DEA AND SOCIAL POLICY
– A PERFORMANCE EVALUATION OF JAPANESE LOCAL WELFARE OFFICES–

Masayoshi Hayashi
Graduate School of Economics/Faculty of Economics, University of Tokyo
7−3−1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
hayashim@e.u-tokyo.ac.jp

Abstract: We examine the efficiency of social welfare offices in Japan and explore the external factors that affect their efficiency. We elaborate on the influence of population, surface area, and fiscal capacity and explore the implications of their effects.

Keyword: Social welfare office, public assistance, DEA, Japan

1. INTRODUCTION

Like most OECD countries, Japan is experiencing substantial changes in its socio-economic structure due to the growing number of low-wage workers and the rapid pace of population ageing. While this necessitates a series of reforms in public programs targeted at such disadvantaged households, public funds for such programs are limited, restricting the scope of possible reforms. Given this lack of resources, it is important to achieve higher efficiency in welfare program implementation. We would then naturally be interested in examining the performance of welfare organization and exploring factors that may affect the efficient implementation of its programs. Data Envelopment Analysis (DEA) is one of the standard tools used for examining efficiency. While a number of DEA studies examine public sector activities, those addressing social assistance are limited (Martin 2002, Ayala et al. 2008, Enache 2012, Habibov and Fan 2010, Broersma et al. 2013). This may partly be due to the traditional reluctance to conduct economic evaluation as part of social policy (Ayala et al. 2008). In addition, DEA studies on social assistance address varied concerns. Some are interested in the efficiency of welfare offices providing social assistance to a given size of welfare recipients (Martin 2002, Ayala et al. 2008), while others focus on the efficiency in social expenditure in reducing poverty level (Enache 2010, Habibov and Fan 2012). The efficiency of social assistance spending across different socio-economic environments has also been studied (Broersma et al. 2013). In this study, we examine the efficiency of welfare offices, not the efficiency of welfare spending. This focus means that the studies of Martin (2002) and Ayala et al. (2008) are of direct relevance to our analysis. Martin (2002) may be among the first to have applied DEA to welfare programs, using the data from social assistance offices in Oregon. In the same vein, Ayala et al. (2008) evaluate the efficiency of 41 social services agencies in Madrid, Spain.

We aim to improve on these studies by investigating the efficiency of social welfare offices in the Japanese system of local public administration. The Japanese case indeed merits analysis. First, no studies have utilized DEA to investigate efficiency issues in the Japanese social assistance program. Second, data are available for the Japanese case, pertaining to caseloads by category for multiple output variables and public employment by type for input variables, both at the municipal level. We also take advantage of this availability to conduct a second-stage regression (2SR) analysis with the efficiency score for social welfare offices, to examine the factors influencing the efficiency of municipal programs.

This paper is organized as follows. Section 2 describes the Japanese system of social assistance and activities of its social welfare offices to set up an input–output model that yields

DEA efficiency scores. It then calculates relevant efficiency scores and conducts an analysis. Section 3 then explores the effects of external factors on the efficiency scores. For this purpose, it utilizes the 2SR analysis. In so doing, it elaborates on the issues concerning the 2SR and employ several estimation methods proposed in the literature to obtain a set of estimates and compare the results. Section 4 extends the analysis in Section 3 to perform quantile regressions. It does so with an anticipation that the external factors exert different impacts on the efficiency, depending on the level of the latter. Section 5 concludes this study.

2. EFFICIENCY SCORES

A DEA study starts by specifying decision-making units (DMUs) and variables for inputs and outputs. This section specifies the DMUs and the variables for inputs and outputs for the current analysis. Finding the relevant inputs and outputs for welfare office activities requires an examination of the nature of the actual system. It therefore elaborates on our DEA model, while discussing the institutional mechanism of the Japanese system of social assistance.

2.1. DMUs

Public Assistance (PA), *Seikatsu Hogo* in Japanese, acts as the last social safety net covering those excluded from the upper layers of social programs and is implemented by local governments in the country. The Public Assistance Law (PAL) allows the Ministry of Health, Labour and Welfare (MHLW) to mandate local governments to implement PA programs. There are two levels of local government in Japan: prefectures and municipalities (cities, towns, villages, and Tokyo Metropolitan special wards (TMSWs)). The Social Welfare Law (SWL) requires cities, TMSWs, and prefectures to set up social welfare offices, through which they implement social programs including PA. The SWL does not require towns and villages to do so. In towns and villages that do not have their own welfare offices, prefectural welfare offices cover their respective population. We thus use 658 cities in 2010 as our DMUs, excluding villages and towns that have their own social welfare offices. Note that we could not utilize all those cities that have their own welfare offices, since some of them lack all the necessary data required for our analysis. Notice also that since a small number of cities have more than one social welfare office, the number of DMUs is not necessarily identical to the number of individual welfare offices. Nonetheless, we do not consider this a major problem, as municipalities, not individual welfare offices, make decisions on human resource allocation concerning welfare offices.

2.2. Outputs and Inputs

A function of social welfare offices is to provide PA for those who are in need of it. The PA intends to guarantee the minimum cost of living for Japanese citizens. Through the PAL, the central government sets uniform procedures for localities to follow when they provide PA benefits. That is, local governments do *not* set the eligibility standard or the benefit levels for their PA programs. The PA benefits are equal to the minimum cost of living in excess of what an individual earns with his/her best effort. The MHLW determines the minimum costs of living, allowing for differences in the cost due to regional price differences, the formula for which uniformly applies across the nation. To receive the benefits, applicants are supposed to exhaust their available resources. The PA program therefore requires local welfare offices to conduct a careful examination, or a "means test," of the financial situation of the applicants.

It is then natural to employ welfare caseloads (the number of recipients obtaining assistance) as our choice of the output variable. Indeed, Martin (2002) and Ayala et al. (2008) make analogous choices. However, they also use other variables. In addition to caseloads, Martin (2002) uses the number of job placements, successful exits (the number of recipients who have been off assistance during the last 18 months at least), and child support benefits. Meanwhile, Ayala et al. (2008)

additionally use the median length of processing applications. Our omission of these additional outputs may be justifiable in the Japanese institutional context. First, Japanese welfare offices do not implement active labor market programs that are comparable to those in other OECD countries. Second, child support benefits are irrelevant for the Japanese case since another municipal branch is responsible for them. Third, processing time may be reflected in the caseload size, since shorter processing time leads to a larger caseload size in a given time period. As we mentioned above, the activities of Japanese welfare offices center on means testing, delivery of benefits, and monitoring of the recipients.

While we focus on welfare caseloads as the output in the current study, a single type of caseload may not suffice. Since the needs of the PA recipients vary depending on their characteristics, the services required for different categories of PA recipients must also be different. Fortunately, we have access to disaggregated caseload data for five categories of recipient households: those made only of (i) the elderly (those consisting only of those aged 65 years and above), and those headed by (ii) single mothers, (iii) the handicapped, (iv) the sick and injured, and (v) others. The categorization is lexicographic, starting from (i) and proceeding to (v). In FY2010, elderly households constituted the largest proportion (43%), followed by households headed by the injured and sick (23%) and the disabled (11%). The remaining consist of households headed by single mothers (8%) and others (16%). These categories of recipients apparently require different types of casework. We therefore use their caseloads as different multiple (five) outputs in our analysis.

A natural choice for the inputs is the size of employment at welfare offices, since "labor" is a straightforward input in this production process. Indeed, Ayala et al. (2008) and analogous DEA studies on public employment offices (Sheldon 2003, Vassiliev et al. 2006, Althin et al. 2010) use the sizes of office staff by type as prime inputs. Our data allow us to differentiate them into caseworkers and administrative staff. It is also natural to consider "capital" type production inputs. For example, Martin (2002) considers the number of offices, while Althin et al. (2010) employ office space. However, the data for office space are not available for the current case. We thus have to content ourselves with the use of single types of input, i.e., labor. This may not be a serious problem though, since labor is more relevant than capital for the analysis of welfare programs, given the labor-intensive character of social services (Ayala et al. 2008)

2.3. Results

Given the nature of the PA system in Japan, the size of the needy is largely exogenous for municipalities. Although it might be possible for welfare offices to implement programs to reduce these needs, such changes are indeed slim. Thus, the efficiency concept in this analysis concerns how efficiently welfare offices manage a given level of caseloads without reducing the services for the recipients. We thus employ the input-oriented efficiency score E. Formally, the score E is the maximal contraction of all inputs $\boldsymbol{x}$ ((i) caseworkers and (ii) other staff) that allows us to produce a given combination of outputs $\boldsymbol{y}$ (caseloads for (i) the elderly, (ii) single mothers, (iii) the handicapped, (iv) the sick and injured, and (v) others). That is, $E \equiv \min\{E > 0 \mid (E\boldsymbol{x}, \boldsymbol{y}) \in T\}$ where T is a technology set. The measurement thus utilizes the Charnes–Cooper–Rhodes (CCR) model (Charnes et al. 1978) to obtain the score E based on the Farrell index of input efficiency.

A linear programing exercise yields three types of efficiency scores (E_{VRS}, E_{CRS}, E_{DRS}), being based respectively on the concepts of variable returns to scale (VRS), constant returns to scale (CRS), and decreasing returns to scale (DRS). **Figure 1** shows their kernel distributions. There are noticeable differences between the scores based on variable returns to scales (E_{VRS}) and those based on the other two types of returns (E_{CRS}, E_{DRS}). As expected, E_{VRS} tend to yield higher efficiency scores. The distribution of E_{VRS} has an average of 0.459 with a standard deviation (s.d.) of 0.191 and ranges from 0.133 to unity. Meanwhile, the distribution of E_{CRS} (E_{DRS}) has an average of 0.376 (0.387) with an s.d. of 0.183 (0.196) and ranges from 0.064 (0.064) to unity.

We perform the bootstrapped tests for returns-to-scale as indicated by Simar and Wilson (2002, 2011), using the (i) ratio of means, (ii) mean of ratios, and (iii) mean of ratios of DEA scores less unity (the number of replications is 3,000). The tests reject both the null hypotheses of constant returns and decreasing returns to scale at the standard levels of statistical significance. In the next analysis, we therefore proceed with E_{VRS}, the scores based on variable returns to scale. There are indeed large differences in the efficiency score among municipalities. Only 29 municipalities (4.4%) are located on the frontier with a full score of unity, while the other 629 municipalities (95.6%) are off the frontier. As such, **Figure 1** suggests sizable room for efficiency improvement.

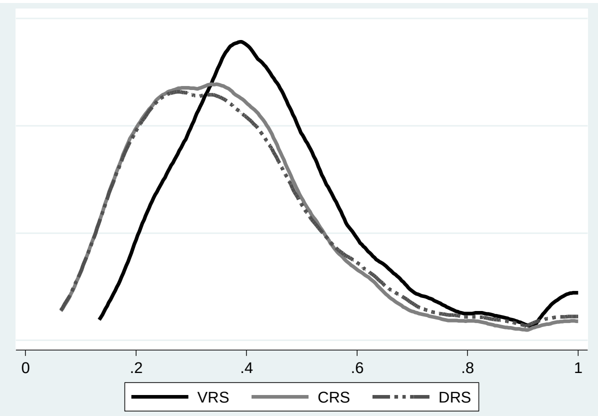


Figure 1: Distributions of efficient scores

3. EXTERNAL EFFECTS ON EFFICIENCY

3.1. Adjustments for Environmental/External Factors

In the previous suction, we noted the large variations in the efficiency scores among municipalities. We might then rank individual DMUs according to the efficiency scores so that the less efficient DMUs could use them as a benchmark to improve their efficiency. However, doing so may not be appropriate since Japanese social welfare offices apparently operate in non-homogeneous environments, which are likely to exert different impacts on the performance of the DMUs. In other words, the “inefficiency” may originate from factors outside the control of the DMUs, thus limiting the scope of efficiency improvement that **Figure 1** may otherwise indicate. The ranking of the DMUs would then require adjustments of the efficiency scores that allow for different external (environmental) factors.[1]

While the evaluation of individual DMUs is the primary objective of DEA, it may also be meaningful to examine if and how external factors affect the efficiency. Indeed, it is important to examine factors that affect PA implementation at social welfare offices and to find the directions and degrees of their impacts on their efficiency. For example, if the central government can change such external factors, it could improve the overall efficiency of municipal PA programs. For this reason, Section 3-4 discusses two-stage models but leaves the adjustment of efficiency scores to future studies.

3.2. The Second-stage Regression Model

We specify our 2SR model as a linear-in-parameters form.

$$E_i = \boldsymbol{z}_i \cdot \boldsymbol{\beta} + u_i \qquad (1)$$

where E_i is an efficiency measure obtained in the first stage, $\boldsymbol{z}_i$ is a vector of external factors, $\boldsymbol{\beta}$ is a vector of coefficients, u_i is an error term, and i indexes the DMUs. For the efficiency measure (i.e., dependent variable), we alternatively use the efficiency score (E_{VRS}) and its reciprocal (distance function $1/E_{VRS}$). For the factors in $\boldsymbol{z}$, we consider municipal population (in log), surface area (in log), the Fiscal Capacity Index (FCI), the Obligatory Expenses Ratio (OER), the Local Allocation Tax (LAT) received (as the binary variable), and caseload

[1] There are three approaches to such adjustments (Cordero et al., 2009). First, one-stage models regard external factors as additional inputs in the standard DEA model but obtain the efficiency scores with special restrictions on them (Banker and Morey 1986, Ruggiero 1996). Second, two-stage models obtain the efficiency scores without external factors in the first stage and regress the scores on a set of external factors in the second stage (Ray 1991). They then use the second-stage estimates to adjust the efficiency scores. Third, adjusted values models utilize the estimates for the effects on slacks of external factors to adjust the values of discretionary variables and obtain the DEA efficiency scores with these slack-adjusted values (Muñiz 2002, Fried et al. 1999).

growth rate. The variables are defined later in this section. We select these six factors for the following reasons.

First, as the literature on local public finance shows, there exist economies of scale in local expenditures up to a certain level of the population (Duncombe and Yinger 1993). Larger localities tend to provide more categories of services than smaller ones (Oates 1988), yielding economies of *scope* in more populous municipalities. Using the savings from these two scale economies, localities with a higher population could invest more resources so that welfare offices may become more efficient. These lines of reasoning suggest that a larger population would lead to higher efficiencies.

Second, the space of a given locality should also matter. Since caseworkers visit their PA recipients within a given period, the more widespread the residences of the recipients, the more the time spent on their cases. Since more time spent on a case implies that the caseworker will handle a fewer number of cases, spacious jurisdictions denote less efficiency.

Third, fiscal capacity is also a concern. The Japanese government estimates an index for the "fiscal capacity" of localities with the data obtained from its system of central grants (the LAT). The amount of LAT a locality receives is the non-negative difference between its Standard Fiscal Demand (SFD) and Standard Fiscal Revenue (SFR). The SFD estimates the level of local expenditures required to maintain the standard level of public services, while the SFR estimates standardized local revenues. The system defines the FCI as the three-year average of the ratio of the SFR to the SFD. Another index of fiscal capacity is the OER, which shows a percentage of expenses that a locality cannot easily adjust, including personnel expenses, local debt-service payments, and other expenses the central government requires them to spend. Larger fiscal capacity implies more fiscal abundance, from which localities could spare more resources for providing welfare offices with more caseworkers and administrative staff for a given level of caseloads. This may or may not imply that a larger fiscal capacity would lead to lower efficiency.

However, a drawback in using the FCI as fiscal capacity is its negative correlation with the LAT grant a locality receives. The LAT may adversely affect the efficiency of local spending (e.g., Otsuka et al. 2014), which then suggests that a large value of FCI may be associated with a smaller value of the efficiency score. To control the effect of receiving transfers, the regression below allows for the receipt of LAT grants. In addition to its claimed adverse effects, the LAT compensates the local burden of PA expenditures. While the central government disburses 75% of local PA expenditure with matching grants to localities, the SFD allows for the rest of the cost. In other words, while LAT recipients enjoy the increase in PA benefits covered by the central grants, the non-recipients have to meet 25% of that increase out of their own pockets. This would then imply that receiving LAT grants adversely affects the efficiency score.

Lastly, the speed of caseload changes affects the efficiency. Roughly speaking, efficiency is the ratio of output over input. Since inputs are the numbers of caseworkers and other staff members, the efficiency tends to increase if the adjustments of the inputs are slow relative to the changes in the outputs. The analysis seen below allows for this aspect by including the rate of increase in PA caseloads from FY2008 to FY2009 as the measured input at the beginning of FY2010.

3.3. Econometric Issues

It is important when estimating 2SR models to recognize that DEA scores obtained in the first stage are *estimates* (Simar and Wilson 2007). We can frame the issue as a typical case of measurement errors in the dependent variables (Wooldridge 2010, 76–77). The 2SR typically assumes the following data generation process (DGP).

$$E_i^* = \mathbf{z}_i \cdot \boldsymbol{\beta} + \varepsilon_i \tag{2}$$

where E_i^* is the *true* value of the efficiency score. Since E_i^* is not observable, the estimated score E_i surrogates E_i^*. Defining the measurement error in the dependent variable as $s_i \equiv E_i - E_i^*$, we can express Equation (2) as Equation (1) with $u_i \equiv s_i +$

ε_i. If E_i is consistent, s_i approaches zero (i.e., E_i approaches E_i^*) as the sample size approaches infinity. Since E_i is indeed consistent (Banker 1993, Kneip et al. 1998), the existence of s_i does not affect the *asymptotic* distribution of estimators for β. In other words, treating DEA scores as *estimates* does not pose an issue if we have a suitably large sample.

On the other hand, we are not sure how large a suitable sample would be, since the convergence of E_i to E_i^* becomes slower as the number of inputs and outputs in the DEA model increases (Kneip et al. 1998). This "curse of dimensionality" might make the asymptotic approximation fail even with a relatively large sample. In addition, the very calculation of DEA scores for individual DMUs creates correlations among them (Xue and Harker 1999).[2] The curse of dimensionality and the correlation among the scores makes u_i non-spherical though s_i is a finite sample, even if ε_i is spherical (i.e., i.i.d.). Furthermore, it is very likely that ε_i is also non-spherical since we typically use a sample of the cross-sectional data. We thus adjust the covariance matrix of β estimates to arrive at a valid inference. As the pattern of non-spherical error u is unknown, our choices include utilizing a heteroskedastic consistent covariance matrix estimator (McDonald 2009) or bootstrapping the covariance matrix (Simar and Wilson 2007).

Another econometric issue concerns the method to estimate Equation (1). The most straightforward is the ordinary least squares (OLS) estimator (Ray 1991), which does not explicitly allow for the fact that the dependent variable is bounded: $E_{VRS} \in (0, 1]$ or $1/E_{VRS} \in [1, \infty)$. To allow for these bounds, Bjurek et al. (1992) estimate the 2SR model with censoring at unity (Tobit estimator). While a number of studies use the method, the Tobit estimator also has its shortcomings when applied to the 2SR (McDonald 2009). Therefore, Simar and Wilson (2007) model it as a linear model with truncation, whereas Hoff (2007) and Ramalho et al. (2010) depart from the linear specification to utilize the fractional response (FR) model devised by Papke and Wooldrige (1996).

3.4. Estimation Results

To estimate Equation (1), we employ all the four models: (i) OLS, (ii) Tobit, (iii) truncation, and (iv) FR models.[3] The FR model is not applicable to cases that employ $1/E_{VRS}$ as their dependent variable. For the covariance matrices of these estimators, we bootstrap the standard errors with 3,000 replications to allow for their possible inconsistency and finite sample bias. In addition, we utilize the double bootstrap procedure (Algorithm #2) by Simar and Wilson (2007), which essentially replaces the original E_{VRS} with bias-corrected bootstrapped E_{VRS} when bootstrapping the truncated regression of Equation (1). The number of replications for the bias-corrected E_{VRS} and 2SR is 200 and 3,000 respectively. Note that, since this method is not computationally applicable to the efficiency score with bounds (0, 1] (Simar and Wilson 2008, Besstremyannaya and Simm 2015), we apply it only to the case with the distance function.

Table 1 lists the results. The first four columns are cases with the efficiency score and the last four are cases with its reciprocal (distance function). These results are robust in the sense that the statistical significances do not change over different estimation methods for a given dependent variable (E_{VRS} or $1/E_{VRS}$). In all the cases, population, surface area, the FCI, and the LAT received are all statistically significant, while the OER and caseload changes are not. The results show that larger population, smaller surface area, lower fiscal capacity, and the non-receipt of LAT grants tend to increase the efficiency as expected, whereas the other two variables do not affect the efficiency. In particular, the insignificance on the caseload changes imply that the inputs (caseworkers and other staff members) adjust smoothly against the changes in the outputs (PA caseloads).

Furthermore, the marginal effects of these external factors

2 Perturbations of DMUs lying on the estimated frontier change the scores of some other DMUs.

3 We use the logistic distribution for the cumulative distribution that shapes the fractional response.

are almost similar among the E-OLS, E-Tobit, and E-FR and between the D-OLS and D-Tobit models (where E and D denote efficiency and distance respectively). On the other hand, the differences in values with truncated regression and those from other models are rather conspicuous, albeit not so large, with statistically significant coefficients. These differences are likely due to different samples rather than the different DGP or estimation method (truncated or not), as E-TRC and D-TRC exclude 29 DMUs that have full efficiency scores of unity. In addition, while D-TRC/SW does not exclude them, it uses bootstrapped bias-corrected scores whose values are not only different from the standard ones but also differ from unity.

McDonald (2009) argues that the idea of efficiency scores as estimates of "true" scores, as suggested by Simar and Wilson (2007), "would lead to considerable complexity and perhaps only minor changes in inference." Qualitatively, his argument seems to apply to our cases. All the estimation methods, including OLS, show that larger population, smaller surface area, lower FCI, and non-receipt of LAT grants would increase the efficiency, whereas the others do not affect the efficiency. Quantitatively, however, the results are somewhat different. In particular, the effects of population, FCI, and LAT receipt differ with the truncation regressions, although the effect of surface area does not change much.

Table 1: Estimation results

Dependent variable	Efficiency score (E_{VRS})				Distance function ($1/E_{VRS}$)			
Models	E-OLS	E-Tobit	E-FR	E-TC	D-OLS	D-Tobit	D-TC	D-TC/SW
ln(population)	.075*** (.014)	.078*** (.014)	.075*** (.014)	.061*** (.011)	−.269*** (.062)	−.294*** (.067)	−.368*** (.104)	−.362*** (.113)
ln(area)	−.027*** (.009)	−.028*** (.009)	−.027*** (.009)	−.026*** (.007)	.143*** (.043)	.149*** (.046)	.229*** (.071)	.246*** (.083)
Fiscal capacity index (FCI)	−.264*** (.056)	−.270*** (.057)	−.263*** (.056)	−.246*** (.043)	1.267*** (.294)	1.320*** (.301)	1.929*** (.448)	2.070*** (.501)
Obligatory expense ratio (OER)	.143 (.166)	.149 (.173)	.143 (.170)	.109 (.145)	−1.168 (.947)	−1.222 (.979)	−1.751 (1.431)	−1.967 (1.659)
LAT receipt	−.140*** (.049)	−.145*** (.051)	−.140*** (.049)	−.108*** (.064)	.561** (.274)	.602** (.285)	.750* (.436)	1.000** (.506)
Caseload growth	−.021 (.095)	−.016 (.099)	−.022 (.096)	−.064 (.088)	.046 (.472)	.082 (.491)	−.082 (.691)	.098 (.794)
Constant	−.075 (.171)	−.099 (.181)	–	−.039 (.158)	4.538*** (.895)	4.729*** (.949)	4.852*** (1.438)	4.904*** (1.609)

Notes: (1) The sample sizes are 658 for E-OLS, E-Tobit, E-FR, D-OLS, and D-Tobit. (2) Truncation regressions (E-TRC and D-TRC) exclude DMUs with E_{VRS} ($1/E_{VRS}$) = 1, trimming the sample size down to 629. (3) ***, **, and * indicate $p \leq .01$, $.01 < p \leq .05$, and $.05 < p \leq .10$ respectively. (4) Bootstrapped standard errors are in parentheses (3,000 replications). (5) D-TC/SW utilizes the r-DEA package (Simm and Besstremyannaya 2015) to obtain the bootstrapped bias-corrected efficiency scores as suggested by Simar and Wilson (2007) with 200 replications, and then uses these scores for the bootstrapped truncated regression with 3,000 replications. (6) E-FR lists the marginal effects evaluated at the sample averages.

4. QUANTILE REGRESSIONS ANALYSIS

4.1. Different Responses along the Quantiles of Efficiency

All the preceding models, except the FR model, assumed that the marginal effects of the external factors on efficiency are constant on average. Such effects may plausibly differ among DMUs with different levels of efficiency. This section therefore conducts the 2SR using quantile regression (QR) to address the possible different responses of the efficiency scores to the external factors.

With QR, we can estimate the responses of the efficiency score to changes in the external factors across the conditional quantiles of the former. For the QR analysis, we first define a conditional quantile function of E as $Q_\tau(E_i \mid \boldsymbol{z}_i) \equiv F^{-1}(\tau \mid \boldsymbol{z}_i)$, where $F(\tau \mid \boldsymbol{z}_i)$ is a c.d.f. of E at quantile τ, conditioned on a given set of external factors $\boldsymbol{z}_i$. We then specify a linear regression model as $E = \boldsymbol{z}\cdot\boldsymbol{\beta}_\tau + u$, where $\boldsymbol{\beta}_\tau$ is a vector of coefficients that vary across quartiles, and u_i is the error term. The QR estimator for $\boldsymbol{\beta}_\tau$ is a sample analogue of $\boldsymbol{b}_\tau \equiv \mathrm{argmim}_{\boldsymbol{b}}$ $\mathrm{E}\{\rho_\tau(E = \boldsymbol{z}\cdot\boldsymbol{b})\}$, where ρ_τ is the check function defined as $\rho_\tau \equiv$ $1[E - \boldsymbol{z}\cdot\boldsymbol{b} > 0]\cdot\tau\cdot\mid E - \boldsymbol{z}\cdot\boldsymbol{b} \mid + 1[E - \boldsymbol{z}\cdot\boldsymbol{b} \leq 0]\cdot(1-\tau)\cdot\mid E - \boldsymbol{z}\cdot\boldsymbol{b} \mid$. This scheme results in a minimand that selects conditional quantiles.

We use the super-efficiency score as the dependent variable. The super efficiency score of a given DMU, K, is obtained by gauging it against another efficiency frontier calculated with a group of DMUs that *excludes K* (cf. Bogetoft and Otto 2011). It equals the standard efficiency score for DMUs that are *off* the frontier, and takes on a value more than unity for DMUs that are *on* the frontier. Since the score can take on a value exceeding unity, it is convenient for us to use them as the dependent variable of our QR model. Notice, however, we may not be able to calculate the super-efficiency scores for all DMUs. As we could not obtain the score for one DMU, the sample size for the second-stage QR is 657.

4.2. Results

Table 2 lists the results of the QRs at the 0.15, 0.25, 0.50, 0.75, and 0.85 quantiles. As a benchmark, we list the OLS estimates obtained with super-efficiency scores. The results of the OLS model show that the directional impacts of the external factors are qualitatively the same as those of the E-OLS model in **Table 1**. However, their magnitudes (in absolute value) are larger, reflecting the changes in the efficiency scores from unity for the DMUs on the frontier. The coefficient estimates for the five quantiles are indeed different from those of the OLS models. In addition, their coefficient values vary across the quantiles. To better understand the changes in coefficients across quantiles, **Figure 2** plots the coefficient estimates at eighteen quantiles (.05, .10, …, .90, and .95) along with their 95% confidence intervals. In general, the effects tend to increase in absolute value toward the upper quantiles if they are largely statistically significant (population, surface area, FCI, and LAT), while they seem change relatively little if the effects are not statistically significant (OER and caseload growth). These findings suggest that the impacts on efficiency are larger for those municipalities located closer to the edge of the frontier.

5. CONCLUDING REMARKS

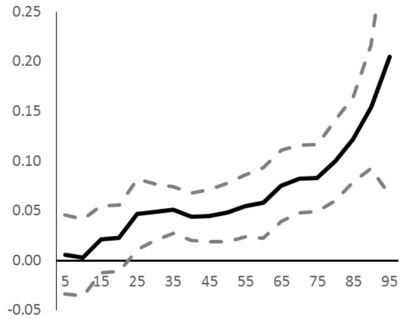
In this study, we obtained the Farrell scores of input-oriented efficiency for municipal PA programs in Japan and explored the effects of a set of external factors on the efficiency scores. We showed that the efficiency varies across municipalities, implying a large potential for efficiency improvement. Nonetheless, such disparities may be due to the variations in external factors that municipalities cannot control. Employing the 2SR analysis with a variety of estimators, we then examined how a set of external factors would affect the efficiency. Our results indicated that surface area, fiscal capacity, and receipt of LAT grants decreased the efficiency, while population improved it. Furthermore, a QR analysis with super-efficiency scores showed that the marginal effects of the external factors on the efficiency would become larger in absolute value for the upper quantiles of the efficiency scores. This then may imply that when we compare efficiency scores among DMUs, we should adjust the scores, taking account of their effects across the quantiles of the efficiency scores. While there is a large body of literature on the adjustment of the efficiency scores to variations in external factors (e.g., Cordero et al. 2009), to the best of our knowledge, no study explicitly allows for such differentiated effects across quantiles. Our next task then would be to elaborate further on the QR approach in a 2SR analysis and possibly construct adjusted efficiency scores that allow for such quantile effects.

Table 2: Estimation results

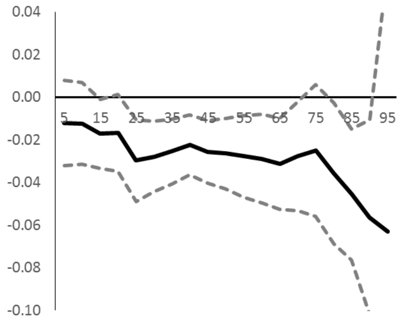
	OLS	Quantile				
		.15	.25	.50	.75	.85
ln(population)	.098*** (.021)	.022 (.017)	.047*** (.018)	.055*** (.016)	.083*** (.017)	.122*** (.022)
ln(area)	−.041*** (.013)	−.017** (.009)	−.030*** (.010)	−.028*** (.010)	−.025 (.016)	−.046*** (.016)
Fiscal capacity index (FCI)	−.361*** (.090)	−.157*** (.052)	−.268*** (.066)	−.194*** (.061)	−.285*** (.082)	−.477*** (.100)
Obligatory expense ratio (OER)	−.242 (.251)	.349* (.172)	.313* (.176)	.026 (.198)	.342 (.264)	.053 (.375)
LAT receipt	−.161*** (.055)	−.065 (.042)	−.097** (.043)	−.099 (.075)	−.256*** (.097)	−.274*** (.095)
Caseload growth	−.037 (.104)	−.039 (.155)	−.083 (.082)	−.076 (.089)	−.064 (.160)	.265 (.185)
Constant	−.252 (.295)	.010 (.230)	−.040 (.237)	.164 (.230)	−.144 (.246)	−.014 (.321)

Notes: (1) Sample size is 658. (2) Standard errors are bootstrapped with 3000 replications. (3) ***, **, and * indicate $p \leq .01$, $.01 < p \leq .05$, and $.05 < p \leq .10$ respectively.

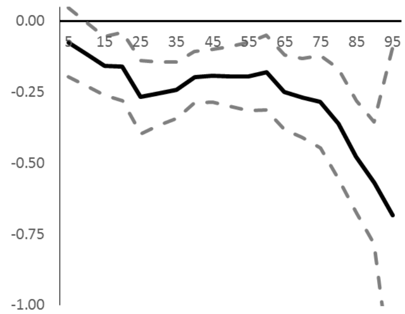
Figure 2 Distributions of efficient scores



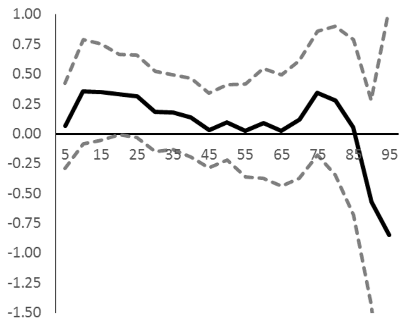
Panel A: Population



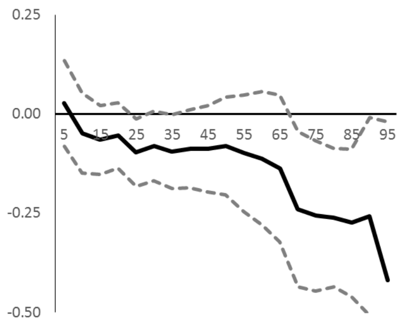
Panel B: Surface area



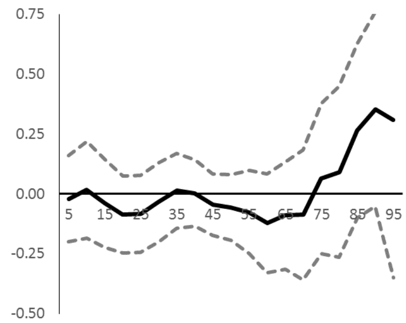
Panel C: Fiscal capacity index



Panel D: Obligatory expense ratio



Panel E: LAT receipt



Panel F: Caseload growth

Notes: (1) Solid lines in panels connect the QR coefficient estimates at eighteen quantiles (0.05, 0.10, …, 0.90, and 0.95). (2) The dotted lines show 95% confidence intervals of the QR estimates.

REFERENCES

[1]. Martin, L.L. (2002) Comparing the performance of multiple human service providers using Data Envelopment Analysis. *Administration in Social Work*, 26 (4), 45–60.

[2]. Ayala, L., Pedraja, F. and Salinas-Jiménez, J. (2008) Performance measurement of local welfare programmes: Evidence from Madrid's regional government. *Environment and Planning C*, 26 (5), 906–923.

[3]. Enache, C. (2012) The efficiency of expenditure-related redistributive policies in the European countries. *Timisoara Journal of Economics*, 5 (18), 380–394.

[4]. Habibov, N. and Fan, L. (2010) Comparing and contrasting poverty reduction performance of social welfare programs across jurisdictions in Canada using Data Envelopment Analysis (DEA): An exploratory study of the era of devolution. *Evaluation and Program Planning*, 33 (4), 45–60.

[5]. Broersma, L., Edes, A.J.E. and van Dijk, J. (2013) Have Dutch municipalities become more efficient in managing the costs of social assistance dependency? *Journal of Regional Science*, 53 (2), 274–291.

[6]. Sheldon, G.M. (2003) The efficiency of public employment services: A nonparametric matching function analysis for Switzerland. *Journal of Productivity Analysis*, 20 (1), 49–70.

[7]. Vassiliev, A., Luzzi, G.F., Flückiger, Y. and Ramirez, J.V. (2006) Unemployment and employment offices' efficiency: what can be done? *Socio-Economic Planning Sciences*, 40 (3), 169–186.

[8]. Althin, R., Behrenz, L., Färe, R., Grosskopf, S. and Mellander, E. (2010) Swedish employment offices: A new model for evaluating effectiveness. *European Journal of Operational Research*, 207 (3), 1535–1544.

[9]. Charnes, A., Cooper, W. W. and Rhodes, E. (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.

[10]. Bogetoft, P. and Otto, L. (2011) *Benchmarking with DEA, SFA, and R*. Springer, New York.

[11]. Simar, L. and Wilson, P.W. (2002) Non-parametric tests of returns to scale. *European Journal of Operational Research*, 139 (1), 115–132.

[12]. Simar, L. and Wilson, P.W. (2011) Inference by the *m* out of *n* bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36 (1), 33–53.

[13]. Cordero, J.M., Pedraja, F. and Santín, D. (2009) Alternative approaches to include exogenous variables in DEA measures: A comparison using Monte Carlo. *Computers & Operations Research*, 36 (10), 2699–2706.

[14]. Banker, R.D. and Morey, R.C. (1986) Efficiency analysis for exogenously fixed inputs and outputs. *Operational Research*, 34 (4), 513–521.

[15]. Ruggiero, J. (1996) On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, 90 (3), 553–565.

[16]. Ray, S.C. (1991) Resource use efficiency in public schools: A study of Connecticut data. *Management Science*, 37 (12), 1620–1628.

[17]. Muñiz, M. (2002) Separating managerial inefficiency and external conditions in data. *European Journal of Operation Research*, 143 (3), 625–643.

[18]. Fried, H., Schmidt, S. and Yaisawarng S. (1999) Incorporating the operating environment into a nonparametric measure of technical efficiency. *Journal of Productivity Analysis*, 12 (3), 249–267.

[19]. Duncombe, W.D. and Yinger, J. (1993) An analysis of returns to scale in public production with an application to fire protection. *Journal of Public Economics*, 52(1), 49–72.

[20]. Oates, W.E., (1988) On the measurement of congestion in the provision of local public goods. *Journal of Urban Economics*. 24 (1), 85–94.

[21]. Otsuka, A., Goto, M. and Sueyoshi, T. (2014) Cost-efficiency of Japanese local governments: effects of decentralization and regional integration. *Regional Studies, Regional Science*, 1 (1), 207–220.

[22]. Simar, L. and Wilson, P.W. (2007) Estimation and inference in two-stage, semi-parametric models of production process. *Journal of Econometrics*, 136 (1), 31–64.

[23]. Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd Ed., MIT Press, Cambridge, MA.

[24]. Banker, R.D. (1993) Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science*, 39 (10), 1265–1273.

[25]. Kneip, A., Park, B.U. and Simar, L. (1998) A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory*, 14 (6), 783–793.

[26]. Xue, M. and Harker, P.T. (1999) Overcoming the inherent dependency of DEA efficiency scores: A bootstrap approach, Working Paper Series 99-17, Financial Institutions Center, Wharton School, University of Pennsylvania.

[27]. McDonald, J. (2009) Using least squares and tobit in second stage DEA efficiency analyses. *European Journal of Operational Research*, 197 (2), 792–798.

[28]. Bjurek, H., Kjulin, U., Gustafsson, B., Efficiency, productivity and determinants of inefficiency at public day care centers in Sweden. Scandinavian Journal of Economics, 94(2), 173–187, (1992).

[29]. Hoff, A. (2007) Second stage DEA: Comparison of approaches for modelling the DEA score. *European Journal of Operational Research*, 181 (1), 425–435.

[30]. Ramalho, E.A., Ramalho, J.J.S. and Henriques, P.D. (2010) Fractional regression models for second stage DEA efficiency analysis. *Journal of Productivity Analysis*, 34 (3), 239–255.

[31]. Papke, L.E. and Wooldridge, J.M. (1996) Econometric methods for fractional response variables with an application to 401(K) plan participation rate. *Journal of Applied Econometrics*, 11 (6), 619–632.

[32]. Simar, L. and Wilson, P.W. (2008) Statistical inference in nonparametric frontier models: recent developments and perspectives, in *The Measurement of Productive Efficiency and Productivity Growth*, (eds Fried, H.O., Lovell, C.A.K. and Schmidt, S.S.) Oxford University Press, New York, pp. 421–521.

[33]. Besstremyannaya, G. and Simm, J. (2015) Robust non-parametric estimation of cost efficiency with an application to banking industry. CEFIR/NES Working Paper Series, No 217, Centre for Economic and Financial Research at New Economic School.

[34]. Simm, J., and Besstremyannaya, G. (2015) *rDEA: Robust Data Envelopment Analysis (DEA) for R*, https://cran.r-project.org/web/packages/rDEA/index.html (accessed 8 April 2016).

Acknowledgements

This study was financially supported by the Grant-in-Aid for Scientific Research (A-15070000033) of the Japan Society for the Promotion of Science.