Incentive Schemes for Local Government : Theory and Evidence from Comprehensive Performance Assessment in England

Ben Lockwood and Francesco Porcelli

No 960

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF WARWICK

Incentive Schemes for Local Government: Theory and Evidence from Comprehensive Performance Assessment in England*

Ben Lockwood† Francesco Porcelli‡

First version: 18 November 2010
This version: 26 April 2011

Abstract

This paper studies Comprehensive Performance Assessment, an explicit incentive scheme for local government in England. Motivated by a simple theoretical political agency model, we predict that CPA should increase service quality and local taxation, but have an ambiguous effect on the efficiency of service provision. We test these predictions using a difference in difference approach, using Welsh local authorities as a control group, exploiting the fact that local authorities in Wales were not subject to the same CPA regime. To do this, we construct original indices of service quality and efficiency, using Best Value Performance Indicators. We estimate that CPA increased the effective band D council tax rate in England relative to Wales by 4%, and increased our index of service quality output also by about 4%, but had no significant effect on our efficiency indices. There is evidence of heterogenous effects of CPA on efficiency, with some evidence that CPA impacted more on less efficient councils, and the "harder test" from 2005-8 having a much bigger effect.

JEL Classification: H10,H70,H77,C21

Keywords: local government, incentives, efficiency, difference in difference, DEA

*We would like to thank Fabian Waldinger, Nicolas Serrano-Verlade, and participants at the conference "New issues on fiscal federalism and tax competition", Lyon, and the 2010 CESifo Political Economy conference in Dresden, for helpful comments.

†University of Warwick, Coventry CV4 7AL, England; Email: B.Lockwood@warwick.ac.uk

‡University of Warwick, Coventry CV4 7AL, England; Email: f.porcelli@warwick.ac.uk

1. Introduction

In recent years, explicit incentive schemes for public organizations, based on quantitative measurement of outputs, have become increasingly commonly used in the UK[1]. For example, school league tables, hospital star ratings, and various schemes for local government, such as Comprehensive Performance Assessment (CPA), have been introduced in the last twenty years or so. The focus of this paper is on CPA, the most important such scheme for local government. This scheme, introduced in 2001, rated local governments in England on the quality of service in six major areas: education, housing; social care; environment; libraries and leisure; use of resources. Hundreds of performance indicators and a variety of audit and inspection reports were collected, summarized, weighted, and categorized so as to arrive at final star ratings between 0 and 4[2].

As well as an evaluation scheme, CPA was also an incentive scheme. The stated objective of the CPA was to target support at those councils that need it most, and to offer a number of benefits for better-performing councils, including elimination of "ring-fencing" grants, and a three-year exemption from subsequent audit inspections[3]. Moreover, because the results of the CPA were widely disseminated in the media, it was also an exercise in providing voters with more information about the performance of their local council, both absolutely, and relative to other councils. In turn, this, in principle, provides *indirect* incentives for good performance. Indeed, there is evidence that councils which performed poorly on CPA were punished by the voters at subsequent elections[4].

CPA is of particular interest because it is, to our knowledge, the only explicit evaluation

[1]Schemes of this type have been little used outside the UK. There are exceptions: in the US, for example, the No Child Left Behind legislation punishes schools financially for poor test results, which are made public to parents.

[2]In fact, from 2002-5, the rankings were designated: "excellent", "good", "fair", "weak", and "poor", changing to zero to four stars during the latter part of CPA - see Tables A1 and A2 below. But, for simplicity, we refer to star ratings thoughout.

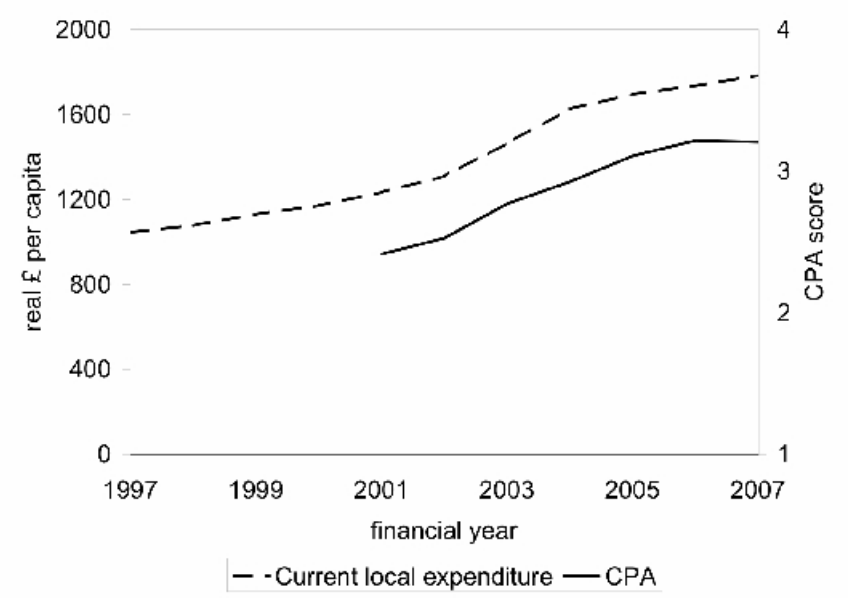
[3]"High scoring councils were Councils that were performing well under CPA would enjoy reduced audit and inspection regimes, and their associated fees, and be granted greater flexibilities and borrowing freedoms by central government. At the other end of the performance spectrum, a combination of audit, inspection and other improvement work was to be commissioned as an outcome of the CPA process, with the aim of transforming failing or poorly performing authorities." (Audit Commission(2009)).

[4]Revelli (2008) finds that an increase in one star rating increaes the probability that the incumbent party retains control of the council by seven percentage points, and Boyne et al. (2009) find "a low CPA score (0 or 1 star) increases the likelihood of a change in political control".

scheme to date, worldwide, that numerically scores and rewards *elected representatives*, as opposed to public service managements. The purpose of this paper is to assess the impact of CPA on local government in three dimensions: quality of service delivery, taxation policy, and the efficiency with which services were provided.

Figure 1 below shows the average CPA score achieved by English local authorities from the beginning to the end of the CPA experience together with the average current local expenditure. There is clearly a steady upward trend in average CPA star ratings. Indeed, the Audit Commission declared officially in 2009 (Audit Commission(2009)) that the CPA had done its job stimulating a continuous improvement in local government performance. However, Figure 1 also shows that at the same time, expenditure by local governments went up, more or less in line with CPA scores.

Figure 1: CPA Scores and Expenditures



So, the key problem is that *we do not observe the counterfactual*; given the large increases in local government spending over this period, it may be that service delivery would have improved anyway, even in the absence of the CPA. To address this, we treat the CPA as a natural experiment by exploiting the fact that it was only introduced in England, whereas in Wales, where the structure of local government is the same, a much weaker performance management scheme was introduced (Haubrich and McLean (2006b) Martin, et al (2010)). In particular, in Wales, there were no quantitative rankings, much less information published, and authorities also had a say with regard to the type of inspections they would like to see for specific services. So, we use local authorities in

Wales as a control group when assessing the impact of CPA on the treatment group, the English councils.

What would we expect the effects of a scheme such as CPA to be on service quality, tax levels, and efficiency? We develop a simple two-period political agency model in the paper to look specifically on the effect on taxation and efficiency of an incentive scheme that both rewards service quality and provides information about this quality to voters. In any period, the quality of a public good or service is determined by the politician's ability, effort, and tax revenue. In this environment, efficiency measures the level of service quality that can be produced at a given level of tax revenue. Voters value service quality and dislike taxes, and thus they care about both service quality and efficiency. The incumbent faces an election against a randomly selected challenger at the end of the first period. Our key predictions are as follows. The larger the direct reward, or the better the information, the more the incumbent taxes, and the higher the effort he makes. While higher effort is not surprising, the prediction of higher taxation, which voters dislike, is a distinctive feature of our theoretical analysis. As both effort and taxes rise, service quality is unambiguously increased by an incentive scheme. But, the effect of either a larger direct reward, or better information, on *efficiency* is ambiguous, because inputs, purchased by taxes, are also higher.

We then test these predictions, using Wales as a control group. Our results broadly confirm the predictions of the theory. First, looking across a number of different measures of revenue, the introduction of CPA appears to have raised council tax revenues in England relative to Wales. For example, we see that the introduction of CPA raised the effective band D council tax rate by about £40, or about 4%, in percent in England. To test the effects of CPA on quality of output and efficiency of local councils, we used specially constructed indices of both, described in more detail below (see also Porcelli(2010) on the efficiency index). We find, consistently with the theory, that the CPA raised our quality of output index by 4% above what it would have been, had English local councils also been subject to the same regime as in Wales.

But, again consistently with the theory, we find that CPA either had no significant effect on efficiency, or lowered it, depending on the efficiency index used, and the estimation method. So, our finding is consistent with the story that local authorities reacted to CPA by performing better, but also spending more. Therefore we conclude that CPA did not boost efficiency overall. This is in stark contrast to the view of the Audit Commission (2009) that CPA has "done its job" effectively.

We then look more closely at the impact of CPA on English councils. We do this in two ways. First, we look for evidence of a "catch-up" effect. That is: did CPA impact

more on less efficient than on more efficient English councils? We find that there is some evidence of a catch-up effect on output; output of English councils that were initially less efficient at the start of CPA increased by relatively more over the CPA period. But, they also raised their taxes by more, and perhaps as a consequence, there appears to be no catch-up effect on efficiency per se.

Our second approach notes that our theory predicts heterogenous treatment effects of CPA on efficiency. With the available data, we are able to test three of these. The first is that the efficiency effect of CPA is more likely to be negative, the larger the percentage of the population paying a reduced or zero rate of property tax. The second is that a less generous reward for service quality should increase the efficiency effect of CPA. The third is that the degree of electoral competition will effect the efficiency effect of CPA; we find that this effect is weakly negative. We find some empirical support for all of these, although the evidence in favour of the third is rather weak.

The rest of the paper is arranged as follows. Section 2 surveys related literature. Section 3 gives a brief overview of CPA. Section 4 develops the theoretical framework, and Section 5 describes our empirical strategy. Sections 6 and 7 give the baseline results and results on heterogenous treatment effects respectively, and Section 8 concludes.

2. Related Literature

There are several related literatures. First, there is an academic literature on the CPA itself. Boyne(2009) and Revelli(2008) have already been mentioned in the introduction. Revelli (2010) is perhaps the most closely related. In this last paper Revelli studies the link between council spending and CPA scores. In particular, he finds that spending in excess of the standard set by central government (standard spending assessment), can have a negative effect on the CPA score. His theoretical explanation for this is that some councils are more efficient than others in transforming expenditure into CPA scores. Moreover, he assumes that all councils have the same relative preferences for CPA scores and spending. In this environment, other things equal, a more efficient council will both spend less and achieve a higher CPA score.

In contrast, our paper constructs an explicit index of efficiency, *independent* of CPA, and asks how the introduction of CPA affects the efficiency of English councils, relative to Welsh ones. So, the two papers are quite different; we are more interested in the incentive effects of CPA, whereas Revelli(2010) is focussing on CPA as a measure of performance or outcome. Basically, as explained in the previous Section, we do not believe that the CPA is a good measure of either output or efficiency: rather, we are studying how it performed

as an incentive scheme.

A second related literature is the wider one on incentives in the public sector. This is surveyed by Burgess and Ratto(2003). Most relevant to our study is very recent and independent work by Burgess et. al. (2010). They use the abolition of school league tables in Wales (but not in England) in 2002 as a natural experiment to estimate the effect of league tables on secondary school performance. This is closely related to our study because one of the output indicators we use is the proportion of secondary school pupils achieving GCSE grades A to C in the local authority. Clearly, as CPA was introduced in England in the same year as school league tables were abolished in Wales, we cannot separately identify the effect of both reforms on school "output". To deal with this problem, we also test whether CPA increased our output index excluding education. We find that the effect of CPA is still significantly positive but smaller in magnitude. This is consistent with a story where both CPA and school league tables have positive effects on output.

Also related are Propper et al (2008), (2008a), and Besley, Bevan and Burchardi(2009), which are papers investigating the effect of the hospital star rating regime in England over 2001-5 on waiting times for hospital treatment, using either Scotland and Wales as control groups. The hospital star rating regime is similar in form to CPA, with good performance closely tied to reducing waiting lists. All three of these papers find strong evidence that the scheme had the desired effect on the targeted "output" i.e. waiting times were reduced in England relative to Scotland and Wales, although waiting times fell everywhere due to higher spending. Note also that all the papers just discussed only focus on single dimensions of local government "output"; unlike us, they do not address efficiency issues, or look at taxation.

Finally, our theoretical model modestly extends a literature on principal-agent problems where the agent has several tasks to perform, initiated by the classic paper of Holmstrom-Milgrom(1991). Holmstrom and Milgrom, however, restrict attention to a static framework, where monetary incentives can be used in an unrestricted way, and where the agent's payoff is exponential in money. Dewatripont, Jewitt, and Tirole(1999) extend that analysis to a career concerns framework, i.e. where the agent is rewarded not explicitly, but in proportion to their ability as inferred by the principal. There have been a few extensions[5] of the multi-task career concerns framework to political principal-agent

[5]Less closely related contributions include Besley(2004), Caselli and Morelli(2004), Messner and Polborn(2004), Mattozzi and Merlo(2008). These papers mostly focus on the effect of pay (fixed, not performance-related) on the incentive for different types of politicians to run for office. Besley(2004) also looks at the effect of varying pay on incentive and selection effects of elections.

problems, notably Gersbach(2008) and Alesina and Tabellini(2008). However, unlike us, neither of these papers allow for a specific reward being offered for one task[6].

3. The CPA - A Brief Overview

Local governments in England and Wales are of two types, unitary and two-tier. Unitary councils are responsible for primary and secondary education, social care, housing and housing benefit payments, waste disposal, transport, and environment, planning, and culture. Two-tier governments (counties) have the same responsibilities, except for housing and housing benefit, and environment, where responsibilities are shared with district councils.

In this institutional setting, the precursor to CPA, introduced in the Local Government Act 1999, was the "Best Value" framework, which "provides a framework for the planning, delivery and continuous improvement of local authority services. The overriding purpose is to establish a culture of good management in local government for the delivery of efficient, effective and economic services that meet the users' needs." (http://www.idea.gov.uk/idk). A key part of this framework were the Best Value Performance Indicators (BVPIs), which were numerical scores measuring the quality of the above services provided by individual councils on various dimensions. Importantly for our purposes, BVPIs were calculated for both English and Welsh councils.

CPA, which started in the 2001/02 financial year, was a move to a stricter assessment regime within the general Best Value framework. In the first three rounds, the method for assessing the current performance of a council was the following. Performance of councils was assessed in seven categories[7] (social care; environment; libraries and leisure; use of resources; education; housing; housing benefit payments). Where available, performance was assessed through already existing judgements from inspectorates and auditors, such as those by Office for Standards in Education (Ofsted) and Department for Education and Skills (DfES) for education. These were augmented with BVPIs. All this information was aggregated to obtain a score between 1 and 4 for each of the service blocks (with 1 being the

[6] Alesina and Tabellini study a sequence of models where the incumbent politician assigns effort to two tasks, and the level of performance on each task is fully observable, and depends additively on effort and ability, as in our setting. But, the main focus is on redistributive policies; each of two voter groups only benefits from the performance on one task, and the politician can make a transfer between these two groups. Finally, Gersbach(2008) considers a political agency model with moral hazard only, i.e. where politicians do not differ in ability, and where voters are able to precommit to a re-election rule.

[7] The CPA did not evaluate transport and planning.

lowest and 4 the highest). The scores were then weighted so that the scores for education and social services count four times, housing and environmental services twice, with the remaining blocks counting only once. These were then added up to produce a performance score of between 15 and 60 points, or 12 and 48 points for shire county councils (because they do not provide, and are therefore not assessed on, housing or benefits services). The performance scores were then aggregated to produce a performance rating of between 1 and 4 for each authority as shown in Table A2 of the Appendix. This score was then combined with an estimate of the councils' ability to improve (1 to 4) as explained in the Table A2 of the Appendix to produce the final CPA score.

In 2005, a new methodology, the "harder test" was introduced. The current performance of the LA was now assessed in the same categories with the exclusion of education, which was dropped. The main innovation, however, involved the aggregation procedure where the ability to improve was replaced by the corporate assessment, a three year period assessment of the council's ability "to lead its local community having clearly identified its needs and set clear ambitions and priorities" (Audit Commission, 2009). Among the service categories social care and use of resources received effectively a higher weighting than the other five through the aggregation rule in Table A3 in the Appendix.

So, what are CPA scores really measuring? Along with some commentators e.g. McLean et. al. (2007), we take the view that CPA is a hybrid measure, partly measuring levels of service quality (thorough the BVPIs), partly measuring operational efficiency (use of resources) and partly broader aspects of corporate health or effectiveness (ability to improve). As McLean et. al. (2007) point out, there may also be "categorization errors" in the aggregation procedure in Table A2, where fine numerical scores are compressed into just four categories. So, we take the view that CPA scores are measuring both service levels (output) and efficiency, and are doing so with some error. In this paper, we are not interested as CPA as a *measurement system*, but as an *incentive scheme*. That is why we construct our own, independent, measures of output and efficiency for local councils, with the aim of studying the effect of the CPA regime on this measures, along with taxation.

We close this section with some very direct evidence which suggests that CPA was acting partly as an incentive scheme. Note from Table A3 that during the harder test period, a score of at least 9 on so-called "Level 1 assessments" was a crucial threshold for getting the highest possible overall score, given a fixed corporate assessment. Also, it was the case that corporate assessment scores were fixed over a three year period , which were also the last three years in our sample (see Audit Commission 2009b). So, over the period 2004/5 to 2006/7 we would expect a "bunching" of scores on level 1 assessments at 9. This is clearly the case, as the histogram below shows.

Figure 3.1. Cut-off points histogram, level 1 assessment.

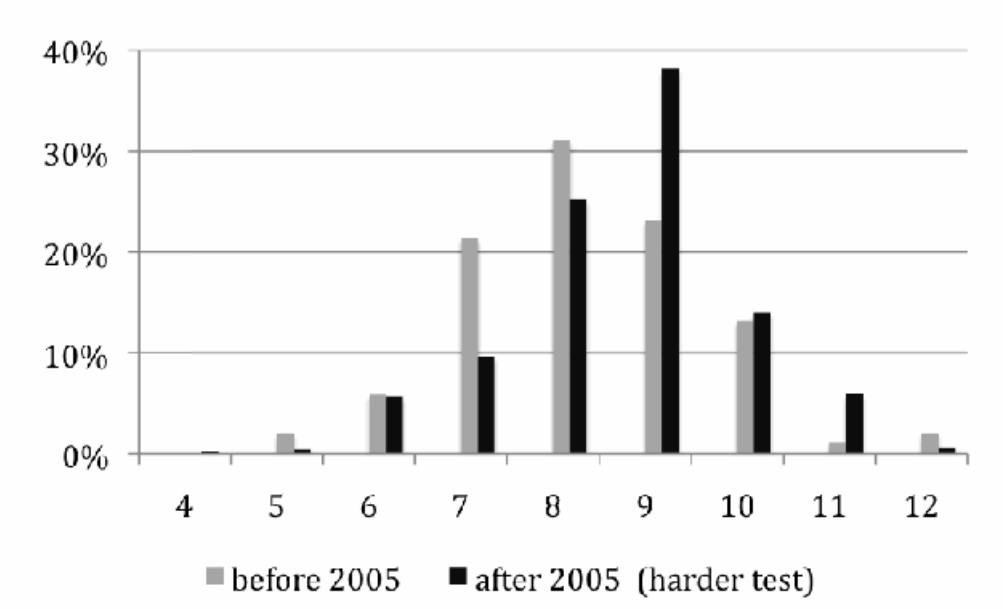


Figure 3.1 shows the levels 1 scores by year and council both before and after 2004/5. The years before 2004/5 constitute a benchmark, because a score of 9 was not a threshold before 2005.[8]

4. A Theoretical Framework

4.1. The Environment

In each of two periods $t = 1, 2$ an incumbent politician produces a local public good at quality level Q_t. This depends on resources, in the form of tax revenue T_t, chosen by the incumbent[9], plus an exogenous grant from central government G_t, the effort input of the

[8] In fact, before 2005, we do not know how the scores on individual services were aggregated to produce the points scores in Table A.2, because this information was never made public by the Audit Commission.

[9] This captures the stylized fact that the Council Tax (a residential property tax) is the only major tax instrument for local government in England and Wales. Over the sample period, due to incoming the Labour government's decision to abandon "rate-capping", local authorities have had in practice considerable autonomy to set their council taxes. The government reserves the right to direct an authority to set a lower budget requirement if it considers that the Council Tax has been increased excessively. However capping took place only in 2004/05 and 2005/06 for 6 and 8 local authorities respectively.

incumbent, a_t, and also his ability parameter η_t :

$$Q_t = \tilde{f}(a_t, S_t) + \eta_t, \ S_t = T_t + G_t \tag{4.1}$$

where S_t is total spending by local government. We assume $\tilde{f}(a_t, S_t)$ is strictly concave. Also, following Rogoff(1988), Alesina and Tabellini(2008), we assume that η_t follows a moving average process i.e. $\eta_t = \theta_t + \theta_{t-1}$ where θ_t is a random draw from a symmetric distribution with mean zero, distribution H,and density h, and support $[-\overline{\theta}, \overline{\theta}]$. Symmetry and zero mean are assumed for convenience only. At the beginning of $t = 1$, both the incumbent and voters know θ_0.

There is a continuum of measure 1 of voters. Voter $i \in [0, 1]$ has linear payoffs over Q_t and tax T_t of the form

$$u_t = Q_t - \mu_i T_t, \ t = 1, 2 \tag{4.2}$$

where μ_i is $i's$ tax price of public spending, and may differ across voters. In England and Wales, the only local tax is the property tax, so the natural interpretation of μ_i would be as that voter's property value relative to the average.

The incumbent politician, while in office, gains some office-related benefits, R, and also incurs a cost of effort. We also assume he puts some weight $\omega \geq 0$ on a weighted average of voter payoffs, either because he himself is a tax-payer and consumer of the local public good, as would be natural in a citizen-candidate setting (Besley and Coate(1997)), or because he is lobbied by special interest groups, or because he cares about his legacy (Maskin and Tirole(2004))[10]. This of course nests the purely office-seeking politician as a special case where $\omega = 0$. So, the politician in office has payoff

$$\omega(Q_t - \mu T_t) + R - ca_t, \ t = 1, 2 \tag{4.3}$$

where μ is the weighted average of the μ_i. As the unweighted average of the μ_i is one, $\mu < 1$ if the politicians put more weight on poorer groups, for example. Finally, following Maskin and Tirole(2004), we assume that when out of office, the politician has zero payoff.

There is an election at the end of period 1, described in more detail below. Also, the incentive scheme is only used in period 1 and is described in more detail below. This simplifies the exposition, and in the two-period model, is without much loss of generality[11].

[10] It is beyond the scope of this paper to provide micro-foundations for these processes.

[11] An infinite-horizon version of this model is available on request where it is possible to distinguish between temporary and permanent incentive schemes; the qualitative effects of the two are similar.

4.2. Political Equilibrium

4.2.1. Equilibrium in Period 2

Substituting (4.1) into (4.3), we see that the incumbent politician's expected payoff is

$$\omega(f(a_2, T_2) + \theta_1 - \mu T_2) + R - ca_2$$

where $f_2(a_2, T_2) \equiv \tilde{f}_2(a_2, G_2 + T_2)$. So, the politician's decision problem is characterized by

$$V(\theta_1) = \max_{a_2, T_2} \{\omega(f(a_2, T_2) + \theta_1 - \mu T_2) + R - ca_2\} \equiv V + \omega\theta_1 \tag{4.4}$$

where $V(\theta_1) = V + \omega\theta_1$ has the interpretation of the incumbent's expected continuation payoff at time 1, given observation of θ_1, but before θ_2 is known.

4.2.2. The Incentive Scheme and Equilibrium in Period 1

We begin by describing the incentive scheme. The politician gets a bonus B per unit of output, i.e. BQ_1. This can be interpreted as monetary or psychological. Obviously, the second interpretation is appropriate in the case of CPA, as local officials - elected or not - do not get any direct personal payment as a result of a good CPA score. Moreover, in view of the important role in practice that CPA and other incentive schemes play in giving voters better information, we assume that in period 1, *voters only observe output* Q_1 *before the election with probability* q. We suppose that this q can be increased by the incentive scheme; we refer to this as the *information effect* of the scheme. As Q_1 appears in the utility function, voters must observe it after the election i.e. at the end of period t, if they do not observe it earlier. Finally, it is assumed that voters always observe T_1 before the election, reflecting the fact that local property taxes are highly "visible".

The order of events in period 1 is then as follows. First, politicians choose a_1, T_1, knowing θ_0. Then, voters vote for incumbent or challenger, having observed T_1 and, with probability q, Q_1. The challenger's productivity is randomly drawn from the same distribution as the incumbent's. Finally, at $t = 1$, voters and the incumbent both know θ_0.

First, consider the voter choice between the incumbent and challenger. We impose the natural condition that this decision must be sequentially rational i.e. the voters cannot precommit to a voting rule. Because distributional concerns, measured by μ, are fixed and the same across all politicians, voters only care about the productivity η_1 of the incumbent and challenger. Given this, voter behavior is easily characterized. First, given knowledge

of θ_0, it is easily seen[12] that at the end of period 1, voters have enough information to infer θ_1. So, once Q_1, T_1 are observed, voters infer that the incumbent's type is

$$\theta_1^e = Q_1 - f(a_1^e, T_1) - \theta_0 \tag{4.5}$$

where a_1^e is the voters' expected value of effort choice by the incumbent.

Now, we assume that voters vote rationally, up to a random error. In particular, if Q_1 has been observed, a voter votes for the incumbent if

$$\theta_1^e + \varepsilon + v \geq 0 \tag{4.6}$$

where ε is an idiosyncratic popularity shock, distributed uniformly on $[-1, 1]$ and ν is an aggregate popularity shock, distributed uniformly on $[-\frac{\sigma}{2}, \frac{\sigma}{2}]$, where σ parameterizes the variance of this shock. In what follows, σ will be an (inverse) measure of electoral competition i.e. the lower σ, the more sensitive is the re-election probability to performance.

What if Q_1 has not been observed? Then, the voters cannot make any inference about θ_1, and so they are indifferent between the incumbent and challenger. In this case, we assume that they randomize between the two, so the incumbent is re-elected with probability 0.5.

Standard computations (see Appendix) then imply that, conditional on θ_t the probability of re-election for the incumbent is

$$p(a_1, T_1; a_1^e, \theta_1) = \frac{1}{2} + \frac{q}{\sigma}\left(\theta_1 + f(a_1, T_1) - f(a_1^e, T_1)\right) \tag{4.7}$$

That is, the probability of re-election is higher, the higher the actual competence of the incumbent, and the higher output relative to output expected by voters. Note that the responsiveness of p to an increase in a_1 is proportional to $\frac{q}{\sigma}$; i.e. voters are more responsive to performance, the greater is electoral competition, or the more informative is the incentive scheme. Note also that in equilibrium, where $a_1^e = a_1$, choice of T_1 does not affect the re-election probability; this is because it is directly observed by voters, not inferred.

So, given the re-election probability (4.7), at time t, the incumbent solves the following problem

$$\max_{a_1, T_1}\left\{(\omega + B)f(a_1, T_1) - \omega\mu T_1 - ca_1 + \delta\int_{-\overline{\theta}}^{\overline{\theta}} p(a_1, T_1; a_1^e, \theta_1)V(\theta)h(\theta)d\theta\right\} \tag{4.8}$$

[12] Assume that voters know θ_0. Then, as voters have observed Q_1, T_1 by the end of period 1, and voters also know that the incumbent has taken equilibrium action a_1, they can infer θ_1 from the relationship $Q_1 = f(a_1, T_1) + \theta_0 + \theta_1$.

where $\delta < 1$ is a discount factor, and $V(\theta)$ is defined in (4.4). This has the interpretation that the incumbent maximizes his current payoff, plus his expected continuation payoff, if re-elected.

Then, the first-order conditions to this problem, evaluated in equilibrium, where $a_1^e = a_1$ are

$$(\omega + B + \delta qV/\sigma)f_a = c \tag{4.9}$$

$$(\omega + B)f_T = \omega\mu \tag{4.10}$$

Note that (4.9) says that there are three motivations for the incumbent to supply effort; some preference congruence with the electorate ($\omega > 0$), career concerns, measured by $\delta qV/\sigma$, and finally the incentive scheme, B. Note also the asymmetry; career concerns affect the choice of effort, but not tax, ultimately because the voters can *directly* observe tax, but can only *indirectly* infer effort.

So, given B and other parameters, the endogenous variables a_1, T_1 are simultaneously determined from the two equations (4.9), (4.10) and V is residually defined by (4.4). This constitutes a political equilibrium.

4.3. An Alternative Interpretation

Although this model has been presented as one of an elected representative being motivated by voters via an election, in the British context, there is an alternative, and possibly more plausible, interpretation[13]. Councils in England and Wales have the following management structure; strategic decision-making is undertaken by an executive comprised of elected officials, typically in the form of a cabinet with the leader elected by council members, with day-to-day operations headed by a full-time CEO. One could argue that CPA is also a management tool for the executive to monitor the CEO. One can therefore re-interpret our model as follows.

Voters can be plausibly re-interpreted as councillors, who live in the council district and who therefore have similar preferences to voters. The "politician" can be re-interpreted as the council CEO, who can be fired or otherwise sanctioned for poor performance. Thus, the election can be reinterpreted as any action that the executive can take to discipline the CEO. CPA is of value to councillors either because it gives them more information about CPA performance (higher q), or because there are direct benefits to the CEO of a higher CPA score i.e. earned autonomy. This re-interpretation is of course, applicable

[13] This was suggested to us by Tim Besley.

to other contexts where there is also a clear division between legislative and executive functions, such council–manager local government in the US.

4.4. Effects of an Incentive Scheme

Here, we establish our main results of the effects of an incentive scheme. We consider the effects of small changes in both B and q on taxation, T_1, output, the expected value of Q_1, and also on "efficiency", defined more precisely below. Note that up to a constant, the expected value of Q_1 is simply $f(a_1, T_1) \equiv f_1$. Our first result, proved in the Appendix, is:

Proposition 1. *If T, a are weak complements i.e. $f_{aT} \geq 0$ then: (a) $\frac{da_1}{dB} > 0, \frac{dT_1}{dB} > 0$, and so $\frac{df_1}{dB} > 0$; (b) $\frac{da_1}{dq} > 0, \frac{dT_1}{dq} > 0$, and so $\frac{df_1}{dq} > 0$.*

So, we see that a stronger incentive scheme, interpreted as an increase in B and/or q, will unambiguously increase both taxes and expected output. Note also that this result does not depend on the relative size of the direct effect and the information effect of the incentive scheme. This is important, because in the empirical work, we cannot estimate the effects of B and q separately.

We now turn to look at efficiency. In our setting, the natural measure of efficiency, and the one that will be used in our empirical work, is the expected output f_1, minus the cost of inputs, T_1;

$$e \equiv f(a_1, T_1) - T_1 \tag{4.11}$$

From (4.11), the effect of B or q on efficiency is :

$$\frac{de}{dB} = f_a \frac{da_1}{dB} + (f_T - 1)\frac{dT_1}{dB}, \ \frac{de}{dq} = f_a \frac{da_1}{dq} + (f_T - 1)\frac{dT_1}{dq} \tag{4.12}$$

So, we see immediately that an increase in B or q has a an *ambiguous* effect on efficiency; there is a positive effect via a_1, but an effect that can be negative via T_1. Specifically, this effect will be negative if the incumbent is already collecting too much tax revenue at the margin $f_T < 1$. In turn, from (4.10), we can see intuitively that this is more likely to be the case if the bonus B is already large, there are strong career concerns, or μ, the politician's disutility of tax, is small enough.

To pin down these effects more precisely, assume that f is Cobb-Douglas. Then we can prove:

Proposition 2. *Assume $f = a^\alpha T^\beta, \ \alpha, \beta > 0, \ \alpha + \beta < 1$. An increase in q increases*

efficiency iff $\frac{\mu\omega}{\omega+B} > \beta$. *An increase in* B *increases efficiency iff*

$$\frac{\mu\omega}{\omega+B} > \frac{(2-\frac{1}{\alpha})+(1-\frac{1}{\alpha})\frac{\delta qV}{\sigma(\omega+B)}}{(4-\frac{1}{\beta}-\frac{1}{\alpha})+(2-\frac{1}{\alpha})\frac{\delta qV}{\sigma(\omega+B)}} \tag{4.13}$$

Note also that by straightforward computation[14], the right-hand side of (4.13) is increasing in $\frac{\delta qV}{\sigma(\omega+B)}$. So, as q, V, σ only appear on right-hand side of (4.13), we see that introducing a small incentive scheme, or increasing the incentive scheme by a small amount, can *decrease* efficiency if: (i) career concerns V are strong e.g. from (4.4), the ego-rent R from office is high; (ii) electoral competition is high i.e. σ is low; (iii) if μ is low, reflecting a e.g. low tax price facing the median voter; (iv) the incentive scheme is more informative to voters i.e. large q. Some of these results may appear counter-intuitive, but they all arise from the fact that the higher "career concerns", measured by $\frac{\delta qV}{\sigma}$, the greater the tendency to set a high tax in the first period, in order to boost output and get re-elected. In turn, from (4.12), if the tax is high enough, it can lead to lower efficiency.

5. Empirical Strategy

5.1. Empirical Specification

Our empirical approach is to estimate the impact of CPA on efficiency in a quasi-experimental setting through difference-in-difference estimation, using Wales, where CPA was not used, as a control group. Welsh local government performance was assessed by an evaluation program called the Welsh Program for Improvement (WPI) since 2001.[15] We believe that Welsh councils can be used to address the counterfactual question of what would have been the path of English councils after 2001 if CPA league tables would not have been produced, for the following reasons. First, Welsh and English local authorities have the same structure and functions.[16] Second, the mean values of our control variables and the input and output variables used to construct our service quality and efficiency indices are very similar in the two countries (see Tables A5 and A6 in the Appendix). Third, as documented by Haubrich and McLean(2006b), WPI was, compared to CPA, a

[14] Let the RHS be $f(x) = \frac{(2-\frac{1}{\alpha})+(1-\frac{1}{\alpha})x}{(4-\frac{1}{\beta}-\frac{1}{\alpha})(\omega+B)+(2-\frac{1}{\alpha})x}$, $x = \frac{\delta qV}{\sigma(\omega+B)}$. Then $f'(x)$ has the sign of $1-\alpha-\beta$, which is positive.

[15] Information and data about the Welsh Program for Improvement can be collected from the web site of the Wales Audit Office www.wao.gov.uk.

[16] All Welsh local authorites are unitary, but they have the same responsibilities as English local authorities, and until 2006, the same funding structure.

much less prescriptive and elaborate assessment regime since only confidential assessments were produced, the evaluation criteria were based only on local self-assessment without quantitative rankings, and no formal rewards or punishments were specified.[17]

As a further check, we test whether our identification assumption holds by testing whether our dependent variables i.e. the Y_{it} in (5.1) follow a common time path in the years before the introduction of CPA in 2001. First, as a "visual" test, Figures A.1, A.2, and A.3 in Appendix D show the common trend followed by the three main dependent variables in England and Wales before the introduction of CPA. Moreover, using a formal test, also reported in Appendix D, we find that with a few exceptions, we cannot reject the hypothesis that they did follow a common time path.

So, we proceed by estimating the following;

$$Y_{it} = \beta_1 CPA_t + \beta_2 (CPA_t \times D_i) + \alpha' \mathbf{X}_{it} + u_i + \varepsilon_{it} \tag{5.1}$$

where CPA_t is a dummy that takes value 1 after 2001 and D_i is a dummy that takes value one for English councils. Also, $X_{\mathbf{it}}$ is a vector of controls. In $X_{\mathbf{it}}$, we included linear and quadratic time trends. These are general, not specific to each council, and the results are basically unchanged if they are omitted. Finally, $Y_{it} = T_{it}, Q_{it}, e_{it}$, where T_{it} is a measure of revenue collected from the council tax, Q_{it} a measure of service quality, and e_{it} a measure of efficiency. The main parameter of interest here is β_2 which captures the treatment effect of the CPA. The theory suggests that if $Y_{it} = T_{it},\ Q_{it}$, then $\beta_2 > 0$ but if $Y_{it} = e_{it}, \beta_2$ has an ambiguous sign theoretically.

We have two different treatments of u_i. First, we treat u_i as a council fixed effect. Then, we treat it as a random effect, estimated using GLS[18]. Finally, some of the variables (all the outputs, and one of the tax variables) are between zero and 1, so also, as a robustness check, we estimate a non-linear model where the dependent variable is transformed to lie between zero and 1 as follows:

$$T_{it} = \Phi(\beta_1 CPA_t + \beta_2 (CPA_t \times D_i) + \beta' \mathbf{Z}_{it} + u_i) + v_{it}$$

using a pooled Bernoulli quasi-MLE,[19]. In this case, u_i is treated as fixed.

[17] It is important to note that the greater regional autonomy obtained in Wales at the end of the 1990s does not interfere with our analysis since the The National Assembly for Wales was created by the Government of Wales Act in 1998 and gained a limited primary legislative powers only in 2007.

[18] The Mundlak (1978) approach will be followed in the estimation of the random effects model. In this approach, the time-average of time-varying regerssors are included as asdditional regressors, in order to tackle the possibility that the unobserved heterogeneity and the regressors may not be orthogonal.

[19] We are using the methodology proposed by Papke and Wooldridge (2008) to tackle the possibility

Standard errors in (5.1) are clustered at the council level, allowing for serial correlation in the ε_{it}. Finally, we take account of the problem, raised by Bertand, Duflo, and Mullainathan (2004) that even with clustered standard errors, there can be downward bias in the standard error in β_2. To deal with this, we follow the procedure recommended in that paper of collapsing the time dimension to before and after the treatment, and re-estimate all of our specifications. As can be seen by comparing tables A.7-A.9 in Appendix C to Tables 6.1-6.3 and 7.2-7.3 below, the results are robust to this alternative estimation method, providing evidence that serial correlation is not a problem.

5.2. Data

First, we discuss our choice of measures of T_{it}, Q_{it}, e_{it} for English and Welsh councils over the period 1997-2007. In the theory, T_{it} is property tax revenue. The closest empirical proxy for this is the tax requirement in the official statistics (CIPFA(2008a)) which is total (real) spending in the financial year minus revenue from the revenue support grant and other grants, and revenue from the business tax rate. We use the tax requirement, both as a raw figure, and normalized in several ways. Specifically, we also measure T_{it} as a percentage of the tax requirement to the budget requirement, where the latter is actual current expenditure that has to be financed by formula grants (which includes the police grant) and property tax revenue. Finally, we divide the tax requirement by the number of equivalent band D dwellings to get an effective council tax rate.

Next, we turn to the measurement of service quality Q_{it}. We need to construct an index of service quality consistently across both English and Welsh local governments. To that end the BVPIs published by the Audit Commission for England and the Audit Office for Wales are the best source of information: first they are broadly accepted by the local governments as measures of output quality; second we are very confident about the comparability of these measures across local authorities since BVPIs were also chosen as one of the building blocks of the CPA procedure.

The first problem to solve was the absence of BVPIs for the housing and benefit sector in case of the counties, where this function is managed by districts. As DEA requires observations for all units in all years, the only possible solution was to drop this sector from the efficiency analysis. A further problem worth discussing is the short life of many BVPIs. Despite the fact that we could count more than 250 BVPIs published on the website of the Audit Commission, almost all of them have been subject to some changes

of non linearity in case of fractional dependent variable. In the non-linear model we also include council fixed effects.

after three or four years, and in many cases replaced with new indicators. There is also the problem that after 2001-2, BVPIs were defined and measured separately in both England and Wales, and there was very little overlap. In the end only five indicators could be used to measure the quality of output consistently for England and Wales; these measure aspects of education, social care of the elderly and children, waste disposal, and central services. These variables are defined in Table A4 of the Appendix, and summary statistics are given in table A5. But, it is important to note that expenditure on these categories accounts for fully 57% of the total local government expenditure, on average[20].

As is clear from that table, four of the five BVPIs are already expressed as percentages; we converted BVPI54 to a percentage also, and averaged it with BVPI49, thus giving an aggregate index for social services. We then calculated Q_{it} as the weighted average of these four indices, where the weights used were the relative expenditures on the four services, in real £ per pupil for education, and real £ per capita for the others, where all monetary amounts were deflated using the 2005 CPI. Summary statistics on these expenditures are given in the bottom panel of Table A5. The source for the expenditure data is from the Finance and General Statistics (FGS) and Local Government Comparative Statistics (LGCS), available on the website of the Chartered Institute of Public Finance and Accountancy (CIPFA) from the 1997/98 to the 2007/08 financial years (CIPFA (2008b) and CIPFA (2008c)).

Our efficiency index e_{it} is constructed as follows. We estimate a Debreu-Farrell[21] efficiency index e_{it} for each council and year in the sample using data envelopment analysis (DEA hereafter).[22] As output measures, we use the same five BVPIs used to construct the output index. As inputs, we use the expenditures already mentioned, corresponding to those outputs. DEA generates two indices. The first, the input index, e_{it}^{IN}, has the following intuitive interpretation. If council i was using the technology efficiently at time t, its inputs could all be scaled down by a fraction $1 - e_{it}^{IN}$ and it would still be able to produce the vector of outputs $\mathbf{y}_{it}$. The second, the output index, e_{it}^{OUT} has a similar interpretation: if council i was using the technology efficiently at time t its outputs could all be scaled up by an amount e_{it}^{OUT}, whilst using the same vector of inputs $\mathbf{x}_{it}$. Formal

[20]Remarkably, if one takes a less demanding view, and only requires identical BVPIs measured in England and Wales in only *one* year before, and *one* year after, the introduction of CPA, which is a minimal requirement for difference in difference analysis, there was just one additional BVPI available, the percentage of recycled household waste that was used to generate energy.

[21]Debreu (1951) and Farrell (1957).

[22]DEA was first developed by Charnes, Cooper, and Rhodes (1978); a complete survey of data evelopment analysis can be found in Ali and Seiford (1993).

definitions are given in Appendix B, and some descriptive statistics are provided in Table A6. Finally, it is also explained in Appendix B that $e_{it}^{IN}, e_{it}^{OUT}$ will generally be upward biased. So, as also explained there, we used bootstrap methods to correct for that bias, yielding bias-corrected versions of both $e_{it}^{IN}, e_{it}^{OUT}$.

Finally, our control variables X_{it} are described in Table A6 of the Appendix, and can be subdivided in the following sub-categories. First, are demographic variables, such as the percentage of the total population below the age of 16 and above the age of 75, the percentage of population that declare themselves religious, the percentage of white people, the population density, and finally the percentage of households who own their house, the number of band D equivalent dwellings per capita that correspond to the tax base of the council tax and has been included as a proxy of the demand for local public services. The second category includes a set of dummy variables to capture the impact of the ruling party and the features of the electoral system ("all out" election every four years, or "by thirds" system which involves more frequent elections). The third group of variables is related to the structure of the local economy and includes: average household disposable income, the percentage of the workforce claiming unemployment-related benefits, the percentage of people below 65 claiming disability living allowance, the percentage of VAT tax payers in the financial and real estate sector, the percentage of high qualified workforce, and the percentage of workforce self employed[23].

6. Empirical Results

The first, empirical prediction of the theoretical model is that CPA should increase council tax revenues (Proposition 1). So, we first estimate (5.1) with $Y_{it} = T_{it}$. As a first pass, Figure A.1 of Appendix D shows that the council effective tax rate (the tax requirement per equivalent band D dwelling) exhibits a clear increase in England relative to Wales after 2002. So, we would expect β_2 to be significantly positive. For each of the three tax measures described above, we estimate three specifications of (5.1), as described in Section 5.1 above.

Table 6.1 shows that irrespective of the estimation method and with all three tax measures, β_2 is positive and significant at the 1% level. According to out linear estimates, (FE and RE), the introduction of CPA raised the tax requirement by about £24 per capita, or 7.5% in England relative to Wales, the tax requirement as a percentage of the budget requirement by about 2.5 percent in England relative to Wales, and finally

[23]Due to the absence of some data on control variables in some years, the panel is unbalanced.

raised the effective council tax rate by about £52, corresponding roughly to a 4 percent increase in England relative to Wales[24]. For the non-linear model, the average partial effect is reported, which is the percentage change in the dependent variable caused by the treatment. So, in this model, the effect of CPA is somewhat smaller.

Table 6.1. Point estimates of the treatment effect of CPA on council tax revenues.

Model	Tax requirement (real £ per capita) (A)	Tax requirement (% of budget requirement) (B)	Effective council tax rate (real £ per band D equivalent dwelling) (C)
FE (linear)	23.98***	2.487***	52.23***
	(4.99)	(0.433)	(11.15)
RE (linear)	23.97***	2.487***	52.23***
	(5.03)	(0.435)	(11.20)
GLM (non linear) (1)	n.a.	1.814***	n.a.
		(0.187)	
Observations	1846	1846	1846
Number of councils	170	170	170
Control variables	yes	yes	yes
Quadratic trend	yes	yes	yes

Clustered standard errors in brackets. *** significant at 1%. Coefficient point estimates are interpreted as follows: £ per capita in column (A), % change in column (B), and £ per dwelling in column (C).

(1) Point estimates are expressed as average partial effects

Second, we estimate β_2 in (5.1) when the dependent variable is our service quality index. As a first step, Figure A.2 in Appendix D shows clearly that the output index rose faster in England than in Wales after the introduction of CPA, so we would expect $\beta_2 > 0$. Point estimates of β_2 are reported in the first column of Table 6.2 using the same econometric specifications as in Table 6.1. Also in this case, irrespective of the estimation method, it is possible to observe a positive and statistically significant effect of CPA on the level of outputs: on average, after the introduction of CPA, the aggregate output increased by 4% in English councils compared to Welsh local authorities.

But, as remarked in Section 2, a concern for us is that secondary school performance, as measured by the percentage of pupils achieving between A and C in GCSE exams, is

[24]Note that the estimates of £24 and £52 are broadly consistent, using the fact that there are on average, according to the latest statistics (CIPFA (2008a)), about 2.3 persons per dwelling in England.

a major component of our output index (with a weight of 63%). Burgess et. al. (2010) show that this measure of performance was impacted by school league tables, which were abolished in Wales it the same year in which CPA was introduced in England. To deal with this problem, we also test whether CPA increased our output index excluding education. The results are in column 2 of Table 6.2: we find that in our baseline fixed effects specification, the effect of CPA is still significantly positive but smaller in magnitude. We believe that this is evidence that both CPA and other "service-specific" performance indicators such as school league tables can have positive effects on output.

The other columns of Table 6.2 display the results of the same exercise conducted considering the quality measures of each sector. So, for English local authorities, in general, there is empirical evidence of a positive effect of CPA on all quality measures, with the exception of the percentage of household waste recycled. Thus, our results are again broadly consistent with the theory.

Table 6.2. Point estimates of the treatment effect of CPA on service quality.

Model	Output measures					
	Aggregated output	Aggregated output (no education)	Education	Social service	Central services	Environment
FE (linear)	4.13***	1.16*	5.63***	1.50**	3.74*	-4.98***
	(0.62)	(0.70)	(0.83)	(0.73)	(2.26)	(0.90)
RE (linear)	4.11***	1.15	5.62***	1.47**	3.77*	-4.97***
	(0.63)	(0.71)	(0.84)	(0.74)	(2.27)	(0.91)
GLM	4.08***	1.32***	5.80***	1.48***	3.54***	-5.26***
(non linear) (1)	(0.39)	(0.18)	(0.50)	(0.52)	(1.16)	(0.73)
Observations	1746	1746	1846	1797	1783	1804
No. of councils	170	170	170	170	170	170
Control variables	yes	yes	yes	yes	yes	yes
Quadratic trend	yes	yes	yes	yes	yes	yes

Clustered standard errors in brackets. *** significant at 1%; ** significant at 5%; * significant at 10%. Coefficient point estimates are interpreted as percentage change in output index due to CPA.

(1) Point estimates are expressed as average partial effects.

We turn to look at the effect of CPA on our efficiency indices. Figure A.3 of Appendix D shows the path of the efficiency index in England and Wales (average between input and output approach) between 1997 and 2007. In both countries the initial decreasing trend

in efficiency reversed its course after the introduction of CPA, and although the initial gap between Welsh and English councils is completely closed in the last year, there is no clear evidence of a positive impact of CPA on the efficiency of English local authorities. This suggests an insignificant β_2, which is in fact what we find. Our econometric specifications are the same as in the previous two tables, except for the third specification. In this case, to account for the possibility of non-linearity, we exploit the fact that the DEA indices of efficiency have an ordinal meaning; therefore we use as a dependent variable a binary indicator that will take value one if the council is ranked above the 50th percentile in the distribution of the DEA efficiency scores, and zero otherwise. This gives a random effect probit model estimated using the unconditional MLE estimator.[25]

Looking at Table 6.3, there is no empirical evidence in favour of a an impact of CPA on the efficiency of English councils. The coefficient of the treatment effect is statistically significant only in case of RE probit model in relation to the input approach, however the magnitude of the estimate tell us that after the introduction of CPA the probability of observing a council ranked in the upper 50th percentile of the efficiency index distribution decreased by 0.3%, a very small number that leads us to the conclusion that the introduction of CPA did not stimulate any change in the efficiency of English local authorities in delivering public services. Finally, our estimates confirm the presence of a generalized quadratic trend in case of the linear model.

[25]It is important to note that also in this case, like for the RE model, the Mundlak (1978) approach will be followed in order to tackle the possibility that the unobserved heterogeneity and the regressors may not be orthogonal.

Table 6.3. Point estimates of the treatment effect of CPA on efficiency.

Model	Input approach		Output approach	
	no bootstrap	bootstrap	no bootstrap	bootstrap
FE (linear) (1)	-2.86	-2.88	1.08	0.68
	(1.91)	(2.20)	(0.88)	(1.42)
RE (linear) (1)	-2.95	-3.32	1.05	0.86
	(1.89)	(2.17)	(0.88)	(1.38)
RE probit (non linear) (2)	-0.04	-0.30***	-0.03	0.03
	(0.11)	(0.11)	(0.10)	(0.13)
Observations	1548	987	1548	932
Number of councils	170	170	170	169
Control variables	yes	yes	yes	yes
Quadratic trend	yes	yes	yes	yes

Clustered standard errors in brackets. *** significant at 1%; ** significant at 5%; * significant at 10%.

(1) Coefficient point estimates are interpreted as percentage change in efficiency index due to CPA.

(2) Dependent variable is 1 in year t iff council is ranked in the upper 50th percentile of the efficiency distribution in year t. Coefficient point estimates are interpreted as percentage change in the probability of being ranked in the 50th percentile of the efficiency index distribution.

7. Heterogenous Treatment Effects

In this Section, we investigate how - if at all - the effect of CPA depends qualitatively on LA characteristics. We investigate this by estimating the following specification:

$$e_{it} = \beta_1 CPA_t + \beta_2(CPA_t \times D_i) + \beta_3(Z_{it} \times CPA_t) + \beta_4(CPA_t \times D_i \times Z_{it}) + \alpha' \mathbf{X}_{it} + u_i + \varepsilon_{it}$$

where Z_{it} is a local government characteristic, that could, for example, proxy for one of the theoretical characteristics described in Proposition 2. So, this allows us to estimate the heterogeneous treatment effects of CPA. Note that this specification, via the inclusion of $Z_{it} \times CPA_t$, and $CPA_t \times D_i \times Z_{it}$ allows for characteristic Z_{it} to have separate effects on e_{it} in both England and Wales before and after CPA.[26] Therefore, in this case, the parameter of interest is β_4.

[26] Note that $\mathbf{X}_{it}$ includes also Z_{it}.

The first possible heterogenous effect that we investigate is whether CPA had a "catch-up" effect - that is, whether it had a greater impact on badly-performing councils in England than on well-performing councils. To do this, we rank councils in England and Wales by their average efficiency score over the period prior to the introduction of CPA i.e. 1997-1999.[27] We then split the sample at the median, with $Z_{it} = 1$ if the council is below the median time-averaged efficiency score. So, in this case, Z_{it} is independent of i. In Table 7.1 below, we present results on the estimates of β_4 for council tax, output, and efficiency. The format of the table is in line with those used previously; that is we report the estimate of β_4 for fixed and random effects specifications. Moreover for the aggregate output and the efficiency index we also consider two different non-linear specifications as we did respectively in Tables 6.2 and 6.3.

In column 1 of table 7.1, we see evidence that relative to "good" English councils, inefficient English councils increased council tax by significantly more (around £14 in terms of the real per capita tax requirement) during the CPA period. As column 2 shows, this is reflected in an increased relative output performance of more that 2% in terms of aggregate output, although it should be said that this effect is less significant if education is excluded (not reported). Perhaps as a result of these two countervailing effects, there is no evidence that "bad" English councils increase their efficiency relative to "good" English councils. So, there appears to be catch-up in output, but no evidence of a catch-up effect on efficiency.

[27] Recall that data for Wales on the efficiency index is missing for the year 2000, because we lack output information.

Table 7.1. Heterogeneous treatment effect of CPA on taxation, aggregate output and efficiency for initially inefficient English LAs.

Model	Tax requirement (real £ per capita)	Aggregated output	Efficiency output approach
	(A)	(B)	(C)
FE (linear)	13.61*	2.19***	-1.40
	(7.90)	(0.64)	(1.65)
RE (linear)	14.50*	2.43***	-1.43
	(7.91)	(0.62)	(1.13)
GLM (non linear) (1)		0.04***	
		(0.01)	
RE probit (non linear) (2)			-0.15
			(0.14)
Observations	1846	1746	1539
Number of councils	170	170	169
Control variables	yes	yes	yes
Quadratic trend	yes	yes	yes

Clustered standard errors in brackets. *** significant at 1%; ** significant at 5%; * significant at 10%.

(1) Point estimates are expressed as average partial effects.

(2) Percentage change in the probability of being ranked in the 50th percentile of the efficiency index distribution after 1% change in the variable interacted with the treatment effect.

Our second approach to heterogenous treatment effects is theory-motivated; we look for reasonable empirical proxies for the variables that according to Proposition 2, might affect the sign of $\frac{de}{dB}$. First, we begin with μ, which is unambiguously predicted to increase the efficiency of an incentive scheme. Recall that this is the politician's perceived tax price. Suppose (reasonably) this is equal, or close to, the median tax price in the jurisdiction. Given the tax is a property tax, then μ is lower, other things equal, if a larger fraction of the adult population who do not pay the property tax, or pay some reduced fraction of the council tax. In the UK, the main groups who do not pay the full amount of council tax on properties they own or rent are the unemployed and those on low incomes, who are eligible for Council Tax Benefit (CTB). For example, in 2010, 68% of those in receipt of CTB were claiming Jobseeker's allowance, incapacity benefit, or income support. We do not have data on CTB or income support recipients by council, so, we proxy μ (inversely) by

the % of the workforce in receipt of Jobseeker's allowance, plus % of the adult population under 65 in receipt of incapacity benefit.

The estimates of β_4 are shown in Table 7.2 below. We expect $\beta_4 < 0$ as our variable is an inverse measure of the tax price. The format of the table is the same as in the case of Table 6.3. That is, we report the estimate of β_4 for fixed and random effects specifications, and for a probit where the dependent variable is 1 if unit of observation i was ranked in the top 50th percentile of the distribution, and zero otherwise. In case of the linear models β_4 point estimates can be interpreted as the percentage increase in efficiency observed in English local authorities, given a 1% increase in Z_{it}, after the introduction of CPA. In case of non linear model β_4 point estimates are displayed in terms of average partial effect, i.e. they exhibit the percentage change in the probability of observing a council ranked efficient (i.e. above the 50th percentile in terms of DEA efficiency scores) after 1% increase in Z_{it} that follows the introduction of CPA in English local authorities. We see that the estimated β_4 is mostly negative, and is certainly negative whenever it is significant.

Table 7.2. Heterogeneous treatment effect, % jobseeker's allowance and incapacity benefit.

Model	Input approach		Output approach	
	no bootstrap	bootstrap	no bootstrap	bootstrap
FE (linear) (1)	-0.041	0.028	0.007	-0.010
	(0.074)	(0.082)	(0.025)	(0.029)
RE (linear) (1)	-0.108*	-0.075	-0.034**	-0.049**
	(0.062)	(0.065)	(0.016)	(0.020)
RE probit (non linear) (2)	-0.260	-0.165	-0.510**	-0.661*
	(0.241)	(0.261)	(0.255)	(0.343)
Observations	1548	987	1548	932
Number of councils	170	170	170	169
Control variables	yes	yes	yes	yes
Quadratic trend	yes	yes	yes	yes

Clustered standard errors in brackets. *** significant at 1%; ** significant at 5%; * significant at 10%.

(1) Percentage change in efficiency after 1% change in the variable interacted with treatment effect.

(2) Percentage change in the probability of being ranked in the 50th percentile of the efficiency index distribution after 1% change in the variable interacted with the treatment effect.

Next, recall that Proposition 2 predicts that an increase in B can decrease efficiency, and generally, will have some impact either way. The introduction, after 2004, of the

"harder-test" procedure for the computation of CPA scores allows to test the impact of B on efficiency. In particular, it is quite plausible to assume that the intensity of the bonus (B) decreased after the new regime became effective since it was more difficult to obtain a high CPA score. In this last case Z_{it} corresponds to a dummy which takes value one after 2004, and the point estimates reported in table 7.3 support our claim providing robust empirical evidence in favour of a positive effect of the "harder-test" on LA's efficiency.

Table 7.3. Heterogeneous treatment effect, "harder test".

Model	Input approach		Output approach	
	no bootstrap	bootstrap	no bootstrap	bootstrap
FE (linear) (1)	2.67*	5.70***	2.06***	1.78***
	(1.58)	(1.84)	(0.48)	(0.66)
RE (linear) (1)	2.65*	5.81***	2.06***	1.765***
	(1.59)	(1.91)	(0.48)	(0.64)
RE probit (non linear) (2)	0.045	0.24***	0.19**	0.22
	(0.07)	(0.08)	(0.09)	(0.14)
Observations	1548	987	1548	932
Number of councils	170	170	170	169
Control variables	yes	yes	yes	yes
Quadratic trend	yes	yes	yes	yes

Clustered standard errors in brackets. *** significant at 1%; ** significant at 5%; * significant at 10%.

(1) Percentage change in efficiency after 1% change in the variable interacted with treatment effect.

(2) Percentage change in the probability of being ranked in the 50th percentile of efficiency index distribution after 1% change in the variable interacted with the treatment effect.

A third variable that can be empirically proxied is the degree of electoral competition. Our available measure of this is the vote share of the party with the largest number of votes at the last election, minus the vote share of the party with the second most number of votes. Call this the *vote gap*. What does this correspond to in the theory? The difference in vote shares between the incumbent and the challenger is $q(2s-1)$, where $s = \frac{1}{2}(1+\nu+\theta_1)$ is the share of voters voting for the incumbent when voting is "non-random" i.e. when Q_1 is observed, from (4.7), using $\theta_t^e = \theta_t$. Conditional on θ_1, the

difference in vote shares, between the winner and the loser is thus

$$\begin{aligned} & qE_\nu[\nu+\theta_1\,|\nu \geq -\theta_1\,] + qE_\nu[-(\nu+\theta_1)\,|\nu < -\theta_1\,] \\ & = \frac{q}{2\sigma}[\frac{\theta_1^2}{2} + \theta_1\sigma + \frac{\sigma^2}{2}] - \frac{q}{2\sigma}[-\frac{\theta_1^2}{2} + \theta_1\sigma - \frac{\sigma^2}{2}] \\ = & \frac{q}{2\sigma}[\theta_1^2 + \sigma^2] \end{aligned}$$

Taking the expectation with respect to θ_1, we see that theoretical equivalent of the observed vote gap is

$$\Delta = \frac{q}{2\sigma}E_{\theta_1}[\theta_1^2 + \sigma^2] = \frac{q}{2\sigma}[\sigma_\theta^2 + \sigma^2]$$

This is non-monotonic in σ : increasing when $\sigma > \sqrt{\sigma_\theta^2}$, and decreasing otherwise. Now, from Proposition 2, efficiency is predicted to be unambiguously increasing in σ. So, our empirical prediction is that the *efficiency effect of CPA should depend on the vote gap,* but may be decreasing or increasing.

Table 7.4. Heterogeneous treatment effect, vote gap.

Model	Input approach		Output approach	
	no bootstrap	bootstrap	no bootstrap	bootstrap
FE (linear) (1)	0.025*	0.039*	0.003	0.008
	(0.013)	(0.022)	(0.006)	(0.008)
RE (linear) (1)	0.010	0.016	0.000	0.006
	(0.010)	(0.014)	(0.004)	(0.005)
RE probit (non linear) (2)	0.095	0.206*	0.058	0.127
	(0.070)	(0.108)	(0.081)	(0.113)
Observations	1045	628	1045	632
Number of councils	170	170	170	169
Control variables	yes	yes	yes	yes
Quadratic trend	yes	yes	yes	yes

Clustered standard errors in brackets. *** significant at 1%; ** significant at 5%; * significant at 10%.

(1) Percentage change in efficiency after 1% change in the variable interacted with treatment effect.

(2) Percentage change in the probability of being ranked in the 50th percentile of efficiency index distribution after 1% change in the variable interacted with the treatment effect.

Some evidence of a significant effect is given in Table 7.4; the effect of the vote gap on efficiency via CPA is always positive, and sometimes significant.

8. Conclusions

This paper has studied Comprehensive Performance Assessment, an explicit incentive scheme for local government in England. A simple theoretical political agency model predicted that CPA should increase service quality and local taxation, but have an ambiguous effect on the efficiency of service provision. We tested these predictions using a difference in difference approach, using Welsh local authorities as a control group, exploiting the fact that local authorities in Wales were not subject to the same CPA regime. We also constructed indices of service quality and efficiency, using Best Value Performance Indicators as well as expenditures on different categories of services. We estimate that CPA increased the effective band D council tax rate in England relative to Wales by 4%, and increased our index of service quality output also by about 4%, but had no significant effect on our efficiency indices. There is evidence of heterogenous effects of CPA on efficiency. Relative to "good" English councils, initially inefficient English councils increased council tax by significantly more during the CPA period. This is also reflected in an increased relative output performance. Perhaps as a result of these two countervailing effects, there is no evidence that "bad" English councils increase their efficiency relative to "good" English councils. So, there appears to be catch-up in output, but no evidence of a catch-up effect on efficiency. Finally, consistently with the theory, the "harder test" from 2005-8 having a much bigger effect, and also the effect of CPA on efficiency seems to be more negative in local authorities where there are larger numbers of voters who face a zero "tax price".

References

Alesina, A., Tabellini, G., (2008), "Bureaucrats or politicians? Part II: Multiple policy tasks", *Journal of Public Economics* 92, 426–447.

Ali, A., Seiford, L. M., (1993), "The mathematical programming approach to efficiency analysis", in Fried, A. O. Lovell, A. K., Schmidt, S. S. (Eds), *The measurement of Productive Efficiency*, Oxford University Press, Ch. 3, 120-159.

Andrews, R., (2004), "Analyzing deprivation and local authority performance: The implications for CPA", *Public Money and Management*, 24(1), 19-26.

Andrews, R., Boyne, G. A., Law, J., and Walker, R. M., (2005), "External constraints on local service standards: the case of comprehensive performance assessment in English local government", *Public Administration*, 83(3), 639-656.

Audit Commission, (2009)." Final score: The impact of cpa of local government, 2002-2008", *www.audit-commission.gov.uk.*

Audit Commission, (2009b)."Scorecard explained", *www.audit-commission.gov.uk.*

Bertrand, M., Duflo, E., Mullainathan, S., (2004), "How much should we trust differences-in-differences estimates?", *The Quarterly Journal of Economics*, 119(1), 249–275.

Besley T. J., Coate S., (1997), "An Economic Model of Representative Democracy", *The Quarterly Journal of Economics*, 112(1), 85-114.

Besley T. J., (2004), "Paying politicians: theory and evidence", *Journal of the European Economic Association* 2(2–3), 193–215.

Besley T. J., (2006), *Pricipled Agents? The Political Economy of Good Government*, Oxford University Press.

Besley, T. J., Bevan, G., Burchardi K.B., (2009), "Naming & Shaming: The impacts of different regimes on hospital waiting times in England and Wales", *CEPR Discussion Papers* 7306.

Boyne, G. A., James, O., John, P., and Petrovsky, N., (2009), "Political and managerial succession and the performance of English local governments", *The Journal of Politics*, 71, 1271-1284.

Burgess, S., and Ratto, M., (2003), "The Role of Incentives in the Public Sector: Issues and Evidence", *Oxford Review of Economic Policy*, 19(2), 285-300.

Burgess, S., Wilson, D., Worth, J., (2010), "A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting", *CMPO Working Paper Series*, No. 10/246.

CIPFA, (2008b), *Finance and General Statistics 2007/08*, The Chartered Institute of Public Finance and Accountancy. London.

CIPFA, (2008c), *Local Government Comparative Statistics 2008*. The Chartered Institute of Public Finance and Accountancy. London.

Charnes, A., Cooper, W. W., Rhodes, E., (1978), "Measuring the Efficiency of Decision-Making Units", *European Journal of Operational Research*, 2(6), 429-444.

Debreu, G., (1951), "The coefficient of resource utilization", *Econometrica*, 19(3), 273-292.

Dixit, A., 1997. "AssociationPower of Incentives in Private versus Public Organizations", *The American Economic Review*. 87(2), 378-382.

Dewatripont, M., Jewitt, I., Tirole, J, (1999), "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*. 66(1), 183-198.

Dewatripont, M., Jewitt, I., Tirole, J, (1999), "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*. 66(1), 183-198.

Farrell, M. J., (1957), "The measurement of productive efficiency", *Journal of the Royal Statistical Society*, 120(3), 253-281.

Gersbacha, H., Liessemb, V., (2008), "Incentive contracts and elections for politicians with multi-task problems", *Journal of Economic Behavior & Organization*, 68, 401–411.

Gutierrez-Romero, R., Haubrich, D., and McLean, I., (2008), "The limits of performance assessments of public bodies: external constraints in English local government", *Environment and Planning C: Government and Policy*, 26, 767-787.

Gutierrez-Romero, R., Haubrich, D., and McLean, I., (2010), "To what extent does deprivation affect the performance of English local authorities?", *International Review of Administrative Sciences*, 76, 137-170.

Haubrich, D. and McLean, I., (2006a), "Assessing public service performance in local authorities through cpa - a research note in deprivation", *National Institute Economic Review*, 197, 93-105.

Haubrich, D. and McLean, I., (2006b), "Evaluating the performance of local government a comparison of the assessment regimes in England, Scotland and Wales", *Policy Studies*, 27(4), 271-293.

Holmstrom, B., Milgrom, P., (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", J*ournal of Law, Economics and Organization*, 7(0), 24-52.

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

Martin, S., Downe, J., and Grace, C., (2010), "Validity, utilization and evidence-based policy: The development and impact of performance improvement regimes in local public services", *Evaluation*, 16(31), 31-42.

Maskin E., Tirole, J., (2004), "The Politician and the Judge: Accountability in Government", *The American Economic Review*, 94(4), 1034-1054.

Mundlak, Y., (1978), "On the Pooling of Time-Series and Cross-Section Data" *Econometrica*, 46, 68-86.

Papke, L. E. and Wooldridge, J. M., (2008), "Panel data methods for fractional response variables with an application to test pass rates", *Journal of Econometrics*, 145, 121-133.

Propper, C., Sutton, M., Whitnall, C., Windmeijer, F., (2008), "Incentives and Targets in Hospital Care: Evidence from a Natural Experiment" *CMPO Working Paper Series*, No. 08/205.

Propper, C., Sutton, M., Whitnall, C., Windmeijer, F., (2008a), "Did 'targets and terror' Reduce English Waiting Times for Elective Hospital Care?", *CMPO Working Paper Series* No. 07/179.

Revelli, F., (2008), "Performance competition in local media markets", *Journal of Public Economics*, 92, 1585-1594.

Revelli, F., (2010), "Spend more, get more? An inquiry into English local government performance", *Oxford Economic Papers*, 62, 185-207.

Rogoff, K,., & Sibert, A., (1988), "Elections and Macroeconomic Policy Cycles" *Review of Economic Studies*, 55(1), 1-16.

Simar, L. and Wilson, P. W., (1998), "Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models", *Management Science*, 44(1), 49-61.

Simar, L. and Wilson, P. W., (2000), "A general methodology for bootstrapping in nonparametric frontier", *Journal of Applied Econometrics*, 27(6), 779-802.

Simar, L. and Wilson, P. W., (2007), "Estimation and inference in two-stage, semi-parametric models of production processes", J*ournal of Econometrics*, 136, 31-64.

Stewart, M. B., (2004), "The impact of the introduction of the uk minimum wage on the employment probabilities of low-wage workers", *Journal of the European Economic Association*, 2(1), 67-97.

A. Appendix

A.1. Appendix A: Proofs of Propositions and Other Results

Computation of Equation (4.7). Assume that Q_t has been observed by voters. Conditional on ν_t, the fraction of voters who vote for the incumbent is

$$\Pr[\varepsilon_t \geq -\nu_t - \theta_t^e + c] = \frac{1 + \nu_t + \theta_t^e - c}{2} = f_t,$$

So, the incumbent only wins if $f_t \geq \frac{1}{2}$, *i.e.* $\nu_t \geq -\theta_t^e$. So, the probability that the incumbent wins is

$$\Pr(\nu_t \geq -\theta_t^e) = \Pr(\nu_t \leq \theta_t^e) = \frac{1}{2} + \frac{\theta_t^e}{\sigma} \tag{A.1}$$

Now from (4.5) and (4.1), we have

$$\theta_t^e = \theta_t + f(a_t, T_t) - f(a_t^e, T_t) \tag{A.2}$$

Finally, the overall probability of a win is

$$p_t = (1 - q)\frac{1}{2} + q.\Pr(\nu_t \geq -\theta_t^e) \tag{A.3}$$

Combining (A.1),(A.2), (A.3), the result then follows.

Proof of Proposition 1. From total differentiation of (4.9), (4.10), and application of Cramer's rule, we have:

$$\begin{aligned} \frac{da_1}{dB} &= \frac{-f_a f_{TT}(\omega + B) + f_{aT} f_T(\omega + B + \delta q V/\sigma)}{D} \\ \frac{dT_1}{dB} &= \frac{-f_T f_{aa}(\omega + B + \delta q V/\sigma) + f_{aT} f_a(\omega + B)}{D} \end{aligned} \tag{A.4}$$

where $D = (f_{aa}f_{TT} - f_{aT}^2 > 0)(\omega + B)(\omega + B + \delta q V/\sigma) > 0$ by the second-order conditions to the incumbent's optimization problem. [This is automatically satisfied as $f_{aa}f_{TT} - f_{aT}^2 > 0$, from strict concavity of f, and $\omega, B, \delta q V/\sigma > 0$]. So, as $f_{aT} \geq 0$, we see that $\frac{da_1}{dB}, \frac{dT_1}{dB} > 0$. Also, in the event of a change in q :

$$\frac{da_1}{dq} = \frac{-f_a f_{TT}(\omega + B)\frac{\delta V}{\sigma}}{D}, \frac{dT_1}{dq} = \frac{f_{aT} f_a(\omega + B)\frac{\delta V}{\sigma}}{D} \tag{A.5}$$

A similar argument then proves that $\frac{da_1}{dq}, \frac{dT_1}{dq} > 0$. □

Proof of Proposition 2. From (4.12), we can write:

$$\frac{de}{dq} \geq 0 \Leftrightarrow \frac{f_a \frac{da_1}{dq}}{f_T \frac{dT_1}{dq}} > \frac{1 - f_T}{f_T} \Leftrightarrow f_T > \frac{1}{1 + \frac{f_a \frac{da_1}{dq}}{f_T \frac{dT_1}{dq}}} \tag{A.6}$$

But from (A.5), and $f = a^{\alpha}T^{\beta}$,we have

$$\frac{f_a \frac{da_1}{dq}}{f_T \frac{dT_1}{dq}} = \frac{1-\beta}{\beta} \tag{A.7}$$

Combining (A.6), (A.7) gives the condition $f_T > \beta$. Again from (4.12), we have:

$$\frac{de}{dB} \geq 0 \Leftrightarrow \frac{f_a \frac{da_1}{dB}}{f_T \frac{dT_1}{dB}} > \frac{1-f_T}{f_T} \Leftrightarrow f_T > \frac{1}{1+\frac{f_a \frac{da_1}{dB}}{f_T \frac{dT_1}{dB}}} \tag{A.8}$$

But from (4.11), and $f = a^{\alpha}T^{\beta}$,we have

$$\begin{aligned} \frac{f_a \frac{da_1}{dB}}{f_T \frac{dT_1}{dB}} &= \frac{f_a^2 f_{TT}(\omega+B) + f_a f_{aT} f_T(\omega+B+\delta qV/\sigma)}{f_T^2 f_{aa}(\omega+B+\delta qV/\sigma) + f_T f_{aT} f_a(\omega+B)} \\ &= \frac{(2-\frac{1}{\beta})(\omega+B)+\delta qV/\sigma}{(2-\frac{1}{\alpha})((\omega+B)+(1-\frac{1}{\alpha})\delta qV/\sigma} \end{aligned} \tag{A.9}$$

Combining (A.8) and (A.9) gives (4.13). □

A.2. Appendix B: Construction of the Efficiency Indices

Assume that the LA has q outputs that can be produced from l inputs, and $\mathbf{y}, \mathbf{x}$ denote the output and input vectors respectively. The production possibility set is $S = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^{l+q} \mid \mathbf{y} \leq F(\mathbf{x})\}$, where F characterizes the efficient frontier. Then the input requirement set $X(\mathbf{y}) = \{x \in \mathbb{R}_+^l \mid (\mathbf{x}, \mathbf{y}) \in S\}$ is the the set of inputs required to obtain a particular output quantity. Then for each input-output combination for LA i at time t, the associated efficiency measure is $e_{it} = \min\{e \in \Re \mid e\mathbf{x}_{it} \in X(\mathbf{y}_{it})\}$. So, in the case of the input approach, e_{it} is the solution of the following linear program providing the efficiency score for the council i in period t:

$$\underset{e\in\Re,\lambda\in\Re^{NT}}{min}\ e \quad s.t.\ \ e\mathbf{x_{it}} \geqslant \mathbf{X}\lambda, \quad \mathbf{Y}\lambda \geqslant \mathbf{y_{it}}; \quad \lambda \geqslant 0; \quad \iota'\lambda = 1 \tag{A.10}$$

where $\mathbf{x_{it}}$ is the vector of inputs of council i at time t, $\mathbf{X}$ is $l \times NT$ the matrix of inputs of all N LAs over all T years, $\mathbf{Y}$ is the $q \times NT$ matrix of outputs of N LAs over all T years, λ is a $NT \times 1$ vector of optimal weights, $\mathbf{y_{it}}$ is the vector of outputs of council i at time t, and ι' is a $1 \times NT$ vector of $(1, ...1)$. The last constraint is important for imposing variable returns to scale. Note that we are taking a pooled approach where only one production frontier is estimated and each region is compared also with itself in another year. In this way it is possible to use all the $N \times T$ observations.

The output approach is similar; the output possibility set $Y(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^q_+ \mid (\mathbf{x}, \mathbf{y}) \in S\}$ is the the set of output vectors that are possible given input vector $\mathbf{x}$. Then the output efficiency measure is $1/e_{it}$, where $e_{it} = \max\{e \in \Re \mid e\mathbf{y}_{it} \in Y(\mathbf{x}_{it})\}$. It is calculated using a similar linear program to (A.10).

The main problem with DEA is that it tends to produce an upward-biased estimate of the true Debreu-Farrell measure of technical efficiency. The bias is due to the piece-wise shape of the DEA frontier that approximates the true unobserved frontier. As a result DEA underestimates the distance of all input/output combinations from the true frontier. Typically the bias, as well as the precision of the its estimation, become smaller as the number of observations increases and becomes larger as we increase the dimensions of the production function (see Kneip et al. 1998). In this study, although more than 1500 observations are available, considering both English and Welsh councils in the production function, a "bias corrected" measure of efficiency, $\tilde{e}_{it}$ along with its interval of confidence at the 95% level of significance, CI_{it}, has been computed following the bootstrap methodology developed by Simar and Wilson (1998, 2000).

After estimating our bias corrected measure of efficiency we found that the Spearman correlation between e_{it} and $\tilde{e}_{it}$ is 0.96 and 0.93 in cases of input and output approaches respectively. Therefore, given the large number of observations, the magnitude of the bias is not a big issue in this case, in fact e_{it} and $\tilde{e}_{it}$ provide very similar regression results. The main concern is that $\tilde{e}_{it}$ may be imprecisely estimated. The precision of the estimate of $\tilde{e}_{it}$ is measured by the width of the CI_{it}. So, to check the robustness of our bias-corrected measure of efficiency, we drop observations where the efficiency index is too imprecisely measured. To do this, first calculate the quartiles of the distribution of the $\tilde{e}_{it}$. We then retain observation $\tilde{e}_{it}$ only if CI_{it} lies entirely in one quartile; otherwise, we drop it. As a result, we have constructed a sub-sample of statistically "significant" bias-corrected indices of efficiency. As shown in Table A1 is possible to keep 64% of the DEA bias-corrected efficiency indices in case of input approach, and 60% in case of output approach.

Table A1. Statistically significant efficiency scores.

Type of local authorities	Total observations	Input Approach		Output Approach	
	e_{it} NxT	Statist. Significant	%	Statist. Significant	%
English Counties	329	203	61%	202	61%
London Boroughs	304	228	75%	190	62%
English Metr. Districts	333	192	57%	196	59%
English Unitary Authorities	407	264	64%	253	62%
Welsh Unitary Authorities	179	102	56%	94	53%
Total	1552	989	64%	935	60%

A.3. Appendix C: Tables

Table A2. CPA aggregation rule, first three rounds.

Counties		London, MD, UA	
Performance score	Category score	Performance score	Category score
Less then 24 points	1	Less then 30 points	1
24 to 29 points	2	30 to 37 points	2
30 to 36 points	3	38 to 45 points	3
More than 36 points	4	More than 45 points	4

Councils' ability to improve	Councils' performance score on core services			
	1	2	3	4
1	poor	poor	weak	n.a.
2	poor	weak	fair	good
3	weak	fair	good	excellent
4	n.a.	good	excellent	excellent

Table A3. The Harder Test aggregation rule.

Corporate assessment	Level 1 assessment (children and adults social care, use of resources)	Level 2 assessment (environment, culture, housing and benefit)	CPA final score
4	None less than 3	None less than 2	4 stars
4	None less than 2	No more than one less than 2	3 stars
4	No more than one less than 2	No more than one less than 2	2 stars
4	Any other combination	Any other combination	1 star
3	None less than 3	None less than 3	4 stars
3	None less than 2	None less than 2	3 stars
3	None less than 2	No more than one less than 2	2 stars
3	Any other combination	Any other combination	1 star
2	None less than 3	None less than 3	3 stars
2	None less than 2	None less than 2	2 stars
2	No more than one less than 2	No more than one less than 2	1 star
2	Any other combination	Any other combination	0 stars
1	None less than 3	None less than 2	2 stars
1	None less than 2	None less than 2	1 star
1	Any other combination	Any other combination	0 stars

Table A4. Description of output variables.

Service	BVPI code	Description	Period
Secondary Education	BVPI38	Percentage of 15 year old pupils in schools maintained by the local education authority achieving five or more GCSEs at grades A*-C or equivalent	Average over the current and the three following academic years
Social services (children)	BVPI49	The percentage of looked after children with no more than three placements during the last financial year (BVPI49)	Financial year
Social services (elderly)	BVPI54	Older people helped to live at home per 1000 population aged 65 or over (BPVI54)	Financial year
Waste disposal	BVPI82a	Percentage of household waste arising which have been sent by the Authority for recycling (BPVI82a)	Financial year
Central services	BVPI8	Percentage of invoices paid by the Authority within 30 days of receipt or within the agreed payment terms (BVPI8)	Financial year

Table A5. Descriptive statistics, output and input variables.

	Mean		Std. Dev.		Observations	
	England	Wales	England	Wales	England	Wales
Output variables						
Secondary education	51.10	51.04	9.51	6.43	1373	179
Social service (children)	11.92	9.36	4.21	4.67	1373	179
Social service (adults)	85.51	100.29	26.97	36.83	1373	179
Waste disposal	11.66	14.35	5.67	8.75	1373	179
Central services	82.80	82.77	10.99	9.62	1373	179
Input variables						
Secondary education (real £ per pupil)	3503	3203	728	392	1373	179
Social service, children and adults (real £ per capita)	205	209	76	56	1373	179
Waste disposal (real £ per capita)	21	22	7	9	1373	179
Central services (real £ per capita)	19	31	11	14	1373	179

Table A6. Control variables and dependent variables, descriptive statistics.

	Mean		Std. Dev.		Observations	
	England	Wales	England	Wales	England	Wales
Tax requirement (real £ per capita)	300	244	71	55	1608	242
Effective council tax rate (real £ per dwelling)	1087	796	197	165	1608	242
e^{IN}	0.77	0.85	0.14	0.10	1373	179
e^{OUT}	0.94	0.96	0.04	0.03	1373	179
e^{IN}(after bootstrap)	0.69	0.76	0.12	0.09	887	102
e^{OUT}(after bootstrap)	0.92	0.95	0.04	0.03	841	94
% age 0 - 16	22.40	22.60	1.75	1.11	1369	179
% age over 75	3.21	8.47	0.65	1.18	1369	179
% religious	77.70	73.82	4.50	4.17	1369	179
% white	89.12	98.34	12.80	1.75	1369	179
% tenure (house ownership)	66.84	70.98	11.29	3.82	1369	179
Band D equivalent dwelling (% per capita)	33.85	35.39	5.23	4.30	1369	179
Population density (persons per hectare)	24.38	4.20	26.84	4.84	1369	179
Conservative dummy (majority of seats)	0.23	0.04	0.42	0.20	1369	179
Labour dummy (majority of seats)	0.42	0.53	0.50	0.50	1369	179
Lib. Dem. dummy (majority of seats)	0.07	0.00	0.26	0.00	1369	179
No overall control dummy	0.28	0.43	0.45	0.50	1369	179
Election by thirds dummy	0.37	0	0.48	0	1369	179
Disposable income (real £ per capita)	12818	11025	3024	1087	1369	179
% firms in the financial sector	29.67	17.81	8.89	6.52	1369	179
% of unemployment related benefit	3.09	2.95	1.62	0.99	1369	179
% attendance allowance below age 65	4.74	8.25	1.95	2.54	1369	179
% high qualified workforce	5.12	3.51	2.32	1.18	1369	179
% self employed work force	7.93	7.98	2.22	3.40	1369	179

Table A7. Treatment effect of CPA on council tax measures, collapse to pre and post reform periods.

Model	Tax requirement (real £ per capita) (A)	Tax requirement (% of budget requirement) (B)	Effective council tax rate (real £ per band D equivalent dwelling) (C)
FE (linear)	15.47*	2.23***	29.87*
	(8.79)	(0.84)	(17.14)
RE (linear)	23.60***	3.38***	39.18**
	(8.94)	(0.88)	(17.46)
GLM (non linear) (1)	n.a.	4.28***	n.a.
		(1.22)	
Observations	340	340	340
Number of councils	170	170	170
Control variables	yes	yes	yes

Clustered standard errors in brackets. *** significant at 1%. Coefficient point estimates are interpreted as follows: £ per capita in column (A), % change in column (B), and £ per dwelling in column (C).

(1) Point estimates are in terms of average partial effect.

Table A8. Point estimates of the treatment effect of CPA on service quality, collapse to pre and post reform periods.

Model	Output measures					
	Aggregated output	Aggregated output (no education)	Education	Social service	Central services	Environment
FE (linear)	5.34***	1.44*	9.01***	2.02*	2.86	-7.04***
	(0.66)	(0.83)	(0.97)	(1.91)	(3.20)	(1.16)
RE (linear)	5.66***	1.35*	8.56***	2.02*	4.52	-6.54***
	(0.65)	(0.82)	(0.95)	(1.22)	(3.18)	(1.29)
GLM	5.67***	1.53**	9.14***	1.79**	2.60	-6.05***
(non linear) (1)	(0.46)	(0.64)	(0.69)	(0.76)	(2.03)	(0.87)
Observations	340	340	340	340	340	340
No. of councils	170	170	170	170	170	170
Control variables	yes	yes	yes	yes	yes	yes

*** significant at 1%; ** significant at 5%; * significant at 10%.

Coefficient point estimates are interpreted as percentage change in output index due to CPA.

(1) Point estimates are in terms of average partial effect.

Table A9.Treatment effect of CPA on efficiency, collapse pre and post reform periods (only raw DEA efficiency indices).

Model	Homogeneous effect (1)		Heterogeneous effect % incapacity benefit and jobseeker's allowance (2)		Heterogeneous effect vote gap (2)	
	Input app.	Output app.	Input app.	Output app.	Input app.	Output app.
FE (linear) (1)	-1.47	1.15	-0.064	-0.013	0.016	0.002
	(2.15)	(1.22)	(0.067)	(0.024)	(0.013)	(0.007)
RE (linear) (1)	-2.16	1.15	-0.102*	-0.052**	0.001	0.001
	(2.12)	(1.03)	(0.061)	(0.019)	(0.015)	(0.004)
Observations	340	340	340	340	340	340
No. of councils	170	170	170	170	170	170
Control variables	yes	yes	yes	yes	yes	yes

*** significant at 1%; ** significant at 5%; * significant at 10%.

(1) Ppercentage change in efficiency index due to CPA.

(2) Percentage change in efficiency after 1% change in the variable interacted with treatment effect.

A.4. Appendix D: Testing the Common Trend Assumption

The fundamental identifying assumption underlying the validity of the quasi-experimental setting is that the variable of interest should follow the same time path in control and the treated group in the absence of the treatment. Figures A.1, A.2, and A.3 show that the effective council tax rate, the aggregate output, and the efficiency indices where following a similar path in England and Wales before the introduction of CPA. We test for this hypothesis more formally by running, for the pre-treatment period from 1997 to 2000, the regression

$$Y_{it} = \eta_t + \theta_t(\eta_t \times D_i) + \alpha' \mathbf{X}_{it} + u_i + v_{it} \tag{A.11}$$

In (A.11) Y_{it} is the variable of interest, η_t is the set of year dummies, D_i is a dummy for English councils, and θ_t is the parameter of interest. So, given that CPA started in 2001, the hypothesis that the variable of interest follows the same time path is simply $H_0 : \theta_{97}, \theta_{98}, \theta_{99}, \theta_{00} = 0$. As reported in the following Table A.10 the null hypothesis of zero interaction can not be rejected in most of our tests. As reported in the table, p-values were below the critical threshold of the 10% significance level only for the output variables related to social services and the environment sector, and for the raw index of efficiency

in case of output approach.

Table A.10. P-values related to the null hypothesis $H_0 : \theta_{97}, \theta_{98}, \theta_{99}, \theta_{00} = 0$.

Variables	p-value*	DiD test
Aggregate output	0.63	ok
Education (BVPI38)	0.83	ok
Social services (BVPI49)	0.01	no
Social services (BVPI54)	0.02	no
Environment (BVPI82a)	0.01	no
Central services (BVPI8)	0.32	ok
Tax requirement (real £ per capita)	0.36	ok
Tax requirement (% of budget requirement)	0.93	ok
Effective council tax rate (real £ per dwelling)	0.62	ok
e^{IN}	0.72	ok
e^{OUT}	0.06	no
e^{IN} + bootstrap procedure	0.15	ok
e^{OUT} + bootstrap procedure	0.27	ok

*Probability of rejecting the null hypothesis of similar time path between England and Wales in the pre-treatment period when the null is true.

Figure A.1: Effective Council tax rate.

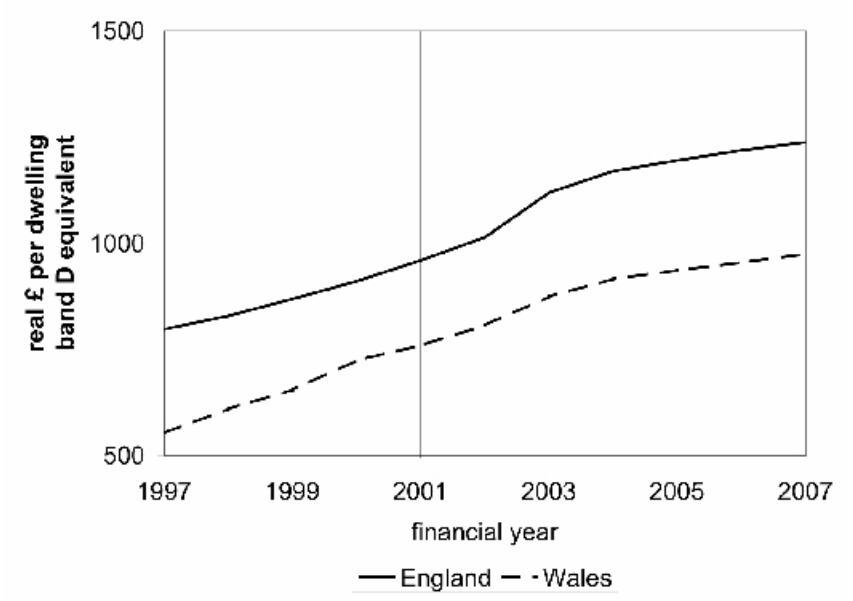


Figure A.2: Aggregate output.

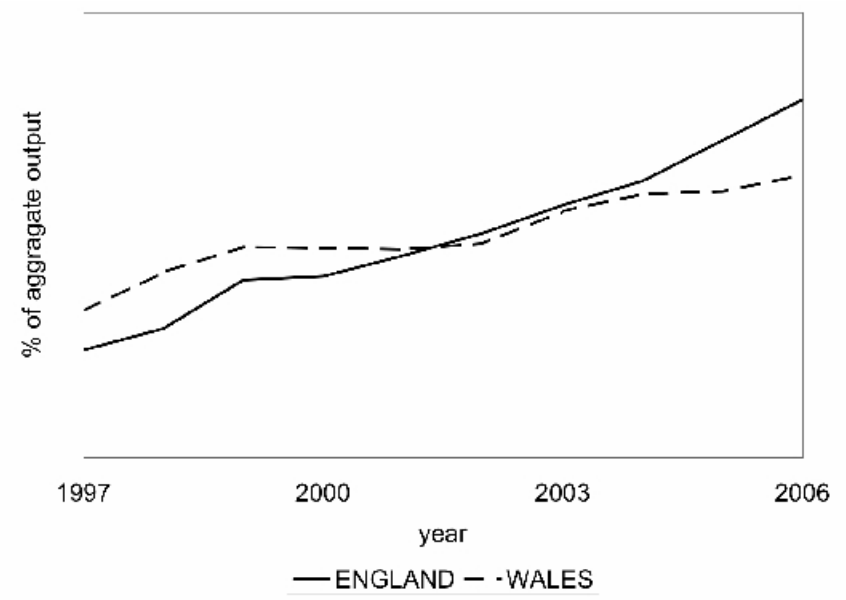


Figure A.3: Efficiency, raw DEA indices,
average between input and output approach.