Essays on total factor productivity (TFP)

by

Pontus Mattsson

Department of Economics and Statistics, Linnaeus University, 351 95 Växjö, Sweden.

E-mail: pontus.mattsson@lnu.se

Contents

Abstract

This thesis consists of two self-contained empirical essays. Essay I investigates the impact of labor subsidies on TFP, and profit per employee is included as a second outcome. Coarsened exact matching (CEM) is performed on the key variables. After matching, a difference-in-difference (DID) model is applied. The study shows that firms employing workers with wage subsidies experience negative and significant effects on both TFP and profit per employee. Heterogeneity is, however, observed; the only sector to show a deficit in both TFP and profit per employee is wholesale. During the second year with a subsidy, a negative impact can be observed on the profit per employee but not on TFP. The policy conclusion from the analysis is that subsidizing individuals from particular groups is necessary to induce firms to hire workers from these groups. However, the time period for which a single firm is subsidized should be considered.

Essay II (with Jonas Månsson from Linnaeus University and the Swedish National Audit Office (SNAO), Christian Andersson from SNAO and Fredrik Bonander from SNAO) measures TFP of the Swedish district courts by applying data envelopment analysis (DEA) to calculate the Malmquist productivity index for 48 Swedish district courts from 2012 to 2015. This study uses a fully decomposed Malmquist index. A bootstrapping approach is further applied to compute confidence intervals for each decomposed factor of TFP as well as for TFP. The study shows an average annual of TFP by 0.7%. However, a substantial variation between years is observed both with regards to the number of statistically significant courts below and above unity. The negative impact is mainly driven by pure technical regress. Large variations are also observed over time where the small courts have the largest volatility. The TFP change is positively correlated with the rate of change in the caseload. Two recommendations are: 1) that district courts with negative TFP growth could learn from those with positive TFP growth and 2) that a back-up force could be developed to enhance flexibility.

Keywords: conditional difference-in-difference (cDID); Data Envelopment Analysis (DEA); District courts; Malmquist index; subsidized employment; Total factor productivity (TFP); Törnqvist TFP index.

Acknowledgment

Being a Ph.D. student is challenging in many phases, and both mental and methodological support are crucial. The most important people in these aspects are the supervisors, Dr. Jonas Månsson and Dr. Lars Behrenz, and I am grateful to for their support. Jonas, as my main supervisor, is undoubtedly the most important source of support in writing this thesis. His relevant questions, patient suggestions and pushes in the right direction have substantially improved the quality of this thesis.

During courses, seminars and conferences I have received many helpful comments. In particular I would like to thank the discussant at my final seminar, Professor Andreas Stephan, for his suggestions. Furthermore, I received many good comments during the course on Theory and Practice of Efficiency and Productivity Measurement in Bandung 2017 and at the Jönköping International Business School, Queensland University of Technology, the Swedish National Audit Office, the 15th International Conference on DEA in Prague and in my seminars at Linnaeus University. Furthermore, I am both honored and grateful for my discussant, Professor Christopher J. O'Donnell, and his valuable comments during the Early Career Research Day prior the European Workshop on Efficiency and Productivity Analysis in London, June 2017.

The department of Economics and Statistics at Linnaeus University has given me the opportunity to take courses, visit universities abroad and go to international conferences – it has been very interesting and helpful for my education, for which I am grateful. The department consists of employees who always help and support if necessary. In particular I am indebted for the comments I have received from Dr. Håkan Locking. Furthermore, I am obliged for all interesting discussions, advice and friendships with the present and previous Ph.D. students; Dr. Abdulaziz Abrar Reshid, Zhizheng Wang, Dr. Luca Fumarco, Dr. Emma Neuman, Dr. Deliang Dai, Dr. Chizheng Miao, Hanna Swahnberg, Dingquan Miao and Abdul Aziz Ali. Regarding administrative matters, I would like to thank Hanna Sandqvist and Anna-Karin Bjuvsjö Johannsson, who are always very helpful with issues related to, for example, forms or receipts.

Last but not least, I want to thank my friends and family. In particular, my parents, Göran and Agneta, have been very important regarding talking with me as good friends and providing mental and practical support and much more.

Växjö, October 2017
Pontus Mattsson

The impact of labor subsidies on total factor productivity

Pontus Mattsson[♣]

Abstract: Lower expected productivity is the motivation for subsidizing labor, but all research, with one exception, focuses on other effects while some investigates the TFP effects of capital subsidies. This study combines methods that, to the best of my knowledge, have not previously been used together to determine the impacts of labor subsidies on total factor productivity (TFP). Further, the profit per employee is included as a second outcome. Coarsened exact matching (CEM) is performed on the key variables; difference-in-difference (DID) is then applied to the matched data. It is found that firms employing workers with wage subsidies experience negative and significant effects on both TFP and profit per employee. Heterogeneity is, however, observed; the only sector to show a deficit in both TFP and profit per employee is wholesale. During the second year with a subsidy, a negative impact can be observed on the profit per employee but not on TFP. The policy conclusion from the analysis is that subsidizing individuals from particular groups is necessary to induce firms to hire workers from these groups. However, the time period for which a single firm is subsidized should be considered.

Keywords: coarsened exact matching (CEM); difference-in-difference (DID); profit per employee; subsidized employment; Törnqvist TFP index.

JEL Classification: C43 · D24 · J08 · J23 · O38

Acknowledgment: I would like to give special thanks to Prof. Christopher J. O'Donnell for his valuable comments during the Early Career Research Day prior to the 15th European Workshop on Efficiency and Productivity Analysis (EWEPA), 2017, London. Furthermore, I am grateful for all the suggestions made in seminars at Linnaeus University, Queensland University of Technology (QUT), the course in the Theory and Practice of Efficiency and Productivity Measurement on Static & Dynamic Analysis at Universitas Padjadjaran, Jönköping International Business School (JIBS) and the Swedish National Audit Office.

[♣] Department of Economics and Statistics, Linnaeus University, 351 95 Växjö, Sweden. E-mail: pontus.mattsson@lnu.se

1. Introduction

A policy used to prevent long-term unemployment is to subsidize the unemployed. The targeted groups in Sweden are for example the already long-term unemployed, unemployed people with disabilities and newly arrived immigrants.

There is extensive research on wage subsidies from a labor supply perspective. In general these studies point to a small but positive impact on the labor market performance of the individuals (see e.g. Gerfin et al. 2005; Jaenichen and Stephan 2011; Kraft 1998), and the re-employment probabilities of the particular group usually increase (see e.g. Calmfors et al. 2002; Kluve 2006). However, only a few studies exist on the labor demand side in relation to subsidies. The previous research at the firm level mainly deals with the total employment effects (see e.g. Bennmarker et al. 2009; Kangasharju 2007; Rotger and Arendt 2011), firm heterogeneity (Bellmann and Stephan 2014), crowding-out effects and deadweight losses (see e.g. Calmfors 1994) or the impacts on the components of the profit function (Månsson and Quoreshi 2015). Only Koski and Pajarinen (2015) consider the productivity effects on firms as a result of hiring subsidized workers. Although only one study can be found regarding the productivity effects from labor subsidies, there is extensive research concerning the total factor productivity (TFP) effects as a result of subsidies on the other input factor of production, capital (see e.g. Bergström 2000; Bernini and Pellegrini 2011; Harris and Trainor 2005; Managi 2010; Skuras et al. 2006). Thus, one motivation for this study is the lack of research on productivity effects regarding non-capital input subsidies.

This paper takes its point of departure from an employer's perspective of subsidized employment and investigates in particular whether there is a change in TFP and profit per employee as a result of hiring employees with wage subsidies. The focus is on the effects of wage subsidies on all subsidy programs together; however, programs targeting the disabled are excluded.

The average treatment effect of the treated (ATT) is estimated for the matched data using the TFP measured by the Törnqvist TFP index and the profit per employee as outcomes. Coarsened exact matching (CEM), developed by Blackwell et al. (2009), is used as the matching procedure. There is only limited literature in the area of TFP that aims to investigate policy impacts based on a causal relationship. Most of the previously mentioned studies regarding capital subsidies do not identify a treatment and a control group to be compared over time. This makes it difficult to evaluate the treatment effects. From a methodological point of view, Greene (2010) proposes a model for sample selection in stochastic frontier analysis (SFA) and Chen et al. (2014) put forward a method to estimate the treatment effect in SFA. Some empirical studies can be found that use a proper impact design to explain the variation in TFP due to political interventions. For example, Bernini et al. (2017) use SFA and a multiple regression discontinuity design, Bernini and Pellegrini (2011) perform matching and difference-in-difference (DID), Bloom et al. (2013) and Lisi and Malo (2017) use DID. However, each of these studies performs impact evaluations to evaluate the treatment effect based on a statistically estimated value of TFP. Nevertheless, the analysis of causal relations is fundamental when evaluating policy changes and deserves more attention in the TFP literature. The contribution of this paper is therefore twofold: 1) to propose a combination of a TFP index and common methods for policy evaluation, such as conditional DID, that is, matching and DID, in the same analysis and 2) to be the first study to investigate the relationship between firm TFP and subsidized employees.

The results show a negative and significant effect on both TFP and profit per employee for the whole sample. On the sector level, only wholesale indicates a negative and significant effect on TFP and the profit per employee. The negative effect on TFP is no longer statistically significant during the second year with a subsidy, which indicates a positive effect of on-the-job training. Conversely, the profit per employee is negative, which can occur regardless of the zero effect on TFP if firms within the treated group do not manage to maximize their profit.

The paper continues as follows. Section 2 describes the Swedish wage subsidy scheme. In section 3 a theoretical framework is presented and hypotheses are derived. Section 4 describes the method. In section 5 the data are presented. The results of the study are presented in section 6. Finally, section 7 concludes and discusses the policy implications.

2. The Swedish wage subsidy scheme[1]

This study focuses on wage subsidies within the programs of Special Recruitment Incentives (*Särskilt anställningsstöd*), New Start jobs (*Nystartsjobb*) and Entry Recruitment Incentives (*Instegsjobb*) on the aggregated level; that is, no separation is made between the programs. The purpose of the included programs is to increase the employability of the particular groups in the regular labor market, since all the individuals included have characteristics that make it difficult to find a job. These can be for example long-term unemployment or poor knowledge of Swedish; the Entry Recruitment Incentives also require employees to study Swedish for immigrants (SFI) part-time. The firms decide who they want to employ with a subsidy as long as the individual fulfills the requirements. The main requirement for all the programs is that the salary and insurances are in line with the collective agreement. Further, termination of an employee's engagement due to job scarcity is not allowed during the previous nine-month period (Regulation 1997:1275); this is not the case for New Start jobs, for which firms do not have to follow the Employment Protection Act either (LAS) (Regulation 2006:1481).

[1] Changes have been made to the programs during the time period; a full description of these can be provided on request.

For New Start jobs, the subsidy is 1 payroll tax, which is 31.42% of the salary if the participant is at least 20 and below 26 years old. If the employee is at least 26 years old, the subsidy is always 2 payroll taxes (Regulation 2009:26). In the case of the Entry Recruitment Incentives, the subsidy is 80% of the salary but maximum compensation of 800 SEK per day, and for the Special Recruitment Incentives it is 85% or a maximum of 890 SEK per day (Regulation 2010:2028).[2]

The period for which an individual can participate in a New Start job is the same duration as he or she has been unemployed but a maximum of 1 year if the individual is below 26 years old and a maximum of 5 years if the individual is at least 26 but less than 55 years old. The maximum time is 10 years if the individual is older, but it can only be given until he or she reaches 65 years of age (Regulation 2009:26). Support within the Entry Recruitment Incentives is given for 6 months to individuals working more than 50% of full-time hours and at most for 24 months if they work less than 50% (Regulation 2007:420). Special Recruitment Incentives can be given for a maximum of 12 months (The Swedish Employment Office 2009).

According to the Swedish Agency for Public Management (2011), the participants in the Entry Recruitment Incentives are relatively young and have the smallest share of secondary school education and the largest share of post-secondary school education. The heterogeneity in this group is therefore large, since it includes both people who are quite close to the labor market with a high education level and conversely low-educated individuals with less/no experience. New Start jobs are similar except that there are fewer participants with less than secondary school education; therefore, this group has less heterogeneity but all the participants are quite distanced from the labor market because of the requirement of a long period of unemployment. The participants in the Special Recruitment Incentives have spent the longest amount of time unemployed.

3. Theoretical framework and hypotheses

The main objective for firms is to maximize their profit. The starting point of the analysis is therefore the profit function in equation 1.

$$[1]\ \pi(w_u, w_s, r, y) = y * p(y) - L_u w_u - L_s w_s - Kr,$$

where π represents the total profit. y is the total production expressed as $f(L_u, L_s, K)$, where L_u represents the amount of unsubsidized labor, L_s the amount of subsidized labor and K the amount of capital. Further, w_u is the wage cost for unsubsidized labor, w_s the wage cost for subsidized labor and r the price of capital. To maximize the profit, the marginal revenue product (MRP) of

[2] The employer can also obtain supervising support of 50 SEK per day for the first 3 months within the programs of particular employment subsidies and Entry Recruitment Incentives.

each input has to equalize its marginal cost. The derivatives of interest, in the scope of this paper, concern unsubsidized and subsidized workers. They are expressed as $\partial \pi(w,r,y)/\partial L_u = p * \partial f/\partial L_u - w_u = 0$ and $\partial \pi(w,r,y)/\partial L_s = p * \partial f/\partial L_s - w_s = 0$, respectively.

The marginal products of labor, that is, $\partial f/\partial L_u$ and $\partial f/\partial L_s$, have means $\overline{\partial f/\partial L_u}$ and $\overline{\partial f/\partial L_s}$ and variances σ_u and σ_s. The marginal productivity of labor is assumed to decrease as the duration without schooling or employment increases. Subsidized employees are therefore less productive than unsubsidized workers on average. The variance components mean that there will be low-productive workers without subsidies as well as some high-productive workers among those who are entitled to subsidies. For simplicity, a fully competitive product market is assumed. If we first suppose that no subsidy is given, this means that the wage from the firm perspective will be equal for the two groups, that is, $w_u = w_s = w$. The consequences for the firm that hires a worker with a subsidy can be viewed from the marginal revenue products (MRPs) of unsubsidized and subsidized workers, respectively. In the case of no subsidy, the selection of employees on average is calculated by equation 2:

$$[2] \qquad w = p * \left[\overline{\partial f/\partial L_u}\right] > p * \left[\overline{\partial f/\partial L_s}\right].$$

According to equation [2], an employer would hire individuals of category L_u, leaving category L_s unemployed, when productivity sorting can be carried out on easily observed characteristics, for example the duration of unemployment (see e.g. Eriksson and Rooth 2014). As a result, without compensation for firms, the targeted groups in wage subsidy schemes are more likely to end up in long-term unemployment. This result, together with other consequences of long-term unemployment, such as health issues (see e.g. Kieselbach 2003), motivates governments to use some kind of wage subsidy as a policy measure. Wage subsidies are used to compensate for the potential deficit in the MRP that firms possibly face (see e.g. Bellmann and Stephan 2014).

Individuals who are eligible for wage subsidies have on average lower productivity than those who are not entitled to subsidies. However, variation around this mean, indicated by the variance components, σ_u and σ_s, will exist. It can be expected that firms to some extent select individuals who are entitled to subsidies based on their expected productivity and subsidy levels. This implies that a non-random sample from the group eligible for wage subsidies will be hired. Depending on how successful firms are in selecting individuals and how successful the employment service is at setting the “correct” subsidy level, there will be different outcomes in terms of the profit per employee for the firms. Conditional on the above-mentioned circumstances, different scenarios and outcomes are possible.

Case 1

If perfect information regarding the subsidized individual is assumed, the productivity deficit is known and the subsidy will be set so that it exactly compensates for the difference in the MRP, that is, $p * \partial f / \partial L_u = (p + s) \partial \mathrm{f} / \partial L_s = w$, where s is the wage subsidies as part of the firm's revenue per entity. In this case the subsidy perfectly compensates for the lower productivity of the subsidized individuals, keeping the profit per employee unaffected. The firm will then, from an economic point of view, be indifferent with regard to hiring a person with or without a wage subsidy. In this case lower productivity but no effects on profits can be expected for firms that hire subsidized workers.

If non-perfect information exists regarding the individuals' marginal productivity, the subsidy might over- or under-compensate for the deficit.

Case 2

A firm is overcompensated if $[\partial f / \partial L_u] > [\partial f / \partial L_s]$ but $p * [\partial f / \partial L_u] < (p + s)[\partial f / \partial L_s]$. There are two effects relating to the subsidies in this case. Firstly, since the marginal productivity is lower for the subsidized individuals, it will reduce the TFP. However, the profit per employee will increase, since the subsidy more than compensates for the productivity deficit. Profit-maximizing firms will in this case replace regular workers with subsidized workers, since there are economic incentives to hire individuals with subsidies. This will generate a high displacement effect (Dahlberg and Forslund 2005), which can lead to positive employment effects for persons entitled to wage subsidies when the subsidy per entity is added to the MRP (see e.g. Kangasharju 2007; Rotger and Arendt 2011).[3] In this case lower productivity and increased profits are expected.

Case 3

The third case is that the subsidy does not compensate enough for the productivity deficit, that is, $p * [\partial f / \partial L_u] > (p + s)[\partial f / \partial L_s]$, indicating that there are no economic incentives to employ a person with a subsidy (see e.g. Pissarides 1992; Snower 1995). A loss in both productivity and profit can be the case when firms fail in the selection process and overestimate the productivity of the subsidized individuals. The reasons for overestimating an individual's marginal productivity could relate to issues such as workplace introduction and training; for example, other employees must spend time introducing the subsidized employee to the work. Another reason might be that TFP depends on the "weakest link," meaning that the average productivity is related to the least productive worker (Kremer 1993). Finally, firms with an altruistic purpose

[3] Potential employment effects are not investigated in this paper.

(Behrenz et al. 2015), rather than profit-maximization behavior, can hire an individual who is eligible for a wage subsidy, which will result in a reduction in both the productivity and the profit per employee (Behrenz et al. 2015).

Case 4

Individuals with above-average productivity in the subsidized group will be chosen if the screening is performed successfully, which will leave the least productive individuals unemployed (see e.g. Montgomery 1991; Pries and Rogerson 2005). Employers will in this case only hire subsidized employees who are at least as productive as non-subsidized ones, that is, $[\partial f / \partial L_u] \leq [\partial f / \partial L_s]$ and further $p * [\partial f / \partial L_u] \leq p * [\partial f / \partial L_s] < (p + s)[\partial f / \partial L_s]$. The compensation is, in contrast to the previous cases, given for the higher risk of employing an individual with an expectation of lower human capital due to the group characteristic of the long-term unemployed (see e.g. Gibbons and Katz 1991; Hartmann 2004; Hujer et al. 2001). When taking higher risk without inducing any deficit in TFP, the profit per employee will increase. If this happens, a deadweight loss will occur. Behrenz et al. (2015) find that 40% of the employers consulted answered that the same individual would be employed even without a subsidy. Case 4 will result in unaffected or increased productivity and higher profits.

Case 5

In the fifth case, it is assumed that firms do not manage to maximize their profits, that is, $p * [\partial f / \partial L] \neq w$. The outcome in that case can be a decline in the profit per employee while productivity increases or remains equivalent (for further descriptions see e.g. Kumbhakar et al. 2015, pp. 280–285).

From a more dynamic perspective, if there is a productivity gap to start with, a closed gap can occur if there is an increase in productivity for the subsidized employees due to individual on-the-job training. Closing of the productivity gap can also occur when firms become more experienced in hiring individuals from the group eligible for wage subsidies and therefore perform better in the selection process. The productivity gap, according to Bellmann and Stephan (2014), is expected to be closed in the medium or long term.

To sum up, the firm effects of employing persons with wage subsidies are not straightforward. If there is full information about the marginal productivity deficit and firms succeed in maximizing their profit, as in case 1, the resulting loss in the MRP will be perfectly compensated for by the subsidy. Assuming less than perfect information regarding the individual's marginal productivity can lead to either too high or too low a compensation level in relation to the decrease in the MRP. Another possible outcome is no negative effect on the marginal productivity, as in case 4, in which the subsidy will generate a positive effect on the MRP. Finally, if firms do not manage to maximize their profits, a situation could arise in which a decline in the profit per employee is observed at the same time as TFP remains constant, as in case 5.

4. Method

4.1. Measuring total factor productivity (TFP)

To measure the impact of hiring unemployed individuals who are entitled to a wage subsidy, Koski and Pajarinen (2015) use labor productivity as the dependent variable. In this study a TFP index is used to measure productivity change. The main reason for using TFP is that partial productivity measures, such as labor productivity, do not capture the potential substitution between labor and capital (see e.g. Brynjolfsson and Hitt 2003). Three different approaches to measuring TFP are used in the literature.

The first approach is to estimate an average production function and compute the Solow residual as an estimate of TFP (see e.g. Solow 1957). The second approach is to estimate a stochastic production frontier (see e.g. Aigner et al. 1977; Skuras et al. 2006). Both these approaches allow comparison over time but are questionable for estimating the treatment effect of a policy intervention. The main problem with the two-step procedure, according to Wang and Schmidt (2002), is that omitted variables in the first step will bias the coefficients in the second step. This uncertainty will produce inefficient estimates of the determinants of TFP (see e.g. Newey and McFadden 1994). Harris and Trainor (2005) try to solve this issue by including TFP, defined as the time trend and the residual, directly in a regression model, with value added as the dependent variable. This is, however, disputable, since value added is in fact a measure of output.

The third way of measuring TFP is to use an index approach. Several indexes are suggested in the literature, such as the Malmquist (Caves et al. 1982), Fisher (1922), Laspeyres (1871), Paasche (1874) and Törnqvist (1936) indexes. An issue with these indexes is that they do not satisfy the circularity and transitivity axioms discussed for instance by O'Donnell (2012). A suggested alternative is the geometric Young index (Young 1812). However, Fisher (1922) concludes that indexes satisfying these axioms are unattractive based on other reasons. Drechsler (1973) summarizes the Fisher dilemma by stating that a conflict is always present between characteristicity and circularity.

In this study the Törnqvist index approach is adopted to measure TFP. The Törnqvist index has the advantage of being computed directly from data on the prices and quantities of a firm's capital and labor (Caves et al. 1982). Another advantage of using the Törnqvist index is that it is non-parametric, which implies that no error term from the first step has to be handled in the second step when comparing the treatment and control groups. The third advantage of the Törnqvist index is that it accommodates changes in the relative prices of inputs over time (Fan and Zhang 2002). Finally, Diewert (1976) recommends index numbers that are exact for the translog revenue functions. The Törnqvist index satisfies this condition.

Törnqvist TFP index

The Törnqvist TFP index is computed in two parts relating to inputs and outputs. The logarithm of the output index ($\ln Q^T_{nit,nit-1}$) is defined and calculated as follows:

$$[4] \qquad \ln Q^T_{nit,nit-1} = \ln \prod_{n=1}^{N} \left(\frac{q_{nit}}{q_{nit-1}}\right)^{\frac{\omega_{nit}+\omega_{nit-1}}{2}} = \frac{1}{2}\sum_{n=1}^{N} [\omega_{nit} + \omega_{nit-1}] [\ln q_{nit} - \ln q_{nit-1}] = [\ln q_{it} - \ln q_{it-1}],$$

where ω_{nit} and ω_{nit-1} are the revenue shares. Further, q_{nit} and q_{nit-1} are the value added of output *n* for firm *i* during *t* and *t*-1 (see e.g. Christensen and Jorgenson 1970; Rungsuriyawiboon and Coelli 2004). In the last part of equation 4, the revenue shares and the subscript *n* are one, since value added is the only output.

The logarithm of the Törnqvist input index ($\ln P^T_{mit,mit-1}$) is defined and calculated as:

$$[5] \qquad \ln P^T_{mit,mit-1} = \ln \prod_{\mathrm{m}=1}^{M} \left(\frac{\mathbf{x}_{mit}}{\mathbf{x}_{mit-1}}\right)^{\frac{\omega_{mit}+\omega_{mit-1}}{2}} = \frac{1}{2}\sum_{m=1}^{M} [\omega_{mit} + \omega_{mit-1}] [\ln \mathbf{x}_{mit} - \ln \mathbf{x}_{mit-1}],$$

where ω_{mit} and ω_{mit-1} are the cost shares of the inputs for firm *i* in periods *t* and *t*-1. The input vector is symbolized as $\mathbf{x}_{mit}$, where *m* stands for inputs, that is, the capital assets or the number of full-time-equivalent employees. When calculating the cost shares, the total salaries are used as the cost of labor.[4] The price of capital is estimated by multiplying the Swedish Treasury Bill plus an expected depreciation rate of 5% with capital assets (see e.g. Lam and Lam 2005; Serafica 1998).[5]

The Törnqvist TFP index is defined as the ratio of the output index and the input index:

$$\text{Törnqvist TFP index} = \exp\left(\ln Q^T_{nit,nit-1} - \ln P^T_{mit,mit-1}\right) = \frac{Q^T_{nit,nit-1}}{P^T_{mit,mit-1}}.$$

[4] Total salaries can include the subsidies in some cases and omit them in others. This can occur due to different accounting practices between firms; thus, there might be an error in the calculation. As a sensitivity test, the total salaries are multiplied by 0.8 (and 0.9) to receive a new value, which is used when calculating the weights of the Törnqvist TFP index, and then the regressions are estimated again with consistent results.

[5] The Repo interest rate decided at the last meeting at the Central Bank of Sweden of each year is used. The results are also consistent for a depreciation rate of 10%.

4.2. Identification

Ideally, the allocation of firms to the treated and untreated groups would have been random. However, this is not the case, as pointed out by both Behrenz et al. (2015) and Bellmann and Stephan (2014). Therefore, this study has to rely on a quasi-experimental design to identify the impacts of subsidized employees. More explicitly, the selection that takes place needs to be modeled.

To be considered as treated, at least 1% of the total number of employees during the post-treatment period is required to be subsidized. The comparison group consists of firms that have never hired a person with a subsidy throughout the observation period of 2007–2013.

4.2.1. Characteristics that influence selection

Behrenz et al. (2015) find that the education level, a large share of employees with a specific gender, the share of immigrants, the share of young workers and the share of older workers correlate with the decision to hire a person entitled to a wage subsidy. Bellmann and Stephan (2014) find that the profit per employee for firms that hire with subsidies is lower than for firms that do not; that the probability of employing a person with a subsidy increases with the labor turnover; and that wage subsidies are most commonly used in retail, repair and wholesale. These characteristics influence the selection into treatment but are present in the data and therefore it is possible to control for them. There are also factors that are unobservable. In the survey conducted by Behrenz et al. (2015), a non-ignorable number of the firms claimed that they would never hire anyone with a subsidy, regardless of the compensation level (Behrenz et al. 2015).

4.3. Estimation strategy

The average treatment effect of the treated is estimated using a conditional DID approach, specifically matching followed by DID. The DID method is used in previous research when investigating the impacts of subsidies, for example labor productivity (Baghana 2010; Koski and Pajarinen 2015), employment effects (see e.g. Bennmarker et al. 2009; Kangasharju 2007; Rotger and Arendt 2011) and TFP effects as a result of capital subsidies (see e.g. Bernini and Pellegrini 2011). The analysis is performed for the whole sample and for each sector separately. Separate estimations between sectors allow for different production functions, which are preferable, since different technologies might influence productivity growth differently. To take care of some of the non-randomness relating to the characteristics previously described, coarsened exact matching (CEM) is performed on the data during the pre-treatment period. This makes a potential correlation among the parameters of interest more likely to be causal (Cobb-Clark and Crossley 2003). CEM has the advantage over other matching methods of being non-parametric, being more transparent, dealing with common support by construction and reducing the sensitivity to measurement error (Iacus et al. 2012). As matching variables, the number of employees, sector dummies and 25 different regional dummies are used. Each of these characteristics is matched exactly. This means that the firms to be compared are within the same sector, have the same number of employees and are located in the same region. In the DID setting, serial correlation in the dependent variable generates standard errors that are too small according to Bertrand et al. (2002). This problem has to be handled, since the Törnqvist TFP index is serially correlated by construction. There are different possibilities to resolve this issue. One way, used in this study, is to consider the mean value of two years before treatment as a pre-treatment period and the mean value of two years after treatment as a post-treatment period. The standard DID model for the comparison of means between the treated and the control group, before and after treatment, to be estimated using the matched data, is:

$$[6]\ \text{Törnqvist TFP index}_{it} = \rho_0 + \rho_1 Treated + \rho_2 After + \rho_3 DID + \rho_i \mathbf{X}_{it} + \mu_i + \varepsilon_{it}.$$

The calculated Törnqvist TFP index and profit per employee taken from the firm's annual report are used as dependent variables, respectively. *Treated* is a dummy variable, which indicates that the firm has at least 1% subsidized employees during each year of the after period and none before.[6] *After* is the period after treatment, and *DID* is the interaction of treated and after that measures the treatment effect. The mean of 2011–2012 is used as the post-treatment period, and the mean of 2009–2010 is considered as the pre-treatment period. To receive a sufficient number of observations on the sector level, another period is aggregated to the sample, specifically the period in which 2010–2011 is used as pre-treatment and 2012–2013 as post-treatment. **X** is a vector of covariates for each firm, *i*, over time (*t*) chosen based on what is concluded to differ between the groups of comparison. The **X** vector consists of information about the number of full-time employees and the share of different education levels, age groups, foreign background, women and the gender of the manager. One advantage of the DID approach is that it handles the potential time-invariant bias, meaning that μ_i cancels out. Time-invariant bias is, according to Koski and Pajarinen (2015), present because firms that choose to use wage subsidies are less productive than others. Finally, ε_{it} represents an error term.[7]

An assumption of the DID method is the existence of parallel pre-treatment trends in the dependent variables, namely TFP and the logarithm of the profit per employee. To investigate this assumption, a statistical test using a method proposed by Mora and Reggio (2015) is performed as well as visual analysis of the TFP change and profit per employee over time.

Another potential source of problems relates to the fact that, according to Behrenz et al. (2015), a substantial number of firms are unwilling to hire subsidized employees, regardless of the

[6] The subsidies can alternatively be included as a continuous variable.
[7] Dynamics over a longer period of time is not investigated due to data limitations.

compensation level. To investigate this matter, a robustness test is performed. In this test firms that do not hire workers with subsidies in the following year are excluded from the control group. This means that firms that are unwilling to hire workers with subsidies, among others, are no longer included as a comparison. The procedure for the post-treatment period of 2011–2012 is as follows: 1) firms in the control group are required to have at least 1% of their total employees with a subsidy during 2013 but no subsidized employees beforehand; 2) to investigate the effect between the pre-treatment period of 2009–2010 and the post-treatment period of 2011–2012, 2013 is deleted. Based on the same procedure, aggregation is performed with the period before the one previously described; specifically, 2008–2009 and 2010–2011 are considered as the pre-treatment period and the post-treatment period, respectively. Thus, the obtained estimates are of the potential differences in TFP and the profit per employee between firms hiring with subsidies during the post-treatment period and firms hiring with subsidies in the following year. This procedure makes all the firms in the control group willing to hire with subsidies but at a later point in time.

To summarize, the combination of the Törnqvist TFP index and conditional DID has not been used before. It has the advantage of not suffering from the issues of estimating in the first stage of a two-stage procedure. Further, the method is both simple and transparent.

5. Data

The data used in this study are linked employee–employer data from Statistics Sweden. The data are combined with information from the Swedish Employment Service about those individuals who received some form of wage subsidy between 2007 and 2013. All individuals employed within the programs of Special Recruitment Incentives (*Särskilt anställningsstöd*), New Start jobs (*Nystartsjobb*) and Entry Recruitment Incentives (*Instegsjobb*), aggregated to the firm level, are included in the data.

The data on subsidies include the first date and the last date on which an individual is employed with a subsidy at a specific firm.[8] The number of days for which a subsidized employee has been employed is recalculated into full-time equivalence to make it comparable with the variable of full-time-equivalent employees from Statistics Sweden (SCB). Descriptive statistics of the share of firms with subsidized employees, the number of full-time subsidized employees in an average firm and the ratio of full-time-equivalent workers with a subsidy compared with non-subsidized employees are reported in table A3.

[8] The same individual can be employed by different firms with a subsidy within one year.

In table 1 the numbers of treated and control firms divided into sectors are presented in columns 5 and 6, both for the full and for the matched sample. In columns 2–4 the number of treated is presented separately for each subsidy program; the number of treated separated into each program sums to more than the number of treated in column 5, indicating that some firms use more than one of the programs. Only firms that existed during the whole time period and did not change sector at any time are included. Further, firms with a Törnqvist output index, that is, $\ln y_t - \ln y_{t-1}$, of more than 1 or less than -1 are considered to be outliers and are excluded from the analysis, because these are small firms that dramatically increase their output. Including them will not generate relevant TFP growth on average. Changes in output drive the index substantially upwards compared with TFP measured for the whole economy.

Table 1 Number of firms in the treated and control groups. SAS = Special Recruitment Incentives, NYJ = New Start jobs and INS = Entry Recruitment Incentives

	(2) Treated SAS	(3) Treated NYJ	(4) Treated INS	(5) Treated	(6) Control
All sectors	444	1603	399	2 040 (1 941)	71 304 (42 101)
Manufacturing, mining and quarrying	47	225	45	270 (246)	6 756 (3 587)
Construction	40	229	34	278 (261)	12 562 (6 933)
Wholesale and retail trade	102	360	87	472 (465)	14 258 (10 368)
Transportation and storage	29	148	27	169 (161)	7 160 (4 363)
Accommodation and food service activities	28	104	28	173 (171)	1 804 (1 291)
Business service activities	88	257	83	325 (317)	13098 (9 243)

The number of firms in the matched sample is in parentheses.

A number of control variables are included in the conditional DID estimation. The size of the firm, measured by the number of employees as in Koski and Pajarinen (2015), is identified as being of importance for hiring persons entitled to subsidies. Identifying two similar groups for comparison is fundamental when the DID method is used. Thus, other firm characteristics that are concluded to affect the decision on hiring workers with wage subsidies according to Behrenz et al. (2015) are included as control variables. These variables are presented in table 2.

Table 2 Description of the included control variables

Variables	**Description**
Number of employees	Number of full-time equivalent employees
Dummies of the gender of the operational manager	Dummy of the gender (male, female, mixed gender) of the operational manager during the period
Share with primary education	Employees with only primary education divided by the total number of employees
Share with college education	Employees with the maximum of college education divided by the total number of employees
Share with post-secondary education	Employees with post-secondary education divided by the total number of employees
Share of young	Employees aged below 25 divided by the total number of employees
Share of older	Employees aged above 55 divided by the total number of employees
Share of immigrants	Employees with at least 1 foreign parent divided by the total number of employees
Share of women	Number of employed women divided by the total number of employees

The control variables are chosen based mainly on the characteristics that the previous literature states to be key characteristics for the decision either to hire or not to hire personnel entitled to a subsidy and potentially influence productivity. For example, the level of education is often considered to be a driving factor of TFP growth when a highly educated workforce is assumed to increase productivity (see e.g. Becker 1964). Age is often used as a proxy for experience, and a more experienced workforce is likely to be more productive (see e.g. Becker 1964). Furthermore, the variable of immigrant background is used as a proxy for language skills. Inferior language skills, according to Kossoudji (1988), have an influence on wage growth, which is closely linked to individual productivity. Finally, differences between males and females are found in relation to giving birth that generate a direct effect on the hours worked, which influences experience and thus productivity (see e.g. Lundberg and Rose 2000).

5.1. Matching results

To evaluate the effect on TFP as a result of hiring subsidized employees, the treated and untreated groups have to be similar in characteristics during the pre-treatment period. With CEM it is possible to match exactly on each observed characteristic. However, the cost of including many factors in the matching stage is loss of external validity, meaning that a substantial number of firms are excluded due to having no match. A number of different matching variables are applied, in different combinations, and the number of employees, sector and region are concluded as being the fundamental ones.

Matching quality

To assess the overall imbalance in terms of the number of employees, sector and region before and the improvement as a result of the matching, the $\mathcal{L}_1$ statistic, introduced by Iacus et al. (2012), is used. If $\mathcal{L}_1 = 1$, it means perfect imbalance, and if $\mathcal{L}_1 = 0$, it means perfect balance. The results show that $\mathcal{L}_1 = 0.488$ before matching and $\mathcal{L}_1 = 0.000$ after matching. Table 3 presents the means during the pre-treatment period for treated and untreated firms before and after matching. The fourth column shows the differences in mean between the treated and the control group before matching is performed.

Table 3 Means before and after matching two years before treatment

	Unmatched			**Matched**		
Variables	Control group (2)	Treated (3)	Difference (4)	Control group (5)	Treated (6)	Difference (7)
Number of employees	5.063	7.066	2.003***	6.518	6.280	-0.238
Man as the operational manager	0.855	0.836	-0.019	0.856	0.835	-0.021
Woman as the operational Manager	0.120	0.131	0.011	0.113	0.134	0.021
Mixed gender as the operational manager	0.025	0.033	0.008	0.031	0.031	0.000
Share with primary education	0.182	0.179	-0.003	0.185	0.181	-0.004
Share with college education	0.553	0.578	0.025***	0.576	0.576	0.000
Share with post-secondary education	0.274	0.253	-0.021***	0.238	0.243	0.005
Share of young	0.099	0.149	0.050***	0.122	0.149	0.027***

Share of older	0.283	0.196	-0.087***	0.246	0.196	-0.050***
Share of immigrants	0.160	0.224	0.064***	0.168	0.228	0.060***
Share of women	0.275	0.319	0.044***	0.294	0.320	0.026***

*** indicates significance at the 1% level and ** significance at the 5% level.

By comparing the magnitude and significance between column (4) and column (7) in table 3, it can be seen that some characteristics become non-significant after matching.[9] The intuitively important characteristics in relation to productivity, namely the education levels, are significant before and non-significant after matching. Further, the differences in magnitude decrease for all except three characteristics after matching, which is apparent from comparing column (4) and column (7). The data still contain small differences, but, as pointed out by Iacus et al. (2012, p. 23), "… no magical method will be able to fix this basic data inadequacy ...." Problems may still be present if an important covariate is not included. This study, however, either matches on or controls for the most important variables concluded to affect selection into treatment according to Behrenz et al. (2015) and Bellmann and Stephan (2014).

6. Results

In the first part of the results, the pre-treatment trend assumption is investigated. It is crucial for the trends to be parallel when using a DID approach. Second, the average treatment effect on all firms is considered. The treatment effect is studied regarding both the Törnqvist TFP index and the logarithm of the profit per employee as the dependent variables. Third, the same analysis is performed on the sector level.

6.1. Test of the pre-treatment trends

The traditional way to investigate whether the pre-treatment parallel trend assumption holds or not is to plot the trends before treatment and make a subjective judgment based on visual analysis.[10] However, Mora and Reggio (2015) provide a test for this assumption. The p-values of the test are reported in table 4.

Table 4 Test of the parallel pre-treatment trend assumption using the matched data

H_0: Parallel pre-treatment trend	**Törnqvist TFP index**	**Log(profit/employee)**
All sectors aggregated	0.374	0.527
Manufacturing, mining and quarrying	0.818	0.910

[9] To be noted is that neither of the compared variables are included in the matching process, except number of employees.

[10] A short illustration of these plots can be found, on the aggregated level, in figures A1 and A2 in the appendix.

Construction	0.528	0.801
Wholesale and retail trade	0.169	0.904
Transportation and storage	0.229	0.708
Accommodation and food service activities	0.178	0.206
Business service activities	0.130	0.621

As table 4 reveals, the parallel pre-treatment trend assumption cannot be rejected for any sector. This means that the development in TFP and the profit per employee before a firm hires a person entitled to a subsidy is the same as for the firms that do not hire a person with a subsidy. This implies that the DID estimates are likely to give a good estimate of the treatment effect.

6.2. Results for all firms

The results for the aggregated sample, when treated firms have at least 1% subsidized employees over two years, are presented in table 5. Data from two years before treatment and a follow-up period of two years after treatment are used in the first row. The second and third rows are the same estimations as in the first except that only either the first year or the second year is included in the post-treatment period. The reason for this separation is to investigate whether the coefficient changes due, for example, to on-the-job training. Throughout the presentation the control group consists of firms that did not employ a person with a subsidy during the period 2007–2013.

Table 5 Changes in TFP for all sectors using the matched sample

Dependent variable:	**Firms with at least 1% subsidized employees**		
Törnqvist TFP index	Treated	After	DID
Matched periods -1 and 0. Treated periods 1 and 2	0.015*** (0.004)	-0.001 (0.010)	-0.022*** (0.006)
Matched periods -1 and 0. Treated periods 1 and 2. Year 2 excluded	0.011 (0.007)	0.009*** (0.002)	-0.030*** (0.009)
Matched periods -1 and 0. Treated periods 1 and 2. Year 1 excluded	0.013*** (0.004)	-0.012 (0.008)	-0.014 (0.008)

Standard errors clustered on the sector level are in parentheses; *** indicates significance at the 1% level and ** at the 5% level. The covariates are reported in table 2.

The DID coefficient on the first row of table 5 shows a statistically significant reduction in TFP by 2.2% for treated firms in relation to matched firms in the control group. The coefficient becomes larger in magnitude when only the first year is included as the post-treatment period, which can be seen in the DID coefficient on the second row. Vice versa, when only the second year of the post-treatment period is included, the DID coefficient on the third row is smaller in magnitude and non-significant. This difference could be due to positive effects of on-the-job training. A deficit in TFP is an expected result relating to the arguments for giving subsidies. A potential problem, as pointed out by Koski and Pajarinen (2015), could be that it is more common for less well performing firms to use subsidized employees. However, based on the sign of the coefficients in column one, "treated," the subsidized firms in our matched data are at least as productive as the non-subsidized firms during the whole time period. This indicates that the results in the present study do not suffer from the potential issue that less well performing firms hire workers with subsidies. To investigate stability, the robustness test described in the methodology is applied. This test indicates the same magnitude of the coefficients, but they are non-significant due to a smaller sample size. However, the interpretation of the robustness test is that the impact of hiring employees entitled to wage subsidies remains consistent with the main results; that is, the results are stable even though there is another control group that consists only of firms that employ workers with subsidies at a later point in time.[11] The second question relates to the profit per employee, which is used as the dependent variable in the second part of the analysis. Since firms with a negative profit per employee exist, the variable is transformed by adding the smallest negative value in the sample to each firm before dividing it by the number of employees. Using this transformation makes the interpretation of the coefficient difficult; however, only the sign and the significance level are needed to relate to the theoretical section. Therefore, the coefficients of the profit per employee are excluded from the tables.[12] These results are reported in table 6.

Table 6 Effects on the profit per employee using the matched sample

Dependent variable:	**Firms with at least 1% subsidized employees**		
Log(profit/employee)	Treated	After	DID
Matched periods -1 and 0. Treated periods 1 and 2		+***	-***
Matched periods -1 and 0. Treated periods 1 and 2. Year 2 excluded		+***	-**
Matched periods -1 and 0. Treated periods 1 and 2. Year 1 excluded		+***	-***
Matched periods -1 and 0. Treated periods 1 and 2. Törnqvist TFP index used as control		+***	-***

*** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

[11] All the robustness test results are provided in tables A1 and A2 in the appendix.
[12] A table that includes the coefficients can be provided on request.

The DID estimate for the matched sample in the first row shows a statistically significant reduction in the profit per employee. This indicates that firms are negatively affected by hiring personnel with subsidies. The DID estimate when only data from the first year of the post-treatment period is used (second row) is also negative and statistically significant at the 5% level. Further, when only the second year is used, post-treatment (third row), the impact is significant at the 1% level. The result indicates on average undercompensation for firms that use subsidized workers if the relationship is a causal effect of the subsidized individuals. However, the Törnqvist TFP index is included as an explanatory variable in the fourth row with consistent results. One year after hiring with a subsidy, there is no observed TFP deficit. This means that the negative impact on the profit per employee is not only an effect of the individuals with a wage subsidy. Instead, these firms perform less well in maximizing their profit with respect to their input prices. Bellmann and Stephan (2014) conclude that the profit per employee on average is lower for firms that use subsidies. In this study's matched data, this is not the case based on the zero coefficients for the treated dummy (see column 1), indicating no difference on average but in the post-treatment period. A robustness test is also performed for the profit per employee with the same conclusion when two years are included as a pre-treatment period.[13]

To conclude, the conditional DID estimates are negative and significant at the 5% level in terms of TFP and at the 1% (or 5%) level for the profit per employee. The conclusion is therefore that a negative impact occurs during the first year for both outcomes.

6.3. Results by sector

In this section the same analysis as in section 6.2 is performed but separated on the sector level. Due to the fact that some sectors are small and contain very few observations, these sectors are excluded. The excluded sectors are agriculture; energy supply and environmental activities; real estate activities; education; human health and social work activities; arts, information and communication; and entertainment and recreation. The results are reported in table 7.

Table 7 Changes in TFP for the matched sample as a result of hiring subsidized employees by sector

Dependent variable:	**Firms with at least 1% subsidized employees**		
Törnqvist TFP index	Treated	After	DID
Manufacturing, mining and quarrying	0.011 (0.011)	-0.008 (0.003)	0.005 (0.016)

[13] However, here there are fewer observations, making each year separately non-significant, since the profit per employee might change considerably from one year to another.

Construction	-0.001 (0.011)	0.005 (0.003)	-0.013 (0.015)
Wholesale and retail trade	0.016 (0.010)	-0.035*** (0.003)	-0.030** (0.014)
Transportation and storage	-0.016 (0.017)	0.008 (0.004)	-0.022 (0.024)
Accommodation and food service activities	-0.002 (0.015)	0.022*** (0.006)	0.015 (0.021)
Business service activities	0.025** (0.011)	-0.002 (0.003)	-0.026 (0.016)

Standard errors are in parentheses; *** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

The results in table 7 indicate that the negative impact of subsidized employees on TFP is driven by firms within wholesale and retail trade. The non-significance in most sectors can either depend on successful selection from the firm perspective or indicate that a small amount of on-the-job training is required to perform equally well as other individuals. Firms within most sectors can therefore hire individuals at a lower cost without any TFP deficit. During the second year with a subsidy, wholesale and retail trade is also non-significant. However, related to the different cases in the theoretical framework, the effect on the profit per employee is ambiguous regardless of whether TFP increases, declines or remains the same.

From a policy point of view, the productivity deficit (or the risk factor for each of the reported sectors except wholesale) should be compensated for by the wage subsidy to equalize the MRP for subsidized and non-subsidized workers, respectively. This assumes perfect information regarding the productivity deficit and that the given subsidy will perfectly compensate for the differences in productivity. Employing an individual with a subsidy in this case will not generate any effect on the profit per employee; if not, the event of case 5 occurs. To investigate this issue, separate estimations are performed for each sector using the profit per employee as the dependent variable. The result is reported in table 8.

Table 8 Changes in the profit per employee for the matched sample. Two years' pre-treatment and two years' follow-up

Dependent variable:	**Firms with at least 1% subsidized employees**		
Log(profit/employee)	Treated	After	DID
Manufacturing, mining and quarrying		+**	
Construction			
Wholesale and retail trade		+***	-**

Transportation and storage	+***
Accommodation and food service activities	
Business service activities	+**

*** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

The results in table 8 show that subsidized employees have a negative and significant impact on TFP within the wholesale sector. For the other sectors, no significant difference can be observed in terms of the profit per employee.

7. Conclusion and recommendations

The focus of this study is twofold: 1) combining well-known approaches, such as index number theory and conditional DID, that is, matching and difference-in-difference, to perform an impact analysis with respect to TFP; 2) applying the method to investigate the effect on TFP of hiring individuals entitled to a wage subsidy. Only one previous study considers firm productivity effects as a result of labor subsidies. In contrast to the previous study, the present paper measures total factor productivity (TFP) using the Törnqvist TFP index rather than labor productivity. The profit per employee is also included as a second outcome. The analysis is performed on both the aggregated and the sector level.

The results on the aggregated level show a negative and statistically significant impact, both in terms of TFP and in terms of the profit per employee, driven by one sector. During the second year with treatment, the TFP deficit disappears, indicating that the subsidized firms are less productive only for a short period of time. On the sector level, the impact on TFP is only negative for wholesale, which also indicates a negative effect in terms of the profit per employee.

From a policy perspective, the analysis shows that subsidizing firms that employ persons entitled to a wage subsidy is necessary, since TFP decreases as a result of employing these individuals. To make the target group for wage subsidies at least not unattractive in terms of performance on the labor market, the subsidies need to compensate for this deficit. However, no productivity deficit can be observed after one year. Firms should therefore, at this point in time, be indifferent between employing previously subsidized individuals and employing non-subsidized ones, both without wage subsidies. Even though no TFP deficit can be observed, the profit per employee is significantly lower for firms that hire with a subsidy during the second year. The reason for this is thus not due to the subsidized employees. Instead, it results from these firms not managing to maximize their profit; that is, the MRP does not equalize the wage, as in case 5 in the theoretical framework.

Based on these findings, the policy maker should consider whether subsidizing the same firm is necessary for a longer period of time. From a firm perspective, it is of interest to keep an equally productive individual at a lower cost. However, the firms that continue to employ individuals with a subsidy become less profitable, without a TFP deficit. This implies that the lower profit is independent of who is employed (subsidized or non-subsidized workers). Hence, it is not an issue for the policy maker to handle.

Finally, the proposed combination of methods, namely the Törnqvist TFP index (or another index that does not rely on estimation) together with conditional DID, can be used in other applications to investigate causal relationships with respect to TFP.

References

Aigner D, Lovell CK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econom 6(1):21–37

Baghana RS (2010) Public R&D subsidies and productivity: evidence from firm-level data in Quebec. UNUMERIT Working Paper Series 2010-055

Becker GS (1964) Human capital theory. Columbia, New York

Behrenz L, Delander L, Frödin O, Mathieub C, Månsson J, Ottosson M (2015) Arbetsgivares perspektiv på subventionerade anställningar: En kartläggning (Employee perspectives of subsidized employment). Report to the Ministry of Labour, Stockholm

Bellmann L, Stephan G (2014) Which firms use targeted wage subsidies? An empirical note for Germany. J Labour Mark Res 47(1–2):165–175

Bennmarker H, Mellander E, Öckert B (2009) Do regional payroll tax reductions boost employment? Labour Econ 16(5):480–489

Bergström F (2000) Capital subsidies and the performance of firms. Small Bus Econ 14(3):183–193

Bernini C, Cerqua A, Pellegrini G (2017) Public subsidies, TFP and efficiency: a tale of complex relationships. Res Policy 46(4): 751–767

Bernini C, Pellegrini G (2011) How are growth and productivity in private firms affected by public subsidy? Evidence from a regional policy. Reg Sci Urban Econ 41(3):253–265

Bertrand M, Duflo E, Mullainathan S (2002) How much should we trust differences-in-differences estimates? No. w8841. National Bureau of Economic Research

Blackwell M, Iacus S, King G, Porro G (2009) Coarsened exact matching in Stata. Stata J 9(4):524

Bloom N, Eifert B, Mahajan A, McKenzie D, Roberts J (2013) Does management matter? Evidence from India. Q J Econ 128(1): 1–51

Brynjolfsson E, Hitt LM (2003) Computing productivity: firm-level evidence. Rev Econ Stat 85(4):793–808

Calmfors L (1994) Active labour market policy and unemployment: a framework for the analysis of crucial design features. OECD Economic Studies No. 22, Organisation for Economic Co-operation and Development

Calmfors L, Forslund A, Hemstrom M (2002) Does active labour market policy work? Lessons from the Swedish experiences. Working Paper no. 2002:4, Institute for Labour Market Policy Evaluation, Uppsala

Caves DW, Christensen LR, Diewert WE (1982) The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica: J Econom Soc:1393–1414

Chen YT, Hsu YC, Wang HJ (2014) Treatment effect stochastic frontier models with endogenous selection (No. 14-A006). IEAS Working Paper No. 14-A006, Institute of Economics, Academia Sinica, Taipei, Taiwan

Christensen LR, Jorgenson DW (1970) US real product and real factor input, 1929–1967. Rev Income Wealth 16(1):19–50

Cobb-Clark DA, Crossley T (2003) Econometrics for evaluations: an introduction to recent developments. Econ Rec 79(247):491–511

Dahlberg M, Forslund A (2005) Direct displacement effects of labour market programmes*. Scand J Econ 107(3):475–494

Diewert, WE (1976) Exact and superlative index numbers. J Econom 4(2): 115–145

Drechsler L (1973) Weighting of index numbers in multilateral international comparisons. Rev Income Wealth 19(1):17–34

Eriksson S, Rooth DO (2014) Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. Am Econ Rev 104(3):1014–1039

Fan S, Zhang X (2002) Production and productivity growth in Chinese agriculture: new national and regional measures. Econ Dev Cult Chang 50:819−838

Fisher I (1922) The making of index numbers: a study of their varieties, tests, and reliability (No. 1). Houghton Mifflin

Gerfin M, Lechner M, Steiger H (2005) Does subsidised temporary employment get the unemployed back to work? An econometric analysis of two different schemes. Labour Econ 12(6):807–835

Gibbons R, Katz L (1991) Layoffs and lemons. J Labor Econ 9:351–380

Greene W (2010) A stochastic frontier model with correction for sample selection. J Prod Anal 34(1):15–24

Harris R, Trainor M (2005) Capital subsidies and their impact on total factor productivity: firm-level evidence from Northern Ireland. J Reg Sci 45(1):49–74

Hartmann J (2004) Lohnkostenzuschüsse und Integration schwer vermittelbarer Personen in den ersten Arbeitsmarkt: Eine Evaluation mit Daten aus Betriebsbefragungen. Bundesagentur für Arbeit, Service-Haus, Geschäftsstelle für Veröff. Vol. 284. Nurnberg

Hujer R, Caliendo M, Radic D (2001) Estimating the effects of wage subsidies on the labour demand in West-Germany using the IAB establishment panel. ifo-Studien 47(2):163–199

Iacus SM, King G, Porro G (2012) Causal inference without balance checking. Political Anal 20(1):1–24

Jaenichen U, Stephan G (2011) The effectiveness of targeted wage subsidies for hard-to-place workers. Appl Econ 43(10):1209–1225

Kangasharju A (2007) Do wage subsidies increase employment in subsidized firms? Economica 74(293):51–67

Kieselbach T (2003) Long-term unemployment among young people: the risk of social exclusion. Am J Community Psychol 32(1–2):69–76

Kluve J (2006) The effectiveness of European active labour market policy. Discussion Paper no. 2018, IZA, Bonn

Koski H, Pajarinen M (2015) Subsidies, the shadow of death and labor productivity. J Ind Compet Trade 15(2):189–204

Kossoudji SA (1988) English language ability and the labor market opportunities of Hispanic and East Asian immigrant men. J Labor Econ 6(2):205–228

Kraft K (1998) An evaluation of active and passive labour market policy. Appl Econ 30(6):783–793

Kremer M (1993) The O-ring theory of economic development. Q J Econ:551–575

Kumbhakar SC, Wang H, Horncastle AP (2015) A practitioner's guide to stochastic frontier analysis using Stata. Cambridge University Press

Lam PL, Lam T (2005) Total factor productivity measures for Hong Kong telephone. Telecommun Policy 29(1):53–68

Laspeyres K (1871) Die Berechnung einer mittleren Waarenpreissteigerung. Jahrbücher für Nationalökonomie und Statistik 16(1):296–318

Lisi D, Malo MA (2017) The impact of temporary employment on productivity. J Labour Market Res (2017) 50: 91.

Lundberg S, Rose E (2000) Parenthood and the earnings of married men and women. Labour Econ 7(6):689–710

Managi S (2010) Productivity measures and effects from subsidies and trade: an empirical analysis for Japan's forestry. Appl Econ 42(30):3871–3883

Månsson J, Quoreshi AS (2015) Evaluating regional cuts in the payroll tax from a firm perspective. Ann Reg Sci 54(2):323–347

Montgomery JD (1991) Social networks and labor-market outcomes: toward an economic analysis. Am Econ Rev 81(5):1408–1418

Mora R, Reggio I (2015) didq: a command for treatment-effect estimation under alternative assumptions. Stata J 15(3):796–808

Newey WK, McFadden D (1994) Large sample estimation and hypothesis testing. Handbook of Econom 4:2111–2245

O'Donnell CJ (2012) Alternative indexes for multiple comparisons of quantities and prices (No. WP052012). School of Economics, University of Queensland, Australia

Paasche H (1874) Uber die Preisentwicklung der letzten Jahre nach den Hamburger Borsennotirungen. Jahrbücher für Nationalökonomie und Statistik 23:168–178

Pissarides CA (1992) Loss of skill during unemployment and the persistence of employment shocks. Q J Econ:1371–1391

Pries M, Rogerson R (2005) Hiring policies, labor market institutions, and labor market flows. J Political Econ 113(4):811–839

Regulation 1997:1275. Om särskilt anställningsstöd (About Special Recruitment Incentives)

Regulation 2006:1481. Om stöd för nystartsjobb (About support for New Start jobs)

Regulation 2007:420. Om ändring i förordningen (1997:1275) om anställningsstöd (About changes in regulation (1997:1275) of employment support)

Regulation 2009:26. Om ändring i förordningen (2006:1481) om stöd för nystartsjobb (About changes in regulation (2006:1481) of support for New Start jobs)

Regulation 2010:2028. Om ändring i förordningen (1997:1275) om anställningsstöd (About changes in regulation (1997:1275) of employment support)

Rotger GP, Arendt JN (2011, April) The effect of a wage subsidy on employment of the subsidised firm. In: Sixteenth Annual Meeting of Labor Economists (SOLE), Vancouver, Canada, April 29–30

Rungsuriyawiboon S, Coelli T (2004) Regulatory reform and economic performance in US electricity generation, Working Paper 06/2004, CEPA Working Papers Series, University of Queensland, School of Economics.

Serafica RB (1998) Was PLDT a natural monopoly? An economic analysis of pre-reform Philippine telecoms. Telecommun Policy 22(4):359–370

Skuras D, Tsekouras K, Dimara E, Tzelepis D (2006) The effects of regional capital subsidies on productivity growth: a case study of the Greek food and beverage manufacturing industry. J Reg Sci 46(2):355–381

Snower MDJ (1995) The simple economics of benefit transfers (No. 95). IMF Working Paper No. 95/5, International Monetary Fund. https://ssrn.com/abstract=883155

Solow RM (1957) Technical change and the aggregate production function. Rev Econ Stat 39(3):312–320

The Swedish Agency for Public Management (2011) Subventionerade anställningar. En kartläggning (Subsidised employment. An overview). The Swedish Agency for Public Management 2011:34, Stockholm

The Swedish Employment Office (2009) *Arbetsmarknadspolitiska program. Årsrapport 2008 (Labor market policy programs. Annual report 2008)*, Stockholm: The Swedish Employment office.

Törnqvist L (1936) The Bank of Finland’s consumption price index. Bank Finl Mon Bull 10:1–8

Wang HJ, Schmidt P (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. J Prod Anal 18(2):129–144

Young A (1812) An inquiry into the progressive value of money in England as marked by the price of agricultural products. Hatchard, Piccadilly

Appendix

A.1. Robustness and sensitivity analysis

To investigate stability, the elimination of the possible differences that might occur from firms claiming that they would never employ anyone with a wage subsidy is performed by changing the control group, as described in the methodology (Behrenz et al. 2015).[14] The results, in terms of TFP, are reported in table A1.[15]

Table A1 Robustness check with TFP as the dependent variable using the matched sample

Dependent variable:	**Firms with at least 1% subsidized employees**		
Törnqvist TFP index	Treated	After	DID
Matched periods -1 and 0. Treated periods 1 and 2	-0.003 (0.011)	0.014 (0.012)	-0.026 (0.015)
Matched periods -1 and 0. Treated periods 1 and 2. Year 2 excluded	-0.004 (0.015)	0.018 (0.016)	-0.036 (0.021)
Matched periods -1 and 0. Treated periods 1 and 2. Year 1 excluded	-0.010 (0.013)	0.001 (0.013)	-0.013 (0.018)

Standard errors are in parentheses; *** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

The DID estimate in the first row of table A1 indicates that firms hiring with subsidies experience a reduction of TFP by 2.6% as a result of employing workers with wage subsidies, which is consistent with table 5 in terms of magnitude. The difference is non-significant at the 5% level. However, since the magnitude remains constant, the same conclusion can be drawn as previously. The interpretation is then that a negative deficit is found, even when excluding firms that are not willing to employ individuals with subsidies.

Each DID estimate is similar in magnitude to those in table 5; that is, when the second year is excluded (second row), the coefficient is larger. When the first year post-treatment is excluded, no statistically significant difference in terms of TFP is found on any reasonable level of significance (the DID estimate in the third row). This non-significance and lower magnitude confirm the result obtained in table 4; specifically, there is no difference in terms of TFP in the medium term.

[14] As a placebo test, the same regression as in the main regression is estimated, that is, two years before and two years after. However, instead of using the correct post-treatment period, the difference between the two years in the pre-treatment period is used, that is, between -1 and 0, which represents 2009–2010 (when treatment is given in 2011–2012) aggregated with 2010–2011 (when treatment is given in 2012–2013), respectively. The results are non-significant as expected. The results can be provided on request.

[15] Here, there are 1162 firms in the control group and 1501 in the treated group for the matched data.

The wage subsidy should compensate for the decline in TFP, so the profit per employee is unchanged. Therefore, the profit per employee should remain the same as if the firm did not hire anyone with a wage subsidy if the subsidy is on the "correct" level, that is, case one in the theoretical framework. However, over- or undercompensation can also be present, generating positive or negative profit per employee effects, respectively. In table A2 the results for the robustness tests when the profit per employee is used as the outcome are reported.

Table A2 Robustness check with the profit per employee as the dependent variable using the matched sample

Dependent variable:	**Firms with at least 1% subsidized employees**		
Log(profit/employee)	Treated	After	DID
Matched periods -1 and 0. Treated periods 1 and 2			-**
Matched periods -1 and 0. Treated periods 1 and 2. Year 2 excluded			
Matched periods -1 and 0. Treated periods 1 and 2. Year 1 excluded			
Matched periods -1 and 0. Treated periods 1 and 2. Törnqvist TFP index as control			-**

*** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

In the first row of table A2, it is apparent that the DID coefficient for the profit per employee is negative and statistically significant. In comparison with the main results, presented in table 6, a statistically significant and negative impact is also observed when firms that are unwilling to hire with subsidies are excluded from the control group but at a lower level of significance.

No statistical significance can be observed when the second or first year is excluded from the post-treatment period, indicated in the column on the right, that is, the DID (the second and third rows). However, when interpreting it in comparison with the DID estimate of the first row, which is significant, one needs to bear in mind that the non-significant DID estimates are only based on one year as a post-treatment period. This generates large standard deviations, since the profit per employee is volatile from one year to another, in comparison with considering a mean of two years. Further, the number of observations is smaller than for the main results in table 6.[16] To control whether the deficit in the Törnqvist TFP index is the explanation for the decrease in the profit per employee, the same estimation as in the first row of table 10 is used, but the

[16] Examples of the reasons for a volatile result can be taxes, profit or losses in exchange rates and interest. It is thus more accounting effects in the short run that are smoothed out in the long run.

Törnqvist TFP index is included as an explanatory variable. The significance from row 1 does not change, which indicates that there is a negative and statistically significant effect on the profit per employee for treated firms that cannot be explained by the decline in TFP; that is, case 5 occurs.

A.2. Visual analysis of pre-treatment trends

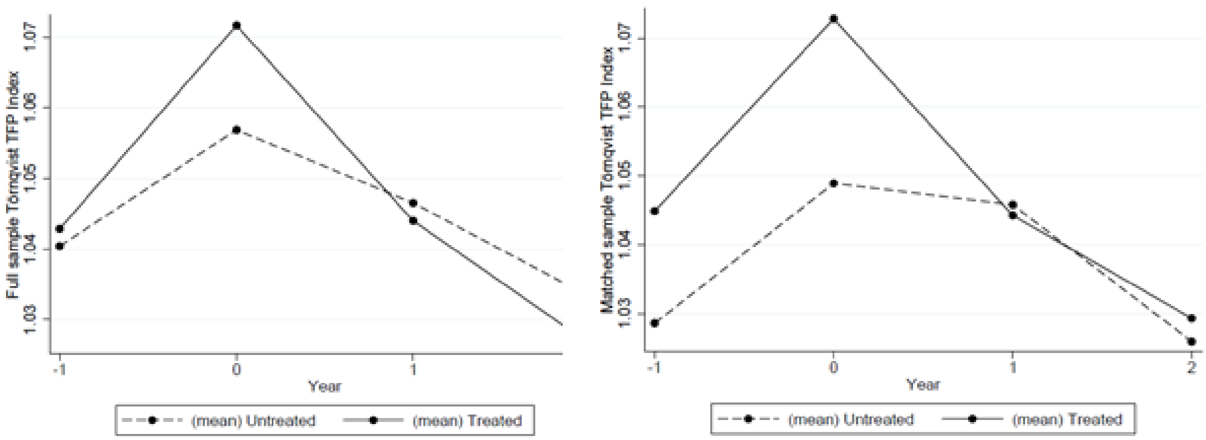

Fig. A1 Törnqvist TFP index using data from two years before treatment and a follow-up period of two years after treatment. The firms have at least 1% subsidized employees during these years and none previously. The full data are on the left-hand side and the matched data on the right-hand side

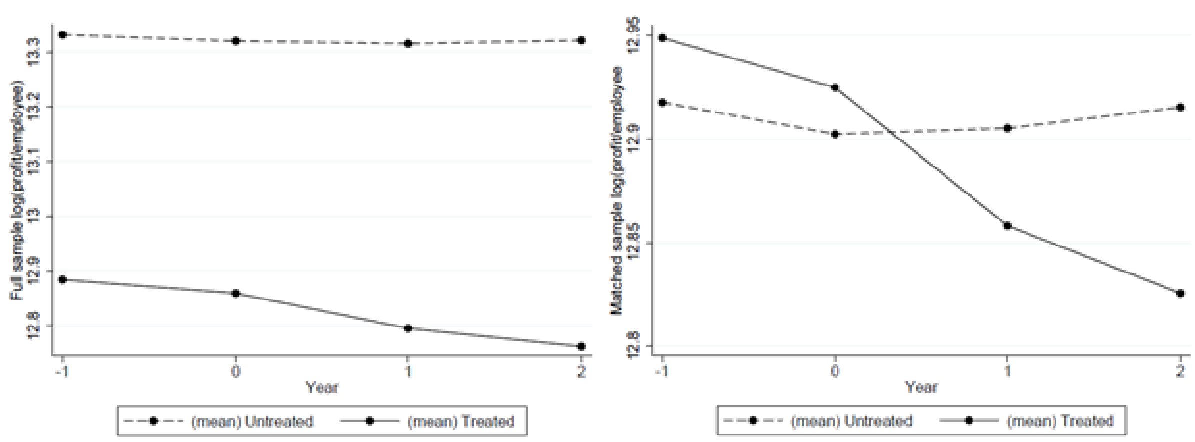

Fig. A2 Log(profit/employee) using data from two years before treatment and a follow-up period of two years after treatment. The firms have at least 1% subsidized employees in these years and none beforehand for all the firms aggregated. The full data are on the left-hand side and the matched data on the right-hand side

A.3. Descriptive statistics

Table A3 Descriptive statistics of the number of firms in total and the firms with wage subsidies

	Total number of included firms	**Share of firms with at least 1 subsidized employee**		**Number of full-time subsidies in an average firm**		**Ratio of full-time-equivalent subsidized to non-subsidized employees**	
Year	Obs.	Mean	S.D.	Mean	S.D.	Mean	S.D.
2008	72 692	0.085	0.279	0.091	0.675	0.012	0.105
2009	72 692	0.094	0.292	0.094	0.696	0.012	0.109
2010	72 692	0.118	0.323	0.127	0.907	0.012	0.099
2011	72 692	0.124	0.333	0.146	0.990	0.012	0.100
2012	72 692	0.114	0.318	0.127	0.995	0.012	0.102
2013	72 692	0.109	0.312	0.120	0.764	0.013	0.105

Appendix for referees only

Table R1 Test of the parallel pre-treatment trend assumption using the full data

H_0: Parallel pre-treatment trend	**Törnqvist TFP index**	**Log(profit/employee)**
All sectors aggregated	0.296	0.878
Manufacturing, mining and quarrying	0.997	0.892
Construction	0.589	0.883
Wholesale and retail trade	0.183	0.864
Transportation and storage	0.326	0.768
Accommodation and food service activities	0.238	0.391
Business service activities	0.138	0.996

Table R2 Effects on the profit per employee using the matched sample

Dependent variable:	**Firms with at least 1% subsidized employees**		
Log(profit/ employee)	Treated	After	DID
Matched periods -1 and 0. Treated periods 1 and 2	0.014 (0.008)	0.013*** (0.004)	-0.040*** (0.010)
Matched periods -1 and 0. Treated periods 1 and 2. Year 2 excluded	0.012 (0.009)	0.024*** (0.003)	-0.027** (0.013)
Matched periods -1 and 0. Treated periods 1 and 2. Year 1 excluded	0.015 (0.009)	0.031*** (0.004)	-0.041*** (0.013)
Matched periods -1 and 0. Treated periods 1 and 2. Geometric Young index used as control	0.013 (0.008)	0.013 (0.004)	-0.038 (0.009)

Standard errors are in parentheses; *** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

Table R3 Changes in the profit per employee for the matched sample. Two-year pre-treatment period and two-year follow-up period

Dependent variable:	**Firms with at least 1% subsidized employees**		
Log(profit/employee)	Treated	After	DID
Manufacturing, mining and quarrying	0.007 (0.022)	0.015** (0.006)	0.005 (0.032)
Construction	0.033 (0.029)	-0.002 (0.008)	-0.051 (0.041)
Wholesale and retail trade	-0.006 (0.015)	0.018*** (0.004)	-0.048** (0.022)

Transportation and storage	0.016 (0.025)	0.027*** (0.006)	-0.020 (0.036)
Accommodation and food service activities	0.035 (0.021)	0.000 (0.008)	-0.006 (0.029)
Business service activities	0.025 (0.021)	0.013 (0.006)	-0.056 (0.030)

Standard errors are in parentheses; *** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

Table R4 Coefficients of covariates. The reference categories are the mixed gender of the leader and the share without education or maximum primary schooling

Covariates	**Törnqvist TFP index**	**Log(profit/employee)**
Number of employees	-0.002*** (0.000)	-0.109*** (0.010)
Man as operational leader	-0.003 (0.008)	-0.001 (0.024)
Woman as operational leader	-0.004 (0.007)	0.104*** (0.026)
Share with college education	0.082*** (0.020)	0.105*** (0.021)
Share with post-secondary schooling	0.088*** (0.028)	0.123*** (0.015)
Share of young	-0.033*** (0.018)	-0.378*** (0.065)
Share of older	-0.011 (0.007)	0.214*** (0.046)
Share of immigrants	0.022 (0.010)	-0.002 (0.023)
Share of women	-0.028*** (0.007)	-0.136*** (0.034)
Constant	0.995*** (0.015)	13.193*** (0.072)

Standard errors are in parentheses; *** indicates significance at the 1% level and ** at the 5% level.

Table R5 Placebo test for the Törnqvist TFP index

Dependent variable:	**Firms with at least 1% subsidized employees**		
Törnqvist TFP index	Treated	After	DID
All sectors	0.001 (0.007)	0.047*** (0.002)	0.008 (0.009)
Manufacturing, mining and quarrying	0.010 (0.018)	0.093*** (0.005)	-0.000 (0.025)
Construction	-0.029 (0.017)	0.008 (0.005)	0.006 (0.023)
Wholesale and retail trade	0.010 (0.014)	0.064*** (0.004)	-0.025 (0.019)

Transportation and storage	0.019 (0.023)	0.048 (0.006)	-0.004 (0.033)
Accommodation and food service activities	-0.019 (0.023)	0.023** (0.009)	0.018 (0.033)
Business service activities	0.001 (0.017)	0.047 (0.005)	-0.003 (0.024)

Standard errors are in parentheses; *** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

Table R6 Placebo test for log(profit/employee)

Dependent variable:	**Firms with at least 1% subsidized employees**		
Log(profit/employee)	Treated	After	DID
All sectors	0.006 (0.008)	0.011*** (0.002)	-0.011 (0.011)
Manufacturing, mining and quarrying	-0.005 (0.020)	0.012*** (0.006)	0.004 (0.028)
Construction	0.001 (0.021)	0.008 (0.006)	-0.004 (0.029)
Wholesale and retail trade	-0.010 (0.012)	0.008** (0.004)	-0.005 (0.017)
Transportation and storage	-0.038 (0.029)	0.016** (0.007)	-0.021 (0.041)
Accommodation and food service activities	0.087*** (0.025)	-0.004 (0.010)	0.019 (0.035)
Business service activities	0.020 (0.021)	0.007 (0.006)	-0.010 (0.029)

Standard errors are in parentheses; *** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

Table R7 Robustness check with the profit per employee as the dependent variable using the matched sample

Dependent variable:	**Firms with at least 1% subsidized employees**		
Log(profit/employee)	Treated	After	DID
Matched periods -1 and 0. Treated periods 1 and 2	-0.002 (0.011)	0.003 (0.012)	-0.035** (0.016)
Matched periods -1 and 0. Treated periods 1 and 2. Year 2 excluded	-0.006 (0.011)	0.018 (0.012)	-0.019 (0.016)
Matched periods -1 and 0. Treated	-0.001	0.022	-0.033

periods 1 and 2. Year 1 excluded	(0.012)	(0.013)	(0.017)
Matched periods -1 and 0. Treated periods 1 and 2. Törnqvist TFP index used as control	-0.002 (0.011)	0.003 (0.012)	-0.034** (0.016)

Standard errors are in parentheses; *** indicates significance at the 1% level and ** at the 5% level. The covariates are presented in table 2.

The Swedish wage subsidy scheme

Entry Recruitment Incentives (*Instegsjobb*)

Entry Recruitment Incentives are given to employers with the purpose of stimulating employment for individuals who find it hard to obtain a job in the regular labor market. An Entry Recruitment Incentive can be given to an individual who is at least 20 years old, who is unemployed and who received a residence permit during the last 36 months. It can be granted for 24 months, but if the employment is more than 50% of full-time, the maximum time is 6 months (Regulation 2007:420). From the beginning of 2013, this time is 12 months. The support cannot be given to immigrants after a time period of 36 months after the individual received a residence permit, if the individual does not qualify on other grounds (The Swedish employment office 2013). The maximum level of subsidy with Entry Recruitment Incentives is 75% of the salary cost and at most 750 SEK per day (The Swedish employment office 2011). From the beginning of 2011, the subsidy is 80% of the salary cost with a maximum level of 800 SEK (Regulation 2010:2028). The subsidy is only given if the employer guarantees that the wage and other advantages are in line with the collective agreements or equivalent within the sector. If a firm has fired anyone based on job scarcity during a period of 9 months, it is not allowed to employ workers with subsidies (Regulation 1997:1275).

Special Recruitment Incentives (särskilt anställningsstöd)

An individual can be assigned to Special Recruitment Incentives if he or she has been participating in the job and development guarantee for 6 months. A person who was assigned to the job and development guarantee after participating in the job guarantee for the young for 15 months can be assigned to the Special Recruitment Incentives at the same time as entering the job and development guarantee (Regulation 2011:586). When a Special Recruitment Incentive has finished, the individual can be assigned again to be employed with a particular recruitment incentive after applying for jobs for 3 months, but for the same employer 6 months are required (Regulation 2007:920).

A Special Recruitment Incentive for individuals who have participated in the job and development guarantee is 85% of the salary cost and mostly 750 SEK per day (Regulation 2008:271), but, since the beginning of 2011, it has been 890 SEK per day (Regulation

2010:2028). From the beginning of 2012, an amount of 50 SEK per day can be given for supervising the employee, but, for people who received a Special Recruitment Incentive from the job and development guarantee, this amount is 150 SEK per day for 3 months and then 100 SEK per day (The Swedish employment office 2013). The support can be given for a maximum of 12 months (The Swedish employment office 2008), but after the period studied here, it was decided that it can be prolonged (The Swedish employment office 2013).

New Start jobs (nystartsjobb)

The purpose of New Start jobs is to stimulate employers to hire individuals who have not had a job for a long time (Regulation 2008:1437). The support is given to an individual who is at least 20 years old and 1) has been unemployed and registered at the employment office for at least one year or participating in a labor market policy program, 2) has had a protected job by Samhall and has been registered at the employment office for one year, 3) has been unemployed and received economic aid and been registered at the employment office for one year, 4) has received sickness payment, rehabilitation payment or sickness or activity payment and has been registered at the employment office for at least one year, 5) a combination of 1–4 and has been registered as unemployed, 6) has spent at least 6 months of the last 9 months with the same conditions as included in 1–4 and registered as unemployed if the individual is 20–24 years old or 7) is a foreigner and received a residence permit during the last 3 years and has been registered as unemployed at the employment office (Regulation 2006:1481). If the individual does not fulfil any of the previous conditions but is included in the job and development guarantee, the support lasts for a maximum of one year (Regulation 2007:421). An individual can also be eligible for support if he or she has been unemployed or part-time for two years, a combination of full-time and part-time or participated in a labor market policy program (Regulation 2007:1366). From March 2009 the youngest age group is from 20 to 25 years old instead, and the regular rules are applicable to individuals who are at least 26 years old (Regulation 2009:26). For individuals who are at least 55 at the beginning of the year, 6 months of unemployment during the last 9 months are required to be eligible for New Start jobs from July 2010 (Regulation 2010:396).

A subsidy is only given if the employer has a collective agreement or without it gives the employee equivalent advantages. The support is given for the same time as the individual has been away from the labor market, or has had a protected job, but for a maximum of 5 years. If the person is more than 55 years old, the support can however be given for 10 years, but only until he or she reaches 65 years of age. For individuals who are at least 20 but below 25 years old (after March 2009 this is 26), the subsidy is given for 1 year maximum, but this is not applicable if 3 of the criteria in the last paragraph are fulfilled (Regulation 2009:26). Foreigners who are at least 20 years old have the possibility to receive a subsidy for 3 years from the day when they received their residence permit but only until they reach the age of 65 years (Regulation

2006:1481). Extra support can also be given if a firm employs an individual eligible for support according to point 4 in the last paragraph (Regulation 2008:1437).

The support in Regulation (2006:1481) is 1 payroll tax as well as the general wage fee that the employer needs to pay. From July 2009 the support is 2 payroll taxes for individuals who are at least 26 years old, and for individuals between 20 and 26 years old it is 1 payroll tax. If the younger age group became eligible for New Start jobs due to sickness payment, rehabilitation payment or sickness or activity payment, the support is 2 payroll taxes (Regulation 2009:26). The employers need to apply for the subsidy with specification of conditions of employment and the employee needs to give the necessary information to the employment office to know that he or she is eligible for a subsidy within New Start jobs. The decision on support is given for 1 year at a time, and the decision needs to be reconsidered if there are changed conditions. The money is given as a credit on the tax account in the month after the wage is paid (Regulation 2006:1481). Now there is a maximum salary to receive the compensation, which is 22,000 SEK per month (Regulation 2015:41); during the period studied here, there is no limit like this.

References for the Swedish subsidy programs

Regulation 2006:1481. Om stöd för nystartsjobb (About support for New Start jobs)

Regulation 2007:420. Om ändring i förordningen (1997:1275) om anställningsstöd (About changes in regulation (1997:1275) of employment support)

Regulation 2007:421. Om ändring i förordningen (2006:1481) om stöd för nystartsjobb (About changes in regulation (2006:1481) of support for New Start jobs)

Regulation 2007:920. Om ändring i förordningen (1997:1275) om anställningsstöd (About changes in regulation (1997:1275) of employment support)

Regulation 2007:1366. Om ändring i förordningen (2006:1481) om stöd för nystartsjobb (About changes in regulation (2006:1481) of support for New Start jobs)

Regulation 2008:271. Om ändring i förordningen (1997:1275) om anställningsstöd (About changes in regulation (1997:1275) of employment support)

Regulation 2008:1437. Om ändring i förordningen (2006:1481) om stöd för nystartsjobb (About changes in regulation (2006:1481) of support for New Start jobs)

Regulation 2009:26. Om ändring i förordningen (2006:1481) om stöd för nystartsjobb (About changes in regulation (2006:1481) of support for New Start jobs)

Regulation 2010:396. Om ändring i förordningen (2006:1481) om stöd för nystartsjobb (About changes in regulation (2006:1481) of support for New Start jobs)

Regulation 2010:2028. Om ändring i förordningen (1997:1275) om anställningsstöd (About changes in regulation (1997:1275) of employment support)

Regulation 2011:586. Om ändring i förordningen (1997:1275) om anställningsstöd (About changes in regulation (1997:1275) of employment support)

Regulation 2015:41. Om ändring i förordningen (2006:1481) om stöd för nystartsjobb (About changes in regulation (2006:1481) of support for New Start jobs)

The Swedish Employment Office (2009) Arbetsmarknadspolitiska program. Årsrapport 2008 (Labor market policy programs. Annual report 2008), Stockholm: The Swedish Employment office

The Swedish Employment Office (2011) Arbetsmarknadspolitiska program. Årsrapport 2010 (Labor market policy programs. Annual report 2010), Stockholm: The Swedish Employment office

The Swedish Employment Office (2013) Arbetsmarknadspolitiska program. Årsrapport 2012 (Labor market policy programs. Annual report 2012), Stockholm: The Swedish Employment office

A bootstrapped Malmquist index applied to Swedish district courts

Pontus Mattsson[♣]
Jonas Månsson[*]
Christian Andersson[†]
Fredrik Bonander[•]

Abstract: This study measures the total factor productivity (TFP) of the Swedish district courts by applying data envelopment analysis (DEA) to calculate the Malmquist productivity index of 48 Swedish district courts from 2012 to 2015. In contrast to the limited international literature on court productivity, this study uses a fully decomposed Malmquist index. A bootstrapping approach is further applied to compute confidence intervals for each decomposed factor of TFP. The findings show a 0.7% average decline of TFP annually. However, a substantial variation between years can be observed in the number of statistically significant courts below and above unity. The averages of the components show that the negative impact is mainly driven by pure technical regress. Large variations are also observed over time where the small courts have the largest volatility. Two recommendations are: 1) that district courts with negative TFP growth could learn from those with positive TFP growth and 2) that a back-up force could be developed to enhance flexibility.

Keywords: Bootstrap; Data envelopment analysis (DEA); District courts; Malmquist index; Total factor productivity (TFP)

JEL Classification D24 · K49 · O33

Acknowledgement We are grateful for received comments during the 15th International Conference on DEA in Prague and the seminar at Linnaeus University.

[♣] Department of Economics and Statistics, Linnaeus University, 351 95 Växjö, Sweden. E-mail: pontus.mattsson@lnu.se

[*] Department of Economics and Statistics, Linnaeus University, 351 95 Växjö, Sweden and Swedish National Audit Office (SNAO), Nybrogatan 55, 114 90 Stockholm, Sweden. E-mail: jonas.mansson@lnu.se

[†] Swedish National Audit Office (SNAO), Nybrogatan 55, 114 90 Stockholm, Sweden. Email: christian.andersson@riksrevisionen.se

[•] Swedish National Audit Office (SNAO), Nybrogatan 55, 114 90 Stockholm, Sweden. Email: fredrik.bonander@riksrevisionen.se

1. Introduction

Efficiency and productivity in the public sector are a major concern for most parliaments and governments. It is explicitly stated in the Swedish Budget Law (2011:203), 1 chapter 3§, that all state services should be provided with a high level of efficiency. When dealing with state-provided production, this means that it should be efficient, that is, use the least amount of inputs at the given output level or produce the maximum amount of production at the given resource level. The motivation for studying productivity development is therefore, from a policy perspective, straightforward. Swedish applications can for instance be found for higher education (Andersson et al., 2017a), employment offices (Andersson et al., 2014), day care (Bjurek et al., 1992), elderly care, primary and secondary education (ESO, 2016). Economic development is concluded to benefit from an efficient judicial system internationally (see e.g. Feld & Voigt, 2003; Messick, 1999). Furthermore, there is an ongoing debate in the media as well as in research regarding punishments, criminality, and the judicial system in Sweden (see e.g. Sturup et al., 2017; Tyrefors Hinnerich et al., 2017). For instance, the share of solved suspicious declined from 2004 to 2014 according to the Swedish National Council for Crime Prevention (2015).

The Swedish Government has launched a number of reforms for the district courts during the last 20 years with the major objective of increasing efficiency and productivity while maintaining a high degree of law and order (The Swedish Agency for Public Management, 2007). One such reform has targeted the size of the courts based on the assumption that scale advantages exist. In 1999 there were 96 district courts in Sweden, but today only 48 exist. Despite the reforms the Swedish National Court Administration (SNCA) reports indications that the productivity has declined. However, problems can occur in this study, since productivity is measured by partial measures, in this case labor productivity, which ignores substitution between inputs. At the same time, new technologies have been introduced in the courts, such as the possibility of conducting hearings through video conferences, and these changes are not captured by the productivity measures used.

The aim of this study is to measure the Swedish district courts' total factor productivity (TFP) from 2012 to 2015. To compute TFP, the data envelopment analysis (DEA) framework proposed by Charnes, Cooper, and Rhodes (1978) is used. TFP is measured by the Malmquist productivity index (MPI) of Färe, Grosskopf, Lindgren, and Roos (1989, 1992). Following Simar and Wilson (1998a), the MPI is decomposed into four parts: changes in 1) pure technical efficiency, 2) scale efficiency, 3) pure technology, and 4) the scale of the technology. A commonly known issue with all types of DEA analysis is that no statistical inference is possible (Simar & Wilson, 2000). In this study a bootstrap approach is used to determine the confidence intervals of the different components presented above (see Efron, 1979). Another issue with DEA is the influence of outliers (see e.g. Kapelko & Oude Lansink, 2015). The analysis of outliers is to a large extent omitted in previous studies on court performance. In this study an outlier detection analysis is performed to investigate whether the results depend on a few extreme observations.[17]

The findings indicate a decline in TFP of 0.7% on average. However, a substantial variation between courts and between years is present. Between 26% and 68% of the courts have a significantly negative change in TFP, while the share of courts with a significantly positive TFP change is 9–41%. Looking at the components, the negative impact is driven by pure technical regress of 3.1% in 2014–2015. During this period the number of courts with a significant decline in TFP is large in comparison with the rest of the years. Furthermore, the number of courts with a significantly positive TFP change is lower at the end of the studied period. The correlation analysis concludes that the rate of change in the caseload has a significantly positive correlation with TFP, which indicates flexibility problems.

The paper is organized as follows. Section 2 provides a brief summary of the Swedish district court system. Section 3 presents the previous TFP and efficiency literature regarding courts. Section 4 describes the methodology and section 5 the data, including outlier detection. The results are reported in section 6. Finally, section 7 concludes and discusses the policy implications.

2. The Swedish judicial system – A short description

The Ministry of Justice is responsible for matters that are related to the judicial system, which include legislation on the fields of civil law and criminal law, for example.[18] However, it is not allowed to interfere in the day-to-day work, since the aim of the Swedish legal justice system is to provide fair trials. This requires independence and autonomy between courts in relation to the Parliament, Government, and other authorities. The judicial process differs depending on whether it is a criminal case, a civil case, or a matter. The different processes are described in figure 1. Each stage has the general purpose of dealing with cases and matters in an efficient manner in compliance with the rule of law.

[17] This study is a development of the report by Andersson et al. (2017b).
[18] This section builds on information presented by the Ministry of Justice (2015).

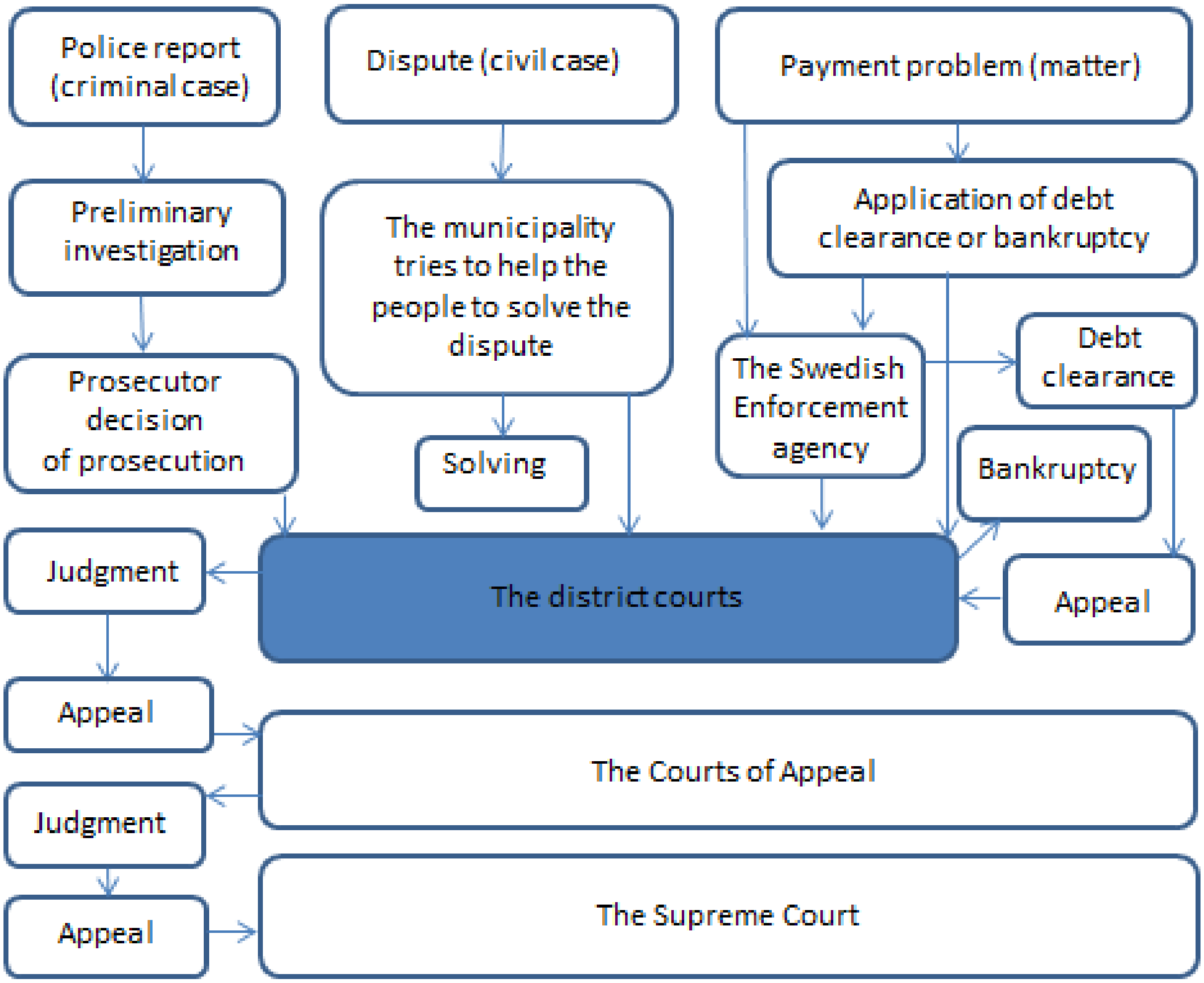

Figure 1: Description of the judicial process

Criminal cases, to the left in figure 1, are first handled by the police. These cases start with a police report followed by a preliminary investigation. In the next step, the case can either be closed or be sent to the prosecutor, who will decide whether the case will be prosecuted or not. If the case continues to prosecution, it will end up in the district court. A dispute is first handled by the municipalities, and if it is not solved it becomes a civil case in the district court. These civil cases are related to a dispute between individuals or business firms. Payment problems are handled by the Swedish Enforcement Agency, which can decide on debt clearance. Another potential route is bankruptcy, in which the decision has to be taken by the district court; this is also the case if the firm itself applies for bankruptcy.

There are three different types of courts that build up the Swedish court system, namely the general courts, the administrative courts, and special tribunals. The general courts consist of the district courts, the Courts of Appeal, and the Supreme Court. Each of the instances is important, since different instances provide possibilities to appeal to achieve a fair trial, which is a fundamental right in any legal justice system. The Supreme Court, that is, the last instance, has the main mission to provide the district court system with legal practice to enhance the uniformity of actions in legal decisions.

This study focuses on the district courts, which have the mission to serve as the first instance in the legal system. Each district court mainly handles cases related to their catchment area, which corresponds to the surrounding geographical area. However, there are five courts that specialize in land and environment cases. These courts deal for example with environmental and water issues, property registration, and building matters. Within each court there are Chief Judges, Senior Judges, and Judges who are considered as permanent judges, the former being the head of the court. Each of these is appointed by the government. There are also law clerks who work as non-permanent judges. Their work tasks normally consist of preparing cases, and they are most likely to be recent law graduates. Finally, Lay Judges have experience from other occupations and politics and are chosen by the Municipal Council but not educated in law. They work as judges for a period of four years.

3. Literature review

There is literature regarding the labor productivity of courts (see e.g. Blank, van der Ende, van Hulst, & Jagtenberg, 2004; SNCA, 2015) but only a limited amount of literature considering court TFP. Kittelsen and Førsund (1992) are the first to investigate TFP change over time. The efficiency scores of Farrell (1957) are used to calculate the MPI, which is decomposed into change in efficiency and change in technology, with the first year as the base (Caves, Christensen, & Diewert, 1982; Malmquist, 1953). In terms of the decomposed factors, the catching-up was 4% and the technology shift 2% from 1983 to 1988. Kittelsen and Førsund (1992) perform an outlier detection analysis in which the MPI and its components are shown in a histogram with the labor share on the x-axis. Based on the diagrams, three courts are considered to be outliers due to a large improvement or decline in TFP.[19] Fauvrelle and Almeida (2016) calculate the Malmquist index and decompose it into technical change and efficiency change. Following Färe, Grosskopf, Norris & Zhang (1994), efficiency change is further decomposed into a pure efficiency change and a scale component.[20] The results show on average a positive TFP change of 1.5% decomposed into a decline of 1.7% in technical change (TC), a pure

[19] Kittelsen and Førsund (1992) also note that an outlier with an exceptionally low efficiency score of 0.35 was present. This court is highly specialized, but it is argued that very efficient courts may be more problematic.

[20] In the USA a PhD thesis by Ferrandino (2010) investigates TFP change and decomposes it into the same components as Fauvrelle and Almeida (2016).

efficiency change of 3.3 %, and a scale efficiency change of 0.7%.[21] Both Fauvrelle and Almeida (2016) and Kittelsen and Førsund (1992) use the averages of TFP change and decompose them into at most three components. However, neither of them investigates whether the changes are statistically significant.

While only a few studies exist on TFP change, efficiency is measured more extensively. This research is important for this study, since it deals with the question of which inputs and outputs to choose to measure performance.[22] Lewin, Morey, and Cook (1982) are the first to investigate inefficiency in district courts using DEA.[23] Lewin et al. (1982), as well as all the other studies, use the number of employees as input. In some studies employees are measured as the number of judges (see e.g. Ferrandino, 2014; Finocchiaro Castro & Guccio, 2014). In other studies the personnel are separated into judges and office staff, for example (see e.g. Major, 2015; Santos & Amado, 2014). The caseload of the court is another input included in some studies. The caseload consists of pending and new cases, that is, the justice demand (see e.g. Kittelsen & Førsund, 1992; Schneider, 2005). For instance, Nissi and Rapposelli (2010) and Schneider (2005) argue the importance of including the caseload, since an underestimation of productivity will occur because the employees cannot perform their job without incoming cases. However, this is a slightly contradicting argument when analyzing TFP, which will be described in the methodology. Moreover, Beenstock and Haitovsky (2004) argue that individual productivity increases if the work pressure is high. However, the caseload can also, as Kim and Min (2016) argue, correlate negatively with quality. For example, if the caseload is low, more time can be spent on each case, which on average generates a more precise judgement. The outputs normally consist of the number of decided cases (see e.g. Nissi & Rapposelli, 2010). In some studies cases are separated into type, for example criminal cases and civil cases (see e.g. Finocchiaro Castro & Guccio, 2016). However, due to data limitations, the studies cannot separate the outputs within each category based on the resources spent on each case. This aggregation equalizes e.g. a murder with a car crime. Different types of crimes require different amounts of resources due to their dissimilarity in complication. A problem with this will occur, in court performance analysis, if there are differences in the mixture of crime types between courts.

Quality variables are argued to be important in some studies (see e.g. Yeung & Azevedo, 2011). Some attempts to investigate the impact of quality variables on performance can be found in the

[21] Moreover, an investigation is performed of what is described as "… productivity or technical efficiency …" (Schneider, 2005, p. 133). It is in fact technical efficiency that is calculated but referred to as court productivity each time, except the previously cited occasion. Furthermore, Schneider (2005) investigates the determinants of efficiency, such as the share of PhD judges and the share of judges older than 60, among others.

[22] For a general descriptive survey of judicial efficiency, see Voigt (2016).

[23] Other previous input–output-oriented studies exist concerning court performance (see Nardulli, 1978) but not concerning technical efficiency. There is also older research on courts that for instance focuses on organizational perspectives (see e.g. Eisenstein & Jacob, 1977; Feeley, 1973).

literature. Examples are judges' salaries and education, of which the former have a significantly positive effect on efficiency (Deyneli, 2012). Furthermore, Schneider (2005) concludes that more PhD holders as judges increase the efficiency. Finally, Falavigna et al. (2015) use court delay, as an undesirable output, in a directional distance function.[24] All these studies focus on technical efficiency, which is basically a measurement of similarity. For instance, Espasa and Esteller-Moré (2015) argue that the efficiency can be high even if the courts perform poorly as long as they are congested. Thus, this is not a good measurement of performance improvements over time; for example, lower inefficiency over time could occur due to a decline in performance of the best district courts.

To sum up, there is no research regarding TFP in Sweden and very little literature internationally. Furthermore, the international studies do not, due to data limitations, investigate the potential heterogeneity in resource spending within the output categories. Moreover, statistical inference is ignored and TFP is at most decomposed into three components.

4. Methodology

Different approaches can be applied in productivity and efficiency studies. Stochastic frontier analysis (SFA) is a widely used parametric methodology (see e.g. Kumbhakar and Lovell (2003) or Krüger (2012) for overviews). SFA has the advantage of allowing for statistical noise directly. However, the disadvantage is that it requires a specific functional form. Another option is the DEA approach, which has the advantages that it relies on few assumptions and can handle multiple outputs and inputs. Furthermore, DEA is relevant when analyzing the public sector, in which the outputs are not sold on the market (Førsund, 2016). However, DEA also has some disadvantages. Firstly, it does not give information about inference. To some extent this can be handled by using resampling methods, such as the bootstrap procedure proposed by Simar and Wilson (1998b). The second disadvantage of DEA is its sensitivity to outliers. This shortage is, to a large extent, neglected in previous literature on court performance.

Outlier detection

There is no optimal procedure to detect outliers, since no generally accepted definition of an outlier can be found (Davies & Gather, 1993). However, plenty of methods are applied in different areas. For example, the outlier detection method by Wilson (1993) is useful when the data checking is costly, that is, when the data set is large. Another approach, proposed by Simar (2003), focuses on detecting what the author refers to as super-efficient observations. Kapelko

[24] The inefficiency is between 9% and 11% in Norway, Denmark, and Germany (Kittelsen & Førsund, 1992; Rigsrevisionen, 2000; Schneider, 2005) and between 18% and 52% in Italy, Portugal, and Brazil (Nissi & Rapposelli, 2010; Pedraja-Chaparro & Salinas-Jimenez, 1996; Peyrache & Zago, 2016; Azevedo & Yeung, 2011).

and Oude Lansink (2015) use a specific deviation from the median. This can be achieved in different ways, for example including each input and output separately, using the ratio between outputs over inputs, or considering the differences in the rate of change over time. The first is just a measure of size; that is, a large (small) court has a large (small) amount of inputs and outputs, making it likely to be eliminated. The focus in this paper is TFP change; therefore, we focus on the deviation from median based on the rate of change in outputs as an outlier detection procedure presented in a figure of boxplots. Finally, when a potential outlier is identified, a closer look at the specific observation should be taken to produce arguments for why it is an outlier (Simar, 2003).

DEA and the Malmquist productivity index

The point of reference can be taken either from an input perspective, that is, minimize the inputs to produce a given level of output, or from an output perspective, that is, maximize the output given the level of inputs. As in most studies of district courts, an output-based perspective is assumed. The production technology in time period *t* for the 48 Swedish district courts is defined as:

[1] $$\mathrm{S}_t = \{(\mathbf{x}_{\mathrm{i}}^t, \mathbf{y}_{\mathrm{i}}^t) | \mathbf{x}_i \text{ can produce } \mathbf{y}_i \text{ at time } t\},$$

where S_t represents the technology. Each court, *i*, uses a vector of inputs, $\mathbf{x}^t$, to produce a vector of outputs, $\mathbf{y}^t$, in period *t*. Using the output distance function,[25] the technical efficiency can, in time period *t*, be written as:

[2] $$D_{\mathrm{o}}^t(\mathbf{x}_i^t, \mathbf{y}_i^t) = inf\{\theta: (\mathbf{x}_i^t, \ \mathbf{y}_i^t / \theta) \in S^t\},$$

where θ is a scalar, the o-subscript represents output distance function and the distance is $D_{\mathrm{o}}^t(\mathbf{x}_{\mathrm{i}}^{\mathrm{t}}, \mathbf{y}_{\mathrm{i}}^{\mathrm{t}}) \leq 1$. From the distance function, a measure of technical efficiency (TE) is obtained as $TE = 1/D_{\mathrm{o}}^t(\mathbf{x}_i^t, \mathbf{y}_i^t)$. If TE is equal to unity, the court is on the frontier, meaning that it is technically efficient. However, if TE is larger than unity, the court is inefficient, for instance TE equal to 1.1 means that the output can be increased by 10%, given the amount of inputs. To calculate the standard Malmquist index introduced by Caves et al. (1982), the same calculation needs to be performed for the following period, that is, *t*+1. This is shown in equation 3.[26]

[3] $$\mathrm{D}_{\mathrm{o}}^t(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) = \inf\{\theta: (\mathbf{x}_i^{t+1}, \ \mathbf{y}_i^{t+1}/\theta) \in S^t\}.$$

This can be written similarly in the VRS case, which is defined as $\Delta_o^t(\mathbf{x}_i^{t+1}, \ \mathbf{y}_i^{t+1})$. Using equations 2 and 3, assuming the technology of period *t* as the reference, Caves et al. (1982) define the MPI as:

[25] See Shephard (1970) for the output distance function. The input distance function is defined in Shephard (1953).

[26] The name comes from the early work by Malmquist (1953).

$$[4] \qquad \mathrm{M}_{\mathrm{o}}^{\mathrm{t}}(\mathbf{x}_i^t,\ \mathbf{y}_i^t,\ \mathbf{x}_i^{t+1},\ \mathbf{y}_i^{t+1}) = \frac{\mathrm{D}_{\mathrm{o}}^t(\mathbf{x}_{\mathrm{i}}^{t+1},\ \mathbf{y}_i^{t+1})}{\mathrm{D}_{\mathrm{o}}^t(\mathbf{x}_{\mathrm{i}}^t,\ \mathbf{y}_{\mathrm{i}}^t)},$$

where the MPI is the ratio of the output distance functions in each period, respectively. To avoid an assumption of the benchmark technology, equation 4 is also defined with *t*+1 as the period of reference. The MPI is then defined as the geometric mean of the two indexes.[27]

Decomposition of TFP

Decomposition of the productivity index is first proposed by Nishimizu and Page (1982), who define TFP as the sum of the efficiency change and technical change.[28] The geometric mean of two indexes is, following Caves et al. (1982) and Färe et al. (1989, 1992), obtained by rewriting equation 4 as:

$$[5] \qquad \mathrm{M}_{\mathrm{o}}(\mathrm{x}_{\mathrm{i}}^{\mathrm{t}},\ \mathrm{y}_{\mathrm{i}}^{\mathrm{t}}, \mathrm{x}_{\mathrm{i}}^{\mathrm{t+1}},\ \mathrm{y}_{\mathrm{i}}^{\mathrm{t+1}}) = \left[\frac{\mathrm{D}_{\mathrm{o}}^{\mathrm{t}}(\mathrm{x}_{\mathrm{i}}^{\mathrm{t+1}},\ \mathrm{y}_{\mathrm{i}}^{\mathrm{t+1}})}{\mathrm{D}_{\mathrm{o}}^{\mathrm{t}}(\mathrm{x}_{\mathrm{i}}^{\mathrm{t}},\ \mathrm{y}_{\mathrm{i}}^{\mathrm{t}})} \times \frac{\mathrm{D}_{\mathrm{o}}^{\mathrm{t+1}}(\mathrm{x}_{\mathrm{i}}^{\mathrm{t+1}},\ \mathrm{y}_{\mathrm{i}}^{\mathrm{t+1}})}{\mathrm{D}_{\mathrm{o}}^{\mathrm{t+1}}(\mathrm{x}_{\mathrm{i}}^{\mathrm{t}},\ \mathrm{y}_{\mathrm{i}}^{\mathrm{t}})}\right]^{1/2},$$

where the distance functions are defined assuming CRS. Based on the geometric mean defined in equation 5, the MPI can be decomposed into technical change (TC) and efficiency change (EC). Following Simar and Wilson (1998a) the decomposition is, while allowing for VRS, written as[29]:

$$[6] \qquad \mathrm{M}_{\mathrm{o}}(\mathbf{x}_i^t,\ \mathbf{y}_i^t, \mathbf{x}_i^{t+1},\ \mathbf{y}_i^{t+1}) = \left(\frac{\Delta_{\mathrm{o}}^{t+1}(\mathbf{x}_i^{t+1},\ \mathbf{y}_i^{t+1})}{\Delta_{\mathrm{o}}^t(\mathbf{x}_i^t,\ \mathbf{y}_i^t)}\right) \times \left(\frac{\Delta_{\mathrm{o}}^t(\mathbf{x}_i^{t+1},\ \mathbf{y}_i^{t+1})\Delta_{\mathrm{o}}^t(\mathbf{x}_i^t,\ \mathbf{y}_i^t)}{\Delta_{\mathrm{o}}^{t+1}(\mathbf{x}_i^{t+1},\ \mathbf{y}_i^{t+1})\Delta_{\mathrm{o}}^{t+1}(\mathbf{x}_i^t,\ \mathbf{y}_i^t)}\right)^{1/2} = \mathrm{EC} \times \mathrm{TC}.$$

EC is interpreted as changes in the relative efficiency of a court, that is, movements towards or away from the frontier, while TC measures the shift of the frontier itself. EC or TC that is larger (smaller) than unity indicates an improvement (decline) in efficiency or technical change between period *t* and period *t*+1.[30] Using the Simar and Wilson (1998a) notation, which allows both TC and EC to have either VRS or CRS, makes the decomposition shown in equation 7 possible.

---

[27] There is a subscript *t* in both the numerator and the denominator, since the technology in period *t* is the reference. Thus, production outside the reference frontier can occur, which would indicate technical change (Färe et al., 1994).

[28] The approach by Nishimizu and Page (1982) requires specification of the functional form.

[29] The inclusion of multiple inputs and outputs when assuming VRS in technical efficiency computations was first proposed by Banker (1984) and empirically applied by Banker, Charnes, and Cooper (1984). A graphical description of the DEA frontier with CRS and VRS presented in figure A1.

[30] Färe et al. (1994) relax the CRS assumption to allow for VRS and decompose EC into a pure effect and a scale effect, respectively. According to Ray and Desli (1997), the decomposition presented by Färe et al. (1994) is wrong, because the efficiency change is assumed to exhibit VRS while the technology has CRS. To clarify, if CRS holds, there will not be any scale effect, since scale optimality pioneered by Frisch (1964) is assumed. On the other hand, if VRS is assumed, technical change, as defined by Färe et al. (1994), does not measure the shift in the CRS frontier. However, the error in calculation that Ray and Desli (1997) mention is, according to Simar and Wilson (1998a), only an error in the definition of equations 6–7 in Färe et al. (1994). In other words the definitions by Färe et al. (1994) assume CRS but their calculations allow for VRS. According to Simar and Wilson (1998a), equation 6 in Färe et al. (1994) should be written as equation 6 in this paper.

[7] $$\mathrm{M_o}(\mathbf{x}_i^t, \mathbf{y}_i^t, \mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) = \left(\frac{\mathrm{D_o^{t+1}}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{\mathrm{D_o^t}(\mathbf{x}_i^t, \mathbf{y}_i^t)}\right) \times \left(\frac{\Delta_\mathrm{o}^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})/\mathrm{D_o^{t+1}}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{\Delta_\mathrm{o}^t(\mathbf{x}_i^t, \mathbf{y}_i^t)/\mathrm{D_o^t}(\mathbf{x}_i^t, \mathbf{y}_i^t)}\right) \times \left(\frac{\mathrm{D_o^t}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})\mathrm{D_o^t}(\mathbf{x}_i^t, \mathbf{y}_i^t)}{\mathrm{D_o^{t+1}}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})\mathrm{D_o^{t+1}}(\mathbf{x}_i^t, \mathbf{y}_i^t)}\right)^{\frac{1}{2}} \times \left(\frac{\Delta_o^t(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})/\mathrm{D_o^t}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{\Delta_\mathrm{o}^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})/\mathrm{D_o^{t+1}}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})} \times \frac{\Delta_\mathrm{o}^t(\mathbf{x}_i^t, \mathbf{y}_i^t)/\mathrm{D_o^t}(\mathbf{x}_i^t, \mathbf{y}_i^t)}{\Delta_\mathrm{o}^{t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t)/\mathrm{D_o^{t+1}}(\mathbf{x}_i^t, \mathbf{y}_i^t)}\right)^{½} = \Delta\mathrm{PureEff} \times \Delta\mathrm{ScaleEff} \times \Delta\mathrm{PureTech} \times \Delta\mathrm{ScaleTech}.$$

ΔPureTech and ΔPureEff are both defined on the best-practice technologies according to Ray and Desli (1997) and Färe et al. (1994), respectively. The scale efficiency change measures the movement towards or away from the technically optimal scale. Finally, the scale of the technology (ΔScaleTech), proposed by Simar and Wilson (1998a), represents the scale bias of technical change, that is, the geometric mean of two scale efficiency ratios. This means that any change in ScaleTech occurs from a change in the shape of the technology. The first ratio consists of the change in the scale of the technology between *t* and *t*+1. The reasoning of the second ratio is similar, specifically a change in the scale of the technology between *t* and *t*+1 relative to the location of the production unit in period *t*.[31] Problems with this decomposition can occur when cross-period distance functions are calculated using the VRS assumption, since it can generate missing values for some components.[32] Finally, this decomposition is criticized slightly for its confusing interpretation. For example, Wheelock and Wilson (1999) interpret what we call ΔScaleTech as the shape of the technology, while Zofio and Lovell (1998) interpret it as the scale bias of the technology (see Balk, 2001; Ray, 2001). A graphical description of ΔPureTech and ΔScaleTech is presented in figure A2 in the appendix.[33]

To examine TFP and its decomposed factors, the efficiency needs to be calculated. The reciprocal to the output-based Farrell (1957) measure of technical efficiency is formulated by Färe et al. (1994) as:

[8] $$[\mathrm{D}_0^t(\mathbf{x}_\mathrm{i}^t, \mathbf{y}_\mathrm{i}^t)]^{-1} = TE = \mathrm{Max}\,\theta$$

subject to

[9] $$\sum_{k=1}^{K} \mathbf{z}_k \mathbf{x}_{k,n} \leq \mathbf{x}_{j,n}, n = 1, .., N$$

[31] This decomposition is simultaneously proposed by Gilbert and Wilson (1998), Wheelock and Wilson (1999), and Zofio and Lovell (1998).

[32] The VRS assumption is more sensitive to infeasible values when the cross-period distance functions are computed (Briec & Kestens, 2009). This can be avoided by only using a CRS technology. However, CRS is only relevant when long-run equilibrium in size appears to be a reasonable assumption (Chambers & Pope, 1996). This is not a valid assumption in our case.

[33] An alternative decomposition can be found in Lovell (2003). In the Lovell (2003) case, both the output and the input mix are considered, respectively. This is, however, not relevant to district courts, since a single court does not have the power to decide the output mix.

[10] $\sum_{k=1}^{K} \mathbf{z}_k \mathbf{y}_{k,n} \geq \theta_j \mathbf{y}_{j,m}, m = 1, .., M$

[11] $\sum_{k=1}^{K} \mathbf{z}_k \geq 0$ (CRS),

[12] $\sum_{k=1}^{K} \mathbf{z}_k = 1$ (VRS),

where z_k is $\mathrm{N} \times 1$ the vector of intensity variables (weights). The objective is to maximize θ, which corresponds to minimizing the value of the distance function, $\mathrm{D}_0^t(\mathbf{x}_i^t, \mathbf{y}_i^t)$.[34] For example, if y_0 is an arbitrarily chosen level of output, the maximum output, given the level of inputs, is calculated as $\mathrm{y}_0 * \mathrm{TE}_i^t$ or similarly $\mathbf{y}_0/\mathrm{D}_0^t(\mathbf{x}_i^t, \mathbf{y}_i^t)$.

To compute the MPI, four single-period problems are required, assuming CRS and VRS, as well as four mixed-period problems, under CRS and VRS.[35] These calculations will generate the average MPI. To improve the robustness of the calculated MPIs and draw conclusions based on statistical inference, a bootstrap approach is applied.

Bootstrapping the Malmquist productivity index

Statistical inference for DEA is most commonly based on bootstrapping (see e.g. Efron, 1979). Bootstrapping, and other resampling techniques, simulates the data-generating process multiple times by resampling from the data and applying the original estimator to each simulated sample. This generates an approximation of the sample distribution that can be used to create inference that is meaningful in a statistical sense, for example the confidence intervals of the DEA efficiency scores. These confidence intervals are based on a large number of bootstrap draws (see e.g. Simar & Wilson, 1998b). Further, the efficiency scores can be bias corrected, as proposed by Simar and Wilson (1999). However, the rule of thumb is not to correct for this bias unless $\mathrm{s}^2 < \frac{1}{3}(Bias_B[\mathrm{M}_i^{t,t+1}])^2$, where s^2 is the variance of the bootstrapped values (Simar & Wilson, 1999). The procedure can be summarized in four steps: 1) calculate the MPIs as previously described, 2) generate an i.i.d. bootstrap sample from the original sample, 3) calculate the MPIs based on the bootstrap sample, and 4) repeat steps 2 and 3 a sufficient number of times, in our study 2000, to generate standard deviations to construct the confidence intervals of the MPI and its decomposed factors.[36]

To summarize the methodology, the MPI including all the decomposed factors and their confidence intervals will be computed and bootstrapped. This provides the possibility to evaluate

---

[34] A distance, $D_0^t(\mathbf{x}_\mathrm{i}^\mathrm{t}, \mathbf{y}_\mathrm{i}^\mathrm{t}) \leq 1$, is fulfilled if the technology is defined on the production set as in equation 3. However, it can be above 1 in the mixed linear problem, indicating technical progress.

[35] The necessary linear problems to be solved are $D_0^t(\mathbf{x}_i^t, \mathbf{y}_i^t)$, $D_0^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})$, $D_0^\mathrm{t}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})$, and $D_0^{t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t)$ for CRS. Similarly they are $\Delta_0^t(\mathbf{x}_i^t, \mathbf{y}_i^t)$, $\Delta_0^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})$, $\Delta_0^t(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})$, and $\Delta_0^{t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t)$ for VRS. These are calculated for each change in time, generating 8*3 problems for each court.

[36] This calculation is performed using the FEAR software in R (Wilson, 2008). A discussion about weaknesses with this bootstrap procedure can be found in Olesen and Petersen (2016).

changes based on statistical significance due to the bootstrapping. The decomposition can serve as a good starting point for investigating the sources of the TFP change. This can be achieved without problems of sample noise, making the results statistically robust.[37]

5. Data

The data used were obtained from the SNCA and are more detailed than the data used in previous research; for example, different cases and matters are reported in 47 categories, which will be taken into consideration even though some aggregation is necessary. The reason is that each input and output that is added to the model reduces the ability to discriminate between courts due to the degree of freedom problem. The choice of input and output variables is based on what representatives of the courts have stated in interviews as reasonable resource and performance measures as well as economic theory and previous research.[38]

In the first step, the outputs are aggregated into decided civil cases, decided criminal cases, and decided matters. Simply adding these three groups together will, however, introduce aggregation errors. There are differences between courts regarding the type of cases and matters that they handle. Some courts handle more resource-intense cases than others. For example, a murder case would most likely require more resources than a traffic case. This heterogeneity within different categories of cases and matters is not taken into account in previous studies (see e.g. Kittelsen & Førsund, 1992; Santos & Amado, 2014; Yeung & Azevedo, 2011). To compensate for these facts, the outputs are weighted by the hearing time, that is, the time in the courtroom.[39] This means that courts with a large share of cases from a complicated category are not negatively affected in terms of TFP. The weights are based on the average hearing time in each sub-category; for example, criminal cases alone consist of almost 40 categories, and each category receives its own weight.[40]

On the input side, labor is the largest cost share for Swedish district courts, with 70%, and the rental cost is about 13% for the period 2012–2015 (SNCA, 2012, 2013, 2014, 2015, 2016). Labor is divided into three categories, specifically the number of hours worked for: 1) judges, 2) law clerks, and 3) other personnel. A measure of capital is omitted from previous research (see e.g. Ferrandino, 2014; Kittelsen & Førsund, 1992), with the exception of Elbialy, García-Rubio,

[37] Førsund (2016) argues that separation of the pure and the scale effects are problematic in the decompositions. Furthermore, the interpretation of multiplicative decomposition is argued to be unclear. However, other authors such as Lovell (2003) and Simar and Wilson (1998a) argue that this decomposition is meaningful.
[38] Wagner and Shimshaks's (2007) strategy of model selection is also used to assess the plausibility of the model.
[39] According to Andersson et al. (2017b), the hearing time could be used as an approximation of the resources spent.
[40] The weights are calculated as an average over the whole time period for which data are available; that is, 2007–2015.

and Zaki (2011), who incorporate computers. To incorporate capital, the office space of the court is used following the assumption that the amount of capital, such as computers and office equipment, is proportional to the size of the premises.[41]

Furthermore, the caseload, as described in the previous literature, is an important source of performance for several reasons; for example, if there is no caseload, there will not be any output. We argue that the caseload is an important variable to incorporate when calculating performance that focuses on improvements of technology, management, and so on. However, it is not recommendable to include the caseload in the main analysis when TFP is investigated, since an important factor of TFP changes is flexibility in inputs, that is, adjustable inputs depending on changes in justice demand. Thus, the caseload is only included in the correlation analysis. The caseload in year *t* is defined as the stock of open cases and matters at the end of year *t*-1 plus the incoming cases and matters in the present year. A potential problem with this method is that the incoming cases in the end of 2015 are not included. The correlation is invariant for addition and therefore only a problem when the difference is non-random between courts. In our case, it is however more relevant to assume randomness between courts, meaning that it does not affect the correlations.[42]

[41] A potential limitation that we cannot control for is differences in the preliminary investigation between regions; for example, a court might perform better if it is has better material with which to work.

[42] A measure of quality can potentially be included in the analysis. Andersson et al. (2017b) incorporates the change frequency in higher instances in a second-stage analysis and concludes zero correlation with the efficiency scores.

Outlier detection

The simplest method of detecting outlier is by using boxplots. A short description of different procedures is presented in the methodology. In this paper the variable of interest is the change in TFP. The outlier analysis is performed using changes in outputs, that is, the rate of change in, for example, criminal cases from 2012 to 2013. The results of this analysis for each output are reported in figure 2.

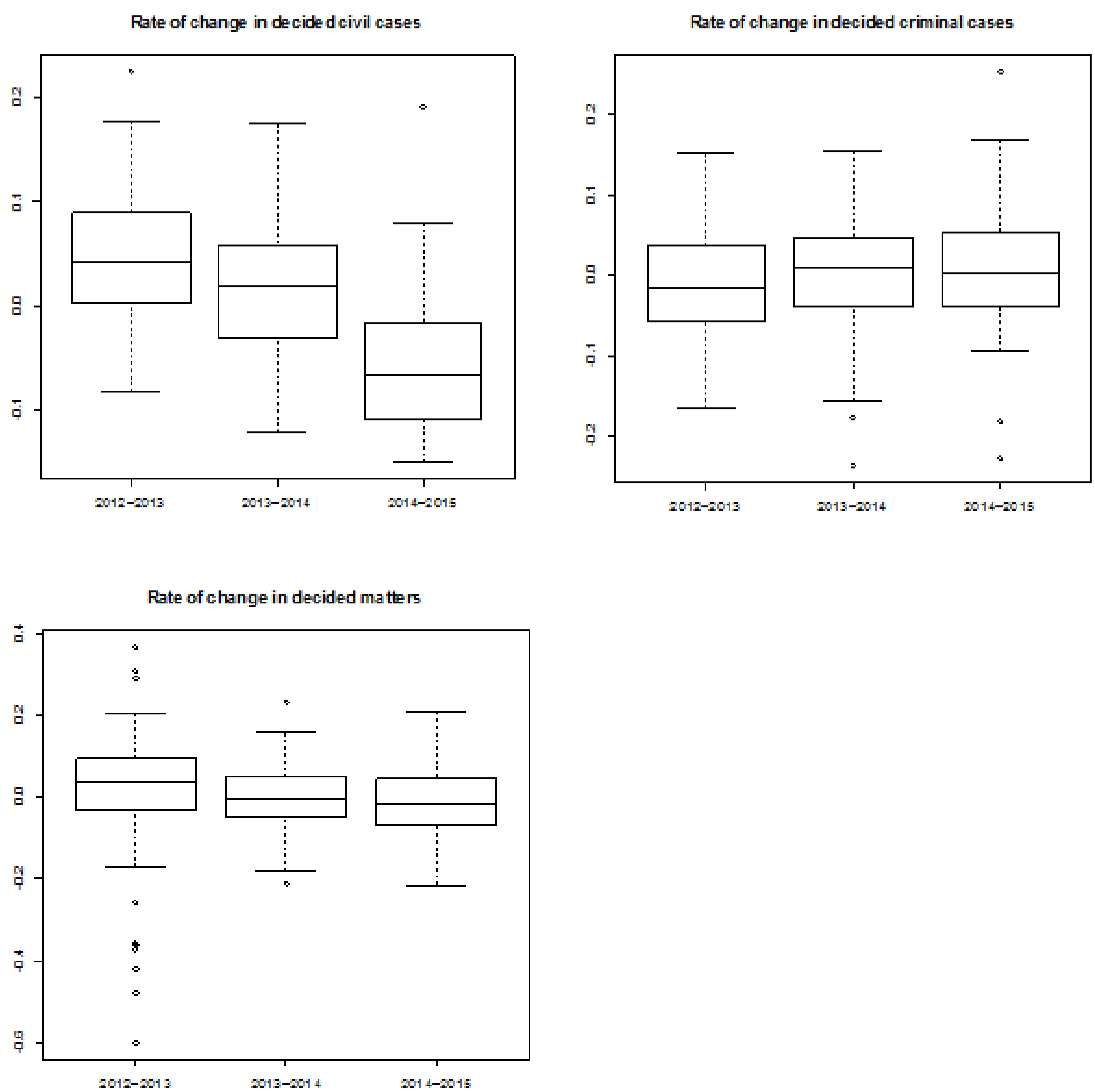

Figure 2: Boxplots of the decided weighted outputs

Figure 2 presents boxplots of the changes in output from one year to the following one, that is, our three investigated time periods: 2012–2013, 2013–2014, and 2014–2015. The upper-left boxplot reports changes in civil cases for the different periods, the upper-right boxplot criminal cases, and the lower-left boxplot decided matters. The whiskers have a standard range showing that courts with a change in at least 1 of the outputs of more than approximately $\pm 20\%$ from one year to the next are eliminated. A total of 14 courts are eliminated, meaning that there are 34 courts left in the sample.[43] The descriptive statistics of the outputs and inputs, after the elimination of outliers, are reported in table 1.

Table 1: Descriptive statistics, excluding outliers.

	Mean (std dev.)			
Outputs	2012	2013	2014	2015
Number of decided civil actions (weighted)	3 877 (3 293)	4 069 (3 529)	4 100 (3 497)	3 837 (3 238)
Number of decided criminal, real estate, and environmental cases (weighted)	4266 (2 991)	4 182 (3 015)	4167 (2 909)	4 207 (3 012)
Number of decided matters (weighted)	901 (597)	942 (619)	955 (654)	944 (634)
Inputs				
Regular and irregular judges of any kind*	19.22 (13.21)	19.59 (14.17)	20.44 (14.93)	20.84 (15.21)
Law clerks and employees of the law clerk training program*	19.83 (15.24)	20.17 (15.35)	20.63 (15.44)	20.17 (15.42)
Other personnel*	33.68 (25.37)	34.75 (26.55)	35.26 (26.78)	35.43 (27.85)
Office space (square meters)	4 493 (2 863)	4 592 (2 932)	4 570 (2869)	4 608 (2838)
Correlation variable				
Caseload	7 447 (5 394)	7 383 (5 332)	7 146 (5 131)	5 430 (3 893)

* Measured as full-time equivalent.

[43] Many courts are outliers based on the matters in 2012–2013 compared with the number of outliers in other periods. The reason is that one type of matters that was handled by seven of the district courts was moved to the Swedish Mapping, Cadastral and Land Registration. This means that the caseload will decline ceteris paribus for these courts during this year, making the TFP change negative if the inputs are not adjusted to compensate. In this study these courts are considered to be outliers.

Table 1 show that the differences in output over time are on average quite small. However, each of the inputs increases in size over time. For example, the number of full-time equivalent judges increases from 19.22 to 20.84 (9%). The caseload decline over time, but the dramatic drop the last period is because all incoming cases are not included as previously described. Also to be noted are the large standard deviations, which are almost as large as the means. For instance, Göteborg district court is, in terms of hours worked for judges, approximately 17 times larger than Gotland. However, the largest and smallest courts are eliminated in the outlier analysis, making the standard deviations smaller than if all the courts are included (reported in appendix table A4). Finally, the non-weighted caseload declines during the studied time period, for example by 4% from 2013 to 2015.

6. Results

The results concerning the MPI and its decomposed factors are reported first, namely efficiency change (EC) and technical change (TC). EC and TC are then further decomposed into a pure and a scale effect. Then a correlation analysis is performed based on the MPI and its components.

6.1. Malmquist index and its decomposed factors

To create possibilities for comparisons, we firstly conclude that the TFP change is negative during each year, driven by technical regress, when all the courts are included.[44] Based on the outlier detection analysis in the data section, the district courts with large volatility in output over time are eliminated. The MPI and its components, after the elimination of outliers, are reported in table 2.[45]

Table 2: Malmquist index and its decomposed factors after eliminating outliers.

Year	Malmquist productivity index (MPI)	Technical change (TC)	Efficiency change (EC)	Number of courts with TFP below unity*	Number of courts with TFP above unity*
2012–2013	1.018 (0.997–1.038)	1.013 (0.951–1.051)	1.005 (0.960–1.069)	9 (26%)	14 (41%)
2013–2014	1.002 (0.985–1.021)	0.999 (0.958–1.045)	1.003 (0.948–1.051)	9 (26%)	11 (32%)

[44] The TFP change is -2.2%, -0.6%, and -3.5% for 2012–2013, 2013–2014 and 2014–2015, respectively. The full results, including all the courts, are reported in the appendix.

[45] The bias is not corrected for since $s^2 = 0.00071 > \frac{1}{3}(Bias_B[\mathrm{M}_i^{t,t+1}])^2 = 0.00026$, calculated as an annual average. This is in accordance with the limit stated by Simar and Wilson (1999).

2014–2015	0.961** (0.940–0.981)	0.963 (0.924–1.006)	0.998 (0.942–1.044)	23 (68%)	3 (9%)
2012–2015	0.993 (0.974–1.013)	0.992 (0.944–1.034)	1.002 (0.950–1.055)	14 (40%)	9 (27%)

The bootstrapped confidence intervals at 5% are reported in parentheses, where ** symbolizes significance at the 5% level. * Below (above) unity means significantly below (above) the 5% level of significance.

In table 2 the TFP change, measured as the MPI, is positive for 2012–2013 and 2013–2014, respectively. Column 3 reports that TC is positive during the first period and approximately zero in the following ones. EC contributes positively to the TFP growth during the period 2012–2014. If the five courts that are considered to be outliers, based on the fact that a type of matter only handled by these courts moved to the Swedish Mapping, Cadastral and Land Registration, are included, the MPI of the first two years is negative, indicating low flexibility, even though this was known at least 1 year in advance.[46] For the last period, 2014–2015, both TC and EC affect TFP negatively, which generates a statistically significant decline in TFP. An interpretation of the results for the first two periods (2012–2014) is that the courts with the largest volatility in output on average explain the negative TFP change. The negative MPI, measured as an annual geometric average over the whole time period, is driven by 2014–2015. A few courts have large volatility in terms of TFP. For example, Gotland has TFP growth of 13.6% and 26.1% for the first periods, respectively. However, it has a 12% decline in TFP during the period 2014–2015. The MPI and its confidence intervals are graphically reported for each court and each year in figures A3–A5 in the appendix.

EC is small during the last period, specifically -0.2% (column 4, row 4), while TC is -3.7% (column 3, row 4). These results are interpreted as the statistically significant decline in TFP having its major source in a movement inwards of the frontier, indicating that the best courts perform worse. In columns 5 and 6, it can be observed that the number of courts with a significantly negative TFP change is stable during the period 2012–2014 but increases in 2014–2015. Furthermore, the number of courts with significantly positive TFP growth decreases in each year.[47] Both the fact that the production frontier moves towards the origin and the fact that many courts perform significantly worse in terms of TFP indicate that this result is not driven by a few observations. To gain more information, TC is decomposed into pure TC and scale TC.

[46] ESO (2016) also concluded low flexibility in the Swedish public sector, e.g. within schools and elder care. For example, the efficiency declined the years with a lower number of pupils.

[47] There is no indication that more or fewer well-performing (or poorly performing) courts are eliminated. Depending on the year, there are for some years larger shares of courts with significantly positive TFP change and for other years the opposite. Looking at the whole period, both the share of well-performing courts and the share of less-well-performing courts, in terms of TFP change, decline.

Decomposition of technological change

The decomposition of TC into pure TC and scale TC is performed according to equation 7 in section 4. The results are reported in table 3.

Table 3: Decomposition of technical change into pure technical change and scale technical change.

Year	Technical change (TC)	Pure TC	Scale TC
2012–2013	1.013 (0.951–1.051)	0.999 (0.927–1.031)	1.010 (0.993–1.044)
2013–2014	0.999 (0.958–1.045)	1.000 (0.960–1.051)	0.996 (0.969–1.015)
2014–2015	0.963 (0.924–1.006)	0.969 (0.929–1.018)	0.996 (0.970–1.017)
2012–2015	0.992 (0.944–1.034)	0.989 (0.939–1.033)	1.001 (0.977–1.025)

The bootstrapped confidence intervals at 5% are reported in parentheses.

Technical change is defined as the product of pure TC and scale TC.[48] Pure TC means that the best firms, assuming CRS, have approximately no change in 2012–2014 but pure TC of -3.1% from 2014 to 2015. The movement from the optimal scale (scale TC) generates regress of around 0.4% for the same time period. As an attempt to interpret these declines, it is worth studying the boxplots used to identify the outliers, shown in figure 1. It can be observed that the average number of decided cases and matters is fairly stable for 2012–2014, that is, a fluctuation between 1% and 2%. However, looking at 2014–2015, the decided criminal cases and matters are in the same range as previously but the civil cases decline by 5.1% on average, with a median of -6.7%. Thus, the produced output decreases in total, driven by a lower number of civil cases. This, however, only concerns decided cases. Regarding the data from the SNCA (2015), the number of incoming civil cases is reported to decline by 5% during the period 2014–2015, and the number of balanced civil cases declined by 7%.[49] Thus, an explanation for the negative TFP change driven by a decline in TC is likely to be due to the decline in the caseload during this period. This in itself should not decrease TFP if the inputs are fully flexible. However, the inputs are not flexible enough to compensate for the lower workload level. To investigate the components of efficiency change, its decomposition is now reported.

[48] As noted in the methodology, a VRS assumption using cross-period distance functions is sensitive to infeasible values. Two courts (Gotland and Haparanda) receive missing values for pure TC and scale TC. Both these courts show the same pattern, specifically large volatility in TFP.

[49] It can also be noted that the incoming cases in this category decrease by 8% from 2013 to 2015 (SNCA, 2016).

Decomposition of efficiency change

Efficiency change is decomposed into pure efficiency change and scale efficiency change according to Färe et al. (1994). The results are reported in table 4.

Table 4: Decomposition of efficiency change into pure efficiency change and scale efficiency change.

Year	Efficiency change (EC)	Pure EC	Scale EC
2012–2013	1.005 (0.960–1.069)	1.012 (0.974–1.091)	0.993 (0.947–1.012)
2013–2014	1.003 (0.948–1.051)	0.998 (0.934–1.047)	1.005 (0.981–1.034)
2014–2015	0.998 (0.942–1.044)	0.996 (0.931–1.046)	1.002 (0.977–1.029)
2012–2015	1.002 (0.950–1.055)	1.002 (0.946–1.061)	1.000 (0.968–1.025)

The bootstrapped confidence intervals at 5% are reported in parentheses.

Efficiency change is positive for 2012–2013 and 2013–2014, respectively. The positive effect has its source in the positive pure EC for the former period and scale EC for the latter. This indicates that district courts on average become more homogeneous, since the efficiency measures the distance from the frontier. During the period 2014–2015, EC is negative, indicating greater heterogeneity between courts; that is, the average court is further away from the production frontier. However, the change is only -0.2%. In comparison with what can be observed in both table 2 and table 3, it can be concluded that most changes in TFP have their source in technical change and that none of the changes in efficiency is statistically significant.

6.2. Correlation analysis

A few of the previous studies argue the importance of incorporating the justice demand to avoid underestimating a court's TFP. However, as stated, the justice demand should not affect TFP if the inputs are fully flexible; that is, there should be a zero correlation if this is fulfilled. In table 5 the MPI and decomposed factors are correlated with the rate of change in the caseload.

Table 5: Spearman correlations between MPI, TC and EC and the rate of change in the caseload.

Year	Malmquist productivity index (MPI)	Technical change (TC)	Efficiency change (EC)
2012–2013	0.454*** (0.007)	0.240 (0.171)	0.286 (0.101)
2013–2014	0.304* (0.081)	0.134 (0.450)	0.137 (0.441)
2014–2015	0.347** (0.044)	0.232 (0.186)	0.176 (0.320)

P-values are reported in parentheses. *** symbolizes significance at 1%, ** significance at 5% and * significance at 10%.

From table 5 it can be observed that the MPI, but not the components, have a positive correlation with the rate of change in the caseload that is statistically significant. A positive correlation can mean either that the inputs do not decrease enough when the demand for justice services declines or that the employees work harder when the demand increases, generating increased output given the inputs. Each of these reasons indicates a slack in the courts; that is, more can be produced without increasing the inputs. Schneider (2005) argues that the exclusion of the caseload generates an underestimation of TFP. However, despite the positive correlation concluded in this section, we argue that the correct measure of TFP is what we reported in the main analysis. Nevertheless, the caseload can at least partly explain the results, meaning that the inputs are not flexible enough. The positive relationship between the MPI and the rate of change in the caseload is also in line with Beenstock and Haitovsky (2004), who argue that individual productivity increases when the work pressure is high. These results strengthen the previous argument of low flexibility in inputs. However, it should be interpreted carefully since no causality can be concluded.[50]

7. Conclusion and policy recommendations

The purpose of this paper was to investigate the development of TFP from 2012 to 2015. The differences in comparison with previous research are: 1) more detailed data are used that allow the outputs to be weighted based on the hearing time, 2) statistical inference is incorporated by adopting a bootstrap approach, and 3) TFP is decomposed into four components in contrast to a maximum of three in the earlier literature.

[50] The results, including the outliers, show the same pattern but the significance is stronger and the correlations higher. These results are reported in table A4 in the appendix.

The findings indicate a 0.7% decline in TFP measured as an annual geometric mean. However, a substantial variation between courts is found; for example, 26–68% of the courts have a negative change in TFP, while 9–41% of the courts have a positive TFP change, depending on the year. The negative impact is mainly driven by technical regress during the last period. The sources of technical change are fairly equally distributed between pure change in the technology and change in the scale of the technology, except for 2014–2015, when most of the decline comes from the pure part. Furthermore, TFP is concluded to have a positive and significant correlation with the caseload, indicating a non-sufficient level of flexibility in inputs.

The policy conclusion is that there is room for improvements. A recommendation is that district courts with negative TFP could learn from those with positive TFP. This can hopefully result in a positive TFP development in the future. Furthermore, it can be observed that the smallest courts have the largest volatility in TFP change which is a source of inefficiency. Therefore, smoother changes can be achieved by merging courts, which would improve TFP. However, the merging is to some extent constrained by the social and geographical issues that need to be taken into consideration. To avoid this issue, a less controversial policy implication that achieves more flexibility in the Swedish district courts is to develop the back-up force, introduced in 2012, to include other personnel than judges. This will allow the inputs to be adjusted when the demand fluctuates which generates a higher degree of flexibility on the regional level. In particular, this will enhance the flexibility of the small courts. The smallest courts have close to the minimum number of employees. Small courts are by construction more sensitive to changes in the workload, since a small change in the justice demand generates a large share of the percentage. Therefore, the issue of large volatility in the justice demand could at least partly be solved by an expansion of the back-up force to enhance flexibility. Furthermore, more flexible inputs across Sweden could potentially make it possible to eliminate the requirement of a minimum number of employees in each court. Instead the volatility in the smallest courts can be served by flexible personnel, for example the back-up force.

Finally, peer comparisons of courts could be used in many potential aspects of the work of improving efficiency and productivity. For example, differences can be present that are not directly possible to determine in this study, such as organizational problems. This is, however, an aspect that can be taken into consideration in future research.

References

Andersson, C., Antelius, J., Månsson, J., & Sund, K. (2017a). Technical efficiency and productivity for higher education institutions in Sweden. *Scandinavian Journal of Educational Research, 61*(2), 205-223.

Andersson, C., Bonander, F., & Månsson, J. (2017b). *Tingsrätters effektivitet och produktivitet (The efficiency and productivity of Swedish district courts), Swedish National Audit Office 2017:6, Stockholm.*

Andersson, C., Månsson, J., & Sund, K. (2014). Technical efficiency of Swedish employment offices. *Socio-economic planning sciences, 48*(1), 57-64.

Balk, B. M. (2001). Scale efficiency and productivity change. *Journal of Productivity Analysis, 15*(3), 159-183.

Banker, R. D. (1984). Estimating most productive scale size using data envelopment analysis. *European Journal of Operational Research, 17*(1), 35-44.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science, 30*(9), 1078-1092.

Beenstock, M., & Haitovsky, Y. (2004). Does the appointment of judges increase the output of the judiciary? *International Review of Law and Economics, 24*(3), 351-369.

Bjurek, H., Kjulin, U., & Gustafsson, B. (1992). Efficiency, productivity and determinants of inefficiency at public day care centers in Sweden. *The Scandinavian Journal of Economics*, S173-S187.

Blank, J., van der Ende, M., van Hulst, B., & Jagtenberg, R. (2004). *Bench marking in an international perspective: An International Comparison of the Mechanisms and Performance of the Judiciary System*. Rotterdam: Netherlands Council for the Judiciary.

Briec, W., & Kerstens, K. (2009). Infeasibility and directional distance functions with application to the determinateness of the Luenberger productivity indicator. *Journal of Optimization Theory and Applications, 141*(1), 55-73.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: Journal of the Econometric Society*, 1393-1414.

Chambers, R. G., & Pope, R. D. (1996). Aggregate productivity measures. *American Journal of Agricultural Economics, 78*(5), 1360-1365.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429-444.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*: Springer Science & Business Media.

Davies, L., & Gather, U. (1993). The identification of multiple outliers. *Journal of the American Statistical Association*, *88*(423), 782-792.

Deyneli, F. (2012). Analysis of relationship between efficiency of justice services and salaries of judges with two-stage DEA method. *European Journal of Law and Economics, 34*(3), 477-493.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The annals of Statistics, 7*(1), 1-26.

Eisenstein, J., & Jacob, H. (1977). *Felony justice: An organizational analysis of criminal courts*: Boston: Little, Brown

Elbialy, N., García-Rubio, M. A., & Zaki, A. (2011). *The 2007 Egyptian judicial reform and its efficiency implications on first instance courts. Joint discussion paper series in economics*. Retrieved from Effectius Newsletter: https://www.econstor.eu/bitstream/10419/56541/1/657907855.pdf

ESO (Expertgruppen för studier i offentlig ekonomi). (2016). *Mer än tur i struktur – en ESO-rapport om kommunal effektivitet (More than luck in structure - an ESO report regarding efficiency within municipalities).* Stockholm: Government office: Ministry of finance.

Espasa, M., & Esteller-Moré, A. (2015). Analyzing judicial courts´ performance: Inefficiency vs. congestion. *Revista de Economía Aplicada, 23*(69), 61.

Falavigna, G., Ippoliti, R., Manello, A., & Ramello, G. B. (2015). Judicial productivity, delay and efficiency: A Directional Distance Function (DDF) approach. *European Journal of Operational Research, 240*(2), 592-601.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General), 120*(3), 253-290.

Fauvrelle, T. A., & Almeida, A. T. (2016). Determinants of Judicial Efficiency Change: Evidence from Brazil (March 1, 2016). Available at SSRN: https://ssrn.com/abstract=2745269.

Feeley, M. M. (1973). Two models of the criminal justice system: An organizational perspective. *Law & Society Review, 7*(3), 407-425.

Feld, L. P., & Voigt, S. (2003). Economic growth and judicial independence: Cross-country evidence using a new set of indicators. *European Journal of Political Economy, 19*(3), 497-527.

Ferrandino, J. (2014). Testing the Packer Theorem: The Efficiency of Florida's Criminal Circuit Courts. *American Journal of Criminal Justice, 39*(2), 375-393.

Ferrandino, J. A. (2010). *Integrating theory, practice and policy: The technical efficiency and productivity of Florida's circuit courts.* Orlando, FL: University of Central Florida

Finocchiaro Castro, M., & Guccio, C. (2014). Searching for the source of technical inefficiency in Italian judicial districts: An empirical investigation. *European Journal of Law and Economics, 38*(3), 369-391.

Finocchiaro Castro, M., & Guccio, C. (2016). Measuring potential efficiency gains from mergers of Italian first instance courts through nonparametric model. *Public Finance Review*, 1091142116652723.

Frisch, R. (1964). *Theory of production*: Chicago: Rand McNally. .

Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1989). Producitvitiy Developments in Swedish Hospitals: A Malmquist Output Index Approach (Discussion Paper No. 89:3). Sourthern Illinois University.

Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 66-83.

Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. . (1992). Productivity changes in Swedish pharamacies 1980–1989: A non-parametric Malmquist approach. *International Applications of Productivity and Efficiency Analysis*, 17.

Førsund, F. R. (2016). Productivity Interpretations of the Farrell Efficiency Measures and the Malmquist Index and Its Decomposition *Advances in Efficiency and Productivity* (pp. 121-147): Springer.

Gilbert, R. A., & Wilson, P. W. (1998). Effects of deregulation on the productivity of Korean banks. *Journal of Economics and Business, 50*(2), 133-155.

Kapelko, M., & Oude Lansink, A. (2015). An international comparison of productivity change in the textile and clothing industry: a bootstrapped Malmquist index approach. *Empirical Economics, 48*(4), 1499-1523.

Kim, D., & Min, H. (2016). Appeal rate and caseload: Evidence from civil litigation in Korea. *European Journal of Law and Economics*, 1–22.

Kittelsen, S. A., & Førsund, F. R. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis, 3*(3), 277-306.

Krüger, J. J. (2012). A Monte Carlo study of old and new frontier methods for efficiency measurement. *European Journal of Operational Research, 222*(1), 137-148.

Kumbhakar, S. C., & Lovell, C. K. (2003). *Stochastic frontier analysis.* Camebridge: Cambridge university press.

Laspeyres, K. (1871). IX. Die Berechnung einer mittleren Waarenpreissteigerung. *Jahrbücher für Nationalökonomie und Statistik, 16*(1), 296-318.

Lewin, A. Y., Morey, R. C., & Cook, T. J. (1982). Evaluating the administrative efficiency of courts. *Omega, 10*(4), 401–411.

Lovell, C. K. (2003). The decomposition of Malmquist productivity indexes. *Journal of Productivity Analysis, 20*(3), 437–458.

Major, W. (2015). Data envelopment analysis as an instrument for measuring the efficiency of courts. *Operations Research and Decisions, 25*(4), 19–34.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de estadística, 4*(2), 209-242.

Messick, R. E. (1999). Judicial reform and economic development: a survey of the issues. *The World Bank Research Observer, 14*(1), 117-136.

Ministry of Justice. (2015). *The Swedish judicial system*. Stockholm Retrieved from http://www.regeringen.se/49e9fa/contentassets/24ef36d9206a41fb80d6b74030fd9ef6/the-swedish-judicial-system.pdf.

Nardulli, P. F. (1978). *The courtroom elite: An organizational perspective on criminal justice*: Cambridge, MA: Ballinger Publishing Company.

Nishimizu, M., & Page, J. M. (1982). Total factor productivity growth, technological progress and technical efficiency change: Dimensions of productivity change in Yugoslavia, 1965-78. *Economic Journal, 92*(368), 920–936.

Nissi, E., & Rapposelli, A. (2010). A data envelopment analysis of Italian courts efficiency. *Statistica Applicata – Italian Journal of Applied Statistics, 22(2)*, 199–210.

Olesen, O. B., & Petersen, N. C. (2016). Stochastic data envelopment analysis—A review. *European Journal of Operational Research, 251*(1), 2-21.

Pedraja-Chaparro, F., & Salinas-Jimenez, J. (1996). An assessment of the efficiency of Spanish courts using DEA. *Applied Economics, 28*(11), 1391–1403.

Peyrache, A., & Zago, A. (2012). *Large courts, small justice! The inefficiency and the optimal structure of the Italian justice sector*. Retrieved from

Ray, S. C. (2001). *On an extended decomposition of the Malmquist productivity index.* Paper presented at the 7th European Workshop on Efficiency and Productivity Analysis, Oviedo, Spain, September.

Ray, S. C., & Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *American Economic Review, 87*(5), 1033–1039.

Rigsrevisionen. (2000). *Report to the public auditors about the productivity of the judiciary (Beretning til statsrevisorerne om retternes produktivitet)*. Retrieved from http://www.rigsrevisionen.dk/media/1877192/5-99.pdf

Santos, S. P., & Amado, C. A. (2014). On the need for reform of the Portuguese judicial system – Does data envelopment analysis assessment support it? . *Omega, 47*, 1-16.

Schneider, M. R. (2005). Judicial career incentives and court performance: An empirical study of the German labour courts of appeal. *European Journal of Law and Economics, 20*(2), 127–144.

Shephard, R. W. (1953). *Cost and production functions* (Vol. 194). Princeton, NJ: Princeton University Press: Princeton University Press.

Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton, NJ: Princeton University Press.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis, 20*(3), 391–424.

Simar, L., & Wilson, P. W. (1998a). *Productivity growth in industrialized countries*: Working Paper, Department of Economics, University of Texas, Austin, TX 78712, USA.

Simar, L., & Wilson, P. W. (1998b). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management science, 44*(1), 49–61.

Simar, L., & Wilson, P. W. (1999). Estimating and bootstrapping Malmquist indices. *European Journal of Operational Research, 115*(3), 459–471.

Simar, L., & Wilson, P. W. (2000). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis, 13*(1), 49-78.

SNCA (Domstolsverket). (2012). *Årsredovisning för Sveriges domstolar (Annual report for Swedish district courts)*. Jönköping.

SNCA (Domstolsverket). (2013). *Årsredovisning för Sveriges domstolar (Annual report for Swedish district courts)*. Jönköping.

SNCA (Domstolsverket). (2014). *Årsredovisning för Sveriges domstolar (Annual report for Swedish district courts)*. Jönköping.

SNCA (Domstolsverket). (2015). *Årsredovisning för Sveriges domstolar (Annual report for Swedish district courts)*. Jönköping.

SNCA (Domstolsverket). (2016). *Årsredovisning för Sveriges domstolar (Annual report for Swedish district courts)*. Jönköping.

Sturup, J., Rostami, A., Gerell, M., & Sandholm, A. (2017). Near-repeat shootings in contemporary Sweden 2011 to 2015. *Security Journal*, 1–20.

Swedish National Council for Crime Prevention. (2015). *The development of solved suspicious and approved prosecutions 2004–2014 (Utvecklingen av personuppklarade misstankar och bifallna åtal 2004–2014)*. Stockholm: Brottsförebyggande rådet, Enheten för kommunikation, Retrieved from https://bra.se/download/18.31d7fffa1504bbffea0a7ff1/1450358617656/2015_Utvecklingen+av+personuppklarade.pdf.

The Swedish Agency for Public Management. (2007). Sammanslagna tingsrätter – en utvärdering. *2007:09*.

Tyrefors Hinnerich, B., Palme, M., & Priks, M. (2017). Age-dependent court sentences and crime bunching: Empirical evidence from Swedish administrative data.

Wagner, J. M., & Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European Journal of Operational Research, 180*(1), 57–67.

Wheelock, D. C., & Wilson, P. W. (1999). Technical progress, inefficiency, and productivity change in US banking, 1984-1993. *Journal of Money, Credit, and Banking*, 212–234.
Wilson, P. W. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business & Economic Statistics, 11*(3), 319–323.
Wilson, P. W. (2008). FEAR: A software package for frontier efficiency analysis with R. *Socio-economic planning sciences, 42*(4), 247-254.
Voigt, S. (2016). Determinants of judicial efficiency: A survey. *European Journal of Law and Economics, 42*(2), 183–208.
Yeung, L. L., & Azevedo, P. F. (2011). Measuring efficiency of Brazilian courts with data envelopment analysis (DEA). *IMA Journal of Management Mathematics*, dpr002.
Zofio, J., & Lovell, C. (1998). Yet another Malmquist productivity index decomposition. *Department of Economics, University, University Autonomy de Madrid (mimeo)*.

Appendix

Table A1: Malmquist index and its decomposed factors including all courts. The bootstrapped confidence intervals at 5% are reported in parentheses.

Year	**Malmquist productivity index (MPI)**	**Technical change (TC)**	**Efficiency change (EC)**	**Number of courts with TFP below unity***	**Number of courts with TFP above unity***
2012–2013	0.978 (0.955–1.002)	0.923** (0.852–0.960)	1.063** (1.008–1.151)	21 (44%)	13 (27%)
2013–2014	0.994 (0.974–1.013)	0.977 (0.931–1.019)	1.018 (0.962–1.070)	16 (33%)	17 (35%)
2014–2015	0.965** (0.941–0.987)	0.981 (0.940–1.035)	0.984 (0.914–1.032)	28 (58%)	5 (10%)
2012–2015	0.979 (0.957–1.000)	0.960 (0.907–1.004)	1.022 (0.961–1.083)	22 (45%)	12 (25%)

** symbolizes significance at the 5% level. * Below (above) unity means significantly below (above) the 5% level of significance.

In the first column of table A1, the years used are indicated. The MPI is reported in the second column. As seen from column 2, the MPI is below unity during each year. Looking at the whole period, TFP decreases by 2.1% measured as an annual average, which can be seen from column 2, row 5. The decomposition into TC and EC makes it possible to some extent investigate the source of this decline. For each of the studied periods, TC is below unity, on average. The interpretation of this is that technical regress occurred. Technical regress can occur for example if reinvestments of capital are not made to compensate for capital use. Another potential explanation is that the cases have become more complicated over time, therefore taking more time to handle. However, the turnaround time decreased during this time period, in terms of both criminal cases and civil cases, according to the SNCA (2014, 2016). Moreover, technical regress can occur due to a reduction in the demand for justice services with either partial or no adjustment in inputs. While technological regress is observed, efficiency improves during two of the periods. The efficiency improvement is 6.3% and 1.8% for 2012–2013 and 2013–2014, respectively. Moreover, from column 5 it can be observed that 16–28 courts, that is, 33–58%, experience a negative change in TFP depending on the year. This can be compared with column 6, which shows that between 5 and 17 courts, or 10–35%, have a positive TFP change. In detail, 7 courts have a significantly negative TFP change during each year.[51] On average, 45% of the courts have a negative TFP change and 25% of the courts a positive and significant TFP change.

Decomposition of technical change

The decomposition of TC into pure TC and scale TC is performed according to equation 7 in the methodology. The results are reported in table A2.[52]

Table A2: Decomposition of technical change into pure technical change and scale technical change including all courts. The bootstrapped confidence intervals at 5% are reported in parentheses.

Year	Technical change (TC)	Pure TC	Scale TC
2012–2013	0.923** (0.852–0.960)	0.959** (0.873–0.997)	0.969 (0.928–1.018)
2013–2014	0.977 (0.931–1.019)	0.989 (0.942–1.037)	0.985 (0.953–1.015)
2014–2015	0.981 (0.940–1.035)	0.979 (0.932–1.040)	1.005 (0.971–1.035)
2012–2015	0.960 (0.907–1.004)	0.976 (0.915–1.024)	0.986 (0.951–1.023)

** symbolizes significance at the 5% level.

Technical change is defined as the product of pure TC and scale TC. Pure TC means that the best firms, assuming CRS, have technical regress of approximately 4% from 2012 to 2013 that is due

[51] These are the courts of Alingsås, Mora, Norrtälje, Sundsvall, Uddevalla, Varberg, and Västmanland.

[52] There are courts with missing values for pure TC and scale TC. Each of them has a large volatility in TFP presented in parentheses: Lycksele (19.2%, 5.7%, and -23.3%) is missing during each time period. Gotland (+15.1%, 11.8%, and -10.9%) and Haparanda (3.4 %, 2.6%, and -8.9%) are missing from 2013-2015. Gällivare (3.9%, 4.5%, and -15.4%) is missing for the last time period. These courts are also the four smallest, which might be a reason for their large volatility.

to non-optimal use of inputs. The movement from the optimal scale (scale TC) generates regress of around 3% for the same time period. Basically the same pattern can be observed for the other time periods but is smaller in magnitude, meaning less regress. During the period 2014–2015, the district courts move towards the optimal scale, which can be seen in the column for scale TC. However, the negative pure technical change generates regress in total for this period.

Decomposition of efficiency change

Efficiency change is decomposed into pure efficiency change and scale efficiency change. The results are reported in table A3.[53]

Table A3: Decomposition of efficiency change into pure efficiency change and scale efficiency change including all courts. The bootstrapped confidence intervals at 5% are reported in parentheses.

Year	**Efficiency change (EC)**	**Pure EC**	**Scale EC**
2012–2013	1.063** (1.008–1.151)	1.024 (0.972–1.122)	1.039 (0.964–1.092)
2013–2014	1.018 (0.962–1.048)	1.004 (0.942–1.059)	1.014 (0.974–1.048)
2014–2015	0.984 (0.914–1.032)	0.997 (0.909–1.053)	0.987 (0.952–1.028)
2012–2015	1.022 (0.961–1.076)	1.008 (0.941–1.078)	1.013 (0.963–1.056)

** symbolizes significance at the 5% level.

Efficiency change is positive for 2012–2013 and 2013–2014, respectively. The positive effect has its source in both positive pure EC and positive scale EC for both time periods, respectively. This indicates that district courts on average become more homogeneous, since efficiency measures the distance from the frontier. However, during the period 2014–2015, both components of EC are negative, indicating larger heterogeneity between courts; that is, the average court is further away from the production frontier. Based on the results in table A2, regarding TC, it can also be concluded that this change in discrepancy means that the courts on the frontier perform worse, since a technical regress occurs. To investigate whether the reported results are related to the rate of change in the caseload, a correlation analysis is performed.

Correlation analysis

[53] There are no missing observations in the decomposition of EC.

The relationship between the MPI and the rate of change in the caseload should theoretically be zero if the inputs are fully flexible. The results, including all the courts, are reported in table A4.

Table A4: Spearman correlations between MPI, TC and EC and the rate of change in the caseload. All courts included

Year	**Malmquist productivity index (MPI)**	**Technical change (TC)**	**Efficiency change (EC)**
2012–2013	0.715*** (0.000)	0.216 (0.140)	0.471*** (0.001)
2013–2014	0.471*** (0.001)	0.288** (0.047)	0.483*** (0.001)
2014–2015	0.580*** (0.000)	0.202 (0.168)	0.535*** (0.000)

P-values are reported in parentheses. *** symbolizes significance at 1%, ** significance at 5% and * significance at 10%.

The same patterns as shown in table 5 in the main analysis can be observed from table A4 when all the courts are included. However, the level of significance is slightly stronger when all the courts are included, which is intuitive, since it should be more difficult to adjust inputs for large changes in percentage of the justice demand.

Descriptive statistics

Table A5: Descriptive statistics including all the courts

	Mean (std. dev.)			
Outputs	2012	2013	2014	2015
Number of decided civil actions (weighted)	3 587 (3 454)	3 746 (3 617)	3 782 (3 593)	3 619 (3 530)
Number of decided criminal, real estate, and environmental cases (weighted)	3 838 (3 196)	3 778 (3 197)	3 736 (3 063)	3 780 (3 142)
Number of decided matters (weighted)	865 (637)	881 (764)	898 (827)	883 (789)
Inputs				
Regular and irregular judges of any kind*	17.17 (14.63)	17.58 (15.58)	18.48 (16.39)	18.76 (16.36)
Law clerks and employees of the law clerk training program*	18.41 (17.80)	18.79 (17.86)	19.12 (18.05)	18.82 (18.17)
Other personnel*	31.45 (29.38)	32.54 (30.59)	33.01 (30.71)	33.36 (31.94)
Area (square meters)	4 371 (4 196)	4 465 (4 218)	4 461 (4 180)	4 511 (4 175)
Correlation variable				
Caseload	6 900 (5 877)	6 817 (5 952)	6 577 (5 696)	5 023 (4 355)

* Measured as full-time equivalent.

Illustration of DEA

Figure A1 below shows the production, under CRS and VRS, when one input is used to produce one output.

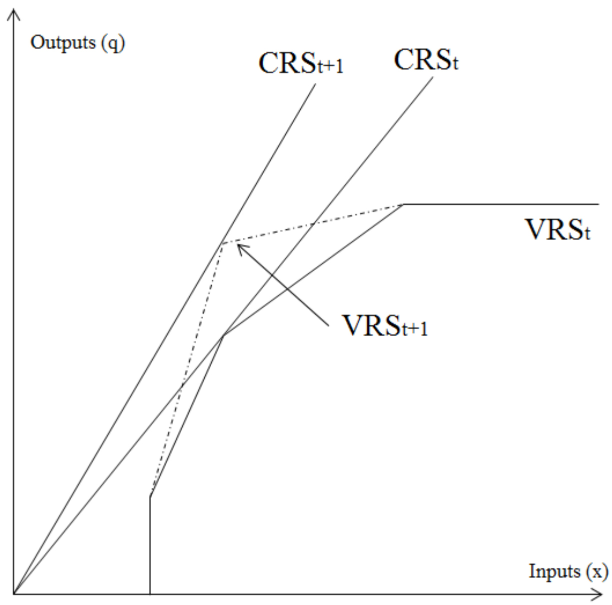

Figure A1: Graphical illustration of DEA (Drawn based on Coelli et al. 2005, figure 11.1)

Figure A1 shows the production frontier for CRS and VRS during year *t* and year *t*+1, respectively. Year *t*+1 is on a higher level of output, given the level of input, meaning that technical change has occurred. Furthermore, increasing returns to scale (IRS) can be observed, assuming VRS, to the left of the tangency between the CRS and the VRS frontier, indicating that the firms using less than this level of input are too small. Similarly, to the right-hand side of the tangency point, there are decreasing returns to scale, meaning that firms observed there would increase their productivity by becoming smaller.

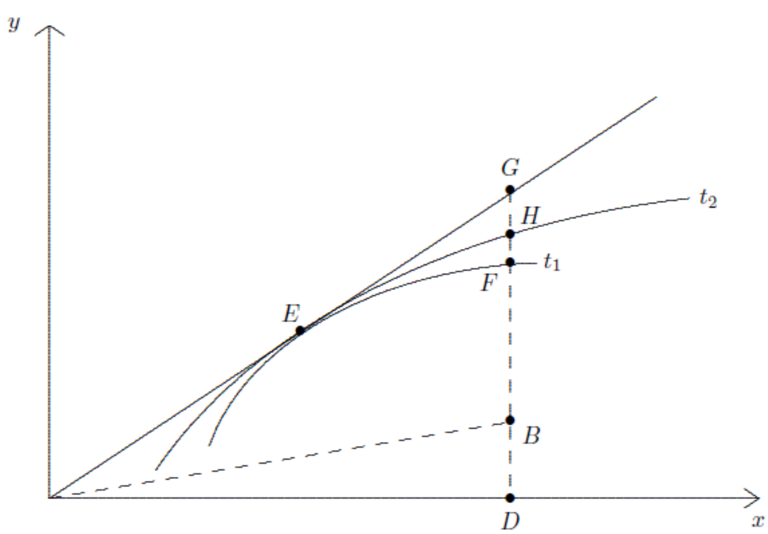

Figure A2: Description of PureTech and ScaleTech (taken from Simar and Wilson, 1998a).

In Figure A2 the change in pure and scale technology can be calculated as:

$$\Delta PureTech = \left(\frac{(DB/DF)/(DB/DF)}{(DB/DH)/(DB/DH)}\right)^{1/2} = \frac{DH}{DF} > 1$$

$$\Delta ScaleTech = \left(\frac{(DB/DG)/(DB/DF)}{(DB/DH)/(DB/DH)}\right)^{1/2} = \frac{DF}{DH} < 1$$

Both the $\Delta PureTech$ and the $\Delta ScaleTech$ consist of a geometric mean of two ratios. Looking at the points, it can be observed that TC has not changed, since the technology, assuming CRS, is constant. However, the VRS frontier in the second period is above the VRS frontier in the first period, given input amount D. Thus, the pure technology has increased and a flattening shape of the VRS frontier indicates a loss in scale.

Finally, the TFP change and its confidence intervals are reported for each court and year in figure A3-A5.

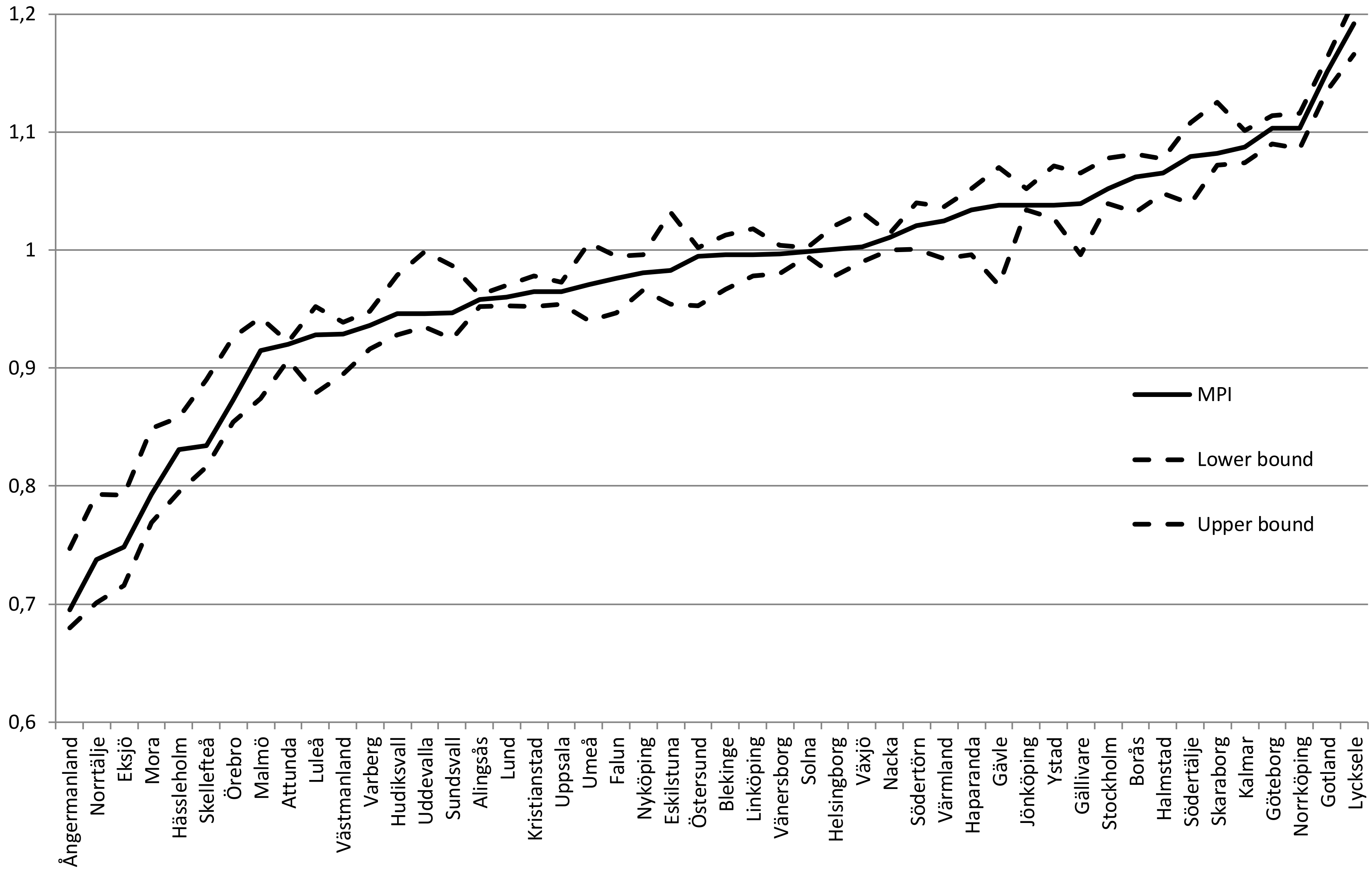

Figure A3: Computed MPI and confidence intervals 2012–2013

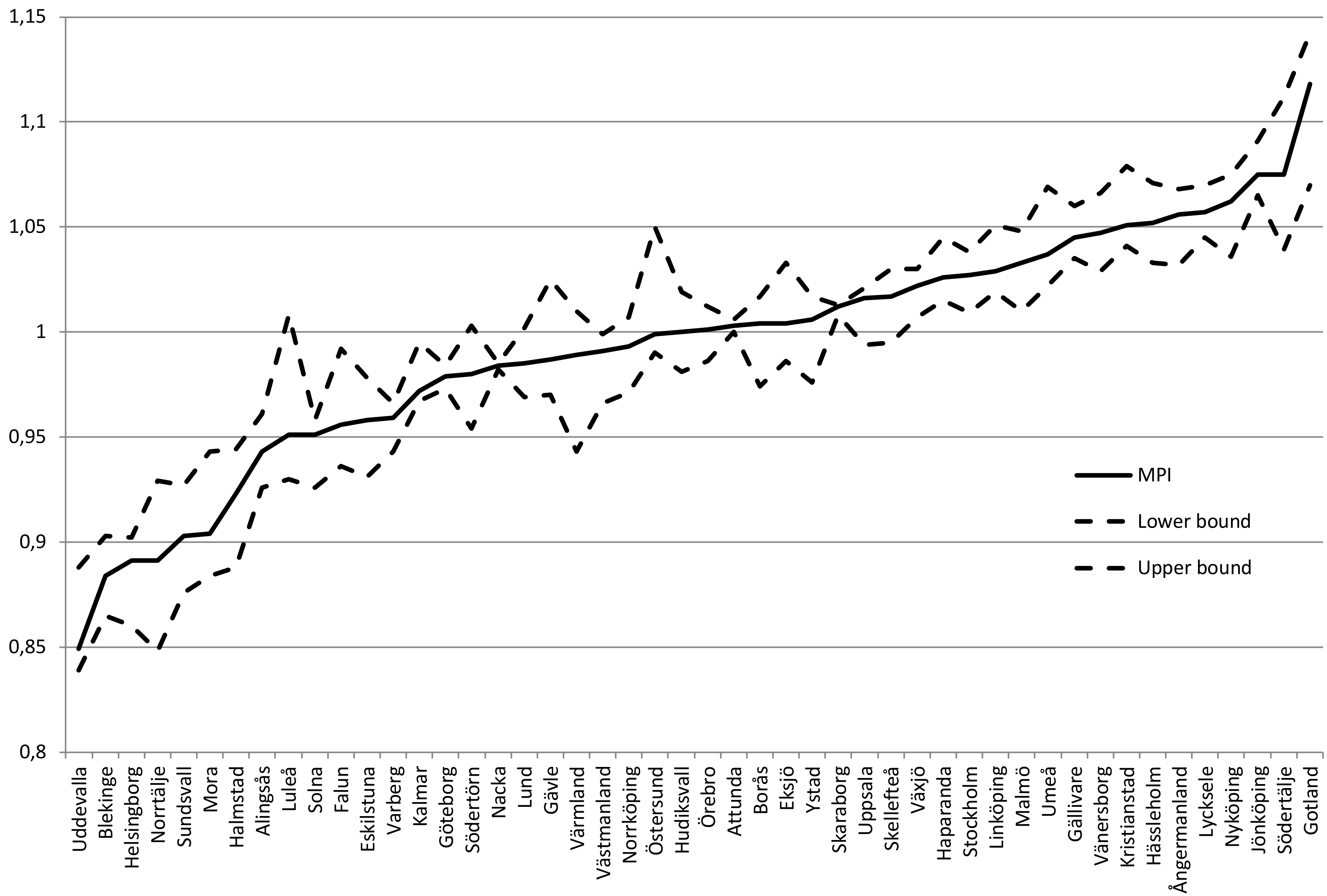

Figure A4: Computed MPI and confidence intervals 2013-2014

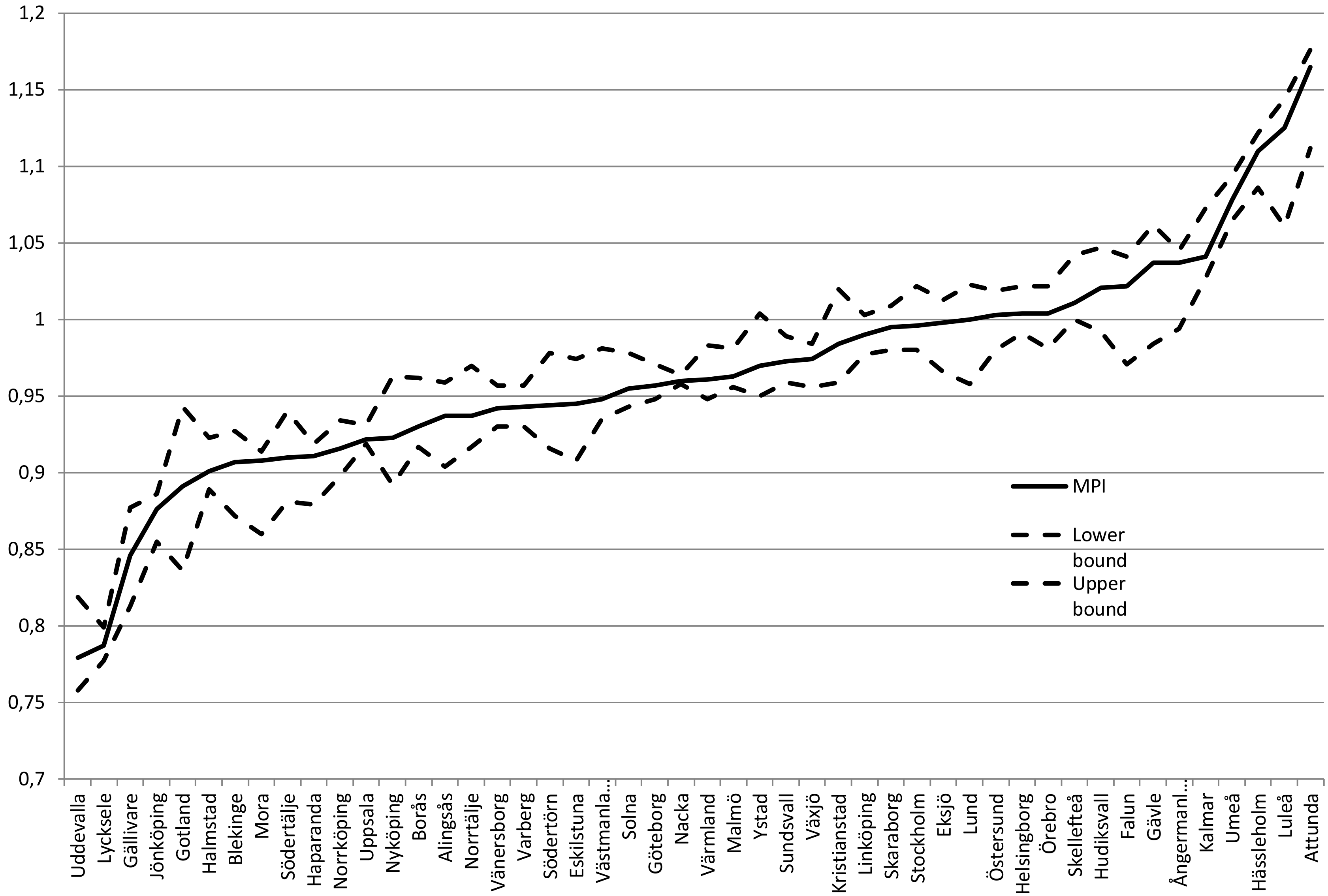

Figure A5: Computed MPI and confidence intervals 2014-2015