Quantile approach for distinguishing agglomeration from firm selection in Stata*

Keisuke Kondo†

March 18, 2017
(`estquant`: version 1.04)

Abstract

Firms and workers are, on average, more productive in larger cities. One possible explanation is that firms and workers in larger cities benefit from agglomeration economies. Another is that the higher concentration of economic activities in larger cities creates tougher competition, after which less-productive firms are forced to exit the market. To distinguish agglomeration from firm selection, Combes et al. (2012, "The productivity advantages of large cities: Distinguishing agglomeration from firm selection," *Econometrica*, vol. 80) suggested a quantile approach. Therefore, the present study introduces the `estquant` command, which implements the quantile approach in Stata. In addition, our Monte Carlo experiments emphasize the importance of simultaneously considering agglomeration and selection.

Keywords: `estquant`, quantile approach, agglomeration, selection

1 Introduction

It is well known that firms and workers are, on average, more productive in larger cities, and many economists have attempted to explain this stylized fact. One main strand in the literature of urban economics emphasizes the role of agglomeration economies. As observed by Marshall (1890), workers and firms enjoy the benefits from agglomeration economies through stronger input-output linkages, better matching between workers and firms in labor markets, and more active knowledge spillovers (e.g., Rosenthal and Strange, 2004). However, recent studies in this literature have shed new light on

*I thank Diego Puga for his helpful advice on the program code. I also thank Luigi Aldieri for his bug reporting. The `estquant` command fundamentally relies on the SAS program code of Combes et al. (2012). This is an outcome of research conducted under the project "Data Management" at the RIETI. The views expressed in this paper are solely those of the author, and do not represent those of the RIETI.

†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`).

another hypothesis of selection (e.g., Melitz, 2003; Melitz and Ottaviano, 2008). In this framework, selection refers to how less-productive firms are forced to exit the market as a result of tougher competition. Extending the firm selection model with the endogenous markup of Melitz and Ottaviano (2008), Combes et al. (2012a) show that the higher concentration of economic activities in larger cities forces tougher competition across firms, after which less-productive firms cannot survive there. Consequently, more-productive firms operate in larger cities, creating the impression that agglomeration economies, on average, make productivity higher. As such, there is an increasing need to develop a new empirical methodology in order to distinguish agglomeration from firm selection.

Combes et al. (2012a) developed a quantile approach to compare the productivity distributions between larger and smaller cities. Consistent with the multi-city setting and simultaneously incorporating agglomeration externalities, and extending the firm selection model of Melitz and Ottaviano (2008), they showed that the productivity distributions between larger and smaller cities are comparable via shift, dilation, and truncation components. In their model, agglomeration benefits are expressed by shifting productivity distribution of larger cities rightward and by making it dilated. Meanwhile, stronger selection leads to the left-truncation of the productivity distribution. In other words, the right-shift of distribution indicates that all firms in larger cities enjoy the same benefits from agglomeration, while the dilation effect indicates that more-productive firms in larger cities enjoy greater benefits from agglomeration. Moreover, the left-truncation of distribution indicates that less-productive firms are forced to exit the market due to tougher competition.

In this study, we introduce a new Stata command `estquant`, which implements the quantile approach suggested by Combes et al. (2012a). Although Combes et al. (2012a) concluded that firm selection does not play a crucial role in determining spatial productivity differences in the French manufacturing sector, this literature requires more empirical studies worldwide, since there may be counterexamples in other countries. Considering why differences arise across countries will deepen our understanding of agglomeration economies and firm selection. Thus, the `estquant` command is expected to expand this literature by providing significant evidence.

In order to promote intuitive interpretations of this quantile approach, this study presents useful numerical examples. In particular, our Monte Carlo experiments offer a good insight into how model misspecification leads to biases for estimates of shift, dilation, and truncation parameters. It has been shown that, if the selection is omitted when it matters in the true model, then agglomeration economies are estimated with an upward bias for the shift parameter and a downward bias for the dilation parameter.

In the existing literature, after applying the quantile approach to the Japanese manufacturing sector between 1986 and 2013, Kondo (2016) could hardly find evidence that stronger selection effects in larger cities, which is consistent with the findings of Combes et al. (2012a). Agglomeration economies better explain spatial productivity distributions, especially via the right-shift of productivity distribution for larger cities in the Japanese manufacturing sector in recent decades. Conversely, Arimoto et al. (2014) concluded that higher productivity in the prewar Japanese silk-reeling industrial clusters could be explained by firm selection. Using the Italian firm-level dataset, Accetturo et al. (2013) found that agglomeration economies play a key role in explaining spatial productivity differences. However, they emphasized that selection is also an important factor to explain such differences in some industries, especially when different spatial scales are considered (e.g., comparison between cities with larger and smaller market potentials).

Interestingly, one can use this quantile approach when comparing any two distributions in wide-ranging studies of economics. In fact, Combes et al. (2012b) applied a quantile approach to workers' skills measured from wages using the French workers' panel dataset. They found that, similar to firm productivity, workers' skills are, on average, higher in larger cities. The study of de la Roca and Puga (2017) also applied a quantile approach to compare distributions of workers' skills between larger and smaller cities. Although they obtained similar results to Combes et al. (2012b) in their static framework using the Spanish workers' panel dataset, their findings showed that workers' skills do not differ considerably between larger and smaller cities, after controlling for the dynamic learning effects of working in larger cities.[1] Thus, this quantile approach offers greater possibilities in applied economics, and the `estquant` command will help future studies easily utilize this new quantile approach.

The reminder of this paper is organized as follows. Section 2 reviews the quantile approach suggested by Combes et al. (2012a). Section 3 describes the `estquant` command. Section 4 conducts the Monte Carlo experiments to determine how model misspecification affects parameter estimation. Section 5 replicates the estimation results of Combes et al. (2012a). Finally, Section 6 presents the conclusions.

2 Quantile approach

As stated earlier, Combes et al. (2012a) proposed a new quantile approach to distinguish agglomeration effects from firm selection. Although the overall theory relates to the differences in distributions between larger and smaller cities, their quantile approach can be applied to a more general case of any two distributions. This section presents a brief review of their quantile approach.

2.1 Theoretical implications

Combes et al. (2012a) extended the firm selection model of Melitz and Ottaviano (2008) by introducing agglomeration economies as technological externalities proportional to the size of cities, including neighboring cities. Their model nests selection and agglomeration economies, both of which affect aggregate productivity through different channels. Their key theoretical predictions can be summarized by comparing the entire productivity distributions between larger and smaller cities, based on

[1] Combes et al. (2012b) did not strictly distinguish whether workers are *ex ante* or *ex post* skilled in their wage distribution analysis. Conversely, after focusing on the dynamic benefits from agglomeration economies, de la Roca and Puga (2017) found that, even if workers are *ex ante* identical, workers in larger cities acquire skills faster than those in smaller cities. In other words, workers in larger cities become *ex post* more-skilled via valuable experiences there, which can be considered as the *dynamic* benefits from agglomeration economies, rather than the spatial sorting of skills.

three key parameters (i.e., shift, dilation, and truncation). Agglomeration economies are captured by the right-shift of productivity distributions between smaller and larger cities (i.e., the right-shift captures the difference in average productivities between smaller and larger cities). Furthermore, Combes et al. (2012a) considered the dilation effect of agglomeration economies, which indicates that more-productive firms enjoy greater benefits from agglomeration (i.e., the dilation parameter indicates whether productivity distributions in larger cities are more dispersed than those in smaller cities). The estimation issue related to these two effects is that, when stronger truncation exists in the productivity distribution in larger cities, right-shift and dilation effects of agglomeration economies are overestimated and underestimated respectively, due to the omitted truncation parameter.[2]

Figure 1 presents four selective cases regarding agglomeration and selection. Panel (a) illustrates the case of the agglomeration economies through the right-shift of productivity distribution for larger cities, but the strength of selection is the same between the cities. In addition to Panel (a), Panel (b) includes the dilation effect of agglomeration economies, which means that the productivity distribution of larger cities is more dispersed than that of smaller cities. Panel (c) exhibits stronger selection in larger cities, but there are no agglomeration economies, while Panel (d) is the case in which larger cities show the right-shift and dilation effects of agglomeration economies as well as stronger selection. As mentioned earlier, omitting the stronger left-truncation of distribution for larger cities leads to an upward bias for the shift parameter (through a higher average) and a downward bias for the dilation parameter (through a smaller variance). One advantage of the quantile approach suggested by Combes et al. (2012a) is that it simultaneously estimates not only the relative strength of selection between larger and smaller cities but also the relative degrees of shift and dilation effects arising from agglomeration economies.

2.2 Basic assumption

At this point, we begin the discussion with Lemma 1 in Combes et al. (2012a).[3] Consider that there are two cumulative distribution functions of F_i and F_j for category i and category j, respectively (e.g., larger cities belong to category i and smaller cities belong to category j) and both F_i and F_j have some common underlying distribution $\tilde{F}$. Then, F_i can be obtained by shifting $\tilde{F}$ rightward by A_i, dilating $\tilde{F}$ by D_i, and left-truncating a share $S_i \in [0, 1)$ of $\tilde{F}$. In the same manner, F_j can be obtained by shifting $\tilde{F}$ rightward by A_j, dilating $\tilde{F}$ by D_j, and left-truncating a share $S_j \in [0, 1)$ of $\tilde{F}$. Furthermore, under the condition that F_i and F_j have some common underlying distribution $\tilde{F}$,

[2] A seminal paper on selection in productivity distribution is Syverson (2004), who showed that the strength of selection increases (which indirectly implies that the dispersion of productivity distribution decreases due to the selection) as the local market size increases. His indirect identification approach to selection using the inter-quantile range of distribution crucially depends on the assumption that the left-truncation of distribution leads to smaller variance. However, the smaller variance of distribution in larger cities is not necessarily driven by selection. This occurs when agglomeration economies benefit relatively less-productive firms greater than relatively more-productive firms. For example, relatively less-productive firms enjoy greater benefits through transactions with more-productive firms in locally segmented markets. Furthermore, Arimoto et al. (2014) also pointed out the possibility of faster technological catch-up through imitation.

[3] For simplicity of explanation, we skip details of their theoretical model with firm selection and the agglomeration economies developed by Combes et al. (2012a).

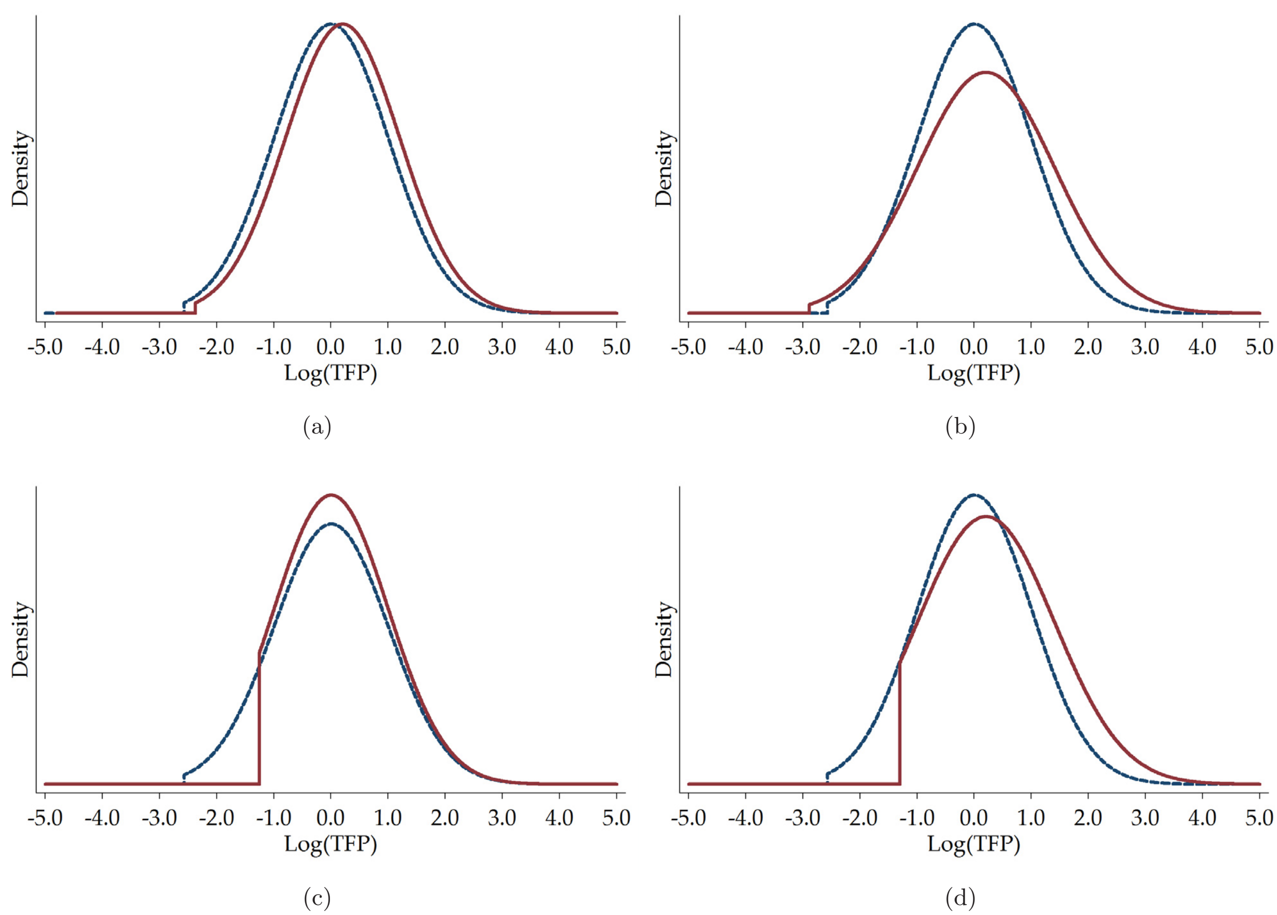


Figure 1: Comparing distributions between large and small cities

Note: Created by author. The solid and dashed lines denote productivity distributions for larger and smaller cities, respectively. Panel (a) shows agglomeration economies (right-shift only) and the same selection. Panel (b) shows agglomeration economies (right-shift and dilation) and the same selection. Panel (c) shows the same level of agglomeration economies and stronger selection in larger cities. Panel (d) shows agglomeration economies (right-shift and dilation) and stronger selection in larger cities.

we have the following relationship between F_i and F_j:

$$F_i(\phi) = \max\left(0, \frac{F_j\left(\dfrac{\phi - A}{D}\right) - S}{1 - S}\right), \quad \text{if} \quad S_i > S_j, \tag{1}$$

$$F_j(\phi) = \max\left(0, \frac{F_i\left(D\phi + A\right) - \dfrac{-S}{1-S}}{1 - \dfrac{-S}{1-S}}\right), \quad \text{if} \quad S_i < S_j, \tag{2}$$

where $D \equiv D_i/D_j$, $A \equiv A_i - DA_j$, $S \equiv (S_i - S_j)/(1 - S_j)$. The first equation shows that F_i can be obtained by dilating F_j by D, shifting F_j by A, and left-truncating a share S of F_j. In addition, the second equation shows that F_j can be obtained by dilating F_i by $1/D$, shifting F_i by $-A/D$, and left-truncating a share $-S/(1-S)$ of F_i.

This relationship helps us compare the two cumulative distribution functions without directly specifying an ad hoc underlying distribution $\tilde{F}$.[4] In addition to shift A and dilation D, we can examine the relative strength of truncation S of category i, compared to category j. Furthermore, parameter A measures how much stronger the right shift in category i is relative to category j, while parameter D measures the ratio of dilation in category i relative to category j. Finally, parameter S measures how much stronger the left truncation in category i is relative to category j.

2.3 Quantile transformation

To estimate Equations (1) and (2), we transform them into quantile functions. Suppose that the cumulative distribution function (CDF) is invertible. Let $\lambda_i(u) = F_i^{-1}(u)$ and $\lambda_j(u) = F_j^{-1}(u)$ denote the quantile functions of categories i and j, respectively, and u is the uth quantile. The quantile function is defined for all $u \in [0, 1]$.

If $S > 0$, then the quantile function can be obtained from Equation (1) as follows:

$$\lambda_i(u) = D\lambda_j\big(S + (1-S)u\big) + A, \quad \text{for} \quad u \in [0,1]. \tag{3}$$

If $S < 0$, then the quantile function can be obtained from Equation (2) as follows:

$$\lambda_j(v) = \frac{1}{D}\lambda_i\left(\frac{v-S}{1-S}\right) - \frac{A}{D}, \quad \text{for} \quad v \in [0,1].$$

Then, we use the transformation of the variable by $u = (v-S)/(1-S)$ and thus, the quantile function can be rewritten as follows:

$$\lambda_j\big(S + (1-S)u\big) = \frac{1}{D}\lambda_i(u) - \frac{A}{D} \quad \text{for} \quad u \in \left[\frac{-S}{1-S}, 1\right] \tag{4}$$

Combining Equations (3) and (4) for all S yields

$$\lambda_i(u) = D\lambda_j\big(S + (1-S)u\big) + A \quad \text{for} \quad u \in \left[\max\left(0, \frac{-S}{1-S}\right), 1\right].$$

This equation cannot be directly estimated since the set of ranks u includes the unknown true parameter S. Thus, additional variable transformation provides the following equation:

$$\lambda_i\big(r_S(u)\big) = D\lambda_j\big(S + (1-S)r_S(u)\big) + A \quad \text{for} \quad u \in [0,1], \tag{5}$$

[4] A quantile approach does not need to specify an ad hoc underlying distribution beforehand, but the assumption that two distributions need to have some common underlying distribution might be strong in empirical analyses.

where $r_S(u) = \max\big(0, -S/(1-S)\big) + \big(1 - \max\big(0, -S/(1-S)\big)\big)\, u$.

An intuitive understanding of Equation (5) is as follows. Consider the case in which $S = 0$, and then Equation (5) becomes

$$\lambda_i\big(u\big) = D\lambda_j\big(u\big) + A, \quad \text{for} \quad u \in [0,1]\,.$$

This equation expresses how well the relationship between the quantile functions $\lambda_i(u)$ and $\lambda_j(u)$ is explained by the relative parameters of dilation D and shift A. Figure 2(a) presents this case, in which a common underlying distribution is standard normal distribution and the true parameters are set as $A_1 = 0.2$, $A_2 = 0$, $D_1 = D_2 = 1$, and $S_1 = S_2 = 0.005$. Thus, the relative shift parameter is $A = 0.2$, the relative dilation parameter is $D = 1$, and the relative truncation parameter is $S = 1$. In Figure 2(a), the upper shift of quantile function for Category 2 is expressed by the positive value of parameter A, but the steepness is the same between the two quantile functions between Categories 1 and 2. When $A = 0$, $D = 1$, and $S = 0$, both distributions are identical. Hence, parameters A and D capture the differences between two distributions.

An empirical issue raised by Combes et al. (2012a) is that firm selection, if it exists, generates biases in parameters A and D. Particularly, firm selection offers a misleading interpretation that agglomeration economies, on average, increase productivity. Figure 2(b) presents the case of $S = 0.1$, in which a common underlying distribution is the standard normal distribution and the true parameters are set as $A_1 = 0.2$, $A_2 = 0$, $D_1 = D_2 = 1$, $S_1 = 0.005$, and $S_2 = 0.1045$. Thus, the relative shift parameter is $A = 0.2$, and the relative dilation parameter is $D = 1$. The quantile approach endogenously matches the quantile ranges between the two distributions in accordance with relative selection S, while simultaneously examining the relative shift parameter A and the relative dilation parameter D. For example, in Figure 2(b), the quantile approach compares the logarithm of total factor productivity (TFP) at the 0th quantile of distribution for smaller cities with the logarithm of TFP at the 0.1th quantile for larger cities, since $S = 0.1$. The quantile approach conducts such adjustment at each quantile while estimating S.

2.4 Estimating quantile functions

Let $\boldsymbol{\theta} = (A, D, S)'$ denote the parameter vector. In order to estimate $\boldsymbol{\theta}$, we define the infinite set of equalities:[5]

$$m_{\boldsymbol{\theta}}(u) = \lambda_i\big(r_S(u)\big) - D\lambda_j\big(S + (1-S)r_S(u)\big) - A, \quad \text{for} \quad u \in [0,1]. \tag{6}$$

To consider the asymmetric relationship between the two distributions arising from the opposite

[5] One difficulty is that we have an infinite set of equalities due to the continuous quantile u. Thus, we need to approximate them by finite set of equalities for estimation. See Combes et al. (2012a) for the details of the estimation methodology.

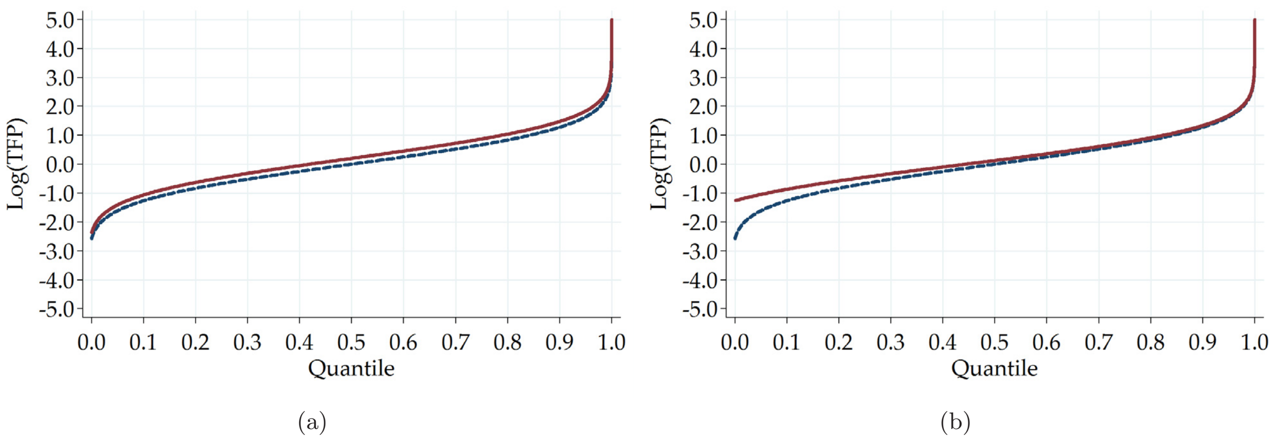


Figure 2: Comparing empirical quantile functions between two categories

Note: Created by author. The solid and dashed lines denote productivity distributions for larger and smaller cities, respectively. Panel (a) shows agglomeration economies (right-shift only) and the same selection. Panel (b) shows the same level of agglomeration economies and stronger selection in larger cities. Panel (a) and Panel (b) here correspond to Panel (a) and (c) in Figure 1, respectively.

transformation, we also define the following infinite set of equalities:

$$\tilde{m}_{\boldsymbol{\theta}}(u) = \lambda_j\big(\tilde{r}_S(u)\big) - \frac{1}{D}\lambda_i\left(\frac{\tilde{r}_S(u) - S}{1-S}\right) + \frac{A}{D}, \quad \text{for} \quad u \in [0,1], \tag{7}$$

where $\tilde{r}_S(u) = \max\big(0, S\big) + \big(1 - \max(0, S)\big)u$.

The estimator of $\boldsymbol{\theta}$ can be obtained by minimizing the criteria function $M(\boldsymbol{\theta})$, which is defined as the sum of the squared values of $\hat{m}_{\boldsymbol{\theta}}(u)$ and $\hat{\tilde{m}}_{\boldsymbol{\theta}}(u)$, as follows:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta}} \; M(\boldsymbol{\theta}), \\ M(\boldsymbol{\theta}) &= \int_0^1 \big[\hat{m}_{\boldsymbol{\theta}}(u)\big]^2 \mathrm{d}u + \int_0^1 \big[\hat{\tilde{m}}_{\boldsymbol{\theta}}(u)\big]^2 \mathrm{d}u, \end{aligned} \tag{8}$$

where $\hat{m}_{\boldsymbol{\theta}}(u)$ and $\hat{\tilde{m}}_{\boldsymbol{\theta}}(u)$ are obtained from the empirical quantile functions $\hat{\lambda}_i$ and $\hat{\lambda}_j$ by data. Finally, to measure the fitness of the model, we define the pseudo R^2 as follows:

$$R^2 = 1 - \frac{M(\hat{A}, \hat{D}, \hat{S})}{M(0, 1, 0)}.$$

Note that this pseudo R^2 can take a small value even if the model specification better fits the data. For example, R^2 becomes 0 in the case where $\hat{A} = 0$, $\hat{D} = 1$, and $\hat{S} = 0$ are obtained (i.e., the two distributions are identical).

2.5 Bootstrap standard errors

In order to estimate the standard errors of the estimated parameters $\hat{\boldsymbol{\theta}}$, we use the bootstrap method. We draw the observations of the same sample size as data with replacement and then estimate $\boldsymbol{\theta}$ for each bootstrap replication. In other words, when we have B bootstrap replications, there are B estimates for $\boldsymbol{\theta}$. Hence, the bootstrap standard errors $\widehat{\mathrm{SE}}_B(\hat{\theta}_k)$ are calculated as follows:

$$\widehat{\mathrm{SE}}_B(\hat{\theta}_k) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_k^{(b)} - \bar{\theta}_k)^2}, \quad k \in (A, D, S),$$

where $\bar{\theta}_k$ is the mean of $\hat{\theta}_k^{(b)}$ obtained from each bootstrap sample ($\bar{\theta}_k = B^{-1}\sum_{b=1}^{B}\hat{\theta}_k^{(b)}$). Note that $\bar{\theta}_k$ is not equal to the $\hat{\theta}_k$ observed in the sample.

The confidence interval can also be constructed by the bootstrap method. One method is to use the normal-based confidence interval. Using the bootstrap standard errors, the confidence interval at the $1-\alpha$ level is calculated as $(\hat{\theta}_k + \widehat{\mathrm{SE}}_B(\hat{\theta}_k)z_{\alpha/2}, \hat{\theta}_k - \widehat{\mathrm{SE}}_B(\hat{\theta}_k)z_{\alpha/2})$, where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution. Note that $z_{\alpha/2}$ takes the negative value and that $z_{\alpha/2}$ is equivalent to the negative of $z_{1-(\alpha/2)}$, due to the symmetric distribution. Another method is to construct the bootstrap-based confidence interval. When we have sufficiently large B values of $\hat{\theta}_k$ from bootstrap samples, we can construct a $1-\alpha$ level confidence interval using the $\alpha/2$ quantile and the $1-(\alpha/2)$ quantile of distribution for $\hat{\theta}_k^{(b)}$.

3 Implementation in Stata

3.1 Syntax

`estquant` *varname* [*if*] [*in*] , `category(`*varname*`)` [`shift` `dilation` `truncation` `qrange(#)` `bvariable(on|off)` `breplication(#)` `bsampling(#)` `strata` `maxiteration(#)` `eps1(#)` `eps2(#)` `ci(normal|bootstrap)` `level(#)`]

3.2 Options

`category(`*varname*`)` specifies the variable classifying the sample into two categories. This category variable must be binary, but it can take any values (e.g., 0 and 1 or 1 and 2).

`shift` estimates the relative shift parameter A. When this option is not specified, the relative shift parameter is constrained as $A = 0$.

`dilation` estimates the relative dilation parameter D. When this option is not specified, the relative dilation parameter is constrained as $D = 1$.

`truncation` estimates the relative truncation parameter S. When this option is not specified, the relative truncation parameter is constrained as $S = 0$.

`qrange(#)` specifies the range of quantile function. The quantile range $[0, 1]$ is divided into # ranges. The default value is 1,000.

`bvariable(on|off)` specifies whether the bootstrap uses variables prepared beforehand in the dataset. If this option is `on`, then the bootstrap replications are conducted using the *varname* named in a sequential order at each iteration. If this option is `off`, then the bootstrap replications are conducted by resampling *varname* at each iteration. The default setting is `off`.

`breplication(#)` specifies the number of the bootstrap replications. If this option takes the value of 0, bootstrap replication is skipped and bootstrap standard errors are not calculated. If `bvariable(on)` is specified, then the `breplication(#)` must be the last number of *varname* named in sequential order. The default value is 50.

`bsampling(#)` specifies the percentage of the sample size for bootstrap sampling. The default value is 100 (%), meaning that observations of the same sample size are drawn for bootstrap sampling. This option is ignored when `bvariable(on)` is specified.

`strata` fixes the number of observations in each category in each bootstrap replication. The `strata` option is not used in the default setting.

`maxiteration(#)` specifies the maximum number of iterations in numerical optimization. The default value is 1e+3.

`eps1(#)` specifies the convergence tolerance in numerical optimization. The stopping rule of `eps1(#)` is defined as the relative difference in parameter θ_k at g iteration and at $g-1$ iteration, $(\theta_k^{(g)} - \theta_k^{(g-1)})/(\theta_k^{(g-1)} + 1)$. The default value is 1e-6.

`eps2(#)` specifies the convergence tolerance in numerical optimization. The stopping rule of `eps2(#)` is defined as $\boldsymbol{g}'\boldsymbol{H}^{-1}\boldsymbol{g}$, where $\boldsymbol{g}$ is the gradient vector of $M(\boldsymbol{\theta})$ and $\boldsymbol{H}$ is the Hessian matrix of $M(\boldsymbol{\theta})$. The default value is 1e-6.

`ci(normal|bootstrap)` specifies types of confidence interval. The `ci()` option allows one to use the normal- and bootstrap-based confidence intervals. If bootstrap-based confidence interval is constructed, then a large number of bootstrap replications should be specified in the `breplication(#)` option. The default setting constructs the normal-based confidence interval.

`level(#)` specifies the level of the confidence interval. The default level is 95.0 (%).

❑ Technical note

The numerical optimization is conducted by a two-step iteration. The vector $\boldsymbol{\theta}$ is divided into two blocks: $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \theta_2)'$, where $\boldsymbol{\theta}_1 = (A, D)'$ and $\theta_2 = S$. The first step in minimizing $M(\boldsymbol{\theta})$ is to concentrate it with respect to parameter $\boldsymbol{\theta}_1$ with the remaining parameter θ_2 (some initial value is used). The second step is to minimize $M(\boldsymbol{\theta})$ with respect to θ_2 given the parameter $\boldsymbol{\theta}_1$ obtained in the first step. The two-step iteration continues until the two stopping rules described for `eps1(#)` and `eps2(#)` are satisfied.

The numerical optimization may fail during the bootstrap replications. Especially, the numerical optimization of S may be unstable. There are some practical solutions. First, it is advisable to change the seed of random number generator. For example, the `set seed` command changes the seed on

Stata. Second, it is possible to change the convergence tolerance by the `eps1()` and `eps2()` options. However, the weak convergence tolerance is not recommended.

❏

3.3 Stored results

The `estquant` command stores the following results in `eclass`.

Scalars

`e(N)`	number of observations	`e(N1)`	# of observations (category 1)
`e(N2)`	# of observations (category 2)	`e(brep)`	# of bootstrap replications
`e(bsample)`	bootstrap sample size (%)	`e(r2)`	pseudo R-squared
`e(cri)`	convergence criteria (total)	`e(cr1)`	convergence criteria 1
`e(cr2)`	convergence criteria 2	`e(mean_y)`	mean of variable
`e(mean_y1)`	mean of variable (category 1)	`e(mean_y2)`	mean of variable (category 2)
`e(sd_y)`	s.d. of variable	`e(sd_y1)`	s.d. of variable (category 1)
`e(sd_y2)`	s.d. of variable (category 2)		

Matrices

`e(b)`	estimates of parameters	`e(V)`	variance matrix
`e(B)`	estimates from bootstrapping		

Macros

`e(cmd)`	`estquant`	`e(varname)`	name of variable

Function

`e(sample)`

3.4 Basic manipulation

The `estquant` command requires two variables for estimation. The main variable is specified as a variable of distribution, while the other variable is a category variable, which divides the sample into two categories. Users need to add `shift`, `dilation`, and `truncation` options depending on the models. The `estquant` command only displays the criteria if not any of `shift`, `dilation`, or `truncation` options are specified.

Here, we offer an example using basic commands. Consider a sample with two variables: `lntfp` (logarithm of TFP) and `cat` (binary category, 0 and 1). The `estquant` command estimates the relative shift, dilation, and truncation parameters by the following command:[6]

(*Continued on next page*)

[6]This result is obtained from one sample simulated with parameter setting ($A = 0.1$, D=1.2, S=0) in Figure 2.

```
. estquant lntfp, cat(cat) sh di tr
Bootstrap replications (50)

Completed:  10%
Completed:  20%
Completed:  30%
Completed:  40%
Completed:  50%
Completed:  60%
Completed:  70%
Completed:  80%
Completed:  90%
Completed: 100%

         cat |     Obs.       Mean       S.D.         Number of obs =     10000
-------------+---------------------------------       Replications  =        50
           0 |     5000       1.01      .9879         BS Sample (%) =       100
           1 |     5000      1.294      1.203         Pseudo R2     =    0.9966
-------------+---------------------------------
       Total |    10000      1.152       1.11

------------------------------------------------------------------------------
             |  Observed  Bootstrap                         Normal-based
       lntfp |     Coef.  Std. Err.      z    P>|z|    [95.0% Conf. Interval]
-------------+----------------------------------------------------------------
     Shift A |  .06683554   .037371    1.79   0.074   -.00641097     .140082
  Dilation D |   1.216367   .017782   12.17   0.000      1.18152     1.25122
Truncation S | -.00035072  .0021036   -0.17   0.868   -.00447374    .0037723
-------------+----------------------------------------------------------------
Constrained parameters: nothing
------------------------------------------------------------------------------
```

In the default setting, the estimation process includes 50 bootstrap replications. The upper portion of the table presents the summary statistics for the two categories (i.e., mean and standard deviation), while the lower portion presents the estimation results of the quantile functions. If either `shift`, `dilation`, or `truncation` is omitted, then the "Constrained parameters" section in the table presents these conditions. In the above example, there are no restrictions on the parameters. Note that hypothesis testing for parameters A and S is based on the null hypothesis $A = 0$ and $S = 0$, while parameter D is based on the null hypothesis $D = 1$.[7]

❏ Technical note

In this example, the bootstrap replications are conducted by resampling `lntfp` in each bootstrap replication. However, Combes et al. (2012a) re-estimate TFP from the bootstrapped firms' sample in each iteration. Similarly, de la Roca and Puga (2017) re-estimate workers' fixed-effects of wages from the bootstrapped workers' sample in each iteration. The `estquant` command can estimate bootstrap standard errors using variables previously prepared in the dataset. The `bvariable(on|off)` option

[7] The quantile approach includes numerical optimization, which takes time when bootstrap replications are conducted. In the above example, it takes approximately six minutes for 50 bootstrap replications (OS: Windows 10 64 bit, CPU: Intel Core i5-6500, Memory: 16GB, Stata: Version 14.2 2-Core).

includes variables bootstrapped in the dataset. Note that the variables named in sequential order, such as `lntfp1`, `lntfp2`, ..., `lntfp#`, are required until the number of bootstrap replications specified by the `breplication(#)` option. Suppose that the dataset includes the variable `lntfp`, bootstrapped variables `lntfp1`–`lntfp50`, category variable `cat` (binary category, 0 and 1), and category variables corresponding to bootstrapped variables `cat1`–`cat50`. Then, the following command estimates the bootstrap standard errors using variables bootstrapped beforehand:

```
. estquant lntfp, cat(cat) sh di tr bvar(on) brep(50)
```

See the sample program available online for the details.

❏

4 Monte Carlo experiments

In this section, we perform the Monte Carlo experiments to determine how this new quantile approach captures the truncation in distribution, after which we discuss the possible biases in parameters A and D when constraining the truncation parameter as $S = 0$.

We assume that a common underlying distribution of the logarithm of TFP is standard normal distribution $\mathrm{N}(0, 1)$. Suppose that Categories 1 and 2 correspond to smaller and larger cities, respectively. We assume that the productivity distribution for category 1 follows a truncated normal distribution with parameters $A_1 = 1$, $D_1 = 2$, and $S_1 = 0.001$. Specifying the distribution of Category 1 and the true values of parameters A, D, and S automatically determines the distribution of Category 2, based on the definitions of relative parameters. In the Monte Carlo experiments, we consider the data generating process (DGP) of Category 2 with the following three cases of true parameter settings:

1. DGP_1: shift + dilation ($A = 0.1$, $D = 1$, $S = 0$)
2. DGP_2: truncation ($A = 0$, $D = 1$, $S = 0.1$)
3. DGP_3: shift + dilation + truncation ($A = 0.1$, $D = 1.2$, $S = 0.1$)

Table 1 summarizes the parameter settings for the Monte Carlo experiments. The process was replicated 1,000 times with a sample size of 10,000. In addition, each category includes 5,000 observations and the quantile range is set to 1,000.

4.1 Generating samples

Based on the distribution assumptions, we draw the values from the corresponding distributions. As for Category 1, the logarithm of TFP (ϕ_{t1}) is drawn from the normal distribution with mean $\mu_1 \equiv A_1$ and variance $\sigma_1^2 \equiv D_1^2$, which is truncated at $\sigma_1 \Phi^{-1}(S_1) + \mu_1$:

$$\phi_{t1} \sim \mathrm{TN}(\mu_1, \sigma_1^2) \quad \text{for} \quad \phi_1 \in [\sigma_1 \Phi^{-1}(S_1) + \mu_1, \infty),$$

Table 1: Parameter setting for the Monte Carlo experiments

	Relative Parameters			Parameters for Category 2		
	A	D	S	A_2	D_2	S_2
Data Generating Process	(1)	(2)	(3)	(4)	(5)	(6)
Shift + Dilation	0.1	1.2	0	1.3	2.4	0.001
Truncation	0	1	0.1	1	2	0.1009
Shift + Dilation + Truncation	0.1	1.2	0.1	1.3	2.4	0.1009

Note: The common underlying distribution is standard normal distribution $\mathrm{N}(0,1)$. For the three cases, it is assumed that the distribution for Category 1 follows truncated normal distribution with parameters $A_1 = 1$, $D_1 = 2$, and $S_1 = 0.001$. From the definition of relative shift, dilation, and truncation parameters, and the true values of these parameters, we obtain the parameters for Category 2 as $A_2 = DA_1 + A$, $D_2 = DD_1$, and $S_2 = S_1 + (1 - S_1)S$.

where t is index of observations, μ_1 and σ_1^2 are the mean and variance of a normal distribution, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The distribution is left-truncated by the share of S_1 from the standard normal distribution with mean μ_1 and variance σ_1^2. The distribution for Category 1 is common for all three cases of DGPs.

Next, we draw the logarithm of TFP for Category 2 from the three different distributions. In the first case, the distribution of Category 2 is right-shifted and dilated compared to Category 1, and the selection level is the same between the two categories. Then, the logarithm of TFP for Category 2 is drawn from the normal distribution with mean $\mu_2 \equiv \mu_1 + A$ and variance $\sigma_2^2 \equiv (D\sigma_1)^2$, which is truncated by the share of S_1 as follows:

$$\phi_{t2}^{\mathrm{SD}} \sim \mathrm{TN}(\mu_2, \sigma_2^2), \quad \text{for} \quad \phi_2^{\mathrm{SD}} \in [\sigma_2 \Phi^{-1}(S_1) + \mu_2, \infty),$$

where the superscript SD indicates shift and dilation.

The second case considers the distribution for Category 2, which only includes the truncation compared to the distribution of Category 1. The logarithm of TFP for Category 2 is drawn from the normal distribution with mean $\mu_2 \equiv \mu_1$ and variance $\sigma_2^2 \equiv \sigma_1^2$, which is truncated by the share of S_2 as follows

$$\phi_{t2}^{\mathrm{T}} \sim \mathrm{TN}(\mu_2, \sigma_2^2) \quad \text{for} \quad \phi_2^{\mathrm{T}} \in [\sigma_2 \Phi^{-1}(S_2) + \mu_2, \infty),$$

where the superscript T indicates truncation

The third case considers the distribution for Category 2, which includes the right-shift, dilation, and truncation compared to the distribution of Category 1. The logarithm of TFP for Category 2 is drawn from the normal distribution with mean $\mu_2 \equiv D\mu_1 + A$ and variance $\sigma_2^2 \equiv (D\sigma_1)^2$, which is truncated by the share of S_2 as follows

$$\phi_{t2}^{\mathrm{SDT}} \sim \mathrm{TN}(\mu_2, \sigma_2^2) \quad \text{for} \quad \phi_2^{\mathrm{SDT}} \in [\sigma_2 \Phi^{-1}(S_2) + \mu_2, \infty),$$

where the superscript SDT indicates shift, dilation, and truncation.

Sampling the logarithm of TFP from truncated normal distributions is implemented by the probability inverse transformation method. Thus, the logarithm of TFP for the three DGPs are drawn using random number u_t from the uniform distribution as follows:

$$\begin{aligned}\phi_{t2}^{\mathrm{SD}} &= (\mu_1 + A) + D\sigma_1\, \Phi^{-1}\big(S_1 + (1 - S_1)u_t\big),\\ \phi_{t2}^{\mathrm{T}} &= \mu_1 + \sigma_1\, \Phi^{-1}\big(S_2 + (1 - S_2)u_t\big),\\ \phi_{t2}^{\mathrm{SDT}} &= (D\mu_1 + A) + D\sigma_1\, \Phi^{-1}\big(S_2 + (1 - S_2)u_t\big),\end{aligned}$$

where the same random number u_t is used across the three DGPs.

4.2 Results and bias from model misspecification

Using the simulated samples, we examine how model misspecification affects the estimation of parameters A, D, and S. To numerically evaluate it, we compute the bias of each parameter as the difference between the mean of the estimates obtained from G samples, $\mathrm{E}(\hat\theta_k^{(g)})$, and true value of each parameter, θ_{0k}:

$$\mathrm{Bias}(\hat\theta_k) = \mathrm{E}(\hat\theta_k^{(g)}) - \theta_{0k}, \quad k \in (A, D, S),$$

where (g) indicates gth iteration of G samples.

Table 2 presents results of the Monte Carlo experiments.[8] For each DGP, we consider two specifications of the model: (1) shift + dilation and (2) shift + dilation + truncation.

In the case of DGP_1, we can see that both model specifications correctly estimate parameters A and D. Even if relative truncation parameter S is additionally estimated, it is estimated around 0. Therefore, the overspecification of the model does not generate bias for parameters.

In the case of DGP_2, the first model underspecifies the DGP by missing truncation. Parameter estimates $\hat A$ and $\hat D$ are biased upward and downward, respectively. However, when the model is correctly specified including truncation, parameters are correctly estimated. The results of DGP_2 are qualitatively the same as those of DGP_3. However, if the dilation exists in distribution of category 2, the both biases in A and D are amplified.

Figure 3 illustrates the numerical results of the Monte Carlo experiments for shift parameter A. The case of DGP_1 is shown in Figure 3(a) and Figure 3(b). When the model is correctly specified, estimates $\hat A$ are distributed around the true value of shift parameter A ($A = 0.1$). Even if researchers additionally consider truncation S, the shift parameter A is correctly estimated, suggesting that overspecification of the model does not generate bias for parameters.

The case of the DGP_2 is shown in Figure 3(c) and Figure 3(d). When the model is misspecified, $\hat A$ has an upward bias that is estimated far from the true value of the relative shift parameter A. When the model is correctly specified, the relative shift parameter A is correctly estimated around

[8] We also performed Monte Carlo experiments by changing the number of observations and quantile ranges. In this study, we offer the case of 10,000 observations and quantile ranges partitioned into 1,000.

Table 2: Results of the Monte Carlo experiments

	Mean			Bias		
	$\hat{A}$	$\hat{D}$	$\hat{S}$	$\hat{A}$	$\hat{D}$	$\hat{S}$
Model Specification	(1)	(2)	(3)	(4)	(5)	(6)
	DGP$_1$: Shift + Dilation (True Parameters: $A = 0.1$, $D = 1.2$, $S = 0$)					
Shift + Dilation	0.0993	1.1987		−0.0007	−0.0013	
Shift + Dilation + Truncation	0.0990	1.1988	0.0000	−0.0010	−0.0012	0.0000
	DGP$_2$: Truncation (True Parameters: $A = 0$, $D = 1$, $S = 0.1$)					
Shift + Dilation	0.5382	0.8476		0.5382	−0.1524	
Shift + Dilation + Truncation	−0.0088	1.0010	0.1015	−0.0088	0.0010	0.0015
	DGP$_3$: Shift + Dilation + Truncation (True Parameters: $A = 0.1$, $D = 1.2$, $S = 0.1$)					
Shift + Dilation	0.7477	1.0154		0.6477	−0.1846	
Shift + Dilation + Truncation	0.0905	1.2009	0.1013	−0.0095	0.0009	0.0013

Note: A total of 1,000 replications were conducted in the Monte Carlo experiments. Here, bias is defined as the difference between the mean of the estimated parameters and the true value for the corresponding parameters. The sample size is 10,000, and each category includes 5,000 observations. The quantile range is 1,000.

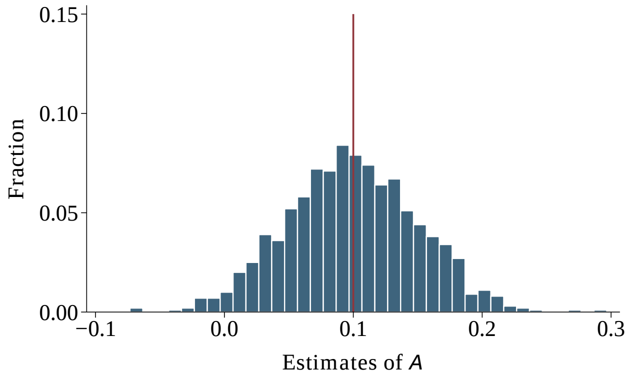
the true value. The case of the DGP$_3$ is shown in Figure 3(d) and Figure 3(d). Here, the results are qualitatively the same as those in DGP$_2$. As discussed earlier, the dilation amplifies the biases for both parameters A and D.

The important conclusion from the Monte Carlo experiments is that, when researchers omit the relative truncation parameter S in distributions (i.e., implicitly imposing the restriction $S = 0$), the relative shift parameter A is estimated with an upward bias and the relative dilation parameter D is estimated with a downward bias. This type of estimation is commonly conducted when researchers simply compare the average productivity between larger and smaller cities through a standard statistical method, after which subsequent studies might draw the misleading conclusion that agglomeration economies lead to higher firm productivity. One advantage of the quantile approach is that it simultaneously estimates the relative truncation, shift, and dilation parameters between the two distributions, which allow us to distinguish agglomeration from firm selection.

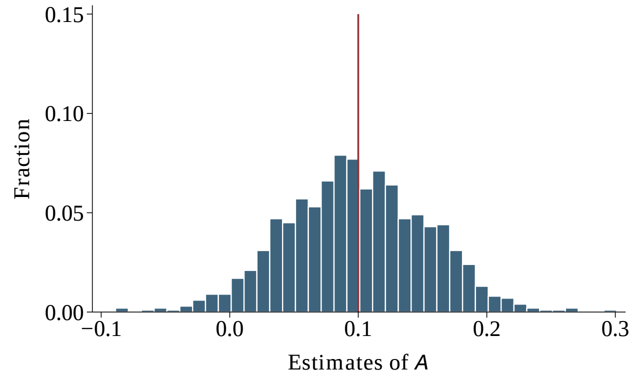
5 Applied example

In this section, we replicate the estimation results of Combes et al. (2012a) for all of the sectors using the `estquant` command.[9] Combes et al. (2012a) examined agglomeration and firm selection using the French establishment-level data. Their findings showed that higher productivity in larger cities could

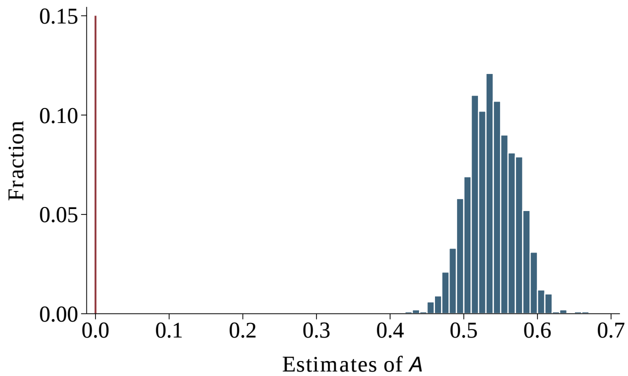
[9]The replication dataset of Combes et al. (2012a) is available from the following website: (URL: `http://www.econometricsociety.org/`).



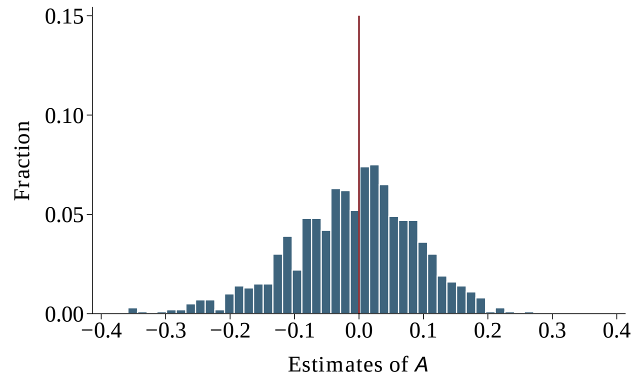
(a) Model: shift + dilation
(DGP$_1$: shift + dilation)



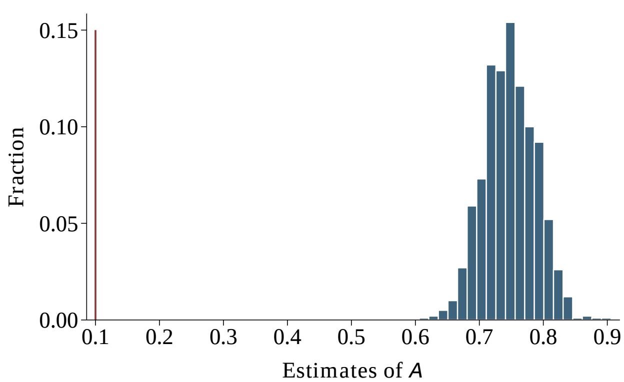
(b) Model: shift + dilation + truncation
(DGP$_1$: shift + dilation)



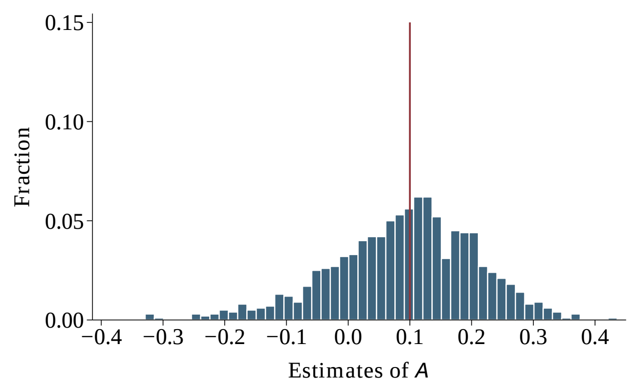
(c) Model: shift + dilation
(DGP$_2$: Truncation)



(d) Model: shift + dilation + truncation
(DGP$_2$: Truncation)



(e) Model: shift + dilation
(DGP$_3$: shift + dilation + truncation)



(f) Model: shift + dilation + truncation
(DGP$_3$: shift + dilation + truncation)

Figure 3: Distributions of estimated parameter $\hat{A}$ and model misspecification

Note: The vertical line indicates the true value of shift parameter ($A = 0.1$ or $A = 0$). A total of 1,000 replications were conducted in the Monte Carlo experiments. Details of the parameter setting and the DGPs are explained in Table 1.

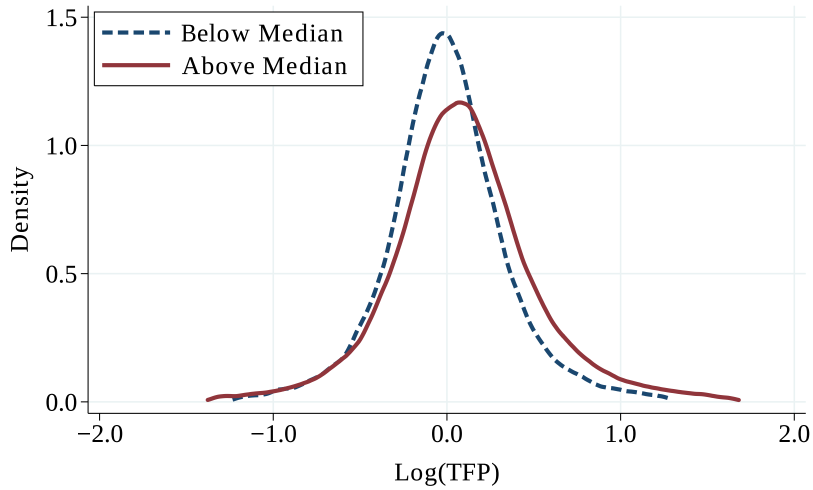


Figure 4: Distributions of TFP between cities with below- and above-median employment densities

Note: Created by the author using the dataset of Combes et al. (2012a). The solid and dashed lines correspond to the cities with above- and below-median employment densities, respectively.

not be explained by firm selection. In other words, firms in larger cities benefit from agglomeration economies, and as a result, there is a positive relationship between productivity and city size.

Before starting the replication, we show two productivity distributions between cities above and below median employment densities in Figure 4. Figure 4(a) indicates that the productivity distribution for larger cities is more right-shifted and flatter, relative to that of smaller cities. Equivalently, this relationship corresponds to the upper shifting quantile distribution with a steeper slope in Figure 4(b).[10]

Table 3 presents the estimation results replicated by the `estquant` command. Although the difference of convergence tolerance generates slight differences in point estimates between Combes et al. (2012a) and the present study, both estimation results are practically similar.[11] As discussed earlier, dropping the truncation from the model specification generates an upward bias in shift parameter A and a downward bias in dilation parameter D. However, the firm selection in lager cities is too weak to explain the higher productivity in such locations. In fact, the positive estimate $\hat{A}$ shows that agglomeration economies, on average, increase firm productivity.

[10]The uppermost and lowermost 0.5 percentile of productivity distributions by category are dropped from the estimation. The truncation of distributions may be detected due to arbitrary data processing or original data structure. In empirical analyses, researchers often conduct data cleaning beforehand, but careless data cleaning (e.g., dropping outliers using arbitrary thresholds) may lead to the truncation of distributions. When researchers use the value added to estimate firm productivities, firms with negative value added are dropped by taking its logarithm, which may also lead to the truncation of distributions (i.e., less-productive firms also exist!). Note that the selection parameter is quite sensitive to such data processing.

[11]One important difference is that Combes et al. (2012a) re-estimated TFP for each bootstrap sample, whereas the present study does not due to data restriction. Unlike the default setting of bootstrap sampling, the bootstrap sampling from the entire sample is conducted in this case.

Table 3: Estimation results replicated by the `estquant` command

	$\hat{A}$	$\hat{D}$	$\hat{S}$	R^2	Obs.
Model Specification	(1)	(2)	(3)	(4)	(5)
Shift	0.0925*			0.5963	13,4275
	(0.0025)				
Shift + Dilation	0.0925*	1.2195*		0.9963	13,4275
	(0.0025)	(0.0077)			
Shift + Truncation	0.1146*		−0.0195*	0.7214	13,4275
	(0.0026)		(0.0021)		
Truncation			−0.0013+	0.0025	13,4275
			(0.0006)		
Shift + Dilation + Truncation	0.0906*	1.2285*	0.0013	0.9974	13,4275
	(0.0023)	(0.0086)	(0.0008)		

Note: Replication of Combes et al. (2012a) for all of the sectors. The bootstrap standard errors are in parentheses, and bootstrap sampling with replacement was conducted 100 times. The same sample size is used for each bootstrap sampling. The parameters not shown in the model specifications are constrained as $A = 0$, $D = 1$, and $S = 0$. * (+) denotes that $\hat{A}$ and $\hat{S}$ are significantly different from 0 at the 5% (10%) level, and $\hat{D}$ is significantly different from 1 at the 5% (10%) level.

In sum, the `estquant` command helps researches apply a new quantile approach suggested by Combes et al. (2012a). It is important to distinguish agglomeration from firm selection in empirical studies. When stronger selection exists in larger cities, it is possible to draw a misleading conclusion that agglomeration economies, on average, increase firm productivity. A new quantile approach enables researchers to simultaneously examine the effects of agglomeration economies and firm selection.

6 Concluding remarks

To answer the question of why productivity is, on average, higher in larger cities, a deeper understanding of the background mechanism is required. One key channel is that agglomeration economies, on average, increase local productivity. Another is that a higher concentration of economic activities in larger cities promotes tougher competition, which forces less-productive firms to exit the market in such locations. If we omit selection in larger cities, it is possible to overestimate the agglomeration effects on productivity and draw misleading policy implications. Thus, the new quantile approach suggested by Combes et al. (2012a) resolves this empirical issue by simultaneously estimating the effects of agglomeration and firm selection. In addition, the newly developed command `estquant` enables researchers to easily implement their quantile approach in Stata.

Furthermore, this quantile approach offers greater possibilities in applied economics exhibiting a selection mechanism. For example, Combes et al. (2012b) applied the quantile approach to evaluate wage and skill distributions in terms of several characteristics of workers (e.g., stayers and migrants). The `estquant` command can be used to examine selection in the wide-ranging field of economics.

References

Accetturo, A., V. D. Giacinto, G. Micucci, and M. Pagnini. 2013. Geography, productivity and trade: Does selection explain why some locations are more productive than others? Temi di Discussione (Working Papers) No. 910, Bank of Italy.

Arimoto, Y., K. Nakajima, and T. Okazaki. 2014. Sources of productivity improvement in industrial clusters: The case of the prewar Japanese silk-reeling industry. *Regional Science and Urban Economics* 46: 27–41.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux. 2012a. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica* 80(6): 2543–2594.

Combes, P.-P., G. Duranton, L. Gobillon, and S. Roux. 2012b. Sorting and local wage and skill distributions in France. *Regional Science and Urban Economics* 42(6): 913–930.

de la Roca, J., and D. Puga. 2017. Learning by working in big cities. *Review of Economic Studies* 84(1): 106–142.

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

Marshall, A. 1890. *Principles of Economics*. London: Macmillan.

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

Melitz, M. J., and G. I. P. Ottaviano. 2008. Market size, trade, and productivity. *Review of Economic Studies* 75(1): 295–316.

Rosenthal, S. S., and W. C. Strange. 2004. Evidence on the nature and sources of agglomeration economies. In *Handbook of Regional and Urban Economics*, ed. J. V. Henderson and J.-F. Thisse, vol. 4, chap. 49, 2119–2171. Amsterdam: Elsevier.

Syverson, C. 2004. Market structure and productivity: A concrete example. *Journal of Political Economy* 112(6): 1181–1222.