 for market potential within 100 km. As the threshold distance increases, the magnitude increases, suggesting that manufacturing establishments face competition in geographically broader markets.

However, the increment gradually decreases with distance. The market-size elasticity is −0.075 for market potential within 200 km, −0.088 for market potential within 300 km, −0.092 for market potential within 400 km, −0.096 for market potential within 500 km, −0.100 for market potential within 600 km, and −0.101 for market potential within 700 km, suggesting that markets in closer proximity to the production location have, on average, the large magnitudes. However, this does not mean that distant markets are of no consequence in the manufacturing sector. The market-size elasticity with respect to market potential within the country, as noted in Column (11), is −0.125, which is the highest value in Table 3.

[Figure 4; Table 3]

4.3 Robustness Check for Endogeneity Bias

Table 4 presents the IV estimation results based on regression (2) with different geographical market range. Overall, the population potential in 1930 shows a large F-value. The overidentification test does not strongly support the exogeneity condition (i.e., the null hypothesis should not be rejected). However, it is modestly supported. Although the qualitative results do not change between the OLS and IV estimation results, the magnitudes of the IV estimates are larger than those of the OLS estimates, suggesting the possibility of endogeneity bias.

The market-size elasticity on markup is significant at the 1 % level and −0.058 for the market potential within 100 km. As previously discussed, the magnitude of the market-size elasticity increases with s the threshold distance. The market-size elasticity on markup is −0.122 for the market potential within the entire country.

This study uses the IV estimates in Table 4 to quantify the extent to which stronger competition in larger markets, on average, generates markup differences across markets, other factors being constant. The estimated percentage change in markups between one market and a market that is twice as large is −4.54 % ($\approx 2^{-0.067} - 1$) for market potential within 100 km. If the markup charged by an establishment in one market is 1.5, then another similar establishment in the market that is twice as large charges 1.43. For two markets with a difference in size of ten times, the percentage change is −14.30 % ($\approx 10^{-0.067} - 1$). That is, if the markup charged by an establishment in one market is 1.5, another similar establishment in the market that is ten times larger charges 1.29.

As for market potential within 500 km, the estimated percentage change in markups between one market and a market that is twice as large is −9.31 % ($\approx 2^{-0.122} - 1$). If the markup charged by an establishment in one market is 1.5, another similar establishment in the market that is twice as large charges 1.38. For two markets with a difference in size of ten times, the percentage change is −27.72 % ($\approx 10^{-0.122} - 1$). That is, if the markup charged by an establishment in one market is 1.5, another similar establishment in the market that is ten times larger charges 1.13.

As for market potential within the entire country, the estimated percentage change in markups between one market and a market that is twice as large is −8.09 % ($\approx 2^{-0.141} - 1$). If the markup charged by an establishment in one market is 1.5, another similar establishment in the market that is twice as large charges 1.36. For two markets with a difference in size of ten times, the percentage change is −24.43 % ($\approx 10^{-0.122} - 1$). That is, if the markup charged by an establishment in one market is 1.5, another similar establishment in the market that is ten times larger charges 1.08.

The quantification exercises of this study shows that market competition factors have non-negligible impacts on markups, suggesting that manufacturing establishments in larger markets face stronger competition. This, in turn, results in lower markups. Importantly, this study provides evidence that both the size of market in closer proximity to the production location and the size of the distant market affect manufacturing establishments' price-setting behavior.

[Table 4]

4.4 Markup and Market Size by Sector

Tables 5 and 6 presents OLS and IV estimation results of regression 2 by two-digit level sector. The sector list is provided in Appendix C. This study finds sectoral heterogeneity in the market-size elasticity of markups. Importantly, all two-digit level industries show significantly negative market-size elasticities of markups. Sector 2 (textile mill products, leather tanning, leather products, and fur skins) shows the largest magnitude for the negative elasticity. In contrast, sector 8 (ceramic, stone and clay products) shows the smallest magnitude.

Based on IV estimation results in Table 6, this study discusses the industrial heterogeneity in the market size effect on markup. For example, sector 4 (pulp, paper and paper products) shows a large magnitude on markups for market potential within 100 km, but not as large a magnitude for the market potential within the entire country, suggesting that the geographical range of market competition is relatively localized in this industry. Compared with other sectors, sector 1 (food, beverages, tobacco, feed) shows large differences in the market size elasticity between d = 1000 km and $d = \infty$ km (−0.082 and −0.119, respectively), suggesting that geographical range of markets within the country is expansive.

In summary, the present study finds that larger markets lead to stronger competition, and this increase in competition results in lower markups. This finding is consistent with theoretical and empirical studies, such Melitz and Ottaviano (2008), Behrens et al. (2014), Bellone et al. (2016), Hottman (2016), and Lu et al. (2014).[14] The core contribution of this study is to exemplify how geographical range of markets affects the markups. Specifically, this evidences that it is important to consider the geographical boundary of markets in empirical studies because the manufacturing goods are tradable and the production location is not identical to the demand location. This study provides evidence that manufacturing establishments face stronger competition in geographically wider markets.

[Tables 5–6]

5 Conclusion

This study has investigated the effect of market size on markups in the Japanese manufacturing sector. Monopolistic competition models with endogenous markup, such as Melitz and Ottaviano (2008), Behrens et al. (2014), and Bellone et al. (2016), have shown that productive firms charge higher markups, whereas stronger competition in larger markets decreases markups.[15] A key concern of this study is to uncover the geographical range of markets in the tradable goods sector. Knowing the geographical boundary of a market for manufactured goods is not easy because the production location is not identical to the demand location. This study has proposed a market potential approach with a distance boundary to determine how the geographical extension of markets affects the markups.

Focusing on the markup distributions, this study finds that manufacturing establishments face stronger competition in geographically wider markets—not only neighboring markets within 100 km but also distant markets more than 100 km away from the production location. The estimation

[14] Anderson et al. (2018) investigate spatial and temporal variations in markups using gross margins in the U.S. retail sector and find that gross margins are positively correlated with the log of income in terms of cross-sectional variations. Importantly, the authors identify that the assortment of goods sold across regions is different, especially between rich and poor regions. Handbury and Weinstein (2015) also note that considering differences in variety across cities is an important factor for explaining regional price variations.

[15] Appendix B provides additional analysis for the theoretical predictions focusing on the relation between markup and wage.

results show that markups in markets that are twice as large are, on average, 3.95 % lower for market potential within 100 km, 7.07 % lower for market potential within 500 km, and 8.09 % lower for market potential within the entire country. Interestingly, this increment decreases with distance, suggesting that the marginal effects of markets in closer proximity are, on average, larger.

This study provides important implications for the empirical literature on firm selection. Combes et al. (2012) conclude that the selection can barely explain the spatial difference in productivity. However, competition in tradable sectors is not limited to the city of the place of production and, as discussed by Accetturo et al. (2018), the market potential is an alternative measure of market size. This study complements their findings by focusing on markup distributions and find that larger markets lead to stronger competition. However, this does not mean that less productive firms cannot survive in larger markets (i.e., left-truncation of the productivity distribution), even if they face stronger competition. There are possible explanations for this in a real economic context. Less productive firms might be small and medium-sized enterprises that the government tends to support through financial policies. It may also be the case that less productive firms leave larger markets, and new entrants are also less productive. As studied by Asplund and Nocke (2006) and Nocke (2006), this continuous turnover makes it difficult to conduct an empirical analysis on firm selection. The limitation of this study is that the entry and exit is not directly considered. Further studies are needed to fill the gap with regard to firm selection between the theoretical analyses and empirical findings.

References

Accetturo, Antonio, Valter Di Giacinto, Giacinto Micucci, and Marcello Pagnini (2018) “Geography, productivity, and trade: Does selection explain why some locations are more productive than others?” *Journal of Regional Science* 58(5), pp. 949–979.

Ackerberg, Daniel, Kevin Caves, and Garth Frazer (2015) “Identification properties of recent production function estimators,” *Econometrica* 83(6), pp. 2411–2451.

Anderson, Eric, Sergio Rebelo, and Arlene Wong (2018) “Markups Across Space and Time.” NBER Working Paper No. 24434.

Asplund, Marcus and Volker Nocke (2006) “Firm turnover in imperfectly competitive market,” *Review of Economic Studies* 73(2), pp. 295–327.

Atkin, David, Azam Chaudhry, Shamyla Chaudhry, Amit K. Khandelwal, and Eric Verhoogen (2015) "Markup and cost dispersion across firms: Direct evidence from producer surveys in Pakistan," *American Economic Review* 105(5), pp. 537–544.

Behrens, Kristian, Giordano Mion, Yasusada Murata, and Jens Südekum (2014) "Trade, wages, and productivity," *International Economic Review* 55(4), pp. 1305–1348.

Behrens, Kristian, Giordano Mion, Yasusada Murata, and Jens Suedekum (2017) "Spatial frictions," *Journal of Urban Economics* 97, pp. 40–70.

Bellone, Flora, Patrick Musso, Lionel Nesta, and Frederic Warzynski (2016) "International trade and firm-level markups when location and quality matter," *Journal of Economic Geography* 16(1), pp. 67–91.

Ciccone, Antonio and Robert E. Hall (1996) "Productivity and the density of economic activity," *American Economic Review* 86(1), pp. 54–70.

Combes, Pierre-Philippe and Laurent Gobillon (2015) "The empirics of agglomeration economies," in Duranton, Gilles, J. Vernon Henderson, and William C. Strange eds. *Handbook of Regional and Urban Economics* Vol. 5, Amsterdam: Elsevier, Chap. 5, pp. 247–348.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux (2010) "Estimating agglomeration economies with history, geology, and worker effects," in Glaeser, Edward L. ed. *Agglomeration Economics*, Chicago: University of Chicago Press, Chap. 1, pp. 15–66.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux (2012) "The productivity advantages of large cities: Distinguishing agglomeration from firm selection," *Econometrica* 80(6), pp. 2543–2594.

de la Roca, Jorge and Diego Puga (2017) "Learning by working in big cities," *Review of Economic Studies* 84(1), pp. 106–142.

De Loecker, Jan and Paul T. Scott (2016) "Estimating market power Evidence from the US Brewing Industry." NBER Working Paper No. 22957.

De Loecker, Jan and Frederic Warzynski (2012) "Markups and firm-level export status," *American Economic Review* 102(6), pp. 2437-2471.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik (2016) "Prices, markups, and trade reform," *Econometrica* 84(2), pp. 445–510.

Dixit, Avinash K. and Joseph E. Stiglitz (1977) "Monopolistic Competition and Optimum Product Diversity," *American Economic Review* 67(3), pp. 297–308.

Dobbelaere, Sabien and Kozo Kiyota (2018) "Labor market imperfections, markups and productivity in multinationals and exporters," *Labour Economics* 53, pp. 198–212.

Donaldson, Dave and Richard Hornbeck (2016) "Railroads and American economic growth: A 'market access' approach," *Quarterly Journal of Economics* 131(2), pp. 799–858.

Georgiev, Boris (2018) "Essays on Firms in International Trade," Ph.D. dissertation, Department of Economics and Business Economics, Aarhus University. https://www.forskningsdatabasen.dk/en/catalog/2434870616 (As of October 15, 2018).

Hall, Robert E. (1986) "Market structure and macroeconomic fluctuations," *Brookings Papers on Economic Activity* 1986(2), pp. 285–322.

Hall, Robert E. (1988) "The relation between price and marginal cost in U.S. industry," *Journal of Political Economy* 96(5), pp. 921—47.

Handbury, Jessie and David E. Weinstein (2015) "Goods prices and availability in cities," *Review of Economic Studies* 82(1), pp. 258–296.

Hottman, Colin J. (2016) "Retail markups, misallocation, and store variety in the US." http://www.columbia.edu/~cjh2164/HottmanJMP.pdf (Accessed on December 12, 2017).

Kondo, Keisuke (2016) "Testing for agglomeration economies and firm selection in spatial productivity differences: The case of Japan." RIETI Discussion Paper No. 16-E-098.

Kondo, Keisuke (2017) "SPGEN: Stata module to generate spatially lagged variables." Boston College Statistical Software Components s458105. https://ideas.repec.org/c/boc/bocode/s458105.htm.

Kugler, Maurice and Eric Verhoogen (2012) "Prices, plant size, and product quality," *Review of Economic Studies* 79(1), pp. 307–339.

Levinsohn, James and Amil Petrin (2003) "Estimating production functions using inputs to control for unobservables," *Review of Economic Studies* 70(2), pp. 317–341.

Lu, Yi and Linhui Yu (2015) "Trade liberalization and markup dispersion: Evidence from China's WTO accession," *American Economic Journal: Applied Economics* 7(4), pp. 221–253.

Lu, Yi, Zhigang Tao, and Linhui Yu (2014) "The markup effect of agglomeration." Mimeo http://ylu6.weebly.com/uploads/8/6/4/2/8642496/markup_effect_of_agglomeration.pdf (Accessed December 26, 2017).

Lu, Yi, Yoichi Sugita, and Lianming Zhu (2017) "Wage markdowns and FDI liberalization." Mimeo. http://www.int.otaru-uc.ac.jp/swet/swet2017-1/9m6hz8/tf2qs9/sugita (Accessed February 6, 2018).

Melitz, Marc J. (2003) "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica* 71(6), pp. 1695–1725.

Melitz, Marc J. and Gianmarco I. P. Ottaviano (2008) "Market size, trade, and productivity," *Review of Economic Studies* 75(1), pp. 295–316.

Ministry of Land, Infrastructure, Transport and Tourism (2020) "National Land Numerical Information."
`http://nlftp.mlit.go.jp/ksj-e/index.html` (As of February 10, 2020).

Nocke, Volker (2006) "A gap for me: Entrepreneurs and entry," *Journal of the European Economic Association* 4, pp. 929–956.

Olley, G. Steven and Ariel Pakes (1996) "The dynamics of productivity in the telecommunications equipment industry," *Econometrica* 64(6), pp. 1263–1297.

Saito, Hisamitsu and Toshiyuki Matsuura (2016) "Agglomeration economies, productivity, and quality upgrading: Evidence from Japan." RIETI Discussion Paper No. 16-E-085.

Zaiki, Masumi, Takashi Oguchi, Yuichi Kagawa, Akiko Takahashi, Shiro Koike, and Masakazu Yamauchi (2005) "Relationships between distribution of Japanese residential areas and topography." CSIS Discussion Paper No. 68. (in Japanese).

Appendix A Theory and Practice for Markup Estimation

A.1 Theoretical Background for Markup Estimation

The theoretical background of markup estimation is based on the study by De Loecker and Warzynski (2012). In general, economic theory defines markup as the ratio between the price and marginal cost of production. Under conditions of perfect competition with the constant returns-to-scale technology, at profit maximization (or at cost minimization), markup takes the value of 1, which becomes the baseline value in the analysis. However, markup is greater than 1 under imperfect competition, such as monopoly, oligopoly, and monopolistic competition, in which firms have the market power.

Suppose that firms demand two factors (labor and capital) for production. Firm i minimizes costs of L_{it} and K_{it} to produce Q_{it} with production technology $Q_{it}(\cdot)$ as follows:

$$\begin{aligned} \min_{L_{it},K_{it}} \quad & w_{it}L_{it} + r_{it}K_{it}, \\ \text{s.t.} \quad & Q_{it} = Q_{it}(L_{it}, K_{it}), \end{aligned}$$

where w_{it} is the wage rate and r_{it} is the rental price of fixed capital. Then, the Lagrangian function for cost minimization problem is given as follows:

$$\mathcal{L}(L_{it}, K_{it}, \lambda_{it}) = w_{it}L_{it} + r_{it}K_{it} + \lambda_{it}\Big(Q_{it} - Q_{it}(L_{it}, K_{it})\Big),$$

where λ_{it} is the Lagrangian multiplier. Solving cost minimization with respect to labor and capital leads to the following first-order conditions (FOCs):

$$\lambda_{it} = \frac{w_{it}}{\partial Q_{it}(L_{it}, K_{it})/\partial L_{it}} \quad \text{and} \quad \lambda_{it} = \frac{r_{it}}{\partial Q_{it}(L_{it}, K_{it})/\partial K_{it}}, \tag{A.1}$$

where the Lagrange multiplier λ_{it} $(= \partial\mathcal{L}_{it}/\partial Q_{it})$ represents the marginal cost of production at a given level of output Q_{it}. Multiplying both numerator and denominator of the FOC with respect to labor by L_{it}/Q_{it} yields the following equation:

$$\lambda_{it} = \frac{1}{\theta^L_{it}} \frac{w_{it}L_{it}}{Q_{it}}, \tag{A.2}$$

where $\theta^L_{it} \equiv (\partial Q_{it}/Q_{it})/(\partial L_{it}/L_{it})$ represents the labor elasticity of the output.

Using the Lagrange multiplier λ_{it} in the general setting of the cost-minimization problem, the markup μ_{it} can be defined as $\mu_{it} \equiv P_{it}/\lambda_{it}$, where P_{it} is the output price.[16] The Lagrange multiplier measures marginal costs of production. Inserting λ_{it} in Equation (A.2) into the markup definition $\mu_{it} \equiv P_{it}/\lambda_{it}$ yields the explicit form of markup for each establishment i as follows:

$$\mu_{it} = \frac{\theta_{it}^{L}}{\alpha_{it}^{L}}, \tag{A.3}$$

where $\alpha_{it}^{L} \equiv w_{it}L_{it}/P_{it}Q_{it}$ represents the share of expenditures on labor L.

The aforementioned definition of markup based on cost minimization takes the general form without assuming a specific production technology. Note again that this specification of the markup takes the value of 1 under perfect competition with constant returns-to-scale technology. For example, profit maximization based on Cobb–Douglas production technology under perfect competition shows that labor share is identical to labor elasticity of output.

Importantly, markup can be easily estimated within the TFP estimation framework. The TFP estimation studies propose some robust estimation methods for the labor elasticity of output θ_{it}^{L}. An additional requirement for markup estimation is to obtain the labor share α_{it}^{L}; however, these shares can be directly observed from the data.[17]

A.2 Markup Estimation Procedure

De Loecker and Warzynski (2012) propose the markup estimation procedure by extending the TFP estimation approach developed by Ackerberg et al. (2015). Unlike the Cobb–Douglas production function, De Loecker and Warzynski (2012) consider the trans-log production function to capture firm heterogeneity in labor elasticity of output θ_{it}^{L} since the markup variation is derived only from the labor share α_{it}^{L} if the Cobb–Douglas production function is used.

The "value-added" trans-log production function is estimated as follows:[18]

$$y_{it} = \beta_0 + \beta_\ell \ell_{it} + \beta_k k_{it} + \beta_{\ell\ell}\ell_{it}^2 + \beta_{kk}k_{it}^2 + \beta_{\ell k}\ell_{it}k_{it} + \omega_{it} + \varepsilon_{it} \tag{A.4}$$

[16] There are other mathematical definitions of markup, such as a difference between price and marginal cost of production. The advantage of markup ratio is that the markup always takes a positive value, which allows to take logarithm. This is important since it can be empirically seen that the distribution of markup ratio nearly follows log-normal distribution.

[17] Theoretically, it is possible to use the ratio of capital elasticity of output θ_{it}^{K} and the capital shares α_{it}^{K} for markup estimation. However, in practice, there is no guarantee that both values of the markup calculated from labor and capital sides are identical.

[18] The constant term β_0 is explicitly added in the production function.

where y_{it} is the logarithm of value-added $P_{it}Q_{it}$, ℓ_{it} is the logarithm of the labor L_{it}, and k_{it} is the logarithm of the fixed capital K_{it}. The error term is assumed to consist of two components: ω_{it}, which is a productivity shock and is unobserved by econometricians but observable to the establishment i, and ε_{it}, which is a sequence of idiosyncratic shock that is not observable by the establishment i before the input decision-making.

De Loecker and Warzynski (2012) rely on material demand to solve for simultaneous problem on input decision-making and productivity, as follows:

$$m_{it} = m_t(k_{it}, \omega_{it}, z_{it}), \tag{A.5}$$

which depends on the capital, productivity shock, and other demand choice of inputs z_{it} (in vector form). To proxy for productivity shocks in the production function, the material demand function is inverted with respect to ω_{it} as follows:

$$\omega_{it} = h_t(k_{it}, m_{it}, z_{it}), \tag{A.6}$$

where an important assumption of the inverse is the monotonicity of intermediate inputs in productivity.

Using the control function of productivity, the trans-log production function in Equation (A.4) can be written as follows:

$$y_{it} = \phi_t(\ell_{it}, k_{it}, m_{it}, z_{it}) + \varepsilon_{it}, \tag{A.7}$$

where $\phi_t(\ell_{it}, k_{it}, m_{it}, z_{it})$ includes the trans-log form of inputs and the control function $h_t(k_{it}, m_{it}, z_{it})$ for productivity. In general, the form of control function $h_t(\cdot)$ takes the higher-order polynomial function.[19]

Labor elasticity of output θ_{it}^{L} is estimated in two steps. First, the trans-log production function in Equation (A.7) is estimated with the productivity control function $h_t(\cdot)$. This first-step estimation does not estimate the consistent estimators for labor and capital elasticities of the output but generates an expected output $\hat{\phi}_{it}$ from the following equation:

$$\phi_{it} = \beta_0 + \beta_\ell \ell_{it} + \beta_k k_{it} + \beta_{\ell\ell}\ell_{it}^2 + \beta_{kk}k_{it}^2 + \beta_{\ell k}\ell_{it}k_{it} + h_t(k_{it}, m_{it}, z_{it}).$$

[19]This study considers the third-order polynomial function with respect to labor, capital, and material including their cross-terms.

The second-step regression estimates labor and capital elasticities of the output by assuming the following law of motion for productivity:

$$\omega_{it} = g_t(\omega_{it-1}) + \xi_{it},$$

where AR(1) process in productivity is assumed as $g_t(\omega_{it-1}) = \rho_0 + \rho_1\omega_{it-1}$ including the constant term ρ_0. Using the expected output $\hat{\phi}_{it}$ and any value of $\boldsymbol{\beta} \equiv (\beta_0\ \beta_\ell\ \beta_k\ \beta_{\ell\ell}\ \beta_{kk}\ \beta_{\ell k})'$ (in vector form), productivity ω_{it} can be obtained as follows:

$$\omega_{it}(\boldsymbol{\beta}) = \hat{\phi}_{it} - \beta_0 - \beta_\ell \ell_{it} - \beta_k k_{it} - \beta_{\ell\ell}\ell_{it}^2 - \beta_{kk}k_{it}^2 - \beta_{\ell k}\ell_{it}k_{it}.$$

The moment condition for estimation of the parameter vector $\boldsymbol{\beta}$ can be obtained as follows:

$$\mathrm{E}\left[\xi_{it}(\boldsymbol{\beta})\begin{pmatrix} 1 \\ \ell_{it-1} \\ k_{it-1} \\ \ell_{it-1}^2 \\ k_{it-1}^2 \\ \ell_{it-1}k_{it-1} \end{pmatrix}\right] = \mathbf{0} \tag{A.8}$$

where $\mathbf{0}$ is a zero vector.[20] To consider heterogeneity in production technology across sector, the trans-log production function is estimated by a two-digit-level sector.

In the trans-log production function (A.4), labor elasticity of output for establishment i is calculated using the estimated coefficients $\hat{\boldsymbol{\beta}}$ as follows:

$$\hat{\theta}_{it}^L = \hat{\beta}_\ell + 2\hat{\beta}_{\ell\ell}\ell_{it} + \hat{\beta}_{\ell k}k_{it}.$$

Note that, unlike the Cobb–Douglas production function, the trans-log production function generates different values of labor elasticity of output across establishments.[21]

[20]The estimation procedure of the parameter vector $\boldsymbol{\beta}$ is as follows. First, the productivity $\omega_{it}(\boldsymbol{\beta})$ is computed by any initial values of $\boldsymbol{\beta}$. Second, the parameters ρ_0 and ρ_1 in the law of motion for productivity are estimated as the AR(1) regression model. Third, $\xi_{it}(\boldsymbol{\beta})$ is estimated as a residual in the second regression. Forth, the parameter vector $\boldsymbol{\beta}$ is estimated from the moment condition (A.8). Fifth, the productivity $\omega_{it}(\boldsymbol{\beta})$ is computed by the parameter vector $\boldsymbol{\beta}$ estimated in the forth step. The second to fifth steps in the procedure are iterated until the convergence is achieved for the parameter vector $\boldsymbol{\beta}$.

[21]Similarly, capital elasticity of output for establishment i can be calculated as $\hat{\theta}_{it}^K = \hat{\beta}_k + 2\hat{\beta}_{kk}k_{it} + \hat{\beta}_{\ell k}\ell_{it}$.

This study computes the logarithm of TFP as follows:

$$\log(\widehat{\text{TFP}}_{it}) = y_{it} - \hat{\beta}_\ell \ell_{it} - \hat{\beta}_k k_{it} - \hat{\beta}_{\ell\ell} \ell_{it}^2 - \hat{\beta}_{kk} k_{it}^2 - \hat{\beta}_{\ell k} \ell_{it} k_{it}.$$

Furthermore, this study divides the overall $\log(\widehat{\text{TFP}}_{it})$ into $\omega_{it}(\hat{\boldsymbol{\beta}}) + \hat{\beta}_0$ and $\hat{\varepsilon}_{it}$. Bellone et al. (2016) use $\omega_{it}(\hat{\boldsymbol{\beta}}) + \hat{\beta}_0$ (i.e., residuals from the expected values of the value-added $\hat{\phi}_{it}$) to calculate TFP. Olley and Pakes (1996) use the definition of this study (i.e., the residuals from the raw value-added y_{it}). Importantly, the former component of TFP, $\omega_{it}(\hat{\boldsymbol{\beta}}) + \hat{\beta}_0$, captures cost-saving productivity related to inputs (input efficiency), whereas the latter component, $\hat{\varepsilon}_{it}$, captures another type of productivity (e.g., positive externalities within establishments) and idiosyncratic shocks. Olley and Pakes (1996) discuss these two differences in their study. Following De Loecker et al. (2016), in order to distinguish cost-saving productivity factors from any other type of productivity factors, this study separately introduces $\log(\widehat{\text{TFP}}^{\omega}_{it})$ and $\log(\widehat{\text{TFP}}^{\varepsilon}_{it})$, which denote cost-saving productivity related to inputs and with any other type of productivity, respectively.[22]

The final step is to calculate the markup μ_{it}. Instead of using the raw data of labor shares α^L_{it}, De Loecker and Warzynski (2012) propose a method for error correction, and this study computes labor shares as follows:

$$\hat{\alpha}^L_{it} = \frac{w_{it} L_{it}}{\widehat{P_{it} Q_{it}}},$$

where $\widehat{P_{it} Q_{it}}$ is calculated from the expected values, $\exp(\hat{y}_{it})$, of the regression of log value-added y_{it} on the third-order polynomial function with respect to labor ℓ_{it}, capital k_{it}, and material m_{it}, including their interaction terms since there is a lack of information on price P_{it} and quantity Q_{it}, separately. If the labor share $\hat{\alpha}^L_{it}$ takes a value of less than 0 or more than 1, these observations are excluded from the analysis. Consequently, the markup of establishment i is computed as $\hat{\mu}_{it} = \hat{\theta}^L_{it} / \hat{\alpha}^L_{it}$.[23]

[22] Note that the control function approach for productivity ω_{it}, which is derived from the material demand, rules out unobserved productivity factors, if $\omega_{it}(\hat{\boldsymbol{\beta}}) + \hat{\beta}_0$ and $\hat{\varepsilon}_{it}$ are used as TFP. In other words, the term $\omega_{it}(\hat{\boldsymbol{\beta}}) + \hat{\beta}_0$ captures only the productivity related to factors in the productivity control function (A.6).

[23] One limitation of the markup estimation approach based on the production function is that markups are estimated at the firm level. If firms produce multi-product, this approach cannot capture markup heterogeneity across products within firms. As mentioned above, another approach for markup estimation is to estimate the price elasticity of demand using product-level data. Using the product-level data in the Japanese manufacturing sector, Saito and Matsuura (2016) estimate the price elasticity of demand and complement results of this study. In addition, Behrens et al. (2014) emphasize the difference between markups on consumers' and firms' sides when differentiated goods are traded across regions. The microeconometric approach by De Loecker and Warzynski (2012) estimates markups on firms' side, whereas markups on consumers' side are essential for welfare analysis.

A.3 Markup and TFP Estimation Results

Figure A.1 presents the estimation results of labor and capital elasticities of output based on the trans-log production function (A.4). In contrast to the Cobb–Douglas production function, the trans-log production function makes a variation in these elasticities. The markup estimation approach proposed by De Loecker and Warzynski (2012) is based on the labor elasticity of the output, which is shown as the circle marker in Figure A.1. The average value in each sector tends to lie between 0.8 and 0.9. The solid line in Figure A.1 represents the 5–95 percentile interval of the estimated labor elasticities of the output.

Figure A.1 presents the estimation results of labor shares, which are directly calculated from the data, as the ratio of total wage payment and the value-added. The average values of labor shares vary and tend to lie between 0.4 and 0.6. The solid line in Figure A.1 represents the 5–95 percentile interval of the labor shares, which has a longer interval than that of the labor elasticity of output in Figure A.1.

The markup estimation is based on the ratio of labor elasticity of output and labor share for each establishment. The markup distribution for all sectors is shown in Figure 4, and that for each sector is shown in Figure 3. Table A.1 presents descriptive statistics of markups in the two-digit level sector.

Figure A.2 presents TFP distributions estimated by the approach of De Loecker and Warzynski (2012) and compares them between large and small markets. Consistent with the earlier empirical studies in the urban economics literature, such as Combes et al. (2012), this study also shows that establishments in larger markets are more productive. This study confirms almost the same TFP distributions as those obtained by Kondo (2016) using establishment-level panel data of the CM, although the TFP estimation approach is different.

[Figures A.1–A.2; Table A.1]

Appendix B Relation between Markup and Wage

As a robustness check of the theoretical prediction, this study also confirms the relation between markup and wage at the establishment level. Instead of the markup estimation in Equation (A.3), merging the definition of markup and FOC for cost minimization in Equation (A.1), the markup can be related with the wage rate as follows:

$$\mu_{it} = \underbrace{\frac{1}{w_{it}}}_{\substack{\text{Inverse of}\\ \text{Marginal costs}\\ \text{of labor}}} \times \underbrace{P_{it}\frac{\partial Q_{it}(L_{it}, K_{it})}{\partial L_{it}}}_{\substack{\text{Marginal revenue}\\ \text{of labor}}},$$

where the first term on the right-hand side indicates the inverse of marginal costs of labor and the second term indicates the marginal revenue of labor. Importantly, markup has a negative relation with wage rate, which suggests that firms' markups decrease when the wage increases in the labor market while other factors are equal (i.e., no employment adjustment by the increase in wage).

Lu et al. (2017) consider firms' wage-setting powers in the labor market, which also decreases the wage. Although the wage-setting powers are not considered, only price-setting powers also lead to wage markdowns, which are defined as the ratio of wage to the marginal revenue of labor. Since the definition of the wage markdown includes the inverse of price-cost markups, it is important to control for the markups in the product market. To assess this further, Dobbelaere and Kiyota (2018) explicitly distinguish both price- and wage-setting powers in their theoretical and empirical analyses.

Figure B.1 shows the relation between average monthly wage per worker and markup at the establishment level. The wage is computed as the annual total wage payment divided by the number of workers for 12 months. Clearly, there is a negative relation between markup and wage.

[Figure B.1]

Appendix C Sector List

Table C.1 presents a list of two-digit level manufacturing sectors used in this study. Sector numbers used in Table 6, Table A.1, Figure 3, and Figure A.1 correspond to those in Table C.1.

[Table C.1]

Table 1: Descriptive Statistics, 2001–2019

Variables	Obs.	Mean	S.D.	p25	p50	p75
Establishment-Level						
Markup	711,049	1.961	0.736	1.478	1.823	2.279
Log(TFP), Overall	711,049	7.978	0.646	7.626	7.973	8.341
Log(TFP), Input Efficiency	711,049	7.995	0.172	7.885	7.979	8.087
Log(TFP), Other Efficiency	711,049	−0.017	0.629	−0.351	−0.018	0.333
D(1=Multi-Establishments)	711,049	0.431	0.495	0.000	0.000	1.000
Log(Employment)	711,049	4.361	0.794	3.761	4.159	4.762
Log(Capital)	711,049	8.912	2.155	7.438	8.468	9.393
D(1=Export)	711,049	0.125	0.331	0.000	0.000	0.000
Area-Level (Municipality-Level)						
Area-Year Mean Log(Markup)	28,954	0.043	0.187	−0.052	0.038	0.136
Area-Year Mean Log(Markup)	28,954	0.041	0.159	−0.046	0.032	0.121
Log(Market Potential)	28,954	13.619	0.692	13.153	13.582	14.091
Log(Market Potential within 100km)	28,954	12.429	1.387	11.466	12.281	13.553
Log(Market Potential within 200km)	28,954	12.923	1.213	12.128	13.015	13.837
Log(Market Potential within 300km)	28,954	13.137	1.132	12.481	13.278	13.959
Log(Market Potential within 400km)	28,954	13.251	1.086	12.692	13.385	14.026
Log(Market Potential within 500km)	28,954	13.333	1.044	12.809	13.477	14.055
Log(Market Potential within 600km)	28,954	13.404	0.990	12.884	13.532	14.067
Log(Market Potential within 700km)	28,954	13.450	0.948	12.971	13.551	14.076
Log(Market Potential within 800km)	28,954	13.486	0.906	13.044	13.560	14.081
Log(Market Potential within 900km)	28,954	13.530	0.852	13.107	13.569	14.085
Log(Market Potential within 1000km)	28,954	13.566	0.793	13.130	13.575	14.088
Log(Population Potential in 1930)	28,954	13.171	0.811	12.606	13.027	13.637
Log(Altitude)	28,954	4.160	1.317	3.301	4.142	5.100
Log(Steep)	28,954	1.217	1.132	0.687	1.488	2.043

Note: The lowermost and uppermost 1 percent of the markup and TFP distributions are trimmed as outliers.

Table 2: Markup and Establishment Factors, 2001–2019

	Dependent Variable: Log(Markup)	
Explanatory Variables	(1)	(2)
Log(TFP), Input Efficiency	0.603***	0.414***
	(0.007)	(0.008)
Log(TFP), Other Efficiency	−0.143***	−0.095***
	(0.001)	(0.001)
D(1=Multi-Establishment)	0.010***	0.001
	(0.002)	(0.001)
Log(Employment)	0.139***	0.191***
	(0.002)	(0.003)
Log(Capital)	0.004***	0.010***
	(0.001)	(0.001)
D(1=Export)	−0.019***	−0.006***
	(0.003)	(0.002)
Industry Dummy	Yes	Yes
Area-Year Dummy (η_{at})	Yes	No
Year Dummy	No	Yes
Establishment Fixed Effects	No	Yes
Number of Observations	711,049	711,049
Number of Establishments	77,580	77,580
AdjustedWithin R^2	0.310	0.104

Note: Heteroskedasticity-consistent standard errors clustered by establishments are in parentheses. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3: OLS Estimation Results for Markup and Market Size, 2001–2019

	Dependent Variable: Area-Year Mean Log(Markup), $\hat{\eta}_{at}$										
	$d = 100$ km	$d = 200$ km	$d = 300$ km	$d = 400$ km	$d = 500$ km	$d = 600$ km	$d = 700$ km	$d = 800$ km	$d = 900$ km	$d = 1000$ km	$d = \infty$ km
Explanatory Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Log(Market Potential)	−0.054***	−0.075***	−0.088***	−0.092***	−0.096***	−0.100***	−0.101***	−0.101***	−0.104***	−0.112***	−0.125***
	(0.003)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Prefecture Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	28,954	28,954	28,954	28,954	28,954	28,954	28,954	28,954	28,954	28,954	28,954
Number of Municipalities	1,610	1,610	1,610	1,610	1,610	1,610	1,610	1,610	1,610	1,610	1,610
Adjusted R^2	0.499	0.503	0.507	0.503	0.502	0.502	0.499	0.497	0.499	0.502	0.510

Note: Heteroskedasticity-consistent standard errors clustered by municipalities are in parentheses. The unit of observation is municipality. The market potential is measured by regional income including those of neighboring municipalities located within the circle of d km radius. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 4: IV Estimation Results for Markup and Market Size, 2001–2019

	Dependent Variable: Area-Year Mean Log(Markup), $\hat{\eta}_{at}$										
	$d = 100$ km	$d = 200$ km	$d = 300$ km	$d = 400$ km	$d = 500$ km	$d = 600$ km	$d = 700$ km	$d = 800$ km	$d = 900$ km	$d = 1000$ km	$d = \infty$ km
Explanatory Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Second Stage											
Log(Market Potential)	−0.067***	−0.097***	−0.107***	−0.115***	−0.122***	−0.126***	−0.128***	−0.131***	−0.132***	−0.135***	−0.141***
	(0.004)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Prefecture Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First Stage	Endogenous Variable: Log(Market Potential)										
Log(Pop. Potential in 1930)	0.986***	0.692***	0.625***	0.582***	0.548***	0.531***	0.521***	0.507***	0.505***	0.491***	0.467***
	(0.031)	(0.023)	(0.020)	(0.019)	(0.018)	(0.017)	(0.017)	(0.017)	(0.016)	(0.014)	(0.013)
Log(Altitude)	0.039***	0.042***	0.032***	0.019**	0.014*	0.012*	0.009	0.005	0.007	0.009	0.007
	(0.014)	(0.011)	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.006)
Log(Altitude) Squared	−0.045***	−0.033***	−0.025***	−0.016**	−0.012	−0.011	−0.009	−0.006	−0.008	−0.012*	−0.015**
	(0.015)	(0.011)	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)
Number of Observations	28,954	28,954	28,954	28,954	28,954	28,954	28,954	28,954	28,954	28,954	28,954
Number of Municipalities	1,610	1,610	1,610	1,610	1,610	1,610	1,610	1,610	1,610	1,610	1,610
First-Stage F-value	447.255	439.683	456.127	441.511	443.832	453.023	472.607	461.829	480.984	535.218	614.149
Overidentification (p-value)	0.429	0.166	0.167	0.182	0.178	0.176	0.171	0.157	0.172	0.245	0.362

Note: Heteroskedasticity-consistent standard errors clustered by municipalities are in parentheses. The unit of observation is municipality. The market potential is measured by regional income including those of neighboring municipalities located within the circle of d km radius. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 5: OLS Estimation Results for Markup and Market Potential by Industry, 2001–2019

	Dependent Variable: Area-Year Mean Log(Markup), $\hat{\eta}_{at}$, Controlled for Establishment Factors				
	Coef. of Log(Market Potential)				Obs.
	d = 100 km	d = 500 km	d = 1000 km	d = ∞ km	
Sectors	(1)	(2)	(3)	(4)	(5)
Sector 1	−0.045***	−0.069***	−0.082***	−0.119***	6,865
	(0.004)	(0.008)	(0.010)	(0.008)	[1,198]
Sector 2	−0.068***	−0.127***	−0.166***	−0.172***	2,132
	(0.006)	(0.011)	(0.013)	(0.013)	[647]
Sector 3	−0.056***	−0.090***	−0.103***	−0.118***	1,051
	(0.007)	(0.014)	(0.016)	(0.015)	[376]
Sector 4	−0.079***	−0.135***	−0.150***	−0.152***	882
	(0.010)	(0.013)	(0.015)	(0.015)	[374]
Sector 5	−0.060***	−0.101***	−0.121***	−0.127***	932
	(0.008)	(0.014)	(0.014)	(0.014)	[378]
Sector 6	−0.057***	−0.101***	−0.110***	−0.113***	1,065
	(0.006)	(0.010)	(0.011)	(0.011)	[436]
Sector 7	−0.045***	−0.099***	−0.110***	−0.112***	2,322
	(0.006)	(0.009)	(0.009)	(0.009)	[778]
Sector 8	−0.046***	−0.080***	−0.090***	−0.107***	1,263
	(0.006)	(0.013)	(0.016)	(0.013)	[513]
Sector 9	−0.084***	−0.143***	−0.158***	−0.159***	515
	(0.010)	(0.020)	(0.021)	(0.021)	[257]
Sector 10	−0.052***	−0.093***	−0.095***	−0.096***	386
	(0.009)	(0.019)	(0.021)	(0.021)	[199]
Sector 11	−0.053***	−0.091***	−0.104***	−0.108***	2,222
	(0.005)	(0.009)	(0.009)	(0.009)	[812]
Sector 12	−0.057***	−0.115***	−0.128***	−0.131***	2,863
	(0.005)	(0.008)	(0.009)	(0.009)	[926]
Sector 13	−0.067***	−0.134***	−0.140***	−0.141***	489
	(0.008)	(0.014)	(0.015)	(0.015)	[263]
Sector 14	−0.056***	−0.111***	−0.129***	−0.131***	3,532
	(0.006)	(0.009)	(0.010)	(0.010)	[986]
Sector 15	−0.050***	−0.097***	−0.106***	−0.110***	1,819
	(0.005)	(0.009)	(0.010)	(0.010)	[612]
Sector 16	−0.054***	−0.114***	−0.132***	−0.136***	616
	(0.009)	(0.016)	(0.017)	(0.017)	[260]

Note: Heteroskedasticity-consistent standard errors clustered by municipalities are in parentheses. The unit of observation is municipality. Numbers in brackets are the number of municipalities. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 6: IV Estimation Results for Markup and Market Potential by Industry, 2001–2019

	Dependent Variable: Area-Year Mean Log(Markup), $\hat{\eta}_{at}$, Controlled for Establishment Factors				
	Coef. of Log(Market Potential)				Obs.
	d = 100 km	d = 500 km	d = 1000 km	d = ∞ km	
Sectors	(1)	(2)	(3)	(4)	(5)
Sector 1	−0.058***	−0.099***	−0.115***	−0.134***	6,865
	(0.005)	(0.010)	(0.012)	(0.011)	[1,198]
Sector 2	−0.081***	−0.164***	−0.195***	−0.198***	2,132
	(0.008)	(0.015)	(0.017)	(0.017)	[647]
Sector 3	−0.073***	−0.150***	−0.168***	−0.172***	1,051
	(0.007)	(0.018)	(0.020)	(0.020)	[376]
Sector 4	−0.099***	−0.161***	−0.175***	−0.176***	882
	(0.011)	(0.017)	(0.018)	(0.018)	[374]
Sector 5	−0.082***	−0.137***	−0.150***	−0.152***	932
	(0.010)	(0.016)	(0.017)	(0.017)	[378]
Sector 6	−0.070***	−0.120***	−0.130***	−0.132***	1,065
	(0.007)	(0.011)	(0.012)	(0.012)	[436]
Sector 7	−0.056***	−0.109***	−0.118***	−0.119***	2,322
	(0.006)	(0.010)	(0.011)	(0.011)	[778]
Sector 8	−0.061***	−0.112***	−0.126***	−0.134***	1,263
	(0.008)	(0.015)	(0.017)	(0.017)	[513]
Sector 9	−0.108***	−0.176***	−0.190***	−0.190***	515
	(0.015)	(0.023)	(0.024)	(0.024)	[257]
Sector 10	−0.062***	−0.114***	−0.122***	−0.122***	386
	(0.012)	(0.023)	(0.025)	(0.025)	[199]
Sector 11	−0.063***	−0.112***	−0.122***	−0.124***	2,222
	(0.006)	(0.011)	(0.012)	(0.012)	[812]
Sector 12	−0.069***	−0.132***	−0.141***	−0.142***	2,863
	(0.006)	(0.012)	(0.012)	(0.012)	[926]
Sector 13	−0.084***	−0.153***	−0.160***	−0.160***	489
	(0.009)	(0.016)	(0.017)	(0.017)	[263]
Sector 14	−0.069***	−0.134***	−0.147***	−0.148***	3,532
	(0.008)	(0.013)	(0.014)	(0.014)	[986]
Sector 15	−0.057***	−0.110***	−0.117***	−0.119***	1,819
	(0.006)	(0.011)	(0.012)	(0.012)	[612]
Sector 16	−0.072***	−0.125***	−0.135***	−0.136***	616
	(0.011)	(0.017)	(0.018)	(0.018)	[260]

Note: Heteroskedasticity-consistent standard errors clustered by municipalities are in parentheses. The unit of observation is municipality. Numbers in brackets are the number of municipalities. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.1: Descriptive Statistics of Establishment-Level Markups by Two-Digit Level Sector, 2001–2019

Sector	Obs.	Mean	S.D.	p25	Median	p75
Sector 1	131,350	2.307	0.818	1.736	2.178	2.721
Sector 2	39,476	2.000	0.760	1.473	1.861	2.362
Sector 3	19,160	1.936	0.772	1.405	1.782	2.252
Sector 4	25,305	2.510	0.996	1.802	2.274	2.970
Sector 5	32,800	1.777	0.705	1.281	1.620	2.084
Sector 6	30,930	1.881	0.879	1.254	1.652	2.271
Sector 7	62,611	2.081	0.765	1.548	1.947	2.449
Sector 8	25,836	2.043	0.953	1.383	1.797	2.404
Sector 9	18,232	2.172	0.893	1.523	1.987	2.604
Sector 10	10,509	1.824	0.953	1.149	1.553	2.198
Sector 11	65,804	1.855	0.672	1.392	1.716	2.151
Sector 12	85,268	1.633	0.581	1.219	1.509	1.901
Sector 13	12,871	1.634	0.584	1.217	1.525	1.904
Sector 14	81,281	1.944	0.833	1.350	1.764	2.322
Sector 15	56,336	1.604	0.611	1.193	1.480	1.852
Sector 16	13,280	1.707	0.673	1.239	1.557	1.992

Note: The unit of observation is establishment. Sector number is listed in Table C.1.

Table C.1: List of Two-digit Level Manufacturing Sectors

Num.	Sectors
1.	Food, beverages, tobacco, feed
2.	Textile mill products, leather tanning, leather products, and fur skins
3.	Lumber, wood products, furniture, and fixtures
4.	Pulp, paper and paper products
5.	Printing and allied industries
6.	Chemical and allied products
7.	Plastic products and rubber products
8.	Ceramic, stone and clay products
9.	Iron and steel
10.	Non-ferrous metals and products
11.	Fabricated metal products
12.	General-purpose machinery
13.	Business oriented machinery
14.	Electrical machinery, equipment and supplies, electronic parts, devices and electronic circuits; Information and communication electronics equipment
15.	Transportation equipment
16.	Miscellaneous manufacturing industries

Note: Table 6, Table A.1, Figure 3, and Figure A.1 use the sector numbers presented in this table.

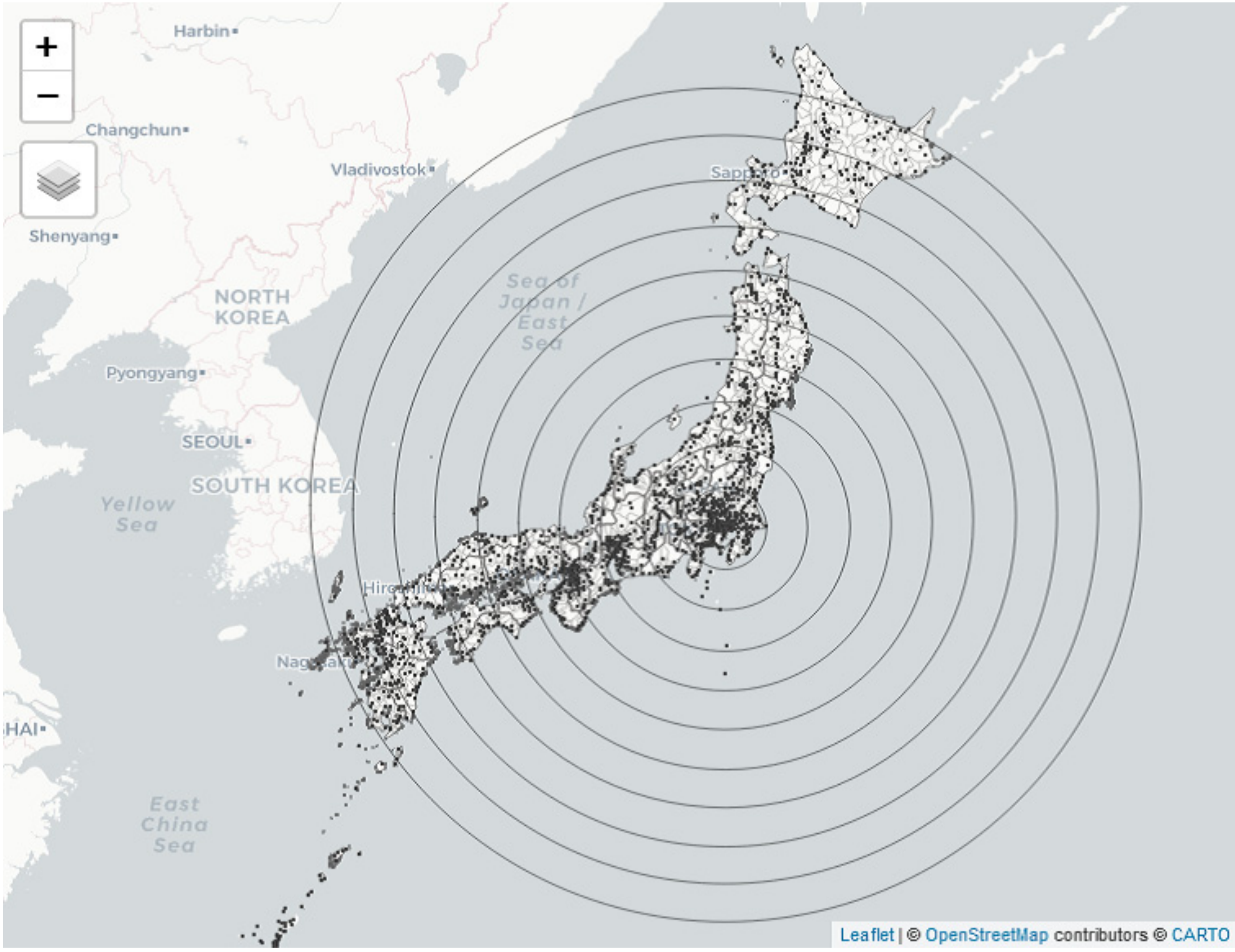


Figure 1: Market Potential within from 100 km to 1,000 km (Case of Chiyoda-ku,Tokyo)

Note: Created by author. Each circle is depicted in 100 km as the center of Chiyoda-ku, Tokyo. Market potential is calculated within the circle of $d \in (100, 200, \ldots, 1000)$ km radius from the location of city hall of each municipality, which is depicted as a black marker.

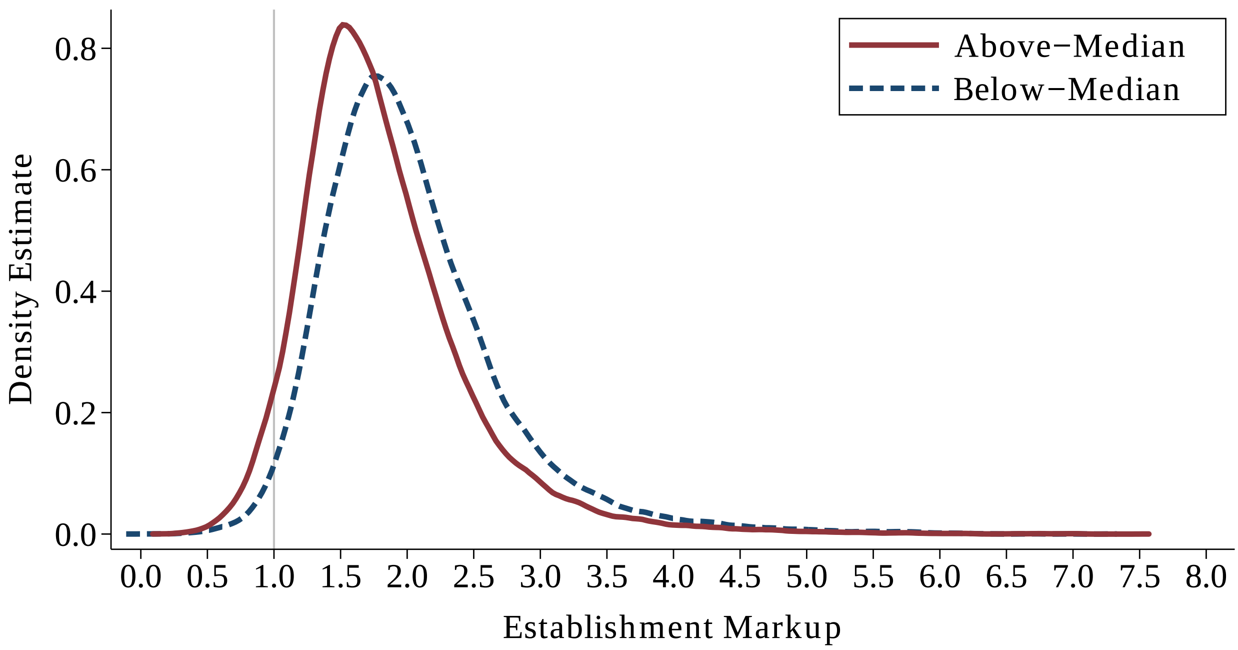
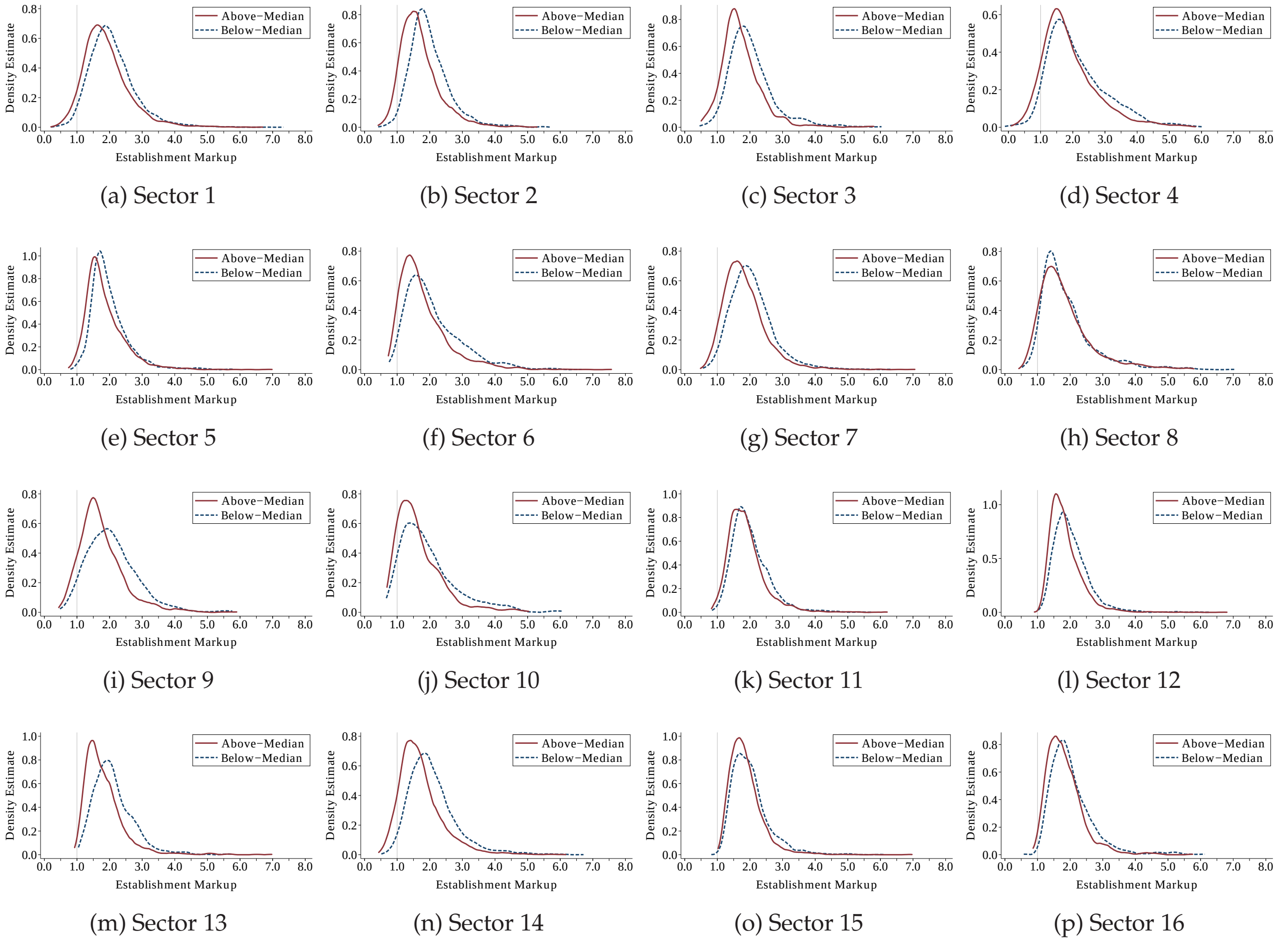


Figure 2: Markup Distribution between Large and Small Markets, 2001–2019

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Average markup of establishment i in the corresponding period is calculated as $\mu_i = 1/T_i \sum_{t=1}^{T_i} \mu_{it}$. Average market size in city a where establishment i is located is calculated as $\mathrm{MP}_{a(i)} = 1/T_i \sum_{t=1}^{T_i} \mathrm{MP}_{a(i)t}$. The solid and dashed lines denote the markup distributions in markets with above-median and below-median market potential, respectively.



(a) Sector 1 (b) Sector 2 (c) Sector 3 (d) Sector 4

(e) Sector 5 (f) Sector 6 (g) Sector 7 (h) Sector 8

(i) Sector 9 (j) Sector 10 (k) Sector 11 (l) Sector 12

(m) Sector 13 (n) Sector 14 (o) Sector 15 (p) Sector 16

Figure 3: Markup Distributions between Large and Small Markets by Sector, 2001–2019

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Establishment-mean markup is averaged across years as $\hat{\mu}_i = 1/T_i \sum_t \hat{\mu}_{it}$, where T_i is the number of years for establishment i observed in the sample. Sector numbers correspond to those used in Table C. The solid and dashed lines denote the markup distributions in markets with above-median and below-median market potential, respectively.

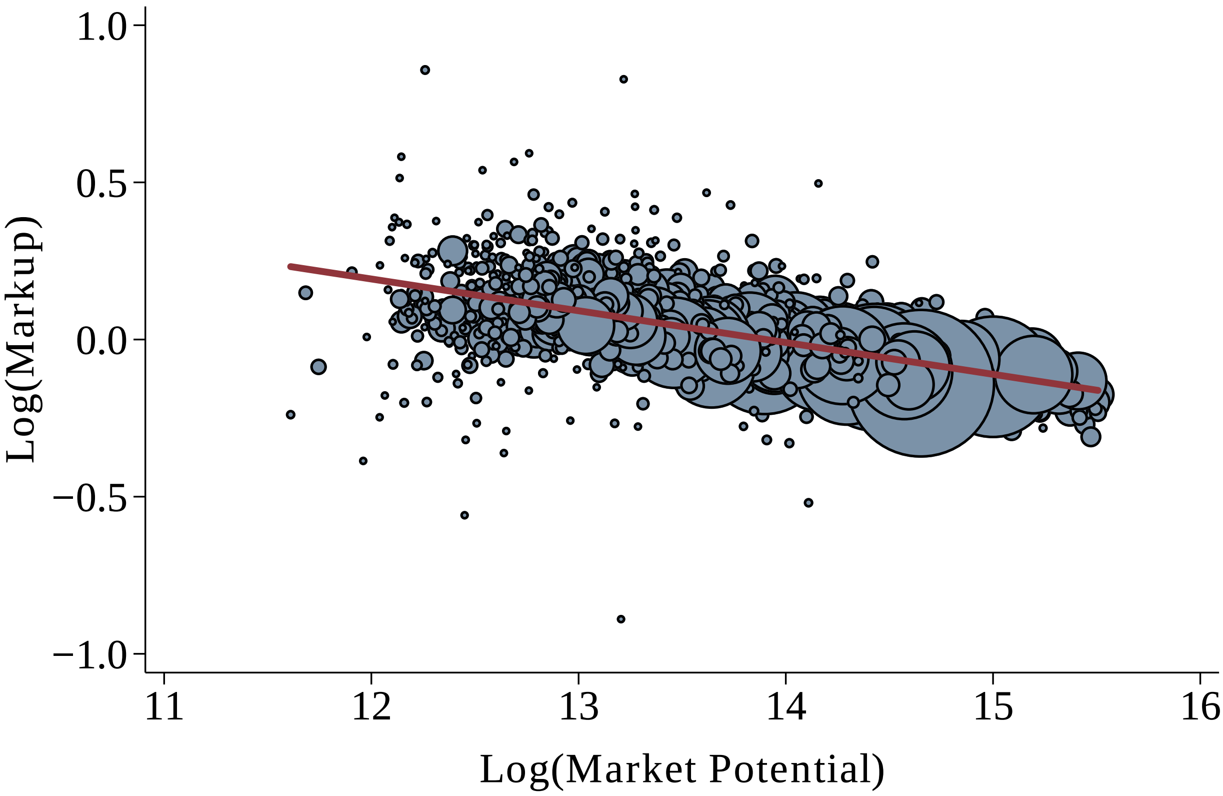
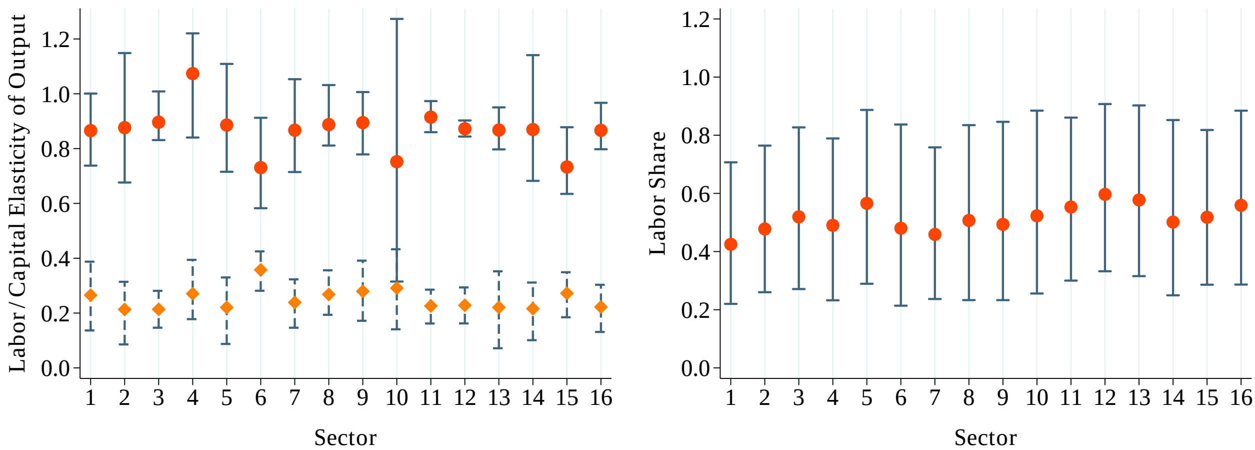


Figure 4: Markup and Market Size

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Area-mean logarithm of markup is averaged across years as $\hat{\eta}_a = 1/T_a \sum_t \hat{\eta}_{at}$, where T_a is the number of years for area a observed in the sample. Market size is averaged across years as $\log(\mathrm{MP}_a) = 1/T_a \sum_t \log(\mathrm{MP}_{at})$. The circle size represents the the number of manufacturing establishments in each geographical unit.



(a) Labor and Capital Elasticities of Output

(b) Labor Shares

Figure A.1: Markup Estimation

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). In Panel (a), the circle and diamond markers represent the average labor and capital elasticities of output, respectively. The solid and dashed lines represent the 5–95 percentile interval of the estimated output elasticities with respect to labor and capital, respectively. In Panel (b), the circle markers represent the average labor shares. The solid lines represent the 5–95 percentile interval of the estimated labor shares.

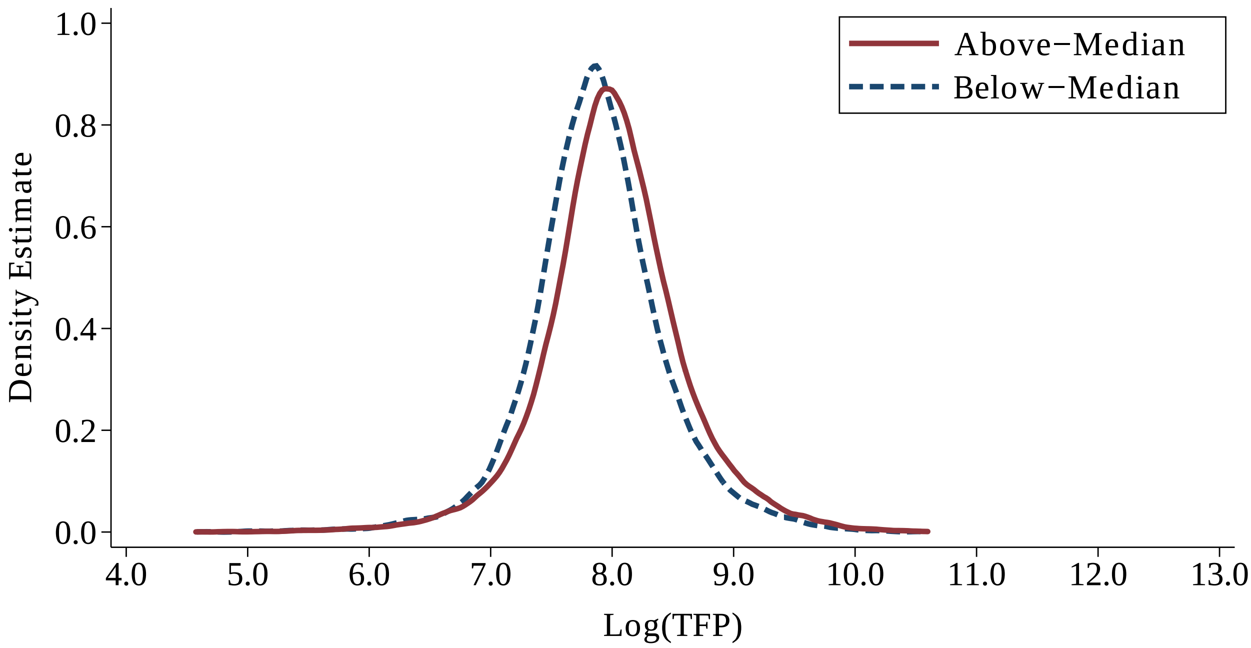


Figure A.2: TFP Distributions between Large and Small Markets

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Average TFP of establishment i in the corresponding period is calculated as $\log(\widehat{\mathrm{TFP}}_i) = 1/T_i \sum_{t=1}^{T_i} \log(\widehat{\mathrm{TFP}}_{it})$. Average market size in city a where establishment i is located is calculated as $\mathrm{MP}_{a(i)} = 1/T_i \sum_{t=1}^{T_i} \mathrm{MP}_{a(i)t}$. The solid and dashed lines denote the markup distributions in markets with above-median and below-median market potential, respectively.

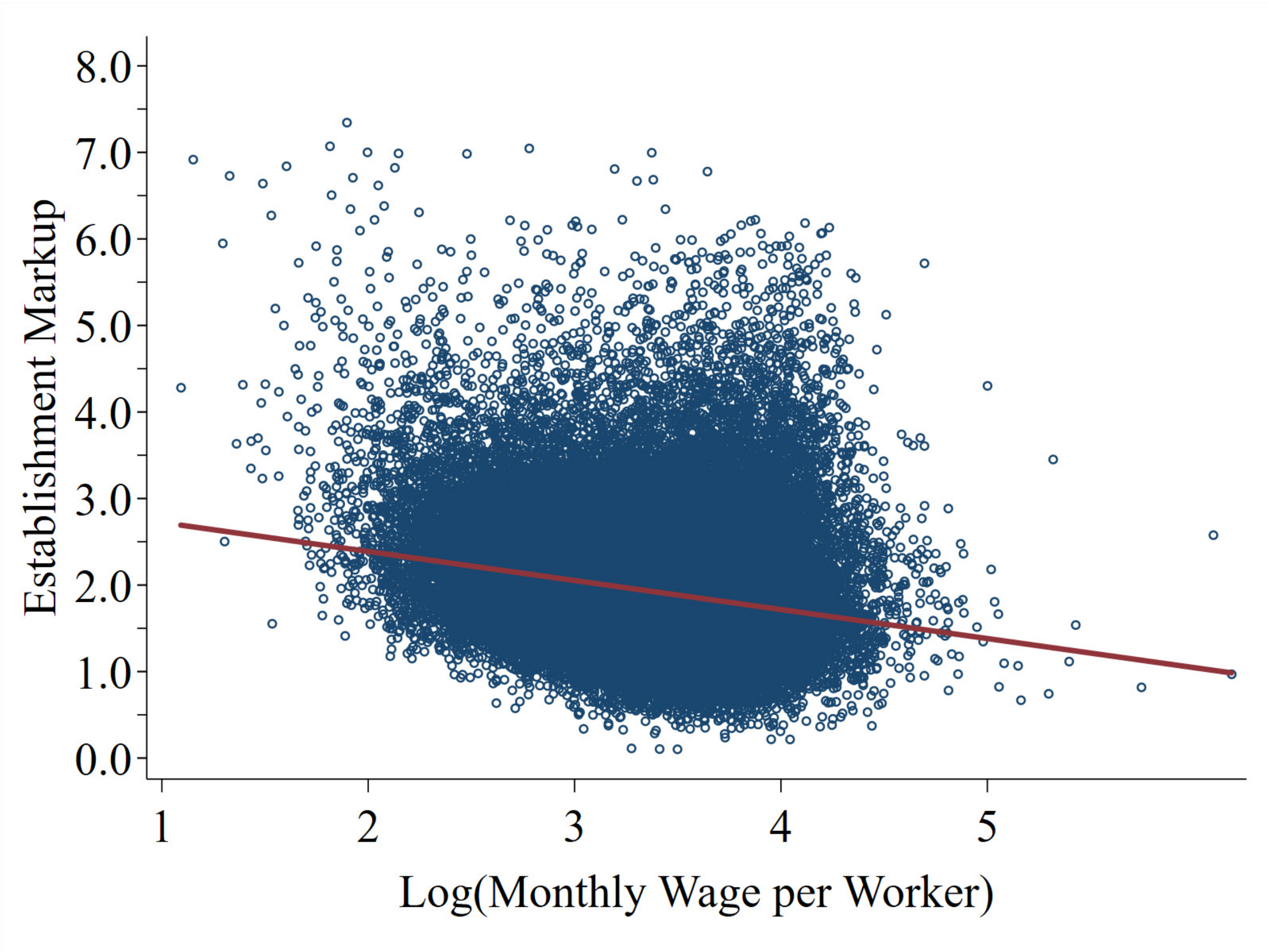


Figure B.1: Markup and Wage

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Average markup of establishment i in the corresponding period is calculated as $\mu_i = 1/T_i \sum_{t=1}^{T_i} \mu_{it}$. Logarithm of average monthly wage per worker of establishment i in the corresponding period is calculated as $\log(w_i) = 1/T_i \sum_{t=1}^{T_i} \log(w_{it})$.

Online Appendix for

Markups and Market Size: Evidence from Japan

Keisuke Kondo*

This online appendix provides additional estimation results.

Contents

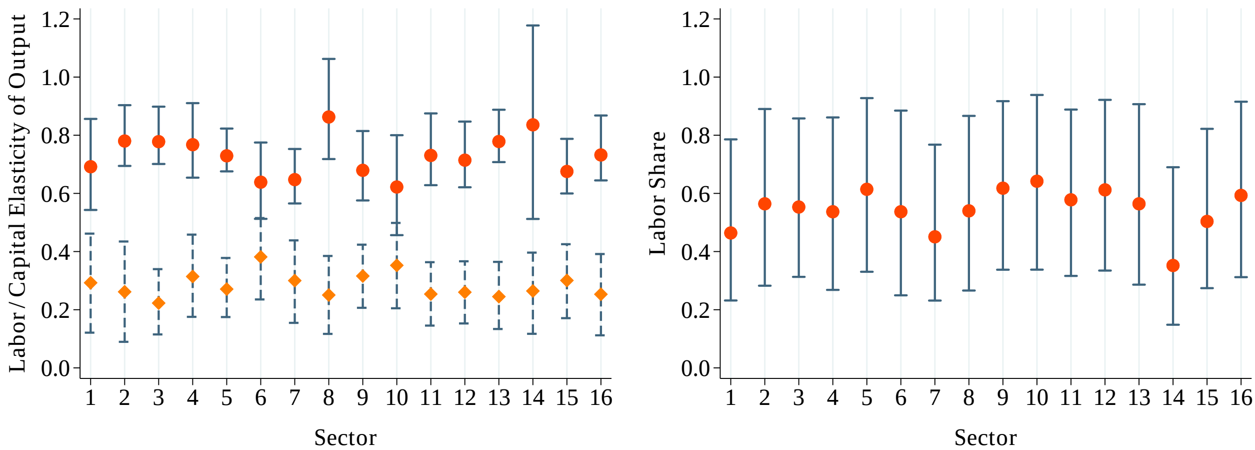
*Research Institute of Economy, Trade and Industry (RIETI). 1-3-1 Kasumigaseki, Chiyoda-ku, Tokyo, 100–8901, Japan. (e-mail: kondo-keisuke@rieti.go.jp).

Online Appendix A. Markup and TFP Estimation in 1986–2000

Figure OA.A. 1 presents estimation results for markups in terms of labor elasticity of output and labor share in the 1986–2000 period. Panel (a) of Figure OA.A. 1 shows labor and capital elasticities of output based on the trans-log production function. Panel (b) of Figure OA.A. 1 shows the estimation results of labor shares, which are directly calculated from the data, as the ratio of total wage payment and the value-added.

Figure OA.A. 2 presents TFP distributions estimated by the approach of De Loecker and Warzynski (2012) and compares them between large and small markets, which corresponds Figure A.1 in the main text.

[Figures OA.A. 1–OA.A. 2]



(a) Labor and Capital Elasticities of Output

(b) Labor Share

Figure OA.A. 1: Markup Estimation, 1986–2000

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). In Panel (a), the circle and diamond markers represent the average labor and capital elasticities of output, respectively. The solid and dashed lines represent the 5–95 percentile interval of the estimated output elasticities with respect to labor and capital, respectively. In Panel (b), the circle markers represent the average labor shares. The solid lines represent the 5–95 percentile interval of the estimated labor shares.

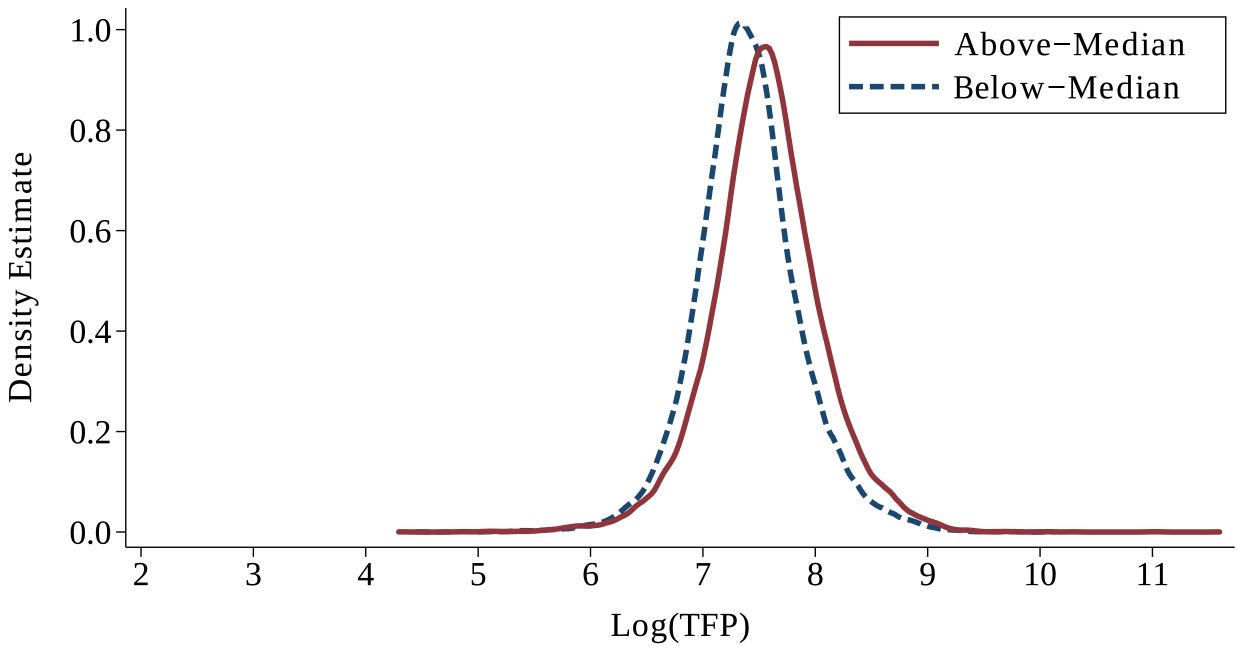


Figure OA.A. 2: TFP Distributions between Large and Small Markets, 1986–2000

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Average TFP of establishment i in the corresponding period is calculated as $\log(\widehat{\mathrm{TFP}}_i) = 1/T_i \sum_{t=1}^{T_i} \log(\widehat{\mathrm{TFP}}_{it})$. Average market size in city a where establishment i is located is calculated as $\mathrm{MP}_{a(i)} = 1/T_i \sum_{t=1}^{T_i} \mathrm{MP}_{a(i)t}$. The solid and dashed lines denote the markup distributions in cities with above-median and below-median population, respectively.

Online Appendix B. Estimation Results for Markup and Market Size in 1986–2000

Tables OA.B. 1–OA.B. 5 presents estimation results for markup and market size in the 1986–2000 period, which correspond to Tables 1–5 in the main text. Note that the Census of Manufacture (Ministry of Economy, Trade and Industry) did not survey export status in the 1986–2000 period.

[Tables OA.B. 1–OA.B. 5]

Table OA.B. 1: Descriptive Statistics, 1986–2000

Variables	Obs.	Mean	S.D.	p5	p25	p50	p75	p95
Establishment-Level								
Markup	599,885	1.602	0.680	0.849	1.134	1.437	1.876	2.975
Log(TFP), Overall	599,885	7.506	0.599	6.562	7.181	7.510	7.852	8.467
Log(TFP), Input Efficiency	599,885	7.549	0.204	7.191	7.433	7.566	7.679	7.857
Log(TFP), Other Efficiency	599,885	−0.043	0.585	−0.967	−0.351	−0.035	0.285	0.889
D(1=Multi-Establishments)	599,885	0.428	0.495	0.000	0.000	0.000	1.000	1.000
Log(Employment)	599,885	4.250	0.650	3.466	3.738	4.094	4.615	5.572
Log(Capital)	599,885	8.400	1.948	6.215	6.908	8.006	9.105	12.715
D(1=Export)	-	-	-	-	-	-	-	-
Area-Level (Municipality-Level)								
Area-Year Mean Log(Markup)	23,419	0.046	0.166	−0.209	−0.062	0.042	0.147	0.315
Log(Market Potential)	23,419	13.513	0.700	12.373	13.026	13.496	13.990	14.708
Log(Population Potential in 1930)	23,419	13.151	0.814	12.047	12.604	13.015	13.628	14.698
Log(Altitude)	23,419	4.181	1.312	2.172	3.323	4.156	5.131	6.298

Note: The lowermost and uppermost 1 percent of the markup and TFP distributions are trimmed as outliers. Export status is not surveyed in the 1986–2000 period.

Table OA.B. 2: Markup and Establishment Factors by Fixed-Effect Estimation

	Dependent Variable: Log(Markup)	
Explanatory Variables	(1)	(2)
Log(TFP), Input Efficiency	0.515***	0.418***
	(0.005)	(0.006)
Log(TFP), Other Efficiency	−0.146***	−0.094***
	(0.002)	(0.001)
D(1=Multi-Establishment)	0.030***	0.007***
	(0.002)	(0.002)
Log(Employment)	0.142***	0.129***
	(0.002)	(0.003)
Log(Capital)	0.006***	0.007***
	(0.001)	(0.001)
D(1=Export)	-	-
	-	-
Industry Dummy	Yes	Yes
Area-Year Dummy (η_{at})	Yes	No
Year Dummy	No	Yes
Establishment Fixed Effects	No	Yes
Number of Observations	599,885	599,885
Number of Establishments	61,022	61,022
Adjusted R^2	0.513	0.346

Note: Heteroskedasticity-consistent standard errors clustered by establishments are in parentheses. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level. Export status is not surveyed in the 1986–2000 period.

Table OA.B. 3: OLS Estimation Results for Markup and Market Size, 1986–2000

	Dependent Variable: Area-Year Mean Log(Markup), $\hat{\eta}_{at}$										
	$d = 100$ km	$d = 200$ km	$d = 300$ km	$d = 400$ km	$d = 500$ km	$d = 600$ km	$d = 700$ km	$d = 800$ km	$d = 900$ km	$d =$ 1000 km	$d = \infty$ km
Explanatory Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Log(Market Potential)	−0.064***	−0.091***	−0.108***	−0.116***	−0.122***	−0.128***	−0.129***	−0.129***	−0.129***	−0.143***	−0.165***
	(0.005)	(0.006)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
Prefecture and Year Dummies	Yes		Yes								
Number of Observations	23,419	23,419	23,419	23,419	23,419	23,419	23,419	23,419	23,419	23,419	23,419
Number of Municipalities	1,611	1,611	1,611	1,611	1,611	1,611	1,611	1,611	1,611	1,611	1,611
Adjusted R^2	0.725	0.727	0.730	0.729	0.729	0.730	0.728	0.727	0.727	0.730	0.737

Note: Heteroskedasticity-consistent standard errors clustered by municipalities are in parentheses. The unit of observation is municipality. The market potential is measured by regional income including those of neighboring municipalities located within the circle of d km radius. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table OA.B. 4: IV Estimation Results for Markup and Market Size, 1986–2000

	Dependent Variable: Area-Year Mean Log(Markup), $\hat{\eta}_{at}$										
	d = 100 km	d = 200 km	d = 300 km	d = 400 km	d = 500 km	d = 600 km	d = 700 km	d = 800 km	d = 900 km	d = 1000 km	d = ∞ km
Explanatory Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Second Stage											
Log(Market Potential)	−0.083***	−0.122***	−0.136***	−0.149***	−0.158***	−0.163***	−0.166***	−0.171***	−0.171***	−0.175***	−0.184***
	(0.006)	(0.009)	(0.010)	(0.011)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)
Prefecture and Year Dummies	Yes		Yes								
First Stage	Endogenous Variable: Log(Market Potential)										
Log(Pop. Potential in 1930)	0.899***	0.615***	0.550***	0.503***	0.470***	0.456***	0.447***	0.431***	0.432***	0.423***	0.400***
	(0.032)	(0.023)	(0.020)	(0.019)	(0.018)	(0.017)	(0.017)	(0.017)	(0.016)	(0.015)	(0.013)
Log(Altitude)	0.035*	−0.001	−0.002	0.000	0.001	−0.001	−0.002	−0.003	−0.003	−0.006	−0.007
	(0.020)	(0.012)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Log(Altitude) Squared	−0.004	0.003*	0.003*	0.001	0.000	0.001	0.000	0.000	0.000	0.001	0.001
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number of Observations	23,419	23,419	23,419	23,419	23,419	23,419	23,419	23,419	23,419	23,419	23,419
Number of Municipalities	1,611	1,611	1,611	1,611	1,611	1,611	1,611	1,611	1,611	1,611	1,611
First-Stage F-value	327.004	276.309	285.364	253.223	261.113	266.463	272.659	257.416	270.411	302.144	347.426
Overidentification (p-value)	0.066	0.002	0.001	0.007	0.020	0.019	0.025	0.042	0.038	0.023	0.036

Note: Heteroskedasticity-consistent standard errors clustered by municipalities are in parentheses. The unit of observation is municipality. The city size is measured by population including those of neighboring municipalities located within the circle of 30 km radius. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table OA.B. 5: IV Estimation Results for Markup and Market Potential by Industry, 1986–2000

	Dependent Variable: Area-Year Mean Log(Markup), $\hat{\eta}_{at}$, Controlled for Establishment Factors				
	Coef. of Log(Market Potential)				Obs.
	d = 100 km	d = 500 km	d = 1000 km	$d = \infty$ km	
Sectors	(1)	(2)	(3)	(4)	(5)
Sector 1	−0.065***	−0.116***	−0.133***	−0.221***	4,156
	(0.008)	(0.015)	(0.018)	(0.032)	[1,007]
Sector 2	−0.085***	−0.197***	−0.224***	−0.257***	4,197
	(0.008)	(0.017)	(0.018)	(0.034)	[985]
Sector 3	−0.071***	−0.167***	−0.199***	−0.209***	1,227
	(0.008)	(0.021)	(0.025)	(0.045)	[440]
Sector 4	−0.102***	−0.173***	−0.181***	−0.190***	705
	(0.011)	(0.017)	(0.018)	(0.031)	[340]
Sector 5	−0.088***	−0.139***	−0.165***	−0.280***	630
	(0.012)	(0.020)	(0.021)	(0.067)	[312]
Sector 6	−0.081***	−0.139***	−0.145***	−0.188***	671
	(0.014)	(0.022)	(0.023)	(0.021)	[315]
Sector 7	−0.086***	−0.169***	−0.180***	−0.212***	1,512
	(0.008)	(0.013)	(0.014)	(0.026)	[661]
Sector 8	−0.070***	−0.135***	−0.154***	−0.163***	1,577
	(0.011)	(0.020)	(0.023)	(0.051)	[622]
Sector 9	−0.095***	−0.176***	−0.189***	−0.163***	311
	(0.015)	(0.025)	(0.027)	(0.024)	[183]
Sector 10	−0.073***	−0.127***	−0.131***	−0.155***	235
	(0.018)	(0.028)	(0.029)	(0.050)	[140]
Sector 11	−0.090***	−0.164***	−0.176***	−0.177***	1,569
	(0.010)	(0.017)	(0.019)	(0.033)	[689]
Sector 12	−0.091***	−0.174***	−0.182***	−0.199***	2,097
	(0.009)	(0.016)	(0.017)	(0.027)	[779]
Sector 13	−0.051***	−0.093***	−0.100***	−0.115***	486
	(0.013)	(0.022)	(0.023)	(0.029)	[250]
Sector 14	−0.073***	−0.130***	−0.139***	−0.232***	2,408
	(0.013)	(0.021)	(0.022)	(0.026)	[859]
Sector 15	−0.088***	−0.167***	−0.175***	−0.146***	1,146
	(0.010)	(0.018)	(0.019)	(0.036)	[489]
Sector 16	−0.066***	−0.139***	−0.146***	−0.178***	492
	(0.015)	(0.027)	(0.028)	(0.049)	[259]

Note: Heteroskedasticity-consistent standard errors clustered by municipalities are in parentheses. The unit of observation is municipality. Numbers in brackets are the number of municipalities. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Online Appendix C. Markup by Sector and Market Size in 1986–2000

Table OA.C. 1 presents descriptive statistics of markups by two-digit level sector in the 1986–2000 period, which corresponds to Table A.1 in the main text.

Figure OA.C. 1 compares the markup distributions for all sectors between large and small markets in the 1986–2000 period, which corresponds to Figure 2 in the main text.

Figure OA.C. 2 compares the markup distributions by two-digit level sector between large and small markets. in the 1986–2000 period, which corresponds to Figure 3 in the main text.

Figure OA.C. 3 shows the scatter plot of area-mean markup and market size in the 1986–2000 period, which corresponds to Figure 4 in the main text.

[Table OA.C. 1; Figures OA.C. 1–OA.C. 3]

Table OA.C. 1: Descriptive Statistics of Establishment-Level Markups by Sector, 1986–2000

Sector	Obs.	Mean	S.D.	p25	Median	p75
Sector 1	96,557	1.684	0.602	1.251	1.584	2.003
Sector 2	78,520	1.531	0.592	1.124	1.396	1.778
Sector 3	24,750	1.511	0.460	1.193	1.444	1.747
Sector 4	22,538	1.599	0.600	1.189	1.459	1.858
Sector 5	29,541	1.316	0.467	0.998	1.219	1.511
Sector 6	19,651	1.382	0.520	1.012	1.263	1.613
Sector 7	44,282	1.626	0.544	1.240	1.542	1.913
Sector 8	32,878	1.768	0.646	1.310	1.624	2.064
Sector 9	10,359	1.173	0.338	0.942	1.106	1.330
Sector 10	6,377	1.088	0.310	0.881	1.026	1.222
Sector 11	53,723	1.379	0.445	1.071	1.295	1.588
Sector 12	64,543	1.286	0.435	0.985	1.196	1.485
Sector 13	12,687	1.553	0.612	1.111	1.429	1.848
Sector 14	55,084	2.585	0.948	1.854	2.477	3.221
Sector 15	36,769	1.486	0.490	1.143	1.405	1.735
Sector 16	11,626	1.364	0.494	1.019	1.258	1.598

Note: The unit of observation is establishment. The list of sector number is in Table C.1 of the main text.

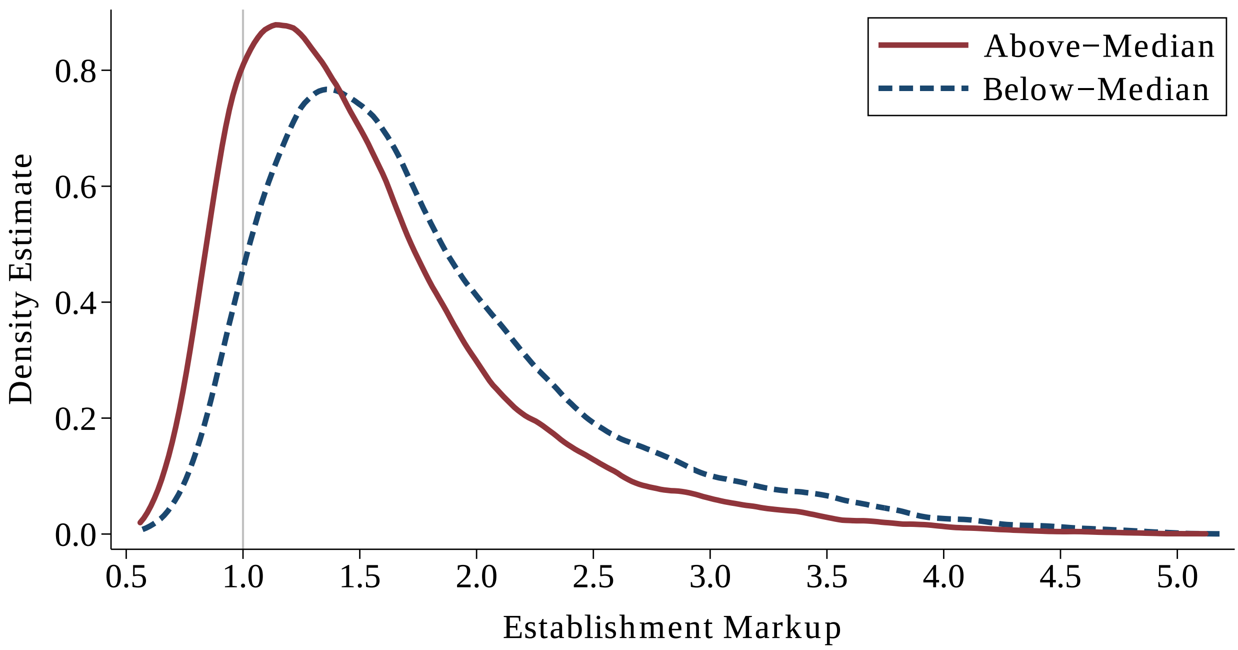
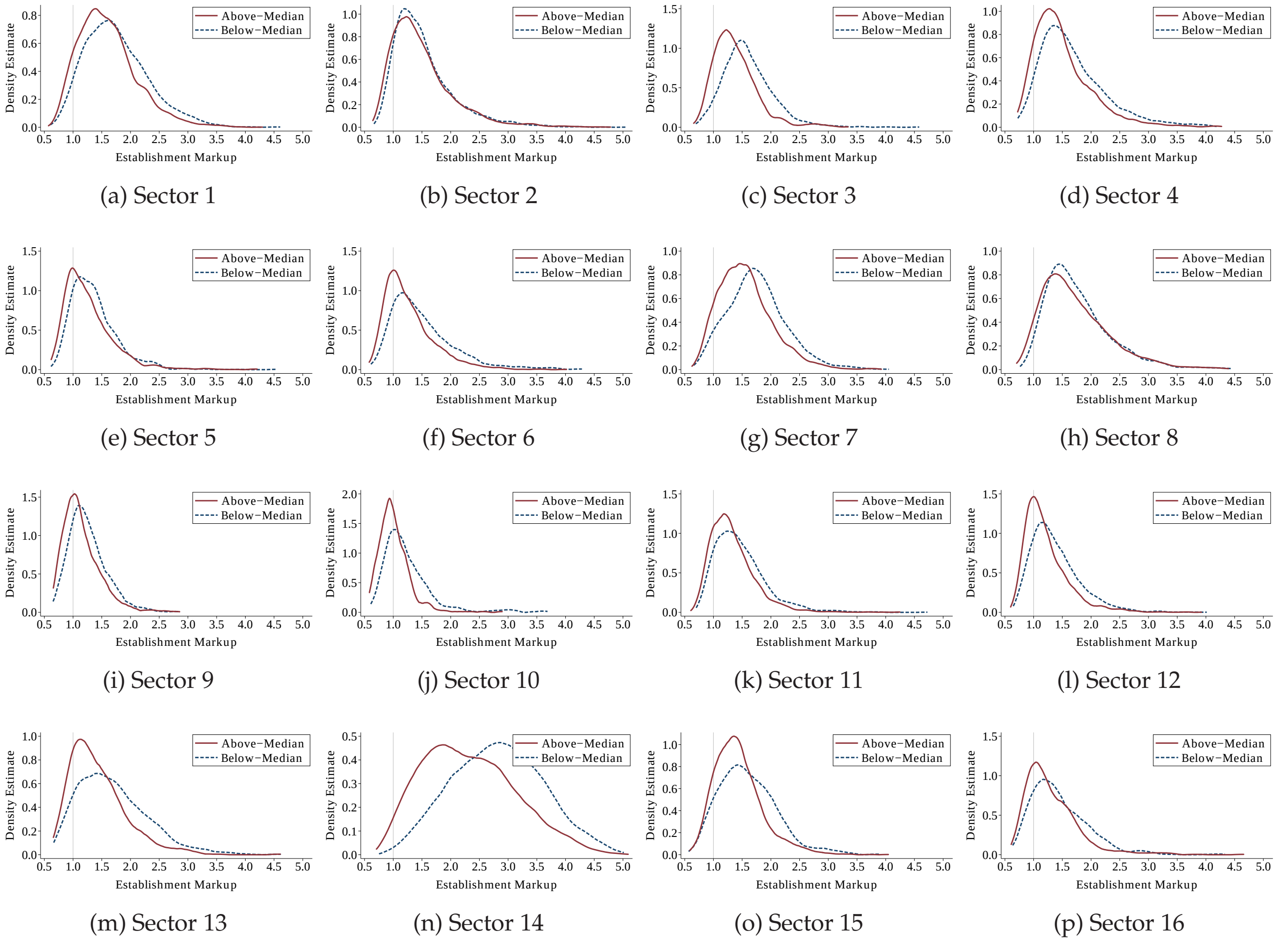


Figure OA.C. 1: Markup Distribution between Large and Small Markets, 1986–2000

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Average markup of establishment i in the corresponding period is calculated as $\mu_i = 1/T_i \sum_{t=1}^{T_i} \mu_{it}$. Average market size in city a where establishment i is located is calculated as $\mathrm{MP}_{a(i)} = 1/T_i \sum_{t=1}^{T_i} \mathrm{MP}_{a(i)t}$. The solid and dashed lines denote the markup distributions in cities with above-median and below-median population, respectively.



(a) Sector 1 (b) Sector 2 (c) Sector 3 (d) Sector 4

(e) Sector 5 (f) Sector 6 (g) Sector 7 (h) Sector 8

(i) Sector 9 (j) Sector 10 (k) Sector 11 (l) Sector 12

(m) Sector 13 (n) Sector 14 (o) Sector 15 (p) Sector 16

Figure OA.C. 2: Markup Distributions between Large and Small Markets by Sector, 1986–2000

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Establishment-mean markup is averaged across years as $\hat{\mu}_i = 1/T_i \sum_t \hat{\mu}_{it}$, where T_i is the number of years for establishment i observed in the sample. Sector numbers correspond to those used in Table C.1 in the main text. The solid and dashed lines denote the markup distributions in cities with above-median and below-median population, respectively.

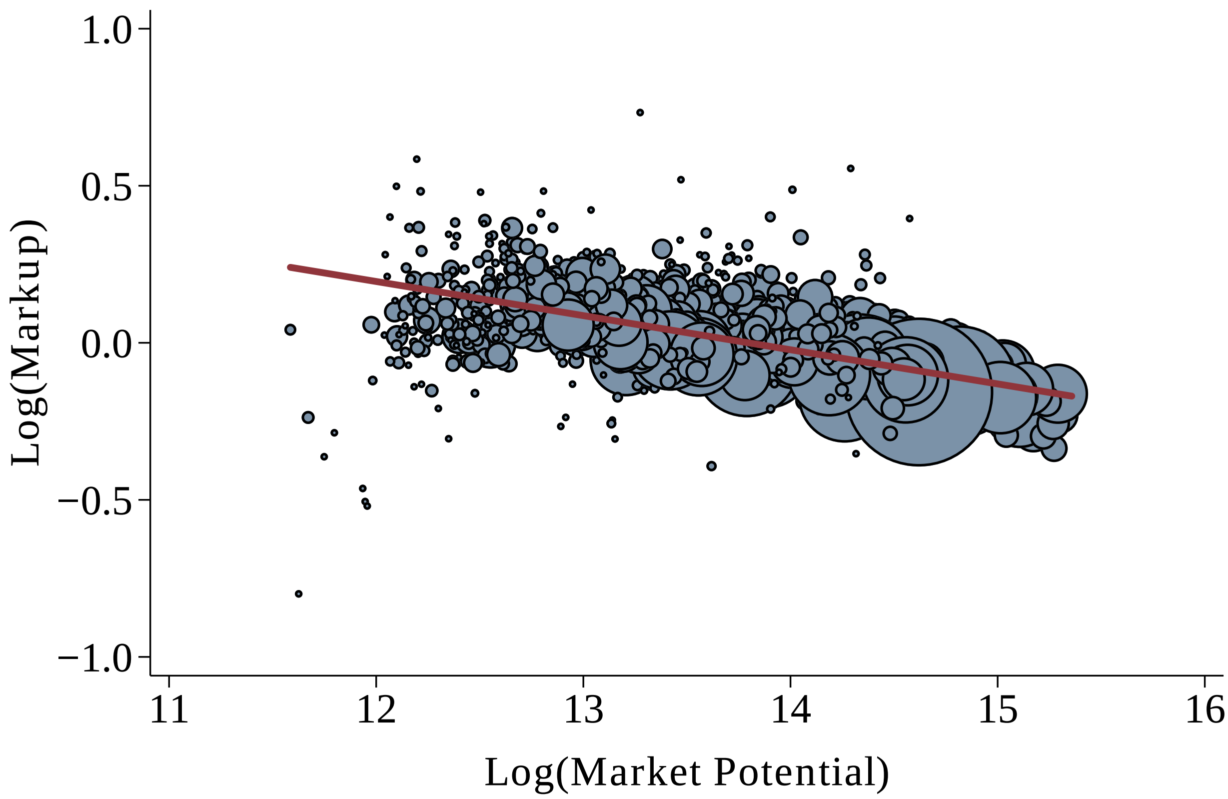


Figure OA.C. 3: Markup and Market Size

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Area-mean logarithm of markup is averaged across years as $\hat{\eta}_a = 1/T_a \sum_t \hat{\eta}_{at}$, where T_a is the number of years for area a observed in the sample. Market size is averaged across years as $\log(\mathrm{MP}_a) = 1/T_a \sum_t \log(\mathrm{MP}_{at})$. The market size size is measured by regional income including those of neighboring municipalities located within the circle of d km radius. The circle size represents the the number of manufacturing establishments in each geographical unit.

Online Appendix D. Markup and Wage in 1986–2000

Figure OA.D. 1 shows negative relationship between markup and wage at the establishment level in the 1986–2000 period, which corresponds Figure B.1 in the main text.

[Figure OA.D. 1]

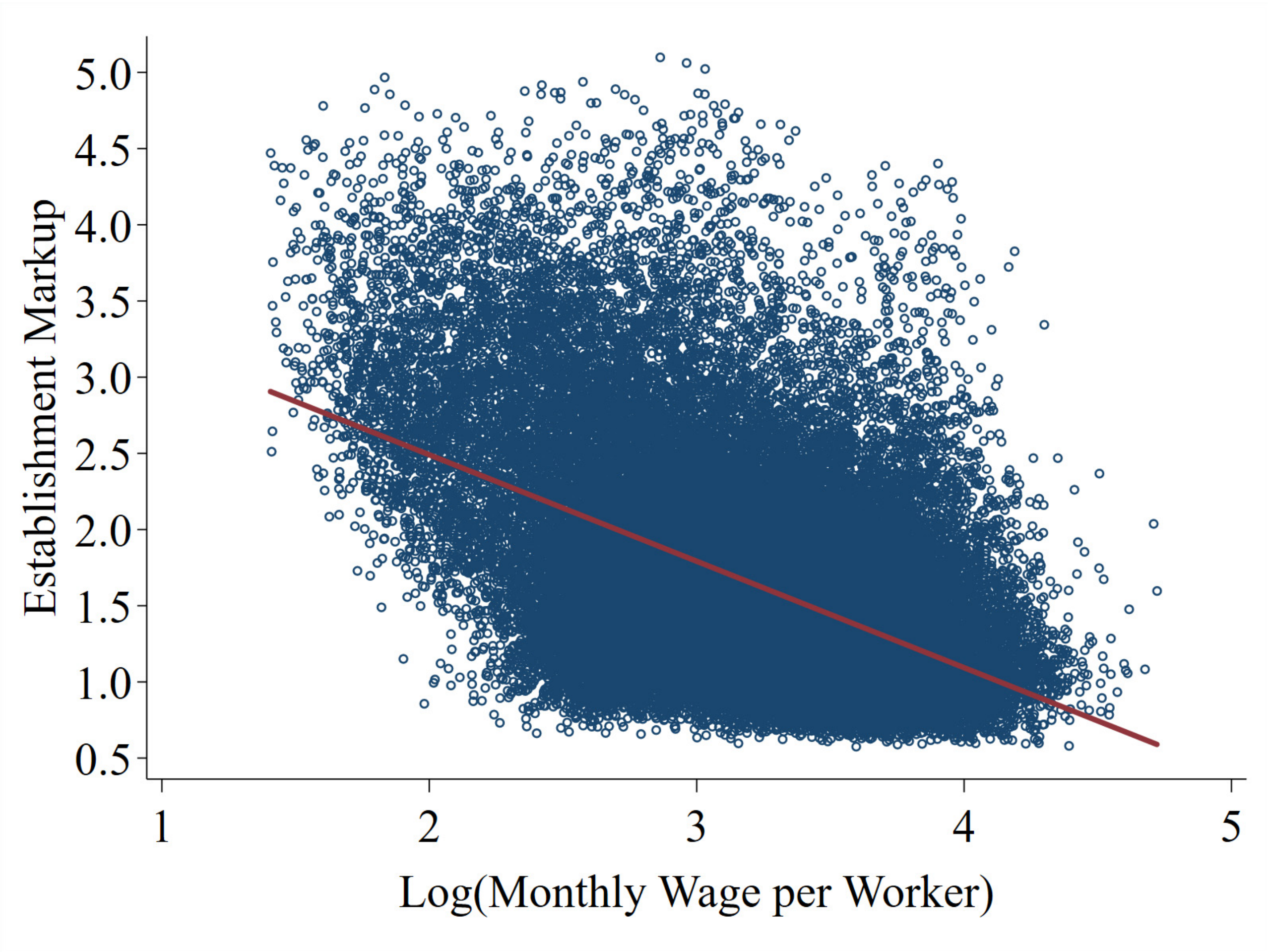


Figure OA.D. 1: Markup and Wage, 1986–2000

Note: Created by author using the microdata (questionnaire information) of the Census of Manufacture (Ministry of Economy, Trade and Industry). Average markup of establishment i in the corresponding period is calculated as $\mu_i = 1/T_i \sum_{t=1}^{T_i} \mu_{it}$. Logarithm of average monthly wage per worker of establishment i in the corresponding period is calculated as $\log(w_i) = 1/T_i \sum_{t=1}^{T_i} \log(w_{it})$.

References

De Loecker, Jan and Frederic Warzynski (2012) "Markups and firm-level export status," *American Economic Review* 102(6), pp. 2437-2471.